Initial Data Management Plan (DMP)

Deliverable D6.2

INNOVATIVE DECISION-MAKING TOOL FOR DEFINING THE MOST SUITABLE MANURE MANAGEMENT STRATEGIES TO ACHIEVE A SUSTAINABLE LIVESTOCK FARMING SYSTEM DURING THE WHOLE VALUE CHAIN

Proposal number: 101135400-2

#HorizonEU

Deliverable D6.2 – Initial Data Management Plan (DMP)			
Deliverable Number	D6.2	**Lead Beneficiary**	1-MEDRAR
Deliverable Name	Initial Data Management Plan (DMP)		
Type	R – Document, Report	**Dissemination Level**	PU-Public
Due Date (month)	M6	**Work Package No**	WP6

Grant Agreement No:	101135400	Project acronym:	NUTRITIVE
Project Title:	**INNOVATIVE DECISION-MAKING TOOL FOR DEFINING THE MOST SUITABLE MANURE MANAGEMENT STRATEGIES TO ACHIEVE A SUSTAINABLE LIVESTOCK FARMING SYSTEM DURING THE WHOLE VALUE CHAIN**		
Financing scheme:	HORIZON-CL6-2023-ZEROPOLLUTION-02		
Project coordinator:	**MEDRAR**		
Principal beneficiary:	1-MEDRAR		
Project start date:	11/07/2024	**Duration of the project:**	48 month
Deliverable:	D6.2- Initial Data Management Plan (DMP)		
Contractual delivery date:			
Actual delivery date:			
Type of deliverable	R – Document, Report		
Dissemination Level	PU-PUBLIC		
Authors:	1-MEDRAR		
Contributors:			
Version:			

History of change

Version:	**Author:**	**Date:**	**Comments:**
1.0	1-MEDRAR	30/11/2024	Final Version
1.1	1-MEDRAR	17/02/2026	Reopened Version. Table 3 added and sections 3.2, 3.4 and 4 updated accordingly.

"NUTRITIVE Project - 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 [name of the granting authority]. Neither the European Union nor the granting authority can be held responsible for them."

TABLE OF CONTENTS

1. INTRODUCTION

1.1. ABOUT NUTRITIVE PROJECT

Livestock farming is a key sector that involves 40 % of the total agricultural activity in Europe. representing a total value for products equal to € 170 billion. However, there is an increasing concern due to livestock farming's contribution to environmental pollution since it generates more than 1.4 billion tonnes/year of manure leading to significant greenhouse gases (GHG) and air pollutants emissions(NH3, NOX) as well as to soil and water contamination caused by hazardous manure chemicals and biological contaminants (called here emerging contaminants). In this context extensive effort has been carried out for years to assess the detrimental effects of farming systems and to develop abatement methods to be implemented. However, despite major advancements, many fundamental issues are beyond the scope of existing legislation.

The main objective of NUTRITIVE is to develop a decision-making tool (DSS, decision support system) able to define the most efficient and sustainable (in its three pillars: environmental, economic, and social) manure management strategies for a given livestock farm limiting manure air emissions as well as soil and water contaminants. This will allow for the formulation of technical guidelines and recommendations that will support policy makers with enhanced knowledge to establish requirements for future European policies.

To fulfil this objective, the project is divided into six work packages (WP): WP1 Up-to-date inventory; WP2 Novel management strategies/technologies investigation; WP3 Modelling and Life Cycle Assessment (LCA); and WP4 Guidelines formulation; WP5 Communication, dissemination, and exploitation; WP6 Management (Figure 1).

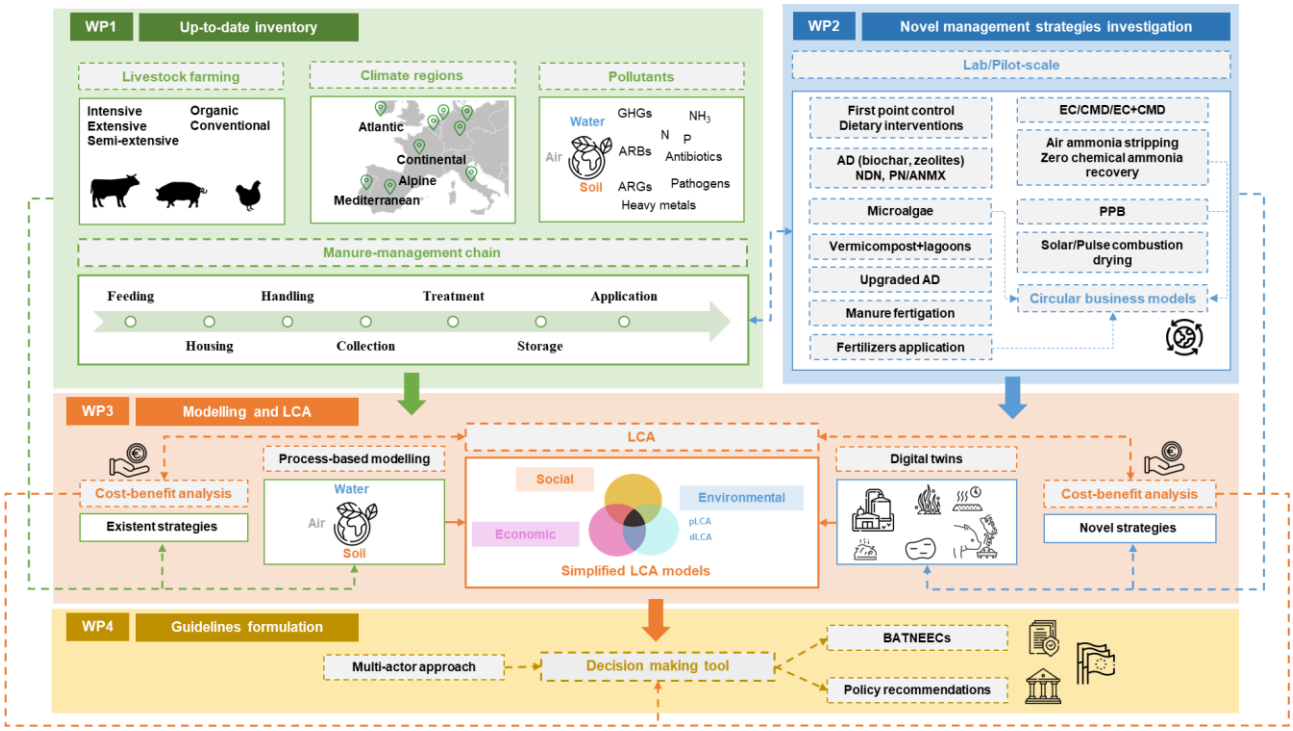

Figure 1. NUTRITIVE methodology.

NUTRITIVE anticipates a wide spread of the project outcomes, with the synthesis of the consortium as a baseline: 22 partners (4 Chinese) from 8 different countries across Europe, covering 6 climatic regions (2 Chinese ones), representing the whole supply chain experts, from animal feed to soil application.

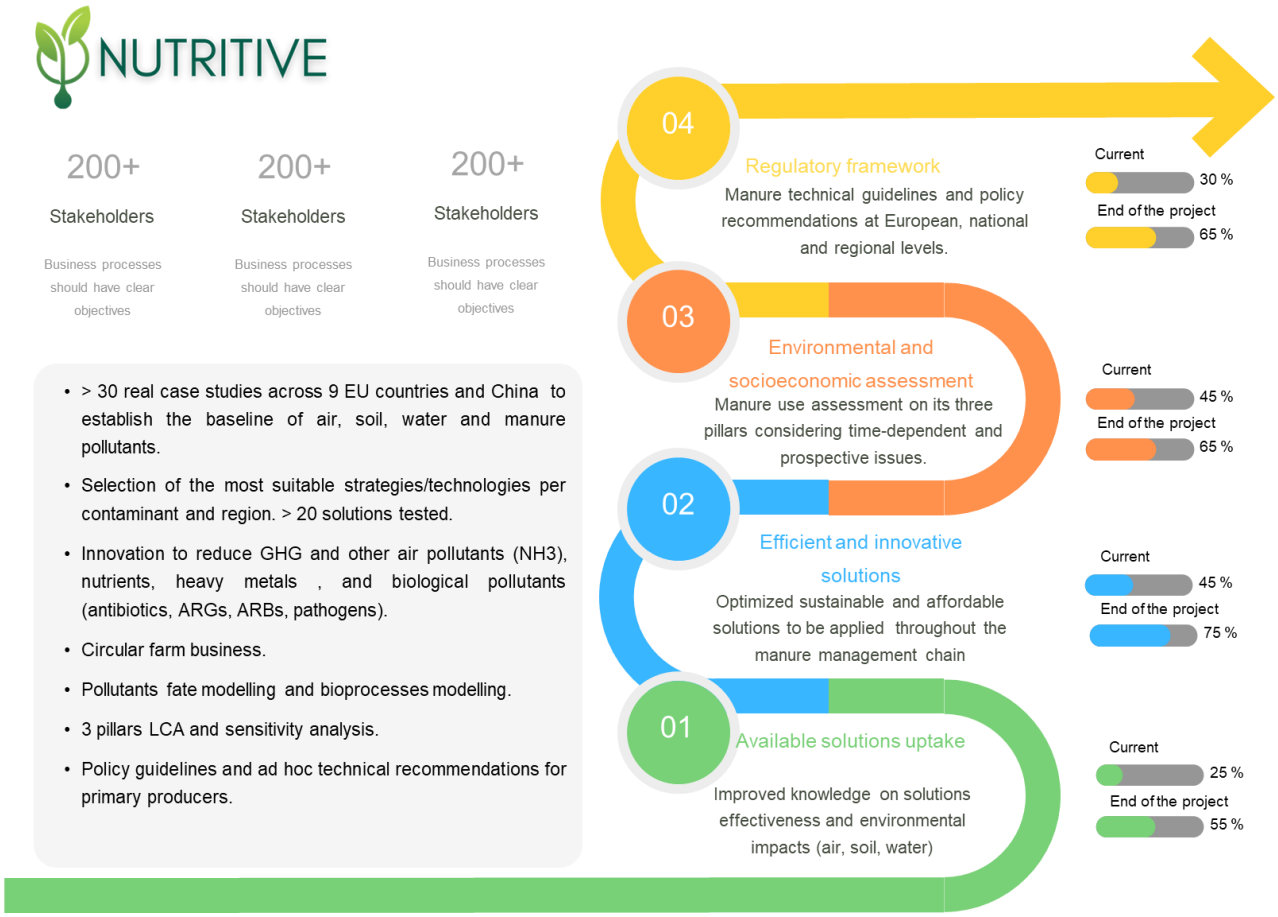

Figure 2. NUTRITIVE overview.

1.2. PURPOSE OF THIS DOCUMENT

This report is the first version of the NUTRITIVE Data Management Plan (DMP), which will be revised throughout the project, and updated in the upcoming Deliverables D6.3 (Mid-term Data Management Plan (DMP), M36) and D6.4 (Final Data Management Plan (DMP), M48). This initial DMP is an indicative plan as to what kind of data the project beneficiaries expect to generate during the project, and how these data will be managed. This DMP follows the standard DMP template for Horizon Europe (HE) projects (version 1.1).

1.3. RELATION WITH OTHER ACTIVITIES IN THE PROJECT

The DMP is related with all the project Work Packages (WPs), tasks, and subtasks. Thus, this document will provide information about data used, gathered, and shared.

1.4. CONTRIBUTION OF PARTNERS

MEDRAR, as the coordinator entity, with the support of all NUTRITIVE partners, is the main contributor of the content provided in the DMP.

2. DATA SUMMARY

2.1. DATA REUSE

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

In *Task 1.3. Air pollutants assessment*, CO_2, N_2O, and NH_3 emission factors collated to the global database DATAMAN through the FACCE ERA GAS project MELS, will be combined with the data collected in *Subtask 1.1.2. Case studies selection and diversity characterization*, regarding manure solutions, diets, animal categories, etc. DATAMAN database, together with collected data will be used to create detailed site-specific emission calculations for the whole manure management chain by using modelling approaches. Also, regarding in *Subtask 1.1.2. Case studies selection and diversity characterization* the case studies selected by ILVO will also be used in other projects (namely RAMBO/INT6B016/Interreg and PAS-boerenbond). These projects might feed the Nutritive project with additional information regarding emission measurements and manure management.

2.2. DATA TYPES AND FORMATS

What types and formats of data will the project generate or re-use?

A wide range of data types and formats are expected to be generated in the project. While these cannot be listed in advance, some general remarks can be made, and will be updated in the new versions of the DMP. Data types will vary according to the source application in which they are generated although in general, the consortium will endeavour to utilise common data formats. In some cases, particular laboratory equipment or software may output data in proprietary formats, but where sharing of these data with other partners is necessary, a conversion to or a summary in common formats may be possible. It will be used open and lossless file formats to retain all data and accessible across platforms. Data can be divided into different categories, depending on their scope, their origin as well as their processing. Data might be classified as follows:

- Observational data, captured in real-time (i.e. using sensors or surveys).
- Experimental data from e.g., WP1 analysis, and WP2 strategies/technologies operation.
- Simulation data, created using numerical models and computer simulations;
- Derived data, generated from the elaboration of already existing data.

In this context, the type and format of data which will be used may include: text documents (questionnaire, transcripts, protocols, etc.); laboratory/field notebooks; spreadsheets (for the ordered storage, analysis and elaboration of data); audios, videos, pictures, photographs; test responses; samples; data files (i.e. outputs from laboratory instrumentation); database contents; models, algorithms, software codes, scripts, log files, simulations; methodologies and workflows; standard operating procedures and protocols, slides, etc. It will be prioritised the use of open and standard formats to enhance data sharing and ensure long-term reusability.

Table 1 gives an overview about the project WPs. Table 2 gives a summary of the expected types of research data expected to be generated in the project whereas Table 3 summarises the non-research data.

Table 1. NUTRITIVE WPs.

WP	NAME	LEADER
1	Up-to-date inventory	EV ILVO
2	Novel management solutions	AINIA
3	Modelling and LCA	USC
4	Guidelines and policies	TEAGASC
5	Communication, dissemination, and exploitation	MEDRAR
6	Project management	MEDRAR
7	Ethics requirements	MEDRAR

Table 2. Types and formats of research data to be generated and used in the project.

WP	TYPE OF DATA	DATA FORMAT
1	Data collected from case study farms, literature review information, Standard Operating Procedures (SOP) and other reports generated, analytics results.	Lists and tables (.xls, .txt, .csv, .tab) Text (.rtf, .txt, .pdf)
2	Analytics results, data from strategies/technologies performance, reports generated.	Lists and tables (.xls, .txt, .csv, .tab) Text (.rtf, .txt, .pdf)
3	Inventory data, simulation models, reports generated.	Models scripts (.htm) Lists and tables (.xls, .txt, .csv, .tab) Text (.rtf, .txt, .pdf) Simulation (.spf, .py,.ipynb)
4	Regulations review information, qualitative research data, surveys data, reports generated.	Lists and tables (.xls, .txt, .csv, .tab) Text (.rtf, .txt, .pdf)

Table 3. Types and formats of non-research data to be generated and used in the project.

WP	TYPE OF DATA	DATA FORMAT
5	Project website and social media statistics, project videos and other materials and reports generated.	Text (.rtf, .txt, .pdf) Video (.mp4)
6	Project management data (e.g., financial information, resources), reports generated.	Lists and tables (.xls, .txt, .csv, .tab) Text (.rtf, .txt, .pdf)
7	Reports generated	Text (.rtf, .txt, .pdf)

Also, it was prepared a table for the register of datasets and other results for their report in the Grant Management Services (Table 4).

Table 4. Register of datasets/other results aligned with the Continuous Reporting ("Results" tab).

CONTENT REQUESTED	
Type of result	Dataset Software Workflow Protocol Prototype Other
Description	
If the result is needed to validate the conclusions of a publication, briefly describe the provisions whereby you intend to make your output available, either in digital or physical form? * Please elaborate	It does not underpin publication Other * Open access
Type of PID (if available)	DOI Hande ARK URI Other None
PID (if available)	
URL to repository landing page for the result service/webpage hosting the result (required for datasets; recommended for other results)	

2.3. PURPOSE OF DATA GENERATION OR RE-USE AND RELATION TO PROJECT OBJECTIVES

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

It is necessary to generate and reuse data in order to achieve the project's Key Objectives (KO).

- **KO#1. To create an up-to-date inventory including available and experimental solutions covering the entire manure management chain and considering different types of livestock across European regions**. Data is necessary to be able to generate this inventory.
- **KO#2: To investigate circular, innovative, and cost-effective solutions to reduce GHG emissions, and air, water and soil pollutants produced by livestock manure management throughout its chain.** It is necessary to have data on emissions, animal diet, energy consumption, storage facilities and pollutants from manure management.
- **KO#3: Optimize management strategies to achieve highest efficiencies with lowest environmental emissions; simulating the biological activity of the bioprocesses and the transport and persistence of manure-derived contaminants in air, soil, and water among different climate regions.** Models and data associated with the simulations will be generated.
- **KO#4: To assess manure management across all areas of sustainability (environmental pollution, cost-benefit of practices/technologies, and social acceptance) overcoming the uncertainty of emerging technologies and including time-depending issues.** In order to carry out this assessment, it is necessary to generate Life Cycle Inventories for each scenario and run different simulations. The different technologies will be compared from a Life Cycle Assessment approach and data generated from simulation software will be used to perform the techno-economic analysis.
- **KO#5: To obtain simplified LCA models that can help non-expert users to easily obtain estimates of environmental impacts to be considered in the decision-making tool.** Related to the previous point is the need to obtain models that can allow non-expert users to make estimates.
- **KO#6: To integrate all the generated knowledge into a decision-making tool (DSS tool).** All the data generated will be integrated into a tool that will provide valuable information for decision making in manure management.
- **KO#7: To formulate technical guidelines and recommendations for manure management based on multiactor approach engagement.** On the basis of the results obtained, recommendations will be drawn up.

2.4. EXPECTED SIZE OF THE DATA

What is the expected size of the data that you intend to generate or re-use?

At this state of the project, the expected size of the data cannot be defined precisely. However, an overall estimate of the expected volume of research data shows that current storage capacity will not be exceeded.

- **WP1** (Expected order of magnitude: MB)
  - Characterisation of > 40 case studies across EU.
  - SOP for soil, water, and manure.
  - Air pollutants assessment in > 40 case studies.
  - Analytics results from > 150 soil, water, and manure samples from > 40 case studies (2 seasons, duplicate).
  - Analytic data from literature review carried out in *Subtask 1.1.1. Literature review.*
- **WP2** (Expected order of magnitude: MB)
  - Data from the operational performance of 22 manure management solutions, plus strategies/technologies intercomparison.
- **WP3** (Expected order of magnitude: GB)
  - At least 5 digital twins of the manure management biotechnologies demonstrated in WP2, and 3 combinations of solutions.
  - Fate models of around 20 different pollutants, and at least simulations of 5 types of emissions regarding soil and water contaminants.
  - LCA in at least 20 full-scale, and 5 laboratory/pilot scale processes.
  - Simplified models for the different technologies and impact categories.
- **WP4** (Expected order of magnitude: TB)
  - Qualitative information from 15- 20 semi-structured interviews, and 3-5 focus groups; > 300 surveys with farmers for awareness and acceptability analysis.
  - Portfolio of manure management solutions.
  - DSS tool integrating WP1, WP2, and WP4 outcomes with the APIs needed.

2.5. ORIGIN/PROVENANCE OF THE DATA

What is the origin/provenance of the data, either generated or re-used?

Most of the data in the project will be generated in-project. The re-used data will come from external and open sources databases as well as from internal databases from the project partners. Databases provided will be assessed beforehand for their suitability by the respective partner. Only technically and legally suitable data from external databases will be used in the project.

2.6. DATA UTILITY

To whom might your data be useful ('data utility'), outside your project?

Data generated in NUTRITIVE will be useful for the project (consortium); but also for other audiences mainly the scientific community (university/research centres) in the field of manure management, pollutants addressing, novel solutions development, and sustainability assessment that are developing other projects in this field; policymakers, public institutions, and other agencies and bodies; and also companies of the green technological industry, waste managers and extension services. Nevertheless, many data generated in this project are subject to intellectual property rights or are part of confidential deliverables and therefore cannot be published. However, data published in public deliverables are made freely available if there is consent of all project partners.

3. FAIR DATA

3.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

Digital Object Identifier (DOI)

Will data be identified by a persistent identifier?

All open data, publications and open source software produced in NUTRITIVE will be identifiable and locatable by means of a persistent Uniform Resource Locator (URL). If possible, open P2R results will be assigned a Digital Object Identifier (DOI) in order to make content easily and uniquely citable. Open results deposited in the Open Access repository Zenodo will be assigned a DOI automatically and will benefit also from Zenodo's DOI versioning support. Open results deposited in institutional repositories, repositories of scientific publishers or other data and research repositories will be at least definable by a persistent URL. If the institution is a DOI registrant that has an agreement with a DOI registration agency, a DOI will be assigned, too. Whether scientific publications will be assigned a unique identifier like DOI, Publisher Item Identifier (PII), International Standard Serial Number (ISSN), etc. depends on the open access strategy (green or gold) chosen by the editors and thus also on the respective scientific publisher and the chosen research repository.

Rich metadata

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

There are many different metadata standards for many different types of data and it may not be possible to find one that fits all purposes. Therefore, a pragmatic and feasible approach is to agree on a common and minimal catalogue metadata schema for those datasets that are published in public catalogues and data repositories and to use data-type specific schema extensions, if necessary. For example, README files and data dictionaries should be part of this metadata both while the project is active and when it comes to sharing data on Zenodo's repository.

In general, Zenodo's deposition metadata domain model which is based on DataCite's Metadata Schema minimum and recommended terms will be used for open data generated by the project and deposited in appropriate repository.

Search keywords

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Data must be findable easily, rapidly and identically. Therefore, exact and standard measures have to be used to identify the data sets. This will include the definition and use of naming conventions, search keywords, version numbers, metadata standards and standard data identifiers.

Data harvesting and indexing

Will metadata be offered in such a way that it can be harvested and indexed?

Datasets published in NUTRITIVE will be harvested and indexed using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) in such a way to allow exchange with other repositories. At the same time, the metadata provided will be as detailed as possible to allow it to be indexed so that the data is searchable and retrievable.

3.2. MAKING DATA ACCESIBLE

Repository

Will the data be deposited in a trusted repository?

In general, research data produced within WP1, WP2, WP3, and WP4, will be deposited in a trustworthy repository (e.g., Zenodo, which allows for providing open access, embargoed access, or closed access).

All the data associated with scientific publications will be made openly available as the default unless there is a specific reason not to publish the data. Datasets which cannot be shared may be made available on a case-by-case basis if it is relevant for third parties.

The following data will not be made publicly available:

- Data obtained with the permission of third parties but these have not agreed to make the data publicly available.
- Data that discloses the identity of a farm/exploitation;
- Data that compromises the protection of a partner(s) intellectual property. The level of data made available will also be considered, for example, pre-processed data will not be provided unless there is a clear reason for doing so.

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Different storage solutions were evaluated, and others are still under evaluation. The current solution was chosen as the best solution in terms of cost/benefits.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

Zenodo applies DOIs to datasets.

Table 3 will be considered to list all datasets and other results reported in the Grans Management Services. For each item, it will provide, when not publicly available, a clear justification will be provided according to the principle “as open as possible as closed as necessary”.

Data

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

All of the data associated with scientific publications will be made openly available as default unless there is a specific reason not to publish the data. Other data may be made available on a case-by-case basis if it is relevant for third parties.

The following data will not be made publicly available:

- Data obtained with the permission of third parties but these have not agreed to make the data publicly available.
- Data that discloses the identity of a farm/exploitation;
- Data that compromises the protection of a partner(s) intellectual property. The level of data made available will also be considered, for example, pre-processed data will not be provided unless there is a clear reason for doing so.

For restricted datasets, data will remain openly available, while files will be deposited with restriceted access; access may be granted upon request and evaluation by the data owner (as should be described in the list of datasets according to Table 3).

In general, research data produced within WP1, WP2, WP3, and WP4, will be deposited in a trustworthy repository (e.g., Zenodo, which allows for providing open access, embargoed access, or closed access). Personal and sensitive data will be anonymised if it is going to be shared. Besides, consent forms will be prepared to communicate that participant’s anonymised data will be shared on an open data repository for other to use.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

It is possible that data may need an embargo period. In Zenodo users may deposit content under an embargo status and provide end date for the embargo. The repository will restrict access to the data until the end of the embargo period; at which time, the content will become publicly available automatically.

Will the data be accessible through a free and standardized access protocol?

As previously indicated, in Zenodo files may be deposited under closed, open, or embargoed access. Files deposited under closed access are protected against unauthorized access at all levels. Access to metadata and data files is provided over standard protocols such as HTTP and OAI-PMH. Besides, contact details of the data repository will be made available to enhance accessibility.

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

Some files may be restricted. Zenodo allows for depositing restricted files with the ability to share access with others if certain requirements are met. These will not be made publicly available and sharing will be made possible only by the approval of the depositor of the original file. Contact details of the data deposited will be made available for this purpose.

How will the identity of the person accessing the data be ascertained?

- Data Centre. Zenodo's data centres is located on CERN premises and all physical access is restricted to a limited number of staff with appropriate training and who have been granted access in line with their professional duties (e.g. Zenodo staff do not have physical access to the CERN Data Centre) .
- Servers. Zenodo's servers are managed according to the CERN Security Baseline for Servers, meaning e.g. remote access to our servers are restricted to Zenodo staff with appropriate training, and the operating system and installed applications are kept updated with latest security patches via Zenodo's automatic configuration management system Puppet.
- Network. CERN Security Team runs both host and network based intrusion detection systems and monitors the traffic flow, pattern and contents into and out of CERN networks in order to detect attacks. All access to zenodo.org happens over HTTPS, except for static documentation pages which are hosted on GitHub Pages.
- Data. Zenodo stores user passwords using strong cryptographic password hashing algorithms (currently PBKDF2+SHA512). Users' access tokens to GitHub and ORCID are stored encrypted and can only be decrypted with the application's secret key.
- Application. Zenodo employs a suite of techniques to protect sessions from being stolen by an attacker when you are logged in and run vulnerability scans against the application.
- Staff. CERN staff with access to user data operate under CERN Operational Circular no. 5, meaning among other things that: staff should not exchange among themselves information acquired unless it is expressly required for the execution of their duties; access to user data must always be consistent with the professional duties and only permitted for resolution of problems, detection of security issues, monitoring of resources and similar; staff are liable for damage resulting from any infringement and can have access withdrawn and/or be subject to disciplinary or legal proceedings depending on seriousness of the infringement.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

It is difficult to speculate as to whether a data access committee will be required. It seems unlikely that committee will be required. Personal identifiable information must be stored in a separate, password-protected folder to research data.

Metadata

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

In NUTRITIVE the metadata will be made openly available, and datasets will be licenced under a public domain dedication CC0. If data has been reused from other sources it may be considered the license applied to that datasets when applying a license of own data. Information about the accessibility of the dataset is also included in the metadata.

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Items will be retained for the lifetime of the repository. Data will be retained for a minimum of 10 years after the project end. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

If software is needed, documentation will be included in the dataset, as well as the open source code software.

3.3. MAKING DATA INTEROPERABLE

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

Data produced in the project will be interoperable as the datasets will adhere to standardised formats: ASCII, .txt, .csv, .xml, etc. If MS Office, pdf viewer or image viewer cannot be used, a text (ASCII) file will be provided with the dataset that explains where a free reader can be obtained.

Data will be deposited in Zenodo, which uses JSON Schema as internal representation of metadata and offers export to other popular formats such as Dublin Core or MARCXML. This data repository is OpenAIRE compliant and meets all the requirements of metadata required by the European Commission.

Persistent IDs will be provided for each document (DOI) and author identifiers (ORCID and contact details) will be included in the metadata. The metadata standard used to describe the dataset is the DDI's metadata schema compatible with the Dublin Core, a flexible and commonly used standard that is also adopted by the European OpenAIRE repository.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

In the project common terminology and standard vocabulary will be used in the field to describe data.

Will your data include qualified references1 to other data (e.g. other data from your project, or datasets from previous research)?

(Meta)data will include qualified references to other (meta)data. In Zenodo each referenced external piece of metadata will be qualified by a resolvable URL. To ensure interoperability, DOIs of research outputs will be linked whenever possible, and all published datasets will include qualified references to the broadest level possible.

3.4. INCREASE DATA RE-USE

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

All documentation needed to validate data analysis and facilitate data re-use will accompany the data via a README file created in accordance with the NUTRITIVE template and guidelines.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Open datasets will be released under a standard licence (CC BY 4.0 or CC0, as appropriate and indicated per dataset in Table 3). Software/code outputs will be released under an appropriate open-source licence (to be specified per item in Table 3). For datasets that cannot be made publicly available (e.g., third-party restrictions, farm identification, or protection of partners' intellectual property), no public licence applies; access conditions and the justification for restricted access are provided in Table 3, following the principle "as open as possible, as closed as necessary".

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the Horizon Europe rules. Raw data collected through the interviews from externals may be available to the whole institution or specific partners.

Will the provenance of the data be thoroughly documented using the appropriate standards?

[1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

The documentation and metadata of each dataset recognize the data provenance through proper citation of the source of information and entities using the formats usually accepted by the relevant scientific community.

Describe all relevant data quality assurance processes.

The data quality will be ensured by different measures. For NUTRITIVE research data collection, the quality control of the data can happen at various stages during the quality assurance process. Initial quality control is needed and early in the collection process. Additional controls will take place at a later stage of the data lifecycle. Final quality control of metadata takes place during its input into the DSS. The quality of the data collection methods used strongly influences data quality, and documenting in detail how data are collected provides evidence of such quality. Errors can also occur during data entry. Data are digitised, transcribed, entered in a database or spreadsheet, or coded. Here, quality will be ensured by standardised and consistent procedures for data entry with clear instructions. Also, NUTRITIVE will adhere to the European Code of Conduct for Research Integrity

4. OTHER RESEARCH OUTPUTS

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

NUTRITIVE is expected to generate additional research outputs beyond datasets (e.g., software/code, models, workflows and protocols such as SOPs). These outputs will be registered and managed according to FAIR principles, and reported according to Table 3 together with their access conditions, repository landing pages and PIDs where applicable.

5. ALLOCATION OF RESOURCES

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

Specific costs for making research outputs FAIR in NUTRITIVE project, have been included within specific WPs (e.g., cloud platform, open access publications). However costs associated with data management are not included. OpenAire cost matrix can be considered to estimate this costs. Besides, it needs to be considered long term costs of data preservation since although Zenodo may be free, there may be other data that needs to be archived for long term preservation that is not suitable for this platform.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

The costs will be covered through the funds from the Horizon Europe grant. Data management will be internally supported by staff from the different institutions that are part of the consortium free of charge.

Who will be responsible for data management in your project?

Each partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

- The Project Coordinator (MEDRAR) has a particular responsibility to ensure that data shared through the website are easily available, but also that backups are performed and that proprietary data are secured.
- WP1, WP2, WP3, and WP4 leaders (EV ILVO, AINIA, USC, and TEAGASC, respectively), will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.
- Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP.
- Backing up data for sharing through open access repositories is the responsibility of the partner possessing the data.
- Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator (MEDRAR).

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Regarding the question of long-term data preservation, no specific arrangements has been done in the consortium yet. Nonetheless, it needs to be considered that all data will need to be preserved (even data that is not shared on Zenodo after the project finishes) for at least 10 years. In this regard, issues like space, size, or security need to be considered.

6. DATA SECURITY

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

During the course of the research project, all data will be stored on local servers maintained and automatically backed up by MEDRAR. Data can be recovered with the help of MEDRAR services in the event of an incident. Only team members have access to the designated server, limited to the principal investigator of the project. The storage security is ensured by MEDRAR services. It will provide additional advice, as needed, on data storage during the research project. This data storage solution offers secure storage and transfer. If personal data are processed adherence to data protection and security legislation will be maintained. It will be provided a managed access procedure in place for authorised users of personal data.

The access to NUTRITIVE cloud is controlled by MEDRAR. A General Data Protection Regulation notice to access project contained a deed of designation of the persons authorized to process personal data was shared with those appointed as authorised subjects to process personal details

in terms of and pursuant to art.29 of the GDPR and to the applicable national data protection law. It can be consulted in this link: NUTRITIVE GDPR notice to access project sharepoint.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Data will be stored in Zenodo according to the following issues.

- **Versions**. Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and record are preserved.
- **Replicas**. All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- **Retention period**. Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
- **Functional preservation**. Zenodo makes no promises of usability and understandability of deposited objects over time.
- **File preservation**. Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- **Fixity and authenticity**. All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
- **Succession plans**. In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

7. ETHICS

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

All the activities carried out under the NUTRITIVE project comply with ethical principles and relevant national, EU and international legislation, for example the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. Ethical issues will be addressed in WP7, focused on ensuring compliance with ethics requirements that the project must comply with. An External Ethics Advisory Group (EEtAB) has been constituted for monitoring the development of the project, with expertise in the major ethical issues identified (i.e., Artificial Intelligence (AI), and human health and safety). Relevant ethics reports covering all ethics issues identified will be submitted at the end of each reporting period.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

Research uses questionnaires dealing with personal data, so informed consent will be included with the exception of data that discloses the identity of a farm/exploitation. The consent form, where

necessary, will mention explicitly that respondents' anonymised data will be publicly shared on an open repository.

8. OTHER ISSUES

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

As well as general policies on open data management, project partners must also adhere to their own institutional policies and procedures for data management. The project will adhere to the European Code of Conduct for Research Integrity.

9. CONCLUSIONS

During the first months of project implementation, work began on the DMP in accordance with the characteristics of the project and the data management requirements of the European Commission. However, this is a very preliminary version that will be expanded and updated during the course of the project.

The typology and format of the data being or planned to be used (and reused) in the project were identified in each work package, taking into account both research and non-research data. The need to generate and reuse data in relation to the different Key objectives of the NUTRITIVE project; the origin and provenance of the data; and their utility outside NUTRITIVE was also assessed. In addition, although at this stage it is difficult to determine the size of the data expected to be generated, an estimate of the order of magnitude associated with each work package was made.

On the other hand, fair data principles were analysed in order to ensure that NUTRITIVE data will be findable, accessible, interoperable, and reusable. Finally, further to the FAIR principles, during the development of the initial version of the NUTRITIVE DMP, it was also addressed other research outputs, and considered aspects related to the allocation of resources, data security and ethical aspects.

During M1-M6, data management was mainly focused on WP1, and attempts were made to carefully apply the practices considered throughout this document. For example, for the elaboration of Deliverable 1.1, the owners of different livestock farms were contacted by WP1 partners in order to carry out the final identification of the farms to be worked with throughout the project.

For this purpose, a questionnaire was prepared and farmers were informed that personal and sensitive data was going to be always anonymised before sharing it. In addition, consent forms were be prepared so that (if necessary), partners can communicate that participant's anonymised data will be shared on an open data repository for others to use. Initial research uses questionnaires developed in WP1 did not deal with personnel data, so informed consent was not needed to be included.