Innovative Entrepreneurship in Cities and Regions: Knowledge Connectivity, Incumbent Appropriation Strategies, and Multinational Firms

A Data Management Plan created using DMPonline.be

Creators: Qian Zhou https://orcid.org/0000-0001-9028-5860 René Belderbos http://orcid.org/0000-0002-4083-3387

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Data Manager: Qian Zhou https://orcid.org/0000-0001-9028-5860

Project Administrator: Rene Belderbos

ID: 200506

Grant Number: G046823N

Start date: 01-01-2023

End date: 31-12-2026

Project abstract:
Innovative entrepreneurship introduces new products, processes, or business models that form the basis of the development of new markets and long term prosperity of countries and regions. Incumbent (multinational) firms may however leverage their hold on global intangible assets and appropriation strategies to fence off the competitive threat of new innovative entrants. Innovative entrepreneurship is particularly salient in metropolitan areas with strong international knowledge connections ('global cities'). We examine the influence of knowledge connectivity, ownership concentration of innovation, foreign and domestic multinational firm activity, and appropriation strategies in facilitating or hampering innovative entrepreneurial entry, growth of innovative entrants, and economic growth and resilience in (metropolitan) regions. To this end, we construct and analyse a comprehensive database on corporate innovative entry and growth in European regions and industries at a fine-grained (NUTS-3 and NACE-4) level.

Last modified: 23-06-2023

Innovative Entrepreneurship in Cities and Regions: Knowledge Connectivity, Incumbent Appropriation Strategies, and Multinational Firms
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.

Patent data will be drawn from the PATSTAT database developed by the European Patent Office, which is available at the department. Trademark data will be retrieved from EUIPO, including both the European and national trademarks. Geocoding of inventor locations has also been completed at the department and additionally uses the GIS format. This covers 40 million patent records and in its current form covers about 10 GB. The concentrated ownership dataset covers the 8500 largest patent-owing firms and is available at the department.

Dataset Name	Description	New or reused	Digital or Physical	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:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • 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
Orbis firms	Business register data including new entrants, incumbents	Generate new data	Digital	Compiled/aggregated data	STATA (.dta)	<100GB
Patent data	Retrieved from PATSTA, including applicants, inventors, and technology classifications for both European patents and national patents	Generate new data	Digital	Compiled/aggregated data	STATA (.dta)	<100GB
Trademark data	Retrieved from EUIPO, including applicants, NICE code for both European trademarks and national trademarks	Generate new data	Digital	Compiled/aggregated data	STATA (.dta)	<100GB
Geocoded inventors and applicants for patents	Geocoded inventors and applicants data based on functional urban areas	Reused	Digital	Compiled/aggregated data	STATA (.dta)	<100GB
Patent ownership	Consolidated patent ownership of 8500 largest patent-owing firms	Reused	Digital	Compiled/aggregated data	STATA (.dta)	<100GB

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:

Geocoded investors and applicants for patents (https://lirias.kuleuven.be/3655250?limo=0)
Patent ownership concentration data. Available on MSI
server(X:\MSI_MSI\Geon\Concentration (HHI))

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.

- No

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.

- Yes

We use ORBIS data of Bureau van Dijk, among other sources. Since this database is accessed under license (KU Leuven library), any future user of the data is likely to require to have similar access rights.

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).

Metadata is available inside the STATA database in the form of variable names and labels. A readme file with information on the variables and data structure is also available. A STATA program indicating the data cleaning and processing is available.

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

The data structure and variables are relatively easy to document with labels and a readme file, and this should be sufficient for reuse.

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

Where will the data be stored?

Data will be stored on the servers of the MSI department at KU Leuven and on KU Leuven Onedrive.

How will the data be backed up?

Backups are generated every day by the ICT unit at FEB and ICTS at KU Leuven.

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

There is sufficient space for data storage on the FEB servers.

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

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?

This is a service of the FEB KU Leuven and has no additional cost

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 retained for at least five years.

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

The data will be stored on FEB servers of the department.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No costs are expected for the internal use of the server for data storage.

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 a restricted access repository (after approval, institutional access only, ...)

We intend to use the data for follow-up studies and PhD projects, but we do not intend to make the data publicly available. Researchers can request access and we may share the data depending on the intended use. The availability of Orbis data is conditional on the users' access to Orbis.
The data will be available in STATA and excel formats.

If access is restricted, please specify who will be able to access the data and under what conditions.

Upon request by mail.

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 Orbis database is under the license of KU Leuven library, so any further user of the data is likely to have similar access rights.

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

Upon request by mail. The data will be available in STATA and excel formats.

When will the data be made available?

Upon publication of the research results.

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

We will ask for an agreement to share improvements of the data. If the user is willing to share his/her dataset and is able to add or improve the current datasets, we will share our datasets for their use.

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?

No costs.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Qian Zhou

Who will manage data storage and backup during the research project?

Qian Zhou

Who will manage data preservation and sharing?

René Belderbos

Who will update and implement this DMP?

René Belderbos bears the end responsibility of updating & implementing this DMP