Plan Overview

A Data Management Plan created using DMPonline.be

Title: HYBRID

Creator: Stijn Spaepen

Principal Investigator: n.n.

Affiliation: KU Leuven (KUL)

Funder: Vlaams Agentschap Innoveren & Ondernemen (VLAIO)

Template: VLAIO cSBO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n.

Project abstract:

Harnessing pYrolysis & Biotechnology to Recycle mIxed plastic waste to Dicarboxylic acids - HYBRID

In HYBRID, we will design an innovative chemical/biological upcycling process, able to convert difficult-to-recycle mixed plastic waste (MPW) from packaging to high-value building blocks for polymers & their additives. A catalytic pyrolysis will convert MPW to tailor-made alkene/alkane mixtures with desired C lengths. An engineered yeast will funnel this heterogeneous mixture to specific dicarboxylic acids (DCAs). To maximize efficiency, an advanced biphasic fermentation process will be developed.

The world is facing a growing plastic waste crisis, with a huge imbalance between yearly plastic production and recycling. Of the 57 Mtons production in Europe in 2021, only 10% originate from recycled plastics. Moreover, >40% of the collected plastic waste is incinerated for energy recovery. While MPW represents an abundant (roughly 40% of total plastic waste) and largely untapped feedstock for upcycling, its chemical diversity and complexity limit its current value. This is mainly because **no satisfying solution exists for the closed loop of mixed polyolefin (PO) streams** containing low density polyethylene (LDPE) foils and residual fractions which contain additives, multiple polymers and/or impurities. Yet, when combined with a matching downstream biological process, this purification hurdle can be solved as all components can be funnelled by engineered microbes into a single product class.

In HYBRID, we will develop a hybrid process consisting of pyrolysis, biological funnelling, fermentation technology and innovative product recovery and purification with maximal resource recycling, for converting MPW with more than 70% polyolefins (PO) in high-value building blocks for nylon-type and polyester polymers & polymer additives. Chemical recycling has emerged as an effective way to tackle previously unrecyclable waste. However, many chemical recycling approaches focus on selective depolymerization of single plastic streams and struggle with heterogeneity in monomer and bond types, or with the presence of heteroatoms and impurities. Microbiology has high potential to complement chemical recycling. Biorefinery pipelines consisting of an chemical treatment, followed by biological funnelling to high-value compounds have gained increasing interest but these are mainly restricted to the conversion of sugars from lignocellulose to biofuels and lignin to aromatics. However, microorganisms have an incredible capacity to catabolize a wide array of substrates. **In HYBRID, we will uniquely combine rational cracking of MPW with bioconversion of a heterogeneous stream to DCAs using a *Yarrowia lipolytica* platform.** While bioconversion of sugars/polyols to DCAs has been studied, the use of alkane/alkene mixtures is underexplored. We will focus on production of adipic acid (AA), a DCA with the highest industrial relevance but produced using fossil fuels.

In a first research line we will focus on the**pretreatment and catalytic pyrolysis of MPW to liquid olefins.**There will be a special focus on the impact of the impurities on the formed products, and whether further pretreatment will be required. Various product compositions are possible, however, catalytic cracking of PO generally leads to gaseous and liquid fractions, composed of paraffins, iso-paraffins, olefins, naphthenes & aromatics. The aim is to maximize the liquid outcome of the reaction containing alkanes/alkenes with a desired C length distribution between C6-C14. The gaseous phase of molecules < 5C can be used as fuel for energy during the process steps.

In the second research line, subsequent biological funnelling of the pyrolyzed fraction into building blocks for polymers is tackled. This includes the development of a novel biological process (using the yeast *Y. lipolytica*) to convert liquid olefins into polymer building blocks. To increase versatility of the pipeline and to de-risk the most critical elements, various parallel strain engineering pipelines will be explored. We will screen an in-house strain collection to select the most suitable chassis strain and construct multiple yeast variants using state-of-the-art synthetic biology to increase flexibility of the process and input material. In addition to strain development, advanced fermentation strategies and intensified product recovery techniques will be combined. These include the development of a biphasic fermentation process alleviating substrate and/or product toxicity and the integration of in situ product recovery (ISPR) while targeting maximal resource recycling. As such reactive extractants can be applied to boost product titers and productivity, yet require careful composition of the extractive solvent mixture to balance extraction and biocompatibility. In addition, efficient back extraction techniques will allow to obtain purified AA with maximal resource recycling.

ID: 209838

Start date: 31-03-2024

End date: 30-03-2028

Last modified: 03-10-2024

HYBRID
VLAIO 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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Phenotypic and genotypic data set on Y. lipolytica strains	WGS, plate screening, small scale fermentations, metabolite profiles	New	Digital	Experimental compiled data	.xlsx .txt .fastq	< 100 GB	
Engineered Y. lipolytica strains	Design and construction of optimized pathways	New	Digital Physical	Experimental compiled data	.fasta		100's of yeast and E. coli strains in microwell plates or cryo vials
Metabolite production of engineered yeast strains	Comprehensive metabolite analysis under relevant fermentation conditions	New	Digital	Experimental	.xlsx .txt	< 10 GB	
Pyrolysis oil	Pyrolysis oil production from different sources & conditions, and characterization (GC, CHNSO, TGA)	New	Digital	Experimental	.xlsx	< 1 GB	

Development of (biphasic) fermentation and product extraction	Process and analytical data and reports	New	Digital	Observational, experimental, compiled data	.xlsx, .csv, .txt, .pdf	< 100 GB	
Process and analytical data and reports	Process samples and materials	New	Physical	Experimental			per experiment 0.2 - 20 g
Risk assessment	Preliminary hazard and operability (HAZOP) study	New, existing (e.g. MSDS, ATEX directives)	Digital	Compiled data, experimental, observational	.xlsx, .csv, .txt, .pdf	< 100 GB	
Sustainability and valorisation	Economic and climate impact studies	New	Digital	Compiled data, simulation data	.xlsx, .pdf	< 10 GB	
Sustainability and valorisation	Technology roadmap	New	Digital	Compiled data, simulation data	.xlsx, .docx, .pptx, .pdf	< 100 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:

1) Literature study. Data will be compiled from existing data available through journal publishers websites (DOI or URL) and exposed as RIS, .txt or .pdf files to online (or desktop) library manager or reference manager (Google Scholar, EndNote, Mendeley or Zotero). Processed literature data wil be summarised in a review document (.docx, .xlsx, .pptx).
2) Physical samples (liquid and solid) available from preliminary experiments can be reused for analysis and testing.

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

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

This work might have potential for tech transfer and valorisation. Therefore, there will be restrictions for data disclosure as it may contain IP-sensitive information. All data will be subjected to their patentability prior to any publication. If applicable, patent applications will be filed. IP management will be conducted as described in the research collaboration agreement (RCA) between all (research) parties. Patent applications will be made for relevant innovative results/procedures/products and the costs will be shared according to ownership that shall be subject to negotiation an a pro rata basis. The IP can be filed by one partner on exclusive project results or by multiple partners to protect comment project results (joined IP). Following the RCA, the intention for patent filing should be reported to the consortium upfront. Furthermore, to avoid publication of IP-sensitive results, all partners will be responsible for transfer of anticipated publications.

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

For this Moonshot ESI project, a research collaboration agreement (RCA) is drawn up. Furthermore, material product samples will be transferred between partners. Common project results and data will be shared in (bilateral) technical meetings (.pdf). Future (joined) anticipated publications or patent applications will be discussed in consortium meetings. The agreements concerning IPR are, briefly, reported in the previous question and described in the RCA. In addition, dissemination of data and related experiments with material obtained from industrial partners, are subjected to the agreements that are made bilaterally between project partners and the industrial partners. These are defined in the RCA and additional documents (e.g., MTA or NDA) signed by both parties. For all base strains used in this project, Verstrepen lab will check, together with René Custers (Regulatory & Responsible Research Manager at VIB) whether we need to comply with the Nagoya protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization. In case we need to comply, we will meet certain diligence obligations under EU Regulation (EU) No 511/2014 (apply for 'Prior Informed Consent' document and negotiate 'Mutually Agreed Terms" with Competent National Authority of the provider country).

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.

- Yes

Intellectual property rights and ownership are defined in the RCA and, possibly, additional documents (e.g., MTA or NDA) signed by partners within the project and industrial partners.

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).

Documentation levels and types:

Project-Level Documentation: Comprehensive documentation outlining the overall project objectives, methodologies, and key milestones. (.docx, .ppt, .xlsx)

Experiment-Level Documentation: Detailed records for each experiment, including the purpose, methods, variables, and results embedded in documents of each experiment. (.docx, .ppt, .xlsx)

Collaborative Documentation:

Collaboration Platforms: Utilizing collaborative platforms (e.g., Google Workspace, Microsoft Teams) to facilitate real-time communication, shared documentation and overview of sample flows among team members.

Regular Meetings: Holding regular meetings to discuss and update project documentation collaboratively, ensuring alignment with project goals.

The project coordinator (Verstrepen lab) has established a shared folder in the Teams application. This folder contains an overview table that needs to be completed for each sample shared among partners to maintain clarity, along with a brief description. Specific numbers will be assigned to each sample shared, which will be used as references in future discussions.

At the Verstrepen lab (KU Leuven - VIB), data (digital files) generated in this project will be stored in a Dropbox Business Advanced account for processing and analyses; following secure data transfer, modern data encryption standards, and encrypted block storage (256-bit AES and SSL/TLS encryption). For more details see: https://www.dropbox.com/business/trust. Additionally, project data and sequencing data will be backed up to KU Leuven servers.

At CSCE (KU Leuven), digital data files (raw data and processed data) from analytical instruments will be backed up to KU Leuven servers. Moreover, all data will be stored in OneDrive Cloud managed by KU Leuven, which allows easy access.

Digital data files will be accompanied with a read me text file that contains relevant metadata for understanding and re-use of data.

All the relevant scripts driving the project will be stored on a secure Dropbox account. Scripts used for analysis will also be stored in Jupyter notebook (jupyter.org - an open source web application to store and share scripts), in GitHub or in the GitLab

service of KU Leuven.
At BBEPP, batch records with standardized outline and formatting will be prepared and completed with description of novel process technologies and optimized conditions, including a sample plan, analytical procedures and non-conformities during trials. Analytical data of obtained product streams from HPLC analysis will be stored as data files obtained directly from the Chemstation software (version C.01.05). All data and documents at BBEPP are stored on a Hyper-V server (RAID-10) within the domain controller (DC) Virtual Machine.

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

The metadata for the data in this project entails:

- Creator of the dataset
- Name of the dataset
- File type of the dataset (depending on the employed software different file formats will be generated).
- Date of generation
- Data type (experimental or modelled)
- Software employed to generate the data (in case of modelled)

Verstepen lab (KU Leuven - VIB): Since there is no formally acknowledged metadata standard specific to our discipline, Dublin Core Metadata will be used. Moreover, we will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards that are more specific to our data.

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

Where will the data be stored?

All data will be stored in a cloud service offered by the KU Leuven university or others used by partners as specified below:
J-drive is used at CSCE for raw data collection and transfer
Personal I-drives will be used by CSCE members working on the project to have a personal back-ups of their files OneDrive of KU Leuven is used as storage of all data related to project
BBEPP online data server
Verstrepen lab (KU Leuven VIB): data (digital files) generated in this project will be stored in a Dropbox Business Advanced account for processing and analyses; following secure data transfer, modern data encryption standards, and encrypted block storage (256-bit AES and SSL/TLS encryption. For more details see: https://www.dropbox.com/business/trust
Sequencing data will be stored on an internal lab server (present in host lab) as well as on a secure Dropbox Business account for processing and analyses.
All the relevant algorithms, scripts and software code driving the project will be stored on a secure Dropbox account. Scripts used for analysis will also be stored in Jupyter notebook (jupyter.org - an open source web application to store and share scripts), in github or in the GitLab service of KU Leuven.
Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database or the European Nucleotide Archive (nucleotide sequences from primers / new genes / new genomes), NCBI Gene Expression Omnibus (microarray data / RNA-seq data / CHIPseq data), the Protein Database (for protein sequences), the EBI European Genome-phenome Archive (EGA) for personally identifiable (epi)genome and transcriptome sequences.

How will the data be backed up?

At KU Leuven, the data will be stored on the university's central servers with automatic daily back-up procedures.
At BBEPP online data are stored on a Hyper-V server (RAID-10, capacity 3 TB) within the domain controller (DC) Virtual Machine.
Verstrepen lab (KU Leuven - VIB): data (digital files) are automatically backed up by the secure Dropbox Business Advanced account cloud backup services. Additionally, project data and sequencing data will be backed up to KU Leuven servers.

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.

- Yes

Each KU Leuven employee has 2 TB on storage facility.
At BBEPP online data are stored on a Hyper-V server (RAID-10, capacity 3 TB) within the domain controller (DC) Virtual Machine. A server capacity of 3 TB is available which provides sufficient storage and backup capacity during the project.
Verstrepen lab (KU Leuven - VIB): Dropbox Business offers unlimited storage and back-up capacity in their clouds.
There is sufficient storage and back-up capacity on all KU Leuven servers:
The “L-drive” is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end.
The “J-drive” is based on a cluster of NetApp FAS8040 controllers with an Ontap 9.1P9 operating system.

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

The KU Leuven central storage is secured and the access to the data can be strictly authorized to only researcher involved directly in project execution.
BBEPP: The data is stored on a Hyper-V server (RAID-10) within the domain controller (DC) Virtual Machine. This virtual machine is password protected and can only be managed by users that are member of the “Admin Function Group” within the BBEU-domain.
Verstrepen lab (KU Leuven - VIB): Access to data stored on the Dropbox Business Advanced cloud is granted based on role based access control and all access requires layers of authentication that includes strong passwords, SSH keys, 2 factor authentication, and one time passcodes. Dropbox safeguards data with document watermarking, granular content permissions and policies, document watermarking, and legal holds.
Both the “L-drive” and “J-drive” KU Leuven servers are accessible only by laboratory members, and are mirrored in the second ICTS datacenter for business continuity and disaster recovery so that a copy of the data can be recovered within an hour.

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

During the course of the project - there are no additional costs related to the data storage and backup. KU Leuven and BBEPP's data storage costs are both covered by each partner.

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...).

Question not answered.

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

At KU Leuven, the data will be stored on the university's central servers for at least 10 years, conform the KU Leuven RDM policy (e.g., K-drive of CSCE: K:\SET-CSCE-Archive-Data-D0771)
Biological data: yeast and bacterial strains will be stored locally in the laboratory (-80°C). Other biological and chemical samples: storage at 4°C and/or as frozen samples as appropriate.
At BBEPP, after the research project has ended, all data will be kept on the Hyper-V server (RAID-10, capacity 3 TB) within the domain controller (DC) Virtual Machine.

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

K-drive long-term storage costs are covered by project budgets. Running projects covers the costs of finished projects.
For BBEPP, all costs regarding data storage and backup are fully covered by BBEPP.

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 a restricted access repository (after approval, institutional access only, …)

Due to confidentiality, proprietary considerations, or other constraints, access to the data will be restricted and accessible to only KU Leuven employees and specific collaborators (VIB, BBEPP) to ensure that it is shared exclusively with those who participated in the project.
Data will be available only in a format of publications (PhD, IP, master thesis) or strictly for internal use within research group internally.

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

1. In a restricted access repository
2. Upon request by mail and approval by responsible PI (main data owner)
3. Via publications, patents

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 reported above, this work might have potential for tech transfer and valorization. Therefore, there will be restrictions for data disclosure as it may contain IP-sensitive information, applicable on the experimental, analytical and sample storage dataset. In the section 'article 6 Publication' of the collaboration agreement, the procedure of publication of results generated in this project covers the necessary requirements for publication.

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

The experimental and analytical data will be stored in a restricted access repository. This data and insights will be, later, made available via publications or patents, which are accessible via LIMO (KU Leuven) and search engines like patentscope. More detailed information and data can be shared upon request by mail (corresponding author).
For KU Leuven, data related to published manuscripts will be published in RDR (subject to restriction due to IP).
As a general rule, datasets will be made openly accessible via existing platforms that support FAIR data sharing (www.fairsharing.org). Sharing policies for specific research outputs are detailed below:
Biological data: Bacteria and yeast strains will be shared upon simple request following publication, unless we identify valuable IP. In this case, we will first protect commercial exploitation, either through patenting or via an MTA that restricts the material from commercial use. Plasmids will be made available via Addgene (non-profit plasmid repository), unless we identify valuable IP.
Datasets will be deposited in open access repositories.
Research documentation: All protocols used to generate published data will be described in the corresponding manuscript(s), and the related documentation will be included as supplementary information. These data and all other documents deposited in lab notebook are accessible to the PI and the research staff involved in the project, and will be made available upon request.
Manuscripts: We opt for open access publications where possible. Publications will be automatically listed in our institutional repository, Lirias 2.0, based on the authors name and ORCID ID.

Nucleic acid and protein sequences: Upon publication, all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database or the European Nucleotide Archive (nucleotide sequences from primers / new genes / new genomes), NCBI Gene Expression Omnibus (microarray data / RNA-seq data / CHIPseq data), the Protein Database (for protein sequences).

When will the data be made available?

Upon publication of the research results of the associated manuscripts and/or vetting by LRD or VIB.

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

For Verstrepen lab (KU Leuven - VIB): Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. Metadata will contain sufficient information to support data interpretation and reuse, and will be conform community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and License, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. For data shared directly by the PI, a material transfer agreement (and a non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.
For others: N/A.

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

Once the research results will be published, the DOI will be linked to the dataset including the project results. Other data stored (e.g., at KU Leuven archive), will not.

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

Publications in specific sources might be a subject of additional costs that will be paid form running projects.

6. Responsibilities

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

The postdocs and PhD students will collectively manage their own data and data documentation during the project based on mutual agreements about data sets. They will be assisted by the supervising professors/innovation managers and project manager.

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

Project manager will take responsibility to organize data collection (set-up One-drive), coordinate of data management activity (incl. manual back-ups) and assist people involved in the project in proper handling of collected data (documentation of data, sharing data etc.). Professors/innovation managers will take a helicopter view on data management (assign roles, specify access permissions etc.). Project manager of current project (Stijn Spaepen) will be always backed-up by other group members involved in project management (such as innovation manager Bert Lagrain, PI Bert Sels).

Who will manage data preservation and sharing?

Innovation manager Stijn Spaepen, main PI/professor (Kevin Verstrepen). Innovation manager Bert Lagrain (CSCE, KU Leuven).

For BBEPP: Evelien Uitterhaegen (R&D Coordinator), Koen Quataert (Innovation manager gas fermentation), Michiel De Middelaer (Automation Engineer BBEPP), Brecht Van der Beken (ICT support BBEPP).

Who will update and implement this DMP?

The end responsibility for updating and implementing the DMP is with the supervisor (promotor) and project manager.

HYBRID
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

HYBRID
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable