

stretchBIO

D7.2 – DATA MANAGEMENT PLAN

StretchBio – DMP

Identifier:	StretchBio_D7.2_ Data Management Plan
Work package:	WP7
Dissemination level:	Public - ORDP
Keywords:	Data, Open Research, Pilot
Abstract:	This deliverable is the first version of the Data Management Plan (DMP) of the StretchBio project. The overall purpose of this document is to support the data management life cycle for all data that will be collected, processed or generated by the project.



Document history:

Version	Date	Reason of change
1	30/01/2022	Preliminary draft
1.1	13/02/2022	Complete draft
1.1	23/02/2022	Deliverable review and validation by Coordinator
1.2	28/02/2022	Final version

Document author(s):

Entity	Contributor
UB-IN2UB	Francisco Hernández
UB-IN2UB	Albert Romano Rodríguez
Consortium	All partners have contributed to the preparation of the Annex I

Disclosure Statement:

This document has been produced by consortium partners of the *StretchBio* Horizon 2020 project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 964808. The content of this document, the information contained herein, and the views expressed are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained.



Executive Summary

This deliverable is the first version of the Data Management Plan (DMP) of the StretchBio project and it follows the H2020 DMP template¹ that was designed by the European Commission to be applied to Horizon 2020 projects which produce, collect or process research data.

The overall purpose of this document is to support in a practical approach the data management life cycle for all data that will be handled by the project following the FAIR² principles as a general guideline.

This version of the DMP remains incomplete due to the initial stage of StretchBio implementation, and as a living document, will be progressively updated as the project moves forward and significant changes occur.

¹ H2020 templates: Data management plan v1.0 – 13.10.2016

² FAIR = Findable, Accessible, Interoperable and Re-usable



Table of Contents

EXECUTIVE SUMMARY	3
1 INTRODUCTION	5
1.1 PURPOSE OF THE DOCUMENT	5
1.2 STRUCTURE OF THE DELIVERABLE	5
1.3 DMP GUIDING PRINCIPLES AND STRATEGY	5
2 DATA SUMMARY	7
3 FAIR DATA	8
3.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA	8
3.2 MAKING DATA OPENLY ACCESSIBLE.....	9
3.3 MAKING DATA INTEROPERABLE	10
3.4 INCREASE DATA RE-USE (THROUGH CLARIFYING LICENCES).....	10
4 ALLOCATION OF RESOURCES	12
5 DATA SECURITY	13
6 ETHICAL ASPECTS	14
7 CONCLUSIONS	15
8 GLOSSARY	16
ANNEX 1 – DATA INVENTORY PLAN	17



1 Introduction

1.1 Purpose of the document

StretchBio DMP has been conceived to support the management life cycle for all data, in the broadest sense, which will be generated, collected and processed by this H2020 project.

The deliverable aims to identify best practices and specific standards applicable to the project data and assess their suitability for sharing and reuse in accordance with official guidelines.

Actually, the StretchBio consortium is committed to the EC objective of making research data findable, accessible, interoperable and re-usable in line with the so-called FAIR principles.

1.2 Structure of the deliverable

The document is structured following the guideline of the H2020 programme on FAIR Data Management in Horizon 2020³ including the following information:

- Data Management Plan (DMP) guiding principles and strategy
- Description of StretchBio type of data
- Introduction to FAIR DATA
- Allocation of resources
- Data Security
- Ethical Aspects

1.3 DMP guiding principles and strategy

All H2020-funded projects must disseminate as much information as possible. On top of that, the StretchBio project participates in the “Open Research Data Pilot”, which means that we are committed to giving open access to all data generated during the project unless it goes against the consortium’s legitimate interests.

In this regard, the main purpose of the DMP is to contribute to the accessibility and intelligibility of the data generated during the project implementation with the aim to comply with the Guidelines⁴ of the “Open Research Data Pilot”.

StretchBio DMP is realized within the Work Package 7 on coordination and management.

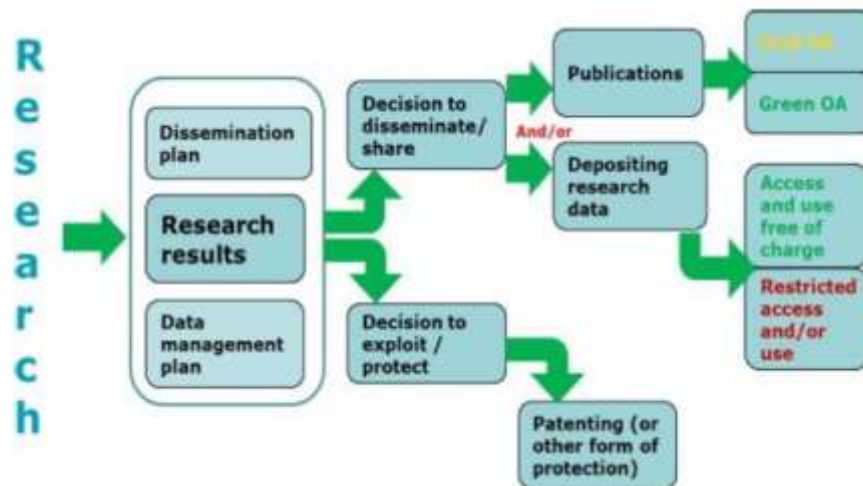
As far as data are concerned, the consortium rationale to use research results will accommodate the diagram shown below.

³ H2020 templates: Data management plan v1.0 – 13.10.2016

⁴ ENLACE



All partners are committed to implementing this workflow during the project implementation.



With this in mind, the main principles for the StretchBio project are the following:

- The DMP has been prepared using the template of the “Guidelines on Data Management in Horizon 2020”⁵
- The DMP is an official project Deliverable (D7.2) due in month 6, but it will be updated throughout the project implementation, when necessary. This first initial version will evolve depending on the research activity and the feedback received from the EC representatives and other significant project stakeholders.
- The DMP is subject to the Regulation (EU) 2016/679 and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (GDPR), as well as other applicable laws on this matter.
- Type of data, storage, confidentiality, ownership, management of intellectual property and access: procedures that will be implemented for data collection, storage, access, sharing policies, protection, retention and destruction will be in line with the EU standards as described in the Grant Agreement and the Consortium Agreement.

⁵ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf



2 Data Summary

As a project participating in the Open Research Data Pilot (ORDP) in Horizon 2020, the DMP' strategy of StretchBio project is focused on the observation of FAIR (Findable, Accessible, Interoperable and Reusable) "Data Management Protocols".

Each data set created during the project will be assessed and categorized as open, embargoed or restricted by the owners of the content. The project coordinator will closely monitor the process to guarantee a homogeneous approach to the decision-making process.

All the data sets, regardless of their categorization, will be stored in each of the participating entities databases. In addition, those categorized as opened or embargoed will be publicly shared (in the case of embargo, after the period is over) through the ZENODO project site⁶ and the project website.

ZENODO is an open access repository for all fields of science that allows uploading any kind of data file format, which is recommended by the Open Access Infrastructure for Research in Europe (OpenAIRE).

Those data of particular interest for dissemination purposes will be published on the project website and the adequate social media channels.

Among all project datasets, deliverables and associated documentation, the following categories of the project outputs are already declared as "ORDP" and as a consequence, they will be made, as a rule, "Open Access" (to be provided free of charge for public sharing):

- Project deliverables labelled as public in the Grant Agreement (after the formal approval by the EC services)
- Articles published in Open Access scientific journals (after acceptance for publication)
- Conference and Workshop abstracts/articles

In annex A of this deliverable, the types of data that can be generated per work package are early identified by the consortium members. This list is not exhaustive and may be enriched during the project implementation.

The follow-up and regular update of the table content will be carried out by the project coordinator using a centralized register available upon request. The status of each project dataset and document will be categorised as "opened, embargoed or restricted".

⁶ <https://zenodo.org/communities/stretchbio/?page=1&size=20>



3 FAIR Data

In 2016, the [‘FAIR Guiding Principles for scientific data management and stewardship’](#) were published in *Scientific Data* ⁷. The authors intended to provide guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets.

The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

Keeping these principles as a general framework to organize the management of R+D data resulting from StretchBio project, the following considerations and actions will be taken by the consortium:

3.1 Making data findable, including provisions for metadata

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for the automatic discovery of datasets and services, so this is an essential component to generate FAIR data.

Metadata are data on the research data themselves. They enable other researchers to find data in an online repository and is, as such, essential for the reusability of the dataset. By adding rich and detailed metadata, other researchers can better determine whether the dataset is relevant and useful for their own research.

In the frame of StretchBio activity, metadata (type of data, location, etc.) will be uploaded in a standardized form. This metadata will be kept separate from the original raw research data. As described in the project Grant Agreement (Article 29.2), the StretchBio bibliographic metadata will include all of the following:

- the terms “European Union (EU)” and “Horizon 2020”;
- the name of the action, acronym and grant number;
- the publication date, and length of embargo period if applicable
- a persistent identifier

As stated in Section 2, StretchBio open data will be collected in an open online research data repository, ZENODO, whose structure, facilities and management are compliant with FAIR data principles.

ZENODO allows researchers to deposit both publications and data, providing tools to link them to these through persistent identifiers and data citations; it is set up to facilitate the finding,

⁷ Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).



accessing, re-using and inter-operating of data sets, which are the basic principles that H2020 projects must comply with.

This CERN-based repository enables researchers, scientists, EU projects and institutions to:

- Share research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science
- Display their research results and get credited by making the research results citable and integrating them into existing reporting lines to funding agencies like the European Commission
- Easily access and reuse shared research results
- Integrate their research outputs with the OpenAIRE portal.

ZENODO also allows to perform simple and advanced search queries using keywords.

To facilitate the searches, StretchBio consortium will use a straightforward naming convention. Files and folders will be versioned and structured as follow: **StretchBio_ID_YYYYMMDD_Vzz**, where *ID* is a unique identifier, *YYYYMMDD* the creation date and *Vzz* the version number that will be incremented at each revision.

3.2 Making data openly accessible

To maximise the impact of StretchBio research data, the results will be freely distributed within and beyond the consortium, unless it goes against the partners' legitimate interests.

Selected data and results will be shared with the scientific community and other stakeholders through publications in scientific journals and presentations at conferences, as well as through open access data repositories.

As said before, StretchBio consortium is committed to making the project data FAIR whenever possible. As far as scientific publications are concerned, priority shall be given to high-rank scientific journals from the participating disciplines.

Keeping this in mind, Green Open Access will be the preferred formula to disseminate the project results, but the editorial policy of well-known journals such as Nature will force the consortium to opt for Gold Access in some particular cases.

The fees associated with the publication of "Gold" papers have already been foreseen in the project budget. The cost-sharing, in the case of multiple authors, shall be decided among the authors on a case-by-case basis.

Green and Gold manuscripts will be freely available in the ZENODO-project dedicated site (see Section 2) as well as curated and preserved in the institutional repositories of UB, DTU and ALU-FR following the provisions of the standard policies of these academic partners and the journal editors' requirements.

To protect the copyright of the project knowledge, a Creative Commons license will be used in some cases. If data are not labelled as free, potential users must contact the project IPR team or the data owner in order to gain access. If necessary, appropriate IPR procedures (such as non-disclosure agreement) will be used.



3.3 Making data interoperable

Data interoperability is crucial to push transversal, collaborative and interdisciplinary research: data usually need to be integrated with other data. To that end, the data need to interoperate with applications or workflows for analysis, storage, and processing.

In the frame of StretchBio project:

- (Meta)data will use a formal, accessible, shared, and broadly applicable language for knowledge representation
- (Meta)data will use vocabulary that follows FAIR principles
- (Meta)data will include qualified references to other (meta)data

These generic considerations will be implemented by observing OpenAIRE guidelines for online interoperability, including OpenAIRE Guidelines for Literature Repositories, OpenAIRE Guidelines for Data Archives, OpenAIRE Guidelines for CRIS Managers based on CERIF-XML⁸.

As the project progresses and data are collected, further and more detailed information on making data interoperable will be outlined in subsequent versions of the DMP taking into account the real nature of the generated data.

3.4 Increase data re-use (through clarifying licences)

Creative Common Licensing⁹ will be used to protect the ownership of the datasets, where considered necessary. However, an embargo period may be applied if the data (or parts of data) are used in published articles in “Green” open access and “Gold” journals.

The recommended maximum embargo period length by European Commission is 6 months, and as a consequence, the participants will do their best to meet this requirement whenever possible.

For datasets deposited on a public data repository (ZENODO project site), the access of free data will be unlimited in time, but potentially subject to restrictions of use. These restrictions could include:

- the “embargo” period imposed by journals publication policy (Green Open access or Gold journals);
- some or all of the following restrictions may be applied with Creative Commons licensing of the dataset:
 - **Attribution:** requires users of the dataset to give appropriate credit, provide a link to the license, and indicate if changes were made
 - **NonCommercial:** prohibits the use of the dataset for commercial purposes by others, if not previously agreed with the data owners

⁸ <https://guidelines.openaire.eu/en/latest/>

⁹ <https://creativecommons.org/licenses/>



- **ShareAlike:** requires the others to use the same license as the original on all derivative works based on the original data.

These issues will be further developed in the coming versions of the DMP once the IPR strategy of the project is more mature.



4 Allocation of Resources

The active management of StretchBio data is resource consuming. The partners will bear the associated costs by using the EC-grant as well as their own resources. As a rule, each partner is responsible for the data they produce.

In a simplified manner, the costs for making data FAIR includes:

- **Fees associated with the publication of scientific articles** containing project’s research data in “Gold” Open Access journals. Any fee incurred for Open Access will be the responsibility of the data owner (authors) partner(s). The cost-sharing, in the case of multiple authors, shall be decided among the authors on a case-by-case basis
- **Project Website** operation
- **Data archiving** at ZENODO and on other online databases
- **Copyright licensing** with Creative Commons (in principle free of direct charges)

The Project Coordinator on behalf of the StretchBio General Assembly is responsible for the follow-up of the Data Management of research data, which will always be aligned to the guidelines of the Data Protection Officers (DPOs) appointed by the partners.

Finally, we should remember that costs related to Open-Access to research data in Horizon 2020 are eligible for reimbursement under the conditions defined in the H2020 Grant Agreement, in particular Article 6 and Article 6.2.D.3. Project beneficiaries will be responsible for applying for reimbursement for costs related to making data accessible to others beyond the consortium.



5 Data Security

Securing confidential and restricted data is globally one of the main challenges that many R+D organisations face in their daily activities. StretchBio data are per se valuable, and if the project succeeds, it could sustain a significant step forward in the current state-of-the-art knowledge with potential business impact. For this reason, the security of the project data is key for the consortium, the EC and the European society.

With this in mind, the following guidelines will be applied to guarantee the confidentiality, availability and integrity of the StretchBio data:

- Store data in at least two separate locations to avoid loss of data
- Encrypt data if it is deemed necessary by the participating researchers
- Limit the use of USB flash drives
- Label files in a systematically structured way in order to ensure the coherence of the final dataset
- Follow the security guidelines of the partners' IT departments

Further details on the actual security actions to prevent and mitigate cyberattacks are labelled as confidential. The consortium will follow the strategy known as “security through obscurity”¹⁰.

¹⁰ https://en.wikipedia.org/wiki/Security_through_obscurity



6 Ethical Aspects

StretchBio workpackage 1 aims to ensure that ethical requirements are met for the research undertaken in the project, including data management aspects, in compliance with H2020 ethical standards.

All partners assure that the EU standards regarding ethics and data management will be fulfilled in compliance with the ethical principles (see Article 34) and confidentiality (see Article 36 as set out in the Grant Agreement).

In addition:

1. In accordance with the General Data Protection Regulation 2016/679, the data controllers and processors are fully accountable for the data processing operations
2. Templates for informed consent forms for research participants are available. Other ethical aspects have been evaluated and correctly addressed from the early stages of the project.
3. An independent ethics advisor (Dr Albert ORIOL, head of the Unit for Clinical Investigation of the Catalan Institute of Oncology), has been appointed following the Grant Agreement provisions and will monitor the project implementation and the adequate use of research data.

See the following deliverables for more details:

- D.1.1 H - Requirement No. 1
- D.1.2 POPD - Requirement No. 3
- D.1.3 OEI - Requirement No. 12



7 Conclusions

This first version of the DMP, in the form of a project deliverable, describes the main principles and guidelines for the Data Management in StretchBio.

As a living document, it will be enriched throughout the project lifetime. Further updating of the Data Management Plan, as data are gradually generated and collected, will include an overview of the data inventory with full details of which concrete data are collected, their description and how they can be accessed by external stakeholders.



8 Glossary

Abbreviation	Explanation
DMP	Data Management Plan
DX	Deliverable X
EC	European Commission
EU	European Union
GDPR	General Data Protection Regulation
ORDP	Open Research Data Pilot
WP	Work package



Annex 1 – Data Inventory Plan

WP1 & WP7 (Management)						
Type of data generated	Origin	Total Number of Elements	Data Purpose	Format	Openly accessible	Comments / description
Project deliverables and reports	Inputs from partners & generated by WP leader	Those indicated in the GA	EC reporting obligations	.doc & .pdf	Y	Those reports labelled as Public will be published in the project website and Zenodo
Images	Inputs from partners & generated by WP leader	<500	EC reporting obligations	.jpg, .TIFF, .png, .pdf	Y	Those figures not labelled as confidential or submitted for publication will be potentially disclosed
Other	Inputs from partners & generated by WP leader	NA	EC reporting obligations	.pdf, .doc, other	N	Documentation requested by the EC services to clarify some doubts (i.e., consent form template for patients involved in the research)



D7.2 – Data Management Plan

WP8 (Dissemination & Exploitation)						
Type of data generated	Origin	Total Number of Elements	Data Purpose	Format	Openly accessible	Comments / description
Project deliverables and reports	Inputs from partners & generated by WP leader	Those indicated in the GA	EC reporting obligations	.doc & .pdf	Y	Those reports labelled as Public will be published in the project website and Zenodo
Images	Inputs from partners & generated by WP leader	<500	EC reporting obligations	.jpg, .TIFF, .png, .pdf	Y	Public images used for communication and dissemination activities
Conference / workshops articles and abstracts	Inputs from partners & generated by WP leader	<50	EC reporting obligations	.doc & .pdf	Y	All information not labelled as confidential or submitted for publication will be potentially disclosed. Embargoed papers will be processed as indicated in the DMP
Other	Inputs from partners & generated by WP leader	NA	EC reporting obligations	.pdf, .doc, other	Y	Additional actions and documents resulting from workshops, trainings, short-term missions and exploitation strategies



D7.2 – Data Management Plan

Technical WPs (WP2 - WP6)						
Type of data generated	Origin	Total Number of Elements	Data Purpose	Format	Openly accessible	Comments / description
Project deliverables and reports (all WPs)	Inputs from partners & generated by WP leader	Those indicated in the GA	EC reporting obligations	.doc & .pdf	Y	Those reports labelled as Public will be published in the project website and Zenodo
Photonic crystal propagation results (WP2)	Post docs and PhD student	< 100	Internal documentation, data sharing	dat, pdf	Y	At request - Announced at Zenodo
Photonic crystal characterization results (WP2)	Post docs and PhD student	< 100	Internal documentation, data sharing	dat, pdf	Y	At request - Announced at Zenodo
Fabrication process flows (WP3)	Cleanroom	Unknown	Documentation of functioning processes	.doc & .pdf	Y	At request - Announced at Zenodo
AFM, SEM, optical microscope images (WP3)	Laboratories involved	Unknown	Performance characterisation and documentation	jpeg, png, etc	Y	At request - Announced at Zenodo
Experimental logs (WP3)	Post docs, Phd students and other PIs - Personal research data	Unknown	Internal documentation, data sharing	.doc & .pdf	Y	At request - Announced at Zenodo
Experimental images from patient samples	Post docs, Phd students and other PIs - Personal research data	Unknown	Internal documentation, data sharing	png, pdf	Y	At request - Announced at Zenodo
Simulation results (WP4 & WP5)	Post docs, Phd students and other PIs - Personal research data	< 100	Internal documentation, data sharing	mph, dat, pdf	Y	At request - Announced at Zenodo
AFM, SEM, optical microscope & contact angle images (WP5)	Laboratories involved	Unknown	Performance characterisation and documentation	jpeg, png, etc	Y	At request - Announced at Zenodo
Experimental data sets (WP5)	Post docs, Phd students and other PIs - Personal research data	Unknown	Internal documentation, data sharing	doc, pdf	Y	At request - Announced at Zenodo



D7.2 – Data Management Plan

Experimental images from patient samples (WP5)	Post docs, Phd students and other PIs - Personal research data	Unknown	Internal documentation, data sharing	png, pdf	Y	At request - Announced at Zenodo - Ethical scrutiny needed
Culture of cells and living tissues (WP6)	Postdoc, Technician, students	Unknown	Data sets internal documentation	TIFF, JPEG	Y	Those reports labelled as Public will be published in the project website and Zenodo
Experimental data sets (WP6)	Postdoc, Technician, students	Unknown	Internal documentation, data sharing	doc, pdf	Y	At request - Announced at Zenodo - Ethical scrutiny needed
Gene expression patterns, tumour generation and tests of viability (WP6)	Postdoc, Technician, students	Unknown	Internal documentation, data sharing	doc, pdf, Tiff, jpeg	Y	At request except those classified as confidential - Announced at Zenodo - Ethical scrutiny needed



[END OF THE DOCUMENT]

