DMP title

Project Name My plan (FWO DMP) - DMP title
Grant Title 1S20822N
Principal Investigator / Researcher Marit Hendrickx
Project Data Contact marit.hendrickx@kuleuven.be
Description The scope of this research is to optimize irrigation management in horticulture. Available soil water and irrigation need will be simulated and predicted, combining real time soil moisture measurements, weather forecasts and a water balance bucket model calibrated with a model optimization technique, inverse modelling, considering their uncertainty. An existing algorithm (DREAM) will be implemented and evaluated, documenting and treating predictions errors and optimizing irrigation management.
Institution KU Leuven

1. General Information

Name applicant

Marit Hendrickx

FWO Project Number & Title

1S20822N - Using data assimilation and weather forecasts in irrigation scheduling

Affiliation

- KU Leuven
- Other

Co-promotors of this research project are (also) affiliated with the Soil Service of Belgium, and the Agrosphere Institute, IBG-3 (Forschungszentrum Jülich).

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Type of data	Format	Volume	Origin
Sensor data: time series (numerical)	xlsx, csv, txt	1-2 GB	Continuous (automatic) soil sensor measurements with different types of soil sensors; raw & processed (calibrated, aggregated)
Python scripts: model, data analysis & visualisation, data collection via API's	py	5 MB	Created python scripts
Python scripts: DREAM algorithm framework	py	2 MB	Reuse of DREAM algorithm; provided by Eric Laloy (licensed under GNU GPL3)
Matlab code: data analysis	m	1 MB	Created Matlab code
HYDRUS-1D simulations	out, IN, txt, DAT, h1d, csv	100 MB	Generated soil moisture simulations with HYDRUS-1D
Soil measurements and other field data (numerical & textual)	xlsx, csv	100 MB	Soil measurements: soil moisture samples, Kopecky samples; other field information including coordinates & soil texture
Climate data (numerical)	xlsx, csv	50 MB	Collected from KMI via Soil Service of Belgium and from (private) local weather stations
Weather forecasts (numerical)	grib	20 GB	ECMWF medium-range weather forecasts collected via python API (research licence)
Satellite data (numerical, images)	csv, xlsx png, jpeg	1-5 GB	Collected Sentinel-2 satellite data (NDVI, LAI, FCOVER) & screenshots; via Terrascope (https://viewer.terrascope.be/)
Photographic images (images)	png, jpeg	10-20 GB	Generated images: crop cover, experimental setup
Visual outputs (images)	png, jpeg gif	10-20 GB	Generated data visualisations, data analysis results, model outputs, image analyses
Existing soil datasets (numerical)	txt	10 MB	Virtual spatially variable soil datasets & Soil-Can lysimeter experiments; via Agrosphere Institute (Open Access data)

The total maximum estimated volume is around 70-80 GB.

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

N/A

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, add the reference to the formal approval by the relevant ethical review committee(s)

- No

N/A

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- Yes

The DREAM algorithm (python scripts) with application modifications can be redistributed under the terms of the GNU General Public License, either version 3 of the License, or any later version.
As this work is in collaboration with a VLAIO project, data/tech transfer and valorisation is in accordance with the project's cooperation agreement. A patent on exclusive research outcomes can be applied for after communicating to all partners of the VLAIO project. For common research outcomes, this is decided on jointly.
Possible research outcomes with potential valorisation include a method & module to combine uncertain sensor measurements and weather forecasts with a soil water balance model for soil moisture prediction, with the purpose of irrigation planning.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

As this work is in collaboration with a VLAIO project, data (of all partners within the project) can be used and published in consultation with the partners according to the project's cooperation agreement.

DREAM code:

A Python 3 implementation of the DREAMzs MCMC sampler (Vrugt et al., 2009, Laloy and Vrugt, 2012) based on the 2013 DREAMzs Matlab code (version

1.5, **licensed under GPL3**) written by Jasper Vrugt.
This program is free software: it can be redistributed and/or modified under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

- Code (Python scripts): Information for understanding the script and its functions will be provided in the script as docstrings. A README file is added for a group of (linked) scripts on how to use them (script chaining), needed inputs and script outputs.
- Images will be organised in logical folders, each with a README file, stating context, location, date, cropping.
- (Raw) data will be organised in logical folders, each with a README file, stating context, location, date, parameters, units, processing.
- Experiments: logbooks (including start and end date, location, people involved, setup, materials and methods, operations and measuring times (timeline), all measured data).

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

'README' style metadata (as stated above)

5. Data storage and backup during the FWO project

Where will the data be stored?

- All data is stored in a university's provided **personal OneDrive** environment (continuous back-ups, 2 TB)
- Copies of all data that concerns Python scripts, script inputs and outputs are saved in a personal (private) **GitHub** environment
- Copies of (finished) bulk data will be saved on the **university's personal server** (I:\ drive, 50 GB), readying for long term storage, eventually moving to the Archive (K:\) drive (1 TB)
- Since we will collaborate with other research units, we use a **shared OneDrive** for active use and sharing of the data during the project

How is backup of the data provided?

- OneDrive: automatic (continuous) back-up
- GitHub: manual back-ups on a regular basis (depending on activity, weekly to monthly) (back-up retention: version control)
- University's servers: automatic daily back-up procedures (back-up retention for 1 year)

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

- OneDrive: 2TB of free storage
- University's servers: 50 GB personal network storage, and 1 TB archive
- University's storage can be expanded, but storage exceedance is not expected. The estimated storage volume that is needed is 80 GB. If this volume is exceeded, we can still count on a large margin before exceeding 1 TB.

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

Storage exceedance is not expected, so extra storage cost is not expected.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data will be stored in the university's secure environment.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All data will be retained for a 5 year period at least after the end of the project.

Where will the data be archived (= stored for the longer term)?

- The data will be stored on OneDrive & the university's servers (K:\ Archive) (automatic back-up procedures).
- GitLab (KU Leuven service: gitlab.kuleuven.be) will be used as repository to store and share source codes (all people involved).

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

- OneDrive: 2 TB is included in the account; this can be extended for free up to 5 TB.
- The KU Leuven GitLab service is free.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

As this work is in collaboration with a VLAIO project, data sharing is in accordance with the project's cooperation agreement. Exclusive research outcomes can be shared after communicating to all partners of the VLAIO project. For common research outcomes, this is decided on jointly. Names and field coordinates of the participating farmers will be restricted (expect when written permission).

All other used (existing) data is Open Access, free to modify and publish.

Which data will be made available after the end of the project?

All non-confidential data will be made available. This includes all data except for names and field coordinates of the participating farmers (except for those with written permission) and potentially patented modules.

Where/how will the data be made available for reuse?

- In an Open Access repository

Non-patented code will be released on the KU Leuven GitLab repository (https://gitlab.kuleuven.be/).
All data will be released on an Open Access repository such as he KU Leuven community: Open Science Framework.

When will the data be made available?

- Immediately after the end of the project
- Upon publication of the research results

Who will be able to access the data and under what conditions?

The data will be Open Access under a CC-BY(-SA) license. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators.

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

No costs are expected.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Marit Hendrickx (the FWO fellow), will be responsible for data documentation & metadata.

Who will be responsible for data storage & back up during the project?

Marit Hendrickx (the FWO fellow), together with supervisors and ICTS, will be responsible for data storage & back up during the project.

Who will be responsible for ensuring data preservation and reuse ?

Marit Hendrickx (the FWO fellow), together with supervisors, will be responsible for ensuring data preservation and reuse.

Who bears the end responsibility for updating & implementing this DMP?

Marit Hendrickx (the FWO fellow) bears the end responsibility of updating & implementing this DMP, while the supervisor will be responsible in the long term.