Ref. Ares(2024)4595944 - 26/06/2024

# D1.2 Data Management Plan

Due date: 30/06/2024
Responsible partner: Enspire

**Disclaimer**:

The content reflects the views of the authors only. The European Commission is not liable for any use that may be made of the information contained herein.

This document contains information, that is proprietary to the Bio4HUMAN consortium. Neither this document nor the information contained herein shall be used, duplicated, or communicated by any means to any third party, in whole or in parts, except with the prior written consent of the Bio4HUMAN consortium. This restriction legend shall not be altered or obliterated on or from this document. Neither the European Commission nor the Bio4HUMAN project consortium are liable for any use that may be made of the information that it contains.

## Document information

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **DELIVERABLE NUMBER** | D1.2 | | | | | | |
| **DELIVERABLE NAME** | Data Management Plan | | | | | | |
| **ANNEXES** | - | | | | | | |
| **TYPE*** | R: | | DMP: | x | DATA: | | DEC: |
| **DISSEMINATION LEVEL**** | PU: | x | SEN: | | EU CLASSIFIED: | | |
| **LEAD BENEFICIARY** | Enspire Science Ltd. | | | | | | |
| **DUE DATE** | 30/06/2024 | | | | | | |
| **SUBMISSION DATE** | 26/06/2024 | | | | | | |

*R — Document, report; DMP — Data Management Plan; DATA — data sets, microdata, etc; DEC —Websites, patent filings, videos, etc.

**Public — fully open (automatically posted online); Sensitive — limited under the conditions of the Grant Agreement; EU classified — restreint-ue/eu-restricted, confidentiel-ue/eu-confidential, secret-ue/eu-secret under decision 2015/444

## Document history

| ID | DATE | AUTHOR | CHANGE |
|---|---|---|---|
| 1. | 13/06/2024 | A. Ratkosova, Enspire | First draft |
| 2. | 20/06/2024 | A. Ratkosova, Enspire | Second draft after inputs and comments |
| 3. | 26/06/2024 | A. Ratkosova, Enspire | Final draft |

## Contents



## List of Tables

# 1. Executive Summary

This document presents the first version of Bio4HUMAN Data Management Plan (DMP). The DMP is a living document that outlines the main elements of the data management that will be used by the Bio4HUMAN Consortium regarding the project data. It includes descriptions of the way that data is handled during and after the project, the type of data to be collected, processed, or generated, the methodology and standards to be applied to ensure that the data management meets the FAIR data principle and whether, and how the data will be shared.

The document follows the template provided by the European Commission in the Participant Portal. The main topics covered in this document are:

- The scope and objectives for data management in the project;
- Overview of what data may be gathered, processed, and/or generated in the project;
- How data will be handled according to the FAIR Data principle, making data: findable, accessible, interoperable, and reusable;
- The resources and costs of making data FAIR in this project;
- How the consortium members intend to keep the data secure;
- Ethics and privacy strategy in Bio4HUMAN.

# 2. Introduction

## 2.1 Scope and Objectives

The purpose of this Data Management Plan (DMP) is to provide an analysis of the main elements of data management during and after the Bio4HUMAN project, and how data will be handled in line with the European Commission's requirements. The DMP describes the data expected to be gathered, processed, and/or generated with the metadata attached to ensure their compliance with the FAIR data principles. The data management practices will ensure that the project's data can be preserved, exploited, and shared in a consistent manner.

Bio4HUMAN aims provide humanitarian aid operators and bio-based sector stakeholders, with science-based information on the application potential, sustainable performances, and circularity of bio-based products and systems, suitable for humanitarian purposes. This will be accomplished by examining the applicability of sustainable bio-based products and systems in the humanitarian context across different geographical regions, focusing on solid waste-related challenges. Parallel to that the project aims to explore the socio-economic, innovative and governance aspects of the proposed solutions, not omitting their potential for future replication that brings about efficiency, effectiveness, sustainability, cost-effectiveness, knowledge sharing, etc. The main findings of the project will be translated into a set of guidelines and recommendations and a replication roadmap of solutions identified applicable to diverse humanitarian contexts. Therefore, sharing and ensuring accessibility to databases of key stakeholders such as humanitarian organisations, donors, clusters, authorities and public service operators, bio-based companies and research institutions as well as national and international organisations advocating for the bio-product and bio-services compiled during the project that can support the identification and following strategy for replication of the bio-based solutions within humanitarian context globally (at a national, EU and international levels is essential.

This first version of the Data Management plan is submitted to the EU in June 2024. It is important to note that the document is a living document, meaning that it will evolve during the project. A history of versions will track the changes with a clear version number and a date. The Coordinator (Enspire) which is the primary responsible for data management will plan an internal review of this deliverable 12 months after its submission and also by the end of the project (D2.1, M29, June 2026).

## 2.2 List of Abbreviations

| | |
|---|---|
| AIMPLAS | Asociación de Investigación de Materiales Plásticos y Conexas |
| BIC | Bio-based Industries Consortium |
| CBE | Circular Bio-based Europe |
| CFRM | Community feedback and response mechanisms |
| DRC | Democratic Republic of Congo |
| DG ECHO | Directorate-General for European Civil Protection and Humanitarian Aid Operations |
| DMP | Data Management Plan |
| FAIR | Findable, accessible, interoperable, reusable |
| HO | Humanitarian Organisation |
| IBF | Irish Bioeconomy Foundation |
| ICRC | International Committee of the Red Cross |
| IFRC | International Federation of Red Cross and Red Crescent Societies |
| ITENE | Instituto Tecnologico del Embalaje, Transporte y Logística Paterna |
| MEAL | Monitoring, evaluation, accountability, and learning |
| OCHA | United Nations Office for the Coordination of Humanitarian Affairs |
| PAH | Polska Akcja Humanitarna Fundacja (Polish Humanitarian Action) |
| PIN | People in Need |
| PRO CIVIS | PRO CIVIS - Fundacja Edukacji i Dialogu Społecznego |
| RCRC | Red Cross and Red Crescent Movement |
| SWM | Solid Waste Management |
| UC | Universidad de Cantabria |
| UNEP | UN Environment Program |
| UNICEF | The United Nations International Children's Emergency Fund |
| WASH | Water, Sanitation and Hygiene |
| WREC | Waste Management Measuring, Reverse Logistics, Environmentally Sustainable Procurement and Transport, and Circular Economy |
| WWF | World Wide Fund for Nature |

# 3 Data Summary

The data that may be collected, generated, and used will be necessary to meet the objectives of the Bio4HUMAN project. The project data can be divided into two groups:

- Primary and secondary data collected/generated/used by the consortium or the involved stakeholders.
- Data collected/generated/used by the consortium or the involved stakeholders during dissemination and communication activities, training activities, and other events.
- Data from literature, previous projects of the consortium or the involved stakeholders

## 3.1 Re-use of Existing Data

In addition to data to be generated in the project, we will use existing data, most of it owned by the partners, including:

- Internal databases of the project’s partners;
- Databases of humanitarian clusters, Reliefweb (a comprehensive repository of information on global crises and disasters) EHA Connect (online repository of environment in humanitarian action (EHA) tools and guidance);
- Publicly available data on: 1) national, European, international legislation on topics relevant for the bio - circular part of Project's activities 2) programs and strategies on bioeconomy / circular economy
- Databases of reports and publications resulting from the EU funded projects on the bioeconomy (CORDIS)
- Databases of Intellectual Property Rights (European Union Intellectual Property Office, European Patents Office)
- Potential additional existing datasets that will be used for reports and activities;
- Data from any platform relevant to the project.
- Data from generic LCA databases (Industry 2.0, European Commission EF 3.1 database, ecoinvent Sphera, etc.) and previous sustainability assessments projects

This data has been used to identify key stakeholders. It is being used to conduct humanitarians sector needs assessment. Furthermore, we foresee the usage of specific bio-based stakeholders’ datasets in order to conduct the scoping exercise and following LCAs of the identified solutions.

## 3.2 Data types and formats

The outputs produced in Bio4HUMAN will be in the form of text resources such as reports, manuals, and descriptions of procedures and occasionally will be accompanied by databases in Excel format (e.g., T3.1 and its D3.1 internal project database of the key stakeholders, T4.2 internal database of Bio-based solutions). While the full scope of data to be collected, processed, and/or generated is not clear yet, we foresee the involvement of different kinds of data related to the following areas:

- Stakeholder engagement
- Humanitarian sector needs assessment
- Scoping exercise
- LCA (environmental and social)
- Training content
- Replication and capacity building

The following data formats will be used:

- Research data and outputs - raw and processed data collected/generated through scientific work, including publications, specific data from companies while available
- Project internal databases
- Project operational data – reports (Word, Excel, PDF, etc.)
- Video recordings from training sessions
- Photographs of waste from case studies' locations

The data collected via the website is subject to the privacy policy, presented here: https://bio4human.eu/privacy-policy/

## 3.3 Data Utility

The data collected/generated/re-used in the Bio4HUMAN project may be useful for several groups:

- Humanitarian organisations (HOs) and donors,
- Civil society organisations (NGOs, community-based organisations, social movements)
- Scientific community (bio-based researchers & their laboratories, circular economy researchers,
- EU funded projects
- Business community (SMEs including start-ups, entrepreneurs, companies, trade associations, business chambers)
- Policymakers (EU, national, regional)

Members of all these groups can use the data and outputs of the project in order to take part in the replication of identified bio-based solutions, further scientific research and collaborative projects and actions. For example, data integrated into the deliverables and different outputs such as List of bio-based solutions, Gap analysis report, Hotspot analysis of the current and innovative solutions can be used by potential aid organisations and donors to further push on the replication of the identified solutions. Furthermore, bio-based research institutions and companies will be able to use results of the scoping and LCA to strengthen collaboration between bioeconomy sector and humanitarian sector with guidelines to be implemented in other contexts. Policymakers will potentially use the data and outputs provided in the Bio4HUMAN project to adjust policies or develop new policies that support addressing solid waste management challenge in humanitarian context.

# 4 FAIR Data

The FAIR data principles describe four key concepts in research data management. In accordance with the FAIR principles, data should be:

- **Findable** – Easy to find by humans and computers and based on mandatory description of the metadata that allows data discovery;
- **Accessible** – Long-term storage so data can be easily accessed with defined access conditions to actual data and metadata;
- **Interoperable** – Ready to be combined with other datasets by humans or computers;
- **Reusable** – Ready to be used and further processed for research or other purposes.

This section describes the measures taken to ensure the alignment with the FAIR data principle.

## 4.1 Making data findable, including provisions for metadata

To allow the reuse of non-confidential Bio4HUMAN data, metadata and data should be easy to find. The data to be generated/collected/reused in the Bio4HUMAN project will be identifiable and locatable by means of unique identification mechanisms. Files will be uniquely identifiable by using standardized name conventions and clear versioning.

The following common rules for dataset/output naming and versioning will be used to ensure data visibility, discoverability, citation, and permanent online tracking. The recommended title for each dataset consists of:

Project acronym: Task title or short description of specifying Task aims: additional information specifying coverage and nature of data/output (optional: version number in case of revisions or updates)

For example:

Bio4HUMAN: Data Management Plan V.1

Bio4HUMAN: List of bio-based solutions

Furthermore, specific, and consistent keywords will be associated with each dataset or output to enhance semantic discoverability.

## 4.2 Making data accessible

Bio4HUMAN will follow the principle "as open as possible, as closed as necessary." Namely, all data will be made available, except for the type of data that will be defined as confidential for various reasons, including commercial, privacy, and security considerations. In case relevant, scientific publications will be published with green/golden Open Access.

To facilitate access, data and other outputs will be reposited and published through project platforms such as the website, social media, as well as self-archived by the Bio4HUMAN consortium members' websites, repositories, and other platforms.

The consortium will store the research data in a format that is suited for long-term preservation and accessibility. Furthermore, in the updated version of this deliverable, the following issues will be addressed:

- Specification of the data that will be made openly available and data that will be confidential, including reasoning;
- Information on where the data and associated metadata, documentation, and code are deposited;
- Specification of restrictions and how access will be provided in such cases.

## 4.3 Making data interoperable

Datasets will follow the convention of the repository they are hosted at. All relevant documentation explaining methodologies and data, including data collection procedures, processing, and data quality information will be made available along with the data in order to guarantee reproducibility and the validation of the findings. All data sets will be described using standard descriptive metadata, such as Dublin Core and DataCite Metadata Schema.

For T3.3 survey and direct observations, PIN team will use KoBoCollect. Kobo uses metadata vocabulary associated with Open Data Kit (ODK). For direct observations, PIN & PAH will also use Geopoint (GPS coordinates).

Data will be collected and shared in a standardized way using a standard format for that data type. Reference will be made to the software(s) required to run it. We do not foresee barriers or interoperability issues. The table below describes the type of metadata to be included.

**Table 1 Bio4HUMAN Metadata types**

| General information | Sharing and Access | Methodology |
|---|---|---|
| Title<br><br>Identifier<br><br>Author information<br><br>Date of production<br><br>Project information (acronym, number) | Access restrictions<br><br>Link to data repository<br><br>Links to publications citing the data<br><br>Link to data source (when relevant) | Used content<br><br>Description of methods<br><br>Instrument / software used<br><br>Quality assurance |

## 4.4 Increase data reuse

In general, data and outputs will be uploaded to the Bio4HUMAN website to make it more accessible for reuse. Relevant documentation explaining data collection/generation/reuse procedures (e.g., methodologies) will be made available with the data, to guarantee reusability.

When relevant, the data generated will be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement.

The examples of data that are anticipated to be reused: humanitarian sector needs assessment data, the list of bio-based solution and its sustainability scoring matrix, gap analysis report on supply chain management in the humanitarian context, images and videos collected during the project.

# 5 Allocation of resources

The activities related to providing the data and outputs in Open Access are anticipated to be covered within the allocated budget for each WP, being acknowledged as eligible costs under Horizon Europe. The costs of making scientific publications, hosting a project website, and the data and outputs repositories are contained within the Bio4HUMAN budget as eligible costs.

Further investigation of the potential cost related to the data and outputs repository needs to be done. The repository will ensure that data is stored safely and securely and in compliance with European Union data protection laws and in accordance with Article 27 (Appendix A).

As part of WP1 and WP2, Enspire, the coordinator of Bio4HUMAN, is responsible for the data and outputs management and the Data Management Plan, in collaboration with all other consortium partners.

As per Grant Agreement of Bio4HUMAN (p. 11), it is mandatory to maintain the project records, including project data for up to 5 years after the final payment.

Resources for long-term preservation, associated costs, and how data will be kept beyond the project and for how long, will be regularly discussed by the Consortium.

# 6 Data security

During the project, data and outputs will be stored on the reliable and secured storage system of the partners. Every partner is responsible for ensuring that the data are stored safely and securely and in compliance with European Union data protection laws and in accordance with Article 27 (Appendix A). After the completion of the project, all the responsibilities concerning secure storage will go to the repository storing the dataset.

Data and outputs will be stored for the long-term in a certified repository that is in line with standard security requirements.

Only authorized personnel working directly with the Bio4HUMAN project will be granted access to the files containing personal data.

Electronic data will be stored on encrypted platforms (KoBo Toolbox, ELO, SharePoint), while physical records of data will be stored in locked cabinets on PIN DRC/HQ premises, with access restricted to authorized persons only. PIN and PAH's internal data protection and information security policies that use GDPR-compliant tools to collect, process, store data subjects' personal data and include the pseudonymisation and encryption of personal data, and policies and procedures to ensure the confidentiality, integrity, availability will be applied. When needed, DPOs will be consulted for advice on how to achieve a level of data security that is commensurate to the potential risks to Bio4HUMAN data subjects.

## 6.1 Data security of HOs

### 6.1.2 PAH Data security

PAH'S IT structure is based on the Microsoft Azure hybrid environment configured with the Azure AD connect tool. The centralized IT structure, i.e. the on-premises environment along with local resources, is located at the HQ and is available in the internal network or via VPN. In addition, most of the resources are in the Azure environment with access to all authorized domain users including country offices. In total, data is backed up in 3 different places with different frequency depending on data priority and criticality. For daily backups PAH uses tapes and the IBM TS3100 3573 tape library. Tapes are stored outside the server room in a fireproof strongbox.

The same data is also backed up to external disk drive storage, which is part of the PAH's IT infrastructure. In addition, once a week data is backed up to the Azure cloud. In summary, PAH keeps data backup on tapes (manual backup) on external disk drive storage (automatic backup) and in the Azure cloud (automatic backup). The software used for this purpose is MABS and Veeam. In principle PAH does not have a data recovery plan, however every single backup solution is occasionally tested. PAH is in progress of implementation of a data recovery plan, but so far because of the relatively small size of the IT infrastructure it was not officially in usage. For now, PAH's IT service provider is responsible for data recovery.

### 6.1.3 PIN Data Security

PIN stores project quantitative data directly in KoboCollect that uses the Humanitarian Server, which is a cloud server hosted by UN Office for the Coordination of Humanitarian Affairs (OCHA). PIN also stores their qualitative data in restricted folders on SharePoint, which also stores data in an online cloud. This means that even if PIN's work devices face technical challenges, their data is safely accessible from any other device given that the user has access rights and credentials.

# 7 Ethics

All the ethics issues related to the data collection and use during the Bio4HUMAN project are being covered by WP9 Ethics and its deliverables. Namely:

D9.1 H - Requirement No. 1 covering the procedures and criteria that will be used to identify/recruit research participants (including inform consent and preservation of data)

D9.2 POPD - Requirement No. 2 covering the processing of personal data (including a description of the security measures that will be implemented to prevent unauthorised access to personal data and Description of the anonymisation/pseudonymisation techniques that will be implemented, personal data transfers from a non-EU country to the EU or another third state)

The following general practices will be applied:

- If personal data is collected, the General Data Protection Regulation will be respected.
- In cases where access to personal data by the partners is granted, they are responsible for the treatment of said data and shall comply with the rules. PIN and PAH have signed the Joint Data Controller. If there is another partner who will need access to the personal data, another Joint Data Controller will be signed with this beneficiary. The informed consent forms will include the question to the respondents of they accept sharing their data with project partners for the purpose of the project.
- The informed consent for data sharing and long-term preservation will be included in any surveys or questionnaires dealing with personal data.