Plan Overview

A Data Management Plan created using DMPonline.be

Title: EntoBOOST FWO-SBO

Creator: Nate Sibinga

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Project abstract:

In the past ten years, the production of insects up-cycling organic wastes into high-quality biomass has gone from fiction to reality in Europe. Several of the main players leading this industrial revolution and exporting their know-how to other parts of the world are headquartered close to Flanders: Innovafeed (Nesle, France), Agronutris (Rethel, France), Ynsect (Amiens, France), Protix (Bergen op Zoom, The Netherlands). To date, these four companies have attracted an estimated €1.1 billion of investment capital, and their production capacity is growing rapidly. Flanders is strategically positioned to harness this emerging source of regionally produced biomass, aligning with the objectives outlined in its "Protein Strategy 2030" aimed at enhancing the sustainability of its protein production system. To do this responsibly, a number of key societal concerns on insect well-being during industrial production and the safety and added value of the end-products have to be addressed. This research initiative has three primary objectives: i) Explore if and how stress in insects can be measured by monitoring insect metabolome and behavior to contribute to the development of science-based guidelines to score insect welfare, ii) Fill research gaps on safety of pre-consumer organic waste streams containing microplastics and/or spore-forming food pathogens to enable safe use of insect feed substrates that do not compete with food or feed production, and iii) Identify which bio-active insect fractions drive observed beneficial effects on gut health in farm animals and define how these effects can be better conserved during processing. This proposal will be carried out by a consortium of the leading Flemish research groups in the field of edible insects working together to resolve several urgent challenges for insect production and usage with direct relevance for multiple potential users.

ID: 211474

Start date: 01-10-2024

End date: 30-09-2028

Last modified: 17-01-2025

EntoBOOST FWO-SBO
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
Insect Samples	insect-derived samples (includes whole insects, insect tissues, insect homogenates, insect fractions, insect extracts including DNA, RNA, proteins, lipids, chitin, etc.) substrate and frass samples	N	P				Mostly frozen storage; for some samples it may be possible to store at ambient temperature after drying/extraction (<50 kg)
Food waste (swill) samples	microplastic samples microbiological samples food waste (swill) samples	N	P				Microplastic samples will be stored either frozen or at ambient temperature (<5 kg) Microbiological samples (e.g. isolated/cultured bacterial strains) will be stored frozen (1 small box in -80º C freezer) Swill samples will be stored frozen until analysis, and small subsamples (<50 kg total) may be retained for up to 5 years, but long-term storage of large samples is not feasible
Rearing Data	animal care logs for insects, observational data including: temperature pH gas emissions larval density, growth, size, mortality, larval weight, total weight behavioral observations	N	P/D	N, T	.csv .xls .docx .txt	<1 GB	paper notebooks
Analytical Data	gene expression gas emissions chemical data (qualitative and quantitative) (micro)biological data (qualitative and quantitative)	N	P/D	N, T	.csv .xls .docx .txt	<1 GB	paper notebooks
Experimental (Meta)Data	recorded information on experimental conditions, methods, reagents, protocols, etc.	N	P/D	T	.csv .xls .docx .txt	<1 GB	paper notebooks

Images and videos	hyperspectral, multispectral, thermographic, RGB HR image/video data microscopic/SEM image data biological image data (e.g. photographs of microbiological plates, protein gels, etc.)	N	D	I	.raw	Hyperspectral: ~65 GB Multispectral: ~1.5 TB Thermal: ~60 GB RGB HR: ~15 TB	
Insect Stress Response Model	model integrating data from observations, sensors, computer vision, and metabolomics to monitor and detect stress	N	D	M, S	ipynb SQL tflite	data processing scripts: ~1 GB CV model: ~1 GB	
Metabolomics data	WP1 metabolomics: LC/GC-LR and HRMS data	N	D	T / N / I	.csv or .xlsx	<1TB	paper notebooks
DNA/RNA sequence data	sequence data (e.g. metagenomic sequencing of microbiome samples in WP4)	N	D	T / N / Sequence	.fasta	<100 GB	
Experimental Results	interpreted data, including statistical analysis	N	D	N, T, M, I	.csv .xls .docx .txt .ppt .tiff	<1 GB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Existing data on the occurrence, distribution, analysis and chemical properties will be collected from previous publications which are openly accessible or can be obtained from licensed journals through the libraries of the partner universities.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

This project contains animal experiments, however (as of the beginning of this project) none of these experiments require ethical approval per KU Leuven ethics guidelines (https://research.kuleuven.be/en/integrity-ethics/ethics/committees/ku-leuven-flowchart-on-ethics-issues.pdf). Insects and Artemia are non-cephalopod invertebrates. The samples used to inoculate the in vitro microbiome models (faecal material or intestinal content of chickens, pigs, or fish) will be collected either non-invasively (e.g. fresh faeces collected from a barn floor) or post mortem. If the guidelines should change during the course of this project to implicate any of the planned experiments, we will seek ethical approval.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

As an SBO project, many aspects of this research are intended to have commercial valorization potential. The sensor platform and associated models generated in WP1 have valorization potential for tech transfer in the management of insect farms. In WP3, (successful) remediation strategies for spore-forming bacteria in swill also have potential to be adopted commercially. Finally, WP4 may identify value-added products derived from insects that could hold appeal for commercial valorization.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

As an FWO SBO project, there are multiple third party members of the advisory committee. These members are all subject to a written agreement governing material and data transfer. Wherever possible, we have used a standard template provided by LRD that minimally restricts our use and dissemination of data.
However, in consultation with LRD, this agreement has been modified in some cases to meet the needs of individual advisory committee members. In these cases, we might be limited in some ways in our ability to disseminate results related to materials/samples obtained from the company.

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

This project is structured around PhD students/researchers, who will generate most of the data. Each researcher will be responsible for organizing and documenting their data in a (shared) standard format utilized by the research group for insect production and processing (P1). This includes standard organization and backup of electronic materials, explanatory cover sheets for all excel files, and other relevant procedures. Within the first six months of the project, a training session will be conducted with all researchers working on the project to ensure that standards for metadata and data management are clear.
As many of the experiments in this project utilize insect products produced internally, a standardized metadata format describing insect rearing and processing conditions will be developed and shared among project partners to ensure traceability of all insect-derived samples.
Further metadata files will be assembled at the work package level as needed to ensure that the setup and procedures used to obtain the data are clear. Particularly for WP1, which is the most data-intensive WP in this project, an extensive metadata file will be assembled which describes in detail the setup and the algorithms used to obtain the data, the generated source code and a comprehensive description of the collected data. The algorithms themselves will not be described in this metadata file but will be available and documented on a GitLab repository, allowing for full replicability of all results. Processed data and final code will also be deposited on a preprint server, such as Zenodo, to allow reproduction of results.
Regarding metabolomics experiments (WP1), detailed lab journal will document daily experiments, including methods, file names, and experiment details. Instrument-specific logbooks will record all measurements. Data analysis outputs will be systematically named based on subtasks, and SOPs will outline experimental and analysis workflows. Publications will serve as the primary project deliverables, adhering to DataCite metadata standards (Creator, Publication Year, Title, Publisher, resourceType General, Identifier), with DOIs provided as permanent URLs.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- Yes

Standard attribution/citation metadata (author, title, date, keywords, institutional affiliation, funding source, etc.) will be attached to all disseminated data from this project. Authors will link published work to their unique ORCID identifiers wherever possible.
Where applicable, specific metadata standards of relevant databases will be adhered to - e.g. NCBI SRA submissions of sequence data.

3. Data storage & back-up during the research project

Where will the data be stored?

☐ Shared network drive (J-drive)

☐ Personal network drive (I-drive)

☒ **OneDrive (KU Leuven)**

☒ **Sharepoint online**

☐ Sharepoint on-premis

☐ Large Volume Storage

☐ Digital Vault

☒ **Other: Teams-site (KU Leuven), Lab-hosted server**

Physical samples will be stored in suitable laboratory space - e.g. freezers or refrigerators for temperature-sensitive samples, BSL2 laboratory for relevant microbiological samples.

Physical records (e.g. lab notebooks) will be stored securely on premises (either in the lab or in offices kept locked when empty) and will be converted to digital backups when full.

WP1 will generate a large amount of raw data (video) from continuous monitoring of insects by automated sensors, as well as metabolomic datasets. The Imaging data is too large to store on KUL standard storage. It will instead be stored on ssd drives on a synology disk station. The data is backed up with raid5. For that storage the infrastructure is already available in the LT group.

This raw data will then be processed to develop the insect stress model. Software Scripts, Models and Manuscripts will be stored in the standard KUL OneDrive storage. Software code is stored on the laptop of the researcher and pushed to the researcher's KU Leuven-hosted Gitlab repository.

The metabolomics raw data will be stored on the local hard drives of two workstations at the Toxicological Centre (University of Antwerp). Relevant subsets may be published as Supplementary Data in associated manuscripts, ensuring long-term preservation. Additionally, the master copy of the processed digital data will be securely maintained on the UA MST cloud.

Electronic data from WPs 1, 2, 3, and 4 will be stored on a combination of OneDrive, Sharepoint and a shared Teams site accessible either internally (OneDrive or Sharepoint) or by invited external partners (Teams site: Non-KUL members of the research consortium, including specific researchers at UAntwerp, UGent, Inagro, Thomas More, and University of Insubria).

How will the data be backed up?

☒ **Standard back-up provided by KU Leuven ICTS for my storage solution**

☐ Personal back-ups I make (specify)

☒ Other (specify) **(raid5)**

As described above, raw image data generated in WP1 will be stored on a synology disk station by the LT group (P2) and backed up using a raid5 configuration.

For the raw metabolomics data in WP1, the internal server at the Toxicological Centre utilizes an automatic backup tool (CobianSoft 11) to ensure regular data backups. Additionally, summarized and processed data will be backed up on the researcher's personal hardware

For WP2-4, standard back-up provided by KU Leuven is envisioned to be sufficient for protection of digital data stored at KU Leuven in this project. In addition, paper documents (e.g. lab notebooks) will be digitized and stored as scanned backups upon completion. Any potentially important or useful isolated bacterial strains will be stored as aliquoted stocks at -80 C.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.
If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- No

The raw image data generated in WP1 cannot be stored at once on the device controlling the experiments. It will be collected at suitable time intervals and stored on the LT synology disk station. The server needs to be upgraded to hold enough storage capacity. At the moment there are ~10TB of storage capacity available, but this will not be sufficient for an estimated raw data amount of 16.81TB in WP1. Options for ensuring adequate storage capacity are described below in the "expected costs for data storage" section.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Electronic data will be stored using a combination of Sharepoint and Teams-sites. Both of these methods are recommended as suitable for confidential data by the KU Leuven storage guide (https://icts.kuleuven.be/storagewijzer/en). Teams allows granting of individual access to users both inside and outside of KU Leuven, thus facilitating easy collaboration between the research consortium while maintaining security of data access.

Metabolomics experiment (WP1). The internal server (10TB) at the Toxicological Centre, where data will be stored and processed, is accessible exclusively to members of the research group. Access is secured by one or more passwords, ensuring that data remains protected from unauthorized access.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

Space and costs associated with storage of physical samplesand most of the digital data are envisioned within the normal operations of the labs involved.

For digital data generated in WP1:

If the LT group synology server should be upgraded to hold the capacity of storing 16.81 TB of data, there are variable costs per TB as well as variable raid-efficiency, depending on the number of drives. This calculation is based on 4*20TB storage drives from IronWolf. For transferring the data efficiently, an additional portable drive is needed.

- Cost of Physical storage (1TB) ~ 100€ (portable SSD-drive)
- Upgrading LT server ~21.875€/TB * 16.81TB / 0.75(raid 5 efficiency) = 490€

As an alternative the data can also be stored at KU Leuven ManGo service, for a price of 35€ per TB per year, or at the KU Leuven large data storage at a price of 95.14€ per TB and year. With a linear approximation of the data acquisition the following amounts would be accumulated over the 4 year ENTOBoost project:

- ManGO: 1,176.7€
- KUL large volume: 3,198€

No extra costs are expected for WP1 at UAntwerpen. In case that the storage capacity is not available, an external hard-drive of 2 TB will be acquired.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

☒ All data will be preserved for 10 years according to KU Leuven RDM policy

☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans

☒ Certain data cannot be kept for 10 years (explain)

UA policy on data management will be followed which entails a preservation term of 10 years. Data will be stored in the central UA storage facility. Relevant subsets will be published as Supplementary Data to the manuscripts originating from the work, thus ensuring their preservation in perpetuity.

Processed digital data will be preserved for at least 10 years in accordance with KU Leuven RDM policy.

Raw image data will be preserved on the LT group server. It will be used in future research and therefore retained for a minimum period of 5 years, but given the volume of data the storage space may be needed for future projects after that.

For certain biological samples, storage for 10 years is not feasible due to space limitations in freezers and the likelihood of slow degradation (even when frozen). Swill and insect samples (specifically whole insects, insect tissues, and insect fractions) fall into this category

Where will these data be archived (stored and curated for the long-term)?

At UA, data will be stored on the local hard drives of two workstations available at the research unit and on the university's central servers (with automatic back-up procedures) for at least 10 years, conforming to the UA policy. In addition, relevant subsets will be published as Supplementary Data to the manuscripts originating from the work, thus ensuring their preservation in perpetuity.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

Sharepoint and Teams sites are free within KU Leuven and can also be used by other members of the research consortium. Space and costs associated with storage of physical samples are envisioned within the normal operations of the labs involved.

For image data generated in WP1, there will be some costs but these can be covered under the project budget if needed. If the hardware of the LT group is upgraded as described above, there will be no ongoing costs for retaining the data for five years. If KU Leuven storage services like ManGo are used the expected costs sums to approx. 2940€ over 5 years.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository
- Yes, in a restricted access repository (after approval, institutional access only, …)

After the ENTOBoost Project follow-up research is planned on the raw image data generated in WP1. For that reason there is no sharing planned at the moment. However, the data will be shared upon request.
Data resulting from the entometabolomics in WP1 will primarily be shared through publications. Supplementary information will include more detailed data, provided in formats such as .XLSX and DOCX. Upon approval, additional detailed and raw data will be accessible in formats such as comma-separated values (.csv), compound exchange format (.cef), and plain text (.txt).

If access is restricted, please specify who will be able to access the data and under what conditions.

Access to data will generally not be restricted, with the caveat that members of the advisory committee may expect to see certain results before they are made publicly available. In such situations, access to data may be restricted temporarily. In addition, data that is the subject of ongoing follow-up research (e.g. raw image data may be restricted while this work is in progress).

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Intellectual Property Rights

As stipulated in advisory council agreements with some advisory council members, data generated using material or expertise provided by a company may be subject to intellectual property rights of the company in question which could restrict dissemination. Such material is envisioned as only a minor part of the project, however, and thus potentially restricts only a small amount of the total data that will be produced.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

Data will be published in peer reviewed journals. Where applicable, raw data (e.g. sequence data) will be uploaded to relevant repositories (e.g. NCBI SRA).

When will the data be made available?

Upon publication of research results

Which data usage licenses are you going to provide? If none, please explain why.

CC-BY 4.0

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

In most cases applicable to this project, such identifiers are provided automatically via the sharing platform (e.g. DOI for articles published in peer-reviewed journals, SRA references for sequence data submitted to NCBI, etc.).

What are the expected costs for data sharing? How will these costs be covered?

Publication fees associated with peer-reviewed academic journals are anticipated to be the primary cost associated with data sharing. These costs can be significant (usually more than 1000 euros, and commonly several thousand), especially for open access articles. These costs will be covered by the PI(s) of each publication.
Sharing of sequence data via NCBI is free.

If physical samples are requested by other research groups (and approved), the requesting group will be asked to cover costs associated with transport.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

To ensure consistent organization, management of documentation and metadata will be divided by work package as follows: WP1: Prof. Ben Aernouts (sensor data) and Dr. Giulia Poma (metabolomics data) WP2: Prof. Mik Van Der Borght WP3: Prof. Dries Vandeweyer WP4: Dr. Nate Sibinga

Who will manage data storage and backup during the research project?

Prof. Mik Van Der Borght

Who will manage data preservation and sharing?

Prof. Mik Van Der Borght

Who will update and implement this DMP?

The project coordinator (Mik Van Der Borght) bears the end responsibility of updating and implementing this DMP