Global MNC Panel Dataset (2010–2024): Firm–Country Data on Strategy, Institutions, and Contexts

Integrated Panel Data on Multinational Corporations and Country-Level Environments

FINAL REPORT
AND DATA DOCUMENTATION

Version 1.0 – August 2025

Sergey Morgulis-Yakushev (Principal Investigator)
Uppsala University / Mälardalen University

Olivia Kang (Project Lead)
Mälardalen University

Emre Yildiz, Aswo Safari, Sara Melén Hånell (Advisory Group)
Mälardalen University

sergey.morgulis-yakushev@fek.uu.se

https://mjserg.github.io/mnc-panel-dataset/

Mälardalen University, Sweden

1. Executive Summary

This technical report introduces a novel, multi-level dataset designed to support empirical research on multinational corporations (MNCs) and the institutional contexts in which they operate. Built with academic rigor and cross-disciplinary usability in mind, the dataset integrates firm-level and country-level data for the years 2010–2024, enabling longitudinal, comparative, and multi-level analyses of international business phenomena.

Objectives and Scope

The primary goal of the project is to provide a panel-structured, globally comparable empirical resource that bridges two traditionally separated levels of analysis:

- **Firm-level data** on financial performance, corporate governance, employment, and internationalization.
- **Country-level data** on institutional quality, political regimes, macroeconomic conditions, cultural values, and environmental sustainability.

The dataset comprises:

- **560 publicly listed multinational manufacturing firms** from 35 countries;
- **Country-level indicators** for up to **85 countries**, totaling more than **75 macro-level variables**;
- A **15-year observation period** (2010–2024), covering pre- and post-COVID years, political transformations, and major global shocks.

Key Features

- **Two-level panel structure**: Firm-year and country-year data are linked via standardized country codes, enabling firm–environment interaction studies.
- **Multi-source integration**: Data are compiled from reputable academic and commercial databases, including *Refinitiv Eikon, Orbis, Forbes Global 2000, World Bank, V-Dem, Hofstede Insights, WVS*, and others.
- **Longitudinal design**: The 15-year time span allows researchers to investigate dynamic changes, institutional shifts, and long-term performance patterns.

- **Standardized format and documentation**: All variables follow a uniform naming convention (snake_case), with accompanying data dictionaries, metadata, and reproducibility guidelines.
- **Theoretically grounded selection**: Variables are curated to align with key concepts in institutional theory, transaction cost economics, comparative capitalism, and liability of foreignness.

Novelty and Contribution

This dataset fills a major gap in the field by offering one of the few globally structured, panel-ready datasets that combine:

- detailed firm-level metrics,
- rich country-level institutional indicators,
- and explicit design for multi-level modeling.

Unlike many existing datasets focused on single countries, static snapshots, or sector-specific metrics, this resource provides the breadth and depth necessary for high-quality comparative research.

Use Cases and Applications

The dataset is intended to support a wide range of academic and applied research, including but not limited to:

- Effects of institutional quality on firm performance;
- Cross-national differences in governance and ownership structures;
- The role of political regimes in shaping internationalization strategies;
- Cultural distance and adaptation of MNC strategies;
- Environmental or ESG strategies in different institutional contexts.

It is particularly relevant for scholars and students in international business, political economy, comparative institutions, and strategic management.

2. Introduction

This project was developed with the aim of constructing a multi-level, panel-structured, and theoretically informed dataset that enables researchers to analyze the strategies and behaviors of multinational corporations (MNCs) across diverse institutional environments. It combines firm-level information—including financial performance, corporate governance structures, and internationalization metrics—with country-level indicators on institutional quality, political regimes, culture, and environmental conditions for the period 2010–2024.

The dataset is designed to support multiple research agendas across fields such as:

- international business and global strategy,
- institutional theory and comparative political economy,
- strategic management and corporate governance.

The motivation for building this resource stems from both academic and practical challenges. In an era of geopolitical uncertainty, regulatory divergence, and institutional flux, scholars increasingly require tools to compare firm behavior across countries and regimes. Previously available datasets have typically suffered from limitations such as narrow geographic scope, lack of longitudinal structure, or insufficient linkage between firms and their operating environments.

This dataset addresses these gaps by providing:

- a globally comparable panel of 560 publicly listed manufacturing MNCs from 35 countries;
- a country-level panel spanning up to 85 countries, with indicators across macroeconomic, institutional, political, and cultural domains;
- a 15-year observation window capturing both stable and turbulent periods;
- a two-level structure suitable for multi-level and cross-level empirical modeling.

The resource is intended for use in:

- academic research, doctoral dissertations, and peer-reviewed publications;
- graduate-level coursework in international strategy, institutional economics, and comparative research;
- applied policy and business analysis requiring cross-national comparability.

In this sense, the dataset functions not merely as a collection of variables, but as a comprehensive research infrastructure built to the standards of academic rigor, methodological transparency, and reproducibility.

3. Conceptual Motivation and Research Use Cases

This dataset is explicitly designed to support empirical research grounded in leading theoretical frameworks from international business, institutional theory, and comparative political economy. By integrating firm-level and country-level indicators, it enables scholars to investigate how multinational corporations (MNCs) adapt their strategies, structures, and performance in response to heterogeneous institutional, political, and cultural environments.

3.1. Theoretical Foundations

1. Institutional Theory

The dataset facilitates exploration of how formal and informal institutions shape firm behavior and outcomes. Country-level variables include measures of rule of law, regulatory quality, corruption, political freedoms, and cultural norms, offering an empirical basis to assess institutional embeddedness, legitimacy-seeking behavior, and institutional complementarity.

2. Liability of Foreignness (Zaheer, 1995)

This framework highlights the additional costs faced by foreign firms operating abroad. Using proxies such as the share of international sales and assets, researchers can assess how political instability, authoritarianism, or weak legal institutions amplify these liabilities and affect internationalization strategies.

3. Transaction Cost Economics

Indicators of institutional reliability and transparency allow for empirical testing of transaction costs associated with cross-border operations. The data support research into governance mode choices (e.g., wholly owned subsidiaries vs. joint ventures) and the institutional determinants of firm boundaries and investment intensity.

4. Mixed Embeddedness Perspective

The two-level dataset structure — linking firm-specific characteristics to national contexts — aligns with the mixed embeddedness approach, which emphasizes the dual anchoring of firms in both global markets and local institutional environments. This is especially relevant for research at the intersection of international business and comparative sociology.

5. Varieties of Capitalism / Comparative Capitalism

The inclusion of political regime types, legal institutions, and indicators of state capacity enables comparative analysis of different capitalist models (liberal, coordinated, state-led). Researchers can examine how firms adapt across these systems and how institutional complementarities affect firm outcomes.

3.2. Empirical Use Cases

The dataset supports a wide array of empirical research questions, such as:

- **How does institutional quality influence corporate performance (e.g., ROA, revenue growth)?**
- **Are levels of firm internationalization (foreign assets or sales) correlated with host-country political institutions (e.g., democracy, corruption control)?**
- **Does cultural distance — measured by power distance, individualism, or uncertainty avoidance — affect market entry strategies?**
- **How do firms respond to global shocks (e.g., COVID-19, sanctions, trade wars) under different institutional regimes?**
- **Can board characteristics (size, diversity, tenure) buffer institutional risks in volatile environments?**
- **Which country-level factors most strongly predict firm internationalization intensity across regions and industries?**

4. Data Architecture

The dataset is built on a two-level panel structure that integrates firm-level and country-level information over a 15-year period (2010–2024). This design enables researchers to model both intra-firm dynamics and cross-national institutional variation with methodological rigor and conceptual clarity.

4.1. Two Interconnected Levels

Firm-Level Panel

This level contains annual observations for 560 publicly listed multinational manufacturing firms. Variables cover financial indicators (revenue, assets, net income), corporate governance (board structure, compensation, diversity), internationalization proxies (foreign assets and sales), and workforce characteristics.

Country-Level Panel

The second level consists of approximately 75 macro-level indicators per country-year, covering political regimes, institutional quality, cultural values, economic development, environmental performance, innovation capacity, and state capacity. Data span up to 85 countries, providing rich cross-sectional and temporal variance.

4.2. Linkage and Integration

The two levels are linked via:

- **Country identifiers**: ISO 2-letter and 3-letter codes (iso2c, iso3c)
- **Temporal variable**: calendar year (year)

This architecture allows for precise merging of contextual variables with firm-level records and supports designs where both within-country and between-country effects are of interest.

4.3. Why the Structure Matters

The dual-level structure enables a range of advanced empirical strategies:

- **Panel regression models**: Firm-level data are organized in long format to support fixed-effects, random-effects, and time-series cross-sectional models.
- **Multilevel modeling (HLM)**: The nested data architecture allows for the estimation of cross-level interactions — for example, how the effect of board diversity on performance may depend on the country's gender norms or regulatory quality.
- **Institutional interaction analysis**: Researchers can examine how firm outcomes are moderated by institutional environments — including differences across democracies, authoritarian regimes, or hybrid contexts.
- **Longitudinal policy analysis**: The 15-year span allows for the study of institutional reforms, crises (e.g., COVID-19), or geopolitical events and their firm-level consequences.

By combining firm-level heterogeneity with macro-level context, the dataset reflects the complexity of real-world international business environments. It enables both depth (detailed firm attributes) and breadth (institutional diversity), making it particularly well suited for theory-driven, methodologically robust research in international strategy, comparative institutions, and global political economy.

5. Sample Design

The sample was constructed with the goal of creating a globally representative, theoretically meaningful, and methodologically consistent panel of multinational corporations (MNCs), with a specific focus on the manufacturing sector. The selection process was designed to ensure both cross-national diversity and depth of firm-level data over time.

5.1. Sampling Frame and Inclusion Criteria

The initial sampling frame was based on the **Forbes Global 2000** list (2024 edition), which ranks the world's largest publicly traded firms based on a composite score of revenues, profits, assets, and market value. From this list, firms were included in the sample if they met the following criteria:

- **Manufacturing sector affiliation**, defined by SIC codes 2000–3999
- **Publicly listed status**, to ensure data transparency and availability
- **Headquartered in one of the major world regions**
- **Availability of longitudinal financial and governance data** (2010–2024)

The Forbes ranking was used as a global benchmark of economic significance, while financial and organizational variables were retrieved from **Refinitiv Eikon** and **Orbis Global** to validate and enrich the sample.

5.2. Sectoral Focus: Why Manufacturing?

The decision to focus on **manufacturing MNCs** was both conceptual and empirical:

- Manufacturing firms typically engage in **asset-intensive internationalization strategies**, such as establishing subsidiaries, production plants, and distribution networks abroad.
- These firms face **higher exposure to host-country institutions**, regulatory risk, and liability of foreignness — making them especially suitable for institutional analysis.
- Compared to service or extractive industries, manufacturing offers greater **data consistency**, comparability, and sectoral homogeneity across countries.

5.3. Geographic and Sectoral Coverage

The final sample includes **560 multinational manufacturing firms** headquartered across five macro-regions:

Region	Firms	Share (%)
United States	171	30.5

Region	Firms	Share (%)
Asia	225	40.2%
Americas	189	33.8%
Europe	143	25.5%
Oceania	2	0.4%
Africa	1	0.2%
Total	**560**	**100%**

The top countries by firm count include the United States (171), China (89), and Japan (64), while the sectoral composition reflects major industrial categories such as consumer durables, capital goods, pharmaceuticals, chemicals, electronics, and food manufacturing.

5.4. Representativeness and Limitations

While the sample does not claim to be exhaustive, it captures a **substantial share of global MNC activity in manufacturing**, both in terms of revenues and market capitalization. It is **not a random sample**, but a **purposive selection** intended to maximize analytical relevance, international comparability, and data richness.

Limitations include:

- Underrepresentation of small and mid-sized firms
- Limited coverage of emerging-market firms outside of Asia
- Exclusion of non-manufacturing sectors due to lack of consistent firm-level metrics

The resulting dataset provides a **balanced and analytically coherent sample** of global manufacturing MNCs. It allows researchers to explore firm strategies, performance, and adaptation in relation to a wide spectrum of national institutional environments, while maintaining methodological consistency and sectoral focus.

6. Data Sources

This dataset combines **firm-level** and **country-level** data from a curated set of high-quality, internationally recognized sources. The integration of these two tiers enables multi-level and comparative analysis, allowing researchers to examine how firm behavior is shaped by and interacts with the broader institutional and macroeconomic context.

All sources were selected based on four core criteria:

- **Transparency and methodological clarity**
- **Longitudinal availability**
- **Cross-national comparability**
- **Academic credibility and citation footprint**

6.1. Firm-Level Data

Firm-level data were primarily obtained from the following sources:

- **Refinitiv Eikon**: A leading commercial database providing structured financial and organizational data. Eikon was the main source for:
  - Revenue, net income, total assets
  - Market capitalization
  - Return on assets (ROA), return on equity (ROE)
  - Number of employees
  - Board size, compensation, tenure, and gender composition
- **Orbis Global (Bureau van Dijk)**: Used to supplement missing governance attributes, verify identifiers, and cross-validate firm structures where Eikon data were incomplete.
- **Forbes Global 2000 (2024 Edition)**: Served as the sampling frame and auxiliary source for:
  - Industry classification
  - Global firm ranking
  - Headquarters location

- Year of founding and CEO identity

Rather than attempting to collect fragmented subsidiary-level information, we included **internationalization proxies**:

- Share of international sales (as % of total sales)
- Share of international assets (as % of total assets)

These indicators offer scalable and comparable measures of cross-border exposure and are especially useful in the absence of detailed subsidiary-level disclosures.

6.2. Country-Level Data

Country-level indicators were retrieved from widely used public datasets, covering political institutions, economic development, culture, environmental performance, and public sector capacity:

- **World Bank (World Development Indicators, Worldwide Governance Indicators)**
- **V-Dem Institute (Varieties of Democracy Project)**
- **Hofstede Insights**
- **World Values Survey (WVS)**
- **Freedom House (Freedom in the World Index)**
- **Environmental Performance Index (Yale)**
- **Global Health Security Index (GHSI)**
- **World Economic Forum (Ease of Doing Business)**
- **World Intellectual Property Organization (WIPO)**

These sources were chosen for their:

- Regular and systematic updates (typically annual or biannual)
- Clear methodology and transparent documentation
- Consistency across countries and over time
- Established usage in international business, political economy, and comparative institutional research

The resulting **country-level panel** spans up to **85 countries** over **15 years (2010–2024)** and includes over **75 indicators**, harmonized to support multilevel and longitudinal analysis.

The use of established data providers across both tiers ensures that the dataset meets high standards of **validity, reliability, and replicability**. All variables are fully documented in separate data dictionaries, including information on units, sources, missing data, and coding logic.

This dual-source architecture forms the empirical backbone of the project, enabling research at the intersection of **corporate strategy** and **institutional context**.

7. Data Collection and Cleaning

The data collection and cleaning procedures were designed to ensure **reproducibility**, **cross-source coherence**, and **longitudinal consistency**. Given the dataset's dual structure—firm-level and country-level—each layer underwent a dedicated yet coordinated workflow of data extraction, standardization, and validation.

7.1. Collection Strategy and Automation

Firm-level and country-level data were retrieved using a combination of:

- **Manual queries** via subscription platforms (Refinitiv Eikon, Orbis Global)
- **Structured imports** from public data portals (World Bank, V-Dem, EPI, etc.)
- **Automated pipelines** built in R and Python for:
  - Reading and reshaping Excel/CSV files (e.g., readxl, pandas, tidyr)
  - Long-format panel transformation
  - Variable recoding, type validation, and naming enforcement

  All scripts are version-controlled and documented, allowing for reproducibility and scalable future updates.

7.2. Harmonization Procedures

To merge firm-level and country-level data:

- **Country codes** were aligned using ISO 2-letter (iso2c) and 3-letter (iso3c) standards
- **Date and year formats** were standardized to Gregorian calendar years (2010–2024)
- **Region and capital fields** were harmonized to enable regional analyses and geographic mapping

Where necessary, country names were manually reconciled across sources to prevent merge conflicts (e.g., “United States” vs. “USA” vs. “US”).

7.3. Handling of Missing Data

A unified NA-handling policy was applied:

- All missing values were encoded as NA, following R and Python conventions
- At the **firm level**, missingness generally reflects lack of public disclosure (e.g., non-reported board data)
- At the **country level**, gaps originate from source-level availability, particularly for smaller countries or earlier years
- Variables with systemic missingness are flagged in the variable dictionaries

This approach supports listwise deletion, imputation, or sensitivity checks depending on the research design.

7.4. Validation and Outlier Screening

To ensure data integrity, several validation routines were implemented:

- Detection of impossible values (e.g., negative employees or revenues)
- Time-series continuity checks (e.g., abrupt changes >10x year-over-year)
- Logical cross-checks between data sources (e.g., asset figures from Eikon vs. Forbes)

Outliers were **retained but flagged**, leaving final inclusion to the discretion of the user.

7.5. Naming Conventions and Coding Standards

All variables follow a consistent snake_case naming convention to support:

- **Transparency** in variable structure
- **Automation** in data processing
- **Compatibility** with statistical software

Variable names encode:

- Year of observation (e.g., y2024)
- Unit of measurement (e.g., millions_usd, percent, count)
- Data source (e.g., eikon, forbes, wdi)

Examples:

- revenue_y2020_millions_usd_eikon
- board_size_y2023_count_eikon
- voice_accountability_y2016_score_wgi

All coding logic is documented in the accompanying data dictionaries.

Together, these procedures ensured that the final dataset is:

- **Panel-ready**
- **Longitudinally consistent**
- **Cross-source validated**
- **Fully documented for academic use**

The combination of automated scripts and manual quality checks offers both scalability and precision for future dataset updates and extensions.

8. Data Structure and Documentation

This section outlines the format, architecture, and documentation of the final dataset. Particular attention was given to ensuring **compatibility**, **transparency**, and **usability** across a broad range of empirical applications and analytical platforms.

8.1. File Formats and Encoding

All dataset components are provided in open, interoperable formats:

- **CSV format** for both firm-level and country-level datasets
- **UTF-8 encoding** to ensure compatibility across operating systems, software environments (R, Python, Stata), and international characters
- **File structure:** Each dataset is delivered in **long format**, optimized for panel and time-series analysis

8.2. Two-Level Panel Structure

The dataset consists of two analytically connected layers:

- **Firm-Level Panel**
  - 560 multinational manufacturing firms
  - Observations span 15 years (2010–2024)
  - Variables include: financials, corporate governance, internationalization, board attributes, and identifiers
  - Unique identifiers: firm_id, year, iso3c
- **Country-Level Panel**
  - Approximately 85 countries (coverage varies by year and variable)
  - 75+ macro-level indicators (institutional, economic, political, cultural, environmental)
  - Unique identifiers: iso3c, year

  The two panels can be seamlessly merged using iso3c (country code) and year, enabling hierarchical or multi-level analysis.

8.3. Metadata and Core Identifiers

Each dataset includes key metadata fields for sorting, merging, and grouping:

- country_name — full name of the country
- iso2c and iso3c — ISO-standard country codes

- region_name — macro-regional classification (e.g., Europe, Asia)
- capital_city — provided for country-level dataset only
- firm_id — anonymized unique identifier for each firm

These identifiers support geographic filtering, cross-dataset alignment, and subgroup analysis.

8.4. Variable Dictionaries

Each variable in both datasets is fully documented in **accompanying CSV dictionaries**, which contain the following fields:

- variable_name — standardized name in snake_case
- label — human-readable description
- unit — unit of measurement (e.g., millions USD, %)
- category — thematic grouping (e.g., financial, institutional, cultural)
- source — original source institution (e.g., Eikon, WDI, Hofstede)
- notes — definition, coding logic, special cases

The dictionaries are structured to enable both human review and programmatic parsing. They are essential for transparent variable selection, replication studies, and data transformations.

8.5. Treatment of Missing Values

- All missing values are encoded as NA
- Missingness is flagged in dictionaries where relevant (e.g., variables unavailable for certain countries or years)
- Users are encouraged to handle missing data explicitly depending on the analytical method (e.g., multiple imputation, listwise deletion, fixed-effects modeling)

The dataset's structure combines:

- **Cross-sectional depth** (detailed firm-level metrics)
- **Longitudinal continuity** (15-year time frame)
- **Cross-national variation** (85+ countries)

This architecture makes the dataset highly suitable for:

- **Multilevel modeling**
- **Fixed-effects and random-effects estimation**
- **Time-series cross-sectional analysis**
- **Comparative institutional research**

Full technical documentation and metadata are distributed alongside the datasets, ensuring academic rigor and reproducibility.

9. Limitations

While the dataset was developed with a strong emphasis on coverage, transparency, and methodological rigor, several **inherent limitations** should be acknowledged by users. These relate to data availability, granularity, and conceptual interpretation, and should be taken into account when designing empirical models and interpreting results.

9.1. Missing Values and Uneven Coverage

Despite extensive data collection efforts, some variables exhibit **missingness**, particularly in earlier years or for firms from countries with limited disclosure requirements.

- At the **firm level**, governance-related indicators—such as board compensation or CEO tenure—are not consistently reported across jurisdictions.
- **Financial data** (e.g., revenue, ROA, employee counts) may be partially missing for certain firms in earlier time periods (2010–2012), especially from emerging markets.
- At the **country level**, gaps exist in several institutional and macroeconomic indicators, particularly for **low-income or small states**. For example, certain V-Dem or EPI indicators are unavailable for parts of the panel.

These missing values are systematically marked as NA and fully documented in the data dictionaries. Users are encouraged to consider robust strategies such as **listwise deletion**, **multiple imputation**, or **sensitivity analyses** depending on the empirical framework.

9.2. Lack of Subnational Disaggregation

All country-level indicators are provided at the **national level**, which imposes several analytical constraints:

- The dataset does not capture **within-country heterogeneity**, such as institutional variation across federal states, provinces, or cities.
- This is particularly relevant in large or decentralized countries like **China, India, Brazil, or the United States**, where firm exposure to institutional environments can vary significantly across subnational jurisdictions.
- Consequently, any analysis that assumes uniform country conditions should be interpreted with appropriate caution.

9.3. Interpretive Nature of Composite Indices

Several variables—especially in the **cultural and political** domains—are based on **composite indices or subjective surveys**, which, while widely used in academic literature, come with conceptual and interpretive limitations:

- Hofstede's dimensions (e.g., power distance, individualism) are based on legacy surveys with limited periodic updates.
- World Values Survey and Freedom House scores reflect **subjective assessments** that may vary depending on methodology or framing effects.
- Political regime classifications from V-Dem involve **expert-coded data**, which, despite methodological robustness, are subject to changes in interpretation over time.

Researchers should be aware of these limitations when using such variables for causal inference or time-series analysis. Whenever possible, **triangulation** with multiple sources or robustness checks is recommended.

In sum, while the dataset is suitable for a wide range of empirical applications, its limitations should guide responsible usage:

- **Not all variables are available for all years or countries**
- **Country-level data lacks subnational granularity**
- **Some indicators rely on interpretive or subjective constructs**

These constraints are common in large-scale international datasets and are mitigated through careful documentation, variable flagging, and recommended best practices for analysis.

10. Access, Licensing, and Use Conditions

This section outlines the legal and institutional framework governing the use, sharing, and redistribution of the dataset. It ensures transparency regarding licensing, academic usage rights, and contact protocols for support or clarification.

10.1. Access and Licensing

The dataset has been compiled using both **publicly accessible sources** and **restricted commercial databases** available under academic licenses held by **Mälardalen University**, specifically by the *School of Business, Society and Engineering* (Division of Marketing and Strategy).

- **Refinitiv Eikon** and **Orbis Global (Bureau van Dijk)** were accessed through institutional subscriptions for academic use only.
- Country-level indicators were obtained from open-access and academically endorsed sources such as the **World Bank**, **V-Dem**, **WVS**, **EPI**, **Freedom House**, and others.

The dataset, as a compiled and cleaned academic product, is intended for **internal research, education, and non-commercial analysis** within the university community or approved academic collaborations.

10.2. Permitted Use

Users are permitted to:

- Apply the dataset for academic research, teaching, and thesis projects;
- Reference or summarize findings in academic papers, presentations, and dissertations;
- Conduct replication or robustness checks using the dataset's public components.

Users are **not permitted to**:

- Distribute the raw dataset externally without written permission;
- Use the dataset for commercial purposes;
- Extract or redistribute proprietary fields sourced from commercial databases (e.g., firm-level financials from Eikon or Orbis) outside the licensed institution.

All users must adhere to the terms of data access set forth by the respective providers and the university's internal data governance policies.

10.3. Citation Guidelines

When using the dataset in publications or academic work, please cite it as follows:

> ***Morgulis-Yakushev, S., Kang, O., Yildiz, E., Safari, A., & Melén Hånell, S.*** *(2025). Global MNC Panel Dataset (2010–2024): Firm–Country Data on Strategy, Institutions, and Contexts. Mälardalen University & Uppsala University. Version 1.0.*
>
> Additionally, users are encouraged to cite the technical appendix and relevant data sources as documented in the variable dictionaries.

10.4. Contact for Support

For questions related to:

- Data structure or content
- Interpretation of variables
- Collaboration or extended access requests

Please contact:

Dr. Olivia Kang
Project Lead
School of Business, Society and Engineering
Mälardalen University
Email: *olivia.kang@mdu.se*

11. Version History and Future Plans

This section provides a structured overview of the dataset's versioning system, its current status, and planned enhancements. It is intended to ensure long-term usability, transparent documentation of changes, and consistency across future updates.

11.1. Versioning Framework

The dataset follows a **semantic versioning system** (Major.Minor.Revision), where:

- **Major** versions reflect significant structural or conceptual changes;
- **Minor** versions introduce new variables or sources while maintaining core compatibility;
- **Revisions** fix errors or update metadata without altering the data structure.

Each version is accompanied by:

- A detailed changelog,
- Updated documentation (data dictionaries, technical notes),
- A reproducible snapshot of the dataset.

11.2. Current Release — v1.0.0 (August 2025)

The first public academic release includes:

- 560 publicly listed multinational manufacturing firms (2010–2024);
- A harmonized country-level panel covering ~85 countries and 75+ macro, institutional, cultural, and environmental indicators;

- A two-tier panel structure (firm-year and country-year) with full metadata;
- Complete documentation: technical appendix, two data dictionaries, summary statistics, and licensing notes;
- Dataset validated and archived for internal academic use at **Mälardalen University** under institutional licensing agreements.

11.3. Planned Updates — v1.1.0 (Tentative)

Expected extensions in the next version include:

- Addition of data for fiscal year 2025 (pending availability);
- Expansion of firm-level variables (e.g., institutional ownership, ESG indicators, CEO tenure);
- Introduction of sector-level aggregates for benchmarking;
- Enhanced coverage of institutional indicators for underrepresented countries;
- Incremental integration of subnational data (starting with federated countries such as the U.S., Brazil, India);
- Partial automation of the update workflow using reproducible R/Python pipelines.

11.4. Long-Term Roadmap

Subject to data availability and institutional support, future iterations may include:

- **v2.0.0** — Structural redesigns (e.g., modular architecture for ESG, innovation, or trade data);
- **Annual updates** extending time-series coverage beyond 2024;
- Web-based tools for data visualization and interactive querying;
- Public metadata repository for variable definitions and code mappings;
- Formal registration of the dataset in an academic data repository with DOI assignment.

The update strategy is aligned with the broader vision of the **School of Business, Society and Engineering** to support reproducible, data-driven research in international business, institutional theory, and comparative economic systems.

Appendix 1: Summary Statistics Tables

This appendix presents key descriptive statistics that characterize the firm-level and country-level datasets. These tables are intended to help users quickly understand the scale, distribution, and completeness of core variables across time and geography.

A1. Firm-Level Sample Composition (n = 560 firms)

Category	Value
Observed Firms	560
Observed Years	2010–2024 (15 years)
Total Firm-Year Observations	8,400
Industries	30+ (manufacturing sectors)
Countries	39
Regions	5 (Asia, Americas, Europe, Africa, Oceania)

A2. Top Countries by Firm Count

Country	Number of Firms	Share (%)
United States	171	30.5%
China	89	15.9%
Japan	64	11.4%
Germany	22	3.9%
South Korea	21	3.8%
United Kingdom	19	3.4%
Switzerland	18	3.2%
France	17	3.0%
India	17	3.0%
Sweden	14	2.5%
Others (28 countries)	108	19.3%

A3. Sectoral Composition (Top Industries)

Industry Category	Number of Firms	Share (%)
Consumer Durables	74	13.2%
Food, Drink & Tobacco	73	13.0%
Capital Goods	67	12.0%
Materials	49	8.8%
Technology Hardware & Equipment	49	8.8%
Drugs & Biotechnology	45	8.0%
Chemicals	44	7.9%
Oil & Gas Operations	41	7.3%
Health Care Equipment & Services	31	5.5%
Semiconductors	31	5.5%
Other	56	10.0%

A4. Summary of Key Firm-Level Variables (2024)

Variable	Mean	Median	Min	Max	Missing (%)
Revenue (USD mn)	30889	15824	0	480446	1.1
Net Income (USD mn)	2769	1093	1	106123	0.5
Assets (USD mn)	44274	22080	920	655249	0.4
Market Cap (USD mn)	64645	23063	223	3766500	0.2
Employees	55624	32248	48	968872	16.2
ROA (%)	0.1	0.1	0	0.7	2.9
International Assets (%)	0.2	0.1	0	0.9	48.9

A5. Distribution of Revenue Data Availability (2010–2024)

Years with Revenue Data	Firms	Share (%)
15 (complete)	423	75.5%
13–14	45	8.0%
10–12	57	10.2%
< 10	35	6.3%

A6. Country-Level Coverage

Coverage Type	Value
Countries	~85
Years	2010–2024 (up to 15 years)
Total Country-Year Observations	>1,000
Total Country-Level Variables	75
Missingness (average)	~5–10% depending on source

Appendix 2: Company Profiles by Region

This appendix offers a curated selection of brief company profiles, organized by global region. Each profile includes key indicators from the dataset, offering users a glimpse into the diversity of firm characteristics and strategic orientations represented in the panel. The purpose is to provide tangible illustrations of the dataset's content and encourage regionally contextualized analyses.

Asia

Samsung Electronics (South Korea)

• Industry: Technology Hardware & Equipment
• Founded: 1938
• Revenue (2024): $206.68B
• Employees: 267,860
• International Sales: 86.8%
• International Assets: 10.6%
• Board Size: 11
• Average Board Tenure: 2.8 years
• CEO (2025): Young-Hyun Jun

Toyota Motor (Japan)

• Industry: Consumer Durables
• Founded: 1937
• Revenue (2024): $329.22B
• Employees: 380,793
• International Sales: 77.7%
• International Assets: 62.4%
• Board Size: 10
• Average Board Tenure: 5.9 years
• CEO (2025): Koji Sato

Americas

Apple (United States)

- Industry: Technology Hardware & Equipment
- Founded: 1976
- Revenue (2024): $391.04B
- Employees: 164,000
- International Sales: 63.6%
- International Assets: NA%
- Board Size: 9
- Average Board Tenure: 11.8 years
- CEO (2025): Tim Cook

Johnson & Johnson (United States)

- Industry: Drugs & Biotechnology
- Founded: 1886
- Revenue (2024): $88.82B
- Employees: 144,000
- International Sales: 43.4%
- International Assets: 16.9%
- Board Size: 13
- Average Board Tenure: 6.5 years
- CEO (2025): Joaquin Duato

Europe

Nestlé (Switzerland)

- Industry: Food, Drink & Tobacco
- Founded: 1866
- Revenue (2024): $101.17B
- Employees: 277,000
- International Sales: 46.2%
- International Assets: 47.4%
- Board Size: 13

- Average Board Tenure: 5.8 years
- CEO (2025): Laurent Freixe

Siemens (Germany)

- Industry: Capital Goods
- Founded: 1847
- Revenue (2024): $79.95B
- Employees: 327,000
- International Sales: 82.1%
- International Assets: 30.8%
- Board Size: 20
- Average Board Tenure: 6.9 years
- CEO (2025): Roland Busch

Africa

Sasol (South Africa)

- Industry: Chemicals
- Founded: 1950
- Revenue (2024): $15.46B
- Employees: 28,141
- International Sales: 49.9%
- International Assets: 35.7%
- Board Size: 12
- Average Board Tenure: 5.2 years
- CEO (2025): Simon Baloyi

Oceania

CSL (Australia)

- Industry: Drugs & Biotechnology
- Founded: 1916
- Revenue (2024): $14.80B
- Employees: 32,000

- International Sales: 93.9%
- International Assets: 41.5%
- Board Size: 9
- Average Board Tenure: 8.3 years
- CEO (2025): Paul McKenzie

Notes

- All financial values are reported in USD (millions).
- Data are based on 2024 values unless otherwise stated.
- Companies were selected to reflect sectoral diversity and global representation.

Appendix 3. Sources and References

This appendix lists all major data sources used in the construction of the dataset, including commercial and public-access databases. Each entry includes:

- Full citation
- Access information (URL or DOI)
- Date of last access
- Suggested citation format for academic use

Firm-Level Data Sources

1. Refinitiv Eikon

Description: Financial and corporate governance data for publicly listed firms
Provider: LSEG (London Stock Exchange Group)
Access: Institutional subscription via Mälardalen University
Last accessed: July 15, 2025
Citation format:

> Refinitiv Eikon. Financial and Governance Data. London Stock Exchange Group. Accessed via institutional subscription, July 2025.

2. Orbis Global (Bureau van Dijk)

Description: Ownership structure, company identifiers, and cross-verification of financials
Provider: Bureau van Dijk (Moody's Analytics)
Access: Institutional subscription via Mälardalen University
Last accessed: July 17, 2025
Citation format:

> Bureau van Dijk. Orbis Global Database. Moody's Analytics. Accessed via institutional subscription, July 2025.

3. Forbes Global 2000 (2024 Edition)

Description: Firm rankings, revenue, profit, assets, market value

Publisher: Forbes Media

URL: https://www.forbes.com/global2000

Last accessed: July 10, 2025

Citation format:

> Forbes Media. (2024). *Global 2000 List of the World's Largest Public Companies*. https://www.forbes.com/global2000

Country-Level Data Sources

4. World Development Indicators (WDI)

Provider: World Bank

URL: https://databank.worldbank.org/source/world-development-indicators

Last accessed: June 20, 2025

Citation format:

> World Bank. (2025). *World Development Indicators*. https://databank.worldbank.org/source/world-development-indicators

5. Worldwide Governance Indicators (WGI)

Provider: World Bank

URL: https://info.worldbank.org/governance/wgi/

Last accessed: June 20, 2025

Citation format:

> Kaufmann, D., Kraay, A., & Mastruzzi, M. (2024). *The Worldwide Governance Indicators*. https://info.worldbank.org/governance/wgi/

6. Varieties of Democracy (V-Dem)

Provider: V-Dem Institute, University of Gothenburg
DOI: https://doi.org/10.23696/vdemds25
Last accessed: June 25, 2025
Citation format:

> Coppedge, M., Gerring, J., Knutsen, C. H., et al. (2025). *V-Dem [Country–Year Dataset] v13*. V-Dem Institute. https://doi.org/10.23696/vdemds25

7. Hofstede Insights

Provider: Hofstede Insights and Hofstede Centre
URL: https://www.hofstede-insights.com/product/compare-countries/
Last accessed: June 18, 2025
Citation format:

> Hofstede Insights. (2025). *National Culture Comparison Tool*. https://www.hofstede-insights.com

8. World Values Survey (WVS)

Provider: World Values Survey Association
URL: https://www.worldvaluessurvey.org
Last accessed: June 28, 2025
Citation format:

> World Values Survey Association. (2024). *World Values Survey Wave 7 (2017–2022)*. https://www.worldvaluessurvey.org

9. Environmental Performance Index (EPI)

Providers: Yale University & Columbia University

URL: https://epi.yale.edu
Last accessed: June 21, 2025
Citation format:

Wendling, Z. A., Emerson, J. W., Esty, D. C., et al. (2024). *2024 Environmental Performance Index*. Yale Center for Environmental Law & Policy. https://epi.yale.edu

10. Freedom House – Freedom in the World Index

Provider: Freedom House
URL: https://freedomhouse.org
Last accessed: June 27, 2025
Citation format:

Freedom House. (2024). *Freedom in the World 2024*. https://freedomhouse.org

11. Global Health Security Index (GHSI)

Providers: NTI, Johns Hopkins Center for Health Security, Economist Impact
URL: https://www.ghsindex.org
Last accessed: June 22, 2025
Citation format:

Nuclear Threat Initiative (NTI), Johns Hopkins Center for Health Security, Economist Impact. (2024). *Global Health Security Index*. https://www.ghsindex.org

12. World Economic Forum – Ease of Doing Business (Historical)

URL: https://www.weforum.org
Last accessed: June 24, 2025
Citation format:

World Economic Forum. (2023). *Global Competitiveness Reports* [historical ease of doing business]. https://www.weforum.org

13. World Intellectual Property Organization (WIPO)

URL: https://www.wipo.int/statistics/en/

Last accessed: June 26, 2025

Citation format:

World Intellectual Property Organization. (2025). *WIPO IP Statistics Data Center*. https://www.wipo.int/statistics/en/

Appendix 4. Glossary of Concepts

Term	Definition	Usage in Dataset	Source
Board Diversity	Gender or demographic composition of a corporate board	Included as firm-level governance indicator	Carter et al. (2003); Adams & Ferreira (2009)
Comparative Capitalism	Theory linking national institutional configurations to economic outcomes and firm behavior	Justifies inclusion of institutional clusters and regulatory indicators	Hall & Soskice (2001)
Cultural Dimensions	National value systems (e.g., individualism, power distance, masculinity)	Hofstede's dimensions used as cultural variables	Hofstede (2001)
Democracy Index	Level of political rights, civil liberties, and institutional integrity	V-Dem and Freedom House ratings used in regressions	V-Dem; Freedom House
Ease of Doing Business	Index of regulatory efficiency and administrative burden in the private sector	Country-level proxy for institutional friction	World Bank (Doing Business Project)
Environmental Performance	Index of environmental health and ecosystem vitality	EPI used as proxy for ecological institutions	Yale EPI
FDI	Foreign direct investment — cross-border investment in physical or controlling interests	Included as macroeconomic control	UNCTAD; World Bank

Term	Definition	Usage in Dataset	Source
Fixed Effects	Panel data estimation method controlling for unit-level unobserved heterogeneity	Enabled by dataset structure (firm × year)	Wooldridge (2010)
Gini Index	Income inequality measure from 0 (equality) to 1 (inequality)	Drawn from World Bank WDI	World Bank (WDI)
Institutional Environment	Formal and informal rules shaping market behavior	Captured via rule of law, corruption, regulatory quality	North (1990)
Internationalization	Expansion of firms beyond borders through trade, investment, or subsidiaries	Measured via % of international sales and assets	Johanson & Vahlne (1977)
ISO Codes	International country codes — ISO2 (2-letter), ISO3 (3-letter)	Used to merge firm- and country-level datasets	International Organization for Standardization
Liability of Foreignness	Disadvantages faced by foreign firms due to unfamiliarity, legitimacy gaps, or discrimination	Motivates inclusion of internationalizatio n and governance indicators	Zaheer (1995)
NA (Missing Value)	Indicator used in R, Python, and statistical tools to denote missing or unavailable data	All missing values in dataset coded as NA	R / Python Conventions
Panel Dataset	Dataset with repeated observations on the same units over time	Structure supports longitudinal and multilevel models	Baltagi (2005); Hsiao (2014)

Term	Definition	Usage in Dataset	Source
ROA	Net income divided by total assets	Financial performance metric	Palepu & Healy (2008)
ROE	Net income divided by shareholders' equity	Used to assess profitability and capital efficiency	Palepu & Healy (2008)
Rule of Law	Extent to which laws are enforced fairly and predictably	Core institutional variable used in multiple analyses	Kaufmann et al. (2009)
Snake Case Naming	Variable naming convention using lowercase and underscores (e.g., revenue_y2023_millions_usd_eikon)	Applied across all variable names for consistency and automation	Tidyverse / Data Science Practice
Transaction Costs	Costs of organizing economic activity due to uncertainty, enforcement, and information asymmetries	Underpins rationale for institutional quality variables	Williamson (1985)
Voice and Accountability	Degree of citizen participation, freedom of expression, and press freedom	Included in governance dimensions from WGI	World Governance Indicators (WGI)

References

- Adams, R. B., & Ferreira, D. (2009). *Women in the boardroom and their impact on governance and performance*. Journal of Financial Economics, 94(2), 291–309.
- Baltagi, B. H. (2005). *Econometric Analysis of Panel Data* (3rd ed.). John Wiley & Sons.

- Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). *Corporate governance, board diversity, and firm value*. Financial Review, 38(1), 33–53.
- Hall, P. A., & Soskice, D. (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford University Press.
- Healy, P. M., & Palepu, K. G. (2008). *Business Analysis and Valuation* (4th ed.). Cengage Learning.
- Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations* (2nd ed.). Sage Publications.
- Hsiao, C. (2014). *Analysis of Panel Data* (3rd ed.). Cambridge University Press.
- Johanson, J., & Vahlne, J.-E. (1977). *The internationalization process of the firm – A model of knowledge development and increasing foreign market commitments*. Journal of International Business Studies, 8(1), 23–32.
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2009). *Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996–2008*. World Bank Policy Research Working Paper No. 4978.
- North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press.
- Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. Free Press.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). MIT Press.
- World Bank. (Various Years). *World Development Indicators (WDI)*. https://databank.worldbank.org/
- Yale Center for Environmental Law and Policy. (Various Years). *Environmental Performance Index (EPI)*. https://epi.yale.edu/
- Zaheer, S. (1995). *Overcoming the liability of foreignness*. Academy of Management Journal, 38(2), 341–363.

Appendix 5. Data Use and Citation Policy

To promote responsible use, ensure academic integrity, and facilitate proper attribution, this dataset is governed by the following citation and usage policy.

Citation Guidelines

When using this dataset in any publication, academic presentation, or research output, please cite it as follows:

> **Morgulis-Yakushev, S., Kang, O., Yildiz, E., Safari, A., & Melén Hånell, S.** (2025). *Global MNC Panel Dataset (2010–2024): Firm–Country Data on Strategy, Institutions, and Contexts.* Mälardalen University & Uppsala University. Version 1.0.
>
> You may also reference the technical documentation as:
>
> **Morgulis-Yakushev, S., Kang, O., Yildiz, E., Safari, A., & Melén Hånell, S.** (2025). Technical Appendix: Data Construction and Sources for the Global MNC Panel Dataset. Mälardalen University.
>
> If you use either the firm-level or country-level component independently, please indicate this clearly in your methodology section and still cite the full source.

Conditions of Use

- **Academic use only:**

The dataset is intended strictly for academic research, teaching, and scholarly inquiry. Commercial use or redistribution without consent is not permitted.

- **Licensed data sources:**

Data from Refinitiv Eikon and Orbis (Bureau van Dijk) was accessed through institutional academic licenses held by Mälardalen University. Use must comply with those license terms.

- **Derivative work disclaimer:**

Users creating derived variables or subpanels are expected to cite the original version and clearly document their modifications.

Replication and Reproducibility

- **Replication packages encouraged:**

Authors are strongly encouraged to provide replication files (e.g., code, models, or subsamples) when publishing findings based on the dataset.

- **Attribution in collaborative projects:**

In institutional projects or multi-author publications, proper attribution must include the full author list and institutional affiliation.

- **Version control:**

Each version of the dataset is archived with a changelog and fixed schema. Deprecated variables are retained to preserve comparability.

Contact Information

For access requests, technical queries, or research collaboration, please contact:

Principal Investigator and Contact Lead
Dr. Sergey Morgulis-Yakushev
Uppsala University
sergey.morgulis-yakushev@fek.uu.se

Project Lead
Dr. Olivia Kang
School of Business, Society and Engineering
Mälardalen University
olivia.kang@mdu.se