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	Christian Pröbsting 0000-0003-4885-9327
Contributor name(s) (+ ORCID) & roles	Alberto Pavia Soto 0000-0003-0312-3092 (contributor)
Project number [1] & title	Why Greece is not Nevada - The Role of Risk-Sharing in Attenuating Business Cycles Within Currency Unions (3H220797)
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

[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.

Please provide a short project description	When several European countries decided to adopt a common currency more than 25 years ago, many economists warned that the currency union would face severe problem because euro area countries lacked alternative mechanisms to smooth out country-specific business cycles. In contrast, the United States have developed mechanisms to cope with asymmetric fluctuations within their national boundaries, such as a high level of labor mobility, income diversification, fiscal transfers and integrated credit markets. This project aims to answer to what extent these risk-sharing channels attenuate output and consumption volatility. The first part sets up an empirical framework to identify the response of risk-sharing channels and economic aggregates such as GDP and consumption to local demand shocks in U.S. counties and European regions. The second part then uses these identified responses to set up and calibrate a structural model that features all four risk-sharing channels. The model is then used as a laboratory to answer 'what if?' questions. For instance, by how much would business cycles in the euro area be less pronounced if member countries had the same risk-sharing mechanisms as observed across U.S. states? What would happen to U.S. business cycles if we shut down one risk-sharing channel at a time?

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	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	
BEA national	Dataset obtained from the US Bureau of Economic Analysis containing	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model	.csv .dta	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

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

	country-wide consumption data, 1929-2021.			☐ Software ☐ Other:			
BEA state	Dataset obtained from the US Bureau of Economic Analysis containing state-level employment (1969-2021), GDP (1963-2022), and consumption data (1997-2021)	☐ 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	
BLS LAUS	Dataset obtained from the Local Area Unemployment Statistics of the US Bureau of Labor statistics containing county-level labor force and employment data, 1976-2019.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.xlsx .dta	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

CBP	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	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	
QCEW	Dataset obtained from the Quarterly Census of Employment and Wages of	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model	.csv .dta	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

	the US Bureau of Labor statistics containing industry-level employment data, 1998-2021.			☐ Software ☐ Other:			
State-level trade	Dataset obtained from the US Census Bureau containing state-industry-destination-level import 2008-2021) and export (2002-2021) trade data.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.csv .xml .dta .png .eps	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
UN Comtrade	Dataset of the United Nations Comtrade program obtained from the BACI of the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII)	☐ 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	

	containing country-industry- import trade data, 1995-2020.						
Imputation	Dataset of county-industry-level export data obtained by imputing the "state-level trade" dataset using the county-industry-level "CBP" employment dataset, 2002-2018.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.dta	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Proxies	Dataset of county, commuting zone and state level consumption obtained from combining the "CBP" employment data with the "CBP state" and "CBP data"	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.dta	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

	consumption data, 1969-2020.						

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: https://apps.bea.gov/regional/downloadzip.cfm BEA national: https://apps.bea.gov/iTable/?reqid=19&step=2&isuri=1&categories=survey BEA state: https://apps.bea.gov/regional/downloadzip.cfm BLS LAUS: https://www.bls.gov/lau/data.htm CBP: http://fpeckert.me/cbp/ IRS: https://www.irs.gov/statistics/soi-tax-stats-migration-data QCEW: https://www.bls.gov/cew/downloadable-data-files.htm State-level trade: 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

[4] See Glossary Flemish Standard Data Management Plan

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).
 	
RDM guidance on documentation and metadata.
	Documentation and metadata will be added to all raw and processed datasets. A README file for each dataset will contain all the necessary information about the context of the data generation, the research project to which it belongs, and an interpretation of the data.
	
	A data dictionary/code book will be created as specified in the following section.
	
	The code used for data analysis and transformation will be documented in-file with detail, adding key comments at each particular step. Every code file will include a documentation section on top indicating:
	• Goal: a short description of the code.
	• Input: name, location, and format of all datasets/files needed to execute the code, clearly specifying their key identifiers.
	• Transformations: short bullet points detailing the critical steps/transformations conducted in the code.
	• Output: name, location, and format of all the datasets/files produced by the code.
	
	All the documentation (data, corresponding documents, methodologies) will be kept in the folder where the dataset is stored. The project 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 will be:
	  ○ Code: the code files used for data transformation, which follow standard naming conventions and indicate with numbers the order of execution.
	  ○ Docs: relevant documentation/methodology.
	  ○ 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.
	• Literature: relevant articles and bibliography related to the research.
	• Notes: meeting notes relating to the project, with participants and dates clearly stated.
	• 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, which follow standard naming conventions and indicate with numbers the order of execution. ○ 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. All good practices in terms of file and variable naming will be followed. Every variable will be stored together with a descriptive label. The variable type (string, float, integer,…) will be clearly indicated and kept consistent across datasets. GitHub will be used as a task organization and versioning system. All standard naming conventions for variables, controlled vocabularies, and ontologies will be followed (main reference: Gentzkow, M., & Shapiro, J. M. (2014). Code and data for the social sciences: A practitioner's guide. Chicago, IL: University of Chicago)
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: A data dictionary (.xslx) of the derived datasets used for analysis is to be kept and regularly updated. It contains all file names, variables, and descriptions. It clearly indicates the data keys (= variables defining dimensions of dataset/unique identifiers). 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) 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.
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*	The data will be stored using KU Leuven OneDrive. As such, it is subject to KU Leuven's network security structure. The GitHub accounts of the coauthors will be protected by multi-factor authentication.
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:
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?	The responsible for supervising the process of data documentation and metadata is prof. Christian Pröbsting. An authorized staff member is PhD student Alberto Pavia Soto.
Who will manage data storage and backup during the research project?	The responsible for supervising the process of data storage and backup is prof. Christian Pröbsting. An authorized staff member is PhD student Alberto Pavia Soto.
Who will manage data preservation and sharing?	The responsible for supervising the data preservation and sharing is prof. Christian Pröbsting. An authorized staff member is PhD student Alberto Pavia Soto.
Who will update and implement this DMP?	The responsible for supervising the updating and implementation of this DMP is prof. Christian Pröbsting. An authorized staff member is PhD student Alberto Pavia Soto.