FWO DMP Template - Flemish Standard Data Management Plan

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.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	**Willem De Cort - https://orcid.org/0000-0001-6106-4890**
Contributor name(s) (+ ORCID) & roles	**Kristof De Witte - https://orcid.org/0000-0003-0505-8642**
Project number[1] & title	11D0723N - Improving Educational Quality by Understanding and Nudging Pupils' and Teachers' Choices
Funder(s) GrantID[2]	11D0723N
Affiliation(s)	KU Leuven
Please provide a short project description	A quality education matters, especially to low socio-economic status (SES) pupils. But the quality of many vocational education programs, which attract a disproportionately large share of low SES pupils, is in doubt. Additionally, shortages in (quality) teachers are most prevalent in schools with more low SES pupils. Increasing the role of workplace learning in vocational education and improving the job attributes of the teaching profession are considered to be promising solutions to these challenges. Implementing these solutions in a cost-effective way requires understanding how pupils in vocational education choose a study program and how teachers(-in-training) choose a career path. This project's overall objective is to advance our understanding of how these two choices are made and how they can be nudged. Earlier research on the determinants of both choices uses either outdated stated-preference methods or observational data suffering from endogeneity and insufficient independent variation. We use discrete choice experiments to circumvent their methodological challenges and provide causal evidence on the importance of previously unexamined determinants. No earlier research has examined how to nudge pupils' study program choice or teachers-in-trainings' career choice. We develop randomized controlled trials to address this gap, and assess the treatments' long terms effects and its underlying causal mechanisms.

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

2. 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.

				Only for digital data	*Only for digital data*	*Only for digital data*	*Only for physical data*
Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
Survey data project 1	Data from self-designed survey amongst Flemish high school students in vocational education to measure preferences for workplace learning.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .xlsx ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Admin data project 1 & project 2	Administrative school-level data on students population (study program chosen, socio-economic	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB	

	status, etc.)			☐ NA	☐ .tab ☐ .gml ☒ other: .xlsx ☐ NA	☐ NA	
Survey data project 3 & 4	Data from self-designed survey amongst teacher-in-training school students to measure job preferences relevant to their career choice	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .xlsx ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Admin data project 3 & 4	Data on students' grades and other background characteristics, obtained from KU Leuven.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .xlsx ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

Guidance: *Data can be digital or physical (for example biobank, biological samples, …). Data type: Data are often grouped by type (observational, experimental etc.), format and/or collection/generation method.* *Examples of data types: observational (e.g. survey results, sensor readings, sensory observations); experimental (e.g. microscopy, spectroscopy, chromatograms, gene sequences); compiled/aggregated data[3] (e.g. text & data mining, derived variables, 3D modelling); simulation data (e.g. climate models); software, etc.* *Examples of data formats: tabular data (.por,. spss, structured text or mark-up file XML, .tab, .csv), textual data (.rtf, .xml, .txt), geospatial data (.dwg,. GML, ..), image data, audio data, video data, documentation & computational script.* *digital data volume: Please estimate the upper limit of the volume of the data per dataset or data type.* *physical volume: Please estimate the physical volume of the research materials (for example the number of relevant biological samples that need to be stored and preserved during the project and/or after).*	
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.	We obtained 'Admin data project 1 & project 2' from the Flemish Department of Education. We will obtain 'Admin data project 3 & project 4' from the KU Leuven .
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, please describe these issues further and refer to specific datasets or data types when appropriate.	☒ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☐ No If yes, please describe: This research project involves randomized controlled trials with human participants (in projects 2 and 4). However, the treatments are limited to information provision. Participants are always informed about the treatment after the experiment (as informing them before in certain cases threatens the scientific validity of the results). To ensure that we deal with all relevant ethical issues appropriately, we will ask for approval from SMEC for each project.

[3] These data are generated by combining multiple existing datasets.

Will you process personal data[4]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☒ Yes ☐ No If yes: - Short description of the kind of personal data that will be used: Survey data project 1: Emailadress is collected to allow us to distribute rewards for participation if students give their permission to do so. Survey data project 3 & 4: Emailadress is collected to allow us to distribute rewards for participation and/or recontacting if students give their permission to do so. Students' KU Leuven identification number ('r-number') is collected to allow us to link survey data with 'Admin data for project 3 & 4' if students give their permission. - Privacy Registry Reference: /
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 ☒ No If yes, please comment:

[4] See Glossary Flemish Standard Data Management Plan

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:
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 to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

3. 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).	**1)** I use README.txt files to describe the contents of the raw datasets, including how they were collected. **2)** I extensively structure and document my Stata do files both for data preparation and data analysis to ensure replicability.

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. *Repositories could ask to deliver metadata in a certain format, with specified ontologies and vocabularies, i.e. standard lists with unique identifiers.*	☐ Yes ☒ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: If no, please specify (where appropriate per dataset or data type) which metadata will be created: README.txt files per dataset. README.txt files per project to describe aim and content of different do files. Comments in each Stata do files to describe aim and content of different lines of code.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	On my physical computer and on OneDrive servers linked to my KU Leuven account.
How will the data be backed up? *What storage and backup procedures will be in place to prevent data loss? Describe the locations, storage media and procedures that will be used for storing and backing up digital and non-digital data during research.*[5] *Refer to institution-specific policies regarding backup procedures when appropriate.*	All data will be stored on OneDrive servers of KU Leuven as stated above. These servers have automatic back-ups according to KU Leuven policy.

[5] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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 yes, please specify concisely: KU Leuven staff has access to 2TB of storage space. I will not come close to this given the nature of my data. If no, please specify:
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *Clearly describe the measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe.* [7]	1) At our research group, we have a strict locked-door policy. 2) Both my laptop and KU Leuven account regularly change password according to KU Leuven policy. 3) My account requires two-factor authentification to log in to. 4) KU Leuven ensures the safety of these files from other attempts at unauthorized access.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	My files will not exceed the capacity that KU Leuven offers for free to KU Leuven staff. No additonal costs besides the standard KU Leuven overhead thus apply.

5. Data Preservation after the end of the Research Project

Which data will be retained for at least five 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.
Where will these data be archived (stored and curated for the long-term)?	It is KU Leuven policy to store relevant data of its staff on central servers for at least 10 years.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Given the very small size of these datasets, no additonal costs need to be expected.

6. 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. *Note that 'available' does not necessarily mean that the data set becomes openly available, conditions for access and use may apply. Availability in this question thus entails both open & restricted access. For more information: https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeurepo-AccessRights*	☒ Yes, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only, …) ☒ No (closed access) ☐ Other, please specify: Open access: survey data from project 1 and survey data from project 3 & 4; both after being anonymized. Closed access: admin data from project 1 & 2 and admin data from project 3 & 4
If access is restricted, please specify who will be able to access the data and under what conditions.	
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, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☐ No If yes, please specify: Both the Flemish Department of Education and the KU Leuven only offer access to this data for the specific use of these projects for privacy reasons.
Where will the data be made available? If already known, please provide a repository per dataset or data type.	Not known as of yet.

When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	The data that can be made available will be available upon publication of research results.
Which data usage licenses are you going to provide? If none, please explain why. *A data usage license indicates whether the data can be reused or not and under what conditions. If no licence is granted, the data are in a grey zone and cannot be legally reused. Do note that you may only release data under a licence chosen by yourself if it does not already fall under another licence that might prohibit that.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."* [6]	**None.**
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *Indicate whether you intend to add a persistent and unique identifier in order to identify and retrieve the data.*	☐ Yes ☒ No If yes:
What are the expected costs for data sharing? How will these costs be covered?	None. This will happen through the publisher of the research results (i.e. the journal).

[6] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Willem De Cort
Who will manage data storage and backup during the research project?	Kristof De Witte
Who will manage data preservation and sharing?	Kristof De Witte
Who will update and implement this DMP?	Kristof De Witte