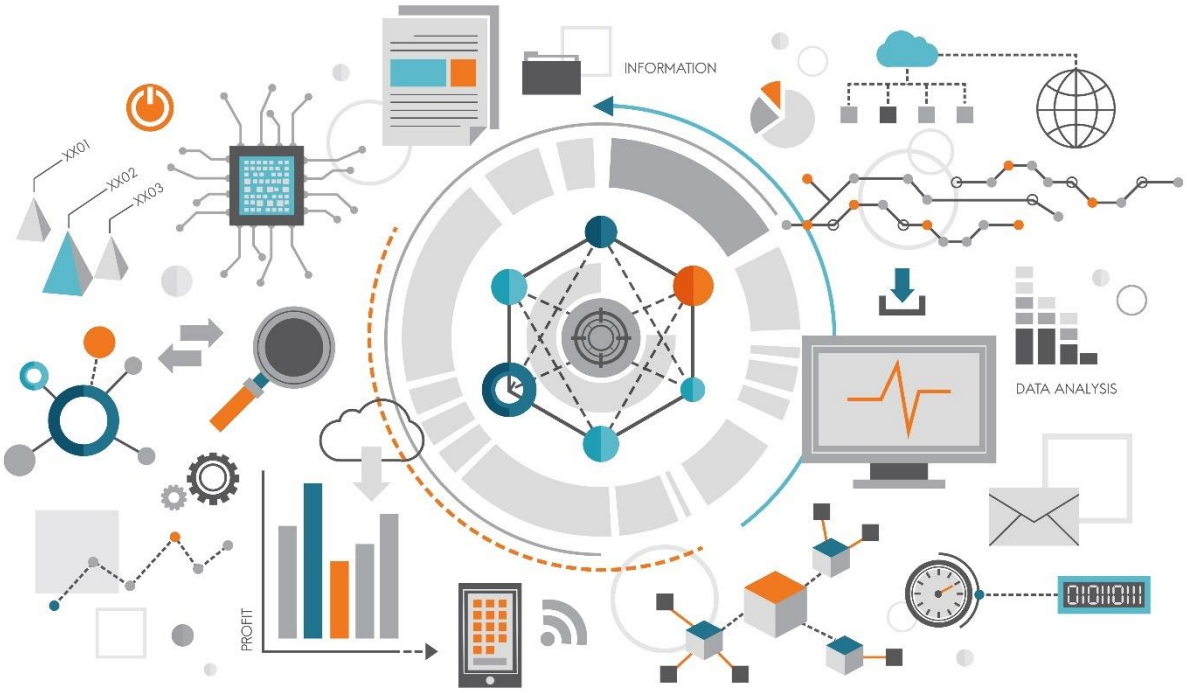
IBEC

Institute for Bioengineering of Catalonia

Guide: Managing research data in IBEC research groups

Draft 1 / July 2022



Performance icon vector created by rawpixel.com - www.freepik.com

Author: Fidel Bellmunt Montagut
(Knowledge Manager at IBEC)

July 2022

Contents

Executive summary

This guide pretends to be a helpful resource to IBEC researchers to manage their research data to achieve better ways of working and to comply with IBEC's policy and funders requirements.

Data management implies several issues and agents. This guide wants to give a general overview of each issue and to help researchers to find the best way of implementing data management at each laboratory based on its specific characteristics.

Further help and collaboration from the Knowledge Manager at IBEC, to implement data management is available, and can be contacted at datamanagement@ibecbarcelona.eu

The guide is divided in different parts, that can be used individually (you may jump there from here):

· **Introduction and definitions**
There are general explanations and definitions about research data management, to fully understand it.

· **Research data management overview**
Shows the general overview of the issues developed at the specific worksheets.

· **Research data management worksheets: recommendations and examples**
Includes 6 sections (each box is a link to the worksheet):

1. Data needs for research To determine data needs at research.	**2. Responsibilities and tasks** To organize the team on data management.	**3. Organization, formats, and description of data** To organize data itself during research.
4. Storage, backup and access To manage preservation and access to data.	**5. Control and publishing datasets, open or closed** To decide how to assure long term preservation of data and if it will be open.	**6. Data management costs** To determine project costs related to data management.

Introduction and definitions

As stated in IBEC's code of conduct for research integrity and deployed in IBEC's Industrial and Intellectual Property Regulations, all data arising from research carried out in IBEC or by people linked to IBEC is the property of IBEC and has to be at the disposal of IBEC staff if required. IBEC has the right to choose how to protect, publish and share the data, unless otherwise stated in agreements with third parties, and should be informed in advance.

Why research data management is necessary?

- Excellence: managing research data properly allows to work more efficiently, improving research quality and productivity. Saves time and resources in the long run. Good management helps to prevent errors, increases the quality of the analyses, and avoids data loss.

- To comply with IBECs policies: At IBEC the *Code of conduct for research integrity* (https://ibecbarcelona.eu/wp-content/uploads/2021/06/2021-IBEC-Code-of-conduct-for-research_web.pdf) must be known and followed by all his researchers. The Code establishes the basis for data management principles at the centre, which are further developed at the *Research Data Management Policy* (https://ibecbarcelona.eu/wp-content/uploads/2021/11/IBEC-Data-Management-Policy.pdf).

- To comply with funders' mandates, especially the European Commission, which are increasingly demanding data management and to open them when possible.

- Open Science: Having well-managed research data allows to make decisions easily and reasonably about how and where they can be shared or why not, making research results more understandable and enabling reuse, allowing others to validate and replicate findings.

Research data is a set of information, digitalised in files or on any physical medium, which in the research process contributes to the production of a scientific result. According to the CODATA-CASRAI (2022), which proposes the establishment of standards to describe research, this type of data is defined as the ***evidence in a research process that validates its findings and results***.

Infographic by Science Nature:
https://researchdata.springernature.com/documents/web_a92645_what-are-research-data-revision

Research Data Management (RDM) comprises the decisions regarding activities that have to do with the life cycle of research data. In other words, the collection, organisation, processing, analysis, preservation, and publication of the data used in a research project.

Managing data properly, could allow to:

- Have more transparency to validate search results.
- Ensure that data is findable, accessible, interoperable , and reusable.
- Improve the profile of the researcher, the impact and visibility of the projects.
- Encourage data protection and minimize the risk of data loss.
- Comply with the legislation on the subject and the requirements of the financing entities.
- Save time and make efficient use of available resources.
- Ensure research integrity and replication.
- Ensure research data and records are accurate, complete, authentic and reliable.

What will be necessary to determine?

- the handling of research data during & after the end of the project;
- what data will be collected , processed and / or generated;
- which methodology & standards will be applied;
- whether data will be shared / made open access and;
- how data will be curated & preserved (including after the end of the project).

When designing research, first is identified what type of data is needed and then are analysed: the reuse of existing data from the most appropriate source, or the generation of own data. Once this decision has been made, it is necessary to consider what formats will be used, how the data will be organized, where and how it will be stored, and finally whether it will be published. If all this is done following the **FAIR principles** (for Findable, Accessible, Interoperable, Reusable), it increases the quality of management and enables the consideration of data sharing without additional efforts at the end of the projects.

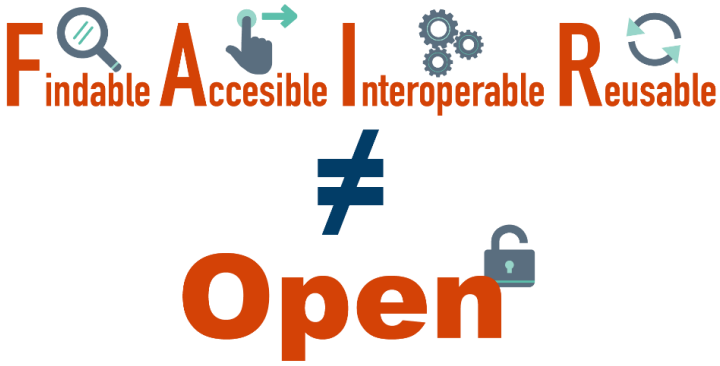
FAIR principles are a set of guidelines that must be followed so that research data is accessible, understandable and reusable, also in studies in other disciplines. In short, the FAIR principles help to standardize and improve the management of research data. Data is considered FAIR when it is cured according to these principles.

To ensure that the institution's data complies with the FAIR principles at the highest possible level, it is necessary to define:

- What are the most appropriate formats to work with and preserve data?
- Which repository or infrastructure should be used?
- What metadata should be used to describe the data?
- What permanent identifiers will be used when publishing the data?

To achieve all this, it is necessary to have: tools, support staff, methodology and training.

The concepts of *data managed according to the FAIR principles* and *open data* are not equivalent or exclusive, but complementary. FAIR principles do not guarantee that data is open, just as open data is not necessarily curated data under FAIR principles. The level of accessibility is what determines whether the data is open or restricted to certain users. The degree of compliance with these principles is what determines whether the data is FAIR.



Credit: vector graphics in the image are licensed for free use by rawpixel.com / Freepik

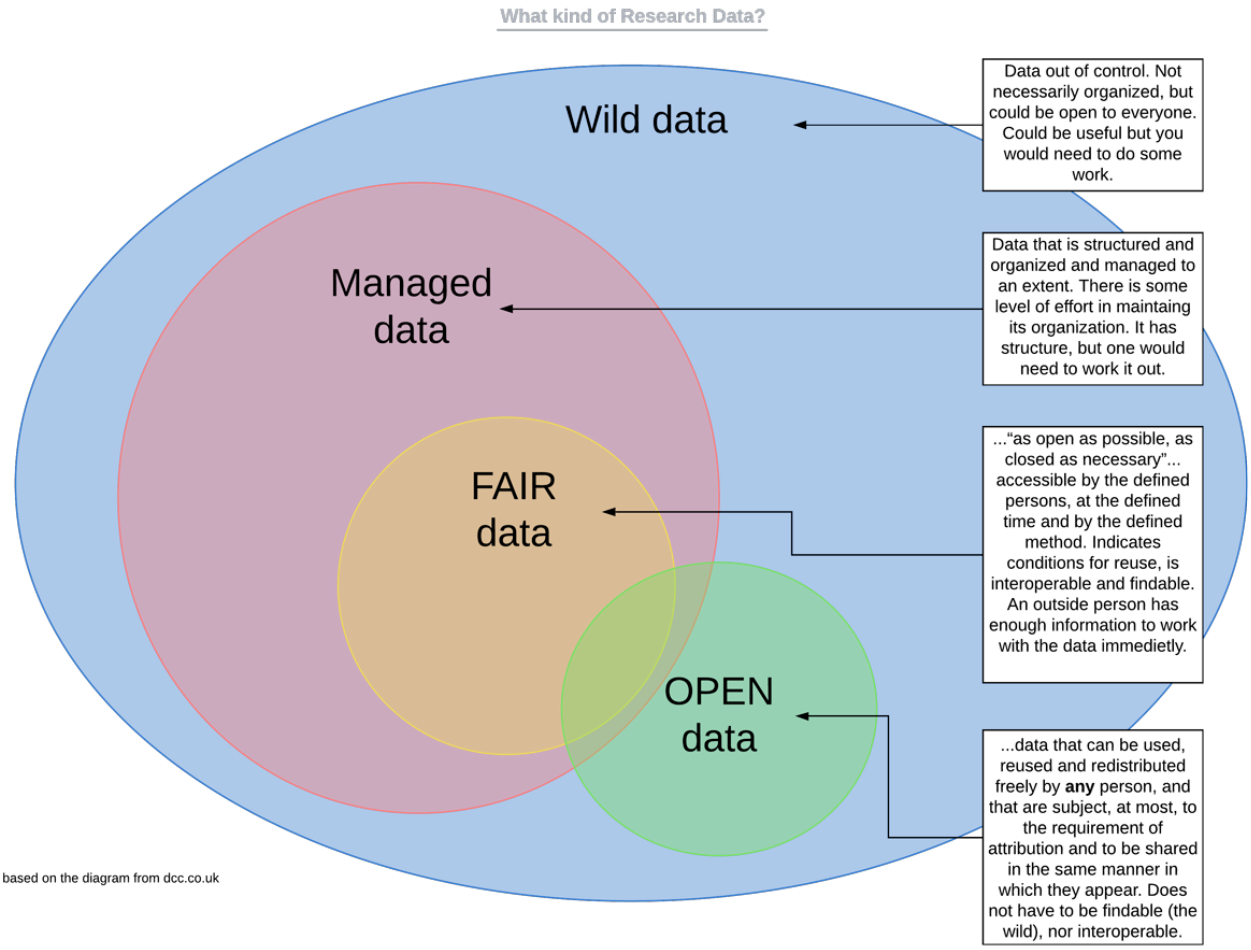
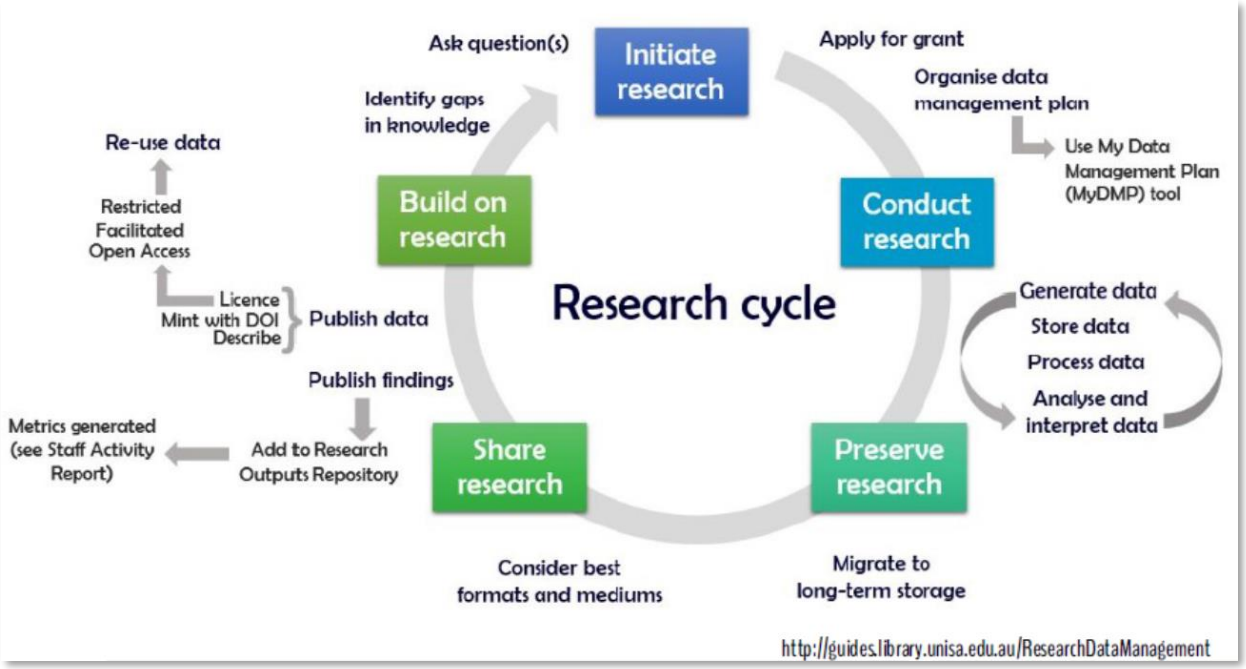


Diagram at http://www.digitalservices.lib.uct.ac.za/dls/documentation/what-data , based on: All Research Data – Sarah Jones (DCC)

The document that shapes the planning of the research data management of each project is the data management plan (DMP), and it is increasingly a mandatory deliverable for many of research funding entities and an obligation at IBEC according to its own policy on data management, in line with the Open Science paradigm.

Research life cycle and actions

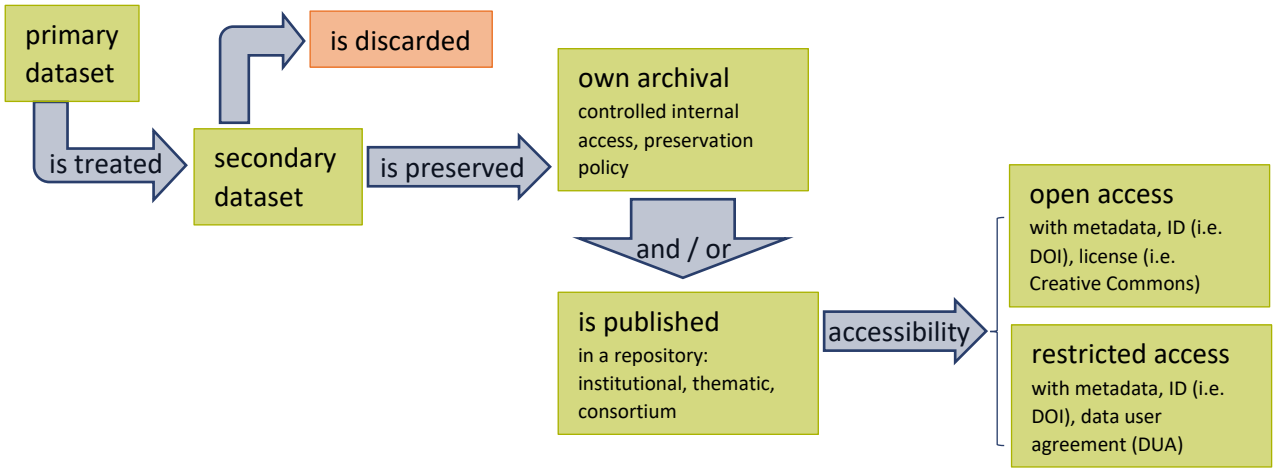
During research cycle there are several moments related with data that must be addressed with the proper actions and planification:



Infographic adapted by: Ignasi Labastida, Universitat de Barcelona

Basic scheme of research data preservation

Decisions related to data generated during research, regarding to it's preservation:



Research data management overview

All these considerations should be addressed at the beginning of the projects, planning all the issues described below, that may change during the development of research. Some are structural decisions for the research group and may be the way to work in all projects. All these decisions will be summarised at the **Data Management Plans** (obligation at IBEC), following the proper template depending on the funders of the project, and the Knowledge Manager together with the Projects Office may assist you on doing it. At the Annex 2 there is a template of a DMP of the H2020 and an example.

1. Data needs for research

Which data will be generated or reused from existing data: see availability (own data or from others) and the licenses that determine its use.

2. Responsibilities and tasks

Define who will oversee what in relation to the research data within the group, and other supports at IBEC.

3. Organization of data and format

- Set data organization methodology, file naming and version assignment.
- Define action guidelines for data management. Disseminate, train and raise awareness among the staff of the centre about the importance of monitoring.
- Use standard formats (non-proprietary).

4. Storage and access: proper equipment, access control (permissions, profiles), backups.

- Determine which devices will be used to work with the data.
- Have a workspace where you can store data with access control, which allows you to set permissions and determine action roles to work with the data.
- Define a preservation system, with data backups.
- Select possible alternative solutions for storing data and provide research staff with information on how to select the most appropriate one in each case.

5. Publishing datasets (open or closed, complying with FAIR principles)

- Deposit datasets at repositories, using a unique identifier.
- Use a standard metadata schema.
- Allow the use with licenses for data
- Closed data (restricted access): Report the reason for the restriction (personal data, confidential data, patents, etc.) and how they can be reused. Possibility of anonymization, etc.

6. Data management costs

Identify all the processes related to the data management of the centre, and estimate the cost of data centre management:

- Hours spent by research staff and expert staff in charge of guiding and supervising the work
- Infrastructure costs and associated services

Research data management worksheets: recommendations and examples

1. Data needs for research

a. Determine what data is expected to be collected or generated during research and explain their basic characteristics.

> Kind of data (images, video, audio, text, etc.) and obtention methods.

> Formats: depending on the tools used (microscopes, etc).

> Size estimation: expected volume of each dataset.

> Storage needs (during research).

> Protection issues: determine if the collected data must be protected for intellectual property issues or sensitive data issues.

> Another considerations.

b. Check if there is already data created and available to use for this research.

b1. Firstly, the own data from previous research (that should be findable and reusable if it was properly managed).

b2. Then, from the scientific community globally. There are several databases of datasets, some are thematic, and others are transversal. Re3data is a global registry of research data repositories that covers research data repositories from different academic disciplines: https://www.re3data.org/

When reusing data from others, openly available or requested to the authors, must be known and respected the licenses that determine how that data can be reused.

2. Responsibilities and tasks

Define who will oversee what in relation to research data within the group or project, and from other support units at IBEC.

At the lab (possible roles, depending on each project needs and lab organization)	> Overall **coordination**: usually the PI of the research. > **Data controller/manager**: they look for the correct use of the policies and criteria set at the research group, guiding research fellows to accomplish them and controlling the processes related with data (naming conventions, access, backups, etc.). > **Data analyst** or **Data scientist** within the group (may be the same person as the Data manager). They establish the guidelines on how to collect, process and prepare the data for analysis. Their RDM responsibilities range from collaborating with other researchers in data collection and analysis, conducting technology watch, creating information analysis and visualisation strategies, identifying problems and providing data analysis solutions, using statistical techniques and data mining models for these purposes, evaluating data collection sources and techniques, staying up to date regarding technologies, techniques, and methods, etc.
At other IBEC units	> Data Management and Advice: **Knowledge Manager**. Coordination and support at IBECs level in the elaboration of the DMPs , in data curation and in the process to select and to deposit data in repositories, datamanagement@ibecbarcelona.eu > Sensible data: **Data Officer**. Validation and monitoring of the management of sensitive data, which may be protected, dataprotection@ibecbarcelona.eu > Intellectual property issues: **Technology Transfer Office**, techtransfer@ibecbarcelona.eu > Digital storage and computer security: **IT**, it@ibecbarcelona.eu

3. Organization, formats, and description of data

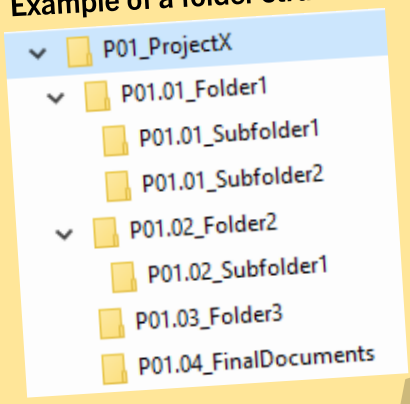
3.1. Organization of data

Establish a way of working the data in the group with a **procedure** or protocol of data management, choose **tools** to do so, and set **criteria**. Research data must be stored in a correct, complete, unadulterated, and reliable manner.

Procedure	Establish the way data must be generated, stored, and curated during research in the lab. Consider if there must be a workflow that determine how to handle data at every stage of research.
Tools	**Laboratory notebooks**
	The use of laboratory notebooks as a primary research record and as an organizing tool, allowing to document and then trace the steps, methods, observations, and conclusions drawn during research in the laboratory. *Types of lab notebooks:* > **Paper:** establish how to use them and how to keep them safe (avoiding loss, theft, preventing being damaged or destroyed, etc.). > **Digital:** using digital lab notebooks (Electronic Lab Notebook: ELN), allows the possibility to integrate and link the record of the work of the researchers with the results (datasets) that are being obtained. **OneNote** the ELN available at IBEC: It is integrated within the Windows cloud system (OneDrive). It is necessary to plan how to apply the use to the daily life of the laboratory in a practical way (tablets or similar devices), and to adapt it for a useful work. Some IBEC groups are already using it. Quick guide for using it: Guerrero S, López-Cortés A, García- Cárdenas JM, Saa P, Indacochea A, et al. (2019) *A quick guide for using Microsoft OneNote as an electronic laboratory notebook* . PLOS Computational Biology 15 (5): e1006918. https://doi.org/10.1371/journal.pcbi.1006918 More complete manual: Montgomery, I (Babraham Bioinformatics). *Using OneNote as an Electronic Laboratory Notebook* (2020). https://www.bioinformatics.babraham.ac.uk/training/OneNote%20manual.pdf
	Databases for data management and analysis
	Databases to collect and analyse data are common tools too. There are several tools and solutions available, form the simple spreadsheet manager to online tools for collaborative works. For example: - **REDCap** (Research Electronic Data Capture): is a secure web cloud-based application for building and managing online surveys and databases. https://www.project-redcap.org/

Criteria	**Folder and File Organization** Decide what to keep, where, when and how during research. Having a clear and well-structured folder and file system, helps to achieve a successful data management. Following best practices for managing your research data can ensure it will be available to other researchers in the long term and help prevent data loss. *Folder structure:* Determine the tree levels and naming criteria for the folders. Choose a consistent organizational structure for all your project folders. Although it may seem obvious, thinking about the structure of your folders and planning effectively makes navigation much easier. Minimize the number of clicks necessary to reach files. In conjunction with a consistent file naming convention, an efficient structure saves a lot of time. A hierarchical structure is a very common model for file organization, shared by most operating systems (i.e., Windows, Mac, etc.). Folders are nested within subfolders. The hierarchy is much like a traditional outline, and it can be helpful to sketch out your hierarchy before creating it. Guidelines: - Be consistent. - Structure your hierarchy logically. Follow the logic that makes the most sense for your project. - Keep folders and subfolders separate to reduce overlap. However, don't make an excessive number of subfolders. - Keep subfolder categories narrow to restrict the number of files in each. - Your Desktop is meant to be temporary storage. Never keep files there for longer than necessary. - When naming folders, think about information you might want when looking up files, and use numerals to keep the desired order (the system sorts automatically by names). - Consider a final project folder with all the relevant final documents and data, once closed. *File naming:* - Keep file names short but meaningful. - Avoid special characters except when swearing, e.g. “£$%!”¬&*^()+=[]{}~@:;#,.<> (These may be interpreted incorrectly by computers, especially if used in different operating systems or by scripts). - Reverse dates so filenames sort usefully YYYYMMDD e.g. 20200131. - Consider whether the folder will always be there to provide more context or not. - Ensure document version control. Use sequential number system e.g., FileName_Date_v1, _v2, _v3 - Avoid using ‘final’ in the filename. Inevitably it is wishful thinking. - Use CamelCase-hyphens-or underscores_ but be consistent. - Change spaces in filenames to underscores_ or else %20 may replace it when it goes online.

Example of a folder structure:



Example of a file or folder name as an audit trail:

File name:
20220713_db_starman_exp05_gel_004
Parts:
- date: 20220713
- creator initials: db (for David Bowie)
- project ID: starman
- experiment ID: exp05
- experiment methodology: gel (for gel chromatography)
- number in the sequence of files generated during the experiment: 004

3.2. Formats of data

The specific formats how data is saved determine the use and further reuse of it.

Better use:

- Common formats (may be de facto standards)
- Standard formats accepted in your field
- Interchangeable or open (published) formats for long-term preservation
- Recommended formats: see the Table of recommended formats at Annex 1

Avoid:

Dependency on proprietary software to render your data

Consider data format conversions, to:

- Save storage space: some machines for data collection (microscopes, etc.), produce data with some associated information that can be erased depending on the intended use. Scripts or other automatic tools may be set to clean them and make them less huge (specially in videos, etc): **check with IT for secure processes**.
- Make it interoperable.
- Long term preservation.

3.3. Description of data

To understand how data was collected, saved, and used, it must be described and documented.

Data description: use of metadata allowing interoperability

Machine-readable, standardised fields that allow discovery through search engines, or mark up the structure of a database, or show relationships between different digital objects.
There are several metadata sets and vocabularies. The RDA Metadata Standards Catalog helps to look for the existing one at each field: https://rdamsc.bath.ac.uk/

Data repositories are structured following a metadata set to describe datasets.

Usual metadata sets: *Dublin Core*, *MINSEQE* (MINimal information about high throughput SEQeuencing Experiments), *MIBBI* (Minimum Information for Biological and Biomedical Investigations)
Format: often *XML* or *JSON*

Example: Dublin Core Metadata Element Set

Contributor – "An entity responsible for making contributions to the resource".
Coverage – "The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant".
Creator – "An entity primarily responsible for making the resource".
Date – "A point or period of time associated with an event in the lifecycle of the resource".
Description – "An account of the resource".
Format – "The file format, physical medium, or dimensions of the resource".
Identifier – "An unambiguous reference to the resource within a given context".
Language – "A language of the resource".
Publisher – "An entity responsible for making the resource available".
Relation – "A related resource".
Rights – "Information about rights held in and over the resource".
Source – "A related resource from which the described resource is derived".
Subject – "The topic of the resource".
Title – "A name given to the resource".
Type – "The nature or genre of the resource".

Example of a Dublin Core Metadata set encoded in XML:

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description>
    <dc:title>Homage to Catalonia,</dc:title>
    <dc:creator>Orwell, George, 1903-1950.</dc:creator>
    <dc:type>text</dc:type>
    <dc:publisher>London, Secker and Warburg</dc:publisher>
    <dc:date>[1938]</dc:date>
    <dc:language>eng</dc:language>
  </rdf:Description>
</rdf:RDF>
```

Documentation of data
The human-readable stuff that contextualises research outputs and processes so your future self or others can understand how you got your findings and/or how the data can be repurposed. Kinds of documents:

- Data dictionary
- Readme file (a template is available at Annex 3.)
- Study protocol
- Methodology statement
- Sampling frame description

Format: often text or PDF

Example: fragment of a Readme file (full at: https://doi.org/10.34810/data154/1)

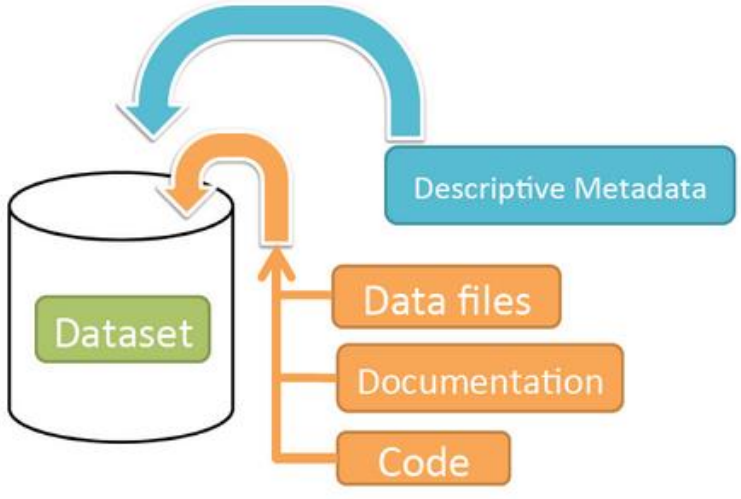
Readmefile_MedAid_WP2_IRTA_AE_Sea_Cages
Generated on 2021-11-22

****
PROJECT INFORMATION
****
1. Title of dataset: Seabream cages in Greece and Spain
2. Author information

Principal Investigator
Name: Alicia Estevez
Address: IRTA-Spain
Email: alicia.estevez@irta.es
ORCID: 0000-0002-7776-0521

3. Data of data collection 8/6/2018 to 31/10/2019

Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, and code.

https://guides.dataverse.org/en/latest/user/dataset-management.html

4. Storage, backup and access

Research data should be stored as soon as it is obtained in a restricted access space available to the researchers involved (local/consortia).

At IBEC, the minimum archive time for research data and records is 10 years after either the assignment of a persistent identifier or publication of a related work following project completion, whichever is later.

If research data and records are to be deleted or destroyed, either after expiration of the required archive duration or for legal or ethical reasons, such action will be carried out only after considering all legal and ethical perspectives and with prior approval of IBEC. The interests and contractual stipulations of third-party funders and other stakeholders, employees, and partner participants in particular, as well as the aspects of confidentiality and security, must be taken into consideration when decisions about retention and destruction are made. Any action taken must be documented and be accessible for possible future audit.

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

- Do you have sufficient storage, or will you need to include charges for additional services?
- How will the data be backed up?
- Who will be responsible for backup and recovery?
- How will the data be recovered in the event of an incident?

How will you manage access and security?

- What are the risks to data security and how will these be managed?
- How will you control access to keep the data secure?
- How will you ensure that collaborators can access your data securely?
- If creating or collecting data in the field, how will you ensure its safe transfer into your main secured systems?

Determine which devices will be used to work with the data, and together with IT set up the informatic system needed to run research.

Required IT infrastructure (under IT supervision):

Disk drives

Group servers

- Define the use and guidelines to follow (what is stored and how).
- Dimension space needs and make predictions.
- Define the use and guidelines to follow (what is stored and how).
- Set an access policy defining access levels and profiles. Provide an arrival and departure protocol for the group.
- Set a final results folder, define what it should contain and drop-down criteria> include supporting documentation or descriptive data (.txt , or similar)
- Criteria for identifying versions and naming files.
- Short, medium and long term backup and preservation system (consider migrations).

Personal Hard Drives (laptops, PCs, external hard disks -usb, etc.-)

Define the permitted use and group guidelines when using local storage. When data stored there, must be copied at the shared devices.

5. Control and publishing datasets, open or closed

Since all data arising from research carried out in IBEC or by people linked to IBEC is the property of IBEC and must be at the disposal of IBEC staff if required, a coordinated register of it is being set. So please **inform the Knowledge Manager for internal control of new datasets** as an output of your research.

Datasets must be individually identified and be findable. **Internally** we can guarantee that by the correct management at each lab and with the new centralised control system. **Externally**, publishing them in the journal's online platforms with the articles which they relate with, It isn't enough to make them FAIR. Posting them at data repositories (either open or closed) gives them an internationally standardised ID and makes them findable, since they are described in normalised schemas of metadata. Consider also if your dataset have reuse potential, and if so, is it reusable?

When open data access is not possible

Before sharing Research Data during or after a project, it is essential to consider whether this is permissible considering IPR ownership, ethical, privacy, confidentiality requirements or any legal, regulatory or funding restrictions. In addition, Researchers must consider in consultation with IBEC's Tech Transfer Unit whether Research Data has commercial potential and if it is suitable for protection and/or transfer under IBEC's Industrial and Intellectual Property Regulations.

- Ethics and security: where data access must be restricted for ethical or security reasons.
- Data protection: where human data cannot be de-identified, so data cannot be shared to protect patient/participant privacy.
- Large data: where data is too large to be feasibly hosted by a recommended repository.
- Third party data: where data has been obtained by a third party, and restrictions apply to the availability of the dataset.
- Industrial interest: closed data to protect patenting, commercialising, etc. Always check with Tech Transfer before deciding to open data.

In any of these cases, datasets can be described at repositories with the explanation of why they are closed.

Choosing where to publish

Data repositories: Most repositories will expect data to be deposited in preferred preservation formats (to enable reuse also in the long term) and for accompanying high-quality documentation to enable correct use of the data. Good repositories will assign a unique permanent identifier, display a clear reuse license and data citation format. The general recommendation is to publish at an specific discipline repository, and if it doesn't exist, then choose the institutional or general purpose ones.

➔ **IBECs institutional repository**: CSUC Data repository: **CORA. Repositori de Dades de Recerca** https://dataverse.csuc.cat/. All CERCA centres are committed to publish in this repository. The Knowledge Manager will help you on publishing in it under the institutional profile.

➔ Discipline-specific repositories can be found via the registry http://re3data.org , contains over 2500 research data repositories. Examples on genomics:
  - **GEO (Gene Expression Omnibus)**, for genomics: http://www.ncbi.nlm.nih.gov/geo/
  - **EGA (European Genome-phenome Archive)**: https://ega-archive.org/

➔ Trusted cross-disciplinary repositories:
  - **Zenodo**, the EU repository: https://zenodo.org/
  - **GitHub**, for software code: https://github.com/

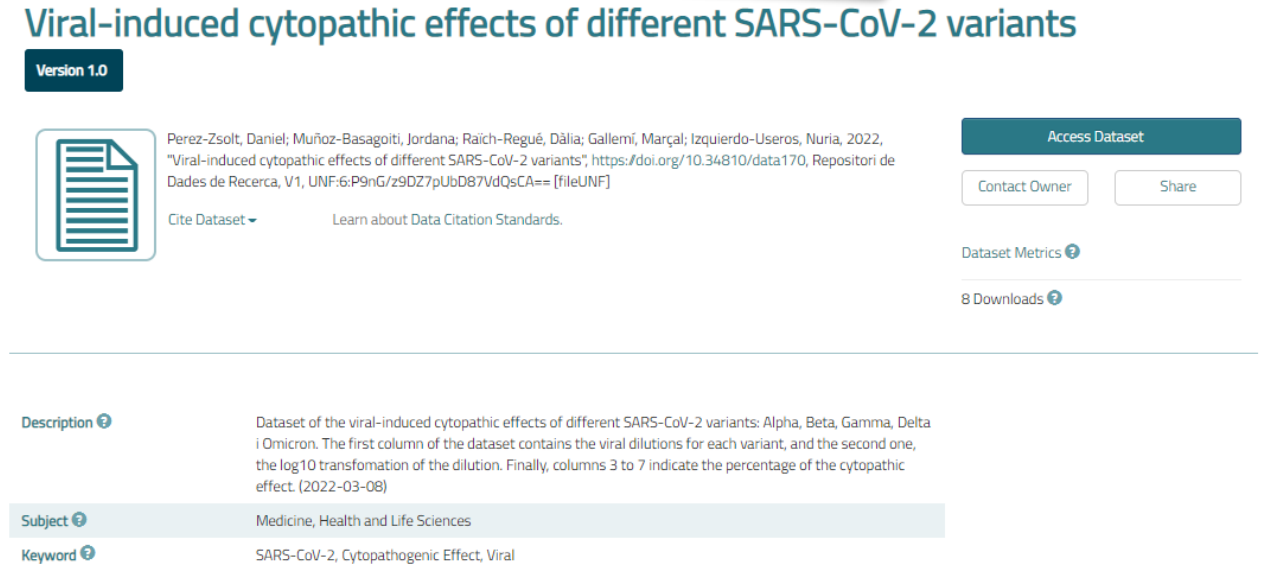
Licenses and copyright

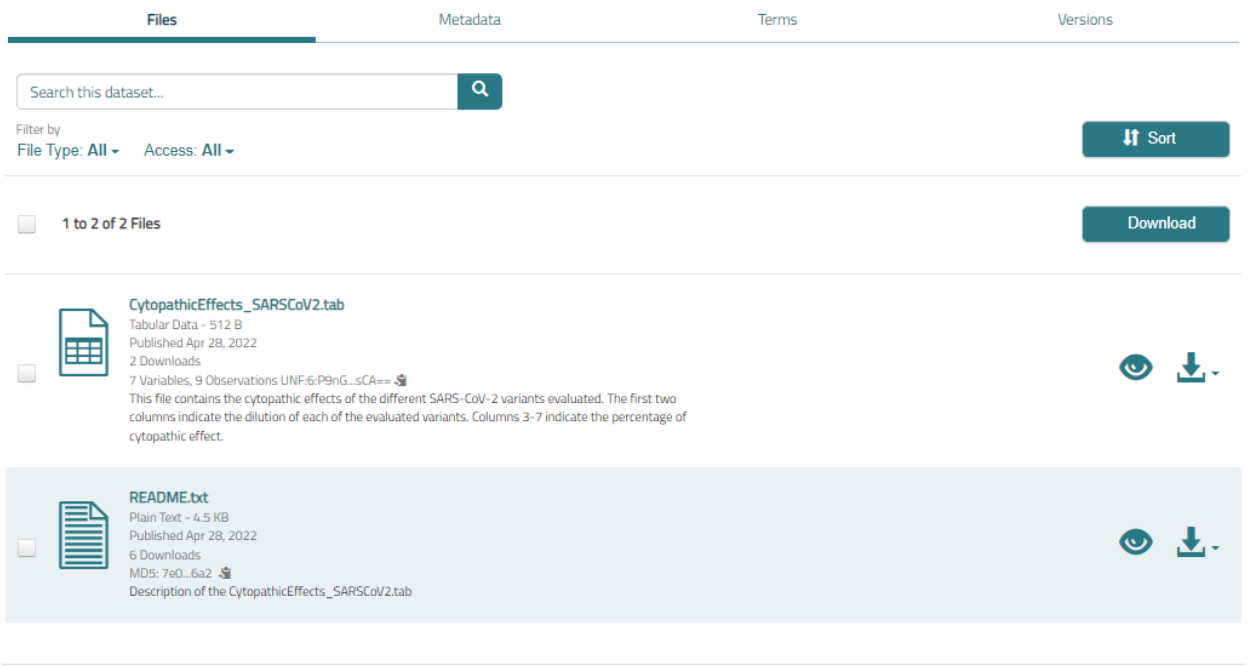
At IBEC and in compliance with intellectual property rights, and if no third-party rights, legal requirements, or property laws prohibit it, research data should be assigned a licence for open use such as **CC0 or CC-BY from Creative Commons**.

Allow the use with licenses for data checking these possibilities:
- Open Data Commons (ODC): https://opendatacommons.org/
- Creative Commons (CC): https://creativecommons.org/ BY, BY SA, BY ND, BY NC, BY NC SA, BY ND SA.

Closed data (restricted access): Report the reason for the restriction (personal data, confidential data, patents, etc.) and how they can be reused. Consider also the possibility of anonymization , etc.

Dataset Example at CORA Repositori de Dades de Recerca (CORA.RDR): https://doi.org/10.34810/data175

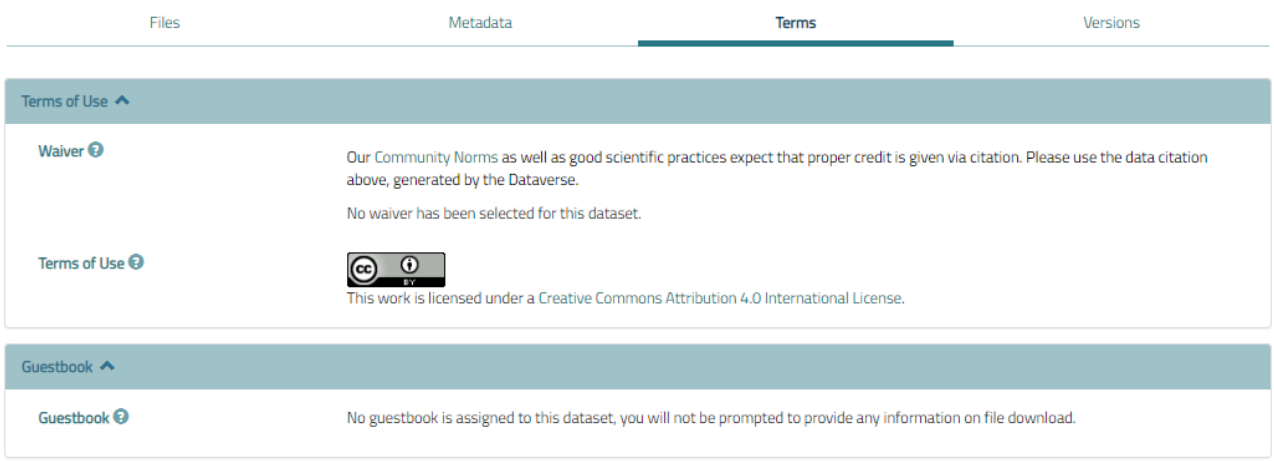




Citation Metadata

Dataset Persistent ID	doi:10.34810/data170
Publication Date	2022-04-28
Title	Viral-induced cytopathic effects of different SARS-CoV-2 variants
Author	Perez-Zsolt, Daniel (IrsiCaixa, AIDS Research Institute,) - ORCID: 0000-0003-4192-7622 Muñoz-Basagoiti, Jordana (IrsiCaixa, AIDS Research Institute,) - ORCID: 0000-0002-0384-5928 Raïch-Regué, Dàlia (IrsiCaixa, AIDS Research Institute,) - ORCID: 0000-0001-7656-5700 Gallemí, Marçal (IrsiCaixa, AIDS Research Institute,) - ORCID: 0000-0003-4675-6893 Izquierdo-Useros, Nuria (IrsiCaixa, AIDS Research Institute,) - ORCID: 0000-0002-1039-1821
Contact	Use email button above to contact. Izquierdo-Useros, Nuria (IrsiCaixa, AIDS Research Institute,) IrsiCaixa AIDS Research institute
Description	Dataset of the viral-induced cytopathic effects of different SARS-CoV-2 variants: Alpha, Beta, Gamma, Delta i Omicron. The first column of the dataset contains the viral dilutions for each variant, and the second one, the log10 transformation of the dilution. Finally, columns 3 to 7 indicate the percentage of the cytopathic effect. (2022-03-08)
Subject	Medicine, Health and Life Sciences
Keyword	SARS-CoV-2 (MeSH) https://www.ncbi.nlm.nih.gov/mesh/2052180 Cytopathogenic Effect, Viral (MeSH) https://www.ncbi.nlm.nih.gov/mesh/?term=cytopathic+effect%2C+viral
Topic Classification	Virology (MeSH) https://www.ncbi.nlm.nih.gov/mesh/68014773

Depositor	IrsiCaixa, AIDS Research Institute
Deposit Date	2022-03-08
Kind of Data	Experimental data



6. Data management costs

Identify all the processes related to the data management of the centre and estimate it's cost. There are increasingly more funding opportunities where these costs may be eligible.

- Hours spent by research staff and expert staff in charge of guiding and supervising the work
  - Data documentation: description of data sets, metadata, etc.
- Infrastructure usual costs and associated services
  - Database access
  - Data generation: equipment, appliances, software, …
  - Transcription, standardization, anonymization, or management of informed consent.
  - Stored data (own / external - repository -).

Annex 1: Table of recommended data formats

(File formats recommended by the UK Data Service: https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/)

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata. Variable labels, code labels, and defined missing values.	SPSS portable format (.por) Delimited text and command ('setup') file (SPSS, Stata, SAS, etc.). Structured text or mark-up file of metadata information, e.g. DDI XML file.	Proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb).
Tabular data with minimal metadata. Column headings, variable names.	Comma-separated values (.csv). Tab-delimited file (.tab). Delimited text with SQL data definition statements.	Delimited text (.txt) with characters not present in data used as delimiters. Widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods).
Geospatial data. Vector and raster data.	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional). Geo-referenced TIFF (.tif, .tfw). CAD data (.dwg). Tabular GIS attribute data. Geography Markup Language (.gml).	ESRI Geodatabase format (.mdb). MapInfo Interchange Format (.mif) for vector data. Keyhole Mark-up Language (.kml). Adobe Illustrator (.ai), CAD data (.dxf or .svg). Binary formats of GIS and CAD packages.
Textual data	Rich Text Format (.rtf). Plain text, ASCII (.txt). eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema.	Hypertext Mark-up Language (.html). Widely-used formats: MS Word (.doc/.docx). Some software-specific formats: NUD*IST, NVivo and ATLAS.ti.
Image data.	TIFF 6.0 uncompressed (.tif).	JPEG (.jpeg, .jpg, .jp2) if original created in this format. GIF (.gif). TIFF other versions (.tif, .tiff). RAW image format (.raw). Photoshop files (.psd). BMP (.bmp). PNG (.png). PDF/A, PDF (.pdf).
Audio data.	Free Lossless Audio Codec (FLAC) (.flac).	MPEG-1 Audio Layer 3 (.mp3) if original created in this format. Audio Interchange File Format (.aif). Waveform Audio Format (.wav).
Video data.	MPEG-4 (.mp4). OGG video (.ogv, .ogg). motion JPEG 2000 (.mj2).	AVCHD video (.avchd).
Documentation and scripts.	Rich Text Format (.rtf). PDF/UA, PDF/A or PDF (.pdf). XHTML or HTML (.xhtml, .htm). OpenDocument Text (.odt).	Plain text (.txt). Widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx). XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0.

Annex 2. Data Management Plan resources, template and example

Resources with templates or guides for DMP creation

- DMPOnline, from the UKs Digital Curation Centre (DCC) : https://dmponline.dcc.ac.uk/
  - Catalan version from CSUC: eiNa DMP: https://dmp.csuc.cat/

- ARGOS, from OpenAIRE and EUDAT: https://argos.openaire.eu/splash/

- CaixaHealth Model of the data management plan for research projects: https://fundacionlacaixa.org/en/caixaresearch-management-policy-open-access-research-data-management-model

Horizon 2020 FAIR Data Management Plan (DMP) Template

(https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx)

1. DATA SUMMARY

- What is the purpose of the data collection/generation and its relation to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data?
- To whom might it be useful ('data utility')?

2. FAIR DATA

Making data findable, including provisions for metadata

- Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?
- What naming conventions do you follow?
- Will search keywords be provided that optimize possibilities for re-use?
- Do you provide clear version numbers?
- What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Making data openly accessible

- Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.
- Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.
- How will the data be made accessible (e.g. by deposition in a repository)?

- 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)?
- Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.
- Have you explored appropriate arrangements with the identified repository?
- If there are restrictions on use, how will access be provided?
- Is there a need for a data access committee?
- Are there well described conditions for access (i.e. a machine readable license)?
- How will the identity of the person accessing the data be ascertained?

Making data interoperable

- Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?
- What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?
- Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?
- In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Increase data re-use (through clarifying licences)

- How will the data be licensed to permit the widest re-use possible?
- When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
- Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
- How long is it intended that the data remains re-usable?
- Are data quality assurance processes described?
- Further to the FAIR principles, DMPs should also address:

3. ALLOCATION OF RESOURCES

- What are the costs for making data FAIR in your project?
- How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).
- Who will be responsible for data management in your project?
- Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

4. DATA SECURITY

- What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
- Is the data safely stored in certified repositories for long term preservation and curation?

5. ETHICAL ASPECTS

- Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
- Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

6. OTHER ISSUES

- Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

Example of a DMP Horizon 2020

Call: H2020-SC1-2019-Single-Stage-RTD
Topic: SC1-BHC-07-2019
Funding Scheme: Research and Innovation Action (RIA)
DELIVERABLE 8.4

Check it full at:
https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5d2b42ed6&appId=PPGMS

DATA MANAGEMENT PLAN
Project Number: 874827
Project Acronym: BRAVƎ
Project title: Computational biomechanics and bioengineering 3D printing to develop a personalised
regenerative biological ventricular assist device to provide lasting functional support to damaged hearts

1. Data Summary

The BRAV3 Data Management Plan has been conceived to support the information management life cycle for all data to be collected, processed or generated by the project. This document outlines the strategy for data management within the BRAV3 project as part of the H2020 Open Research Data Pilot. The main aim is to identify best practices for storing and gathering information about the variety of data to be used in the project. This could in turn optimise the development of specific standards and assess the suitability for sharing and reusing it, in accordance with official guidelines. Data security will be also considered. Hereafter we report the information about the contents and the collection/generation of the data provided by all partners. At this early stage of development of the project, the reported information will be processed to determine the general specifications and structures of the metadata that will be generated within the DMP.

This plan covers data produced from any computational, experimental or management activity, which is linked with BRAV3, both raw and processed data. This can include (but is not limited to) exported software files, spreadsheets, images, text files, computer code and videos. The produced

data is expected to be predominantly digital, and an effort should be made to digitise any data, which exists in physical format. It is expected that the output of the project may reach 10s of TBs depending on the format of raw data. It is not expected that any personal data will be produced (e.g. patient data). However, commercially relevant data is foreseen. The origin of data will be *in silico*, *in vitro, ex-vivo* and *in vivo* experimental activities, along with reports from external subcontractors or regulatory agencies. As part of this project, it is anticipated that several classes of data will be produced:

· **Management data -** e.g. files related to consortium meetings (minutes, presentations), financial reports and reports from subcontractors and regulatory agencies.
· **Raw experimental data** – e.g. files and spreadsheets output from experimental activities including initial processing.
· **Processed experimental data** – e.g. experimental data, which has been graphed or processed for presentation.
· **Publication data** – e.g. a subset of raw/processed data, which underpins journal or conference publications.

For the purposes of the Open Research Data pilot, the need for open access applies primarily to the data underpinning scientific publications. Other data can also be provided via open access on a voluntary basis if it does not conflict with the commercial strategy of the project (i.e. institutional repositories and/or Zenodo will be considered).

To prepare the first draft, BRAV3 consortium will use the H2020 Data Management Plan Template. Once the project has advanced, the idea is to use ARGOS, an online tool in support of automated processes to creating, managing, sharing and linking Data Management Plans. ARGOS has the support of the OPENAIRE initiative and for sure will be the best option to address FAIR and Open best practices of BRAV3's data.

BENEFICIARY 4 – UKW

1.1 What is the purpose of the data collection/generation and its relation to the objectives of the project?	UKW is mainly involved in BioVAD fabrication and material synthesis for the regenerative and supportive part including adhesives. CAD files for print preparation and print files will be generated and the constructs will be characterization mainly via optical imaging and SEM. The printing process can be documented via video recordings. This is envisioned for demonstration and not for quality control and thus will not be performed routinely. The materials will be characterized via physicochemical and biochemical material characterization. In addition, the BioVAD will be characterized via rheological, mainly mechanical, testing.
1.2 What types and for-mats of data will the pro-ject generate/collect?	Reports and analysis: .docx. pptx, .xlsx, .opj, .pdf (software: Microsoft Office, Origin, Adobe Acrobat) Images: including .jpg, .tiff, .png, .dng, .czi (Zeiss microscopes) and .lif (Leica microscopes) (software: ImageJ) Videos: .mov, .mp4 (software: Windows movie player, ImageJ)

	NMR: .csv, .xls (Bruker) GPC: .csv, .xls (Omnisec, Malvern) UV/Vis, RAMAN and FTIR: .csv, .xls (Thermo Fischer) Mechanical Testing: .csv, .xls (Zwick and TA) Print files: .gcode, .iso (software: Match3, BioCAD from regenHU) CAD files: .stl, .step, .slddrw, .sldprt, .sldasm (SolidWorks, AutoCAD)
1.3 Will you re-use any existing data and how?	Existing data will be only used in exceptions. Print files will be adapted from existing templates but generated individually for the prints. Existing data or data from literature might be used to benchmarking and comparison of results.
1.4 What is the origin of the data?	Data will mainly originate from the experiments performed at UKW. Data from the BRAV3 consortium will include print files and output from simulations, biochemical analysis of the cell seeded MEW constructs as well as cell material interactions and biological characterization of the gels and adhesives.
1.5 What is the expected size of the data?	Several gigabytes: mainly image and video files, other files are text based and much smaller
1.6 To whom might it be useful ('data utility')?	The data (mainly print files) will be generated and exchanges within the BRAV3 consortium. It will be useful for the scientific community (publications, presentations) but also for industry (patents) and the society (in case of products originating from the project).

2. FAIR data

2. 1. Making data findable, including provisions for metadata

Data description

The description of procedures to generate data is associated to a **dataset** (i.e. collection of data).
At this stage of development of the project, the specific typology and total number of variables in a single dataset table (see Data Summary) cannot be defined *a-priori*.
The procedures for the identification of data are defined as follows:

- Each dataset is initially assigned to a unique ID, automatically generated through a Universally Unique Identifier (UUID) application.
- Each dataset is also associated to a Digital Object Identifier (DOI). The service is provided by the DOI (www.doi.org) community through a request to a local Registration Agency (RA).

The use of a DOI guarantees, at the same time, unique identification of the single dataset and the possibility of automatic data web retrieval.

After this step, the dataset is univocally associated to an identifier.

The implementation of the data description depends on the typology of datum considered. In most cases, a text description is appropriate. In this case, data are described by compiling a form (data description template), available to all users.

Metadata
Metadata should be included with all data (open access or otherwise). Metadata should allow the reader to understand the content of the data, how the data was collected (e.g. referring to a specific protocol), how it can be read (i.e. if specific software is required) and the date of production/edit.

One or more metadata files are generated for each dataset. The metadata are identified by the same unique ID of the related dataset, with a different suffix/extension.

Each metadata file is uploaded in a standardized format, depending on the dataset considered. Appropriate templates will be available for download to all BRAV3's partners in the shared folders. Metadata can be in the format of a text based readme file stored in the same directory as a collection of data. Keywords will be provided to optimize possibilities for re-use. Therefore, a file naming convention (FNC) will be implemented. This will help to organize the data and to identify the information, by grouping files that contain similar information close together. For naming convention, it will be considered several descriptive aspects, such as date, project data, acronym, lab work, author and version, all clearly separated by underscores. The files with a naming convention provide a preview of the content, are organized in a logical way (by date yyyy-mm-dd) identify the responsible party and convey the work history, unlike the files without a naming convention. Clear version history should be maintained on files within the filename (e.g. 20200601_Brav3_LAbTest_Mazo_v1.xls).

2.2. Making data openly accessible

By default and as a first case-study of data management, only data related to publications will be made openly available. In general, the General Assembly will decide on a case-by-case basis which data can be released in order to avoid issues related to IP rights protection or access. Before publication, consideration should be given to the commercial significance of the data. It is recommended to consult the consortium prior to publication to ensure that data of commercial interest is not prematurely published. BRAV3 is a multi-beneficiary project so data will be shared as possible, but the beneficiaries can keep date closed (Consortium Agreement) if the give reasons for opting out.

Data will be accessible through specific repositories. All data and metadata files will be uploaded onto a cloud storage and sharing facility specifically dedicated to BRAV3 project. The unique ID allows the retrieval of data and metadata files to registered users. Underpinning data should be copied to an open access repository for long-term storage and access. The lead author should take responsibility for arranging data storage unless otherwise agreed between authors. Institutional repositories such as UNAV's (DADUN) or Zenodo repository can be utilised (EU based). The repository should provide a DOI number to allow open access to the data.

2.3. Making data interoperable

All data will be made available in standard/open formats compliant with commercial/open software in order to allow as much as possible data exchange between researchers and institutions. Standard vocabulary for metadata description will be used, in case this will not be possible a mapping of more common ontologies (i.e. diagnostics, optoelectronics, plasmonics, assay, calibration, electronics, module, validation, demonstration…) will be provided. In this case, specific technical contribution from specialists in semantics and logics will be considered.

Where possible data should be stored in a format, which is compatible with commonly used operating systems (Windows, MacOS, Linux). For images, this includes JPEG, PNG and TIFF formats and for other documents, this should be in CSV, TXT, PDF or formats associated with Microsoft Office/Open Office. In the case of files which are in less common formats (e.g. output from specific devices or analysis software) then a minimally processed version should also be included (e.g. data imported into an spreadsheet or text file). To increase reusability files should use clear headers and minimise the use of acronyms. Alternatively, a glossary of acronyms can be included within metadata.

2.4. Increase data re-use (through clarifying licences)

The data will be made available according to Open Licenses such as Creative Commons.
The data will be available for re-use upon decision of the General Assembly, in order to avoid issues related to IP rights protection or access. Once the data are made openly available, they will remain open.

Within the strategy of development of DMP, the dataset that will be firstly available are those reported in publications originated from the consortium, thus intrinsically made for being reused. However, specific agreements with the Editors of scientific/technological journals will be considered and provided.

The data will be assigned a DOI for identification and licensed under Creative Commons Attribution (CC-BY) allowing reuse with attribution, even beyond the lifetime of the project. Data from the project must be accessible for 10 years from the last usage. This applies to date stored in repositories and via cloud or local storage.

The data quality is assured by each partner, that bears the responsibility of them. The tools necessary for describing and identifying the dataset and for preparing the metafiles will be provided by the General Assembly in strict collaboration with the Coordinator and the Exploitation and Innovation Committee.

3. Allocation of resources

Data storage and management is the responsibility of the partner producing the data. In terms of data storage costs, all partners have existing IT infrastructure to allow the storage of project data. Institutional repositories such as Enlighten and Pure are free to use for academic staff.
The costs for making data FAIR include the costs of the cloud facility and of personnel involved in collecting and managing data:

- Set up of the data space in the Project Collaborative Platform
- Implementation of the UUID generator
- DOI registration request
- Preparation of templates for: Data descriptor (general, text format, pdf output); Metadata: text template, spreadsheet template
- Data collection
- Generation of the data description
- Generation of the metadata
- Upload to the private cloud server
- If public, upload data to ZENODO/OpenAIRE

UNAV as project's coordinator will be responsible for the data management and will be assisted by the WP leaders.

The overall cost of DMP according to the reported cost-items is considered and covered by the *Open Access* cost the project Budget.

Regarding the resources for long-term preservation will be discussed as soon as the General Assembly make the first monitoring meeting.

4. Data security

As an initial step, only the Consortium Partners will have access to the cloud storage where dataset and metadata are filed. Other options about managing store data according to the inputs from the General Assembly. It is the responsibility of all project participants to ensure secure storage of project data, including implementing a regular backup regime. Data should be stored in a minimum of two separate physical locations, e.g. on a local computer and an institutional storage drive (stored locally on campus). Regular backup (i.e. weekly or monthly should be carried out to minimise opportunities for data loss).

Once the data is curated, will be stored and shared using Cloud storages with restricted access to authorized users. This system will be used to share management and experimental data between all project participants. Storage of publication data via institutional (e.g. DADUN) and public (Zenodo) repositories complies with the needs for long-term storage and accessibility of public data.

5. Ethical aspects

Most of the data that will be collected during the course will not be personal data. All the project data and all research should comply with local institutional and national guidelines on research. In this particular case, the data an informed consent for data sharing and long-term preservation will be strongly considered.

6. Other issues

This data management plan should be followed in parallel to local institutional data management practices.

Annex 3. Readme file template

Template from CSUC.

```
[The instructions in this document are in square brackets and the level of
obligation of each section is included at the beginning of them.]

[When you have completed this document, remove all instructions and delete
sections that you did not complete because they are not applicable in your
dataset.]

[Fill in as many sections as possible in this document to facilitate the
reuse of your dataset.]

[This document has been created by the CSUC Research Support Working Group
based on the staff of the Libraries Service of the Universitat Autonoma
de Barcelona.]

[For any information on this document, contact XXX.]

[This template is released under a CC0 license.]

This readme.txt file was generated on <DD-MM-YYYY> by <Name>

GENERAL INFORMATION
------------------
1.  Dataset title:
[Mandatory]

2.  Authorship:
[Mandatory | Fill in the information of all the authors following the
format below. Repeat the structure, one for each author.]

      Name:
      Institution:
      Email:
      ORCID:

DESCRIPTION
----------
1.  Dataset language:
[Mandatory if applicable]

2.  Abstract:
[Mandatory | Summary of the dataset, not of the research results. This
must be different from the abstract of the article, book or thesis. You
can add the abstract in different languages.]

3.  Keywords:
[Mandatory | Minimum 3 keywords. You can add them in different languages.]

4.  Date of data collection (single date or date range):
[Mandatory | Format DD-MM-YYYY, or DD-MM-YYYY - DD-MM-YYYY]

5.  Publication Date:
[Mandatory if applicable | Format DD-MM-YYYY]
```

6. Grant information:
[Mandatory if applicable | Repeat the scheme for each agency if necessary.]

Grant Agency:
Grant Number:

7. Geographical location/s of data collection:
[Mandatory if applicable | Latitude, longitude, city, region, country, etc.]

ACCESS INFORMATION
---
1. Creative Commons License of the dataset:
[Mandatory | The recommended license is CC-0 or CC-BY.]

2. Dataset DOI:
[Mandatory]

3. Related publication:
[Mandatory if applicable | Bibliographic citation, in the standard style of your discipline, of the publication related to the dataset, including the DOI.]

4. Link to related datasets:
[Recommended if applicable | (E.g. other datasets from the same project) following the format below.]

DOI/URL:

VERSIONING AND PROVENANCE
---
1. Last modification date:
[Recommended if applicable | Format DD-MM-YYYY]

2. Were data derived from another source?:
[Mandatory if applicable | Answer Yes or No; if yes, cite the source following the standard style of your discipline.]

3. Additional related data not included in this dataset:
[Optional]

METHODOLOGICAL INFORMATION

[Description of the methodology used to generate the dataset.]
[Do not copy and paste the text of the methodology section of a document that is pending of publication, unless you know that it is allowed: some publishers may consider it as a previous publication and not accept your manuscript.]

[If you are referring to an unpublished article, please provide as much information as you can.]

[Include DOI in references.]

1. Description of the methods used to collect and generate the data:
[Recommended | Reference, in the standard style of your discipline, the publications or documents that contain the experimental design or the protocols used in data collection.]

2. Data processing methods:
[Recommended | Describes how the published dataset was generated from the collected raw data.]

3. Software or instruments needed to interpret the data:
[Mandatory if applicable | Include the software version. If there is specific software with restricted access, explain how it can be obtained. Please assess whether it is possible to change the dataset to an open format (recommended).]

4. Information about instruments, calibration and standards:
[Mandatory if applicable]

5. Environmental or experimental conditions:
[Mandatory if applicable | E.g.: atmospheric influences, computational environment, etc.]

6. Quality-assurance procedures performed on the data:
[Optional]

FILE OVERVIEW

[All files included in the dataset must be mentioned, with the name and extension (.csv, .pdf, etc.) of each file. Include the directory structure, if applicable.]

1. Explain the file naming conversion, if applicable:
[Mandatory]

2. File list:
[Mandatory | Repeat the scheme for each file.]

File name:
Description:

3. Relationship between files:
[Mandatory if applicable]

4. File format:
[Mandatory]

5. If the dataset includes multiple files, specify the directory structure and relationships between the files:
[Mandatory if applicable]

SPECIFIC INFORMATION FOR TABULAR ATA

[This section should be repeated for each data file that requires the explanation of variables (usually tabular data). All variables are described, including units of measure.]

1. Name file:
[Optional]

2. Number of rows and columns:
[Optional]

```
3.  Variables list:
[Optional | Repeat the structure for each variable.]

      Variable name:
            Description:
            Units of measure or value labels:

4.  Codes or symbols for missing data:
[Optional | Repeat the structure for each code or symbol.]

      Code or symbol:
      Definition:

5.  Special formats or abbreviations used:
[Optional]


MORE INFORMATION
--------------
[Include any other information about the dataset that is not reflected in
this template and that you consider relevant.]
```

References

CODATA-CASRAI (2022). *Research Data Management Terminology*. https://codata.org/initiatives/data-science-and-stewardship/rdm-terminology-wg/rdm-terminology/

CESSDA Training Team (2017 - 2019). *CESSDA Data Management Expert Guide*. Bergen,Norway: CESSDA ERIC. DOI: 10.5281/zenodo.382047. Retrieved from https://dmeg.cessda.eu/

DANS (n.d.) File Formats. Retrieved from https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats

IBEC (June 2021). Code of conduct for research integrity. https://ibecbarcelona.eu/wp-content/uploads/2021/06/2021-IBEC-Code-of-conduct-for-research_web.pdf

IBEC (September 2021). *Research Data Management Policy*. https://ibecbarcelona.eu/wp-content/uploads/2021/11/IBEC-Data-Management-Policy.pdf

UK Data Service (n.d.). *Research data management*. https://ukdataservice.ac.uk/learning-hub/research-data-management/

Xunta de Galicia and the CERCA Institution (2021). *Guide to data management in R&D&I* centres. https://cerca.cat/wp-content/uploads/2022/03/Data-Management-Guide-in-RD-centres.pdf