FWO DMP Template

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

1. General Information	
Name applicant	**Kristof De Witte**
FWO Project Number & Title	**G0D2722N** **The long run impact of the COVID-19 school closures**
Affiliation	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other:
2. Data description	
Will you generate/collect new data and/or make use of existing data?	☐ Generate new data ☒ Reuse existing data

| Describe the origin, type and format of the data (per dataset) and its (estimated) volume

If you **reuse** *existing data, specify the* **source** *of these data.*
Distinguish data **types** *(the kind of content) from data* **formats** *(the technical format).* | ***Dataset 1: Primary Education data***

***Source*: Provided by Katholiek Onderwijs Vlaanderen**
Description: Test results from standardised tests in the last year of primary education from catholic schools in Flanders, which are collected for quality assessment purposes. To allow for the analysis of trends, the data covers the period from 2014 – 2022.
    The analysis is based on standardised tests that are administered every year by the network of catholic schools in Flanders (Katholiek Onderwijs Vlaanderen) in grade 6, that is the last year of primary school. Catholic schools are publicly funded, but privately-run schools and constitute the majority of schools in Flanders. The data comprises data over a time span of six years from 2015.
    The test data is collected at individual level. In each year, the tests of the different subjects can be combined at individual level by the use of anonymous student identification numbers that teachers use when submitting the test results. The tested subjects were slightly adjusted over the six year period under consideration. In all six years, the subjects mathematics and language (Dutch) were tested. From 2016 onward, science and social sciences were introduced in the tests, first as a combined subject (world studies) and from 2018 onward as separate subjects. In 2019, French (second language) was added as an additional subject.

Type: Numerical
Estimated volume: 400MB
Format: Spreadsheets (Excel) that will be imported into Stata and transformed to .dta files

Dataset 2: Administrative data

Source: Provided by the Flemish Ministry of Education
Description: Administrative data with school characteristics. To allow for the analysis of trends, the data covers the period from 2014 – 2022.
    In addition to the test data, the analysis makes use of administrative data at the school level. The administrative data comprises general school characteristics, such as the number of students in the school and the share of girls in the school. The data also contains information if the school is a |

special needs school, that is a school for children with specific education needs for whom the common curriculum with reasonable adjustments in a regular school is not feasible, as well as the share of students with special needs in regular schools, that is the share of children with specific education needs for whom individual adjustments are made to the common curriculum. Regarding teachers, the data contains the number of teachers at the school in absolute terms as well as full time equivalents (FTE) by age group. We use this information to compute the share of teachers above the age of 50, which might matter in the context of COVID-19 as older persons might have a higher expected cost of infection, and hence, put more pressure on the school management to not reopen the schools.

Furthermore, the administrative data contains a rich set of measures of socioeconomic status (SES). These include the share of students coming from a disadvantaged neighbourhood, the share of students with a mother with a low level of education, the share of students who receive financial support from the government and the share of students who speak a different language than the language of instruction at home. This set of measures is used in combination by the government to allocate funds to schools.

As an additional measure of students with an immigration background, the data also comprises the share of newcomers, defined as the students who speak a different language than the language of instruction at home and only moved to Belgium in the last few years.

The administrative dataset also comprises the number of students, the share of girls and the SES indicators for grade 6. In addition, for grade 6, the data contains the share of grade repetition in grade 6 and the share of slow learners, that is students with a backlog who have repeated at least one grade in the past.

Type: Numerical
Estimated volume: 2MB
Format: Spreadsheets (Excel) that will be imported into Stata and transformed to .dta files

Dataset 3: Higher Education data

	Source: Provided by KU Leuven *Description:* Administrative data about study programmes and doctoral research with information about students and their promotors. To allow for the analysis of trends, the data covers the period from 2010 – 2022. From KU Leuven administration, we receive anonymous information at individual level. *Type*: Numerical *Estimated volume*: 500MB *Format*: Spreadsheets (Excel) that will be imported into Stata and transformed to .dta files

3. Ethical and legal issues

Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register.
 	
In case your host institution does not (yet) have a privacy register, a reference is not yet required of course; please add the reference once the privacy register is in place in your host institution.
	☒ Yes
	☐ No
	If yes:
	    - Privacy Registry Reference: **ongoing review: G-2022-5119 (ongoing)**
	    - Short description of the kind of personal data that will be used:
	
	The data contains sensitive personal data of students and teachers/professors, such as age, gender, nationality and test results. The identifying data will be used to match different datasets. Subsequently, all data will be anonymised to prevent identification of individual students, teachers/professors or schools. After matching the different datasets, the identifiers are no longer needed. The observations will be numbered and all personal identifiers (student/teacher identifiers, school names) will be deleted.
	
	As indicated in the written agreement with the providers of the data, all analyses will consider aggregated outcomes for schools or groups of students, but never consider individual students. This means that, in order to avoid that individuals or schools can be identified on the basis of their specific characteristics, the study will only report results about groups of students or schools, and no individual outcomes.
	
	In the data from primary education and higher education, individual students cannot be identified, since anonymous identifiers were used already in the provided data. Identification in that dataset is only possible at the school level (for primary education data) and course level (for higher education data), which is anonymised after matching as well.

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).	☒ Yes ☐ No If yes: Since the data contain sensitive personal data about students and teachers/professors (e.g., age, gender, nationality, test scores), approval from the ethical committee will be requested. - Reference to ethical committee approval: **G-2022-5119 (ongoing)**
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 ☒ No If yes, please comment:
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 ☐ No If yes, please comment: Following the written agreements with the providers of the data (Katholiek Onderwijs Vlaanderen, Flemish ministry of education and KU Leuven), the project supervisor committed to keeping the data strictly private and restrict the use to the analyses for the purpose of this research project. Accordingly, the data will kept private and only shared with members of the project team. After the described research has been conducted, the data will be destroyed.

4. Documentation and metadata

What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	A readme file will be written that describes each dataset as well as the scripts that are used to process the data.
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	☐ Yes ☒ No If yes, please specify:

5. Data storage & backup during the FWO project	
Where will the data be stored?	On password-protected KUL-managed computers and on the KU Leuven OneDrive for Business network drive, which will be protected by multifactor identification.
How will the data be backed up?	The OneDrive for Business provides automatic backups of the data in the cloud.
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 ☐ No If no, please specify: The standard storage offer of the OneDrive is deemed sufficient for the data of this project.
What are the expected costs for data storage and backup during the project? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of **the allocated project budget** to be used to cover the cost incurred.*	No additional costs for data storage and backup are expected, since the existing infrastructure of the OneDrive provided for staff is deemed sufficient.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	The computers on which the data will be stored are password-protected with multifactor authentification and are KUL-managed. Backups will be automatically done in the university's secure OneDrive for Business Cloud. Datasets that contains personal data will be destroyed once not needed anymore after matching all datasets to a final dataset for analysis (school names and student identifiers).

6. Data preservation after the end of the FWO project FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.

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, ...).	As agreed with the providers of the data, all data will be destroyed 10 years after the end of the project to guarantee that the data are kept private and only used for the purpose of this research. Scripts that document the analyses and readme files that describe the scripts will be stored after the end of the project.

Where will these data be archived (= stored for the long term)?	The scripts for the data analyses will be stored on the university servers with automatic back-up procedures for at least 10 years, conform the KU Leuven RDM policy. This will be managed by the supervisor of the project.
What are the expected costs for data preservation during these 5 years? How will the costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of* **the allocated project budget** *to be used to cover the cost incurred.*	No additional costs are expected for the data preservation, as the data must be destroyed 10 years after the project, conform with the privacy agreement with the providers of the data. The preservation of the scripts will be done in the existing framework of the university.

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 ☐ No If yes, please specify: Following the written agreements with the providers of the data (Katholiek Onderwijs Vlaanderen, Flemish ministry of education and KU Leuven), the project supervisor committed to keeping the data strictly private and restrict the use to the analyses for the purpose of this research project. Accordingly, the data cannot be shared with persons outside the project team.

Which data will be made available after the end of the project?	The data will not be made available publicly. Following the written agreements with the providers of the data (Katholiek Onderwijs Vlaanderen, Flemish ministry of education and KU Leuven), the project supervisor committed to keeping the data strictly private and restrict the use to the analyses for the purpose of this research project. Accordingly, the data will not be made available to other researchers outside the project team. After the described research has been conducted, the scripts with the data analyses will be made available upon request for research purposes. The datasets with the identifiable information of students will <u>never</u> be shared.
Where/how will the data be made available for reuse?	☐ In an Open Access repository ☐ In a restricted access repository ☐ Upon request by mail ☐ Other (specify):
When will the data be made available?	After the publication of the results of the project, the scripts from the analyses will be made available upon request and only for research purposes.
Who will be able to access the data and under what conditions?	Only the researchers from the project team and the supervisor of the project will be able to access the data. The datasets with the identifiable information of students will only be used until all datasets have been matched and destroyed afterwards.
What are the expected costs for data sharing? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of **the allocated project budget** to be used to cover the cost incurred.*	No costs are expected for the data sharing, since data sharing is limited by the 3rd party agreement.

8. Responsibilities	
Who will be responsible for the data documentation & metadata?	The PI bears the overall responsibility for updating & implementing this DMP.
Who will be responsible for data storage & back up during the project?	The PI bears the overall responsibility for updating & implementing this DMP.
Who will be responsible for ensuring data preservation and sharing?	The PI bears the overall responsibility for updating & implementing this DMP.
Who bears the end responsibility for updating & implementing this DMP? *Default response: The PI bears the overall responsibility for updating & implementing this DMP*	The PI bears the overall responsibility for updating & implementing this DMP.