DMP title

Project Name ELO-extensions - DMP title
Project Identifier 93381
Principal Investigator / Researcher Wim Van Den Noortgate, Dries Debeer
Project Data Contact Dries Debeer
Institution KU Leuven

1. General Information
Name applicant

Wim Van Den Noortgate

Han van der Maas (University of Amsterdam)

FWO Project Number & Title

G0D4122N. Psychometric innovations for digital learning applications: Enhancing ratings systems for computerized adaptive practice.

Affiliation

- KU Leuven
- Other

KU Leuven will be the principal host institution. The project will be done in collaboration with the University of Amsterdam (UvA).

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	How created
Generated data using Monte-Carlo simulation techniques. The simulation code will automatically summarize each simulated data set using aggregated statistics, that will be stored (rather than the raw data).	text-based format (.csv) + visual representations (i.e., images) of the results (.tiff)	about 5MB	The raw data will be generated in a simulation study (using the statistical software R or SAS). The raw data will be processed, resulting in aggregated results and visualizations of these results.
Existing data from an online learning environment managed by Oefenweb/Prowise (hereafter referred to as the Oefenweb/Prowise data)	text-based format (.csv)	about 50MB	Responses and time stamps from students using an online learning environment. A research agreement with Oefenweb/Prowise will be set up.
Software syntax. For the simulation studies, the source code will be stored as well as the random seed (for reproducibility purposes). All code used for the final analyses (of simulated and real data) will be stored as well.	text-based format (.R) (SAS)	about 1MB	Code written by researchers

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

The log data form Math Garden, an electronic learning environment managed by Oefenweb/Prowise will be ananonimized by employees from Oefenweb/Prowise before they are shared with the principal investigators and the research collaborator.

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

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

Both the simulated and real data will be used for research purposes, this is to evaluate and illustrate new techniques for tracking latent variables in online learning. Whereas conceptual developments made within this project can be implemented in commercial online learning

systems, the ideas are not inherently linked to the specific data used in hte project.

For the real data, a research aggreement will be set up with Oefenweb/Prowise, so that their data can be used during the project.

Sections of this aggreement relevant to IP restrictions are:

"All copyrights, intellectual property rights and other rights with respect to the Facilities that Oefenweb should make available to the Researcher will at all times remain vested in Oefenweb. All Facilities that Oefenweb should make available to the Researcher will be used exclusively by the Researcher for the Researcher's scientific research.

If the nature of the research and the collaboration give rise to this, the parties may make more detailed written arrangements about the intellectual property in respect of the research results.

If the Researcher and Oefenweb together develop exercises or other content that is suitable for use in digital teaching resources, the Researcher and Oefenweb will share the copyrights and where applicable any other intellectual property rights. In that case, the Researcher may only use the work for scientific purposes. Oefenweb is the only one allowed to exploit the work commercially, for example by including the work in a digital teaching resource."

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?

- Yes

A research aggreement will be set up with Oefenweb/Prowise.

Sections of this aggreement relevant to data restrictions are:

- "All Facilities that Oefenweb should make available to the Researcher will be used exclusively by the Researcher for the Researcher's scientific research."

- "Oefenweb endorses the importance of Open Science and wants to give researchers the opportunity to comply with the requirements set for the Open Access of Data of scientific publications. Because we also want to guarantee the privacy of users, one of the most important requirements in this respect is that no more data is shared than is strictly necessary, that the data cannot be traced back to individual persons in any way, and that no data may be published in storage systems outside Oefenweb. Oefenweb provides back-ups of the shared data, and the data will be shared again with researchers for replication research or other purposes."

4. Documentation and metadata
What documentation will be provided to enable reuse of the data collected/generated in this project?

For each simultation study a README file will be added in which the setup of the simulation study as well as the use of the different source-code files will be explained. In addition, the version numbers of all software will be included.

For all (final) real and simulated data analyses, R-scripts will be stored, that include clear references to the datasets used and clear notes that document each step.

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

The data from Oefenweb/Prowise will not be stored by the researchers, and will only be made available by Oefenweb/Prowise.

The raw data generated during simulation studies will not be stored. Only aggregated statistics (for each simulated dataset) and the source code (including the used random seeds) will be stored.

5. Data storage and backup during the FWO project
Where will the data be stored?

With respect to the source code in general, and the result obtained through the simulation studies, a time-stamped master copy of the data will be kept on our research unit central storage facility. Copies can be made by team members of this project and kept on personal devices.

The data from Oefenweb/Prowise will only be made available to the team of this project using a secure web-based cloud service. This data will not be stored or analyzed locally.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures.

This does not account for the data form Oefenweb/Prowise. Oefenweb/Prowise will provide back-ups of the data, and the data will be shared again with researchers for replication research or other purposes.

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

The amount of data that needs to be stored is limited. There is sufficient storage and backup capacity at the itec research group.

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

Given the amount of data that will be stored, the expected cost for data storage will be limited, and covered by the host research institution (i.e., KU Leuven campus KULAK).

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

Regarding the real data: only the Oefenweb/Prowise data holds real person data, and only anonimized data will be used during the project. In addition, the data will only by made available to this project trough a secured web-based cloud service.

Regarding the simulated data: data, codes and results will be safely stored on our research unit central storage facility, and will be pass-word protected such that they are accessible to the team members of the proposed project only.

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, ...).

According to the KU Leuven guidlines the data will be stored for 10 years after the project.

- The raw data generated during simulation studies will not be stored. Aggregated data that are stored during the simulation study will be kept for 10 years after the project. Yet, both the raw and aggregated data will be reproducible by means of the source-code and the used random seeds.

- The Oefenweb/Prowise data will never be used or stored locally by the researchers. It can only be accessed through a secured web-based cloud service. However, Oefenweb/Prowise will made the data available (via this service) upon request, and after a research aggreement is signed by the researchers.

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

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

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

Given the limited size of the data (< 1GB) the cost for data preservation is limited.

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)?

- Yes. Specify:

The Oefenweb/Prowise data will be made available upon request by Oefenweb/Prowise.

The raw simulated data will not be stored. However, it will be reproducible based on the source code and the stored random seeds.

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

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

- The source code and the random seeds necessary to generated the simulated data.
- The aggregated results (both in text as in visualizations) of the simulation studies, as well as the source code to aggregate the results and to make the visualizations.
- All source code used to analyze both real and generated data.

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

Source code to generate and analyze the data in the the simulation studies will be made available upon request by mail.

The source code used to analyze the read data will be included as an appendix / supplementary material to the relevant publications, and/or will be made available upon request by mail.

The data from Oefenweb/Prowise will not be made available by the project team, but can be made available by Oefenweb/Prowise.

When will the data be made available?

- Upon publication of the research results

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

There will generally be no restrictions on who can access the data.

Except the data from Oefenweb/Prowise, which will not be made available by the researcher, but by Oefenweb/Prowise. Other research will be required to agree with the research aggreement of Oefenweb/Prowise.

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

The cost for sharing the data will be limited.

8. Responsibilities
Who will be responsible for data documentation & metadata?

Wim Van Den Noortgate, Dries Debeer, Han van der Maas and the PhD student that will work on the project.

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

Oefenweb/Prowise is responsible for their data. The other data will be backed up and stored automatically by the KU Leuven facilities.

Who will be responsible for ensuring data preservation and reuse ?

Wim Van Den Noortgate, Han van der Maas and the PhD student that will work on the project.

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

Wim Van Den Noortgate