



*Horizon 2020 Work programme*

Food Security, Sustainable Agriculture and Forestry, Marine, Maritime and Inland Water Research and the Bioeconomy

*Call*

H2020-FNR-2020: Food and Natural Resources

*Topic name*

FNR-16-2020: ENZYMES FOR MORE ENVIRONMENT-FRIENDLY CONSUMER PRODUCTS

*FuturEnzyme:*

Technologies of the Future for Low-Cost Enzymes for Environment-Friendly Products

Final ID: 101000327



29/09/2021

# DATA MANAGEMENT PLAN

DELIVERABLE NUMBER D8.4

MANUEL FERRER

CSIC

Marie Curie n2, 28049, Cantoblanco, Madrid, Spain

## Document information sheet

<b>Work package:</b>	WP8, Communication, Dissemination and Exploitation
<b>Authors:</b>	CSIC (Manuel Ferrer, Patricia Molina)
<b>Document version:</b>	1
<b>Date:</b>	29.09.2021
<b>Starting date:</b>	01/06/2021
<b>Duration:</b>	48 months
<b>Lead beneficiary:</b>	CSIC
<b>Participant(s):</b>	CSIC, BSC, Bangor, UHAM, UDUS, IST-ID, CNR, ITB, FHNW, CLIB, Inofea, BioC_Chem, Schoeller, Henkel, Evonik, Eucodis
<b>Dissemination Level:</b>	Confidential, only for consortium's members (including the Commission Services)
<b>Type</b>	ORD: Open Research Data
<b>Due date (months)</b>	4
<b>Contact details:</b>	Manuel Ferrer (mferrer@icp.csic.es), Patricia Molina (patricia.molina@icp.csic.es)

## Summary

History of changes .....	5
List of abbreviations .....	6
Data Management Plan .....	7
1. Scope of Deliverable .....	7
2. Data and Metadata (raw and processed) to be managed: Summary .....	7
Data and Metadata .....	8
3. FAIR data: decision-taken strategy for publishing, repositories and databases .....	14
Decision-taken strategy to publish and to manage the data and metadata .....	17
File extensions to be used for data and metadata storage .....	18
Standardization: making the bio-resources, data and metadata homogeneous and comparable .....	19
Quality controls .....	19
4. Allocation of resources for data and metadata open access .....	20
5. Data security .....	20
Confidentiality .....	20
GDPR .....	21
6. Ethical aspects .....	22
7. Innovation and Intellectual Property Rights Management Strategy .....	28
Annex .....	30
Template for new bioresources .....	30
Template for Microorganism datasets .....	31
Template for enzyme and enzyme-immobilized datasets .....	33
Template for declaration of invention/work form .....	35

## History of changes

Version	Publication date	Changes/Comments
1.0	29.09.2021	Initial version

## List of abbreviations

**DMP** Data Management Plan

**DOI** Digital Object Identifier

**DSP** DownStream Processing

**FAIR** Findable, accessible, interoperable and reusable

**GDPR** General Data Protection Regulation

**GMP** Good Manufacturing Practices

**HMM** Hidden Markov Models

**HR-MS** High Resolution Mass Spectrometry

**IP** Intellectual Property

**KPI** Key Performance Indicator

**LCA** Life Cycle Assessment

**MD** Molecular Dynamics

**MSA** Multiple Sequence Alignment

**MS** MileStone

**NMR** Nuclear Magnetic Resonance

**OA** Open Access

**ORD** Open Research Data

**PELE** Protein Energy Landscape Exploration

**RMSD** Root-Mean-Square Deviation

**RRI** Responsible Research and Innovation

**TEWL** TransEpidermal Water Loss

**SASA** Solvent-Accessible Surface Area

**WP** WorkPackage

# Data Management Plan

## 1. Scope of Deliverable

This deliverable describes the Data Management Plan (DMP) as a key element of good data management. The DMP describes the data management life cycle for the data to be collected, processed and/or generated by the Horizon 2020 project FuturEnzyme. As part of making research data findable, accessible, interoperable and reusable (FAIR), the DMP includes information on: the handling of research data during and after the end of the project; what data will be collected, processed and/or generated; which methodology and standards will be applied; whether data will be shared/made open access; and how data will be curated and preserved (including after the end of the project). In addition to the above, the DMP includes the Innovation and Intellectual Property Rights (IPR) Management Strategy of the project, main potential exploitation routes, the target groups (per project asset) that are potential end-users of the results, general terms of use and relevant IPR provisions, most promising joint exploitation plans for the whole consortium or specific groups of project partners, and the means and procedures for the exploitation of project's assets and a clear action plan to this.

According to the Open Research Data (ORD) strategy, like other EU-funded projects, this project will provide open access to the research data generated and published, ensuring their integration and reuse, considering the necessary balance between openness and protection.

This DMP is based on the template guidelines of the H2020 programme on FAIR Data Management and will be updated when required along with the project duration (see History of changes, pg. 4). At least, it will be reviewed in the reporting periods specified in the Grant Agreement (months 18, 36 and 48).

## 2. Data and Metadata (raw and processed) to be managed: Summary

The data generated along and in the frame of the project is reasonably predicted to be extensive. For example, by counting the bioresources available at the beginning of the project gathered by the partners from previous projects, more than 3000 entries are at the consortium disposal, including enzymes, isolates and metagenomic libraries/shotgun sequences. These bioresources have their origin in the previous EU-funded projects, which generated one of the biggest collections of enzyme bio-resources worldwide and as well as a large number of datasets and knowledge (see document for Deliverable D3.1). They include:

- FP7 project ULIXES (<http://www.ulixes.unimi.it/>),
- FP7 project MAMBA (<http://mamba.bangor.ac.uk/>),
- FP7 project MicroB3 (<https://www.microb3.eu/home.html>),
- FP7 project Kill•Spill (<http://www.killspill.eu/>),
- FP7 project MagicPAH (<https://cordis.europa.eu/article/id/31997-cleaning-up-the-hydrocarbon-act>),
- H2020 project INMARE (<http://inmare.bangor.ac.uk/>), and
- EraNET project MetaCat (DOI: <https://doi.org/10.21820/23987073.2018.5.82>).

The data and information related to these existing bioresources at the disposal of FuturEnzyme need to be managed and stored (done in the frame of Deliverable D3.1). In addition, FuturEnzyme has available sequence motifs and descriptors of enzyme performances and strategic directions for the development of a machine learning platform to rapidly bio-prospect the existing and internal databases for the enzymes which potentially meet the key performance indicators (KPIs) of the detergent, textile and cosmetic sectors (detailed in the Deliverable D2.1). Moreover, the Milestone MS5 in the frame of Work Package WP2 stands

for the compiling of a set of 50,000 homology driven sequences pre-selected with relevance for FuturEnzyme; this has been achieved by searching *in silico* internal and public repositories, using Hidden Markov Models (HMM) and BLAST (DIAMOND TOOL) approaches, followed by network analysis to identify sequences encoding enzymes of interest.

FuturEnzyme's project manager, Patricia Molina (patricia.molina@icp.csic.es), is responsible for Data Management to ensure its efficient and correct implementation.

The purpose of generating and managing data is double: to learn knowledge on how enzymes work, how they can be produced, and how they can be integrated into products, and to feed predictive software tools to be created in the frame of FuturEnzyme. In both cases, the objective is to put the data at the service of the implantation of enzymes in the industry to make processes and products greener and more functional first focused on textiles, cosmetics and detergents, and then on any other interested party or sector. In this latter sense, the effort will be put to reach stakeholders, policy makers, consumer associations, and the general public, as described in the Grant Agreement.

## Data and Metadata

During and after the project's lifetime, multiple types of data and metadata will be generated, resumed in **Table 1** and further described below.

**Table 1.** Category and short description of some of the different data and metadata that will be produced because of FuturEnzyme's activities.

CATEGORY	DESCRIPTION
<b>MICROORGANISMS</b>	Validated strain name Taxonomy Sampling site Enzymatic activity Etc.
<b>SEQUENCE</b>	16S amplicons DNA sequences (genomes, metagenomes, etc.)
<b>GENE CLONING AND PROTEIN SYNTHESIS</b>	PCR amplification Gene synthesis Cell free expression system Host Vector Inductor Antibiotic resistance Fusion tag Origin of the sample (GPS coordinates, microbe, environmental sample, etc.) Activity (esterase, lipase, protease, etc.) Laboratory and person implicated, production date, etc
<b>ENZYME CHARACTERIZATION</b>	Substrate specificity and enantioselectivity Specific activity $V_{max}$ $k_{cat}$ , $K_m$ , $k_{cat}/K_m$ Optimal physical-chemical parameters for activity Stability: half life time, denaturation T, pH and thermal stability Conversion rates Crystallographic structure



	Raw proteomic/mass spectrometry Introduction of mutations and their effects Etc.
<b>ENZYME PRODUCTION</b>	Media composition Fermentation parameters (O <sub>2</sub> , pH, T, fermentation time, etc.) Growth rate during fermentation Productivity Downstream processing (DSP) method tested Enzyme expression and production level Enzyme recovery yield Enzyme production quality controls and production process reproducibility Conditions for enzyme production and formulation of prototypes Conditions for enzyme immobilization and shielding Etc.
<b>ENZYME IMMOBILIZATION</b>	Particle size and polydispersity Particles concentration Immobilization yield Surface available Carrier pore size Shield thickness Enzyme concentration Enzyme activity Enzyme stability Recovered enzymatic activity Etc.
<b>COMPUTATIONAL</b>	Metrics extracted from molecular modelling software. Sequence information by Multiple Sequence Alignment (MSAs), Hidden Markov model (HMMs)... Structural information by homology modelling or others Catalytic positions Substrate-enzyme interaction energy Relevant distances between atoms Relevant angles and dihedrals between atoms Ligand's solvent-accessible surface area (SASA) Ligand's root-mean-square deviation (RMSD) Proteins backbone's RMSD Catalytic residues' RMSD Degree of sequence conservation Conserved motifs Etc.
<b>CHEMICAL</b>	Synthesis yield Nuclear Magnetic Resonance (NMR) datasets High resolution mass spectrometry (HR-MS) datasets Etc.
<b>PATENT AND BIBLIOGRAPHIC EXTENSIVE/MASSIVE SEARCH</b>	Sequences encoding enzymes of use in the 3 sectors relevant for FuturEnzyme Working conditions Conversions and yields Product performance Etc. Consumer stains / soiled fabric used

<b>PRE-INDUSTRIAL VALIDATIONS</b>	Washing buffer solution composition Enzymes' concentrations in the washing solution Stability of the enzyme formulations (washing and storage) Washing conditions Remaining stains in the fabric at the end of washing cycles Degradation products in the washing buffers at the end of washing cycles Textile or fabric used Composition of the treatment/finishing buffer solution Amino, carboxyl and hydroxyl groups released into the treatment liquid Amino, carboxyl and hydroxyl groups remaining in the fabrics Size and composition of enzyme-treated hyaluronic acid Hyaluronic acid degradation product yield Inclusion levels of the hyaluronic acid product in the targeted cosmetic Transepidermal water loss (TEWL) Skin hydration Etc.
<b>LIFE CYCLE ASSESSMENTS (LCA)</b>	Consumption of resources (biotic and abiotic), including energy consumption Air emissions Emissions to water Ground emissions Waste Etc.
<b>ONLINE SURVEY</b>	Information about consumers reaction to new enzymes and new products Contact information (name, address, phone number, and email address) Sensitive data such as gender, income level, and consumer habits
<b>PRODUCT TESTING SURVEY</b>	Information about reactions to the project's real products Survey on the new product Contact information (not released to FuturEnzyme) Etc.

Below, the nature and explanations for the data and metadata to be generated and managed are detailed.

#### **Microorganisms data/metadata**

The information obtained will be transferred to the templates created explicitly for FuturEnzyme (see Annex). These documents will be filled as completely as possible, gathering at least:

- Validated strain name,
- Taxonomy,
- Sampling site,
- Cultivation conditions,
- Access number or link, and/or
- Enzymatic activity.

#### **Sequence data/metadata**

The information obtained will be transferred to the templates created explicitly for FuturEnzyme (see Annex). These documents will be filled as completely as possible, gathering at least:

- 16S amplicons, and

- DNA sequences derived from single cultures, microbial consortia, genomes, metagenomes and fosmic/phage/cosmid clones.

#### **Gen cloning and protein synthesis data/metadata**

- PCR amplification
- Gene synthesis
- Cell free expression system
- Host
- Vector
- Inductor
- Antibiotic resistance
- Fusion tag
- Origin of the sample (GPS coordinates, microbe, environmental sample, etc.)
- Activity (esterase, lipase, protease, etc.)

#### **Enzyme characterization data/metadata**

Enzyme characterisation metadata (including raw data and final calculations). The information obtained will be transferred to the templates specifically created for FuturEnzyme (see Annex). These documents will be filled as completely as possible, gathering at least:

- Substrate specificity and enantioselectivity (enantiomeric excess, enantiomeric ratio),
- Specific activity,
- $V_{max}$ ,
- $k_{cat}$ ,  $K_m$ ,  $k_{cat}/K_m$
- Optimal physical-chemical parameters for activity,
- Stability: half life time, denaturation temperature, pH and thermal stability,
- Conversion rates,
- Crystallographic structure,
- Raw proteomic/mass spectrometry, and/or
- Introduction of mutations and their effects.

#### **Fermentation (from small to upscaling (non-GMP) fermentation) and enzyme production data/metadata**

The information obtained will be transferred to the templates created explicitly for FuturEnzyme (see Annex). These documents will be filled as completely as possible, gathering at least:

- Media composition,
- Fermentation parameters such as oxygen, pH, temperature, fermentation time, induction strategies, and media composition (e.g. addition of trace elements, co-factors, anti-foaming compounds),
- Growth rate during fermentation,
- Productivity,
- Downstream process (DSP) method tested, e.g., simple isolation of supernatants, precipitation or spray drying, separation on specific chromatographic resins, isolation with medium pressure chromatography followed by spray drying, etc.,
- Enzyme expression and production level (including SDS-PAGE gels, etc.),
- Enzyme recovery yield,
- Enzyme production quality controls and reproducibility of the production process,
- Best conditions for enzyme production and formulation of prototypes, and
- Best conditions for enzyme immobilization and shielding.

### Enzyme immobilization data/metadata

The information obtained will be transferred to the templates specifically created for FuturEnzyme (see Annex). These documents will be filled as completely as possible, gathering at least:

- Particle size and polydispersity (nm),
- Particles concentration (mg/mL),
- Immobilization yield (% as mg/mL suspension or mg/g carrier),
- Surface available (m<sup>2</sup>/mg),
- Carrier pore size (in case of porous particles) (nm),
- Shield thickness (nm),
- Enzyme concentration (mg/mL),
- Enzyme activity (U/mL; U/mg; U),
- Enzyme stability (%), and
- Recovered enzymatic activity (U/mL suspension or U/g carrier or %).

### Computational data/metadata

By applying computational methods for calculations, metadata will be generated, gathering at least:

- Metrics extracted from PELE (Protein Energy Landscape Exploration), molecular dynamics (MD), or other molecular modelling software,
- The information that can be extracted at the sequence level with MSAs, HMMs and others,
- The information that can be extracted from the structure obtained with homology modelling or AlphaFold, etc. (which is essential for any kind of molecular modelling data),
- Catalytic positions: PELE positions where the catalytic distances from the active center are as expected/ideal and the distance to the substrate would allow the reaction to take place,
- The substrate-enzyme interaction energy (energy describing the interaction between substrate and enzyme in PELE),
- The distances between atoms that may be relevant (for both MD and PELE),
- Angles and dihedrals between atoms that may be relevant (for both MD and PELE),
- The Solvent Accessible Surface Area (SASA) of the ligand: That part/area of the substrate that is exposed to the solvent normalised between 0 and 1,
- The Root Mean Square Deviation (RMSD) of the ligand with a reference pose (which can be from a crystallographic pose or one obtained by docking) (for both MD and PELE),
- The RMSD of the protein backbone: This tells us about the stability of the protein and its folding/conformation in the solvent,
- The RMSD of the catalytic residues,
- The degree of sequence conservation in the position of the residues, and
- Possible conserved motifs in the enzyme/protein family.

### Chemical data/metadata

The data and metadata will include those describing the synthesis and those describing the molecule characterisation. They will include at least:

- Synthesis yield,
- Nuclear Magnetic Resonance (NMR) datasets: <sup>1</sup>H, <sup>13</sup>C, <sup>31</sup>P, and
- High Resolution Mass Spectrometry (HR-MS) datasets.

### Patent and bibliographic extensive/massive search data/metadata

We will complete extensive bibliographic and patent search work. This is important as it will allow us to evaluate to which extent our developments are ahead of the state of the art. Aside from the corresponding reports, at least the following metadata will be generated and compiled:

- Sequences encoding enzymes found to be of use in the cosmetic, detergent and textile sectors,
- Working conditions (buffers, T, pH, additives, etc.) under which the enzymes have been applied,
- Conversions and yields, and
- Product performance.

#### **Pre-industrial validations data/metadata**

The data and metadata will include those related to the performance of enzymes and enzyme processes to prepare detergents and the bioprocessing of textiles and hyaluronic acid for cosmetics. They will include at least:

- Consumer stains / soiled fabric used,
- Composition of the washing buffer solution,
- Concentrations of the enzymes that will be introduced into the washing solution,
- Stability of the enzyme formulations under washing and storage conditions,
- Washing conditions (time, wash temperature, etc.),
- Residual specific and stubborn stains present in the fabric at the end of washing cycles
- Degradation products in the washing buffers at the end of washing cycles,
- Textile or fabric used,
- Composition of the treatment/finishing buffer solution,
- Composition of amino, carboxyl and hydroxyl groups released into the treatment liquid,
- Composition of amino, carboxyl and hydroxyl groups remaining in the fabrics,
- Size and composition of enzyme-treated hyaluronic acid,
- Hyaluronic acid degradation product yield,
- Inclusion levels (X% active matter) of the hyaluronic acid product in the targeted cosmetic (e.g., commercial Hyacare® product),
- Trans epidermal water loss (TEWL), and
- Skin hydration (Corneometer).

#### **LCA data/metadata**

The Life Cycle Assessment (LCA) that will be performed in Task 7.4 (to assess the 3 new end-consumer products environmental impacts compared to benchmark reference products already on the market), data from primary and secondary sources will be used to build the inventory to perform the LCA study. For each life cycle step, we will collect information to assess the elementary flow, defined as “material or energy entering the system being studied that has been drawn from the environment without previous human transformation, or material or energy leaving the system being studied that is released into the environment without subsequent human transformation”. Therefore, the following data will be collected at least:

- Consumption of resources (biotic and abiotic), including energy consumption,
- Air emissions,
- Emissions to water,
- Ground emissions, and
- Waste.

Data sources will include at least:

- Data provided by the companies through interviews or surveys. This source will be preferred when possible since it provides more reliable and closer to real data, and
- Data from the literature, scientific reports and patents, as well as other LCA studies on similar topics.

This data will be elaborated according to ISO14040 guidelines with the help of specific software to perform the final calculation.

### Online survey data/metadata

A customer survey will be produced based on a Multi-Choice Experiment model, with at least 10 000 people interviewed across Europe, to understand and analyse how consumers react to these new enzymes and the new products and if they are willing to appreciate them and buy more sustainable enzyme-based products.

Participants will be asked to compare multiple products and select the most attractive cost, added-value, the willingness to pay for a greening solution by responding to simple multi-choice questions through the Qualtrics platform. The survey will collect personal data, including the respondents' contact information (name, address, phone number, and email address), using the Qualtrics survey platform. As part of the survey, additional sensitive data such as gender, income level, and consumer habits will be collected to evaluate whether these factors affect consumers' reactions to the new enzyme-based products.

The personal data management and the collection of all questionnaire replies will be entrusted to Qualtrics, an ISO 27001 certified and FedRAMP authorised company. They will manage Personal Data following their Privacy Statement, updated on June 3, 2020, compliant with the GDPR (General Data Protection Regulation). Qualtrics will provide a final report collecting all the questionnaire replies, further analysed to meet the study objective. However, FuturEnzyme won't receive any respondents' contact information, and the replies will remain totally anonymous for Qualtrics' customers (ITB in the case of the projects).

### Product testing survey data/metadata

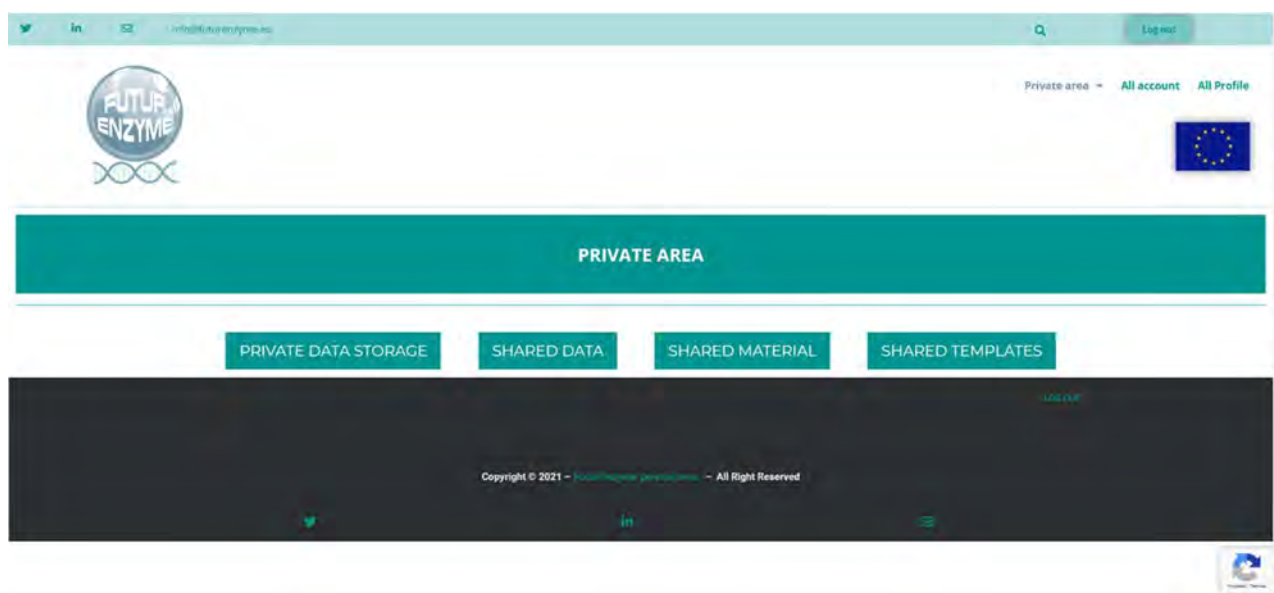
An ITB subcontractor will perform the product testing activity with consolidated experience in tests and investigations on products and services. At least 100 consumers will be selected from the subcontractor's database to participate in a product testing activity which means that participants will receive a sample of the final products and test them or take part in focus groups to ascertain consumer reactions to the project's products. Consumers will also be asked to complete a set of questions/tests/proofs to give their feedback on the new product. This activity will also involve personal data collection (mainly contact information), performed according to the GDPR rules and ethical principles. The data will be collected, stored, and managed entirely by the subcontractor, who will provide a final report with the main results at the end of the test. Personal data, if collected, will remain confidential and will not be shared with ITB or other project partners.

## 3. FAIR data: decision-taken strategy for publishing, repositories and databases

The information, knowledge and data, described in Section 2, that is going to be generated in the frame of the project has to be handled in order to accomplish FAIR (findable, accessible, interoperable and reusable) and RRI (Responsible Research and Innovation) principles by following the procedures described in here.

There will be 3 main repositories/databases for managing FuturEnzyme's data, metadata and information:

- The private area (**Figure 1**) of the project's website (login in at [www.futureenzyme.eu](http://www.futureenzyme.eu), restricted access to the consortium members), where files will be uploaded via the project manager and website developer.
- The repository created in MareNostrum 5 Supercomputer (BSC) for FuturEnzyme (ongoing). The idea of this repository will be double. First, to serve as backup for all the information to be deposited in the private area of the project's website. Second, to have a space to store information related to the project. The type of information/data would be sequencing, computational, biochemical, structure, experiment, etc. Each partner will have at its disposal 1-2 TB, with a total of 15 TB for the full consortium. The files will initially be uploaded via the project manager, but the decision to give access to manage the documents to the consortium members will be further discussed.
- The Zenodo community FuturEnzyme (<https://zenodo.org/communities/futureenzyme>). The consortium members will upload the files which will be curated by the project manager.



**Figure 1.** Private area of the FuturEnzyme’s webpage. Four sections have been created to organise the material to be uploaded. In the menu, pages containing information on the partners’ uploading have been included. The access is exclusive for FuturEnzyme members.

More in detail, all the documents of relevance to be made FAIR will be handled over to the project manager, who will assign them a QR code (the QR code system is extensively described in Deliverable D4.1) and keep a copy. Then, they will be made available in the private area of the project’s website, the MareNostrum 5 repository (BSC), and/or Zenodo, together with their corresponding QR code (unless another identification, such as doi number, is already available). The partners will use the QR code to label the corresponding physical material produced (tubes, plates, etc.), if applicable, so now and in the future, any person who uses the material can access the related information. Each partner will keep a record with a list of all their QR codes and files, and the project manager will keep a list with the whole consortium’s. With this system, the traceability of the material is ensured, and heterogeneous labelling issues are avoided.

A list of the deliverables that imply documents to be uploaded to the private area of the FuturEnzyme’s website and/or MareNostrum’s repository are detailed in **Table 2**. A section inside this private area (*Private data storage*) has been set up to include (via the project manager) those files containing information and data that the partners want to upload but that they decide not to make visible for other consortium members (e.g. not fully characterised enzyme, preliminary results, etc.). Other partners can see that a file has been added, but cannot see the contained information (a password will protect the file). Once the authors decide to share such files with all the consortium members, they will be transferred to the *Shared data* section.

**Table 2.** List of deliverables for which data and metadata need to be uploaded into the private area of the the FuturEnzyme’s website.

Deliverable number	Deliverable title	Type	Dissemination level
D2.2	Set of 250,000 sequences pre-selected	Other	Confidential <sup>1</sup>
D2.3	Set of 1,000 enzymes selected using motif screens	Other	Confidential <sup>1</sup>
D2.4	Set of 180 enzymes for experimental focus	Other	Confidential <sup>1</sup>
D3.1	Bio-resources prepared and exchanged	Other	Confidential <sup>1</sup>

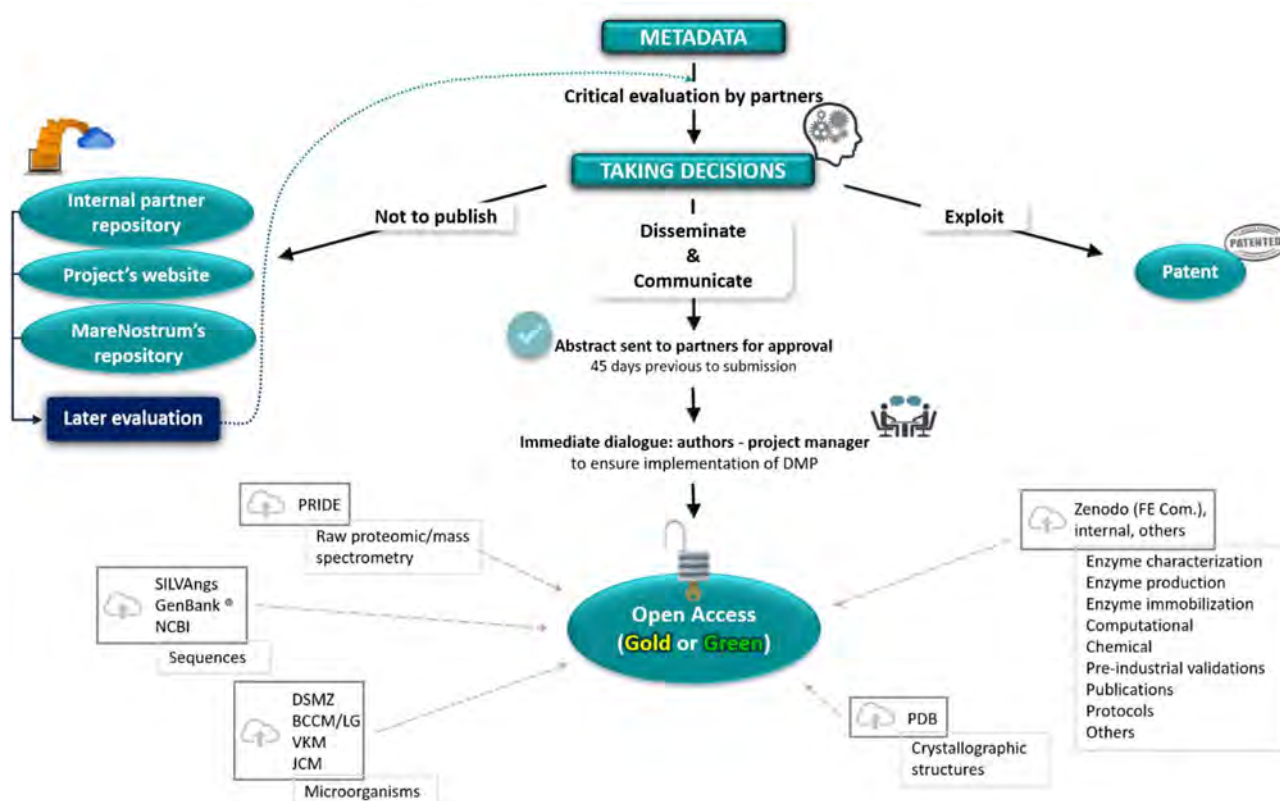
<b>D3.3</b>	Set of 100 best clones, 10 isolates, and 10 enzymes shortlisted for sequencing or transfer to WP2	Other	Confidential <sup>1</sup>
<b>D3.4</b>	Sequence, activity, and stability datasets from best positive bioresources	Data sets, microdata, etc.	Confidential <sup>1</sup>
<b>D3.5</b>	Set of new bioresources to screen or sequence	Other	Confidential <sup>1</sup>
<b>D3.6</b>	Complete set of positive naïve screened enzymes and sequences and their datasets	Report	Confidential <sup>1</sup>
<b>D4.2</b>	The FuturEnzyme Portfolio of 1,000 enzyme (recombinant/ native/biomimetic) material, obtained	Other	Confidential <sup>1</sup>
<b>D4.5</b>	At least 9 enzyme crystal structures	Other	Confidential <sup>1</sup>
<b>D4.6</b>	The metadata on expression yield, activity and stability, available	Data sets, microdata, etc.	Confidential <sup>1</sup>
<b>D4.7</b>	At least 180 enzymes (recombinant, native, biomimetic) with attractive properties, available	Other	Confidential <sup>1</sup>
<b>D5.1</b>	The shortlist of at least 18 enzymes nominated for engineering	Report	Confidential <sup>1</sup>
<b>D5.2</b>	Set of 18 mutants generated by genetic engineering	Other	Confidential <sup>1</sup>
<b>D5.3</b>	Set of 4 PluriZymes with single activities	Other	Confidential <sup>1</sup>
<b>D5.4</b>	Set of 3 multi-purpose PluriZymes	Other	Confidential <sup>1</sup>
<b>D5.5</b>	Set of 18 improved enzymes by supramolecular engineering	Other	Confidential <sup>1</sup>
<b>D5.6</b>	Datasets of engineered variants	Data sets, microdata, etc.	Confidential <sup>1</sup>
<b>D6.2</b>	Report on fermentation, DSP and activity verification for 18 PreLead enzymes	Report	Confidential <sup>1</sup>
<b>D6.3</b>	Best 9 Lead Enzyme Materials obtained at multi-gram/kg scale for real-life tests	Other	Confidential <sup>1</sup>
<b>D6.4</b>	Report on fermentation, optimisation and verification for 9 Lead Enzyme Materials	Report	Confidential <sup>1</sup>
<b>D7.1</b>	Report on small/ medium validation trials of 18 best preselected enzymes	Report	Confidential <sup>1</sup>
<b>D8.13</b>	Two divulgative/ promotional articles in scientific magazines	Websites, patents filling, etc.	Public

<sup>1</sup>Only for members of the consortium (including the Commission Services).



## Decision-taken strategy to publish and to manage the data and metadata

When a whole-sense pool of data is obtained by one/several partner/s, the decision to disseminate and communicate or protect the results has to be taken (**Figure 2**). If publication is chosen, open access (OA) has to be guaranteed (preferably gold, green is possible) in order to accomplish ORD, following the principle of “as open as possible, as closed as necessary”. In this case, the article will always be uploaded to the corresponding repository (see below for further details), even when gold access is selected.



**Figure 2.** Diagram followed after whole-sense research and data are achieved to decide the end of the acquired know-how. Adapted from H2020 online manual.

Publications and files containing information that supports such publications will be properly uploaded to the Zenodo's public repository ([www.zenodo.org](http://www.zenodo.org)), and FuturEnzyme's Community internal repository (selecting the type of data, title, authors, grant number, type of access, and all the information requested that is available). The Zenodo's community has already been created (<https://zenodo.org/communities/futureenzyme/>; identifier: futureenzyme). This makes it easy and without charge for anyone to access the whole pool of information, if/when it is embargo free. The FuturEnzyme's project manager will revise and accept/cure/ask for modifications to the partners who upload material (<https://zenodo.org/deposit/new?c=futureenzyme>). Other categories that can be uploaded to Zenodo when appropriate are: poster, presentation, dataset, image, video, software, lesson, and other. The access rights that can be chosen to match the different needs of the researchers who upload the material at any moment are: open access, closed access (access only by owner), embargoed, and restricted. All uploaded data is given a digital object identifier (DOI) number, and the researcher has to include keywords to optimise possibilities for reusing the information. Besides, Zenodo allows for Creative Commons license. In addition to the functionalities that Zenodo represents as public repository, this platform is integrated into reporting lines for research funded by the European Commission via OpenAIRE ([www.openaire.eu](http://www.openaire.eu)) in collaboration with CERN ("Conseil Européen pour la Recherche Nucléaire", or European Council for Nuclear Research, Geneva, Switzerland). The objective of this infrastructure is to support, boost and measure the correct application of the European policies regarding open access to scientific publications and data.

Other preprint servers will be also used for Open Access publications; they include servers such as bioRxiv (<https://www.biorxiv.org/>), ChemRxiv (<https://chemrxiv.org/engage/chemrxiv/public-dashboard>), MetaArXiv (<https://osf.io/preprints/metaarxiv/>). The scripts used in the published paper will be made available at appropriated server such as *figshare* (<https://figshare.com>), to aid in reproducing the analyses. Raw mass spectrometry (proteomics and metabolomics), sequence and structural datasets, will be made available in MetaboLights (<https://www.ebi.ac.uk/metabolights/>), ProteomeXchange Consortium via the PRIDE (<https://www.ebi.ac.uk/pride/>), SILVAngs platform (<https://ngs.arb-silva.de/silvangs/>), GenBank® (<https://www.ncbi.nlm.nih.gov/guide/all/>), NCBI BioProject (<https://www.ncbi.nlm.nih.gov/bioproject>) and Protein Data Bank (PDB) (<https://deposit-pdbe.wwpdb.org/deposition>, for depositing the coordinates, and <https://www.rcsb.org>, once those are released) repositories, to cite most significant ones. Isolated microbes will be deposited in international culture collections, such as DSMZ (<https://www.dsmz.de/>), BCCM/LG (<https://bccm.belspo.be/about-us/bccm-lmg>), VKM (<http://www.vkm.ru/>), or JCM ([https://jcm.brc.riken.jp/en/depositing\\_e](https://jcm.brc.riken.jp/en/depositing_e)) collections.

The data generated from experimental and computational work will show the performance of the enzymes/microorganisms in terms of identification and characterisation following the templates created for the project (see Annex). These documents will be filled with as much information as available and be as complete as possible, always including the researcher's identification in charge of the experimental work. This information will be uploaded to the FuturEnzyme's website private area, the MareNostrum 5 repository (BSC), and/or Zenodo, depending on the needs and decisions taken by the authors. When possible, protocols will be made publicly available (Open Access) through open repositories (preprint server) of community-contributed protocols, such as the one sponsored by Nature Portfolio (<https://protocolexchange.researchsquare.com/>). Results from the Life Cycle Assessments (LCA), bibliographic (academic and patent), consumer and market evaluations, etc. produced in the frame of FuturEnzyme will also be handled to the project manager to be linked to a QR code. In all cases, the identifiers relative to the deposited datasets will be collected. Those will be available to partners in the FuturEnzyme's web ([www.futureenzyme.eu](http://www.futureenzyme.eu)) private area.

In all cases, every time an entry is uploaded to any repository, the name of the action, acronym and grant number, as well as the DOIs and permanent identifiers, have to be added, so they can be easily sent to the national and European agencies upon request at any time.

All partners are free (and encouraged) to include publications and data also in public repositories from their institutions (e.g. Digital CSIC, <https://digital.csic.es/>).

Regarding data that is not meant to be published nor protected, it is the researcher's decision whether to upload it in the repositories or not. It is strongly recommended to do so in order to achieve FAIR and RRI principles.

With this strategy, the project results will meet Open Access and Open Science status.

All the data uploaded to FuturEnzyme's Zenodo Community will be permanently accessible and managed by the project manager at least until the end of the project. The material uploaded to the project's website and MareNostrum 5 repository for our project will be available and managed as far as the sites are maintained (at least five more years after the end of the project for FuturEnzyme's website, permanently for MareNostrum 5). After the end of the project, a copy of the material will be kept at CSIC's facilities by the Coordinator (Manuel Ferrer).

#### File extensions to be used for data and metadata storage

Scientific datasets from measurement tests/computational analysis will be stored in the original format of the file and in spreadsheet (\*.xlsx), word processor (\*.docx, \*.txt) or a similar format of common use.

Posters, leaflets, brochures, reports, presentations or any other similar material will be stored produced in Portable Document Format file (\*.pdf).

Videos will be created as .mp4 file.

Related files that need to be used together will be compressed (\*.zip), including a Readme.txt describing in English the content and how to retrieve the information.

Standardisation: making the bio-resources, data and metadata homogeneous and comparable

#### Files names

When creating a file for the work done/to be done, there is always a discrepancy in how to name it, moreover when people from different groups and even countries are involved. For this reason, all the files produced by a member of the consortium in the context of FuturEnzyme activities that will be uploaded to any of the repositories will be named following this pattern:

FE\_Partner name\_Deliverable/Task number\_Type of material/data\_Name\_Subname\_Version number

Example 1: FE\_CSIC\_D1.1\_Enzyme\_EH25\_mutb\_v1

Example 2: FE\_ITB\_T4.1\_LCA\_Detergent benchmarking\_sample 1\_v2

This procedure standardises and unifies the designation of the files, helping to find and track any information required.

#### Protocols

All the protocols to be used within FuturEnzyme are fully described in the document for deliverable D3.2. With this strategy, we avoid discrepancies that can occur because of different procedures employed, giving rise to confusion in the data.

#### Analysis programs

It is also common that in the different laboratories composing a consortium, and they use distinct programs to analyse the same type of data. This is why we will unify the software to be used when an activity is carried out in several laboratories within the consortium. With this strategy, we avoid discrepancies that can occur because of differential management of the data.

- Kinetic fits and calculations: GraphPad Prism ([www.graphpad.com/](http://www.graphpad.com/))
- Enzyme (specific) activity calculation: SOFTmax PRO (Device-specific software, [www.moleculardevices.com/products/microplate-readers/acquisition-and-analysis-software/softmax-pro-software](http://www.moleculardevices.com/products/microplate-readers/acquisition-and-analysis-software/softmax-pro-software)) & MS Excel spreadsheet.
- Apparent protein melting temperature: NanoTemper PR.Stability Analysis Software ([www.nanotempertech.com/prometheus-pr-stability-analysis-software/](http://www.nanotempertech.com/prometheus-pr-stability-analysis-software/))
- The collection of data concerning enzyme activity, fermentation yield, parameters for growth of the different microorganisms, condition for storage of microorganisms (ADD OTHER), will be based on formats (operating manuals) agreed and shared by the FuturEnzyme partners. These formats will be stored on an easy to access platform and will allow a uniform system for the collection of data and an immediate interpretation of the collected data by all of the interested partners.
- Etc.

#### Quality controls

Quality control of the deposited datasets will be guaranteed. For example, in case of protein structures deposited, before they can be deposited, the system performs an exhaustive analysis of the quality of the

experiment, with technical criteria of crystallographic analysis and control of the protein geometry. Small mismatches are retouched or corrected and resubmitted. When everything is correct, they are assigned a PDB code, but they can remain inaccessible for a maximum of 1 year or until they are published. Note that the PDB itself analyses the structure and provides a document called "Validation Report" with all the analysis.

When data/publications are uploaded to Zenodo, a curation is required before they are published, which will be carried out by the project manager. With this we ensure that not anyone can upload whatever to FuturEnzyme's community, and that everything that appears within it is legitimate to the project.

#### 4. Allocation of resources for data and metadata open access

Zenodo repository (including Creative Commons) and other institutional repositories chosen by the partners are free of charge (as detailed above, in Section 3). On the other hand, publication in green or gold open access peer-review journals has Article Processing Charges (APC). This costs that are eligible in Horizon 2020 programme as stated in the Grant Agreement Article 6, specifically in section 2.D.3: "Costs of other goods and services ... are eligible ....include, for instance, consumables and supplies, dissemination (including open access), protection of results, certificates on the financial statements (if they are required by the Agreement), certificates on the methodology, translations and publications." Such costs will be covered by the partner responsible of the information to be published. Table 3.4b of the Grant Agreement specifies the budget destined for open access publications and articles in non-scientific periodicals magazines dedicated to the dissemination of knowledge for scientific and non-scientific community. Whenever multiple entities are involved in the same publication, the costs can be distributed as all the parties convene. In this table, it is also specified the cost intended for the project's website creation and maintenance is also specified.

#### 5. Data security

##### Confidentiality

In order to preserve the confidentiality of the results obtained in the project, every partner will decide when to make them public (following the process depicted in **Figure 2**). The documents can be uploaded without being made public both to the private area of the project's web (via the project manager and website developer, the latter under non-disclosure agreement) and to the FuturEnzyme's Zenodo Community. In the private area of the FuturEnzyme's web, a section for private storage is created as previously described to allow to every partner to upload their material and protect it with a password, that will be transferred when needed to the shared data section. Zenodo also allows for different types of access: open, closed, embargoed or restricted. Regarding MareNostrum's repository, it will only be accessible to the consortium members, and the project manager will upload/manage the documents.

Since QR codes will also be generated for the datasets, lists of bioresources, etc., those documents that need to be protected will be locked by a password, only known by the project manager and the responsible of the material or the whole consortium, depending on the specific needs.

The consortium members responsible for each file, dataset, etc. will store them in their own devices and regularly made security copies.

Besides, the project manager will keep a copy of all the documents generated by the partners that will be uploaded to the project's web and/or to Zenodo. This copy will be physically stored in a hard drive, and security copies will be made regularly in two more devices.

## GDPR

As mentioned above, the online customer survey will be produced by Qualtrics, an ISO 27001 certified and FedRAMP authorised company, who will be responsible for the personal data management and the collection of all questionnaire replies. They will manage Personal Data following their Privacy Statement (June 3, 2020) compliant with the GDPR (General Data Protection Regulation). FuturEnzyme won't receive any respondents' contact information, and the replies will remain totally anonymous for Qualtrics customers (ITB in the case of the projects).

The product testing survey will be carried out by a subcontractor with consolidated experience in tests and investigations on products and services. The personal data collected (mainly contact information) will be performed according to the GDPR rules and ethical principles and will not be handled to FuturEnzyme.

To ensure GDPR compliance, a Data Protection Officer has been appointed in case of conflicts from every institution (see **Table 3**). In addition to that, in any case, any of the members of this consortium may withdraw their consent or exercise their rights of access, rectification, deletion, limitation of processing or opposition to it, as well as the right to data portability. These requests must also be sent to the Data Protection Officers (see **Table 3**).

Further information on personal data management and GDPR compliance will be included in D9.2 as described below.

**Table 3.** Contact/link of the Data Protection Officer from the institutions conforming FuturEnzyme.

PARTNER	NAME	EMAIL
<b>CSIC</b>	José López Calvo	jose.lopez.calvo@csic.es; delegadoprotecciondatos@csic.es
<b>BSC</b>	-	dpo@bsc.es
<b>BANGOR</b>	Gwenan Hine	gwenan.hine@bangor.ac.uk
<b>UHAM</b>	Dr. Stefan Thiemann	stefan.thiemann@uni-hamburg.de
<b>UDUS</b>	Dr. Ursula Hilgers	datenschutz@hhu.de
<b>IST-ID</b>	Dr. Tiago Silva Abade	rgpd@ulisboa.pt
<b>CNR</b>	Ing. Roberto Puccinelli Dr. Massimo Virgili	rpd@cnr.it massimo.virgili@cnr.it
<b>ITB</b>	Lanfranco Masotti	presidenza@italbiotec.it
<b>FHNW</b>	Karin Hiltwein	Karin.hiltwein@fhnw.ch
<b>CLIB</b>	Dennis Herzberg	herzberg@clib-cluster.de
<b>INOFEA</b>	Anne Timm	anne.timm@inofea.com
<b>BIOC_CHEM</b>	Fabrizio Beltrametti	fbeltrametti@bioc-chemsolutions.com
<b>SCHOELLER</b>	Eveline Scheidegger	eveline_scheidegger@schoeller-textiles.com
<b>HENKEL</b>	-	<a href="https://www.henkel.com/data-protection-statement#pageID=9388">https://www.henkel.com/data-protection-statement#pageID=9388</a>
<b>EVONIK</b>	-	privacy-policy@evonik.com
<b>EUCODIS</b>	-	office@eucodis.com

## 6. Ethical aspects

In order to guarantee ethical considerations in our research, the ethical aspects detailed in **Table 4** will be guaranteed, an extensive description of which will be delivered at month 6 (Deliverables D9.1-D9.4).

**Table 4.** Ethics deliverables.

Delivable number	Deliverable title	WP number	Lead beneficiary	Type	Dissemination level	Due date (in months)
D9.1	H - Requirement No. 1	WP9	1-CSIC	The procedures and criteria that will be used to identify/recruit research participants must be submitted as a deliverable. - The informed consent procedures that will be implemented for the participation of humans and in regard to data processing must be submitted as a deliverable. - Templates of the informed consent/assent forms and information sheets covering the voluntary participation and data protection issues (in language and terms intelligible to the participants) must be kept on file (to be specified in the grant agreement) and the English version must be submitted as a deliverable. - The applicant must clarify whether children and/or adults unable to give informed consent will be involved and, if so, justification for their participation must be submitted as a deliverable. - In case children and/or adults unable to give informed consent are involved, details on how the consent of the legal representatives (and assent, when applicable) will be acquired must be submitted as a deliverable. - If applicable, the applicant must clarify whether invasive physical procedures will be used and what measures will be taken to minimise possible pain, and submit this information as a deliverable. - If applicable, details on incidental findings policy must be submitted as a deliverable. - If applicable, for each clinical study, the following documents/information must be submitted as a deliverable (in one package) prior to enrolment of first study subject: (i) Final version of study protocol as submitted to regulators/ethics committee(s), (ii) Registration number of clinical study in a WHO-or ICMJE- approved registry (with the possibility to post	Ethics	Confidential, only for members of the consortium (including the Commission Services)

				results), (iii) Approvals (ethics committees and national competent authority if applicable) required for invitation/enrolment of first subject in at least one clinical centre. - If applicable, for each clinical study, a report on the status of posting results in the study registry(s) must be submitted as a deliverable, including timelines if/when final posting of results is scheduled after end of funding period		
	Specific sub-deliverables H - Requirement No. 1			<ul style="list-style-type: none"> <li>• The procedures and measures to analyse in vivo efficacy properties of cosmetics, that prove all tests are performed in a way that no pain or harm can occur</li> </ul>		
				<ul style="list-style-type: none"> <li>• The procedures and criteria for recruiting and processing informed consent of the participants in the research survey and product testing</li> </ul>		
				<ul style="list-style-type: none"> <li>• Templates of the informed consent/assent forms and information sheets covering the voluntary participation and data protection issues (in language and terms intelligible to the participants)</li> </ul>		
				<ul style="list-style-type: none"> <li>• A copy of the Agreement signed by Community of Madrid (General Directorate of Education) and the CSIC, that guarantees that training and teaching activities follow ethics requirements</li> </ul>		

D9.2	POPD - Requirement No. 2	WP9	1-CSIC	<p>The beneficiary must check if special derogations pertaining to the rights of data subjects or the processing of genetic, biometric and/or health data have been established under the national legislation of the country where the research takes place and submit a declaration of compliance with respective national legal framework(s). - Each beneficiary must confirm that a Data Protection Officer (DPO) has been appointed and the contact details of their DPO are made available to all data subjects involved in the research. For beneficiaries not required to appoint a DPO under the General Data Protection Regulation (GDPR), a detailed data protection policy for the project must be kept on file (to be specified in the grant agreement) and submitted to the Agency upon request. The confirmation for each beneficiary must be submitted as a deliverable. - Justification for the processing of sensitive personal data must be submitted as a deliverable. - Each beneficiary must explain how all of the data they intend to process is relevant and limited to the purposes of the research project (in accordance with the 'data minimisation' principle). This must be submitted as a deliverable. - A description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants must be submitted as a deliverable. - A description of the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing must be submitted as a deliverable. - Description of the anonymisation/pseudonymisation techniques that will be implemented must be submitted as a deliverable. - In case personal data are transferred from the EU to a non-EU country or international organisation, confirmation that such transfers are in accordance with Chapter V of the General Data Protection Regulation 2016/679, must be submitted as a deliverable. - In case personal data</p>	Ethics	Confidential, only for members of the consortium (including the Commission Services)
------	--------------------------	-----	--------	--	--------	--



				are transferred from a non-EU country to the EU (or another third state), confirmation that such transfers comply with the laws of the country in which the data was collected must be submitted as a deliverable. - In case the research involves profiling, the beneficiary must provide explanation how the data subjects will be informed of the existence of the profiling, its possible consequences and how their fundamental rights will be safeguarded. This must be submitted as a deliverable. - An explicit confirmation that the data used in the project is publicly available and can be freely used for the purposes of the project must be submitted as a deliverable. - In case of further processing of previously collected personal data, an explicit confirmation that the beneficiary has lawful basis for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects must be submitted as a deliverable		
	Specific sub-deliverables H - Requirement No. 2			<ul style="list-style-type: none"> <li>• Justification for the processing of sensitive personal data</li> </ul>		
				<ul style="list-style-type: none"> <li>• Explanation how all of the data to process is relevant and limited to the purposes of the research project (in accordance with the 'data minimisation' principle)</li> </ul>		
				<ul style="list-style-type: none"> <li>• An explicit confirmation that the data used in the project is publicly available and can be freely used for the purposes of the project</li> </ul>		
				<ul style="list-style-type: none"> <li>• The technical and organisational measures to safeguard the rights and freedoms of the data subjects/research participants</li> </ul>		
				<ul style="list-style-type: none"> <li>• The security measures to prevent unauthorised access to personal data</li> </ul>		
				<ul style="list-style-type: none"> <li>• Description of the anonymisation/pseudonymisation techniques that will be implemented for data processing</li> </ul>		

				<ul style="list-style-type: none"> <li>• Contact details of the appointed Data Protection Officer for each partner, and their confirmations</li> </ul>		
				<ul style="list-style-type: none"> <li>• For beneficiaries not required to appoint a DPO under the General Data Protection Regulation (GDPR), a detailed data protection policy for the project</li> </ul>		
				<ul style="list-style-type: none"> <li>• In case personal data are transferred from the EU to a non-EU country, confirmation that such transfers are in accordance with Chapter V of the General Data Protection Regulation 2016/679.</li> </ul>		
D9.3	NEC - Requirement No. 3	WP9	1-CSIC	<p>If applicable, detailed information to demonstrate that fair benefit-sharing arrangements with stakeholders from low and lower-middle income countries are ensured must be submitted as a deliverable. - Details on the materials which will be imported to/exported from the EU must be submitted as a deliverable. - Copies of import/export authorisations, as required by national/EU legislation must be kept on file (to be specified in the grant agreement) and submitted to the Agency upon request</p>	Ethics	Confidential, only for members of the consortium (including the Commission Services)
	Specific sub-deliverables H - Requirement No. 3			<ul style="list-style-type: none"> <li>• Details on the materials which will be imported to/exported from the EU</li> </ul>		
				<ul style="list-style-type: none"> <li>• Copies of import/export authorisations, as required by national/EU legislation</li> </ul>		
				<ul style="list-style-type: none"> <li>• In case activities undertaken in non-EU countries raise ethics issues, to ensure that the research conducted outside the EU is legal in at least one EU Member State</li> </ul>		
D9.4	EPQ - Requirement No. 4	WP9	1-CSIC	<p>Further information about the possible harm to the environment caused by the research and the measures that will be taken to mitigate the risks must be submitted as a deliverable. - Copies of authorisations for relevant facilities (e.g., security classification of laboratory, GMO authorisation) must be kept on file (to be specified in the grant agreement) and submitted to the Agency upon request. - The applicant must demonstrate that appropriate health and safety</p>	Ethics	Confidential, only for members of the consortium (including the Commission Services)

				procedures conforming to relevant local/national guidelines/legislation are followed for staff involved in this project. This information must be submitted as a deliverable. - If applicable, details on the endangered species and/or protected areas involved in the research, and the measures to minimise the impact of the activities must be submitted as a deliverable		
	Specific sub-deliverables H - Requirement No. 4			<ul style="list-style-type: none"> <li>• Material transfer agreements for biotic resource exchange</li> </ul>		
				<ul style="list-style-type: none"> <li>• Copies of authorisations for relevant facilities (e.g., security classification of laboratory, GMO authorisation)</li> </ul>		
				<ul style="list-style-type: none"> <li>• Information of health and safety procedures conforming to relevant local/national guidelines/legislation, followed for staff involved in this project</li> </ul>		

The consortium has at its disposal the ethic committees of all the members institution, whose contacts can be found in **Table 5**.

**Table 5.** Contact/link of the responsible of the Ethics Committee from the institutions conforming FuturEnzyme.

PARTNER	NAME	EMAIL
CSIC	D. Lluís Montoliu José	montoliu@cnb.csic.es
BSC	Simona Giardina <sup>1</sup>	simona.giardina@bsc.es
BANGOR	Gwenan Hine	gwenan.hine@bangor.ac.uk
UHAM	-	-
UDUS	Prof. Dr. med. Thomas Hohlfeld	ethikkommission@med.uni-duesseldorf.de
IST-ID	-	comissaoetica@tecnico.ulisboa.pt
CNR	Dr. Cinzia Caporale	cnr.ethics@cnr.it
ITB	Lanfranco Masotti	presidenza@italbiotec.it
FHNW	Karin Hiltwein	Karin.hiltwein@fhnw.ch
CLIB	Dennis Herzberg	herzberg@clib-cluster.de
INOFEA	Anne Timm	anne.timm@inofea.com
BIOC_CHEM	Fabrizio Beltrametti	fbeltrametti@bioc-chemsolutions.com
SCHOELLER	Eveline Scheidegger	eveline_scheidegger@schoeller-textiles.com
HENKEL	-	<a href="https://www.henkel.com/sustainability/positions/white-biotechnology">https://www.henkel.com/sustainability/positions/white-biotechnology</a>
EVONIK	-	compliance-officer@evonik.com
EUCODIS	-	-

<sup>1</sup>Secretary of the BSC Internal Board of Revision of projects, in collaboration with BSC experts and legal and DPO departments.

## 7. Innovation and Intellectual Property Rights Management Strategy

FuturEnzyme partners are aware that the rules for participation state clear obligations for beneficiaries:

- "Each participant that has received European Union funding shall use its **best efforts to exploit** the results it owns, or to have them exploited by another legal entity..."
- "... participants shall provide any **information on their exploitation** and dissemination related activities, and provide any documents necessary in accordance with the conditions laid down in the grant agreement", and
- Grant Agreement, Article 28.

In relation to the above, the Grant Agreement states the following Obligations:

28.1 Obligation to exploit the results. Each beneficiary must — up to four years after the period set out in Article 3 — take measures aiming to ensure ‘exploitation’ of its results [...]

- using them in further **research** activities (outside the action),
- developing, creating or marketing a **product** or **process**,
- creating and providing a **service**, and
- using them in **standardisation** activities.

In relation to the above, the DoA states the following Obligations in the WP 8 - Communication, Dissemination and Exploitation (Lead ITB)

“Exploit FuturEnzyme business potential to possible licensees and create the basis for the long-term impacts of the project by a multi-channel platform (including Open-Access, Science and Innovation). Exploit innovation, industrial implementation and market uptake considering gender, rights and ethical issues inside and outside the consortium.”

Task 8.5: Maximising exploitation of project enzymes, products, and knowledge

- Exploitation [...] is a major commitment of the project
- Exploitation and innovation coordinators
  - (T. Schwabe-Marković, CLIB; M. Ferrer, CSIC)
- Exploitation and Innovation Plan (M12, M48)

**To ensure this, an Exploitation and Innovation Plan will be delivered, covering at least:**

- Innovation and IPR Management Strategy
- Existing patents
- Project’s assets and their main potential exploitation routes
- Target groups (per project asset) - potential end-users of results
- Joint exploitation plans
- Means and procedures for exploitation of assets
- Clear action plan towards exploitation of results

**Industrial, manufacturer and SME partners will contribute also to the Exploitation and Innovation Plan, as follows:**

- Industrial partners
  - define an exploitation plan
    - establish directions for work in different WPs
    - facilitate the identification of inventions with market potentials.
- Manufacturers and SMEs
  - contribute definition of market aspects

- Specific directives to direct work in RTD-oriented WPs → coordinate market aspects and research aspects.

Although, an extensive Exploitation and Innovation Plan will be delivered at month 12 (D8.6 Preliminary Exploitation Plan), we would like to mention the following actions to guarantee the managing of the data and metadata.

A declaration of invention/work form (see Annex) will be filled for every data/metadata that after preliminary evaluation by the researcher generating the data is selected for patenting or for evaluating transfer possibilities. This document will state the precise involvement of the consortium member and of external authors if any. With this strategy, possible conflict for authorship will be avoided. This will lay the groundwork for a potential exploitation of the results by a project or external partner. It will also serve to clearly determine the origin of any product that may go to the market, so the consumers can have at their disposal all the needed information about its development process. Once a real possibility for transfer or patenting of a promising enzyme candidate or process is decided by the researchers implicated, its patentability will be evaluated. For that, the coordinator (CSIC) will support partners, if needed, in the worldwide search of prior art documents relating: i) Selection of the most relevant documents that might affect the patentability; ii) analysis of the patentability depending on the selected relevant documents; and iii) advice concerning filing patent applications.

To help in exploitation issues, the consortium can count on the help of their institution's technology transfer and intellectual property rights departments, as also determined in deliverable 8.3, and whose contacts can be found in **Table 6**. The Dissemination, Communication and IP Task Force, coordinated by ITB, will effectively coordinate internal IP relevant issues (besides dissemination and communication activities) and will ensure regularly liaise with the appointed contact persons.

**Table 6.** Contacts of the technology transfer departments.

<b>PARTNER</b>	<b>NAME</b>	<b>EMAIL</b>
<b>CSIC</b>	Sara Junco	s.junco@csic.es; patentes@csic.es
<b>BSC</b>	Anna Escoda	anna.escoda@bsc.es; techtransferoffice@bsc.es)
<b>BANGOR</b>	Gareth Mayhead	gmayhead@bangor.ac.uk
<b>UHAM</b>	Sabine Baars	sabine.baars@uni-hamburg.de
<b>UDUS</b>	Claudia Niemann (contact through Karl E. Jaeger or Stephan Thies)	niemann@zuv.hhu.de
<b>IST-ID</b>	Patrícia Lima	pi@tecnico.ulisboa.pt
<b>CNR</b>	-	segretaria.uvr@cnr.it
<b>ITB</b>	Ilaria Re	ilaria.re@italbiotec.it
<b>FHNW</b>	Marion Rutsche	marion.rutsche@fhnw.ch
<b>CLIB</b>	Tatjana Schwabe	schwabe@clib-cluster.de
<b>INOFEA</b>	Rita Corroero	rita.corroero@inofea.com
<b>BIOC_CHEM</b>	Fabrizio Beltrametti	fbeltrametti@bioc-chemsolutions.com
<b>SCHOELLER</b>	Rainer Roesch	rainer_roesch@schoeller-textiles.com
<b>HENKEL</b>	Sarah Dieckmann	vtp-in.vtp@henkel.com
<b>EVONIK</b>	Meinolf Boettcher	meinolf.boettcher@evonik.com
<b>EUCODIS</b>	Jan Modregger	modregger@eucodis.com

## Annex

### Template for new bioresources



Update Enzyme - New bioresources / Enzymes

ID enzyme	Amino acid sequence or accession number (if public)	Source (microbe, environmental, etc.)	Screen system (homology or functional)	GPS coordinate: Latitude	GPS coordinate: Longitude	Depth (m)	Taxonomic origin (if available)	Cloning or expression system (if available)	Activity	Name, laboratory, date



Insert Enzyme - New bioresources / Enzymes

Sample ID	Validated strain name	Sampling site	GPS coordinates	Taxonomic information	Status Genome (accession number of access link)	Enzyme activity (if tested)	Name, laboratory, date



Insert Enzyme - New bioresources / Antigenomic libraries

LIBRARIES site / material	Sampling date	Sample source (environmental, enrichment, strain, etc.)	GPS coordinates	Type of library	Size of the library (nr. clones and insert lenght)	Cultivation conditions	Name, laboratory, date



Update Enzyme - New bioresources / Enrichment libraries

Sample ID	Starting material	Sampling date	GPS coordinates	Enrichment (substrate, medium, conditions)	Taxonomic information of enriched microorganisms	Further comments	Potential enzyme active present in the enrichment (if known based on the substrate used)	Name, laboratory, date



Update Enzyme - New bioresources / Shotgun metagenomic sequences

Sample description	GPS coordinates	BioSample/BioProject (or link to access the data)	Sequencing method	Name, laboratory, date

## Template for Microorganism datasets

MICROORGANISM DATASET	
QR	
<div>Sample number: Sample name: Author: Institution: Partner number: Date:</div>	
MICROORGANISM IDENTIFICATION	
Microorganism ID:	
Taxonomic information:	
GPS coordinates, location:	
Environmental sample:	
Date of sampling:	
16S RNA sequence or accession number:	
Glycerol or culture QR code:	
CULTIVATION CONDITIONS	
Medium:	
pH:	
Temperature:	
Stirring:	
Other conditions:	
GENOME (IF SEQUENCED)	
Accession number (if public):	
Access link (if non-public):	



## MICROORGANISM DATASET

### ACTIVITIES FOUND (IF SCREENED)

Detailed method and substrate for screening:

Lipase:

Esterase:

Amylase:

Protease:

Hyaluronidase:

Cutinase and related polymer-degrading enzyme:

Oxido-reductase:

Others:



## Template for enzyme and enzyme-immobilized datasets



QR

Sample number:

Sample name:

Author:

Institution:

Partner number:

Date:

### PROTEIN IDENTIFICATION

Protein ID:

Source of enzyme (detailed):

EC number:

Sequence (or accession number if public):

Activity (class/type):

Mutations or any other modification (and method employed):

Other information of interest:

### CLONING, EXPRESSION, PURIFICATION AND PREPARATION

Producer microbe (if produced by a native host)

Organism:

Culture conditions:

Purification method:

Producer microbe (if produced by a heterologous host)

Organism:

Cloning and expression systems:

Induction:

Antibiotic resistance:

Purification method:

Other information of interest:

Glycerol QR code:



## ENZYME and ENZYME-IMMOBILIZED DATASET

### STORAGE CONDITIONS:

Stability buffer:

Storage temperature:

Other information of interest:

### IMMOBILIZATION CONDITIONS:

Detailed method description:

Carrier:

Particle concentration:

Activity:

Other information of interest:

### ASSAY CONDITIONS

Detailed method description:

Temperature:

Buffer and pH:

Substrate:

Product:

Cofactor:

Detection method:

Other information of interest:

### RESULTS: ACTIVITY DETERMINATION AND RAW DATASETS

Detailed method and conditions for calculation:

Specific activity (units/mg) (if calculated):

$V_{max}$  (if calculated):

$K_m$  (if calculated):

$k_{cat}$  (if calculated):

$k_{cat}/K_m$  (if calculated):

Optimal parameter for activity (i.e. pH and T):

Stability (T, pH, solvent, detergents, etc.):

Inhibitors:

Raw data for enzyme parameters determination (either raw data or link to access the data):

## Template for declaration of invention/work form

This form itself does not constitute a prior disclosure but is a confidential note, intended only for the authors, the project coordinator and exploitation task force of the project, which they will keep confidential (only available to the consortium and project officer). In case of an interest from industry in any of the exact enzymes or technologies found in the project the transfer or patenting will remain to be seen, as negotiating over IP might be considered, but this document at least paves the way for this to be possible.



### DECLARATION OF INVENTION/WORK FORM

Title:

Date:

#### 1. DESCRIPTION OF THE INVENTION/WORK

Describe briefly (maximum 1000 words approximately) the object of the invention/work with potential for technology transfer, what it consists of, what technical problem it solves and what advantages it provides with respect to the current state of the art and possible applications. If you need more space, you can include a separate document with the technical information.

Text...

#### 2. AUTHORS

Persons from the FuturEnzyme consortium involved in the generation of the invention/work

Last name, Name	Partner	e-mail	Implication/contribution

Necessary collaborators outside the FuturEnzyme consortium (if any)

Last name, Name	Entity	e-mail	Implication/contribution

#### 3. NOVELTY

Has the invention/work been previously disclosed in whole or in part?

NO ☐ YES ☐ If YES, how?

Has information on the invention/work been searched in patent databases?

NO ☐ YES ☐ If yes, which ones?

Please note that any prior disclosure, either by the inventors/researchers themselves or by a different author, may affect the novelty and inventive step of the invention. In case of prior



## DECLARATION OF INVENTION/WORK FORM

disclosure, it is requested to attach the corresponding information together with this document, regardless of the type of disclosure (oral/written communication at a congress or seminar; doctoral thesis; master's thesis; scientific article; reference manual; patents; utility models; specialized technical journals, commercial catalogs, etc.).

**Describe keywords to search for documents in the state of the art.**

Keywords:

## ORIGIN OF THE INVENTION/WORK

**Describe briefly (maximum 1000 words approximately) the origin of the enzyme, microbe, etc. that constitute the basis of the invention/work.**

If the obtaining of the results to be protected has involved the use of genetic resources, state:

- Country of origin of the genetic resource:
- Date of access to the genetic resource
- Conditions of access to the genetic resource: access *in situ* ☐ access *ex situ* ☐
- Have you applied for Authorization (if applicable)? NO ☐ YES ☐
- Include the Authorization number
- Any other information of relevance

## AUTHOR'S AGREEMENT

Through this document, the authors:

1. Agree with the authorship of the invention/work described.
2. Are committed to collaborate in the elaboration of the communication, dissemination and/or exploitation processes related with the in here described invention/work.
3. Will communicate to the FuturEnzyme's Coordinator/Project Manager any changes in the contact details to enable the necessary communication.
5. That the document has a purely informative character through which to inform the consortium about the technology transfer potential of a project outcome, and the partners/persons implicated.

And so is validated by the authors through signing this authorship form:

Last name, Name	Signature	Date