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	**Aaron Putseys; 0000-0002-0394-4598**
Contributor name(s) (+ ORCID) & roles	
Project number[1] & title	FWO 11P8N24N; The Macroeconomic Effects of Superstar Firms: Why Policymakers Should Care
Funder(s) GrantID[2]	FWO 11P8N24N
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible: 05f950310

---

[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.

[3] Research Organization Registry Community. https://ror.org/

| Please provide a short project description | Over the past few decades, we have seen a rapid decline in the labor share and the rise of so called Superstar Firms. Superstar Firms are categorized as the most productive firms within an industry or country, and have typically above-average markups and below-average labor shares. Despite the ongoing discussion on its source, the empirical literature documented two facts related to the decline in labor share. First, many studies have documented the rise of intangible investment, especially concentrated among Superstar Firms. A second empirical fact is the decline in wage-price pass-through. These two empirical facts have major policy implications.

My research agenda aims to contribute to the development of a profound understanding of the role of Superstar Firms in these two macroeconomic trends. The first work package will investigate how the emergence of intangible investments – which tend to be mismeasured – influences measures of productivity growth at the firm and aggregate level. For this, a novel measure of intangible investment is constructed using a B2B firm-level transactions data set. The second work package will empirically estimate the pass-through of wages to prices at the firm level allowing for heterogeneity in market share. To avoid bias, a unique data set is constructed capturing collective bargained wage changes at the firm level. |

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[4].

Dataset Name	Description	New or Reused	Digital or Physical	*ONLY FOR DIGITAL DATA* Digital Data Type	*ONLY FOR DIGITAL DATA* Digital Data Format	*ONLY FOR DIGITAL DATA* Digital Data Volume (MB, GB, TB)	*ONLY FOR PHYSICAL DATA* Physical Volume
Firm-product level production data (PRODCOM)	It contains information on both values and quantities of sales, which enables us to construct domestic unit values at the firm-product level.	☐ 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: .dta ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	NA
Employer-employee data (RSZ)	The sample includes 20% of all employees in Belgium for the period 1996-2019. For each	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB	NA

[4] Add rows for each dataset you want to describe.

					☐ Software ☐ Other ☐ NA	☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .dta ☐ NA	☐ < 50 TB ☐ > 50 TB ☐ NA	
employee in the sample, we observe wages, working time and other characteristics.								
Company accounts data from the NBB Central Balance Sheet Office	These cover all incorporated firms in Belgium, which includes annual data on value added, sales, labor, intermediate inputs and capital.	☐ 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: .dta ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	NA
Wage Floor Data from FOD WASO	This dataset covers the wage floors (euros), the criteria of wage floor (tenure, function,…), the cause of wage change per joint committee over time.	☐ 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: .dta ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	NA

BTW-aangiftes (VAT declarations)	Data on sales and inputs for large and small firms are provided by the NBB, taken from the quarterly VAT declarations, which the NBB annualized and made consistent with the reporting period of the annual accounts.	☐ 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: .dta ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	NA
ECOOM R&D survey	The Research & Development Survey measures the research and development activities of companies in the past 2 years: spending, personnel, and funding of R&D activities.	☐ 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: .dta ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	NA
B2B Data	The B2B data reports all domestic	☐ Generate new data ☒ Reuse existing	☐ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/	☐ .por ☐ .xml ☐ .tab	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB	NA

transactions over €250 between Belgium firms, annually.	data		aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .dta ☐ NA	☐ < 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[5] (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).

---

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

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.	1) PRODCOM: https://statbel.fgov.be/en/themes/enterprises/industrial-production 2) Company accounts: https://www.nbb.be/en/publications-and-research/research-cooperation/mirdh/available-datasets 3) VAT Declarations: https://www.nbb.be/en/publications-and-research/research-cooperation/mirdh/available-datasets 4) Employer-employee data: https://www.ksz-bcss.fgov.be/nl/dwh/variables_table/variables_description?filter=name 5) B2B data: https://www.nbb.be/nl/articles/belgian-business-business-transactions-dataset-2002-2021 6) ECOOM R&D survey: https://www.ecoom.be/nodes/rd/en
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:
Will you process personal data[6]? 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: - Privacy Registry Reference:

[6] See Glossary Flemish Standard Data Management Plan

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:
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: All data is confidential (either to the FOD WASO, RSZ, NBB). Therefore, I am not allowed to share or publish this data.

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).	**All the files will be named and file names will be listed in the table with the description of the content.** **Each variable is defined in excel file. Units of measurement are also added.**

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: Variable name, source, definition, coverage (time & units).

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	**For research project 1: the data will be stored on the servers of the national bank of Belgium and can only be accessed using a computer at the NBB.** **For research project 2: the data will be stored on the servers of VIVES and can only be accessed when given permission.**
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.[7]* *Refer to institution-specific policies regarding backup procedures when appropriate.*	☒ Standard back-up provided by KU Leuven ICTS and NBB for my storage solution ☒ Personal back-ups I make (specify)

---

[7] 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: There is enough space to store the data on OneDrive (KU Leuven) and the servers at the NBB to make back-up on my personal OneDrive. 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]*	**For research project 1: the data will be stored on the servers of the national bank of Belgium and can only be accessed using a computer at the NBB.** **For research project 2: the data will be stored on the servers of VIVES and can only be accessed when given permission.**
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	NA

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...).	The KU Leuven requires its researchers to store their data for at least 10 years. Typically, the supervisor of a PhD student is responsible for data preservation. Therefore, Jozef Konings (supervisor) & the National Bank of Belgium will be designated as the responsible person. Furthermore, the data files will not be too large in terms of memory and can be stored on the faculty's servers during and after my research.
Where will these data be archived (stored and curated for the long-term)?	Project 1: ☒ Standard back-up provided by the National Bank of Belgium Project 2: ☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☒ Personal back-ups I make ☒ Data also will be backed up on my personal OneDrive storage, password protected and anonymized
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	**NA**

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:
If access is restricted, please specify who will be able to access the data and under what conditions.	Data will be available upon approval of the National Bank, RSZ, FOD WASO,…
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: The national bank and other institutions restricted access due to the confidentiality of the data sources. Only results are allowed to be shared.
Where will the data be made available? If already known, please provide a repository per dataset or data type.	**NA**

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'.*	**NA**
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." [8]*	**NA**
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?	**NA**

---

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

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**For research project 1: the national bank of Belgium** **For research project 2: The researcher (Aaron Putseys) & Jozef Konings**
Who will manage data storage and backup during the research project?	**For research project 1: the national bank of Belgium** **For research project 2: The researcher (Aaron Putseys) & Jozef Konings**
Who will manage data preservation and sharing?	**For research project 1: the national bank of Belgium** **For research project 2: The researcher (Aaron Putseys) & Jozef Konings**
Who will update and implement this DMP?	**For research project 1: The researcher (Aaron Putseys)** **For research project 2: The researcher (Aaron Putseys)**