DMP du projet "Improving education research and medical training through large-scale experimentation"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).

Plan Details

Plan title	DMP du projet "Improving education research and medical training through large-scale experimentation"
Version	First version
Fields of science and technology (from OECD classification)	Educational sciences, Computer and information sciences, Psychology
Language	eng
Creation date	2022-08-30
Last modification date	2022-09-08
License	Creative Commons Attribution 4.0 International

Project Details

Project title	Improving education research and medical training through large-scale experimentation
Acronym	SIDESLAB

Abstract

Although learning science has established general effects such as retrieval practice, spacing, and interleaving, which provide rough guidelines for improving learning, exactly how to fine-tune the learning experience in such a way as to optimise long-term memory retention in a real educational setting remains a vast open question. We will use the SIDES digital learning platform, which is used every day by 45000 French medical students to review their courses, train themselves and take their exams as multiple choice questions. Each semester, volunteer students will be randomised into one of many experimental conditions differing in parameters such as difficulty level, feedback, and the specific algorithms used to model memory retention and forgetting, in order to guide optimal scheduling of retrieval practice. The results will feed fundamental research on learning, and will be used to progressively improve SIDES and offer students a more efficient and evidence-based learning experience.

Funding

- French National Research Agency : ANR-21-CE28-0030

Start date 2022-03-01

End date 2026-02-28

Partners

- Laboratoire dInformatique de Grenoble (200711886U)
- GIP UNESS / Direction ()

Produits de recherche :

1. Moodle interfaces for experiments (Software)
2. Workflow of data between Ontosides, experiments, and outputs (Workflow)
3. Student experimental data (Dataset)
4. Student background and outcome data (Dataset)
5. Analysis scripts of student experimental and outcome data (Software)

Contributeurs

Nom	Affiliation	Rôles
Jouanot Fabrice		• Personne contact pour les données (Data workflow)
Mathes Grégory		• Personne contact pour les données (Experiments)
Ramus Franck		• Personne contact pour les données (External data, Analysis scripts, Experimental data) • Project coordinator
RAMUS Franck	Laboratoire de sciences cognitives et psycholinguistique	• DMP manager

Droits d'auteur :

DMP du projet "Improving education research and medical training through large-scale experimentation"

1. Data description and collection or re-use of existing data

Moodle interfaces for experiments

An experimental SIDESLAB module will be created, by duplicating the BNE (banque nationale d'entrainement), and adding experimental functionalities.
Each experiment will be programmed within this module, putting students in a modified BNE environment.

Each experiment will put students in a particular training regime, offering them to train on certain questions in a certain order or in certain conditions, collecting data on their responses, and potentially adapting the training regime to their responses.
Experiments will be programmed in Moodle, like the BNE.
Experiments will produce Experimental data (product #3).

Workflow of data between Ontosides, experiments, and outputs

A data workflow will be defined that describes data flow between Ontosides (data lake), the experimental modules, student responses, linking with external data, and output for analysis.

A workflow chart and specifications

Student experimental data

Experimental data consists of all the data collected in the experimental module during an experiment: student responses to questions, answers to questionnaires.
They will be linked with external data about the same students and outputted in an appropriate format for analysis.

Data sheets with each student action on a line, with proper identifiers for student, experiment, group, condition, timestamps, questions, answers, score.

Student background and outcome data

Data will be retrieved from the central data lake Ontosides through a user-friendly ontology-based query language (developed within Task 3) for enabling the cognitive scientists in charge of designing the experiments to specify the data needed for computing or learning their model, and then to extract them from OntoSIDES in the appropriate format.
Once an experiment has been launched on volunteer students, the training traces of the involved students will be injected in the OntoSIDES archive on a regular basis (for instance each night).
For allowing comparative experiments but also for reproducibility of experiments, it is very important to keep the *provenance* of the data in OntoSIDES as actionable metadata. Data provenance must enable to distinguish, possibly for each student, the training activities induced by a given experiment from the ones that are not related to the experiment.

Previous learning history of participating students, grades obtained before and after the experiment will be found in the OntoSIDES

archive in which all the data and metadata are in RDF format. RDF is a standard format recommended by W3C to describe any resource in a unified format based on triplets <subject-predicate-object> or <subject-property-value>.

Once an experiment has been launched on volunteer students, the training traces of the involved students will be injected in the OntoSIDES archive on a regular basis (for instance each night).

For allowing comparative experiments but also for reproducibility of experiments, it is very important to keep the *provenance* of the data in OntoSIDES as actionable metadata. Data provenance must enable to distinguish, possibly for each student, the training activities induced by a given experiment from the ones that are not related to the experiment.

Analysis scripts of student experimental and outcome data

Analysis scripts will be written to analyse the data produced in experiments.

Analysis scripts will be written mainly in R, and stored on personal computers.

2. Documentation and data quality

Moodle interfaces for experiments

Each experiment is described in a specifications document (cahier des charges).
Software produced will be documented.

Each experiment will be tested by fictitious student accounts in order to ensure conformity with the specifications, and revised until full conformity.

Workflow of data between Ontosides, experiments, and outputs

The specifications will be self-sufficient.

The workflow will be implemented and revised if necessary.

Student experimental data

Template for experimental data sheet, with documentation.

During pre-launch tests of an experiment, data sheets will be downloaded and their analysis will be simulated to ensure that all the necessary data is included and that the format is appropriate.

Student background and outcome data

All the data are centralized within OntoSIDES in RDF and organized according to a domain-specific e-learning ontology specialized

in Medical studies, which provides a common and structured vocabulary shared by all users (learners and teachers) involved in medical education.

Mappings are used to extract and transform data from other formats into RDF, and SPARQL-based extractors are used to transform targeted RDF data to csv files.

Ensuring that the extracted data are sound and complete for the cognitive scientists' needs is part of Task3. The soundness of the extraction query will be verified by the cognitive scientist through a query form making explicit the different basic components of the query that he/she will be able to modify by relaxing and constraining some filters on the corresponding properties. To help cognitive scientists check whether the extracted data is complete enough, we will provide the result of some statistical tests on the extracted data.

Analysis scripts of student experimental and outcome data

Scripts will be fully documented.

Scripts shared together with anonymised data will be downloaded from repositories and tested to check that they run correctly and produce the same results than in the original analyses.

3. Storage and backup during the research process

Moodle interfaces for experiments

Experiments are stored in OntoSIDES. This database is hosted on a dedicated physical server situated in datacenter of the univesity of Grenoble. This datacenter is compliant with the standards of physical and logical security. There is also a daily backup of the data.

The experimental server (Moodle server) runs on a virtual machine (VM Ware). This VM is hosted on a farm of physical servers spread over several datacenters in order to have a high level of redundancy and disponibility. This server is backed up every day

The experimental server, OntoSIDES server and also all servers of the UNESS environnement (used for exemple for the login) are physically protected in dedicated datacenters compliant with standards of physical security.

The network access to these servers is also controlled by firewalls. The network architecture also guarantees a high level of security and data integrity

Data access is only allowed to users with proper roles permissions.

Workflow of data between Ontosides, experiments, and outputs

not applicable

not applicable

Student experimental data

Data and Metadata are stored in OntoSIDES. This database is hosted on a dedicated physical server situated in datacenter of the univesity of Grenoble. This datacenter is compliant with the standards of physical and logical security. There is also a daily backup

of the data.

The experimental server (Moodle server) runs on a virtual machine (VM Ware). This VM is hosted on a farm of physical servers spread over several datacenters in order to have a high level of redundancy and disponibility. This server is backed up every day

The experimental server, OntoSIDES server and also all servers of the UNESS environnement (used for exemple for the login) are physically protected in dedicated datacenters compliant with standards of physical security.

The network access to these servers is also controlled by firewalls. The network architecture also guarantees a high level of security and data integrity

Data access is only allowed to users with proper roles permissions.

Student background and outcome data

Data and Metadata are stored in OntoSIDES. This database is hosted on a dedicated physical server situated in datacenter of the univesity of Grenoble. This datacenter is compliant with the standards of physical and logical security. There is also a daily backup of the data.

The experimental server (Moodle server) runs on a virtual machine (VM Ware). This VM is hosted on a farm of physical servers spread over several datacenters in order to have a high level of redundancy and disponibility. This server is backed up every day

The experimental server, OntoSIDES server and also all servers of the UNESS environnement (used for exemple for the login) are physically protected in dedicated datacenters compliant with standards of physical security.

The network access to these servers is also controlled by firewalls. The network architecture also guarantees a high level of security and data integrity

Data access is only allowed to users with proper roles permissions.

Analysis scripts of student experimental and outcome data

Scripts will be backed up as part of the general back-up procedure of personal computers.

Scripts are not sensitive and the analyses involve only anonymised data.

4. Legal and ethical requirements, code of conduct

All experiments will be subject to informed consent by research participants.

Experimental data in the experimental platform (Moodle SIDESLAB module) will be pseudonymised (student identifier in place of name).

Once experimental data is merged with external data and exported for analysis, it will be anonymised using adapted k-anonymization techniques. Reinforcing their robustness to additional knowledge that attackers may have or obtain easily by the information exposed by students about themselves through social networks is a problem that will be studied in Task3.

Moreover all data treatments are described in the UNESS PIA. Users are informed of these treatments in the CGU (Conditions Générales d'Utilisation) of the UNESS environment. The CGU are available in all pages of UNESS applications. UNESS has also a DPO who manages all these questions of data protection within the UNESS activities and projects

All developments made during the project are open source. We use only open source softwares and our productions will be under open source licences

Shared research data: CC license

The general project and the experimental protocol for the 1st experiment has been approved by Comité d'éthique de la recherche Paris Descartes on 15/10/2019.
Subsequent experiments will be submitted as extensions of the same project.

5. Data sharing and long-term preservation

Experimental data will be sharable after anonymisation, i.e. removal of student identifiers, timestamps, ranks

Sharing will typically take place only after a given experiment is finished and a first analysis has been completed, at the time of submission of the paper (for reviewing purposes).

After publication of the paper, access will be made entirely open.

One option used in psychology is Open science framework (https://osf.io/).

Another option will be to use the national federated repository for French research data under construction.

A third option could be the Perscido research data sharing platform (https://perscido.univ-grenoble-alpes.fr/) which is hosted by University Grenoble Alpes. We will compare the different options and will choose the one with the best balance between sovereignty and visibility.

The data are csv files that can be processed by many modules or scripts of open data analytics tools such as R.

DOI will be provided by the repository.

6. Data management responsibilities and resources

Franck Ramus, as project coordinator.

Metadata is the key for ensuring that data used and produced by the project will be FAIR.

To ensure the traceability of data resulting from experiments, we will rely on the metadata and named graph facilities offered by Semantic Web standards in order to associate provenance metadata to the learning traces produced within a given experiment. Each experiment will also be assigned metadata that will point to the underlying model and the experimental setting.

The data curation and enrichment with metadata has been planned within Task 3 of the project.