Plan Overview

A Data Management Plan created using DMPonline

Title: Ieuan's Plan

Creator: Ieuan Roberts-Harry

Affiliation: Other

Funder: European Commission

Template: Horizon Europe Template

Project abstract:

This project will investigate the feasibility of using cellulose microfibrils to generate various emulsion based foods, and to understand their impact of product stabilisation and texture. As part of this, we want to understand CMF contribution to both bulk and interfacial stabilisation. We also want to understand the interactions between CMFs and other surface-active materials.

ID: 108902

Start date: 01-09-2022

End date: 31-08-2026

Last modified: 25-10-2022

Grant number / URL: https://pickfood.eu/

Copyright information:

Ieuan's Plan

Data Summary

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

It's possible that data will be re-used. This data would likely come from similar studies, completed at the same institution, on the topic of cellulose microfibrils. The data would be re-used to compare with data from new experiments.

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

Data generated will often be tabular data, generated for the characterisation of emulsions and dispersions. Will be rheological data of many types. Will also include images, taken using CLSM, and other imaging techniques. Currently difficult to list all the formats of data which may be generated. In terms of data processing, some images will processed using image processing softwares.

Naming conventions:

I foresee that I will need to locate data using the project name, and so project name will come first in file names. Then, the measurement technique used will come, and finally the date, in the format YYYYMMDD, to ensure that files are presented in chronological order, and so easier to find.

For images, a concise description of the sample being imaged will be included. For versioning, the format V000 will be used, e.g. (V002), for version 2 of a file.

The naming conventions will need to be expanded upon as new types of data is attained.

Folder structures:

Firstly, folders are the first level will be named after the project they are in, e.g. "CFM and Potato Protein investigation". The next sub-level of files will be named after method which has been used to collect the data, e.g. "Rheometer data", or "CLSM images". Default ordering will be used so that the data appears alphabetically/numerically. Meaning that the data will appear in order of the sample description.

Consistency and quality of data:

At the beginning of each experiment, a strategy that will be used is to make one sample, and to measure it, rather than making all samples at once. This will mean that any issues that may arise can be dealt with quickly, and will only affect one sample, and one set of measurements. This allows for changes to be made quickly, and work can be completed efficiently.

Measurement techniques will go through the various calibrations required, and repeat measuruements will be completed to ensure quality of data.

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

The purpose of re-using data could be to compare it with new data that has been generated. For example, a study which focuses on the use of CMFs in emulsions may be repeated, but this time, a formulation may also include proteins. The re-use of data could allow us to compare and contrast the data, allowing for valuable insights to be made.

Furthermore, re-use of data would allow for re-processing of data. While the re-used data may not be published, it may serve to inspire experimental work.

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

At this point it is difficult to say. With Unilever's sharepoint system, I don't foresee large volumes of data being an issue.

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

How data will be shared:
For the most part, data will be shared using a data repository(Zenodo, Dryad), which will be used when the project ends. However, some data will be business confidential, and so some requests will be handled directly. In addition to the data repository, data will be published in data journals.
When data will be made available:
Data will be made available at the end of the project, or when an article is published. Request which are dealt with directly will be treated on a case-by-case basis, and so difficult to tell if the data will be made available, and how long it will take.
Restrictions on who can access the data:
As some of the work I am completing with Unilever could be business confidential, it may be that some pieces of data cannot be made public. It may be the case that a limited embargo can be arranged.
How data may be re-used:
It is difficult to speculate at this stage, however the work will likely be relevent to researchers studying the use of plant particles for emulsion stabilisation, the shelf-life of formulated products, and researchers using similar measurement methods. Keywords will be used to help other researchers find the data.

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

The data would be useful for people completing work in a similar field (plant-based emulsion stabilisation), but also potentially useful for people using the same measurement and characterisation methods.

FAIR data

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

Yes, data will be mind as easy to find as possible, through the use of a persistent identifier. This will include keywords which are consistent with other works completed in a similar field.

2.1. 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 that will be created is; Project name, publishing date, author and co-author names, the institution(s) that the author(s) are working from, links to the data available, name and page number for the article within a journal if the paper has been printed.

2.1. 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 provided. One thing to be wary of here is that a balance is struck between keywords which are specific, so that the data is easily found, but also more generalised so that it can be found when broader searches are made.

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

Yes, data will be offered in the way described above.

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

Yes, data will be deposited in a repository as much as possible. Due to some info being business confidential, this may not always be possible. Data will be deposited using Zenodo and re3data, and the NTNU open data sytem.

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

Question not answered.

2.2. 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?

Yes, NTNU open data applies digital object identifiers to data sets.

2.2. 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.

Not all data will be made available. Some data will be business confidential, and thus cannot always be shared outwardly. Studies will be altered to use certain variations of a material so that publications can be made, and data can be made as available as possible.

2.2. 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.

It is difficult to speculate how long embargos may apply for at this stage, and so will be considered on a case-by-case basis. It is likely that embargos won't be required.

2.2. Making data accessible - Data:

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

Yes, data will be made accessible as described as much as is possible.

2.2. 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?

There should not be any long-term restrictions to the data and will all be made available. It will be possible to access data independently, without any requests etc needing to be made.

2.2. Making data accessible - Data:

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

Question not answered.

2.2. Making data accessible - Data:

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. However, no

personal/sensitive data will be generated and so at this stage, it seems unlikely that a committee will be required.

2.2. 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?

Yes, metadata will be made openly available and licensed.

2.2. 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?

It is difficult to speculate at this stage, however I see no reason currently as to why data shouldn't be available indefinitely. Should the data be made unavailable, then metadata should still be available.

2.2. 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)?

Documentation and reference to data processing software will be necessary.

2.3. 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?

Every effort will be made to publish data in a format which is the most accessible. it will be ensured that data is formatted in a way which is 'computer readable'. While best practices will be used, which BPs will be used is as yet undecided.

2.3. 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?

Yes, mappings to more commonly used ontologies will be used.

2.3. Making data interoperable:

Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?

[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/)

Yes, my data will include qualified references to other data, as explained above.

2.4. 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.)?

As yet undecided.

2.4. 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?

Yes, it will. It will also be licensed as mentioned.

2.4. Increase data re-use:

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

Yes, it will.

2.4. Increase data re-use:

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

Yes, however the appropriate standards have not yet been decided.

2.4. Increase data re-use:

Describe all relevant data quality assurance processes.

Question not answered.

2.4. 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.).

it is possible that models will be generated, though a plan for the management of models is unlikely to be necessary.

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.

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.) ?

At the current stage, I don't foresee there being any extra costs incurred for the factors mentioned above.

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)

Currently unapplicable.

Who will be responsible for data management in your project?

Primarily responsible: Ieuan Roberts-Harry
Advisor: Krassimir Velikov

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)?

Data will be backed up using Unilever's sharepoint system, as well as being saved on local computers. I (Ieuan RH) will be responsible for the regular backing of this data, and will take place periodically, and regularly during a project. While the work is not sensitive in the sense that it does not include personal data, nor is it politically sensitive, it could however include data which is business confidential. Therefore, we are reliant on the safety of Unilever IT systems, and ensure that such confidential data is not stored externally.

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).

I see no ethical nor legal reasons that could have an impact on data sharing.

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

No questionnaires dealing with personal data will be created.

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)?

Currently there are no plans to use another data management procedure.