



GIANT LEAPS

DATA MANAGEMENT PLAN (DMP)

Project acronym: GIANT LEAPS

Project title: Gap resolution in sAfeTy, NuTritional, aLLergenicity and Environmental assessments to promote Alternative Protein utilization and the dietary Shift

Call: HORIZON-CL6-2021-FARM2FORK-01



Co-funded by
the European Union



Project no. 101059632
Project acronym: GIANT LEAPS
Project title: Gap resolution in sAfeTy, NuTritional, alLergenicity and Environmental assessments to promote Alternative Protein utilization and the dietary Shift
Call: HORIZON-CL6-2021-FARM2FORK-01
Start date of project: 01.09.2022
Duration: 48 months
Deliverable title: D9.1 (Data Management Plan (DMP))
Due date of deliverable: 28.02.2023
Actual date of submission: 28.02.2023
Deliverable Lead Partner: Wageningen Research (WR)
Dissemination level: Public

Name	Organization
HJ van der Fels-Klerx	Wageningen Research (WR)
Paul Vos	Wageningen Research (WR)

Document History

Version	Date	Note	Revised by
01	28.02.2023	Initial submitted version	The authors
02	DD.MM.YYYY	info	Name



Disclaimer

While GIANT LEAPS is funded by the European Union, views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the European Research Executive Agency (REA) can be held responsible for them.



Executive Summary

This document contains the GIANT LEAPS Data Management Plan (DMP), Deliverable 9.1. This is the initial version developed at M6 and the document will be reviewed and updated, if relevant, every 6 months. This DMP is developed using the online resource DMP online¹ using the Horizon Europe template.

¹ dmp.wur.nl



Table of Contents

Disclaimer	2
Executive Summary	3
Table of Contents	4
List of Acronyms	5
Plan Overview	6
Gap resolution in sAfeTy, NuTritional, aLLergenicity and Environmental assessments to promote Alternative Protein utilization and the dietary Shift	7
Data Summary	7
FAIR data	8
Other research outputs	14
Allocation of resources	14
Data security	15
Ethics	15
Other issues	15



List of Acronyms

D	Deliverable
DMP	Data Management Plan
DOI	Digital Object Identifier
DQA	Data Quality Assurance
EC	European Commission
EFSA	European Food Safety Authority
EU	European Union
HE	Horizon Europe
IPR	Intellectual Property Rights
JRC	Joint Research Centre
M	Month
N.A.	Not applicable
WP	Work Package



Plan Overview

A Data Management Plan created using DMPonline

Title: Gap resolution in sAfeTy, NuTritional, aLlergenicity and Environmental assessments to promote Alternative Protein utilization and the dietary Shift

Creator: Paul Vos

Principal Investigator: Ine van der Fels-Klerx, Paul Vos

Affiliation: Wageningen University and Research (Netherlands)

Funder: European Commission

Template: Horizon Europe Template

ORCID iD: 0000-0002-7801-394X

Project abstract:

Accelerating the transition from animal-based to alternative dietary proteins – the dietary shift – is key to reducing the footprint of our food system in terms of greenhouse gas emissions (GHG), energy, water and land use, and other relevant environmental impacts, and for improving the health and well-being of people, animals and the planet. GIANT LEAPS delivers the strategic innovations, methodologies, and open-access datasets to speed up this dietary shift, in line with the Farm-to-Fork strategy and contributing to the Green Deal target of reaching climate neutrality by 2050. Achieving the dietary shift in practice is inherently complex due to the diverse set of actors involved and further hindered by major knowledge gaps, scattered across the various alternative protein sources and the domains of health (safety, allergenicity and digestibility), environment (GHGs and other environmental and climate impacts, biodiversity, circularity), and/or barriers to adoption (technological, sensory, and consumer acceptance). The GIANT LEAPS consortium consists of the key actors and spans all expertise to address relevant knowledge gaps and proactively engages to arrive at optimized future diets based on alternative proteins that are broadly accepted across stakeholder groups. In order to deliver required insights for short-, mid- and long-term decision making and impact, GIANT LEAPS protein sources have been selected for either targeted or full assessment based on their current level of specification. The innovations and improved methods combined with accessible and comprehensive information, generated for a wide collection of alternative proteins, will enable policymakers to prioritise changes in the food system towards the dietary shift based on desired impact, value chain actors to make strategic scientific, business and investment choices, and the general public to make more sustainable and healthy dietary choices.

ID: 108979

Start date: 01-09-2022

End date: 31-08-2026



Last modified: 28-02-2023

Grant number / URL: <https://cordis.europa.eu/project/id/101059632>

Gap resolution in sAFety, NuTritional, aLlergenicity and Environmental assessments to promote Alternative Protein utilization and the dietary Shift

Data Summary

Will you re-use any existing data and what will you re-use it for?

Yes, the project will make use of secondary, already existing data. In addition, primary data collected in several WP's will be re-used in other WP's. Primary data will be collected in WP1, 2, 3, and 4. Secondary data will be collected and used in WP5, 6, and 7 for the overall assessments on the impacts of the dietary shift on health and sustainability.

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

Data collection and integration is an important aspect of GIANT LEAPS, and includes both primary and secondary data. For instance, it covers existing data stored in datasets, and collection of primary data on e.g. nutritional aspects of novel proteins and human trial data.

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

The overall objective of the project is “to deliver the methodologies, datasets and innovations needed to accelerate the adoption of healthy and sustainable alternative proteins and actualise a dietary shift across Europe. The methodologies and datasets will be open (access) and comprehensive, providing stakeholders with a scientific basis to make future-proof and maximum-impact choices, from consumer-accepted policies to new technologies”. Data collection and making data available is thus an integrated part of the project. Regarding data collection the project aims to gather data to create open access, comprehensive datasets. The underlying reason is that although a wide array of alternative protein sources is already available and even more will become available in the future, data related to health and environmental impacts, and critical applicability parameters (technological, nutritional and sensory functionalities) of these protein sources are scattered and incomplete. Therefore, GIANT LEAPS will develop a future-proof, open cloud platform with protein screening AI algorithms to collect and update comprehensive datasets from project-generated and existing data, covering a broad set of alternative proteins.



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

The expected size of data varies from 2-32 terabytes as the size varies widely depending on the type and complexity of the data, as well as the storage format and compression techniques used. For a simple text, the document may be only a few kilobytes in size will be used. At the same time, a high-resolution image or video file can be several gigabytes or even terabytes; a dataset used for statistical analysis may be much smaller. Still, it may also include large amounts of metadata or other associated files that increase the overall size.

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

Primary data – data from project partners (e.g., research results/ outputs generated within Giant Leaps project) and stakeholders (private databases but shared for project use and/or further open access).

Secondary data – publicly available literature, from public funded projects, literature, proprietary databases.

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

Primarily, the data will be used by project partners.

Secondly, for stakeholders, as some data is expected to be provided by involved stakeholders.

Finally, after completing the project, scientists, research institutes, policymakers, food industry and consumers will be able to use the open aggregated data for scientific, industrial and public decision-making activities.

FAIR data

Making data findable, including provisions for metadata: Will data be identified by a persistent identifier?

The metadata will be collected and included in the database (and data platform) of WP6 to make the data searchable. It was not planned to assign DOI to the data as the data platform would define the level of access to the data and data searchability. At the end of the project, it is planned that the platform and database would be integrated with the other Open Data Initiatives and data will be publicly available (after approval of project partners) with potential assigning of DOI.

Making data findable, including provisions for 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.

The metadata includes information about the content, format, quality, and context of research, information about the study- such as its purpose, methods, results, and conclusions; title and abstract of research; authors and contributors; keywords, research design, results (findings, statistical analysis, trends, patterns), conclusions, funding sources, data availability, and access. In addition to these types of metadata, it will include the software or tools used, relevant citations, or details about the data processing and analysis and will be based on FAIR metadata standards.

Making data findable, including provisions for metadata: Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Yes. Keywords will be a part of data search and ontology construction.

Making data findable, including provisions for metadata: Will metadata be offered in such a way that it can be harvested and indexed?

Yes, metadata will be offered according to up to date standards such as the Open Archives Initiative Protocol for Metadata Harvesting to enable automated metadata harvesting and indexing.

Making data accessible - Repository: Will the data be deposited in a trusted repository?

As part of the project objective to develop (in WP6) a future-proof, open cloud platform with protein screening AI algorithms to collect and update comprehensive datasets from project-generated and existing data, the data will be accessible through the developed interface and could be uploaded as well to a research data repository at later stages of the project. The strategy to implement uploading to a repository and the definition of the technical requirements will be developed during the first 3 years of the project.

Making data accessible - Repository: Have you explored appropriate arrangements with the identified repository where your data will be deposited?

No arrangements were explored, as development of the data platform requires keeping the data on protected servers with different levels of access. It is planned to explore the integration possibilities in year 4 of the project. This DMP will be reviewed and/or updated regularly to reflect the developments.



Making data accessible - Repository: Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

At the end of the project, it is planned that the open cloud platform and database would be integrated with the other Open Data Initiatives and data will be publicly available (after approval of project partners) with potential assigning of a DOI. The choice for a repository for data storage will include the consideration for appropriate DOI assignment and handling in relation to the data platform.

Making data accessible - 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.

Data from open data sources (research publications, open data bases etc) will be remained openly by default.

Private data will be restricted and partially available after applying aggregation and anonymization procedures. Private data (e.g., from stakeholder companies and partner companies) could be made available provided permission of data owner.

All data generated in the project by research partners and provided for the integration in the platform database will be available after the confirmation from project partners.

Making data accessible - Data: 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.

The GIANT LEAPS Consortium Agreement states that prior notice of any planned publication shall be given to the consortium parties at least 30 calendar days before the submission for publication. This time is allocated to assess whether the protection of the objecting party's results or background would be adversely affected, or the objecting party's legitimate interests in relation to its results or background would be significantly harmed, or the proposed publication includes confidential information of the objecting party. A publication delay of maximally 90 days can be requested in case of an objection.

Making data accessible - Data: Will the data be accessible through a free and standardized access protocol?

The data will be accessible through the developed interface and could be uploaded as well to a research data repository at later stages of the project.



It is planned that the data platform would allow for data availability through the web-page interface.

No specific documentation is needed to access the data. Moreover, to support the users of the data, there will be a User Manual (D6.2 User manual for GIANT LEAPS cloud data platform, IRIS (M36))

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

During the project completion different levels of access will be defined according to the accessibility levels (user profiles) to assure the potential to work on data and open access to aggregated data with relevant permissions. At the end of the project all data (with permission from project partners) are publicly shared.

Making data accessible - Data: How will the identity of the person accessing the data be ascertained?

not yet defined

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

For the moment, there is no need for a data access committee since there is no personal/sensitive data in the database of WP6.

Making data accessible - 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?

The metadata will be collected and included in the database (and data platform) to make the data searchable. It was not planned to assign DOI to the data as data platform would define the level of access to the data and data searchability. At the end of the project, it is planned that the platform and database would be integrated with the other Open Data Initiatives and data will be publicly available (after approval of project partners) with potential assigning of DOI.

Making data accessible - Metadata: How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

No time use limitations will be provided to all open access data.

Newly generated data will initially be stored locally in the institutional repositories of the respective partner. Metadata of all original datafiles, compliant with the Horizon Europe guidelines, will be made publicly available. Then, newly generated data will be made openly accessible in trusted



public repositories, such as EOSC, whenever possible, while keeping robust IPR management (data linked to IP will be available after only IP protection mechanisms have been put in place), and with the metadata standard of that repository added to the dataset. Source code, and AI and statistical models will be made available in open-source version control systems, like GIT, to make the analysis reproducible (e.g., R or Python Notebooks), let others review it and share it publicly.

Making data accessible - Metadata: 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)?

It is planned that the data platform would allow for data availability through the web-page interface. No specific software would be needed.

The main application through web-based user interface is planned. At the end of the project data platform code will be provided open access and any GIT console might be useful (not necessary) to clone repository.

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?

It is one of the aims of WP6 to create ontology for data interoperability between researchers, institutions, organisations, countries, etc. (i.e., adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins).

Ontologies will be used to define the links between the different types of semantic knowledge which help in formulating the data search strategies, useful for data users with different needs. Further existing ontologies will be integrated to combine available key databases to the maximum possible extent

Links to EFSA FoodEx database like EU funded infrastructures and initiatives will be created to develop a single framework. Interoperability of datasets will be ensured by using controlled vocabularies, keywords or ontologies where possible and by using file formats that are as open and widely used as possible.

A framework and ontology which will be created, will be crossed linked with other digital Farm-to-Fork projects with further integration in EOSC and JRC databases.

Making data interoperable: 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?



Maximum effort will be taken to use the available standard vocabularies for all data types present to maintain interdisciplinary interoperability. However, when suitable standard vocabularies are not available, suitable controlled vocabularies will be created. Connections to already existing ontologies and public databases will be assured.

Making data interoperable: Will your data include qualified references² to other data (e.g. other data from your project, or datasets from previous research)?

Question not answered.

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.)?

Question not answered.

Increase data re-use: 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?

Question not answered.

Increase data re-use: Will the data produced in the project be useable by third parties, in particular after the end of the project?

Question not answered.

Increase data re-use: Will the provenance of the data be thoroughly documented using the appropriate standards?

Question not answered.

Increase data re-use: Describe all relevant data quality assurance processes.

The DQA would be implemented upon the construction of a beta version of the platform and would involve several steps, including data profiling, data validation, data reconciliation, and data

² 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/>)



monitoring. It is planned that the DQA process is repeatable and automated as much as possible, to ensure that data is consistently of high quality and to minimize the time and resources required to manually check and correct data.

Increase data re-use: Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

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.).

Question not answered.

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.

Question not answered.

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.) ?

The budget for WP6, mostly aimed to provide the access to data is approximately 750 000 Euro.

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)

These costs will be covered via Giant Leaps project from the HE grant

Who will be responsible for data management in your project?



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)?

Question not answered.

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)?

Question not answered.

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).

not applicable

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

yes, see the ethics Deliverables of WP10

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)?

Question not answered.



www.giantleaps.eu



www.twitter.com/giantleapseu



www.linkedin.com/company/giantleapseu

