HyPro: Automatic hybrid digital twins for process modelling

A Data Management Plan created using DMPonline.be

Creator: Mumin Enis Leblebici

Affiliation: KU Leuven (KUL)

Funder: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Principal Investigator: Mumin Enis Leblebici, n.n. n.n.

Data Manager: Mumin Enis Leblebici, n.n. n.n.

Project Administrator: n.n. n.n.

Grant number / URL: C3/23/007

ID: 203408

Start date: 01-10-2023

End date: 30-09-2025

Project abstract:
Digital twins are precious tools for the chemical industry to in-silico test and optimizevtheir processes. These industries can benefit from both first-principle information and data-driven knowledge. Hybrid modelling showed excellent capabilities to merge these two sources of information. However, developing a hybrid model is often a tedious and time-consuming process requiring the designer to have both machine-learning and chemical modelling know-how.
In a recent research project of the applicant group, this modelling technique is performed by a facile and efficient patented method for the first time. The efficacy is then
demonstrated for various reaction and distillation processes for several industrial partner use cases.
The first aim of HyPro is to expand the capabilities of the said patented method to advance physical-chemical properties and chemical processes (e.g., solubility estimation). This expansion will vastly strengthen the portfolio of unit operations we can accurately model.
The second aim is to automate the design of hybrid models developed previously in the research group to be used in the fine chemical and pharmaceutical processes without advanced programming expertise.
The final aim of our project is to bring this into the industrial field in the form of a spinoff company or via licensing to various partners.

Last modified: 08-12-2023

HyPro: Automatic hybrid digital twins for process modelling

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 / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
Industrial	Data received from the industrial partners under the NDA	*R*	D	Numerical Digital	.csv .txt .xlsx	<100Gb	CIPT local Data center
Simulation	Data created during in-silico experiments	N	D	Numerical Digital	.csv .txt .xlsx	<1Tb	CIPT local Data center
Models	Data created during in-silico experiments	N	D	Digital	Python libraries	<10Tb	CIPT local Data center

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:

Industrial historical data

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- 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

The models will be evalutaed for IP valorization in a future spinoff.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or 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

The industrial data and the fitting models cannot be disclosed. The models are the IP of KU Leuven

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

The industrial data and the fitting models cannot be disclosed. The models are the IP of KU Leuven

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).

We are keeping all the data in the local 60TB data center in Diepenbeek. The models are all well documented and in the future will also be placed in the KUL private git Library.
The Industrial data will be kept at least for 5 years in the same data center.
All data will be kept in separate folders with text files and spreadsheets clearly explaining all required information for future reference.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

All data is confidential.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- OneDrive (KU Leuven)
- Digital Vault

How will the data be backed up?

- Personal back-ups I make (specify below)

CIPT local data center

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- No (explain solution below)

Purchase of a Synology 120TB data center going on

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

The data is encrypted.

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

There is sufficient budget to cover the 5-6k€ costs

Data Preservation after the end of the Research Project

Which data will be retained for 10 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

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

- Other (specify below)

Local private CIPT server

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

Explained in the previous budget question. The local data center costs close to 6k€.

Data Sharing and Reuse

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

- No (closed access)

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

Prof. Leblebici
Dr. Di Caprio (As long as he is employed)
Prof. Leen Braeken

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 per dataset or data type where appropriate.

- Yes, intellectual property rights

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- Other data repository (specify below)

CIPT local data server

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.

We will valorize the KUL IP and cannot disclose the industrial data.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- No

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

None.

Responsibilities

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

Prof. Leblebici

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

Prof. Leblebici

Who will manage data preservation and sharing?

Prof. Leblebici

Who will update and implement this DMP?

Prof. Leblebici
Dr. Di Caprio

Created using DMPonline.be. Last modified 08 December 2023

5 of 5