Funded by
the European Union

October 25th 2023 – Open Science Event

NYMPHE

NYMPHE Experience – managing data in Horizon project

Elena Biagi, UNIBO
Giulia Caldoni, UNIBO

CC BY

25 October 2023

The project at a glance

Call: HORIZON-CL6-2021-ZEROPOLLUTION-01 (**Destination**: Clean environment and zero pollution)

Topic: HORIZON-CL6-2021-ZEROPOLLUTION-01-10 - Environmental services: improved bioremediation and revitalization strategies for soil, sediments and water

Type of Action: Research and Innovation Action

Starting date: 1 January, 2023

Duration: 48 months

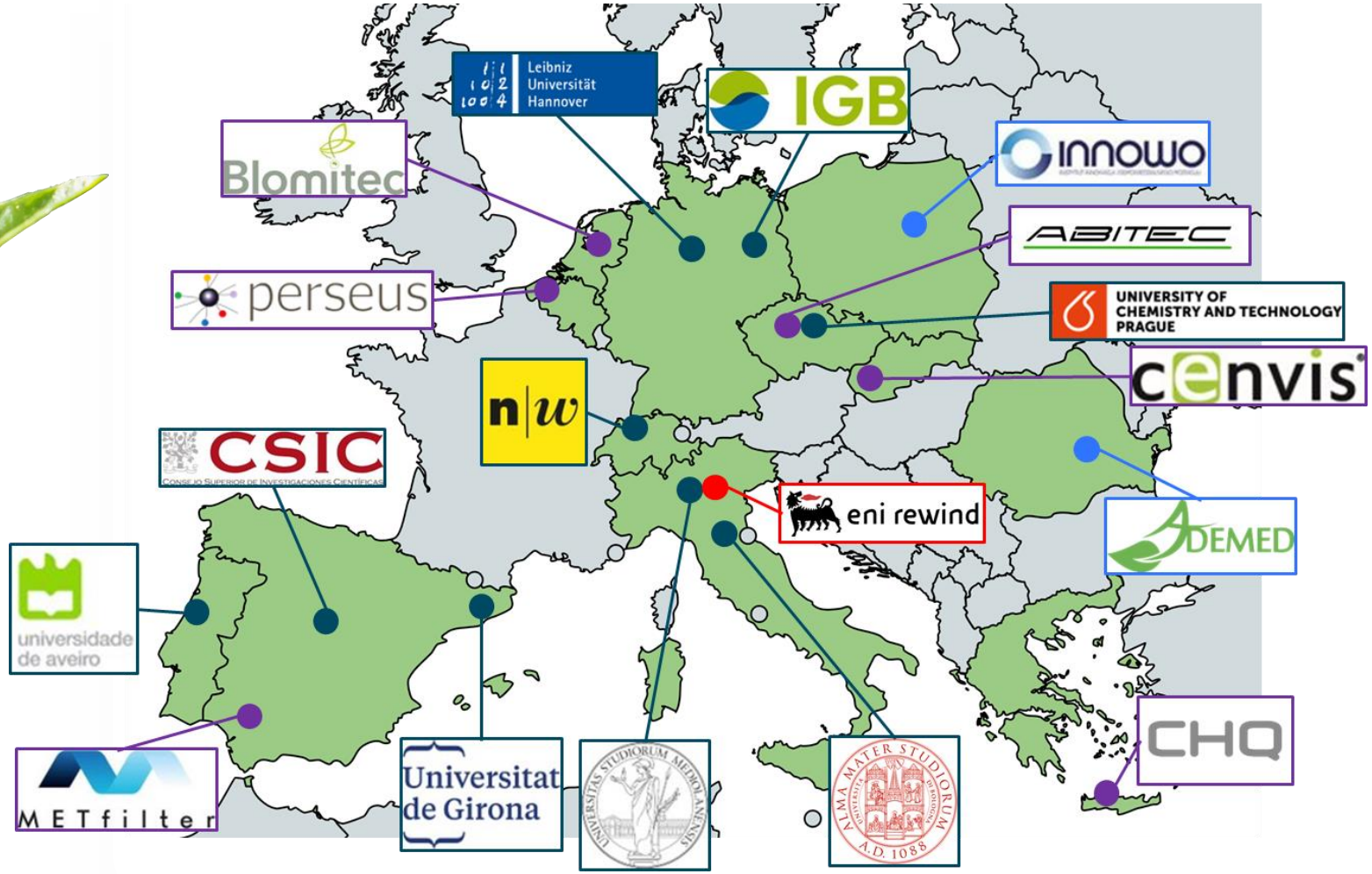
Total cost: 5.702.902,80 €

EU funding: 5.499.973,50 €

Coordinator: Prof. Giulio Zanaroli, University of Bologna

Project Officer: Dr. Lukas Varnas, European Research Executive Agency

NYMPHE consortium



18 Partners

- 9 Universities / Research institutions (●)
- 6 SMEs (●)
- 1 Large company (●)
- 2 NGOs (●)

12 Countries

Objectives

To develop bioremediation / revitalization strategies based on **bioremediation systems** consisting of assemblies of **multiple biologics**,

assess them from technical, environmental, economic and societal perspectives for the revitalization of polluted environments, and

showcase them **at 4 real contaminated sites** and 1 site for confined tests with genomically edited organisms,

with the **ambition of removing at least 90% of the main pollutants of the four polluted sites and improve their biodiversity status**.

Sites & target polluted matrices

IGB site - Hyporheic zone (HZ) (Hoppegarten-Berlin, DE): River Erpe sediment and surface water contaminated with **pharmaceuticals**

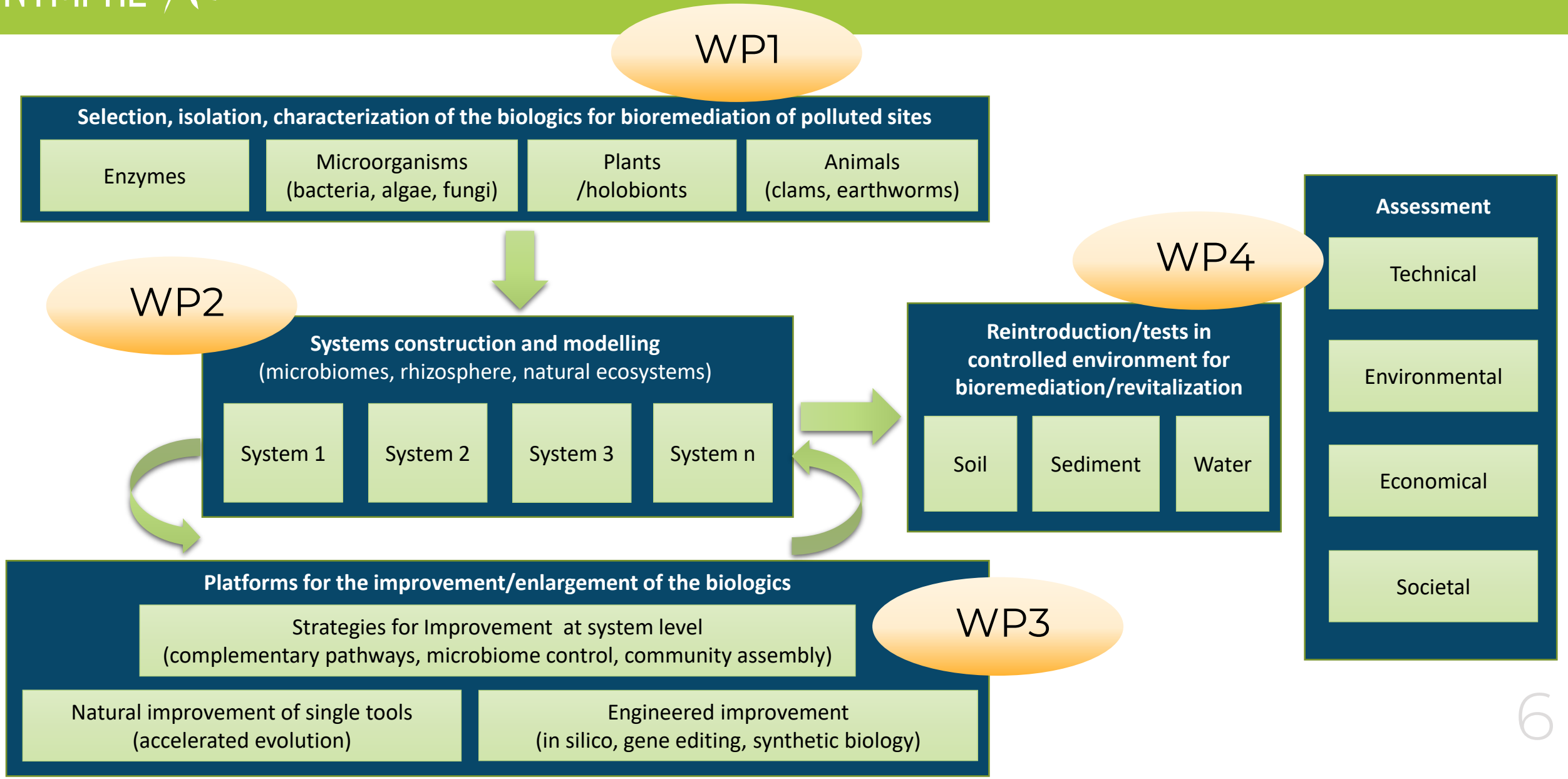
CENVIS site – (Leopoldov, SK) site used as **facility enabling the "confined use" of genomically edited biologics** to remediate the contaminated matrices of ER, CHQ and IGB sites

METFI site – Municipal WW (IMDEA Water, Alcala, ES): Municipal WW with **pharmaceuticals** (antibiotics) and **microplastics** (**MPs**)

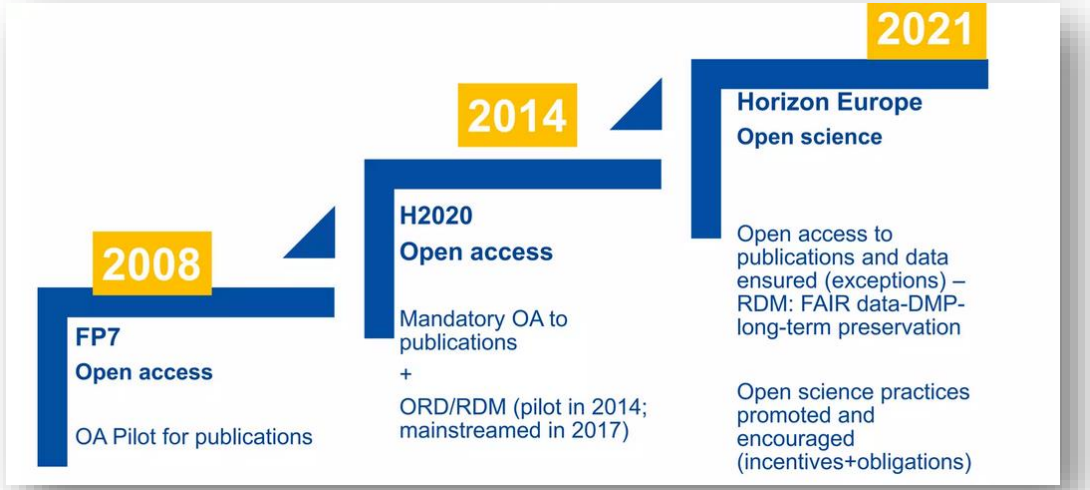
CHQ site - Agricultural field (Falassarna, GR): soil contaminated with **microplastics (MPs)** and **pesticides**

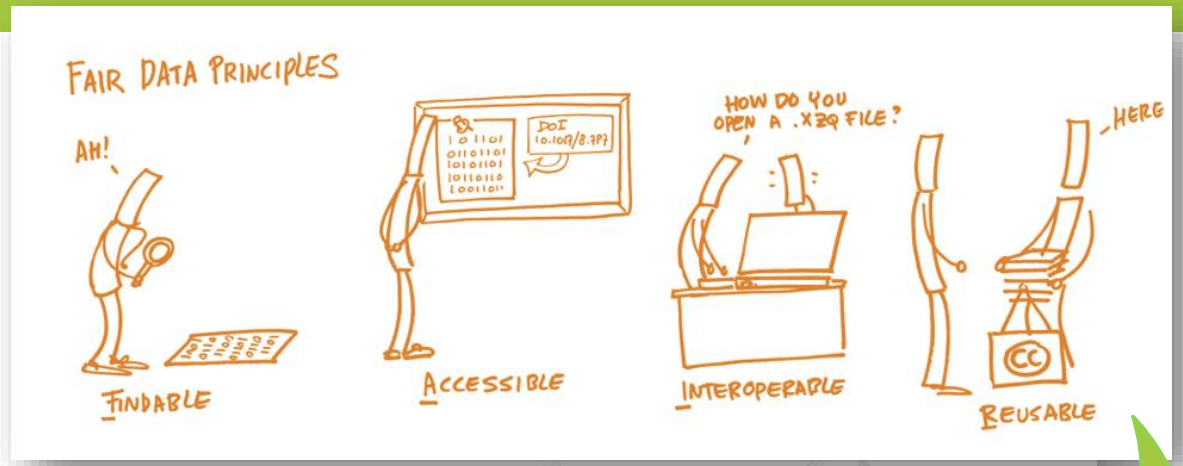
ER site - Industrial site (Ferrara, IT): groundwater (GW), aquifer sediments, and soil matrices contaminated with **Petroleum Hydrocarbons** (**PHC**) and **chlorinated aliphatic hydrocarbons** (**CAH**), **heavy metals** (**HM**)

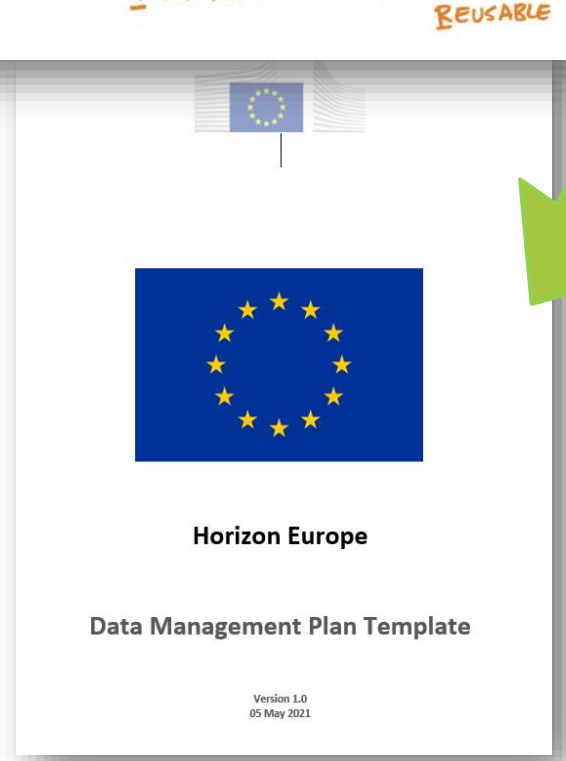
The technical approach



Research Data Management in Horizon Europe Projects





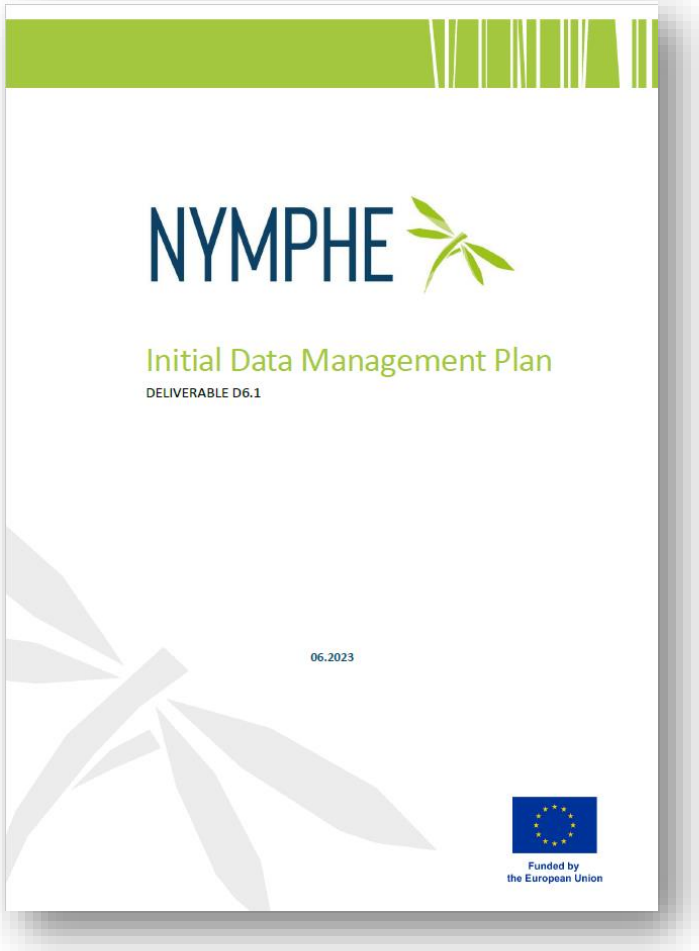


- **Manage** data in line with **FAIR principles**
- **Establish** and **update** a Data Management Plan (**DMP**)
- "**As open as possible, as closed as necessary**"

Tsoukala, V., Open Science in Horizon Europe (2021). https://es.slideshare.net/victsoukala/2021-05-oscytsoukala; Communication, dissemination, open and visibility – article 17, annex 5, grant agreement

The Data Management Plan (DMP)

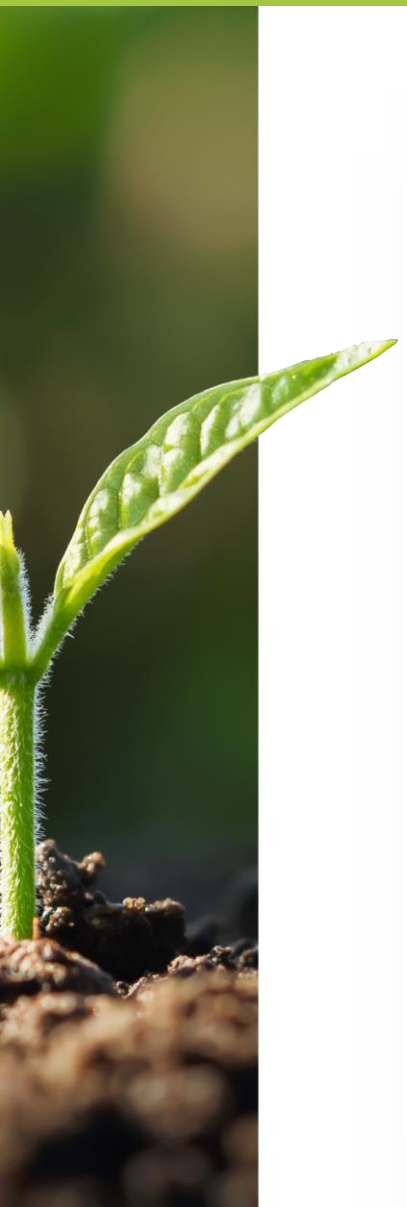
The DMP is a MANDATORY deliverable for project funded by Horizon Europe programme.



Is a roadmap to:

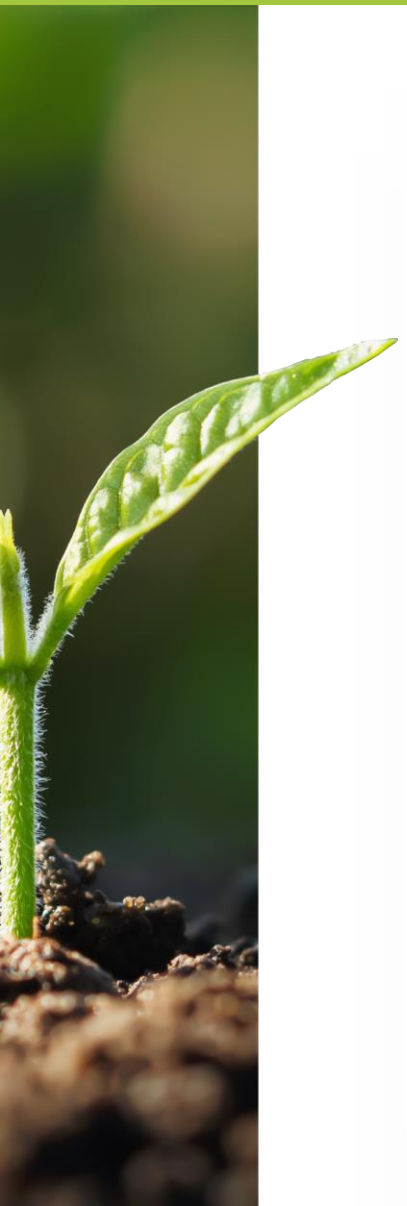
- Develop procedures early on for efficient and consistent research work, saving your time later on.
- Manage data to avoid problems like data duplication, loss, and security breaches.
- Make research materials visible and citable to maximize the impact of your work.
- Be aware of how you are fulfilling all your legal and ethical obligations.

Initial Worries & Challenges in DMP preparation



- High number of partners, both research institutions and companies
- WPs complex design in terms of data used and generated
- Difficulty in translating complex and heterogenous scientific outputs into organized datasets before actually obtaining them
- How to collect and harmonize information from partners
- How to compromise between the requirement of disclosing the content of datasets and the need of keeping data «hidden» before publication

Worries & Challenges in DMP preparation



- High number of partners, both research institutions and companies
- **WPs complex design in terms of data used and generated**
- Difficulty in translating complex and heterogenous scientific outputs into organized datasets before actually obtaining them
- How to collect and harmonize information from partners
- How to compromise between the requirement of disclosing the content of datasets and the need of keeping data «hidden» before publication

Example: WP2 - Systems design, assembly and modelling

Computational methods

Biologics performance data from WP1 and WP3

Metagenomics

Microbiomes modelling

Assembly of biologics systems

Metagenomics data

Performance evaluation data

Metagenomes and MAGs

Microbiome digital twins
Microbiome **networks**

- Systems **composition** and features
- Measurments of **parameters to assess performances**

Example: WP4 - Testing and assessment of the remediation technologies

Regulatory background

Sites preparation

Sites features

- **Physicochemical features** of each evironmental matrix
- **Ecotoxicological features** of each site

Data, models and biologics from WP1-3

Bioremediation field tests

Existing bioremid technology
Societal background on bioremid

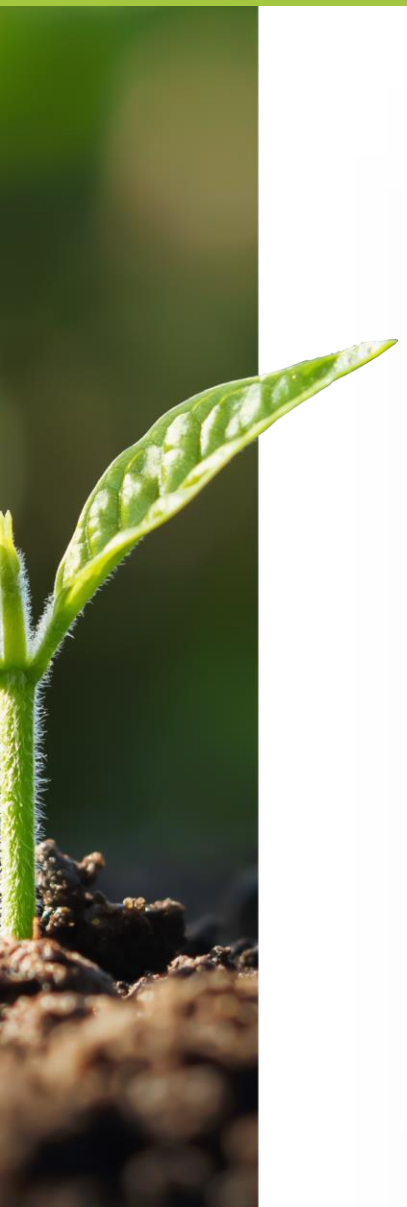
Background data

Assessment data

- Measurments of **parameters to assess performances**
- **LCA, ECA, EA, RA** data and outcomes
- **Societal assessment and stakeholder questionnaire data**

Privacy!

UNIBO activities of coordination



- Provide partners with a **decision tree to understand what type of data each one will have to manage**
- **Collect information from partners using a questionnaire** with all necessary questions to stimulate and guarantee a FAIR data management
- **Harmonize and integrate collected information** to **compile a first version of the DMP** which includes the NYMPHE general approach to data management and a **list of (for now) hypothetical datasets the project is going to produce**

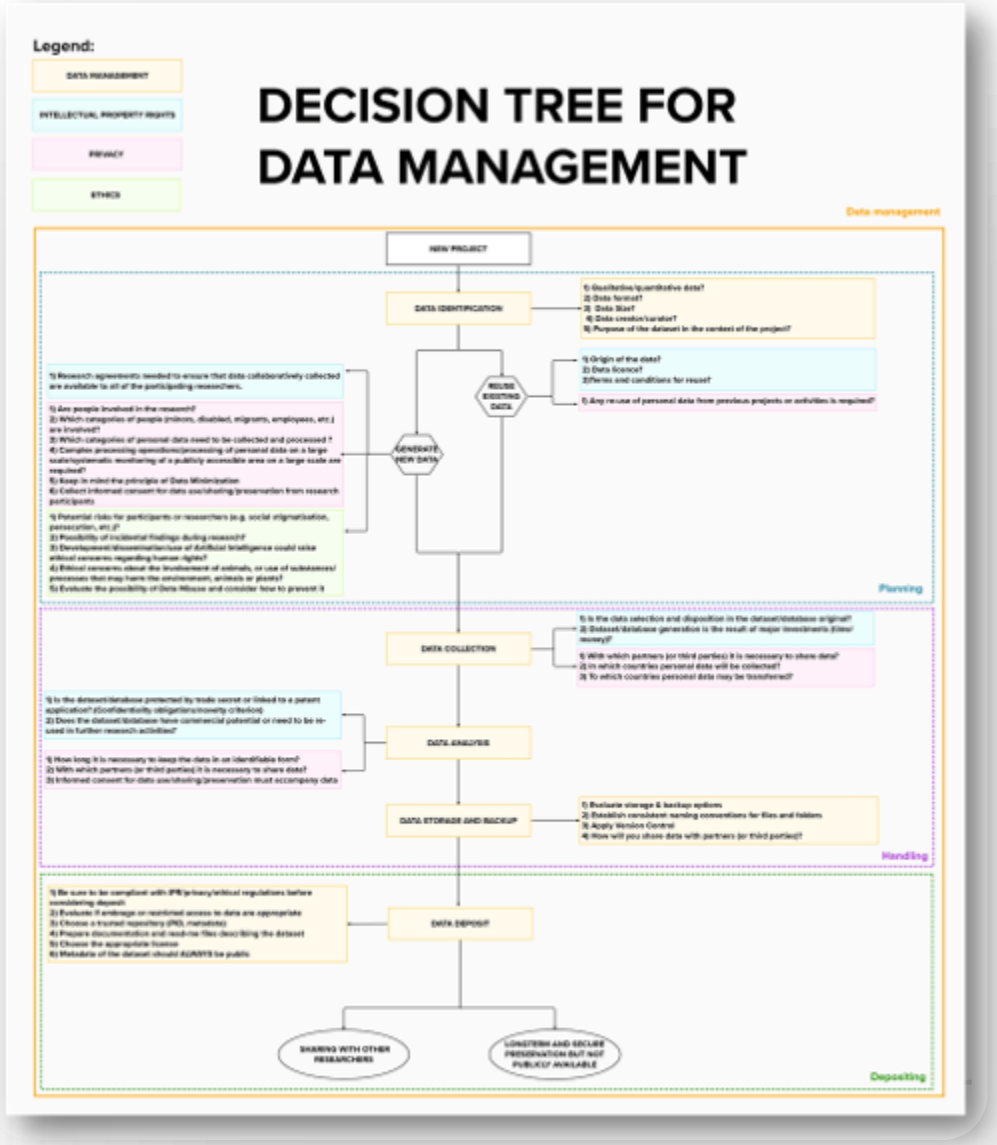
...for the next 4 years, UNIBO will **keep on updating** and integrating the information in the DMP and periodically provide an updated versions.

Define what kind of data a partner will have to manage

Decisional tree to **help researchers ask, and answer, questions they did not know they had.**

Starting from the fundamental bricks of the data lifecycle, integrates a **series of questions** aiming to encourage the researchers address some major attention points, encompassing issues including

- privacy/ethics requirements
- IPR legislation
- Data storage and preservation
- FAIR principles and Data deposit



UNIBO activities of coordination

- Provide partners with a **decision tree to understand what type of data each one will have to manage**
- **Collect information from partners using a questionnaire** with all necessary questions to stimulate and guarantee a FAIR data management
- **Harmonize and integrate collected information** to **compile a first version of the DMP** which includes the NYMPHE general approach to data management and a **list of (for now) hypothetical datasets the project is going to produce**

...for the next 4 years, UNIBO will **keep on updating** and integrating the information in the DMP and periodically provide an updated versions.

Partners questionnaire

English (United Kingd...

Data sets to include in DMP - form for NYMPHE project partners

In order to create the best possible Data Managment Plan (DMP) for o
attached questionnaire, providing details on the data that you collect a
guidance is provided for each question below.

Please bear in mind:

- A data set is defined as **a named collection of data units with th**
- **Each dataset needs to be described separately**; when you finish
button to start the questionnaire again.
- If you have **datasets that are similar** to each other, after filling ou
get PDF of answers" (at the end of the form)->**"Save as PDF"** to
same fields
- We are awa
collection.
what you p

Thank you!

Data summary

12. **Work package number** *

Which WP does the data set belong to? E.g., WP1

13. **Task (and sub-task)** *

E.g., Task 1.1

14. **Data set name** *

Please provide a name or a title for the data set being described. For consistency, data sets should be named according to specific rules. We suggest to follow this convention:

- *Project acronym*
- *Work Package Number*
- *Short title consistent with the content of the data set*
- *Version number (if applicable, to identify updates)*

*e.g., **ACRONYM. WP1. Short title. V2***

21. **Data types** *

Whether re-used or new, data can be categorized in many ways. For example:

- *numerical, textual, graphical, visual or tactile*
- *born digital or digitized*
- *quantitative o qualitative (qualitative tend to be descriptive and cannot be measured; quantitative data provides information about quantities and can, of course, be measured)*
- *raw, secondary, derived, cleaned, processed*

For more information, please see: https://data.library.virginia.edu/data-management/plan/format-types/

Inserisci la risposta

22. **Data formats** *

Please indicate the file formats used in:

- *data collection/generation (raw data)*
- *data processing*
- *data preservation and sharing*

Proprietary formats should be converted into non-proprietary, documented, standard format term preservation. You can find some guidance here: https://ukdataservice.ac.uk/learning-hu management/format-your-data/recommended-formats/

Inserisci la risposta

23. **Size of data**

This can vary widely according to the type of data. To estimate how much storage might be look at: https://en.wikipedia.org/wiki/Orders_of_magnitude_(data) to get a general idea. Bea your estimate needs to be quite precise, as storage space may have to be included in the bu

Please make an estimate as accurate as possible.

Inserisci la risposta

Making data Findable

26. **Persistent identifier (PID)** *

Your data set must have a persistent identifier (PID). Please include t itself, if already available (e.g., DOI, https://doi.org/10.1000/182).

*Persistent identifiers are generally associated to a data set upon arc repository that automatically assigns a PID as soon as (or soon after include **DOI**, **Handle**, **PURL**, **URL**, **URN**, **ISBN**, check the type of per service.*

If the data set has no persistent identifier yet, just type "not ye

Inserisci la risposta

27. **Metadata and documentation** *

*Indicate which metadata **standards**, if any, will be used to describe the data set. (e.g., which controlled vocabulary or thesaurus is used to indicate the subject of the data set or which descriptive domain specific details are provided to represent a data set).*

*Also, please specify which kind of **documentation** will be provided **(e.g., codebooks, readme files provided alongside the data).***

*Specify if **keywords** will be provided in the metadata to optimize discovery and re-used, and if metadata will be made available for harvesting by web browsers and search engines.*

If you don't already have this detailed information, please provide an outline of what you plan to do.

Inserisci la risposta

28. **Repository for long-term preservation** *

*Please include both the **name and URL** of the repository.*

Repositories assume a central role in the Horizon Europe. They can be certified repositories (e.g. CoreTrustSeal, nestor Seal DIN31644, ISO16363) or disciplinary and domain repositories commonly used and endorsed by international research communities. General-purpose repositories or institutional repositories are also allowed but must:

- *be "Trusted", with specific provisions in place and explicit information online about their policies and services (e.g., acquisition, access, security of content, long-term sustainability of service including funding etc.)*
- *assign persistent unique identifiers to contents (e.g., DOIs, handles, etc.) and ensure that contents are accompanied by enough machine-actionable and standardized (e.g., Dublin Core, Data Cite etc.) metadata*
- *meet international and national criteria for security and have different levels of security depending on the sensitivity of the data being deposited to maintain privacy and confidentiality.*

To find a repository, you can ask your institution or use a registry such as https://www.re3data.org/. Alternatively, you can use Zenodo, which is a multidisciplinary repository recommended by the European Commission: https://zenodo.org/.

Inserisci la risposta

29. **Is your repository OpenAIRE compliant?**

*If possible, choose a data repository that is **OpenAIRE compliant (https://explore.openaire.eu/participate/deposit/search)** and/or catalogued in Re3data registry of repositories (https://www.re3data.org/).*

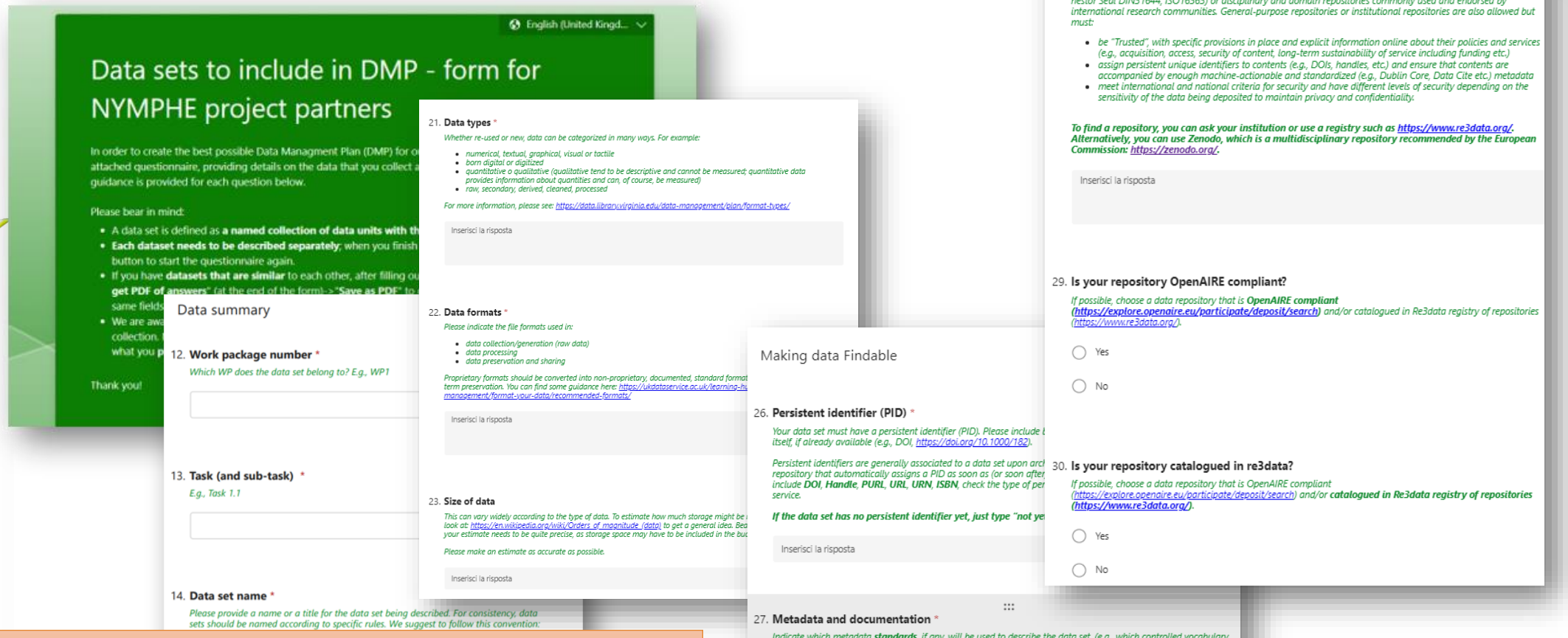
- Yes
- No

30. **Is your repository catalogued in re3data?**

*If possible, choose a data repository that is OpenAIRE compliant (https://explore.openaire.eu/participate/deposit/search) and/or **catalogued in Re3data registry of repositories (https://www.re3data.org/)**.*

- Yes
- No

Partners questionnaire



UNIBO research groups filled in the questionnaire before all other partners and provided the answers to them as an example

Partners questionnaire

Data sets to include in DMP - form for NYMPHE project partners

In order to create the best possible Data Managment Plan (DMP) for o
attached questionnaire, providing details on the data that you collect a
guidance is provided for each question below.

Please bear in mind:

- A data set is defined as **a named collection of data units with th**
- **Each dataset needs to be described separately**; when you finish
button to start the questionnaire again.
- If you have **datasets that are similar** to each other, after filling ou
get PDF of answers" (at the end of the form)->"**Save as PDF**" to
same fields
- We are awa
collection.
what you p

Thank you!

Data summary

12. **Work package number** *

Which WP does the data set belong to? E.g., WP1

13. **Task (and sub-task)** *

E.g., Task 1.1

14. **Data set name** *

*Please provide a name or a t
sets should be named accord*

- *Project acronym*
- *Work Package Numbe*
- *Short title consistent w*
- *Version number (if app*

*e.g., **ACRONYM. WP1. Shor***

21. **Data types** *

Whether re-used or new, data can be categorized in many ways. For example:

- *numerical, textual, graphical, visual or tactile*
- *born digital or digitized*
- *quantitative o qualitative (qualitative tend to be descriptive and cannot be measured; quantitative data provides information about quantities and can, of course, be measured)*
- *raw, secondary, derived, cleaned, processed*

For more information, please see: https://data.library.virginia.edu/data-management/plan/format-types/

Inserisci la risposta

22. **Data formats** *

Please indicate the file formats used in:

- *data collection/generation (raw data)*
- *data processing*
- *data preservation and sharing*

*Proprietary formats should be converted into non-proprietary, documented, standard format
term preservation. You can find some guidance here: https://ukdataservice.ac.uk/learning-hu
management/format-your-data/recommended-formats/*

Inserisci la risposta

23. **Size of data**

Making data Findable

26. **Persistent identifier (PID)** *

*Your data set must have a persistent identifier (PID). Please include t
itself, if already available (e.g., DOI, https://doi.org/10.1000/182).*

*Persistent identifiers are generally associated to a data set upon arc
repository that automatically assigns a PID as soon as (or soon after
include **DOI**, **Handle**, **PURL**, **URL**, **URN**, **ISBN**, check the type of per
service.*

ifier yet, just type "not ye

*y, will be used to describe the data set. (e.g., which controlled vocabulary
ct of the data set or which descriptive domain specific details are provided*

entation** will be provided **(e.g., codebooks, readme files provided

*Specify if **keywords** will be provided in the metadata to optimize discovery and re-used, and if metadata will be made available for harvesting by web browsers and search engines.*

If you don't already have this detailed information, please provide an outline of what you plan to do.

Inserisci la risposta

28. **Repository for long-term preservation** *

*Please include both the **name and URL** of the repository.*

Repositories assume a central role in the Horizon Europe. They can be certified repositories (e.g. CoreTrustSeal, nestor Seal DIN31644, ISO16363) or disciplinary and domain repositories commonly used and endorsed by international research communities. General-purpose repositories or institutional repositories are also allowed but must:

- *be "Trusted", with specific provisions in place and explicit information online about their policies and services (e.g., acquisition, access, security of content, long-term sustainability of service including funding etc.)*
- *assign persistent unique identifiers to contents (e.g., DOIs, handles, etc.) and ensure that contents are accompanied by enough machine-actionable and standardized (e.g., Dublin Core, Data Cite etc.) metadata*
- *meet international and national criteria for security and have different levels of security depending on the sensitivity of the data being deposited to maintain privacy and confidentiality.*

To find a repository, you can ask your institution or use a registry such as https://www.re3data.org/. Alternatively, you can use Zenodo, which is a multidisciplinary repository recommended by the European Commission: https://zenodo.org/.

Inserisci la risposta

29. **Is your repository OpenAIRE compliant?**

*If possible, choose a data repository that is **OpenAIRE compliant** (**https://explore.openaire.eu/participate/deposit/search**) and/or catalogued in Re3data registry of repositories (https://www.re3data.org/).*

- Yes
- No

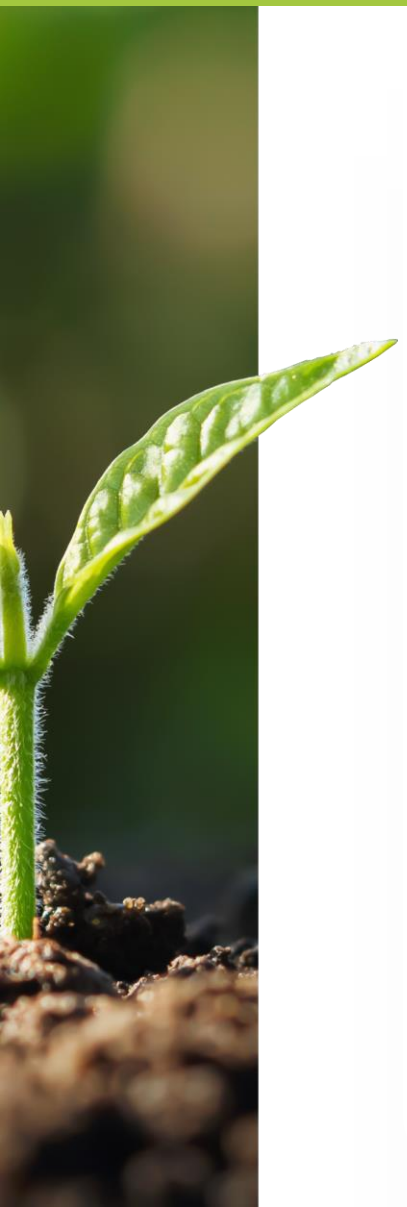
30. **Is your repository catalogued in re3data?**

*If possible, choose a data repository that is OpenAIRE compliant (https://explore.openaire.eu/participate/deposit/search) and/or **catalogued in Re3data registry of repositories (https://www.re3data.org/)**.*

- Yes
- No

The questionnaire approach had a very positive feedback from partners

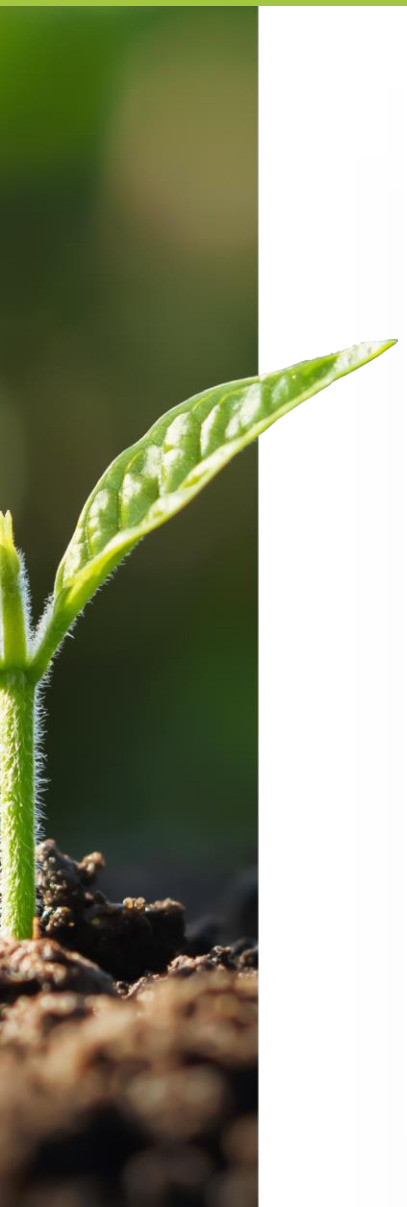
UNIBO activities of coordination



- Provide partners with a **decision tree to understand what type of data each one will have to manage**
- **Collect information from partners using a questionnaire** with all necessary questions to stimulate and guarantee a FAIR data management
- **Harmonize and integrate collected information** to **compile a first version of the DMP** which includes the NYMPHE general approach to data management and a **list of (for now) hypothetical datasets the project is going to produce**

...for the next 4 years, UNIBO will **keep on updating** and integrating the information in the DMP and periodically provide an updated versions.

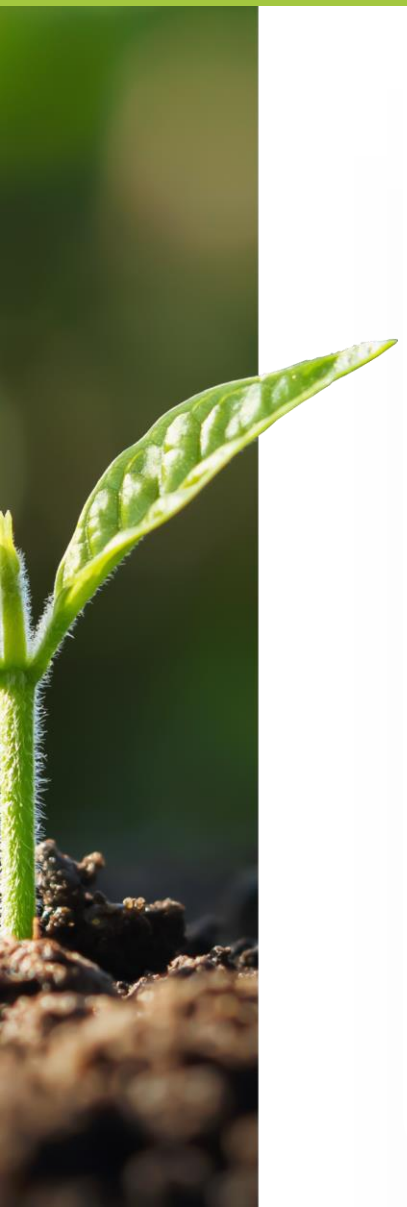
UNIBO activities of coordination



- Provide partners with a **decision tree to understand what type of data each one will have to manage**
- **Collect information from partners using a questionnaire** with all necessary questions to stimulate and guarantee a FAIR data management
- **Harmonize and integrate collected information** to **compile a first version of the DMP** which includes the NYMPHE general approach to data management and a **list of (for now) hypothetical datasets the project is going to produce**

...for the next 4 years, UNIBO will **keep on updating** and integrating the information in the DMP and periodically provide an updated versions.

...keeping the DMP updated



- Keeping track of produced and published data within the consortium activity (thanks to partners meeting twice a year)
- Reminding and helping partners who produced datasets in updating the infos in the DMP and making sure to deposit the data in **appropriate repository**

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.