

Project partners:

1. A4F, Algafuel, SA (A4F)
2. Mikrobiologicky Ustav AV CR V.V.I (IMIC)
3. Forfarmers Corporate Services BV (FF)
4. Instituto de Biologia Experimental e Tecnológica (IBET)
5. International Flavors and Fragrances IFF (Nederland) BV (IFF)
6. Laboratorio Nacional de Energia e Geologia I.P. (LNEG)
7. Phycom BV (PHY)
8. Upfield Research and Development B.V. (UPF)

## MULTI-STR3AM

**A sustainable multi-strain, multi-method, multi-product microalgae biorefinery integrating industrial side streams to create high-value products for food, feed and fragrance**

**BBI-2019-SO1-D2 - Produce components for various materials, including for food and feed, from microalgae**

**Collaborative project**

**Start date of the project: 01/05/2020**

**Duration: 48 months**

### Deliverable 7.3

#### *Knowledge and Data Management Plan*

WP	7	Project management
Task	7.2	Scientific coordination

Dissemination level <sup>1</sup>	PU	Due delivery date	31/10/2020
Nature <sup>2</sup>	R	Actual delivery date	22/04/2021

Lead beneficiary	A4F
Contributing beneficiaries	All partners

<sup>1</sup> Dissemination level: **PU** = Public, **CO** = Confidential, only for members of the consortium (including the BBI), **CI** = Classified, information as referred to in Commission Decision 2001/844/EC.

<sup>2</sup> Nature of the deliverable: **R**: Document, report (excluding the periodic and final reports) , **DEM**: Demonstrator, pilot, prototype, plan designs, **DEC**: Websites, patents filing, press & media actions, videos, etc., **OTHER**: Software, technical diagram, etc.

WP 7:	A4F	Author
	A4F	Approval by WP leader
	A4F	Approval by coordinator

Document Version	Date	Partner	Comments <sup>3</sup>
V0	02/07/2020	A4F	Creation
V1	18/10/2020	A4F	Modification
V2	22/10/2020	A4F	Modification
V3	30/03/2021	A4F	Final version for evaluation
V4	22/04/2021	A4F	Final version

---

<sup>3</sup> Creation, modification, final version for evaluation, revised version following evaluation, final

## Deliverable abstract

This report is a deliverable associated with Task 7.2 “Scientific coordination” and describes the approach taken to knowledge management by the MULTI-STR3AM consortium and provides a Data Management Plan (DMP) to be followed throughout the project.

This first DMP (Deliverable 7.3) follows the structure of the Horizon 2020 DMP template and has been developed according to Horizon 2020 open data requirements and implementation guidelines as well as FAIR data management principles. It includes details on what data will be collected, processed and/or generated by MULTI-STR3AM, how collected and generated data will be handled during and after the project, whether data will be shared/made open access and lastly how data will be preserved so that it remains accessible and usable following the completion of the project.

This publicly available document represents an initial version of MULTI-STR3AM’s DMP. Future iterations will include the results of data collection and generation activities performed during the project and outline the procedures for management of these.

## Table of Contents

---

<b>Table of Contents</b> .....	<b>4</b>
<b>1 Introduction</b> .....	<b>5</b>
1.1 Purpose of this document.....	5
1.2 Intended audience .....	5
1.3 Links to other tasks and WPs within MULTI-STR3AM .....	5
<b>2 Knowledge and data management in MULTI-STR3AM</b> .....	<b>5</b>
2.1 Background and overall knowledge management strategy .....	5
2.2 Role of the Data Management Officer .....	6
<b>3 Data Management Plan</b> .....	<b>6</b>
3.1 Data summary.....	6
3.1.1 Data related to WP1: Market analysis, exploitation and impact.....	7
3.1.2 Data related to WP2: Strain validation and biomass production .....	8
3.1.3 Data related to WP3: Biomass refining .....	8
3.1.4 Data related to WP4: Product demonstration and validation .....	9
3.1.5 Data related to WP5: Social, environmental and economic impact assessment .....	9
3.1.6 Data related to WP6: Communication and dissemination .....	10
3.1.7 Data related to WP7: Project management .....	10
3.2 FAIR data.....	10
3.2.1 Open data produced by MULTI-STR3AM .....	10
3.2.2 Non-open data produced by MULTI-STR3AM .....	11
3.3 Allocation of resources .....	11
3.4 Data security .....	11
3.5 Ethical aspects .....	12

## **1 Introduction**

---

### **1.1 Purpose of this document**

This public document presents the first DMP for the MULTI-STR3AM project. It describes the overall Data Governance in the project including the lifecycle of data to be collected, generated, used or processed within the project and the handling of data, including methodologies, data sharing, privacy and security considerations, legal and regulatory requirements, informed consent, open access, for during and after the project. The goal of considering these aspects of data management is to ensure data is well managed during the project, ensuring its effective implementation, and properly preserved for future use, ensuring the full potential and value of the data is retained.

This document describes the current status of the DMP at the time of delivery, October 2020, and will be refined by future deliverables of the DMP and updates in individual Work Packages (WPs).

### **1.2 Intended audience**

The intended audience of this document is both the participants in the MULTI-STR3AM and third parties wishing to learn more about data management practices within this project.

### **1.3 Links to other tasks and WPs within MULTI-STR3AM**

Task 7.2 on scientific coordination, with which this deliverable is associated, forms part of WP7 on overall project management, but is closely linked with all other WPs of MULTI-STR3AM as each WP entail the collection, generation and handling of data of either private, corporate or scientific nature.

In WP1, data will be generated related to market research (T1.1) and in the form of transcripts of policy meetings (T1.4) and stakeholder databases (T1.2). Additionally, T1.3 is closely linked to T7.2 since it identifies commercially and non-commercially exploitable results and consider their direct and indirect value for different stakeholders. For WP2, chemical, biological and genetic data will be collected into strain dossiers (T2.4) and operational data and documentation will be produced for production and transport manuals (T2.6). WP3 will generate data relating to process optimisation (T3.4), compositional analyses (T3.5) and design specifications for the MULTI-biorefinery (T3.1). WP4 will generate documentation in the form of specification sheets for the microalgae products (T4.1) as well as scientific performance data of the different microalgae ingredients (T4.3-T4.5). WP5 will collate and process data from all other WPs on the social, environmental and economic impacts of MULTI-STR3AM (T5.1) and will use these data to perform a TEA (T5.3) and LCAs (T5.4). In WP6 names and email addresses will be collected as a part of a newsletter signup (T6.2) and finally WP7 will entail collection of financial data as well as personal data of the project participants (T7.3).

## **2 Knowledge and data management in MULTI-STR3AM**

---

### **2.1 Background and overall knowledge management strategy**

MULTI-STR3AM will demonstrate the first large-scale solution for sustainable and cost-effective microalgae production, responding to the growing interest in and demand for sustainable products. It

provides Europe with the first dedicated multi-strain, multi-process and multi-product biorefinery ('MULTI-biorefinery'), which will refine microalgae biomass into high-quality, in-demand products: lipids including omega-3 and omega-6 fatty acids for feed and food applications; protein for feed and fragrance applications; pigments such as carotenoids and phycocyanin for food and feed applications; and low molecular weight (MW) organic compounds for fragrance applications. MULTI-STR3AM covers the entire value chain, from biomass production through to the integration of individual technologies into three main processing streams together with three large end users.

As a result of the complex and interdisciplinary nature of the MULTI-STR3AM project, and the involvement of stakeholders from different sectors, partners have carefully considered the different types of data involved and their different needs in terms of confidentiality. For the overall knowledge management strategy, green open access will be the preferred model of publishing to allow online access to the peer-reviewed scientific publications originating from the project. While the MULTI-STR3AM partners will ensure open access to scientific publications and public deliverables, it is not participating in the Open Research Data Pilot initiative, the project will generate and handle many confidential commercial results that will be exploited after the project. Data management within the project will follow standard quality criteria, as laid out by Project Management Institute (PMI) standards and EU Official guidelines for data management in H2020 for quality assurance, and will respect the principles of accuracy, correctness, currency, completeness and relevance. Data management will also follow standards as set out by FSSC 22.000 and GMP certifications. For results which are not foreseen to be exploited commercially by the consortium, dissemination of knowledge will take place by providing an open access to research data generated by the project, in publications, conference settings and in online repositories.

## **2.2 Role of the Data Management Officer**

All data management, knowledge creation and management activities in MULTI-STR3AM are supervised by the Steering Committee as a part of WP7. The Coordinator has appointed the PMO, Dr. Mariana Doria, as Data Management Officer, who will be a part of the Project Management Office located at the premises of A4F. The Data Management Officer will be supported in the implementation of this DMP by the Exploitation Committee, established as part of WP1 and led by Exploitation Manager (EM) Bruno Sommer Ferreira from A4F. The Data Management Officer will review this DMP on an ongoing basis and revise it at least twice during MULTI-STR3AM's implementation (in April 2022 and April 2024) and at any other time it is required and agreed upon by all partners.

## **3 Data Management Plan**

---

### **3.1 Data summary**

The main categories of data foreseen to be collected or generated by MULTI-STR3AM are:

- **Underlying research data:** This category encompasses the data, including associated metadata, forming the basis of results and conclusions presented in scientific articles and in any potential patents arising from the project. To remove any limitations to review and validation of results by the scientific community, green open access (self-archiving) will be the preferred model of

publication for scientific articles. Additionally, the underlying data will be deposited in an open repository (independent of the project), which will be linked to in the resulting article.

- **Operational data:** This includes raw or curated data arising from the operation of equipment, for example associated with biomass cultivation, fractionation and purification of microalgae components, and routine analyses of the resultant products (e.g., compositional analyses). Data related to the production process will be used to produce guidelines for optimal performances, quality checks and confirmation checks, which will be of use in the project and in future planned production of algae. This category of data is likely to contain commercially sensitive data; careful consideration will be given to which information can be published openly (e.g., for dissemination purposes) and which should be consideration non-open. Some of this data is also of value for scientific or other publications and presentations and will be treated accordingly.
- **Impact monitoring data:** Primarily in WP5, data will be gathered to assess the social, environmental and economic impact of MULTI-STR3AM and to track the performance of the project against the KPIs set out in the proposal. These data include biorefinery process modelling and data gathered on e.g., feedstock, raw materials, energy, waste and emissions to complete life cycle and social life cycle assessments. Such assessments will be performed according to methodology as defined by ISO 14040/44 and the project impacts measured with the help of computer-based tools such as SimaPro v9 (with Ecoinvent v3.5 database, and others).
- **Documentation relating to instruments and methods:** This category covers documentation needed to implement the project and reproduce its results, including SOPs from each partner for their respective processes and details of tools, methods, instruments and software.

This section will describe the kinds of data that each work package will be handling, the partner responsible for collecting and managing the data, the importance and purpose for the project of each specific data, different types and formats, to which stakeholders the data will be more useful, to whom it will be available, and when confidentiality will be requested and disclosure will be needed.

### 3.1.1 Data related to WP1: Market analysis, exploitation and impact

- Market research on specific products.
- Transcripts and notes from policy meetings.
- Stakeholder database.

Work package (WP) 1 will deal with market data, both aggregated from secondary sources and also micro data from individualized primary sources such as interviews. The responsible partner for this data will be A4F.

Copyright protected reports from market consulting companies will be acquired to support the development of the business planning and business strategy. This information will be highly valuable to the consortium with the objective of reaching products with viable commercial success at the end of the project.

Freely available patent data will also be used during this WP, in aggregated form, to identify possible trends as well as market players, and understand the innovative and competitive environment.

Meetings with different stakeholders and market experts for further analyses and comprehension, may also be recorded and transcripts made, or meeting notes will be produced and kept during the

project length. All information will be stored at the data server designated for the project located at A4F site.

All this information will be kept confidential, although the consortium partners will have full access to it given that it will be the basis to validate market numbers, and support assumptions for the different business planning.

The data will be mostly useful to the development of the business planning for algae producers but will also support the decision making with industrial partners, to understand if the future products correspond to the trends mapped.

The consortium partners will also provide strategic market data from their own database. This data will be supplied to the consortium already refined and processed or in aggregated form (not raw data), in order to respect confidentiality and copyrights related.

### **3.1.2 Data related to WP2: Strain validation and biomass production**

- Strain dossiers including chemical, biological and genetic data.

WP2 will deal with chemical, biological and genetic data disclosed in dossiers regarding the microalgae strains. The responsible partner for this data will be PHY.

All information regarding the different microalgae strains and genetical enhancement will be highly confidential and strategic for the technical and economic viability of the products under development.

All this data will be stored at the data server located at A4F, available to all consortium partners. However, this data will be more useful and strategic to the producers of microalgae and technology developers than to the end-users.

### **3.1.3 Data related to WP3: Biomass refining**

- Design and specifications for the MULTI-biorefinery.
- Data related to process optimisation.
- Compositional analyses of different fractions.

WP3 will deal with engineering design and specifications data produced during the development of the MULTI-biorefinery project.

The WP3 will develop the industrial plant building layout, process flow diagrams (PFDs) and equipment specifications, process description, energy and mass balances, and utility requirements. This data will all be compiled in reports, presented visually in drawings, and also through pictures and videos.

Standard operating procedures (SOPs), with step-by-step instructions regarding the MULTI-biorefinery operation, will be presented in a report format.

After the installation and commissioning of the MULTI-biorefinery at A4F site, and until the end of the project, process data will be continuously collected to support the process optimization activities as well as the economic and technical viability study. This data will be presented in several reports format.

The main stakeholders interested in this data will be the producers, A4F and PHY, as well as the research institutions IBET and LNEG, that will be supporting the WP3 developments.

The products derived from the MULTI-biorefinery will be continuously analysed, its fractions and their compositions. From these data will derive important reports that will provide inputs both to the

process and to the formulators at WP4. In this circumstance, the data will be highly important to all industrial partners, A4F, PHY, FF, IFF and UpF.

All data derived from WP3 will be key to the technology under development, highly confidential, and will be evaluated by the Exploitation Committee and may be objective of intellectual property (IP) protection and/ or publication of scientific papers.

The responsible partner for this data will be A4F.

### **3.1.4 Data related to WP4: Product demonstration and validation**

- Data on the purity, quality and safety of microalgae products (specifications sheets).
- Data on the performance of microalgae ingredients in the three different application areas.

WP4 will deal with purity, quality and safety of microalgae products, which will be key to produce the specification sheets. WP4 will also deal with data on the performance of microalgae ingredients in the three different application areas of the project: food, feed and fragrance.

During WP4, different tests will be developed by the three end-users' partners of the consortium – FF, IFF, and UpF, with the support of IBET. The data will be derived from laboratory and field tests to understand the performance, quality and safety behaviour of the new products and its different formulations.

This data will be used in different reports and IBET will be responsible to produce the final specification sheets. Although, during the project the data will be confidential, by the end of the project and afterwards this data may be object of communication and marketing activities from all consortium partners.

It will be highly important for the consortium to have part of this data disclosed, to inform the benefits and safety of the use of microalgae as raw material. Nevertheless, information from the final products formulation will be evaluated by the Exploitation Committee given that it can also be object of intellectual property protection.

The responsible partner for this data will be IBET.

### **3.1.5 Data related to WP5: Social, environmental and economic impact assessment**

- Collected data for the different impact categories.

WP5 will deal with engineering, chemical and biochemical process data. Information regarding equipment and process flow diagrams – mainly from WP3, will be used for the study of technical and economic viability.

Logistics data, such as industry addresses as well as forms of storage and transportation and distribution of the raw material produced and its fractions, will be also needed to develop the life cycle analyses studies and understand the sustainability level and environmental impacts of the processes proposed by the project.

In this WP, LNEG will be responsible for this data and to produce reports with optimum technical, economic, and sustainability conditions. LNEG will also generate possible future scenarios to run simulations and present to the consortium the best conditions to which the simulations will be more accurate.

Mostly, this data will be confidential to the consortium and will support the final business planning as well as understand the better conditions of the value chain – with higher efficiency and lower impact.

### 3.1.6 Data related to WP6: Communication and dissemination

- Names and email addresses collected in connection with newsletter signup.

WP6 will deal with mainly public data, both technical and marketing information. The only sensitive data will regard names and e-mail addresses that will be on the project's newsletter mailing list.

This mailing list will be formed by spontaneously sign-up activity, at a form available in the project's website. This mailing list will not be shared with third parties or disclosed publicly.

A4F will be the partner responsible for this data.

### 3.1.7 Data related to WP7: Project management

- Financial data.
- Personal data of project participants.

WP7 will deal with project management data, such as, the evolution of all WPs tasks as well as financial expenditure throughout the project's years. The only sensitive data will regard the partners professional contacts, such as names, e-mails and phone numbers.

This data will only be disclosed outside the consortium partners to the BBI-JU and the EU funding agency during the annual and periodic reports.

## 3.2 FAIR data

This section brings the concept of FAIR data – findable, accessible, interoperable and re-usable.

It is important to remark that when dealing with advances on the technology frontier, the equilibrium between disclosure and confidentiality is key: to guarantee that products and processes will reach the market, benefiting the society with more sustainable e better quality products, generating taxes; at the same time that the revelation of scientific knowledge will benefit society showing advances and promoting a “fast-track” to more technological developments.

The MULTI-STR3AM consortium will play an effort to reach successful results, launching innovative processes and products. To be economically viable, the consortium partners will evaluate which kind of data will be disclosed and which will be considered strategic for the development of a successful business model.

### 3.2.1 Open data produced by MULTI-STR3AM

Information regarding the regulatory landscape, logistics of transporting and storing microalgae biomass will all be public from the start.

All public information will be made available at the project's website. In the menu there will be a link to 'results', where scientific publications, presentations and public deliverables will be available for download as 'portable document format' (pdf) documents.

Although public data, when disclosed, will be mostly in pdf format, if it is a database that can be useful for future research and receives clearance from the Exploitation Committee, it will be made available in an open-source format, for download at the project's website.

Communication and dissemination activities will be highly incentivised among the consortium partners. Still, dissemination activities will occur after the evaluation of the content data by the Exploitation Committee. Given that it will contain scientific and technological data, if it is considered strategic for commercialization, it will be protected prior to the disclosure. If the protection occurs through patenting, the technological information will be freely disclosed by the Patent Offices after the confidentiality period of 18 months. Nevertheless, the consortium highly values all scientific dissemination activities, and intends to increase the knowledge and access to information of microalgae, processes and products derived.

The scientific and technological data, as described previously, will contain genetic information on different macroalgae strains; chemical and biochemical processes data; process flow diagrams; product safety and performance data; besides technical and economic process viability data, and sustainability evaluation data. Most data on products performance and safety and process sustainability will be disclosed with the achieved success and project conclusion.

At the end of the project, the Exploitation Committee will reunite to evaluate all data – in order to guarantee that what has not been used or considered strategic, that it will be disclosed to the innovation ecosystem.

### 3.2.2 Non-open data produced by MULTI-STR3AM

Ideally, data produced by the MULTI-STR3AM should be open by default, however, one of the following exceptions can preclude open access:

- **Personal data:** Personal data collected in MULTI-STR3AM are not open by default as their publication may pose privacy, ethical or security risks.
- **Commercially sensitive data:** Data that is planned to be exploited commercially by any of the MULTI-STR3AM partners will not be made open.
- **Third-party datasets:** Processing and using input data from sources such as third-party market studies may lead to unclear IPR situations; as a result, analyses using these data will only be made open if the underlying data is open.

## 3.3 Allocation of resources

The costs relating to making data open access where possible have been taken into consideration and will be covered by each partner as part of their respective budgets.

## 3.4 Data security

To ensure the security of all data collected in and generated by MULTI-STR3AM, a secure SSL-connected server will be set up and managed by A4F. It will be accessible to all partners but not to

outside parties. The system will be protected and confidentiality ensured. Security and accessibility will be guaranteed.

Each user will have a unique login and password that will be sent separately to each user. After the first login, the password will have to be changed by the user. After three trials to access the system, in a five minutes period, with an incorrect password, the admission of that user will be blocked. In order to re-establish the access, the user will need to contact the PMO (A4F project coordinator). In case of need to create a new “user ID”, the request will need to be addressed to the PMO. Similarly, the consortium partner will need to inform the PMO in case a user leaves the project, so his/her “user ID” will be disabled.

Also, an encrypted certificate will be associated, that is, an encrypted connection between the user’s computer and the server. Thus, an external user will not be able to identify the data that will be passing through the connection.

All data on “users ID” (names and corporative e-mails from partners) will be treated in accordance to the EU GDPR (General Data Protection Regulation).

### **3.5 Ethical aspects**

Ethical aspects related to collection and use of data in MULTI-STR3AM are mostly related to personal data, for example those collected in connection with stakeholder interviews, discussions with policy makers and names and for the purpose of newsletter signup. The collection and use of such data should be kept to a minimum and only on a “need to know” basis. With guidance from the Data Management Officer, each project partner is responsible for complying with their own Data Protection Policy and GDPR rules.

Informed consent for data sharing and long-term preservation is included in questionnaires dealing with personal data. When personal data is collected from a given individual, the individual must receive adequate information concerning the aims and procedures of the data collection, processing, storage, protection, retention and destruction in writing to ensure they make an informed decision on whether to provide their data. At any given time, the person can withdraw their consent by writing directly to the Data Management Officer, whose contact details will be clearly stated along with the above-mentioned information.



This project has received funding from the Bio-based Industries Joint Undertaking (JU) under the European Union’s Horizon 2020 research and innovation programme under grant agreement No 887227. The JU receives support from the European Union’s Horizon 2020 research and innovation programme and the Bio-based Industries Consortium.