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	Jakob Vanschoonbeek (0000-0001-8405-3430)
Contributor name(s) (+ ORCID) & roles	Jakob Vanschoonbeek (0000-0001-8405-3430)
Project number [1] & title	Anti-Revolutionary Roads - The Distributive and Political Impact of Belgian Workmen's Trains (12A4424N)
Funder(s) GrantID [2]	D-2024-2573
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	The invention of the steam engine in the 18th century started a transport revolution that would gradually erode the costs of moving goods, people and ideas. Belgium was a global frontrunner in this revolution and was among the first to build railroads on such a large scale that, by the end of the 19th century, it had the densest track system in the world. Due to a unique government policy of cheap workmen's tickets, it became one of the first places where the separation of workplace and residence not only became possible, but where it was actively advocated. It is widely accepted that this government policy served at least three policy goals. First, giving rural workers better access to well-paying manufacturing jobs in the city, to spread the benefits of industrialization more evenly across the population. Second, allowing workers to stay in their native villages, to prevent urbanization and its problems that typically accompanied industrialization. Third, slowing the electoral advent of the new socialist party, which could benefit from the concentration of workers in crowded urban centers. This project exploits rich and hitherto unused archival data on population, employment, wages, voting behavior and the light rail network for the period 1846-1914 to for the first time quantitatively assess the success of this unique Belgian transport policy against these three policy goals. It also aims to draw broader policy lessons on how commuting costs affect socio-economic outcomes.

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	**N**ew or **R**eused	**D**igital or **P**hysical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
		☐ Generate new data ☐ Reuse existing data	☐ Digital ☐ Physical	☐ **A**udiovisual ☐ **I**mages ☐ **Sou**nd ☐ **N**umerical ☐ **T**extual ☐ **M**odel ☐ **Sof**tware ☐ **O**ther:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
IC	Industrial Censuses	R	D	N	PDF/EXCEL	<1 GB	
PC	Population Censuses	R	D	N	PDF/EXCEL	<1 GB	
MPEC	Mouvement de la Population et de l'Etat Civil	R/N	D	N	PDF/EXCEL	<1 GB	
SNCV	Annual reports of Société nationale des chemins de fer vicinaux	N	D	N/I/O :geocoded maps	PDF/EXCEL/Shape files	<100 GB	
CLEA	Constituency-level electoral results	R	D	N	Stata .dta	<1 GB	

KKM	Voting results for local elections by constituency	R	P	N/T	PDF	< 1 GB	< 200 pages
KBR	Historical press articles	R	D	T	PDF	< 5 TB	
LER	Local election results	N	D	T/N	Excel/Stata .dta	<1 GB	

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.	• IC/PC: For the available (already digitized) Population and Industrial Censuses (PC/IC), https://bib.kuleuven.be/english/ebib/project-historical-census/project-bht • MPEC: https://www.queteletcenter.ugent.be/en/homepage/ • CLEA: https://electiondataarchive.org/ • KKM: KU Leuven SBIB • KBR: https://www.belgicapress.be/
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:

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

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: The digitized railway and light rail data will be donated to the NMBS and VlaTam archives (SNCV/SNCB), respectively. Interested parties that wish to obtain access to the data will be able to do so upon simple request to these archives.
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: See above.

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*.	In agreement with current best practices put forward by leading economic journals, metadata for each dataset used in the various work packages of the project will include a full description of the data set at hand, its coverage, variable names and descriptions, format and disk size. Additionally, working papers resulting from the work packages will contain a technical appendix detailing all data sources, data handling, variable creation. Upon completion of the project, all data (with the exception of proprietary data), computer code, logbooks, resulting graphs and tables will be made publicly available. All metadata will be accompanied by a corresponding README file that includes the former sources of information as well as a description of how to proceed with a replication analysis. More specifically, this file will explain the project folder structure and the set-up of the empirical analysis (when and where applicable). Embedded computer code (State, R, MATLAB, ...) will contain explanatory in-file documentation to explain basic data manipulations as well as estimation procedures and auxiliary output (tables, graphs and figures) generation for other researchers. More specifically, the following actions will be taken to make the project data FAIR (findable, accessible, interoperable and reusable): 1. Making data **findable**. For each of the datasets used, I will include a full description of the metadata in the coding and data pipelines, and on the project website. This metadata includes data source, link to the original datasets/archives, access requirements (if any), coverage, variable names, formats, and total disk size. A description of each variable (units, string/numeric etc.) will be given. I will also provide manuals and related documents wherever available from the providers. As mentioned below, these metadata and the accompanying descriptions will be documented either in the appendix of corresponding working papers, or in specified Technical Notes (TN) stemming from data-related tasks. 2. Making data openly **accessible**. Several datasets are publicly available, and will be made accessible to the general public through collaborations with several archives. I will provide access to all codes, logbooks and results (graphs, tables, etc.). For all datasets and results, including proprietary datasets, I will provide contact details on how to obtain access to the raw data. 3. Making data **interoperable**. The research team has extensive experience with digitizing data and large and complex project pipelines. To that end, I can follow standardized methods to do quality control and to provide accessible and reproducible research outputs. I will also develop README files explaining how different datasets can be merged and how different units of observation (e.g. municipalities, arrondissements, railway sections, ...) relate to each other.

	4. Increase data **re-use**. All of the newly digitized data will be made available to interested parties, either through the project website or through the relevant archives that will host the data. With respect to research output, all project code with in-file documentation will be made available through the project website and through replication files accompanying published (working) papers. This will allow full replication and further extension of all results
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 Technical Note (TN).

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)
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*	In addition to security measures put in place by the KU Leuven that restrict access to unauthorised users and guarantee secure data storage, internal procedures are implemented to restrict access to work package folders and proprietary data to users on a case-by-case basis to prevent improper use of these data sets.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Research groups at FEB contribute annually by means of a fixed sum paid to ICT-services to cover basic ICT-support which includes data storage, data backup and server maintenance. The ICT-costs stemming from research activities in this project are budgeted on a proportional basis.

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?	Preservation costs will be covered by the annual ICT-contributions discussed above and budgeted on future research projects.

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.	The data will be freely accessible, potentially upon user registration (e.g. VlaTAM).
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 digitized light rail data will be donated to the VlaTam archive (SNCB), respectively. Interested parties that wish to obtain access to the data will be able to do so upon simple request to these archives.
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?	Expected costs for data sharing will be low given the nature (binary files) and size (very small) of the project files that can be posted publicly, and mainly relate to setup and maintenance of corresponding project websites (GitHub). These are also budgeted on the project.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Responsibility for overall management of data documentation and metadata lies with the principal investigator (PI), Dr. Jakob Vanschoonbeek.

Who will manage data storage and backup during the research project?	Data storage and backup are facilitated by means of FEB protocols and managed automatically. The PI bears end responsibility .
Who will manage data preservation and sharing?	The IT-services at the Faculty of Economics are responsible for the data preservation; data sharing is handled by the PI. The PI will bear the end responsibility for data preservation and data sharing for the entire project.
Who will update and implement this DMP?	Updates to this DMP and its effective implementation are handled by the PI. He is assisted in his task by the Research Data Management (RDM) team of the KU Leuven, in particular with respect to follow-up of this DMP, see https://www.kuleuven.be/rdm/en for more information.