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	Norah Kennis https://orcid.org/0000-0002-2813-8875
Contributor name(s) (+ ORCID) & roles	Steven Vanhaverbeke https://orcid.org/0000-0002-0436-8019
Project number [1] & title	11P6Z24N Unravelling European governmental venture capital funds: investment patterns, innovation and brain drain.
Funder(s) GrantID [2]	Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)
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	Policymakers around the world have intervened in venture capital markets to help alleviate the financial constraints start-ups face. For this purpose, governments designed governmental venture capital funds (GVCs). Despite the widespread use of GVCs in practice, research has yet to reach consensus on the treatment effect of GVCs on start-ups. In this research project, I want to clarify the influence of European GVCs on their start-ups. I start from the idea that GVCs operate on the basis of different incentives than Private VCs (PVCs). Contrary to PVCs, GVCs are established to create value beyond the financial objective and focus on impact investing (i.e. foster economic growth, innovation and regional development). The first research project explores differences in investment patterns between PVCs and various types of GVCs. In the second and third research project, I will study the influence of GVCs on their portfolio start-ups respectively regarding innovation (distinguishing between explorative and exploitative innovation) and brain drain (the emigration of high-skilled workers).

[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
		☐ Generate new data ☐ Reuse existing data	☐ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
1. Firm-year level data and venture capital data from **Orbis Global**	Database. Ownership structure data of firms. Venture capital data. Firmographics and financial data (as control variables).	Reused Secondary data	Digital	Compiled/aggregated data Numerical (mostly, sometimes strings)	Data provider is Moody's Analytics. Extracted through Amazon workspaces in .csv file format. Transformed and stored as .dta file format (STATA).	> 1TB	
2. Venture capital data from **VICO**	Database. VICO contains geographical, industry and accounting	Reused Secondary data	Digital	Compiled/aggregated data Numerical (mostly, sometimes strings)	.dta STATA .csv Dataset Access: RISIS https://docs.risis.io/dat	<1GB	

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

	information on start-ups that received venture capital financing. Venture Capital data.				asets/metadata/vico Dataset Owner: Politecnico di Milano http://www.dig.polimi.it Dataset Access Manager: Benedetta Montanaro benedetta.montanaro@polimi.it		
3. **EIC Accelerator data hub**	The European Innovation Council and Small and Medium-sized Enterprises Executive Agency (EISMEA) has developed an interactive online tool that generates information on EU funding programmes like the EIC accelerator. Venture Capital data.	Reused Secondary data	Digital	Compiled/aggregated data Numerical (mostly, sometimes strings)	Online datahub https://eismea.ec.europa.eu/eismea-datahubs_en https://sme-datahub.eismea.eu/		Remark: I did not yet downloaded the data. Will only be potentially used for additional / alternative analyses.
4. Patent data from **PATSTAT**	Database. Data on patent information, like patent citations.	Reused Secondary data	Digital	Compiled/aggregated data Numerical (mostly, sometimes strings)	.dta Access granted to PATSTAT online server through KUL. PATSTAT raw data: 2023 Spring Edition Version 5.21 saved on I-drive and	< 5TB	

					external hard drive.		
5. Patent data from **Orbis IP**	Database. Data on patent information, like patent citations.	Reused Secondary data	Digital	Compiled/aggregated data Numerical (mostly, sometimes strings)	.dta	<1GB	We use Orbis IP database to link patent identifiers (PATSTAT) to firm identifiers (Orbis Global). Can also be potentially used as alternative to PATSTAT for patent data. (But PATSTAT is preferred).
6. Migration patterns inventors (extracted from patent information).	Database consisting of patent and inventor data (for migration patterns). The database is originally constructed by Miguelez and Fink (2013) and further developed by Pellegrino et al. (2023).	Reused Secondary data	Digital	Compiled/aggregated data Numerical (mostly, sometimes strings)	.csv .dta Inventor migration files: publicly available on WIPO. https://www.wipo.int/publications/en/details.jsp?id=3952 publicly available on Harvard dataverse https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/AETFTF	<1GB	Remark: I did not yet saved / stored this data.
7. Employment based migration data. - U.S. Department of Labor - the Office of Foreign Labor	Data on issued labor certifications for permanent employment under several programs. This can be an alternative source for migration data.	Reused Secondary data	Digital	Compiled/aggregated data Numerical (mostly, sometimes strings)	.xlsx https://www.dol.gov/agencies/eta/foreign-labor/performance	<1GB	Remark: I did not yet saved / stored this data. Will only be potentially used as alternative to Harvard dataverse migration patterns inventors (which is

Certification (OFLC)							preferred).
8. STATA scripts	STATA scripts for statistical analysis	New	Digital	Models	.dta	<1GB	

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.	1. Orbis Global & 5. Orbis IP: Moody's Analytics 2. VICO: RISIS - Politecnico di Milano VICO | RISIS Docs 3. EIC: European Commission. European Innovation Council and SMEs Executive Agency EISMEA datahubs - European Commission (europa.eu) 4. PATSTAT: European Patent Office PATSTAT | Epo.org 6. Migration patterns: WIPO Knowledge Repository https://doi.org/10.34667/tind.28871 Measuring the International Mobility of Inventors: A New Database (wipo.int) or Harvard Dataverse doi:10.7910/DVN/AETFTF 7. U.S. Department of Labor: Performance Data | U.S. Department of Labor (dol.gov)
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	☐ Yes (provide PRET G-number or EC S-number below) ☐ No Additional information:

[4] See Glossary Flemish Standard Data Management Plan

Leuven privacy register number (G or S number).	For my third research project, I am considering incorporating personal secondary data concerning inventor migration patterns (via patent information). However, the approach to handling this data remains uncertain at this stage. I intend to ensure compliance with GDPR regulations (PRET / SMEC) at the outset of project 3, once the details become more clear.
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: We use secondary data. Depending on the database, 3rd party agreements restrict exploitation or dissemination: 1. Orbis Global & 5. Orbis IP: databases are accessed under license (KU Leuven library), any future user of the data is required to have similar access rights. 2. VICO: access granted (under conditions) through Politecnico di Milano (via e-mail). 3. EIC: open access (no 3rd party agreements). 4. PATSTAT: database is accessed under license (KU Leuven ECOOM), any future user of the data is required to have similar access rights. 6. Migration patterns: open access (no 3rd party agreements). 7. U.S. Department of Labor: open access (no 3rd party agreements).
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). *RDM guidance on documentation and metadata*.	I will use .docx and .txt files that describe how the data is extracted from data providers (secondary data). The files will contain log-in information, date of data retrieval, query information, version of data retrieved, variable definitions given by data provider, description of data…. This metadata documents how the dataset was created. I will use STATA Do Files as a tool / metadata that documents how the dataset was subsequently processed. I will primarily add comments in my STATA Do Files to explain the commands and procedures executed. These comments (non-executable text lines within the code) will provide explanations, descriptions, or annotations to aid understanding for myself or others who might read the code. Metadata related to variable names (and labels) and data properties is also shown in STATA. I will also include readme files to explain and document all the files within a certain folder (per project or per database).
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: For datasets obtained through a non-standardized procedure and for which no full documentation is available on the website of the data provider, a full description will be stored in a metafile, containing their coverage, variable names, units, formats, and total disk size. These will also be documented in detail either in the appendix of the corresponding working paper, or in an accompanying readme text file. However, in general the data structure and variables are relatively easy documented by the data provider (, for instance by Moody's Analytics).

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: DropBox
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) Standard back-up provided by FEB (Faculty of Economics and Business) and KU Leuven ICTS. Data storage, management and access are handled by the IT-services at the Faculty of Economics and Business as well as by the KU Leuven. I will use KUL OneDrive, personal network drive (I-drive), Dropbox, and an external hard disk.
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*	Security measures put in place by FEB KU Leuven that restrict access to unauthorised users guarantee secure data storage. In other words, the data infrastructure at FEB KU Leuven is secured and only accessible by department members.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	I have an account on Dropbox. Subscription includes offline access to files, stress-free sharing, offline folders on mobile, and 2 TB of storage. Costs €119.88 per year. Renews every year (in May). Costs covered by FWO bench fee.

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) As part of my PhD program requirements at the FEB, all (necessary) data will be preserved for 10 years according to FEB KU Leuven RDM policy. KU Leuven guarantees safe data storage and restricted data access for the duration of the project as well as a minimum period of 10 years after completion of the project.
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): After the ten-year post-completion storage period, project folders and data will be archived on the FEB's central servers (including automatic backup procedures) in agreement with KU Leuven's RDM policy.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Data preservation costs are managed by FEB / KU Leuven.

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: 1. & 2. & 4. & 5.: Secondary data under third party agreements (only accessible under license). I will / can not share this data. 3. & 6. & 7.: Secondary data – open access. I have no ownership rights over this data. I will not share this data. However, I can describe the data and indicate how and where others can find and obtain access to the data. 8. STATA scripts: Primary data. I can make my STATA Do Files (coding) openly available. In this way, researchers can build upon my contribution by utilizing the added value in the STATA scripts (such as newly defined variables & methodologic techniques & ideas of merging databases & ideas for unique identifiers per database). Researchers who also have access to the secondary databases can replicate my study with the STATA scripts. All other questions under 6. Data Sharing will be answered with respect to my primary data, i.e. STATA scripts (not regarding the secondary databases).
If access is restricted, please specify who will be able to access the data and under what conditions.	Other researchers who want access to my STATA scripts can contact me via mail and the scripts will be send in attachment (for instance, researchers who I meet on conferences or researchers who peer review my papers). The researchers who want access to my STATA scripts do not need a license, there seems no scientific or legal reason to create a license.

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
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) Data can be shared via mail.
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☒ Other (specify) For instance, when requested by colleagues for peer review.
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) No specific data usage license will be used. See explanation above.

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?	No costs.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Norah Kennis (norah.kennis@kuleuven.be) is responsible for data documentation and metadata during the research project. Norah Kennis is responsible for day-to-day data management operations.
Who will manage data storage and backup during the research project?	Norah Kennis is responsible for data storage and backup during the research project and will seek advice and help from the supervisor and principal investigator Steven Vanhaverbeke (steven.vanhaverbeke@kuleuven.be). Steven Vanhaverbeke can also potentially back-up data.
Who will manage data preservation and sharing?	Norah Kennis is responsible for data sharing. After the research project ends, preservation of data is the responsibility of Steven Vanhaverbeke and the KU Leuven RDM policy.
Who will update and implement this DMP?	In the course of the research, Norah Kennis is responsible for updating the DMP and implementing research data management. The supervisor, Steven Vanhaverbeke, is responsible for ensuring that the PhD student implements research data management.