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

Integrated Panel Data on Multinational Corporations and Country-Level Environments

TECHNICAL APPENDIX
DATA CONSTRUCTION AND SOURCES

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. Overview

This technical appendix documents the construction, sources, and structure of a **novel empirical dataset** developed to support rigorous academic research on multinational corporations (MNCs) and the institutional environments in which they operate.

The primary objective of the project is to construct a **panel-ready, multi-level dataset** that systematically combines firm-level indicators with institutional, economic, cultural, political, and environmental characteristics at the country level.

This dataset is among the **few international resources** that **integrate firm-level and country-level variables** in a unified analytical framework. Its architecture allows scholars to investigate how contextual environments shape corporate strategies, structures, and outcomes across global markets.

The dataset was designed with a strong emphasis on:

- **reproducibility**,
- **methodological transparency**,
- and **cross-disciplinary usability**, making it suitable for peer-reviewed publications, doctoral research, and graduate-level coursework in international business, political economy, and institutional analysis.

1.1. Dataset Scope

The dataset spans the period **2010 to 2024** and comprises two primary levels:

Firm-level panel:

Includes **560 publicly listed multinational manufacturing firms**, selected from the *Forbes Global 2000 (2024)* and enriched with granular financial and governance data from Orbis Global (Bureau van Dijk), *Refinitiv Eikon*, including revenues, profits, assets, board characteristics, employment figures, and internationalization metrics.

Country-level panel:

Comprises **more than 75 annual indicators** across **80+ countries** representing all major global regions. Variables capture the institutional, economic, cultural, political, and environmental context in which firms operate. Data were sourced from the *World*

Bank, *V-Dem*, *Hofstede Insights*, *World Values Survey*, *Environmental Performance Index*, and others.

1.2. Analytical Use and Research Applications

The integrated structure of the dataset enables a wide range of empirical strategies—including fixed-effects modeling, multilevel analysis, and institutional interaction frameworks. Full variable standardization and documentation enhance its flexibility for use across various academic disciplines.

Example research questions supported by the dataset include:

- **To what extent do institutional conditions influence corporate performance and governance?**
- **How do firms adjust their internationalization strategies in response to political instability?**
- **What is the impact of national culture and values on board structure and decision-making?**

1.3. Future Updates and Extensions

The structure of the dataset is designed to support **future updates and expansions**, including:

- Annual additions of new data (e.g., 2025 and beyond),
- Extension to **additional countries**, industries, or sectors,
- Potential integration of **subnational-level indicators** (e.g., regional data for federal states).

2. Sample Design and Structure

This section outlines the construction, scope, and characteristics of the firm-level sample used in the dataset. The objective was to build a **representative, internationally diversified, and analytically robust** panel of publicly listed

multinational corporations (MNCs) in the manufacturing sector, suitable for studying firm–institution interactions in a global context.

2.1. Sampling Frame and Selection Principles

The initial sampling frame was derived from the **Forbes Global 2000 (2024 edition)** — a widely recognized and academically validated ranking of the world's largest publicly traded companies, based on four indicators: **revenues, profits, assets, and market capitalization**.

This list offers **global coverage**, **transparent inclusion criteria**, and **cross-country comparability**, making it highly suitable for research in international business, corporate strategy, and institutional theory.

> *Note:* The sample covers a significant share of global MNC activity in the manufacturing sector by total revenue and market capitalization.
>
> Given the absence of comprehensive global business registries, a **purposive selection strategy** was used instead of random sampling to ensure analytical comparability and data richness.

2.2. Data Sources: Sampling vs. Variables

It is important to distinguish between the sources used for **sampling** and those used for **data collection**:

- The **Forbes Global 2000** served as the **sampling basis** and provided baseline information such as firm name, country of incorporation, industry classification, and market capitalization.
- **Refinitiv Eikon** and **Orbis Global (Bureau van Dijk)** were used to **extract firm-level variables**, including financials, governance, employment, ownership structure, and other attributes.

2.3. Sectoral Restrictions

To ensure conceptual alignment and comparability, firms primarily operating in the following sectors were excluded:

- Banking and financial services
- Insurance
- Telecommunications
- Energy and extractive industries
- Transportation, construction, retail, and other service sectors

These industries are typically characterized by high regional embeddedness, unique regulatory environments, and internationalization strategies that differ substantially from those in manufacturing.

2.4. Manufacturing Firm Focus

The sample is restricted to firms classified as manufacturing under **SIC codes 2000–3999**, including:

- Food and beverage manufacturing
- Textiles and apparel
- Machinery and equipment
- Chemicals and pharmaceuticals
- Electronics and precision instruments
- Automotive and transport equipment

These industries tend to pursue capital-intensive internationalization strategies such as establishing subsidiaries or asset-heavy operations abroad, making them particularly suitable for comparative institutional analysis.

The dataset's structure aligns with the **mixed-embeddedness perspective**, linking firm-specific capabilities with country-level institutional constraints (cf. Peng, 2003; Meyer & Peng, 2005).

2.5. Geographic and Country Composition

The sample includes firms from all major world regions, ensuring strong **cross-national** and **institutional diversity**:

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 country composition spans liberal democracies (e.g., United States, Sweden, Germany), hybrid regimes (e.g., India), and authoritarian states (e.g., China, Saudi Arabia). This institutional diversity supports robust comparative research on governance, regulation, and strategy.

> The presence of both high- and low-democracy regimes enhances the external validity of institutional interaction models.

2.6. Country Distribution

Country	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

Country	Firms	Share (%)
India	17	3.0
Sweden	14	2.5
Others (28)	108	19.3
Total	560	100

2.7. Sectoral and Firm-Level Characteristics

Industry	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 industries	56	10.0
Total	560	100

2.8. Scale and Data Completeness

Revenue (2024, USD millions):

- Mean: 30,889
- Median: 15,824
- Range: 0 to 480,446

- Missing: 6 firms (1.1%)

Employees (2024):

- Mean: 55,624
- Median: 32,248
- Range: 48 to 968,872
- Missing: 91 firms (16.3%)

Time Series Completeness (2010–2024):

- 75.5% of firms have complete 15-year revenue data
- 94% have at least 10 years
- Similar completeness applies to employee count and asset data

2.9. Temporal Coverage and Panel Structure

The dataset spans **2010 to 2024** and is structured as a firm-year panel. This allows for:

- Fixed-effects and random-effects models
- Difference-in-differences designs
- Time-trend and lagged analysis at the firm level

Longitudinal coverage supports the study of:

- Internationalization trajectories (foreign assets and sales)
- Firm responses to global shocks (e.g., COVID-19, trade disruptions)
- Institutional changes (e.g., regulatory reforms, regime transitions)

2.10. Transparency and Reproducibility

Only publicly listed firms with reliable and consistent disclosures on financial and governance metrics were included. This ensures:

- High source credibility (Forbes, Eikon, Orbis)
- Transparency of sampling and coverage
- Full reproducibility for peer research

All variables are documented in structured data dictionaries, with units, sources, time availability, and coding logic clearly specified.

2.11. Dataset Uniqueness

Unlike most datasets that focus on either financial metrics or institutional conditions in isolation, this dataset combines:

- **Firm-level data**: financials, structure, and governance
- **Internationalization measures**: foreign sales and assets as a share of total
- **Country-level variables**: political, legal, economic, cultural, and environmental indicators

The result is one of the few **globally comparable firm–country panels** designed for institutional and strategy research in international business.

2.12. Summary

Taken together, the sample achieves a careful balance between:

- **Breadth** — global coverage across sectors and regions
- **Depth** — long-term time series and data completeness
- **Transparency** — clear criteria and structured documentation

The next sections outline the procedures used to retrieve, harmonize, and prepare the data from the selected sources for empirical use.

3. Data Sources

This project integrates two tiers of data—**firm-level** and **country-level**—to support the analysis of how multinational corporations (MNCs) interact with their institutional, economic, political, and cultural environments.

This dual-level design allows researchers to analyze not only how firm strategies vary across institutional contexts, but also how firms adapt to changing macro conditions over time. The dataset architecture explicitly supports **multilevel modeling**, including **hierarchical linear models (HLM)**, **fixed-effects regressions with country-year controls**, and **cross-level interaction terms**.

Such integration remains uncommon in empirical international business research. A review of leading journals—*Journal of International Business Studies*, *Global Strategy Journal*, *Journal of World Business*—shows that most studies rely on single-country or region-specific panels and rarely employ harmonized, globally representative firm–country datasets. This project seeks to address that gap.

3.1. Firm-Level Data

Firm-level data cover 560 publicly listed multinational manufacturing companies over the period 2010–2024. Variables include financials, employment, corporate governance, and internationalization. Data were retrieved from licensed sources and harmonized into a clean panel structure.

Primary Sources:

- **Refinitiv Eikon** – primary source of structured financial and corporate data
- **Orbis Global (Bureau van Dijk)** – used to complement governance and ownership variables and validate firm profiles
- **Forbes Global 2000 (2024)** – used as the sampling frame and a validation reference

Scope:

- 560 firms
- 15-year panel (2010–2024)
- Only publicly traded manufacturing companies

Variable Domains:

- **Financial indicators**: revenue, net income, total assets, market capitalization (USD millions)
- **Performance ratios**: return on assets (ROA), return on equity (ROE)
- **Corporate governance**: board size, average tenure, gender diversity, board compensation
- **Workforce**: number of employees (annually)
- **Ownership and structure**: year of incorporation, NAICS/SIC industry codes, country of registration

- **Internationalization**:
  - *Share of international sales* (% of total revenue generated abroad)
  - *Share of international assets* (% of total assets held abroad)

Processing and Cleaning:

All variables were cleaned, standardized, and formatted for panel analysis. Inconsistencies were addressed through cross-referencing across data sources. Variables were named using a consistent snake_case convention and checked for internal consistency over time.

3.2. Country-Level Data

Country-level data include over 75 indicators describing institutional, economic, political, cultural, and environmental conditions across 85 countries. These data were integrated into a 15-year panel (2010–2024) and aligned with firm-level observations.

Primary Sources:

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

Scope:

- 85 countries
- 15-year time frame (2010–2024)
- Coverage varies by indicator; missing data handled via interpolation or conservative imputation when justified

Variable Domains:

- **Institutional quality**: rule of law, regulatory quality, control of corruption, voice and accountability
- **Political regime types**: electoral, liberal, deliberative, participatory, and egalitarian democracy (V-Dem); Freedom House democracy index
- **Cultural values**: Hofstede dimensions (e.g., individualism, power distance), WVS-based indicators on trust, religiosity, tolerance
- **Macroeconomic environment**: GDP per capita, GDP growth, FDI inflows, inflation, Gini index, labor participation, tax revenue
- **Infrastructure and development**: internet access, urbanization rate, public health and education spending, state capacity
- **Environmental and health resilience**: CO_2 emissions per capita, renewable energy share, EPI scores, health security (GHSI)
- **Innovation and competitiveness**: R&D intensity, patent filings, digital infrastructure, entrepreneurship indicators (e.g., new business density)

Variable Selection Logic:

Indicators were chosen not only for availability and temporal coverage but also for **theoretical relevance** to international business and institutional theory. They capture key constructs such as **legitimacy**, **liability of foreignness**, **institutional voids**, and **transaction costs**, enabling multidimensional analysis of firm–environment interactions.

Integration with Firm Data:

All country-level variables were matched to firm-level data using standardized ISO2 and ISO3 codes and the year variable. The resulting structure supports cross-sectional and longitudinal multi-level modeling.

4. Data Collection Procedures

The data collection process was conducted in a **multi-stage, multi-level** fashion, encompassing both firm-level and country-level indicators. The core principles guiding the process were **representativeness, reproducibility, transparency, and**

consistency across years and sources. All data span the period from **2010 to 2024**, enabling both cross-sectional and longitudinal (panel) research designs.

The procedures were carefully documented to ensure **full reproducibility** by other researchers or research teams. The 15-year horizon was chosen to capture **long-term institutional transformations** and the **effects of global disruptions**, such as the COVID-19 pandemic, trade tensions, and geopolitical shifts.

4.1. Firm-Level Data

Firm-level data were retrieved from multiple licensed sources using a combination of manual validation and automated scripting. The focus was placed on financials, governance structures, employment metrics, and internationalization indicators.

Sources:

- **Refinitiv Eikon** – primary source of structured financial, market, and governance data
- **Orbis Global (Bureau van Dijk)** – used to verify ownership structure and complement board-related and employment variables
- **Forbes Global 2000 (2024 edition)** – used to define the sampling frame and provide baseline firm attributes (e.g., sector, headquarter country, year of incorporation, CEO identity)

Key Variables:

- Financial indicators: revenue, net income, total assets, market capitalization
- Profitability: return on assets (ROA), return on equity (ROE)
- Corporate governance: board size, average tenure, gender diversity, board compensation
- Workforce: employee count by year
- Internationalization:
  - *Share of international sales* (% of total revenue generated abroad)
  - *Share of international assets* (% of total assets held abroad)

Licensing Note:

All commercial databases were accessed via **institutional academic licenses held by Mälardalen University**, in accordance with terms of use for scholarly research.

4.2. Country-Level Data

Country-level data were sourced from internationally recognized databases widely used in institutional, economic, political, and cultural research. These indicators provide the macro-level context in which multinational firms operate.

Primary Sources:

- **World Bank (WDI)** – macroeconomic, demographic, and public sector indicators
- **V-Dem Institute** – political regime types and institutional quality
- **Hofstede Insights** – cultural value dimensions
- **Environmental Performance Index (EPI)** – environmental policy and sustainability
- **World Values Survey (WVS)**, **Freedom House**, **GHSI**, **WIPO**, and others

Selection Criteria:

- Multi-year panel availability (preferably ≥10 years)
- International comparability across countries and time
- Academic validity and citation history
- Transparency and open access for external verification

Conceptual Relevance:

Variables were selected not only for coverage but also for **theoretical alignment with international business and institutional theory**, addressing concepts such as **institutional voids**, **legitimacy**, **transaction costs**, and **liability of foreignness**.

4.3. Standardization and Automation

4.3.1. Standardization Procedures

After import, all datasets underwent standardized cleaning:

- Variable names were harmonized using consistent snake_case formatting (e.g., revenue_y2024_millions_usd_eikon)
- Units were converted to standardized formats (e.g., millions USD, percentages)
- Country identifiers were harmonized to **ISO2** and **ISO3** codes
- Missing values were systematically encoded as NA for consistency

4.3.2. Automated Workflow

To ensure efficiency and minimize manual errors, **automated scripts** were developed in **R** and **Python**. Key packages used included:

- **R:** dplyr, tidyr, readr, data.table, janitor
- **Python:** pandas, numpy, pyreadr, openpyxl

Automated procedures included:

- Batch importing of Excel and CSV files
- Wide-to-long data reshaping
- Logical consistency checks across time and firm-level IDs
- Range validation and temporal completeness testing

5. Data Cleaning and Processing

This section documents the standardization, validation, and harmonization procedures applied to the raw data to produce a clean, panel-ready dataset. Given the integration of heterogeneous sources across levels (firm and country), sectors, and time periods, rigorous data cleaning was critical to ensure **analytical validity, internal coherence, and reproducibility**.

All data processing steps were fully scripted and transparently documented to ensure that the cleaning process can be **independently reproduced** by other researchers.

5.1. Standardization of Formats

To ensure comparability across sources and time, the following standardization procedures were applied:

- **Measurement units** were unified across numeric variables (e.g., revenue and assets were expressed in millions of USD; proportions and ratios in percentages);
- **Data tables** were converted into *long format* structures compatible with common panel data methods in R, Python, and Stata;
- **Temporal alignment** was performed to maintain consistency across year-indexed variables at both the firm and country levels.

5.2. Handling of Missing Data

A consistent approach to missing data was applied throughout the dataset:

- All missing entries were encoded as NA;
- For firm-level variables, missing values generally reflect **non-disclosure in financial reports** for a given year;
- For country-level indicators, they typically reflect **gaps in the original international datasets** (e.g., WDI, V-Dem).

This standardized encoding allows for the use of **NA-aware statistical techniques** (e.g., listwise deletion, multiple imputation) and preserves the interpretability of patterns of data absence.

5.3. Country Code Harmonization

To enable seamless merging of firm-level and country-level data, country names and codes were harmonized as follows:

- ISO2 codes (e.g., US) were used for labeling, mapping, and interface purposes;
- ISO3 codes (e.g., USA) served as the **primary merging key** across datasets.

Both **automated and manual procedures** were used to resolve inconsistencies (e.g., resolving alternate country names such as “Korea, South” and “South Korea”), ensuring full alignment across all records.

5.4. Outlier Screening and Logical Validation

A structured set of validation rules was applied to identify potential errors, inconsistencies, and outliers:

- **Logical checks**: Ensured variables such as number of employees or board members were non-negative and within plausible ranges;
- **Outlier detection**: Flagged firms showing abrupt, implausible jumps in financial metrics (e.g., >10× revenue growth in a single year);
- **Cross-source comparison**: Verified unit consistency and scale agreement across Refinitiv Eikon, Forbes, and Orbis data.

Outliers were not deleted but **flagged for further robustness checks** in empirical analysis.

5.5. Variable Naming Convention

All variables were renamed according to a **uniform snake_case convention**, encoding key metadata (year, unit, source, level) directly into the variable name. This promotes:

- Clear interpretation and semantic structure;
- Compatibility with programming tools (R, Python, Stata);
- Ease of scripting and automation.

Examples:

- board_size_y2024_count_eikon: board size in 2024, count, from Eikon
- revenue_y2020_millions_usd_eikon: total revenue in 2020, in millions USD
- int_assets_share_y2023_percent_eikon: international assets as % of total, 2023

5.6. Tools and Reproducibility

All cleaning and transformation steps were performed using **reproducible scripts** developed in:

- **R** (packages: dplyr, tidyr, janitor, data.table);
- **Python** (packages: pandas, pyjanitor, openpyxl).

These scripts are available upon request and ensure that the full data preparation pipeline—from raw input to analytical dataset—is **fully transparent, automated, and replicable**.

6. Dataset Structure

This section describes the layered structure, panel configuration, metadata schema, and accompanying documentation of the final dataset. The data architecture was intentionally designed to support multi-level empirical research, enable reproducibility, and ensure analytic flexibility across a range of software environments and research designs.

6.1. Two-Level Dataset Architecture

The final dataset is structured into two interlinked levels:

- **Firm-level dataset**: Comprises 560 multinational manufacturing firms, each described across 212 variables, including financials, governance metrics, internationalization proxies, and employment indicators.
- **Country-level dataset**: Includes approximately 75 annual indicators per country, capturing macroeconomic, institutional, political, cultural, and environmental dimensions relevant to firm-level strategy.

The two datasets are merged via harmonized country codes and calendar years, forming an integrated, empirically tractable system for cross-level analysis.

6.2. Time-Indexed Panel Configuration

Both datasets are formatted in a **long, panel-ready structure**, with:

- A dedicated year variable serving as the temporal index;
- Merging enabled via a composite key (country + year);
- Compatibility with standard statistical software and panel data libraries (e.g., plm and lme4 in R, xtreg in Stata, pandas and linearmodels in Python).

The panel format allows for use in **multilevel modeling** (e.g., hierarchical linear models), **fixed-effects estimation**, or **time-series cross-sectional designs**, facilitating dynamic and comparative institutional analysis.

Both datasets are provided in **CSV format with UTF-8 encoding** to ensure compatibility across operating systems and analytical platforms.

All missing values are represented as **NA**, in line with best practices and conventions used in R and Python-based workflows.

6.3. Standardized Metadata and Country Identifiers

Each dataset includes a consistent set of standardized metadata fields to enable filtering, merging, and geographic aggregation:

- country_name — full name of the country;
- iso2c and iso3c — standardized ISO 2-letter and 3-letter codes;
- region_name — macro-regional grouping (e.g., Europe, Asia, Americas);
- capital_city — applicable to the country-level dataset only.

These fields support integration with external data and ensure consistency in cross-national analysis.

6.4. Variable Dictionaries and Technical Files

All variables are documented in accompanying **CSV-based variable dictionaries**, which contain:

- Variable name and human-readable label;
- Source attribution;
- Unit of measurement;
- Coding logic and data type;
- Interpretation notes, valid ranges, and missing value treatment.

These documentation files are included in the project archive and are structured for automated loading into data environments.

Taken together, the dataset's structure balances **flexibility**, **transparency**, and **scalability**, enabling robust empirical research across institutional and strategic domains.

7. Limitations

This section outlines known limitations in coverage, conceptual scope, and data structure that users should be aware of when applying the dataset. While the data architecture was designed to maximize reliability and comparability, certain structural constraints and data availability issues remain.

7.1. Missing Values in Firm-Level Variables

Some variables—particularly those related to corporate governance—are not consistently available across all firms or years. For example:

- Board compensation data (board_member_compensation) are frequently missing for firms headquartered in jurisdictions with weaker disclosure mandates.
- Gaps also appear in headcount figures, ROA metrics, or executive tenure variables, especially in earlier years of the panel.

Such gaps reflect real-world discrepancies in reporting standards and practices. Researchers are advised to adopt imputation methods, listwise deletion, or robustness checks based on the requirements of their analytical design.

7.2. Coverage Gaps in Country-Level Indicators

Although country-level variables were drawn from authoritative sources (e.g., WGI, V-Dem), coverage is not uniform across countries and years. Limitations include:

- Incomplete data for smaller states, low-income economies, and the earlier part of the time series (especially prior to 2012).
- Researchers should assess data availability when building country panels or using lagged predictors in regression models.

7.3. Interpretive Nature of Cultural and Political Indices

Several country-level constructs are derived from surveys or composite indices that involve conceptual interpretation. These include:

- Hofstede's cultural dimensions (e.g., power distance, individualism),
- V-Dem and Freedom House indices on democracy and civil liberties,
- WVS indicators of trust, religiosity, and civic engagement.

While theoretically grounded and widely cited, such measures may reflect normative assumptions or vary in meaning across cultural contexts. Caution is advised in causal inference or cross-regional comparisons involving these variables.

7.4. Lack of Subnational Disaggregation

All institutional and contextual indicators are reported at the national level. As such:

- Subnational variation (e.g., state-level governance in federal systems, urban–rural divides) is not captured.
- This limitation is particularly relevant for countries with substantial internal heterogeneity (e.g., the United States, China, India, Brazil), where firm-level exposure to institutions may differ within the same national boundary.

8. Licensing and Data Access Compliance

This project integrates both publicly available and subscription-based data sources. The following section outlines the access rights and usage conditions applicable to each data tier.

8.1. Firm-Level Data

Firm-level data were accessed through an **institutional academic subscription at Mälardalen University**, under the authority of the **School of Business, Society and Engineering (Division of Marketing and Strategy)**.

The primary sources for firm-level data include:

- **Refinitiv Eikon** — a provider of structured financial and corporate data;
- **Orbis Global (Bureau van Dijk)** — a global database of firm-level information;
- **Forbes Global 2000** — a publicly available annual ranking of the world's largest companies.

Access to Refinitiv Eikon and Orbis was granted **exclusively through secured academic channels** and used strictly within the framework of:

- non-commercial academic research,
- thesis and dissertation development,
- internal educational analytics in courses related to international business and strategy.

All use of proprietary data fully complied with the terms of the institutional licenses. **No raw data from restricted sources are published or redistributed**, and all analyses rely solely on aggregated or derived metrics.

8.2. Country-Level Data

Country-level variables were obtained **exclusively from open-access, publicly available databases**, including but not limited to:

- World Bank (World Development Indicators),
- Worldwide Governance Indicators (WGI),
- V-Dem Institute,
- Hofstede Insights,
- World Values Survey,
- Environmental Performance Index (EPI),
- Global Health Security Index (GHSI),
- Freedom House,
- World Intellectual Property Organization (WIPO), and others.

These sources are widely used in academic research and offer full transparency and reproducibility.

8.3. Restrictions on Redistribution

This dataset is intended **solely for internal academic use** within Mälardalen University. Redistribution of the full dataset, transfer to external institutions, or publication of raw data originating from licensed providers (Refinitiv, Orbis) is **strictly prohibited** unless prior authorization is obtained from the respective data vendors.

All users granted access to the dataset are expected to adhere to the relevant licensing agreements and to limit their use to scholarly research, teaching, and academic analysis within the institution.

9. Version History and Updates

This section details the dataset's release timeline, versioning practices, and planned extensions, ensuring transparent maintenance and long-term usability.

The versioning scheme follows semantic versioning principles (vMAJOR.MINOR.PATCH) to distinguish between full releases, incremental feature additions, and minor corrections. All dataset versions are archived internally with associated documentation and changelogs to support reproducibility and scholarly traceability.

Version v1.0.0 – August 2025

Initial public academic release, including:

- Full integration of firm-level data for 560 publicly listed multinational manufacturing firms (2010–2024).
- Construction of country-level panel for up to 85 countries and 75+ institutional, macroeconomic, cultural, and environmental indicators.
- Comprehensive data cleaning, harmonization, and variable standardization procedures applied.
- Delivery of a two-level panel structure: firm × year and country × year.
- Full technical documentation:
    - Technical Appendix
    - Two variable dictionaries (firm-level and country-level)
    - Metadata coding schema and source attribution

    - Summary statistics tables
    - Usage notes and licensing details
- Final dataset stored for institutional academic use at Mälardalen University under university-wide data licensing agreements.

Planned Extensions (Short-Term: v1.1.0)

Subject to institutional priorities and data availability, the following updates are under consideration for the next release:

- Addition of fiscal year 2025 data (firm- and country-level).
- Expansion of governance variables: e.g., institutional ownership, CEO tenure, board independence.
- Inclusion of benchmark aggregates by industry and region.
- Extended coverage of countries with improved access to institutional and macro data.
- Introduction of selected subnational indicators (e.g., U.S. states, German Länder).
- Partial automation of update workflows (via scripted pipelines in R and Python).

Long-Term Roadmap (Major Releases: v2.0.0+)

Longer-term improvements will focus on scalability and academic integration:

- Annual updates with version tracking and changelogs.
- Open-access metadata portal (under development).
- Interactive dashboard tools for internal visualization and pedagogical use.
- Modular expansions: ESG metrics, innovation, trade and investment flows.
- DOI registration and citation guidelines for scholarly referencing.

Versioning Policy

Each new version will be accompanied by:

- A changelog detailing additions, removals, and modifications.
- Archived snapshots of the dataset (CSV, UTF-8 encoded).
- Updated metadata and variable dictionaries.
- Explicit documentation of deprecated variables (retained for backward compatibility).

This versioning protocol is designed to ensure methodological integrity, transparent documentation, and longitudinal comparability across releases.