eTeacher: interactive web-based platform for programming education with automatic feedback and monitoring

A Data Management Plan created using DMPonline.be

Creator: Tom Schrijvers

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Grant number / URL: C3/22/032

ID: 199920

Start date: 01-03-2023

End date: 28-02-2025

Project abstract:

With growing industry-wide demand for IT skills, governments are finally introducing computer science education in schools. Yet, for more technical topics like programming and basic algorithms (for years 3–6 of Flemish secondary school), few teachers have sufficient knowledge, skills or experience to create educational material or teach these subjects. Traditional publishers fail to address the problem, providing print-based textbooks that are disconnected from the interactive nature of programming. With the web-based platform eTeacher, we serve ready-made interactive textbooks that seamlessly switch between explanations, executable examples and programming exercises. We relieve teachers: by providing thorough automatic feedback to students; by monitoring and signaling learning difficulties; and by proposing remedial actions. We will develop this, now a small pilot in use at 15+ schools, into a spin-off.

Last modified: 04-08-2023

eTeacher: interactive web-based platform for programming education with automatic feedback and monitoring

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
		Indicate: ***N****(ew data) or* ***E****(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
1	existing eTeacher platform code	R	D	SO	.py,.js.,.ts,.sql,...	<1GB	
2	new eTeacher platform code	N	D	SO	.py,.js.,.ts,.sql,...	<1GB	
3	existing eTeacher course texts	R	D	T	.tex, .yaml	<1GB	
4	new eTeacher course texts	N	D	T	.tex, .yaml	<1GB	
5	scientific publications	N	D	T	.pdf	<1GB	
6	generic code to support scientific publications	N	D	SO	.rs,...	<1GB	
7	eTeacher submissions	N	D	T	.py	4GB	
8	class observation	N	D	T	.txt, .yaml	<1GB	

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:

The data sets 1 and 3 have been created for the eTeacher project in the two years prior to the C3 project funding.

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).

- Yes (Provide PRET G-number or EC S-number below)

This concerns dataset 7:

G-2023-6932: application in preparation

For the sake of delivering the eTeacher platform service to schools, personal identification information has to be retained. This is governed by the contract with schools and the privacy declaration that LRD has put together.

However, for research purposes, all data is anonymised as personal identification is not relevant.

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 data 1 and 3 are already being valorized through KU Leuven LRD and the new data 2 and 4 will be incorporated in that valorization. The target customers are secondary schools who use the platform in their education.

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.

- No

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

The IP rights of 1 and 3 have been transferred to KU Leuven LRD prior to the C3 project, and LRD will be the owner of the new data as well.

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).

The data 1-8 are in human readable formats and self-documenting.

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

This is not relevant for data 1-8.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Other (specify below)

We store the data in KU Leuven gitlab (gitlab.kuleuven.be) projects in the DTAI / eTeacher group.

How will the data be backed up?

- Other (specify below)

KU Leuven gitlab takes care of backup.

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.

- Yes

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

KU Leuven gitlab provides access control. Only authorized people are given access to the eTeacher projects in gitlab.

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

The gitlab service is free of charge.

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)

In KU Leuven gitlab.

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

The gitlab service is free of charge.

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.

- Yes, as open data
- Other (specify below)
- No (closed access)

The data 1-4 will be used for commercialisation, and not be made available to third parties.
The scientific publications (5) will be made publicly available through the KU Leuven Lirias database. The generic research code (6) to support publications will be made available in a public git repository (a clone of the corresponding gitlab repository).
The datasets 7 and 8 will not bet made available.

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

Access to the data 1-4 will be managed by prof. Tom Schrijvers for KU Leuven LRD, and with the purpose of commercialisation. Datasets 7-8 will be used by Tom Schrijvers and his KU Leuven team.

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

The purpose of the project is to commercialise the eTeacher platform. For that reason all data relevant to the commercialisation will not be shared.

Where will the data be made available?

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

- Other (specify below)

Scientific publications will be made available in Lirias. Public code will be made available on github.com.

When will the data be made available?

- Upon publication of research results

The public data will be made available alongside with publication of the corresponding research results.

Which data usage licenses are you going to provide?

If none, please explain why.

- Other (specify below)

Commercial licenses for data 1-4 are arranged by KU Leuven LRD. Scientific publications (5) are governed by standard copyright. The public research code (6) will be made available under the European Union Public License.

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.

- Yes, a PID will be added upon deposit in a data repository

DOIs are usually assigned to publications (5).

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

Data sharing will not incur any costs.

Responsibilities

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

All involved project staff. Currently:

- Jesse Hoobergs
- Birthe van den Berg
- Tom Schrijvers

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

All involved project staff. Currently:

- Jesse Hoobergs
- Birthe van den Berg
- Tom Schrijvers

Who will manage data preservation and sharing?

During the project all involved project staff. Currently:

- Jesse Hoobergs
- Birthe van den Berg
- Tom Schrijvers

After the project:

- Tom Schrijvers

Who will update and implement this DMP?

Tom Schrijvers