FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

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	Alberto Pavia Soto 0000-0003-0312-3092 (fellow)
Contributor name(s) (+ ORCID) & roles	Christian Pröbsting 0000-0003-4885-9327 (supervisor)
Project number [1] & title	(1132725N) Regional Business Cycles and their Impact on Aggregate Fluctuations
Funder(s) GrantID [2]	D-2023-2150
Affiliation(s)	☑ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	Economic fluctuations are a recurring problem for economists and policymakers. Still, our understanding of the sources driving these economic fluctuations remains incomplete. Traditionally, business cycles are studied at the national level. But recently, the literature has pointed out strong regional disparities over the business cycle, putting forward the hypothesis that a large share of the national economic fluctuations we observe at the national level might originate at the regional level and propagate through the network structure of the economy. My research studies this hypothesis by exploiting novel granular data and state-of-the-art macro models. It proceeds in three steps: First, I construct a dataset of regional business cycle indicators and compare their statistical properties to those found at the national level. Second, I study how regions respond to a specific driver of business cycles: sudden changes in global demand for their products. Third, informed by my empirical estimates I set up a multi-region, currency union model to aggregate my regional estimates and quantify the share of aggregate fluctuations that is driven by regional fluctuations. The model is then used to derive policy implications for both the United States and Europe.

[1] "Project number" refers to the institutional project number. This question is optional. 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 [3].

				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
BEA Regional Economic Accounts	Dataset obtained from the US Bureau of Economic Analysis containing county-level employment, income, nominal GDP, and real GDP data, 1969-2020.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.csv .dta	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Economic Census	Dataset obtained from the US Census Bureau containing, regional	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model	.csv .dat .pdf .dta	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

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

	economic data on production, employment and investment, available every 5 years (years ending in -7 and -2)			☐ Software ☐ Other:			
Annual Survey of Manufactures	State-level dataset on manufacturing output, employment and consumption, available yearly 1963-2023.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.csv .dat .pdf .dta	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Construction spending	State-level dataset on construction spending, available yearly from 1992-2023.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.csv .dat .pdf .dta	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Building Permits Survey	County and state-level dataset on building permits by value, available	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model	.csv .dat .pdf .dta	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

	monthly from 1988m1-2025m1.			☐ Software ☐ Other:			
CBP imputed by Fabian Eckert et al.	Dataset obtained from the County Business Patterns of the US Census Bureau with imputed missing values by Eckert et al. (2020) containing county-industry employment data, 1975-2018.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.csv .dta	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
IRS migration data	Dataset obtained from the US Internal Revenue Service containing county-level in- and out-migration dataflows, 1991-2020.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.csv .xls .dta	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
BLS Quarterly Census of Employment and Wages	Dataset obtained from the Quarterly Census of	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical	.csv .dta	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB	

GUIDANCE: *The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.* *RDM Guidance on data*	
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.	BEA Regional Economic Accounts: https://apps.bea.gov/regional/downloadzip.cfm Economic Census: https://www.census.gov/programs-surveys/economic-census.html Annual Survey of Manufactures: https://www.census.gov/programs-surveys/asm.html Construction Spending: https://www.census.gov/construction/c30/c30index.html Building Permits Survey: https://www.census.gov/construction/bps/index.html CBP imputed by Fabian Eckert et al.: http://fpeckert.me/cbp/ IRS migration data: https://www.irs.gov/statistics/soi-tax-stats-migration-data BLS Quarterly Census of Employment and Wages (QCEW): https://www.bls.gov/cew/downloadable-data-files.htm U.S. Census Bureau international trade data: https://www.census.gov/data/developers/data-sets/international-trade.html, https://www.census.gov/foreign-trade/reference/products/catalog/usatradeonline.html, and API https://api.census.gov/data/timeseries/intltrade/exports/statenaics?get=STATE,CTY_CODE,NAICS,ALL_VAL_YR&COMM_LVL=NA4 UN Comtrade: http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=37
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.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information:

Will you process personal data[4]? 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) ☒ No Additional information:
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:

3. Documentation and Metadata	
Clearly describe what approach will be followed to capture the accompanying information necessary to keep **data understandable and**	To ensure that the data remains understandable and usable both for myself and for external users, I will enact the following procedures.

[4] See Glossary Flemish Standard Data Management Plan

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).
 	
RDM guidance on documentation and metadata.
	I document each dataset with the following documentation:
	• **README.md files:** each dataset (both raw and processed) has a README file that describes the goal and use of data generation, the work package to which the dataset belongs, and any instructions on how to obtain the code from the original sources and how to replicate the results, as well as using the data.
	• **Data dictionary/code book:** a comprehensive data dictionary (maintained in an .xlsx file) and a codebook (maintained in a .md file) detail the name of all files where the data is stored; the name of the variables, their definitions, and types (binary, integer, string, etc.); and unique identifiers or unique data id keys that are common across datasets.
	• **In-code documentation:** all code used for data analysis and transformation will be documented in-file with detail, adding key comments at each particular step. A header section will outline the goal of the code through a short description, the input data sources, key steps/transformations conducted in the code, and all output datasets/files produced by the code.
	
	I will use the following folder structure:
	• Archive: obsolete data/code which merits storage.
	• Data: contains the data dictionary and an individual folder for each dataset. In each dataset folder, the structure is:
	  ○ Code: the code files used for data transformation, numbered with the order of execution and following standard naming conventions.
	  ○ Docs: relevant documentation/methodology/communication with the statistical agencies.
	  ○ Input: raw datasets.
	  ○ Output: processed datasets.
	  ○ Tmp: temporary data files generated by the code.
	  ○ A README file and a main executable file for the code.
	• Tasks: contains folders (with order of execution) indicating the transformations/analysis made in the project. In each task folder, the structure will be:
	  ○ Code: the code files used for data transformation, numbered with the order of execution and following standard naming conventions.
	  ○ Docs: relevant documentation/methodology.
	  ○ Output: output generated by the task (figures, tables, derived datasets).

	○ Tmp: temporary data files generated by the code. ○ Results: a LaTex document indicating the results/analysis conducted in the particular task. ○ A main executable file for the code.
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: **Metadata standards:** Yes, I will use a **DataCite Metadata Schema** for dataset-level metadata to ensure that the datasets are findable and executable. I will maintain all data dictionaries and codebooks regularly updated, clearly indicating the data keys (= variables defining dimensions of dataset/unique identifiers), in accordance to FAIR principles. **Controlled vocabularies and ontologies:** I will use standard ontologies to (macro) economics, as recommended by the FAIR principles (main reference: Gentzkow, M., & Shapiro, J. M. (2014). Code and data for the social sciences: A practitioner's guide. Chicago, IL: University of Chicago) If no, please specify (where appropriate per dataset or data type) which metadata will be created:

4. Data Storage & Back-up during the Research Project

Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☐ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☒ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☐ Large Volume Storage ☒ Digital Vault ☐ Other:
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☒ Personal back-ups I make (specify) ☐ Other (specify) I will automatically back up the data on a regular schedule using the institutional backup system, ensuring that the latest version is always securely stored. Code files documenting the transformation of the raw data to processed data will be version-controlled using GitHub, stored online, and made publicly available together with the published papers. This also ensures that changes can be reverted if necessary, preventing data loss due to accidental modifications.
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:

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.* *Guidance on security for research data*	My data are not private/confidential and as such do not require special security measures. However, I will follow the following standards of network and computer security: • Data transfers occur only using secure protocols (HTTPS, SFTP). • I only access the data from my KU Leuven device, using my network firewalls and intrusion detection systems. • Access to data is controlled via two-factor authentication, and only I and my supervisor prof. Christian Proebsting have write or modification access. • My device follows regular software updates and antivirus protections.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The OneDrive license we will use is free for KU Leuven staff members.

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...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)

Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☒ KU Leuven RDR ☐ Large Volume Storage (longterm for large volumes) ☐ Shared network drive (J-drive) ☐ Other (specifiy):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	There are no costs expected for the data preservation.

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, as open data ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify: The new imputed datasets (regional consumption and investment) will be made available through my website (albertopavia.com), hosted in a ICPR archive. This will enable other researchers to enjoy positive externalities from my work. This will be made available under a **CC-BY 4.0 license**. Replication codes will be made available from my GitHub repository. The code will be offered under an **MIT license**.

If access is restricted, please specify who will be able to access the data and under what conditions.	NA
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:
Where will the data be made available? If already known, please provide a repository per dataset or data type.	☒ KU Leuven RDR ☐ Other data repository (specify) ☐ Other (specify)
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)

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.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☒ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☒ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
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, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☐ No
What are the expected costs for data sharing? How will these costs be covered?	There are no expected costs for the data sharing.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	PhD Fellow Alberto Pavia Soto is responsible for managing data documentation and metadata during the research project. My supervisor is prof. Christian Pröbsting is an authorized supervisor.
Who will manage data storage and backup during the research project?	PhD Fellow Alberto Pavia Soto is responsible for managing data storage and backup during the research project. My supervisor is prof. Christian Pröbsting is an authorized supervisor.
Who will manage data preservation and sharing?	PhD Fellow Alberto Pavia Soto is responsible for managing manage data preservation and sharing. My supervisor is prof. Christian Pröbsting is an authorized supervisor.

Who will update and implement this DMP?	PhD Fellow Alberto Pavia Soto is responsible for updating and implement this DMP. My supervisor is prof. Christian Pröbsting is an authorized supervisor.