Plan Overview

A Data Management Plan created using DMPonline

Title: Not All Collaborators Are Equal: Partner Heterogeneity in Eco-Innovation — Evidence from European Firms

Creator:Xinyue Zhang

Affiliation: University of Birmingham

Template: UoB short template

Project abstract:

This project investigates how inter-organisational collaboration contributes to the breadth of firms' eco-innovation, and how the national institutional environment conditions the value of different types of collaboration partner. Theoretically, it integrates the natural-resource-based view of the firm with a taxonomy of partnership types — vertical (suppliers and customers), horizontal (competitors), and institutional (universities and public research organisations) — and tests whether country-level institutions act as complements to, or substitutes for, firm-level partnerships in driving environmental innovation.

The study is a quantitative secondary analysis of existing, anonymised data. The primary dataset is firm-level microdata from the Community Innovation Survey (CIS), an official harmonised survey of innovation activity conducted by national statistical offices across European countries and distributed for research purposes through Eurostat. The analysis pools the 2008 and 2014 waves — the most recent waves for which a harmonised cross-country environmental-innovation module is available at the micro level — yielding a sample of 38,745 eco-innovating firms across fifteen countries. These firm-level data are matched, at the country and survey-year level, to publicly available institutional indicators drawn from established secondary sources (including OECD product market regulation indices, intellectual property rights indicators, and country-level corporate social responsibility measures).

The project involves no human participants, no primary data collection, and no personally identifiable information; the CIS microdata concern firms rather than individuals and are de-identified prior to release. The data are accessed and used under the relevant Eurostat data-use agreement, which governs confidentiality, permitted use, and storage conditions. All analysis is conducted at the host institution using statistical software, with the data held in accordance with the access agreement's security requirements. No new data are generated other than derived analytical variables and statistical output; the underlying microdata are not redistributed, and any outputs are produced at a level of aggregation consistent with the confidentiality rules of the data provider.

ID: 206407

Start date: 01-09-2025

End date: 30-09-2028

Last modified: 14-06-2026

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Not All Collaborators Are Equal: Partner Heterogeneity in Eco-Innovation — Evidence from European Firms

Data description

What types of data will be used or created?

This project uses and creates exclusively numeric and textual data; it involves no image, audio, video, or mixed-media data, and no interview or recorded data of any kind.

Existing data to be re-used. The project is a secondary analysis built primarily on existing data rather than newly collected data. Two categories of existing data are re-used:

The principal dataset is firm-level microdata from the Community Innovation Survey (CIS), waves 2008 and 2014, obtained from Eurostat under the relevant data-use agreement. These are quantitative, structured survey data supplied in statistical data-file format. The data are received and held in Stata format (.dta), the working format for the analysis; where the original files are distributed as delimited text, they are imported and stored as .dta. Stata format is used because it is the established standard for the analytical software employed, preserves variable and value labels essential to the harmonisation of the two survey waves, and is the format in which the analysis is conducted and version-controlled. For interoperability and longer-term accessibility, key derived datasets are additionally exportable to comma-separated values (.csv), an open, non-proprietary, plain-text format suitable for long-term archiving and re-use across platforms.

Country-level institutional indicators (covering product market regulation, intellectual property rights, and corporate social responsibility) are re-used from established public secondary sources, including the OECD and comparable international datasets. These are numeric data obtained in spreadsheet (.xlsx / .csv) format and stored as .csv for processing, before being merged with the firm-level data.

The project generates no new primary data. The data created are limited to: (i) derived analytical variables (for example, the eco-innovation breadth index and the partner-type indicators) constructed from the existing CIS variables; (ii) analysis code, written and stored as Stata do-files (.do), which is plain-text and openly readable; and (iii) statistical output and results, exported as plain-text logs (.txt / .log), rich-text and document files (.rtf / .docx) for tables, and .csv for tabular results. Code and documentation files are maintained in plain-text or open formats wherever possible to support transparency and reproducibility.

Open and standard formats (.csv, .txt, .do) are preferred throughout for derived data, code, and output, as recommended for long-term archiving and to facilitate sharing and re-use. The one proprietary format retained is Stata's .dta, used for the working analytical files because it is required by the analysis software and uniquely preserves the labelled variable structure on which the multi-wave harmonisation depends; all such files can be, and are, exported to open formats (.csv) for archiving and sharing of any shareable derived material.

No data are collected from human participants, and no recording platforms (e.g. Zoom, Microsoft Teams) are used, as the project involves no interviews, observations, or primary data collection. The CIS microdata are accessed and processed under the confidentiality conditions of the Eurostat data-use agreement; the data are de-identified at source (they concern firms, not individuals, and contain no personal data), and are handled and stored in accordance with the agreement's security requirements throughout.

How will the data be structured and documented?

- Research data will be organized using a structured, shallow hierarchical folder system on the UoB Microsoft cloud, separating raw datasets, processed datasets, analytical scripts, and documentation.
- A systematic naming convention will be applied to all files, incorporating project identifiers, data types, dates, and version numbers to track revisions.
- A formal version control table will be maintained to track major milestone versions of the merged datasets, recording details of changes, authors, and dates.

Data storage and archiving

How will your data be stored and backed up?

All electronic project data, including the restricted Community Innovation Survey (CIS) micro-data, country-level institutional datasets, and analytical scripts, will be stored securely on the University of Birmingham (UoB) Research Data Store and Archive.

Given the highly sensitive and restricted nature of the firm-level CIS data, any data temporarily held on local or removable drives for immediate analytical processing will be strictly encrypted to prevent unauthorized access.

To ensure robust data preservation and disaster recovery, the UoB Research Data Store automatically backs up all project files to two physically separate, geographically distinct data devices. This ensures that no single point of failure can result in data loss.

Storage Volume Management: The project will operate within the free storage volume allocated to researchers by UoB IT services, as the panel data and Stata/EViews scripts do not exceed standard capacity limits. Should the volume of processed data expand unexpectedly, the principal investigator will contact bearstorage@contacts.bham.ac.uk to arrange and budget for additional storage support.

Is any of the data of (ethically or commercially) sensitive nature? If so, how do you ensure the data are protected accordingly?

Yes, the research involves highly confidential and commercially sensitive data. The primary dataset consists of firm-level micro-data from the Community Innovation Survey (CIS), which contains proprietary information regarding business innovation activities, strategic partnerships, and firm performance.

To ensure the data is protected accordingly, the following measures will be implemented in line with UK GDPR, the Data Protection Act 2018, and UoB BEAR Services guidance:

1. Access Control and Secure Storage: The raw CIS micro-data will be stored exclusively on the secure, password-protected, and encrypted University of Birmingham Research Data Store. Access to this data is strictly restricted to the principal investigator and the immediate supervisory team.
2. Data Protection and Informed Consent: While the core dataset is secondary firm-level data, any administrative or project tracking files containing personal data will be managed in full compliance with UK GDPR. Informed consent will be maintained for all research governance purposes.
3. Anonymization for Dissemination: To protect the commercial sensitivity of the participating firms, no individual firm-level observations will ever be disclosed or shared. All regression outputs, summary statistics, and analytical tables will be aggregated at the industry or country level. This ensures the final public data outputs are truly anonymized, removing them from the scope of personal or commercially identifying data.

Where will your data be archived in the long term?

At the publication of each paper, the specific subset of data and materials that underpins the findings—specifically the custom analytical scripts (Stata/EViews do-files) and the aggregated country-level institutional datasets—will be transferred to the University of Birmingham BEAR Archive for long-term preservation. Once transferred, these files will be set to read-only to ensure data integrity and prevent inadvertent alterations. In accordance with the University's data retention policy, these replication materials will be securely stored for a minimum of 10 years, with the retention clock resetting upon any formal access requests.

Crucially, because the raw firm-level Community Innovation Survey (CIS) micro-data is highly restricted and subject to strict third-party licensing and confidentiality agreements, it will not be archived indefinitely or shared in the University's public repository. Instead, the raw micro-data will be securely deleted or managed at the end of the project strictly in compliance with the retention and disposal mandates specified by the original data provider. Only anonymized, aggregated outputs and the code used to generate them will be kept for long-term institutional archiving.

Data sharing

Which data will you share, and under which conditions? How will you make the data available to others?

1. Select the Primary Repository Option

Data will be shared through the University of Birmingham's eData repository (https://edata.bham.ac.uk/) which makes the datasets discoverable through search engines like Google. It provides a range of options for open licensing of data but also allows closed and embargoed deposits where required. Ubira eData uses Dublin Core as a metadata standard and the minimum metadata provided for published datasets will cover amongst others title, type of data, creators, publication date and related publications.

2. Accompanying Explanatory Text to Insert Into the Form

(Copy and paste the text below into the accompanying text box to explain your specific conditions, embargoes, and restrictions.)

Justification for Restricted and Shared Components:

The data management for this project follows a tiered approach based on third-party confidentiality and licensing restrictions:

- Restricted Data (Not Shared): The raw, firm-level micro-data from the Community Innovation Survey (CIS) is highly sensitive and subject to strict third-party data user agreements. Consequently, this raw data component cannot be openly shared or deposited into a public repository.
- Shareable Replication Data & Code: To satisfy institutional transparency requirements, the custom analytical scripts (Stata/EViews .do files) and the aggregated country-level institutional dataset (PMR, IPR, and CSR indices) will be fully shared.

Embargo and Licensing Conditions: At the completion of the project, these shareable replication materials will be archived in the University of Birmingham's eData repository under a one-year embargo to enable final publications to be made without risking premature disclosure of active research models.

Following the expiration of the one-year embargo period, the metadata record will be fully opened, and the replication scripts will be released under an open Creative Commons Attribution (CC-BY 4.0) licence to maximize capacity for academic re-use and verification.