The Fast-Track to Technological Change: Railways, Market Access and Industrialization in Belgium

A Data Management Plan created using DMPonline.be

Creators: Jakob Vanschoonbeek, Anneleen Fastenaekels, n.n. n.n.

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n., Jakob Vanschoonbeek

Project Administrator: Anneleen Fastenaekels

Grant number / URL: G054823N

ID: 199746

Start date: 01-01-2023

End date: 31-12-2026

Project abstract:

The project will for the first time digitize spatially and sectorally disaggregated high-quality census information on labor and output market variables from the various Belgian population, industrial and agricultural censuses; as well as detailed information contained in the historical (light) railway and patent archives for the period 1846-1914. These data are subsequently used to better understand how Belgium’s forerunner role in ‘railwayfication’ has shaped and continues to shape its economic performance. The analysis of the data will combine the development of theoretically founded structural trade and economic geography models that integrate existing insights in the links between transport innovations, technology adoption and technological progress with reduced-form estimating equations that provide insights in main elasticities in the model and allow for a quantitative validation for the Belgian case. The central focus is to analyze the effect of increased market access on the intensive and extensive margins of technological development: (i) on the intensive margin, the extent to which market access and increased firm competition facilitated the adoption of cost-saving technologies for existing firms and (ii) on the extensive margin, the extent to which market access and increased effective demand allowed entrepreneurs to found firms active in new sectors which require large initial capital investments yet push up manufacturing productivity.

Last modified: 30-05-2023

The Fast-Track to Technological Change: Railways, Market Access and Industrialization in Belgium
FWO DMP (Flemish Standard DMP)

1. 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.

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)
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • **O**bservational • **E**xperimental • **C**ompiled/aggregated data • **S**imulation data • **S**oftware • **O**ther • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA
SNCB	Annual reports of the Société nationale des chemins de fer belges	R/N	D	C	.dta / .shp	<100GB
SNCV	Annual reports of Société nationale des chemins de fer vicinaux	R/N	D	C	.dta / .shp	<100GB
IC	Industrial Censuses	R	D	C	.csv	<1GB
AC	Agricultural Censuses	R	D	C	.csv	<1GB
MPEC	Mouvement de la Population et de l'Etat Civil	R/N	D	C	.csv	<1GB
ABIPO	Archives Belgian Intellectual Property Office	N	D	C	.csv	<100GB
AIdM	Archives Inspection de Mines	N	D	C	.csv	<1GB

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:

- For the available (already digitized) Population, Industrial and Agricultural (PC/IC/AC): https://bib.kuleuven.be/english/ebib/project-historical-census/project-bht
- For the Mouvement de la Population et de l'Etat Civil (MPEC): https://www.queteletcenter.ugent.be/en/homepage/
- The raw data from the SNCV, which will be newly digitized, is located in the VlaTam archives, see https://www.vlatam.be/over-ons/.
- The raw data from SNCB, which will be newly digitized, is located in the archive at trainworld, see https://www.trainworld.be/nl/collecties/erfgoeddatabank
- ABIPO: https://www.epo.org/applying/online-services/online-filing/national/be.html

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

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.

- No

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 in the comment section to what data they relate and what restrictions are in place.

- Yes

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 in the comment section to what data they relate and which restrictions will be asserted.

- No

2. 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).

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, we 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. We 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. We will provide access to all codes, logbooks and results (graphs, tables, etc.). For all datasets and results, including proprietary datasets, we will provide contact details on how to obtain access to the raw data.
3. Making data **interoperable**. Our research team has extensive experience with digitizing data and large and complex project pipelines. To that end, we have developed standardized methods to do quality control and to provide accessible and reproducible research outputs. We 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 throught 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 for interested researchers that have cleared access to the respective datasets. Accessible data will be provided in open formats (e.g., .csv) for use by others.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

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).

3. Data storage & back-up during the research project

Where will the data be stored?

In general, data are stored on a private data server at the Faculty of Economics and Business, which handles data storage, data management and data acces in agreement with KU Leuven regulations. Furthermore, data transfer (through Belnet for example) and storage protocols are controlled by the university's Data Protection Officer (DPO). For particular research tasks in specified work packages, data will be stored on dedicated computing servers to increase computational speeds. In this, an additional layer of security is added on top of the university- and faculty-wide security and access protocols. Access to folders and use of specified data sets is then granted on a case-by-case basis for involved researchers by the system administrators of the respective computing servers.

How will the data be backed up?

Standard back-up provided by KU Leuven ICTS. Data storage, management and access are handled by the KU Leuven as well as IT-services at the Faculty of Economics and Business (see also above). The latter have the necessary infrastructure and protocols in place to simultaneously backup data on multiple servers.

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

Storage capacity is provided by (1) the Faculty of Economics and Business through the system of internal disks with allocated slots for separate research groups, (2) disk space on dedicated computing servers, and (3) the KU Leuven through the OneDrive system that provides researchers with both personal and shared storage facilities (should the need arise). All types of storage are sufficient to handle backups of the data used in the project.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

In addition to security measures put in place by the KU Leuven that restrict acces to unauthorised users and guarantee secure data storage, internal procedures are implemented to restrict acces to work package folders and proprietary data to users on a case-by-case basis to prevent improper use of these data sets, see also above.

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.

4. 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...).

All data will be preserved for 10 years according to 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. This basically covers publicly available data; storage of proprietary data however is stipulated by contractual obligations and often involve physical removal from disks after completion of the project.

Where will these data be archived (stored and curated for the long-term)?

After the ten-year post-completion storage period, project folders and data will be archived on the KU Leuven 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?

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

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository
- Yes, in a restricted access repository (after approval, institutional access only, ...)

Most of the newly digitized data will be given to the respective archives (e.g. the SNCV archive at VlaTAM or the Project Historical Census Data of the KU Leuven for the census data), who will be responsible for the further exploitation of this data. The data will be freely accessible, potentially upon user registration (e.g. VlaTAM).

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 in the comment section per dataset or data type where appropriate.

- Yes, Intellectual Property Rights

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.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

See above.

When will the data be made available?

After the completion of the relevant working papers.

Which data usage licenses are you going to provide? If none, please explain why.

CC BY, see https://creativecommons.org/licenses/ for more information.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- 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.

6. 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), Prof. Dr. Joep Konings.

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, Prof. Dr. Joep Konings, 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 respective work package leaders. Prof. Dr. Joep Koning 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, Prof. Dr. Joep Konings. 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.

The Fast-Track to Technological Change: Railways, Market Access and Industrialization in Belgium
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

The research will gather and newly digitize data from the annual reports of the Société nationale des chemins de fer belges (SNCB); the Société nationale des chemins de fer vicinaux (SNCV); the Belgian Intellectual Property Office; and the industrial censuses. The newly digitized data is first scanned to PDF format and then manually converted to a machine-readable format (Excel), which will in turn be converted to .dta files .The project will also georeference yearly maps of the railroad and light rail network from the same sources to GIS files. Finally, the project will use digitized versions of the agricultural and population census by the Quetelet center and population information from of the Mouvement de la Population et de l'Etat Civil.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

Data storage, management and access is provided by the home institution of the KU Leuven, which already stores some of the (partial) datasets to be used and has the necessary protocols and infrastructure in place to back up the data simultaneously on several servers. Data will be stored on secured and restricted-access data servers and will be protected by the security measures already in place, guaranteeing safe data storage for the duration of the proposal as well as a minimal period of 10 years after its completion. The promotors will be individually responsible for data preservation and data sharing. The Principal Investigator, Joep Konings, will bear the end responsibility.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

NA.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

NA.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

The digitized railway and light rail data will be donated to the NMBS and VlaTam archives,
respectively. Interested parties that wish to obtain access to the data will be able to do so upon simple request to these archives.

The Fast-Track to Technological Change: Railways, Market Access and Industrialization in Belgium
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- No

The Fast-Track to Technological Change: Railways, Market Access and Industrialization in Belgium
GDPR

GDPR

Have you registered personal data processing activities for this project?

- No