Organisation for Economic Co-operation and Development

SDD/DOC(2024)1

Unclassified

English - Or. English

2 February 2024

STATISTICS AND DATA DIRECTORATE

The OECD-UNSD Multinational Enterprise Information Platform

SDD Working Paper No. 120

Contact: stat.contact@oecd.org.

JT03536604

OECD STATISTICS WORKING PAPER SERIES

The OECD Statistics Working Paper Series – managed by the OECD Statistics and Data Directorate – is designed to make available in a timely fashion and to a wider readership selected studies prepared by OECD staff or by outside consultants working on OECD projects. The papers included are of a technical, methodological or statistical policy nature and relate to statistical work relevant to the Organisation. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors.

Working Papers describe preliminary results or research in progress by the authors and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed and may be sent to the Statistics and Data Directorate, OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France.

This document, as well as any statistical data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

The release of this working paper has been authorised by Paul Schreyer, OECD Chief Statistician and Director of the OECD Statistics and Data Directorate.

www.oecd-ilibrary.org/economics/oecd-statistics-working-papers_18152031

The OECD-UNSD Multinational Enterprise Information Platform

Graham Pilgrim and Shirly Ang

Abstract / Résumé

The OECD and the United Nations Statistics Division (UNSD) have developed jointly the new Multinational Enterprise Information Platform (MEIP). MEIP is built on past OECD and UN efforts to compile statistics on the scale and scope of the international activities of Multinational Enterprises (MNEs). The new platform uses publicly available data to gather information on the world's 500 largest MNEs in a timely manner, facilitating a comprehensive view of their physical and digital presence. It also includes a monitoring tool for large events such as Mergers and Acquisitions (M&A). The platform also provides a valuable benchmark for National Statistical Offices (NSOs) and researchers, allowing them to compare the national presence of an individual MNE to the global presence. Information on MNEs and their global network can also be visualised in a user-friendly dashboard.

Keywords: multinationals, open source, business register.

JEL codes: F23,C55, C81.

****

L'OCDE et la Division de statistique des Nations Unies (DSNU) ont développé conjointement la nouvelle Plateforme d'information sur les entreprises multinationales (PIEM). PIEM s'appuie sur les efforts passés de l'OCDE et de l'ONU pour compiler des statistiques sur l'ampleur et la portée des activités internationales des multinationales. La nouvelle plateforme utilise des données en accès libre afin de collecter des données récentes sur les 500 plus grandes multinationales du monde, facilitant ainsi une vue complète de leur présence physique et numérique. Elle comprend également un outil de suivi de changements importants concernant les entreprises tels que les fusions et acquisitions. La plateforme fournit également une référence précieuse aux Offices de Statistiques Nationaux et aux chercheurs, leur permettant de comparer la présence d'une entreprise multinationale sur le territoire national et dans le monde. Les informations sur les entreprises multinationales et leur réseau mondial peuvent également être visualisées dans un tableau de bord facile à utiliser.

Mots clés : multinationales, données en libre accès, registre des entreprises.

Codes JEL : F23, C55, C81.

Table of contents

FIGURES

TABLES

The OECD-UNSD Multinational Enterprise Information Platform

By Graham Pilgrim and Shirly Ang[1] [2]

1. Introduction

1. Multinational Enterprises (MNEs) have been at the forefront of changes in the global economy, as transportation and communication costs have declined, technologies have facilitated more complex operations, and trade and investment barriers have lessened over the last few decades. Understanding the structure and behaviour of MNEs is fundamental to the production of consistent global statistics and policy insights. This information facilitates analysis of the impact of globalisation on value chains, cross-border investment, productivity growth, market power and knowledge spill overs, and provides transparency to the public with regards to large enterprises operating in their countries, particularly in regard to the international efforts to ensure that these firms pay a fair share of tax wherever they operate and generate profits.[3]

2. Despite their significant and growing importance, information on MNEs remains limited. The cross-border nature of their activities can make them particularly difficult to measure through traditional statistical infrastructure and methods. Traditionally, national statistical offices (NSOs) are responsible for measuring, aggregating, and disseminating information on economic activity. They typically work on the basis of residence rather than nationality of production units, and as a result data collections generally focus on the activities of MNEs and their affiliates within their economic territory and rarely provide the view of an entire MNE. Therefore, data on MNEs are often fragmented across numerous countries, each NSO holding a single piece of the puzzle, but with no one able to see the full picture. This is further complicated by confidentiality restrictions dictating that data reported to an NSO cannot typically be shared externally, even with other NSOs.

[1] The authors are working in Statistics and Data Directorate, OECD and UNSD. They would like to thank Federica Daniele, Samuel Delpeuch, Peter Horvát, Asa Johansson, Alexander Lembcke, Pierce O'Reilly and Paul Schreyer for valuable comments and Virginie Elgrably for excellent technical assistance.

[2] The compilation of MEIP for this year has benefitted greatly from the contributions of Shirly Ang, Htu Aung, Julian Chow, Aida Diawara, Ilaria Di Matteo, Pedro Farinas, Erifeoluwa Jamgbadi, Covadonga Machicado Alvarez, Annabelle Mourougane, Graham Pilgrim and Zhiyuan Qian.

[3] www.oecd.org/tax/beps/.

3. MNEs also often have vast and complex structures, which move and change over time. They are able to quickly and flexibly augment their operations, merging with or acquiring other companies (i.e. Mergers & Acquisitions), shifting the location of key activities, or otherwise restructuring. MNEs also vary greatly in their physical and digital presence. Alphabet (the parent company of Google), for instance, has a far greater digital presence than a physical presence, with a domain for almost every jurisdiction in the world. At the same time, multinationals like oil and gas companies are highly asset driven, meaning their physical presence is more important. This compounds the measurement challenge for NSOs, with a clear role to be played by international organisations in helping to understand MNEs in greater detail.

4. Against this background, the OECD and the United Nations Statistics Division (UNSD) have developed jointly the new Multinational Enterprise Information Platform (MEIP). MEIP is built on past OECD[4] and UN[5] efforts to compile statistics on the scale and scope of the international activities of MNEs. The new platform uses publicly available data alongside extensive manual validation to gather information on the world's 500 largest MNEs, facilitating a comprehensive view of their physical and digital presence. It also includes a monitoring tool for large events such as Mergers and Acquisitions (M&A).

5. MEIP is more timely than similar structural databases, with data compiled for the 31st December 2022, and provides a one-stop-shop for data users, with the aim of future releases updating this information on a yearly basis and expanding coverage. The platform also provides a valuable benchmark for NSOs and researchers, allowing them to compare the national presence of an individual MNE to the global presence. Information on MNEs and their global network can be visualised in a user-friendly dashboard (https://bit.ly/mne-platform).

6. This paper is structured as follows. Section 2 discusses the challenges faced in the construction of a global business register such as MEIP. Section 3 gives an overview of the workflow, whilst Section 4 details sources, methods and techniques. Section 5 presents examples of analyses that can be drawn from MEIP. Section 6 discusses how MEIP can be improved through further outreach. Section 7 concludes.

[4] OECD Analytical Database of Individual MNEs and their Affiliates: www.oecd.org/sdd/its/ADIMA-Methodology.pdf.

[5] UNSD Global Groups Register: https://unstats.un.org/unsd/business-stat/GGR/.

2. Challenges to building a global business register

7. The goal of a global business register is to collect information on the geographic and economic footprint of the world's largest MNEs and extend this to the level of each individual affiliate. The existence of a global business register that shows the legal structure of the largest MNEs can assist countries to understand the non-national part of the MNEs in their country; facilitate data sharing among countries using the global business register as a common, public source; and, more generally, aid the analysis of globalisation effects and global value chains. In theory official statistics collected by national statistics offices (NSOs) could be pooled together to provide these views. However, in practice, confidentiality restrictions typically prohibit data sharing[6] across NSOs and the dissemination of confidential information on individual affiliates. In addition, there does not exist a system of unique identifiers for MNEs at the global level that poses a serious challenge to the development of a global business register.

Could publicly accessible business registries be the solution?

8. Publicly accessible or open source data can help overcome some of these limitations, removing the constraint of breaching confidentiality restrictions. Jurisdictions are increasingly providing access to business registry data, including in some cases ownership and financial indicators. For example, in the United Kingdom a bulk download of all companies detailing basic information is made available[7] and a supplementary register of People with Significant Control is also made available[8,] which allows for chains of control to be determined. Unfortunately, although promising, coverage is not complete and highly dependent on jurisdiction, and therefore, the information currently available is insufficient to build a global business register alone.

Could Annual Reporting be the solution?

9. One of the best examples of timely open source information is the individual annual reports published by MNEs and related documents such as financial statements and filings for stock exchanges. These will be referred to in this paper as 'Annual Reporting' documents. They provide a snapshot of the current state of the business, including financial statements describing the company's performance during the last year. The notes to the financial statements may include information about subsidiaries and geographic breakdowns of performance. However, Annual Reporting is only a regulatory requirement for publicly traded companies and therefore only provides coverage for a subset of MNEs.

10. In addition, Annual Reporting formats are not standardised, and so the process of harmonising data collection for many MNEs is necessarily a manual and resource intensive process.

11. Furthermore, the type of information required in Annual Reporting is largely determined by the stock exchange where a company is listed, and completeness in a number of areas is not guaranteed. For example, Apple Inc, with a primary listing in the United States, files Financial Reports with the Securities and Exchange Commission (SEC). SEC affiliate reporting standards allow for subsidiaries which correspond to less than 10% of group investment, assets or income to be excluded, and therefore Apple Inc declares only 19 affiliates within 12 jurisdictions for Financial Year 2022,[9] whereas evidence of

[6] However, some efforts leaded by International Agencies are encouraging and providing tools to overcome these restrictions, like the UNECE Task Force on Exchange and Sharing of Economic Data.

[7] Available at: http://download.companieshouse.gov.uk/en_output.html.

[8] Available at: http://download.companieshouse.gov.uk/en_pscdata.html.

[9] Apple Inc. Significant Subsidiaries from Annual Report 2022, available at: www.sec.gov/Archives/edgar/data/320193/000032019322000108/a10-kexhibit21109242022.htm.

a physical presence in most countries is evident from their website. Furthermore, details of the financials of each Apple Inc affiliate are not made available – instead aggregations are often made on geographic levels. These geographic levels are non-standardised, with Apple Inc defining reportable segments of "Americas", "Europe", "Greater China", "Japan" and "Rest of Asia Pacific", where "Europe" consists of "European countries, as well as India, the Middle East and Africa".[10] The example of Apple Inc. is indicative of a wider issue surrounding the depth and consistency of reporting.

Pooling multiple sources can increase coverage

12. Other sources of publicly available information on MNEs can also be mobilised (Table 1). Like Annual Reporting however, these additional sources are also, typically, limited to a subset of companies (by function, jurisdiction or significance) and, ownership and financial indicators are also often limited.

Table 1. A sample of potential sources of data on MNEs

	Source	Coverage	Website/Email	Ownership
PermID (1)	permid.org	Subset of global firms	Yes	No
LEI (2)	gleif.org	Subset of global firms (typically engaging in financial transactions)	No	Yes
Business Registers	National Sources	Full coverage of firms within a jurisdiction	Varies	Varies
Common Crawl (3)	commoncrawl.org	Most important global websites	Yes	No
X.509 Certificates (4)	crt.sh	Websites utilising security certificates	Yes	No
WikiData (5)	wikidata.org	Subset of global firms listed on Wikipedia	Yes	Yes

Notes:
(1) PermID provides a subset of global companies as identified by the London Stock Exchange Group, these tend to be companies which are publicly listed, or commonly referenced in media articles.
(2) The Legal Entity Identifier (LEI) is a unique global identifier to identify the legal entities participating in financial transactions and provides additional information to identify the ultimate owner of each legal entity.
(3) Common Crawl is an open source initiative to provide a monthly snapshot of the internet through web scraping. As part of the project, they provide a quarterly network graph of hyperlinks between domains, which can be used to determine whether two websites are probably linked. For example, if example.com links to example.fr, and example.fr links to example.com, then it is likely that they are a member of the same Parent MNE family.
(4) X.509 Certificates are used in a number of internet protocols to ensure secure communications (such as in e-commerce) and have increased in use as internet search engines have begun to prioritise domains with secure communications in their ranking algorithms. X.509 certificates detail the company and the domains that it operates.
(5) WikiData is a community sourced initiative derived from Wikipedia data, it details the website and subsidiaries of a company.
Source: Authors'compilation.

[10] Apple Inc. Annual Report 2022, available at: www.sec.gov/ix?doc=/Archives/edgar/data/320193/000032019322000108/aapl-20220924.htm.

13. Although all of these sources come with caveats and gaps, pooling them together provides significant scope to overcome the caveats and fill gaps, providing a more comprehensive view of the whole of the MNE and the location of its affiliates, together with an economic view of each affiliate.

14. Pooling works because each source typically contains certain identifying information that allows any given unit to be linked to units identified in other sources. As such, a view of all (or at least as many as the sources allow) of the affiliates within a given MNE can be knitted together. These identifiers include "shared" resources such as websites. A number of initiatives have followed a similar approach (Box 1).

Box 1. Initiatives of building global/regional business registers

The EuroGroups Register (EGR) shows the benefits that can be gained by pooling together data from NSOs within the European Statistical System (ESS). The EGR aims to facilitate the coordination of survey frames for producing high quality statistics on global business activities. However, access to this information is strictly limited to those within the ESS and coverage is not global but limited to the contributors.

Some private sector data vendors do compile information on global business registers. Providers include Dun & Bradstreet and Bureau van Dijk. However, the cost of data access is often prohibitive for NSOs and International Organisations. Licence restrictions often prevent the dissemination of results which reveal individual company level data. In addition, metadata and documentation of data provenance is often missing or incomplete, making appraising data quality difficult.

Initiatives at the global level also exist. The Legal Entity Identifier (LEI)[11] is a unique global identifier of the legal entities participating in financial transactions and provides additional information to identify the ultimate owner of each legal entity. LEIs are a requirement to participate in a number of financial markets. Unfortunately, their issuance tends to be limited to the subset of firms operating within financial markets. Another example of a global initiative includes PermID[12] which provides an identifier for a subset of global companies identified by the London Stock Exchange Group. However, coverage is not universal and tends to be companies which are publicly listed, or commonly referenced in media articles.

[11] More information regarding LEI can be found at: www.gleif.org/en/.

[12] More information regarding PermID can be found at: https://permid.org/.

3. Constructing MEIP – Overview of the workflow

15. In theory, with the range of sources described above, significant scope exists to develop a truly global register. Additional sources will be needed to ensure adequate coverage of privately owned or government controlled firms and of MNEs in developing economies where the sources described above have larger gaps. However even if all this information were available the necessity to undertake resource intensive data capture and validation processes necessarily limits the scope of MNEs that can be covered.

16. The OECD-UNSD effort has been to focus on the world's largest publicly listed MNEs, and to develop a register of these firms. The construction of MEIP follows nine distinct stages (Figure 1 and Section 4).

17. MEIP currently focuses on the world's 500 largest listed MNEs (Stage 1). The top 500 are selected on the basis of market capitalisation, i.e. the total value assigned to a company's outstanding shares of stock. For example, the largest 500 companies by Market Capitalisation for 2022 have aggregated global revenues of over USD 24 trillion, which is roughly equivalent to the GDP of the United States. If considering a list such as the Forbes Global 500, which is ranked by Revenues, this revenue increases to USD 41 trillion, however this tends to focus on industries with a lower profit margin such as retailing and show greater bias. Going forward market capitalisations are subject to movements, which means the MEIP sample will change with each vintage. It is however expected that there will be sufficient overlap to consider longitudinal studies in a number of cases.

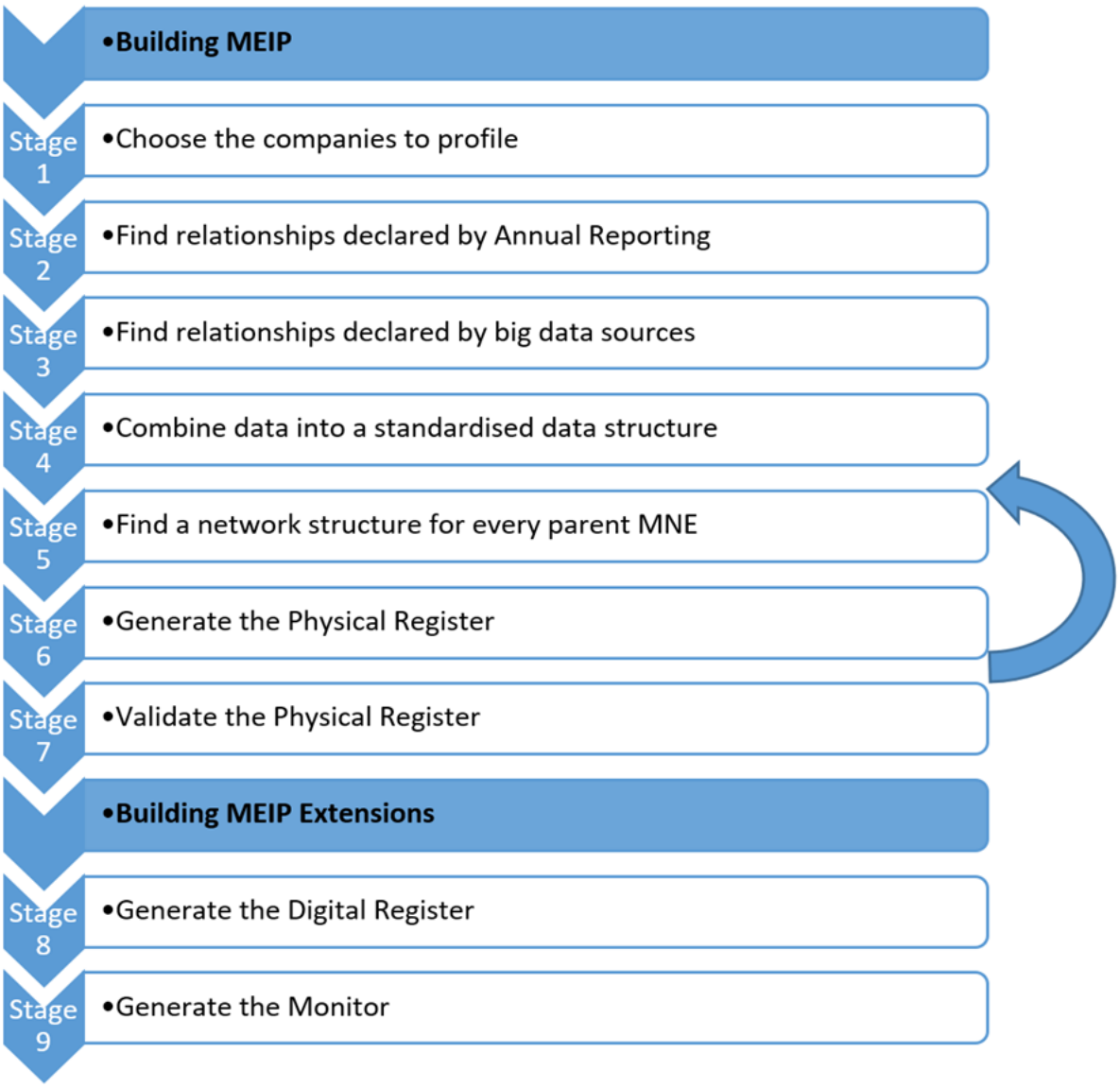
18. Annual Reporting is then used to create the first view (Stage 2) of the structure of each selected MNE (including using manual data capture tools). Other open source big data sets are used to complement this information to provide a more comprehensive view of the MNE structure (Stage 3).

19. The collected data are then standardised into a common data structure (Stage 4, Box 2). This involves extracting metadata for each unit, for example data regarding the location and address. In addition, connections are also extracted, for example, in the case of a website, data regarding the company that has a certificate to communicate securely on the domain and linked websites is collected. This information is then brought together and linked. Because of the multitude and complexity of connections across parents and affiliates, MEIP uses a network database approach, which is better suited to describe the connections between parents and affiliates. Network databases provide scope to perform complex queries which under a traditional database would be computationally complex and expensive. An example of this is the search performed in Section 4 Stage 5 (below) to determine all identifiers which are linked to a Parent MNE. In a network graph architecture, relationships are followed from each identifier until all routes are exhausted.

20. For each MNE group structure, MEIP builds a physical register of affiliates (Stage 6). However, as the physical register relies on the accuracy of a number of data sources, validation is required to ensure accuracy (Stage 7). Validation involves checking the identified and matched link of each affiliate that has been derived by non-standard links such as via websites. This process is continued iteratively until the physical register has been fully validated.

21. As the data sources go beyond those utilised by a traditional business register, it is possible to build extensions, including a digital register of the websites operated by the MNE (Stage 8, see also Section 4) and a monitoring tool (Stage 9) for large corporate events.

Figure 1. MEIP step by step



Source: MEIP.

Box 2. Pooling data in practice

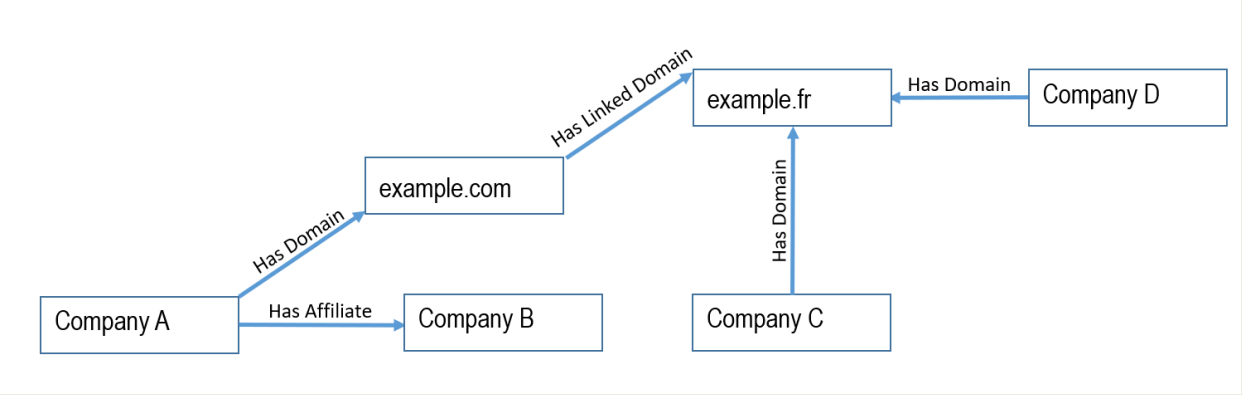
This box illustrates how the various sources of data can identify parts of an MNE that may not be identified in separate sources.

Consider a parent MNE, Company A, with two affiliates Company B and Company C (Figure 2). In the Annual Reporting source Company A reports only its control of company B, but not Company C. However administrative and statistical business registers may provide additional information about Company A, for example its domain website (example.com), and with digital sources (e.g. CommonCrawl) it can be established that this domain name is linked to example.fr, which can be identified as belonging to Company C through X.509 Certificates. As such it is possible to identify links between the parent A and one additionally affiliated firm, C.

Some care is needed in confirming these links and so further validation is needed. For example, the use of security certificates to establish links across domains is not fool-proof, as another Company D may also have a right to operate the domain but Company D may not be controlled by Company A, for example Company C may merely have outsourced the provision of digital services on the domain to Company D.

Validation focuses on determining whether an affiliate is controlled by the parent MNE[13] (i.e. has a voting power greater than 50%), and is performed by desk research into each individual company.

Figure 2. Example of a network for an MNE



Source: MEIP.

[13] A controlled affiliate (or subsidiary) is an enterprise in which an investor owns more than 50% of its voting power (OECD Benchmark definition of Foreign Direct Investment). The same criterion determines the enterprises covered in the Foreign Affiliates' Statistics (FATS) statistical framework.

4. Constructing MEIP – Data sources, methods and techniques

Stage 1. Choose the companies to profile

22. MEIP currently covers 500 MNEs. However, the methods and techniques used to compile this information are scalable and could be expanded to cover a wider sample. The current size was selected to achieve a good balance between relevance and feasibility. This size also aligns well with those currently chosen by large case units (LCUs). LCUs are units that are established by NSOs to work closely with MNEs to improve estimates that are required for compiling national accounts, balance of payments and other statistics. Sample sizes are chosen carefully to cover the most important firms for the production of these statistics whilst managing resource requirements. For example, the MEIP sample for 2022 has aggregated global revenues of over USD 24 trillion, which is roughly equivalent to the GDP of the United States.

23. Multinationals are selected for inclusion according to the following criteria:

- **An entity within the group is publicly traded (listed on a stock exchange):** Annual Reporting is the primary data source for Stage 2, however this means non-listed companies are omitted.

  Relevant example: Samsung Electronics is a publicly traded company and part of the wider Samsung Group (a private company). As a result, the entire Samsung Group was included in the MEIP universe.

- **Market Capitalisation:** Entities were ranked by market capitalisation as of 31/12/2022 to select the top 500. In cases where firms have dual listings, only the primary listing was considered.

- **Group structure not previously covered:** Complex chains of control mean that some firms that could be included within the universe under the previous two filtering conditions would be controlled by a firm that is also a member of the sample. To avoid double counting these cases are analysed on a case by case basis.

  Relevant example: Hindustan Unilever Limited and Unilever are publicly traded and both are within the world's top 500 market capitalisations. However, Unilever is a majority shareholder and has economic control over Hindustan Unilever. As Unilever was already a member of the MEIP universe due to its larger market capitalisation, Hindustan Unilever Limited was not included as a separate corporate entity, but rather as a part of Unilever.

Stage 2. Find relationships declared by Annual Reporting

24. To mirror the traditional business register approach, the controlled affiliates of each company are collected from Annual Reports. To best align with the reference period of 31st December 2022, the Annual Reporting for the financial year ending closest to 31st December 2022 is selected. For each controlled affiliate the following information was extracted: ultimate parent company, parent company, affiliate name, country and state of incorporation, country and state of operation, voting share and ownership share of the parent, type of consolidation in parent's financial statements, source of information and date.

25. There is no standardised format for reporting subsidiary information within Annual Reporting, and therefore data sources were prioritised with regards to the simplicity of data extraction. Sources were consulted in the following order:

- **SEC filings**: Filings with the US Security and Exchanges Commission (SEC) via their web portal Edgar[14] are prioritised, as, in many cases, the format is harmonised, and in other cases the data are provided in an easily extractible text format. The majority of US companies file an Annual Report (10-K) and foreign companies with a US listing file a Foreign Annual Report (20-F).
    - Formatted 10-K subsidiaries are obtained from the CorpWatch API,[15] which uses automated parsers to extract the subsidiary information from Exhibit 21 of the 10-K report.
    - In cases where this is unsuccessful (or for 20-F filings) text is extracted and formatted manually.
- **German Federal Business Register**: German Federal Law obliges companies to publish yearly consolidated financial statements with a full list of their subsidiaries in a legally defined format registered with the German Federal Business Register (*Bundesanzeiger*).[16] These data are extracted and formatted manually.
- **Annual Reporting**: In absence of an alternative source, data are manually extracted from Annual Reporting made available by companies (usually in the Investor Relations sections of their websites). The majority of these reports are available in PDF, and a variety of tools were used in order to streamline this process (for example, automated table extraction and Optical Character Recognition (OCR) for images).

26. Data collection involves overcoming a number of hurdles including:

- **Language**: Data collectors in the MEIP team had a relatively broad language coverage meaning the vast majority of firms could be researched and information found. However, difficulties with non-Latin character sets were present (e.g. Chinese and Arabic character sets), with some companies within the sample only providing Annual Reports in this way, meaning data collectors were unable to use them. In other cases translations were provided, however not at the same level of timeliness as the original Annual Reporting.
- **Inconsistent depth of reporting**: Legislation within the jurisdiction determines whether companies provide only information on their most significant subsidiaries or are required to report all subsidiaries. In cases where significant subsidiaries are only required, practices by competitors is a driving factor and there has been evidence of reducing transparency levels. This means that some companies within MEIP are covered in greater detail than others. There exists no mechanism to determine the actual coverage but the wide range of sources used by MEIP does mean that MEIP has higher coverage than any of the individual sources.

[14] Accessible via: www.sec.gov/edgar/searchedgar/companysearch.html.

[15] Accessible via: http://api.corpwatch.org/.

[16] Accessible via: www.bundesanzeiger.de/.

- **Interpretation issues**: Data often proved difficult to interpret and extensive metadata was needed to validate data. In many cases at least one of the following cases occurred:
  - Unknown consolidation basis.
  - Unknown whether the reference refers to a Legal Entity, Branch or Establishment.
  - Unknown whether location refers to incorporation or physical operations (this was especially true for special purpose entities and branches).
  - Unknown whether 'percentage of ownership' refers to the percentage of a subsidiary/affiliate owned by the parent or the percentage controlled.

27. Due to these hurdles, extensive metadata was attached to each data point, enabling the decisions made during data collection to be challenged during validation (Stage 7).

Stage 3. Extract relationships declared by open source big data sources

28. To enhance the traditional business register approach (Stage 2) a number of additional data sources are also considered (see Table 2 and Section 2).

29. All data sources are collected via a bulk data download or an Application Program Interface (API). With bulk downloads, information is downloaded and scripts executed in order to filter and harmonise the output structure. With APIs, scripts are developed to submit the required queries, and results are returned and harmonised. The output of harmonised data is therefore largely automated within Stage 3.

30. One example of data harmonisation is simplifications at the data collection stage. For example, websites and email addresses are harmonised to extract only the underlying domain. For example, the MEIP website www.oecd.org/sdd/its/mne-platform.htm and email address graham.pilgrim@oecd.org, would both be harmonised to extract only the domain oecd.org.

31. All data sources have a shorter (more timely) update frequency than Annual Reporting, and whilst MEIP currently provides estimates for the end of 2022, this does allow for the potential to produce data on a close to real-time basis in the future. Furthermore, the collection scope of all sources listed exceeds the 500 MNEs and thus provides scope to scale up beyond the top 500 MNEs.

Table 2. Open source big data sources within MEIP

	Source	Collection Method	Average update frequency	Description
PermID	permid.org	API	Real time	PermID provides a database detailing individual firms by the identifier PermID. Coverage is determined by companies deemed of interest by the database issuer. Data for each PermID includes Business Name, Address and website.
GLEIF	gleif.org	Bulk Download	8 hours	GLEIF provides a database detailing individual firms by the identifier LEI. Coverage is determined by those companies requesting an LEI and legislation exists in a number of jurisdictions requiring an LEI for engaging in certain transactions. Data for each LEI includes Business Name, Address and Parent LEI.
Business Registers	National Sources	Bulk Download	Depends on source	Business Registers detail each firm by the national Business ID. When available coverage tends to include all firms within the jurisdiction. Data availability depends on the source but can include Business Name, Address, Website, Email Address and Parent Business IDs.
Selected National Sources	National Sources	Bulk Download/API	Depends on source	A number of individual sources exist at a national level. These range from website ownership records (WHOIS), information on Government contractors and companies requesting visas for their employees. In general, these sources provide a way of giving a link between a company name and a website.
Common Crawl	commoncrawl.org	Bulk Download	3 months	The CommonCrawl compiles a network graph of hyperlinks existing between websites. MEIP processes this data to find linked domains. A linked domain is defined when two domains have a matching second level domain (SLD) and both link to one another. In the case in Figure 2, if example.com links to example.fr and example.fr links to example.com as they have the same SLD (example) they are considered to be linked domains.
X.509 Certificates	Crt.sh	Bulk Download	2 weeks	X.509 certificates are heavily used in the security of online communications. They also determine the identity of the party being communicated, which is verified by a Certificate Issuing Authority (CA). Various levels of verification exist, but MEIP utilises only Organisation Level (OV) and Enterprises Level (EV) certificates as for issuance the CA has to verify the existence of the company and that they have the right to operate the underlying website. MEIP processes this data to find Business Names, associated Business IDs and the operated domains.
WikiData	wikidata.org	API	Real time	WikiData provides structured data sourced from underlying Wikipedia projects. In particular, for every Wikipedia concept (WikiID) which is identified as a business data can be downloaded for the WikiID of affiliates, associated websites and social media profiles.
OpenStreetMap	openstreetmap.org	Bulk Download	Real time	OpenStreetMap provides a geospatial database of places, and provides details of the related website, brand (WikiData) and social media profiles.
Crunchbase	crunchbase.com	Bulk Download	Depends	Crunchbase provides a company database providing information on websites and social media profiles.
PeopleDataLabs	peopledatalabs.com	Bulk Download	Depends	PeopleDataLabs provide a company database providing information on websites and social media profiles.
Web Data Commons	webdatacommons.org	Bulk Download	Yearly	WebDataCommons extract structured metadata from websites. It therefore contains information linking websites to their social media profiles.
Name Matching	permid.org and opencorporates.com	API	Real time	Business Names are declared in the Annual Reporting (Stage 2) and X.509 Certificates, but often company names are not harmonized (for example capitalisation, acronyms and abbreviations) and no link exists to another identifier within the database.

Source: MEIP.

Stage 4. Combine data into a standardised data structure

32. In order to combine the data collected during Stage 2 and Stage 3 a network database is utilised. This involves extracting metadata and connections for each identifier collected. An **identifier** for a company may have metadata regarding its location and address, and connections for the affiliates that it declares within Annual Reporting. However, in the case of an identifier for a website, no metadata exists, but **connections** to linked domains (via CommonCrawl) and companies which are authorised to operate on the domain (via X.509 certificates) do exist. Formally the identifiers are known as "**vertices**" and the connections as "**edges**", and each has a standardised data structure within the MEIP framework (and described in more detail below).

33. Each **edge** has a starting point ("From Vertex"), ending point ("To Vertex") and an explanatory variable for the reason for the connection ("Relationship Type"). A further metadata variable to reflect the direction in which the points are linked ("Direction") allows for Joint Ventures to be determined (discussed later). The aggregate set of edges defines the **vertices,** and metadata is appended to these vertices to provide detail on Name, Location and Addresses for the Vertex (Table 3).

Table 3. Data schema for vertices and edges within MEIP

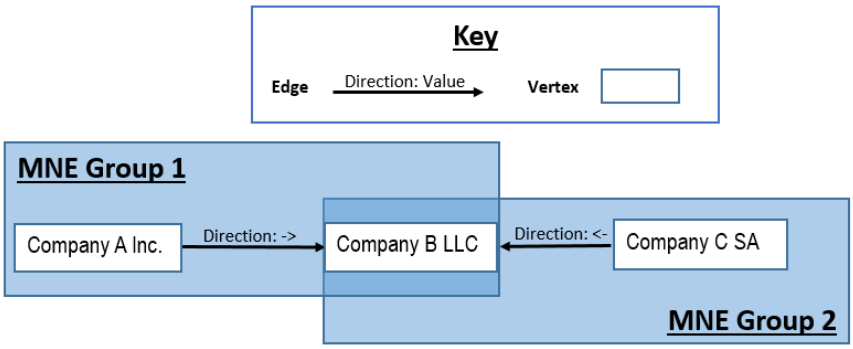
	Variable	Value
Vertices Data Schema		
Vertex	Type	Options: PermID, LEI, Business Name, Business ID, Domain, WikiID, OSM, Crunchbase ID, Social Media
	ID	Text
Metadata		
	Name	Text
	Location	Text
	Addresses	Text
Edges Data Schema		
From Vertex		
	Type	Options: PermID, LEI, Business Name, Business ID, Domain, WikiID, OSM, Crunchbase ID, Social Media
	ID	Text
To Vertex		
	Type	Options: PermID, LEI, Business Name, Business ID, Domain, WikiID, OSM, Crunchbase ID, Social Media
	ID	Text
Metadata		
	Relationship Type	Options: Has Parent, Has Affiliate, Has LEI, Has Domain, Has Business ID, Has Business Name, Has Linked Domain, Has PermID, Has Social Media, Has WikiID
	Direction	Options: ←, →, ↔

Source: MEIP.

34. The "Direction" variable defined within the data schema allows MEIP to implement a directed network structure, which is necessary to be able to handle Joint Ventures and controls the flow of how searches are made within the MEIP network. In general, where control exists the relationship of the edge is defined in both directions ("↔"), however in the case of a joint venture (two parties with joint control of the same company) the direction of the relationship needs to be defined from parent to child ("←"), or child to parent ("→"), to prevent two MNEs from being falsely connected.

35. Figure 3 gives a representation of this methodology in practice. In this case MNE Group 1 has declared that the Parent Company A has child Company B, whereas MNE Group 2 has declared that Parent Company C also has child Company B. With an undirected graph we would determine that Company A and Company C are part of the same MNE Group. However, in this case as a path from Company C to Company B is only allowed, from the perspective of MNE Group 1 Company C is unrelated to Company A, and likewise from the perspective of MNE Group 2 Company A is unrelated to Company C. In practice, the number of these cases are limited and make up a very small share of the database.

Figure 3. Reflecting joint ventures in MEIP



Source: MEIP.

36. Table 4 provides a summary of the type of edges determined from the data found in Stage 2 and Stage 3 and gives a count of the number of edges within MEIP. In total, MEIP contains over 65 million relationships, of which over 99% are found through big data sources (Stage 3). Although the number of edges varies significantly depending on data source, it is difficult to quantify the impact of each data set individually due to the interlinked nature of the data.

Table 4. Summary of relationships (Edges) within MEIP

	Source	From Vertex Type	Relationship Type	To Vertex Type	Count in MEIP
Stage 1	Annual Reporting	Business Name	Has Parent	Business Name	4,311
	Annual Reporting	PermID	Has Affiliate	Business Name	84,334
	Annual Reporting	PermID	Has Parent	Business Name	1,347
Stage 2	PermID	PermID	Has Domain	Domain	1,234,537
	PermID	PermID	Has LEI	LEI	1,734,700
	GLEIF	LEI	Has Affiliate	LEI	1,247,968
	GLEIF	LEI	Has Parent	LEI	111,860
	Business Registers	Business ID	Has Affiliate	Business ID	427,896
	Business Registers	Business ID	Has Domain	Domain	3,194,747
	Selected National Sources	Business Name	Has Business ID	Business ID	230,771
	Selected National Sources	Business Name	Has Domain	Domain	625,085
	Common Crawl	Domain	Has Linked Domain	Domain	1,419,023
	X.509 Certificates	Domain	Has Business ID	Business ID	74,527
	X.509 Certificates	Domain	Has Business Name	Business Name	26,224,544
	WikiData	WikiID	Has Affiliate	WikiID	115,000
	WikiData	WikiID	Has Business ID	Business ID	37,458
	WikiData	WikiID	Has Domain	Domain	202,749
	WikiData	WikiID	Has LEI	LEI	34,226
	WikiData	WikiID	Has PermID	PermID	3,673
	WikiData	WikiID	Has Social Media	Social Media	127,328
	OpenStreetMap	*OSM*	*Has Domain*	*Domain*	*3,733,779*
	OpenStreetMap	*OSM*	*Has Social Media*	*Social Media*	*228,602*
	OpenStreetMap	*OSM*	*Has WikiID*	*WikiID*	*1,819,563*
	Crunchbase	*Crunchbase ID*	*Has Domain*	*Domain*	*2,562,554*
	Crunchbase	*Crunchbase ID*	*Has Social Media*	*Social Media*	*4,246,598*
	PeopleDataLabs	*Domain*	*Has Social Media*	*Social Media*	*9,990,021*
	WebDataCommons	Domain	Has Social Media	Social Media	3,125,048
	Name Matching	Business Name	Has Business ID	Business ID	2,243,593
	Name Matching	Business Name	Has PermID	PermID	372,944

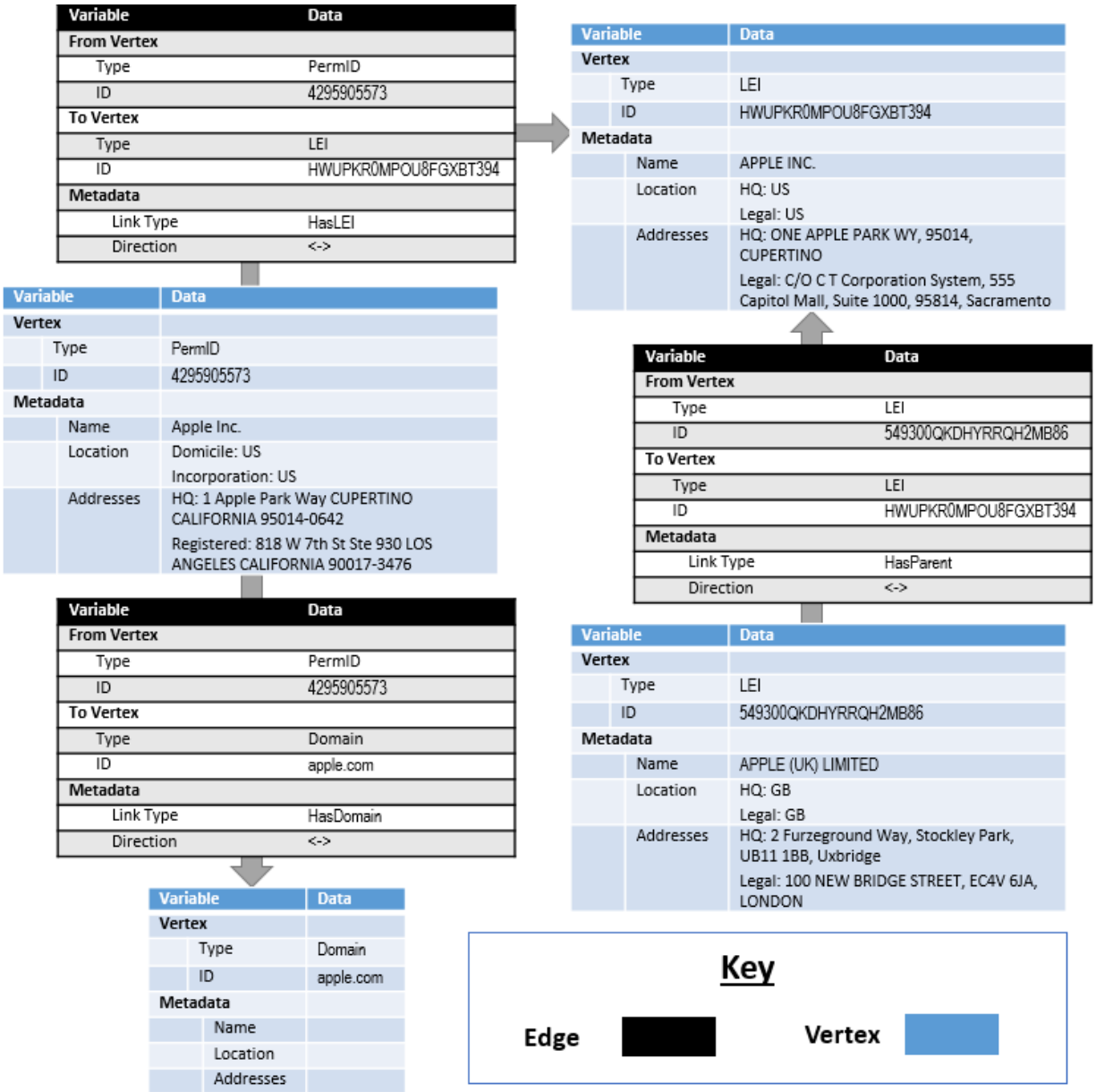
Source: MEIP.

Stage 5: Find a network structure for every parent MNE

37. This stage aims to determine all of the identifiers which are linked to a parent MNE, whether physical or digital. In simple terms, this involves finding all identifiers which can be linked to the parent MNE via the paths declared within MEIP. From a technical perspective, MEIP performs an iterative search through all data listed in Table 4 for connected edges beginning from the vertex of the parent MNE to find vertices of type: Business Name, Business ID, PermID, LEI, Domain, WikiID, OSM, Crunchbase ID and Social Media. The search takes into account the directional variables, meaning that joint ventures are handled. The average network for an MNE contains 1 800 vertices taking 8 steps to obtain, suggesting that MEIP is focussing on some of the most complex cases. The maximum number of vertices in MEIP is 41 655 (Seven & i Holdings Co Ltd), whilst the minimum is 1 (SIT Land Holdings Ltd).

38. A partial implementation of the iterative search is shown for Apple Inc. (Figure 4). The process begins from the vertex of the PermID with the value 4295905573 (Apple Inc). PermID 4295905573 is linked to the domain name apple.com, and also has an associated LEI HWUPKR0MPOU8FGXBT394. In addition, LEI 549300QKDHYRRQH2MB86 defines its parent as LEI HWUPKR0MPOU8FGXBT394. Therefore, a number of identifiers for companies within the MNE Group and a controlled domain have been determined with only three edges.

Figure 4. A subset of the MEIP network for Apple Inc.



Source: OECD MEIP.

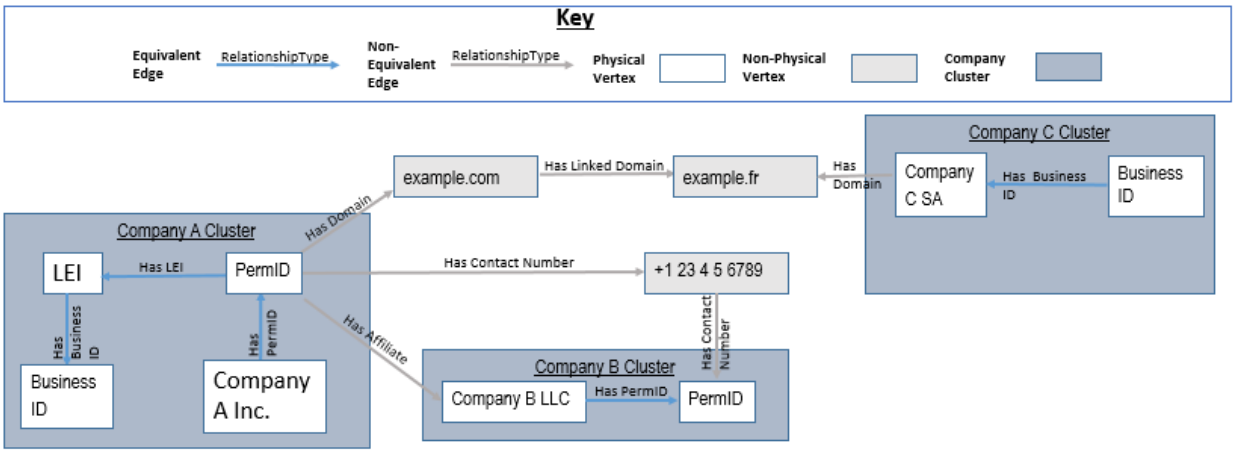
Stage 6: Generate the Physical Register

39. The network determined in Stage 5 for each MNE is used to generate the physical register. As no unique vertex type exists for every affiliate, and MEIP has to be able to map between multiple vertex types and determine equivalent concepts, two concepts are defined:

- "Physical Vertices" and "Equivalent Edges". A vertex is defined as physical if it relates to a physical company. Therefore, vertex types of PermID, LEI, Business Name and Business ID are defined as physical as they relate to physical companies.
- An edge is defined as equivalent if it provides a one-to-one mapping between two identifiers for the same physical company. Therefore, edge relationship types of *Has LEI*, *Has Business ID* and *Has PermID* are defined as equivalent.

40. To identify distinct companies in the absence of non-unique identifiers, clusters of physical nodes which are connected by equivalence relationships for a given MNE network are identified (Figure 5). The level of detail available for each cluster depends on the vertices which form the cluster, with the most detailed cluster containing information for Business Names, PermIDs, LEIs and Business IDs for the given company.

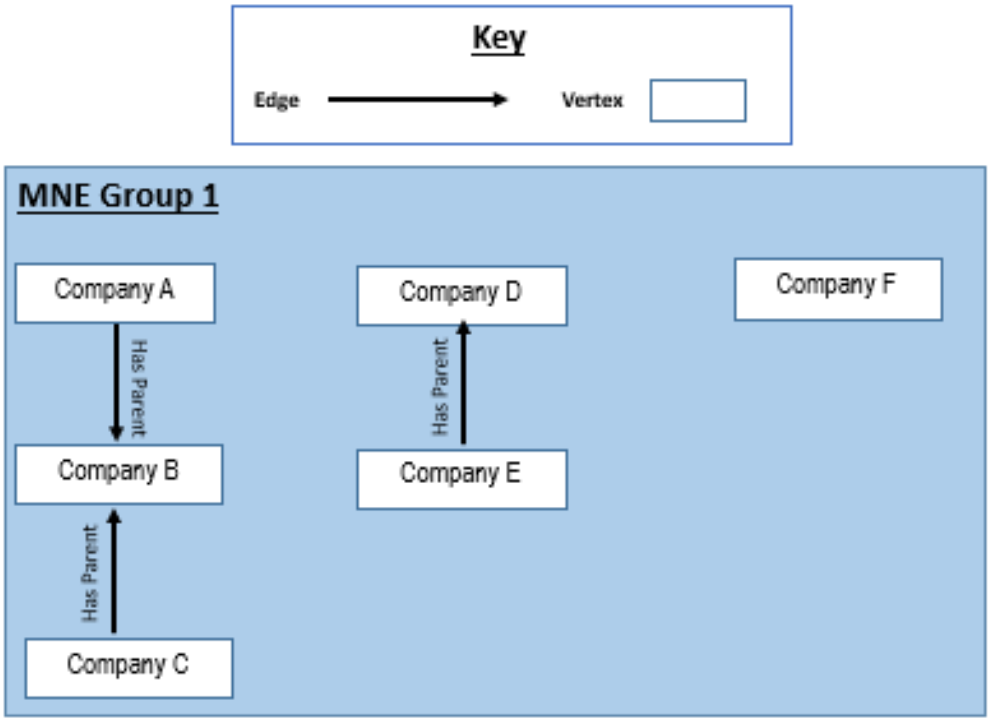
Figure 5. Determining the unique companies belonging to an MNE network



Source: MEIP.

41. Using only the "Has Parent" edge connection type (primarily available in GLEIF and Annual Reporting) and equivalent relationship types to construct company clusters it is also possible to construct a firm hierarchy. Figure 6 gives a visual representation of this process, for a profiled MNE company A. As Company A has parent Company B it is determined that Company B is the "MNE Head" for this family, as Company A and Company C are both connected to the MNE Head these companies are of "Known" Hierarchy within the company. Although relationship data is known for Company D and Company E, it is not possible to connect these firms with the MNE Head and therefore leading to only a "Partial" understanding of their hierarchy within the firm. Company F has no connections and therefore its hierarchy within the firm is completely "Unknown".

Figure 6. Determining company hierarchy within an MNE network



Source: MEIP.

42. Each entry within the physical register relates to an affiliate cluster (Table 5). The level of detail available for each affiliate cluster depends on the vertices which form the affiliate cluster, and therefore a coverage ratio is also presented.

Table 5. Variables within the Physical Register

	Description	Calculation Method	Coverage Ratio
Parent MNE	Name of the Parent MNE	The name of the source vertex used for determining the Group Network	100%
Hierarchy	Level of Hierarchy which is known	Reflects whether the Company is the "MNE Head", or whether its hierarchy within the MNE Family is "Known", "Partial" or "Unknown"	100%
ISO3	Jurisdiction of the subsidiary	The jurisdiction of the subsidiary identified by the most trusted source within the subsidiary cluster	99%
Subsidiary Name	Name of the subsidiary	The name of the subsidiary identified by the most trusted source within the subsidiary cluster. The most trusted source is obtained from the first source with available data in the following order: Country Business Registers, LEI, Annual Reporting, X.509 Certificates and PermID.	100%
Business ID	Business ID for the subsidiary	The Business IDs for vertices of type "Business ID" within the affiliate cluster	50%
OpenCorporates	OpenCorporates reference for the Subsidiary	The OpenCorporates reference for the subsidiary	48%
LEI	LEI for the subsidiary	The LEIs for vertices of type "LEI" within the subsidiary cluster	22%
PermID	PermID for the subsidiary	The PermIDs for vertices of type "PermID" within the subsidiary cluster	60%
Alternative Names	Alternative names which have been used within MEIP to refer to the subsidiary	The names of the subsidiary identified by sources which were not presented in the "Subsidiary Name" variable	27%
Address	Address of the subsidiary	The address of the subsidiary identified by the most trusted source within the subsidiary cluster	70%
Complexity of discovery	The number of steps required in the Network Graph to discover the subsidiary	The minimum distance from the source node within the subsidiary cluster	100%
Parent of Subsidiary	Immediate Parent of the Subsidiary	The "Subsidiary Name" of the identified Parent subsidiary when this information is available	18%

Note: Some coverage ratios are significantly less than 100%, and this reflects available data, for example as the LEI covers only a subset of firms, generally limited to those engaging in financial transactions, it cannot be expected to give complete coverage.
Source: MEIP.

Stage 7: Validate the Physical Register

43. As previously mentioned, the average network for each MNE within MEIP contains roughly 1,800 vertices with 250 company clusters, and a large number of these vertices are only discovered after a large number of edges have been searched. It is therefore necessary to consider how to validate this information to ensure data quality.

44. The primary objective of the validation approach in MEIP is to find a path from the source vertex to all affiliate clusters which is of reasonable quality. A judgement is made about the quality (or reliability) of each edge depending on the source and the relationship type to determine a list of sources and relationship types which are considered of high enough quality to be automatically validated. Table 6 provides an extension to Table 4 to determine which relationships are automatically validated. The general rule is that 'Equivalent Edges" and ownership relationships defined by authoritative sources (Annual Reporting, GLEIF and some business registers) are validated. In total roughly 10% of edges are determined to be of high enough quality to automatically validate. The remaining require further validation checking. However, these can be validated through a combination of already automatically validated links which verify a path (i.e. if B to C is already validated, validating A to B also validates the A to C relationship).

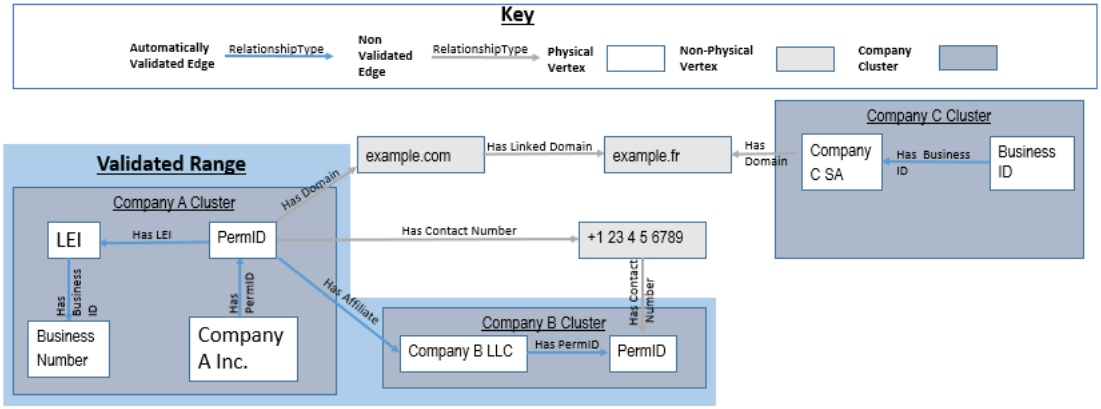
Table 6. Summary of edges within MEIP by validation type

	Source	From Vertex Type	Relationship Type	To Vertex Type	Count in MEIP	Automatic Validation
Stage 1	Annual Reporting	Business Name	Has Parent	Business Name	4311	Yes
	Annual Reporting	PermID	Has Affiliate	Business Name	84334	Yes
	Annual Reporting	PermID	Has Parent	Business Name	1347	Yes
Stage 2	PermID	PermID	Has Domain	Domain	1234537	No
	PermID	PermID	Has LEI	LEI	1734700	Yes
	GLEIF	LEI	Has Affiliate	LEI	1247968	Yes
	GLEIF	LEI	Has Parent	LEI	111860	Yes
	Business Registers	Business ID	Has Affiliate	Business ID	427896	Depends
	Business Registers	Business ID	Has Domain	Domain	3194747	No
	Selected National Sources	Business Name	Has Business ID	Business ID	230771	Yes
	Selected National Sources	Business Name	Has Domain	Domain	625085	No
	Common Crawl	Domain	Has Linked Domain	Domain	1419023	No
	X.509 Certificates	Domain	Has Business ID	Business ID	74527	No
	X.509 Certificates	Domain	Has Business Name	Business Name	26224544	No
	WikiData	WikiID	Has Affiliate	WikiID	115000	No
	WikiData	WikiID	Has Business ID	Business ID	37458	No
	WikiData	WikiID	Has Domain	Domain	202749	No
	WikiData	WikiID	Has LEI	LEI	34226	No
	WikiData	WikiID	Has PermID	PermID	3673	No
	WikiData	WikiID	Has Social Media	Social Media	127328	No
	OpenStreetMap	OSM	Has Domain	Domain	3733779	No
	OpenStreetMap	OSM	Has Social Media	Social Media	228602	No
	OpenStreetMap	OSM	Has WikiID	WikiID	1819563	No
	Crunchbase	Crunchbase ID	Has Domain	Domain	2562554	No
	Crunchbase	Crunchbase ID	Has Social Media	Social Media	4246598	No
	PeopleDataLabs	Domain	Has Social Media	Social Media	9990021	No
	WebDataCommons	Domain	Has Social Media	Social Media	3125048	No
	Name Matching	Business Name	Has Business ID	Business ID	2243593	Yes
	Name Matching	Business Name	Has PermID	PermID	372944	Yes

Source: MEIP.

45. The source vertex is assumed to be the PermID for Company A, as all "Equivalent Edges" are automatically validated, this means that the contents of the cluster for Company A are deemed to be valid (Figure 7). The path from the PermID for Company A to the Business Name Company B LLC is of sufficient quality as it is sourced from Annual Reporting, therefore the cluster for Company B is deemed to be valid as well. However, the path from Company A to Company C can only be reached via digital sources which are not of sufficient quality to validate automatically, as a result the relationship to Company C needs further investigation before it can be included in MEIP.

Figure 7. Determining the scope of validation in MEIP



Source: MEIP.

46. For each company not currently validated two options are possible:

- **Valid:** Company is manually added into the validation scope from the perspective of the parent company. In the example above, if Company C was manually determined to be a child of Company A this relationship would be validated.
- **Invalid:** Reasons for how the link has been determined are investigated, and the cause of the issue is resolved. This normally consists of deleting an incorrect edge, or changing the direction of the edge to reflect a Joint Venture which was not declared in the raw data sources. In the example above, if Company C was determined not to be a child of Company A the paths for determining this relationship would be investigated. A common case is that Company C may provide services to the domain example.fr, and therefore this relationship should be removed.

47. Validation is completed when there does not exist a physical company which has not been validated, Step 5 and Step 6 are repeated and the process continues until no further validations are needed. In total for the MEIP FY 2022 release, roughly 30 000 companies were validated. Given the fact that MEIP FY 2022 consists of over 120 000 subsidiaries, this means that validation was necessary for around 25% of affiliates.

48. The quantity of data validation necessary for MEIP requires a team of data validators, and in order to ensure consistency and reproducibility, each data validator rigorously followed the same process using a purpose-built interface.

49. In order to validate a company, each validator performed desk research attempting to locate two independent and up to date sources confirming the association between legal entities. Due to the extremely large workload, this rule was relaxed in cases where the connection was obvious (e.g. self-explanatory name). The validation team had to employ personal judgement when performing desk research regarding the reliability of each source. However, in most cases validations were achieved using sources which the validators deemed of high quality, such as company websites or online business directories and registers.

50. Certainty was unattainable in a limited number of cases. In these cases, the data point was allowed if it was deemed highly probable that it was part of the MNE group, and metadata associated with the validation explaining the decision process that had been applied was documented, with the aim to make these decisions publicly available in future enhancements of the systems. In general, limited confidence in the relationship was attributable to one of the following reasons:

- **Language**: A lack of knowledge of local languages. In some cases, despite the use of a translation service sufficient information could not be found to validate relationships.
- **Non-existence of information:** In some cases only the data available within MEIP detected the relationship between the parent and associated affiliate. This was particularly applicable to special purpose entities located in jurisdictions with less stringent reporting frameworks.
- **Issues following corporate events:** Following corporate events it was often difficult to determine how the action had been executed. Examples of corporate events include mergers, spin-offs, divestments and name changes. A common source of error was outdated contact numbers following a corporate event. The boundary between assets and the corporate entity was often difficult to determine, and furthermore resulting liquidations and renaming actions were complex to track.

51. Where invalid relationships were discovered, the source of the error was researched and the invalid relationship (edge) deleted. Excluding generic data errors, frequently occurring errors included:

- **Non-distinct business register numbers:** For a number of jurisdictions a harmonised country level business register does not exist. In these cases, there was the possibility that two businesses held the same identification number. An example of this is in Germany, where each region operates its own business register, as a result a number of businesses (particularly from X.509 certificates) were incorrectly matched. Therefore, business numbers from this jurisdiction where regions were not fully described were excluded.
- **Service providers:** Due to the exploratory way in which MEIP derives relationships between companies when services were provided by external service providers this could lead to false connections (such as Company D in Figure 2). Examples include the provision of a social media profile, use of an external company for email services and provision of website hosting or management services.

52. Company specific errors were also present and these were generally attributable to one of the following:

- **Corporate events:** Corporate events were often not immediately reflected by all sources and therefore edits were necessary to reflect these changes.
- **Joint Ventures:** Joint Ventures were often not declared and handled as regular corporate relationships. In these cases, connections between two MNEs were often present and “Direction” variables needed to be applied to edges to reflect the directional nature of the joint venture.

53. Given each case and decision was unique, comprehensive metadata was recorded to justify the decision process. Complex cases were referred to the most senior member of the validation team for in-depth research. Box 3 discusses one such case and its handling within MEIP.

Box 3. Handling of Cereal Partners Worldwide in MEIP

Cereal Partners Worldwide (CPW) is a joint venture between General Mills and Nestlé, which produces a number of well-known breakfast cereals. As a result, both General Mills and Nestlé should include the subsidiaries for CPW within MEIP.

As Annual Reporting was not always fully disclosive of the joint ventures it was necessary to apply a number of directed validations to subsidiaries within CPW, so that the connections which linked Nestlé and General Mills were removed and the two companies were seen as independent within MEIP.

Stage 8: Generate the Digital Register

54. The network determined in Stage 5 for each MNE is also used to generate the Digital Register which consists of vertices from the network which are of type *Domain.*

55. Each entry in the Digital Register relates to an individual domain belonging to a MNE (Table 7), however this provides no measure of how important each of these domains are to the given MNE. Therefore, the data for each domain are extended by joining two measures of global importance at the domain level from the Common Crawl Project[17] and a measure of popularity from Tranco.[18] The first measure, Page Rank looks to reflect the percentage chance that a random web user (randomly clicking links) has of being on a given domain. The second measure, Harmonic Centrality, looks to reflect a measure of the distance that other domains are from the given domain.

[17] Accessible via: https://data.commoncrawl.org/projects/hyperlinkgraph/cc-main-2022-23-sep-nov-jan/index.html.

[18] Accessible via: https://tranco-list.eu/list/X5ZKN/1000000.

Table 7. Variables within the Digital Register

	Description	Calculation Method	Coverage
Parent MNE	Name of the Parent MNE	The name of the source vertex used for determining the Group Network	100%
Domain	Domain controlled by the Parent MNE	The name of digital vertex	100%
TLD	The Top Level Domain (TLD) of the Domain	The top level domain is extracted by removing the SLD. For example, example.com becomes .com and example.fr becomes .fr	100%
Jurisdiction	Jurisdiction of the TLD	A mapping of the TLD to a jurisdiction where it is geographically defined. For example.fr belongs to France, but .com has no mapping to a geographic location (explaining the coverage well below 100%). Exceptions are made for commonly misused TLDs such as .co and .io	64%
Tranco Rank	Website Popularity measure	The Rank determined by the weighted average of 3 online ranking providers smoothed over a monthly period	21%
Page Rank Value	The Page Rank of the domain	Iteratively calculated where: $PR(x) = \sum_{y \in B_x} \frac{PR(y)}{L(y)}$ Where PR(x) represents the Page Rank of domain x, B_x represents the set of all domains linking into Domain x and L(y) is the number of links from domain y	82%
Page Rank Position	The rank of the domains Page Rank relative to all domains covered by the Common Crawl	Global Ranking of Page Rank Value	82%
Harmonic Centrality Value	The Harmonic Centrality of the Domain	Calculated by: $H(x) = \sum_{y \in C_x} \frac{1}{D(x,y)}$ Where H(x) represents the Harmonic Centrality of domain x, C_x represents the set of domains where there exists a path from x and D(x,y) represents the minimum distance path between domain x and y	82%
Harmonic Centrality Position	The rank of the domains Harmonic Centrality relative to all domains covered by the Common Crawl	Global Ranking of Harmonic Centrality Value	82%
Domain Distance	The number of steps required in the Network Graph to discover the domain	The minimum distance from the source node to the domain	100%

Source: MEIP.

Stage 9: Generate the Monitor

56. The network determined in Stage 5 for each MNE is also used to generate the monitoring tool. The tool is based on analysing media coverage on a daily basis in order to identify events of significant interest relating to each MNE within MEIP.

57. To identify search terms for media coverage for each MNE the vertices which are of type WikiData are extracted from the network graph. For each WikiData the corresponding Wikipedia articles are extracted and media coverage is then searched for mentions of concepts which match with the Wikipedia articles. Information on media coverage is extracted from the Global Database of Events, Language and Tone (GDELT)[19] and the structured Global Entity Graph.[20]

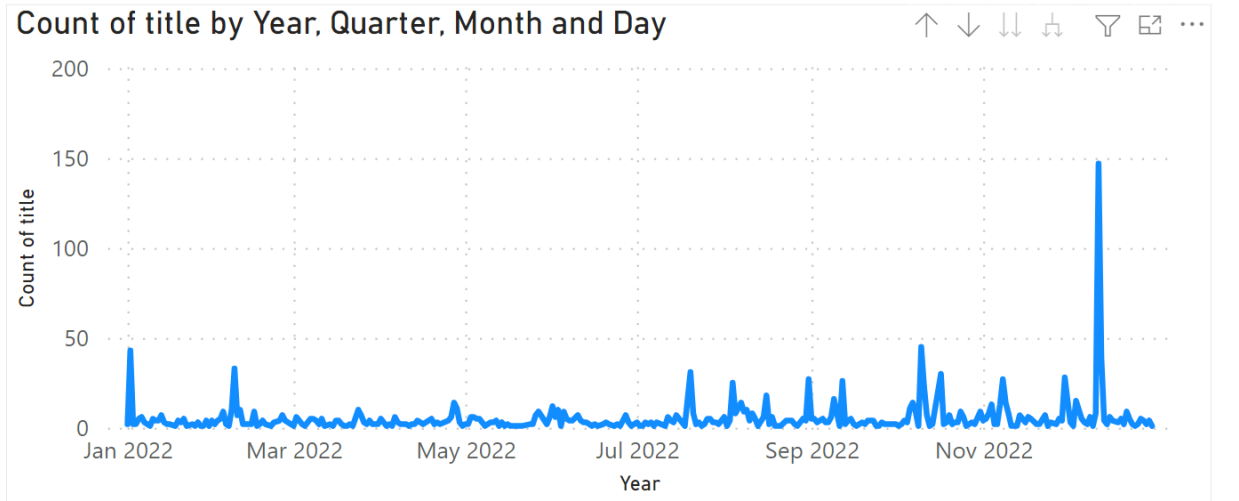
58. Once news articles have been determined to belong to a given MNE, information surrounding the headline is then extracted from the Global Embedded Metadata Graph[21] and presented in a dashboard for further analysis. Box 4 gives an example of how this dashboard can be used to determine a significant event for a given MNE.

Box 4. Example of analysing a corporate event using the MEIP monitoring tool

Amgen to buy Horizon Therapeutics

On 12th December 2022 the MEIP Monitor detected an unusual high level of media coverage for Amgen (a US based pharmaceutical company). Almost 150 articles were discovered on this day, which is far above the usual activity expected (Figure 8).

Figure 8. Unusual media activity for Amgen



Source: MEIP.

[19] More details available at: www.gdeltproject.org/.

[20] More details available at: https://blog.gdeltproject.org/announcing-the-global-entity-graph-geg-and-a-new-11-billion-entity-dataset/.

[21] More details available at: https://blog.gdeltproject.org/announcing-the-global-embedded-metadata-graph/.

Given this unusual level of activity, we can use the MEIP monitor to manually investigate the top headlines (Table 8) for the given day to determine whether the event was significant (i.e. provided an indication of a significant corporate restructuring, which would have an impact on official macro-economic statistics).

Table 8. Top headlines for Amgen (12th December 2022)

	Source
Amgen Delves Further into Immune Therapies with Horizon Deal	www.msn.com/en-us/money/companies/amgen-delves-further-into-immune-therapies-with-horizon-deal/ar-AA15boCe
Horizon Therapeutics to be acquired by Amgen for $28B	www.chicagotribune.com/business/ct-biz-amgen-acquiring-horizon-pharmaceuticals-20221212-a3vorjyiczexlfiuypsbbszfuq-story.html
Cramer shares his outlook on Amgen after the Horizon Therapeutics deal	https://invezz.com/news/2022/12/12/amgen-to-buy-horizon-therapeutics/
AMGEN to host webcast investor call following announced acquisition of horizon therapeutics	www.prnewswire.com/news-releases/amgen-to-host-webcast-investor-call-following-announced-acquisition-of-horizon-therapeutics-301699964.html
AMGEN to host webcast investor call following announced acquisition of horizon therapeutics	www.prnewswire.com:443/news-releases/amgen-to-host-webcast-investor-call-following-announced-acquisition-of-horizon-therapeutics-301699964.html
Amgen to Buy Horizon Therapeutics for $27.8 Billion	www.marketwatch.com/articles/amgen-horizon-therapeutics-acquisition-stock-price-51670843416
Amgen to Buy Horizon Therapeutics in $27.8 Billion Deal	www.msn.com/en-us/money/companies/amgen-to-buy-horizon-therapeutics-in-dollar278-billion-deal/ar-AA15boCe
Amgen to buy Horizon Therapeutics in year's biggest biotech deal	www.biopharmadive.com/news/amgen-horizon-acquire-biotech-drug-deal/638485/
Amgen to buy Horizon Therapeutics in year's biggest biotech deal	www.healthcaredive.com/news/amgen-horizon-acquire-biotech-drug-deal/638535/
NewsNow: Amgen news | Breaking News & Search 24/7	www.newsnow.co.uk/h/Industry+Sectors/Biotech/Amgen

Source: MEIP.

Although announced on 12th December 2022, the transaction completed on 6th October 2023 following regulatory approval. This is an indication of a large corporate event that will likely impact on Irish statistics in particular. Horizon Therapeutics are legally headquartered in Ireland, whilst Amgen is headquartered in the United States, and hence this represents a significant foreign investment by a US firm within Ireland.

This tool therefore provides a mechanism to identify potentially major events, but also a tool to aid in the validation and updating of MEIP in the future.

5. Using MEIP as a framework for MNE analysis

59. Information collected in MEIP allows to examine specific issues related to MNEs and their global network in details. This section provides a few examples from gender diversity in senior leadership positions, reactions to global conflicts and pledges on reducing GHGs emissions. This work, which reflects opinions of the authors, and has been provided unedited from the original form, often look to combine MEIP with information available from other sources. Further outputs will be developed as the project matures and expands.

The long road to gender parity in senior leadership positions[22]

60. The effort in pursuing a more gender equal society must be collective and undertaken in many domains. A number of countries are now adopting targets in an attempt to improve representation among directors of listed companies, one such example is the 2022 European Parliament "Women on Boards" directive, which addresses the need for transparency in the hiring procedures of listed companies, and sets a goal of attributing at least 40% of non-executive director posts (or 33% of all director posts) to women, but took over a decade to conclude and has an implementation target of June 2026. Furthermore, if the rate of progress found by MSCI in recent years continues it will take over a decade to achieve this target globally.

61. When MEIP is paired with data from the commercial database Orbis, which details information on senior leadership, it is possible to create indicators on the progress individual MNEs are making on gender diversity. The resulting dataset allows for breakdowns by industry, country of incorporation, age and position.

Parity remains out of reach, but geographical and industry differences exist

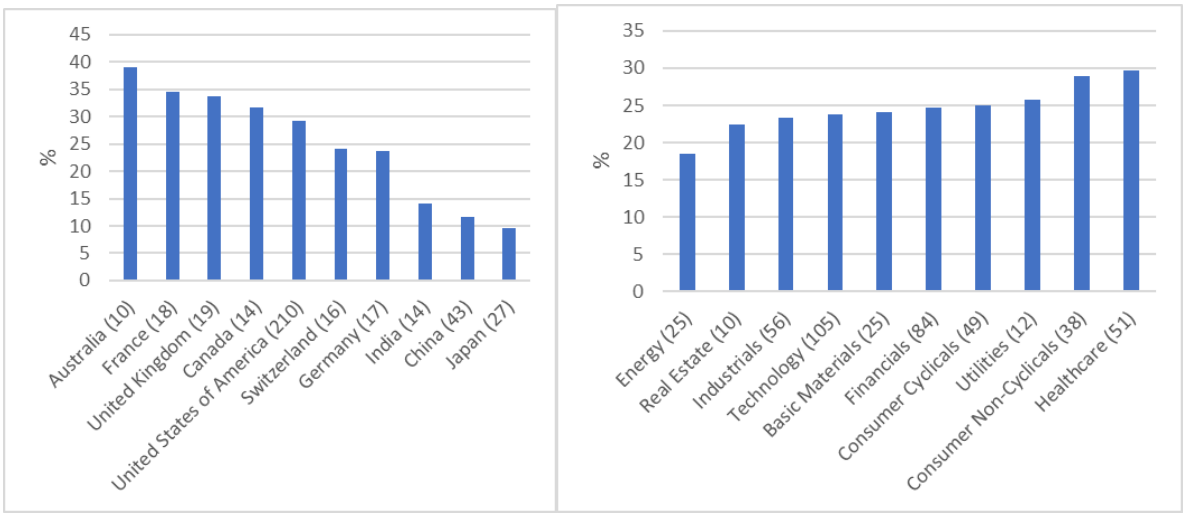
62. The average firm within MEIP has 24.9% of senior leadership roles held by women, which, albeit below the desired threshold, represents an improvement from previous years.

63. Looking at industries, as classified by the Refinitiv Business Classification (TRBC) system, the highest share of women is in the healthcare (29.7%). Of note, there is a persistent under-representation of women in the energy industry (18.5%), the lowest across all industries covered, an industry historically male-dominated as also detailed in the IEA gender employment gap.

64. Focussing on countries, Australia, France, the United Kingdom and Canada all show a share of women in senior leadership roles of above 30% (Figure 9). However, rates are far lower in India (14.2%), China (11.7%) and Japan (9.5%). Seven companies within the sample also have no women in senior leadership positions, with those companies in China (4), Japan (1), the United Arab Emirates (1) and Saudi Arabia (1).

[22] Based on the analysis by Michela Gamba and Graham Pilgrim.

Figure 9. Average share of women in senior leadership positions by country (left) and industry (right), 2022

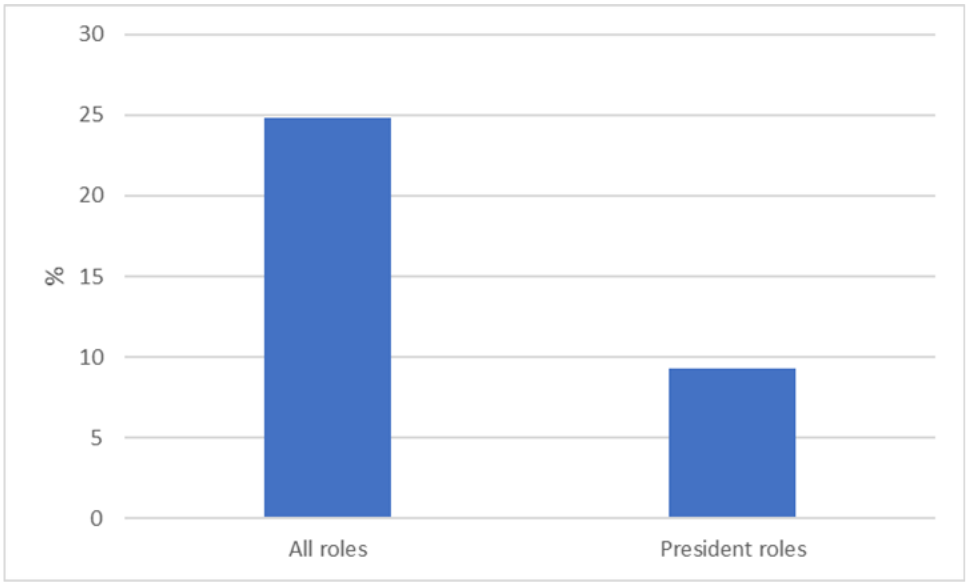


Note: Number of companies shown in brackets countries within under 10 companies are excluded. The findings at an industry and country level are broadly consistent with previous literature.
Source: MEIP.

Senior leadership has an additional glass ceiling

65. Whilst the average share of women lies at 24.9%, as the seniority of the role increases participation rates tend to drop (Figure 10). This is clear in particular for prominent roles such as President and Vice-President roles within the Board of Directors where the rate plummets to 9.3%, showing that there is an additional glass ceiling to access the more prominent leadership roles.

Figure 10. Share of women in president roles, 2022

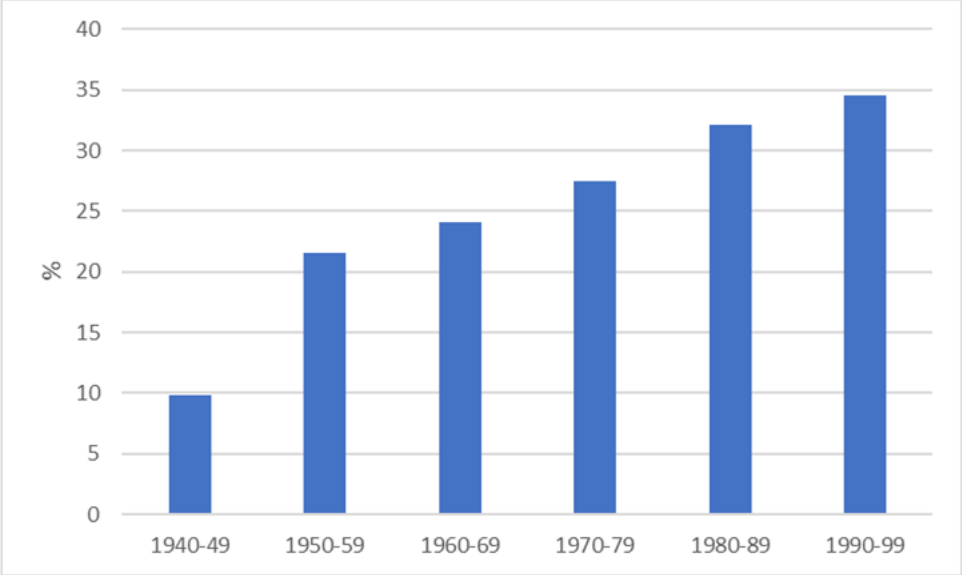


Source: MEIP.

Greater diversity with younger cohorts, showing steps in the right direction

66. Whilst there is a persistent under-representativeness of women across all decade cohorts, those born more recently are seeing a picture which is closer to parity (Figure 11). For example, 34.5% of those holding a board role born in the 90s are women, whereas this reduces to 21.6% for those born in the 50s. Although, a portion of this trend could also be explained by the "motherhood penalty", when career advancements differ when women start having children.

Figure 11. Share of women in senior positions by decade of birth, 2022



Source: MEIP.

67. The data-driven insights provided by MEIP combined with board member data from Orbis reveal that further steps are still necessary to achieve gender parity within some of the largest companies. Not one single industry or country on aggregate has exceeded a representation of above 40%, with the more senior roles being even less diverse. There are signs of improvement within younger cohorts, and with time as these cohorts make up a larger share of senior leadership positions the representation of women will increase – however even the 90s cohort has an aggregate representation below 40% showing more action is necessary.

To leave or not to leave: How are the world's largest MNEs responding to the war in Ukraine[23]

68. Russia's war of aggression against Ukraine has triggered unprecedented economic sanctions with potentially far-reaching implications for doing business in Russia. Difficulties making international payments, reduced access to foreign capital, and logistical challenges are likely to disrupt local operations of companies with a global footprint. Mounting reputational pressure and volatile market conditions may put additional strain on the ease of doing business in the country. These practical challenges, along with ethical considerations, have prompted a large number of multinationals to curtail their activity in Russia.

[23] Based on the analysis by Polina Knutsson.

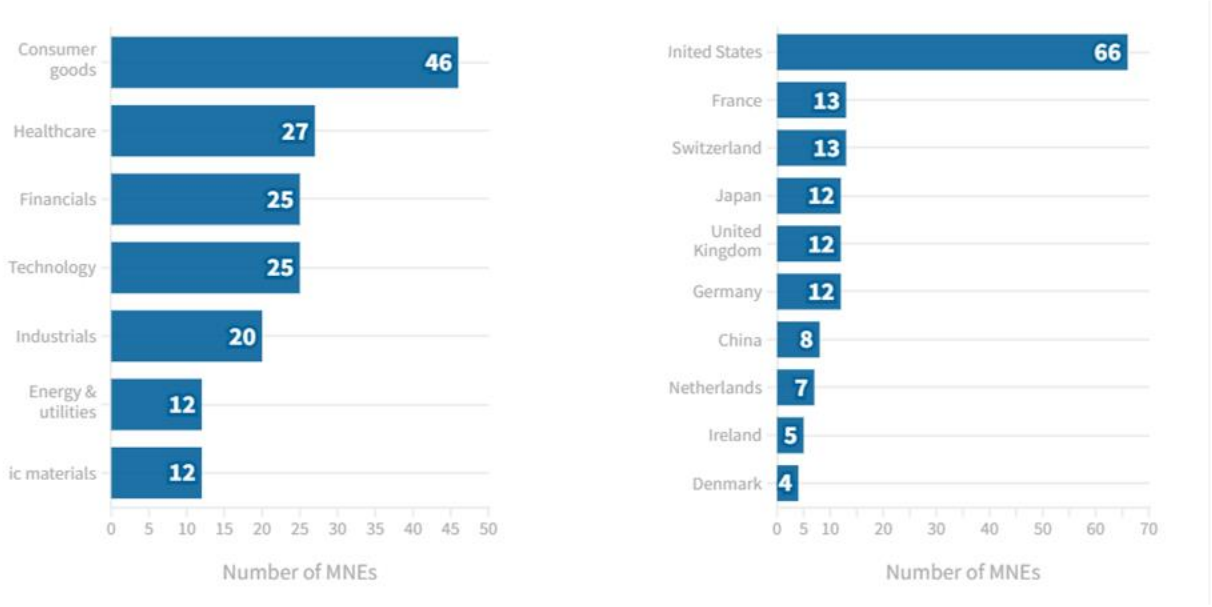
Yet, according to information collected in MEIP, the economic impact of exiting the Russian market is limited for many foreign MNEs, since most of them have only a few affiliates in the country or derive less than 2% of total revenues from Russia.

Many MNEs used to do business in Russia

69. In 2020, 173 MNEs out of the world's largest 500 covered by MEIP had a physical presence in Russia, of which only 6 are Russian-owned, with the remaining 167 being foreign with at least one affiliate in the country. Most of the foreign multinationals operate in either consumer goods or healthcare sector (Figure 12, left). Over one-third are headquartered in the United States, with many other enterprises coming from Europe and Asia (Figure 12, right).

Figure 12. Distribution of MNEs by sector (left) and headquarter country (right)

Per cent



Source: MEIP.

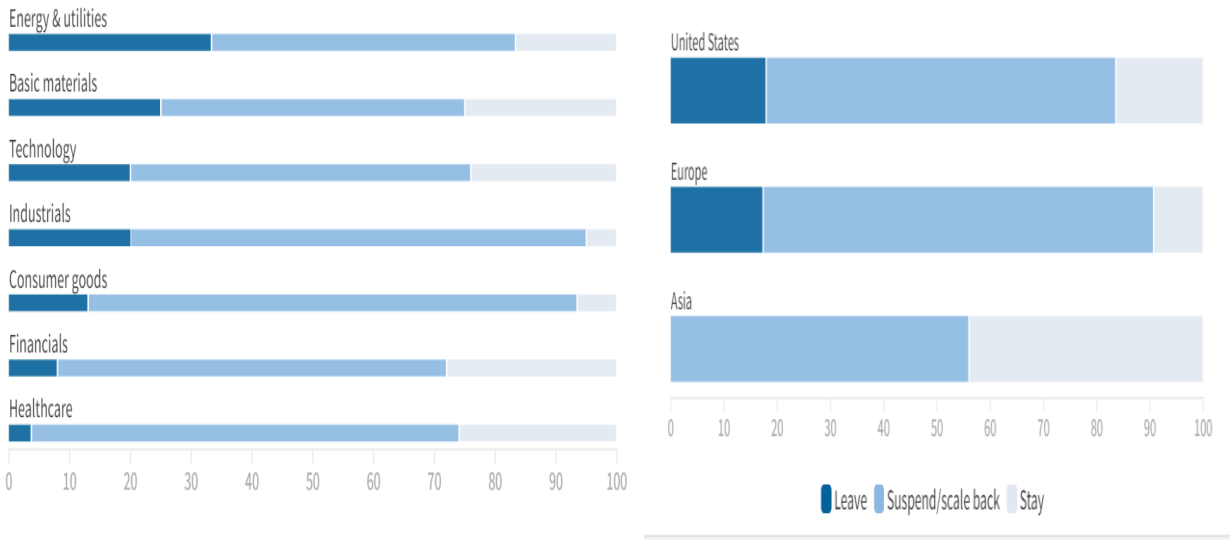
The war has triggered a large-scale response

70. By 29th July 2022, 25 foreign MNEs covered by MEIP had announced their exit from Russia and 113 suspended or scaled down their operations in Russia, while 29 continued business as usual, according to data collected by the Yale School of Management, the Kyiv School of Economics (KSE) institute and the LeaveRussia initiative.

71. The response by sector was uneven (Figure 13, left). Nearly one-third of energy and utilities companies and one-fifth of technology and industrial MNEs announced their complete withdrawal from the Russian market. On the contrary, less than 10% of finance and healthcare companies withdrew completely, and more than 25% continued their operations, a larger share than in other sectors. Most consumer goods multinationals decided to maintain at least some business in the country, with many of them citing the willingness to meet the essential needs of the local population, such as food and hygiene products. The steps taken by foreign-owned enterprises also varied across different source regions (Figure 13, right). Most MNEs headquartered in the United States or Europe have announced their withdrawal or reduced

their operations in Russia since the onset of the war, whereas close to 50% of Asian multinationals did not signal any intention to halt their activity.

Figure 13. The response to the war varied by sector (left) and headquarter region (right)

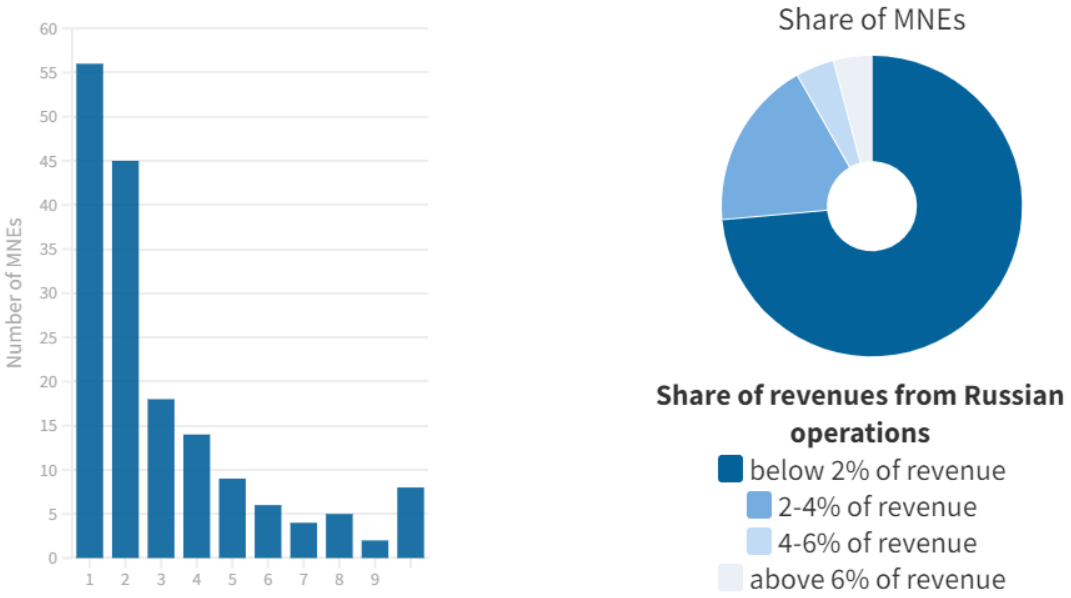


Source: MEIP.

In exiting Russia, many MNEs are leaving behind only a small fraction of their global activity

72. For many multinationals, pulling out of Russia entails losing a relatively small share of their global activity. According to MEIP, most foreign MNEs have few affiliates in the country: a third have only one affiliate, a quarter have two (Figure 14, left). More importantly, the Russian market accounts for a small fraction of total revenues for many enterprises. Estimates based on affiliate-level data suggest that the majority of MNEs covered by MEIP derive less than 2% of total revenues from Russia and less than one-fifth generate between 2 and 4% (Figure 14, right).

Figure 14. Most foreign MNs have few affiliates (left) and generate a small share of revenues (right) in Russia



Source: MEIP.

Unveiling the emission reduction plans of multinational enterprises[24]

73. As the world faces the hottest year on record, national and regional policies to reduce emissions have become a fixture in the public discourse. But for reductions to be achieved, it is important to understand the commitments of the world's largest companies. Much of the burden in reducing emissions is shouldered by multinational enterprises (MNEs), which span borders and whose operations play a substantial role in global emissions.

74. One of the leading frameworks for corporate target setting is the net-zero standard, developed by the Science Based Targets initiative (SBTi), and providing standardised guidelines and requirements for companies to achieve a net-zero GHG emissions target. Like countries, companies are committing at different levels over different timelines, but it can be difficult to find information on companies' progress.

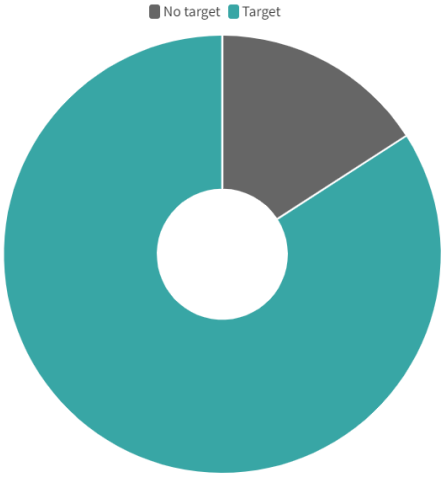
75. When paired with data from the net-zero tracker database, which details the climate reduction plans of the world's largest companies, it can be used to uncover new insights on the emission-reduction progress being made by MNEs as of 2022. Matching these data with the company ID of MNEs in the net-zero tracker database, a total of 415 of companies remained in the dataset. The resulting dataset allows breakdowns by industry, target date, percentage of emissions reduction and scope of reduction.

Most of the world's largest MNEs are making commitments

76. A large majority of MNEs express an interest in cleaning up their operations in the fight against climate change. In fact, 84% of the companies in the dataset have put forth specific greenhouse gas (GHG) emissions reduction targets as of 2022 (Figure 15). However, not all targets are created equal. Breaking this data down by target category and ambition level can help us get a clearer understanding of how much can be expected from multinationals.

[24] Based on analysis by Rodrigo Pazos and Graham Pilgrim.

Figure 15. 84% of MNEs have specified emission-reduction targets

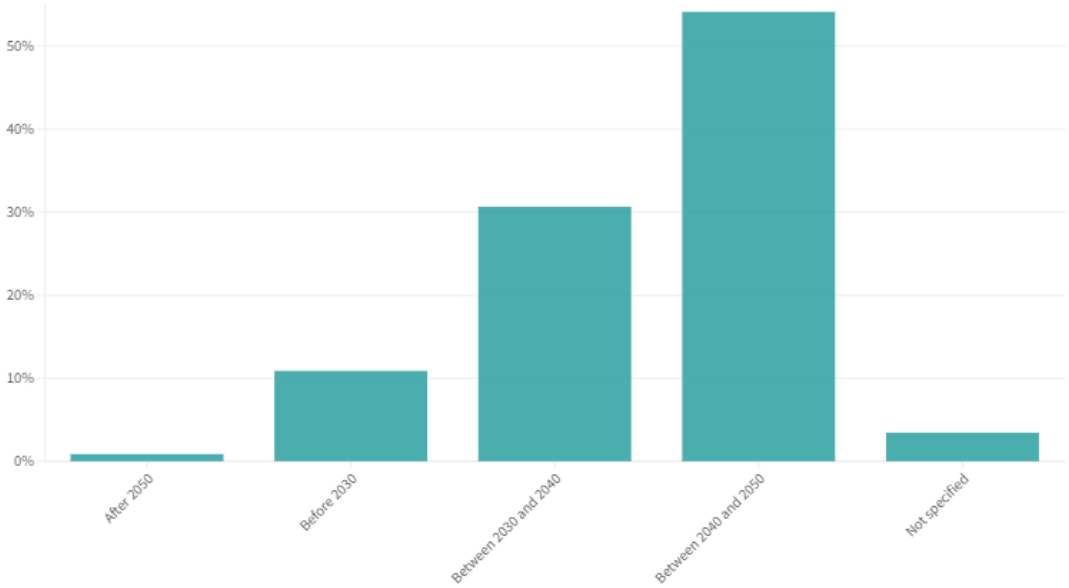


Source: MEIP.

When will these commitments be delivered?

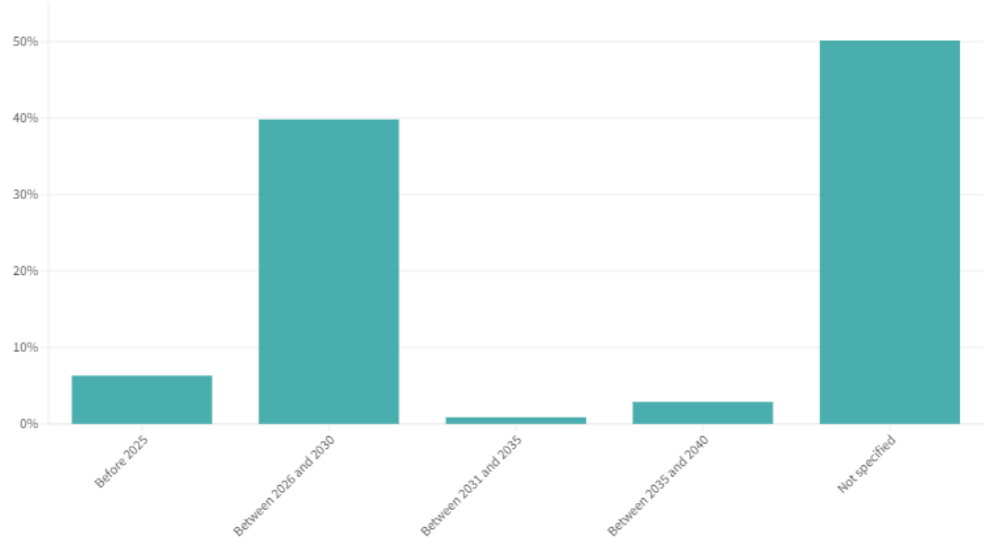
77. While most MNEs have declared emissions reduction targets, deadlines differ between individual companies (Figure 16). These deadlines can be divided into two categories: end reduction targets, and interim reduction targets. Of the companies that have an end reduction target, most aimed for a date between 2040 and 2050. Of those that have intermediate reduction targets, the target date is usually between 2026 and 2030 (Figure 17).

Figure 16. Percentage of MNE that declared end reduction targets



Source: MEIP.

Figure 17. Percentage of firms that declared an interim target



Source: MEIP.

78. While natural to assume that an intermediate target must be set with reference to a specified end target, this is not always the case. In fact, more companies communicate on intermediate targets than end targets, at least according to the available data.

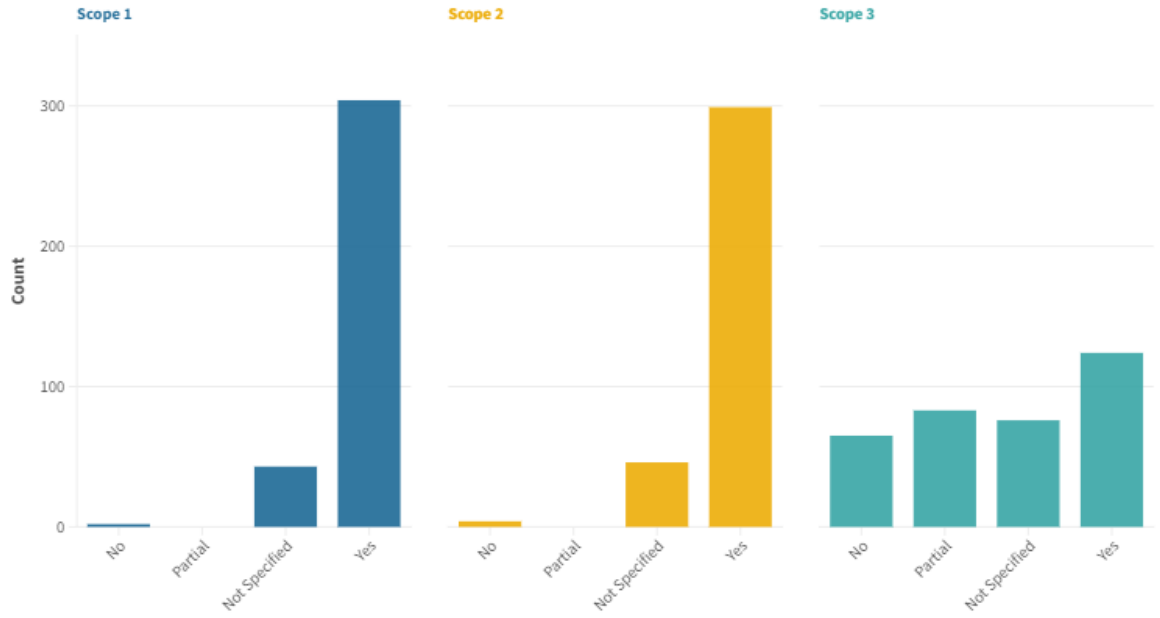
Which emissions and by what means?

79. As well as variation in the timelines and extents of targets, the method for measuring these goals also vary. Companies can address these goals by reducing emissions within three scopes, which set out a way to divide greenhouse gas emissions:

- Scope 1 covers the emissions resulting directly from the company's operations.
- Scope 2 covers the "indirect" emissions created by the production of energy that companies consume, such as electricity used to heat the company's buildings.
- Scope 3 includes both upstream activities (those produced by suppliers) and downstream activities (those produced by the consumers through normal use of the company's products).

80. For scopes 1 and 2, measurement remains within the company's influence and therefore reduction policies are easier to implement. Scope 3 requires a knowledge of supply chains and product usage and is therefore much more difficult to quantify. There were around 300 companies that declared reductions targets for scope 1 and 2 emissions, while this reduces to around 200 companies when considering scope 3 (Figure 18).

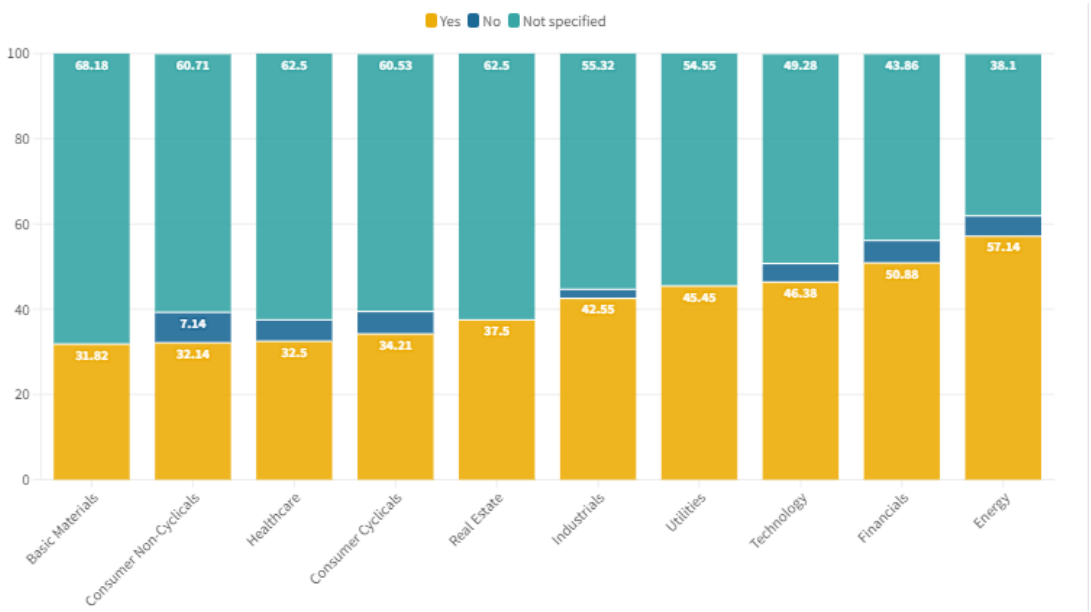
Figure 18. Emission reductions by scope



Source: MEIP.

81. In addition to reducing emissions directly and indirectly throughout their operations, companies may also achieve their climate commitments via carbon credits or offsets. Carbon credits are tradeable permits that can be earned by companies by either reducing their own emissions, investing in projects for others to reduce, or removing emissions from the atmosphere. Within this dataset, less than half of the companies are planning to use carbon credits. The use of carbon credits also varies across industries, as classified by the Refinitiv Business Classification (TRBC) system, being least popular with the basic materials industry (32%) and most popular within the energy industry (57%) (Figure 19).

Figure 19. Use of the carbon credit system by industry



Source: MEIP.

82. The data-driven insights provided by MEIP combined with the net-zero tracker database reveal that a majority of world's largest MNEs are making commitments to reduce their emissions. However, there are substantial differences in companies' commitments on achieving reductions in emissions, both in terms of timing, ambitions and tools used.

6. Improving MEIP through outreach

83. Although the tools currently developed provide good coverage of MNE activities they cannot guarantee exhaustiveness. Strategic outreach through a number of channels is necessary to further improve the coverage of MEIP.

Outreach with National Statistical Offices

84. Adopting the MEIP framework as a data source for information on MNEs is the first step a number of national statistical offices (NSOs) can take to help in the profiling of firms (including for surveys). This can help to improve the quality of foreign affiliate trade statistics, balance of payments statistics and modes of supply statistics in countries.

85. The OECD and UNSD support more active engagement with NSOs to formalise the use of MEIP in national statistical infrastructures but also as a global resource. Countries are encouraged to:

- Compare the scope of MNE data (i.e. coverage of units) they currently have and use with that of the MEIP;
- Consider the possibility of integrating business register data (which is often not confidential) within the MEIP framework;
- Work with the OECD and UNSD to identify non-disclosure mechanisms to quality-assure and validate MEIP data; and
- Link unit records of their national statistical business registers with MEIP.

Outreach with other initiatives

86. A number of international initiatives can benefit from MEIP.

- The Global Legal Entity Identifier Foundation (GLEIF) is seeking to create a harmonised identification number (LEI) of all entities worldwide. MEIP data are providing a framework for assessing current coverage rates and guiding strategies for improving data accuracy and coverage.
- MEIP's monitoring tool complements Eurostat's Early Warning System (EWS) for large corporate events by providing additional indicators and the ability to simulate the impacts of a company restructuring.

87. The United Nations Statistical Commission at its 54th session in March 2023 endorsed the global initiative on unique identifiers for businesses and encouraged countries and relevant organisations to coordinate their activities in this area to provide solid infrastructure for statistical business registers. Strengthening the administrative business registration in countries and the establishment of unique identifiers for legal entities is considered an important first step in improving the statistical business registers, especially in countries where the administrative data system for businesses is not well established. Linking the global initiative of national unique identifiers and the mapping to global identifiers will facilitate the establishment and maintenance of MEIP containing the legal structure of MNEs.

Outreach with Multinational Enterprises

88. MNEs can help to improve the accuracy of the initiative by engaging with it to validate and enhance current data. Engagement can be beneficial for responding firms, as well as for the OECD, UNSD and the NSOs they work with. Increasing the quality and coverage of data improves our collective understanding of MNEs and their global reach, which is fundamental for many of today's global challenges. It is also a tool and resource that provides valuable information about the companies themselves.

7. Conclusion

89. MEIP provides a flexible framework for developing a number of outputs that allow for the profiling of MNEs in greater detail. The level of detail expands upon existing initiatives, such as providing almost four times more information on subsidiaries than GLEIF. It can support research through the granular nature of the data, which provides a means to analyse individual MNEs, and in particular the scale and nature of their cross-border activities.

90. MEIP can also help NSOs to improve their statistics by providing insights into the scale and complexity of international MNE activity and timely information on any restructurings they undertake. MEIP can be seen as a single reference point, free from confidentiality restrictions that may impede discussions between NSOs; and it features a monitoring tool to help identify significant corporate restructurings.

91. But there is still scope to improve the coverage of the project and the objective is to increase the scope of variables covered at the group level (ideally, employment, investment, turnover or profit), but with estimates at a jurisdiction level. This could be done by collecting more data and/or mapping MEIP with other data source. This will require investing in data storage methods (a move to a native graph database).

92. MEIP is evolving and both the OECD and UNSD seek to continuously develop new methods, find new data sources and IT tools for its improvement. In this regard, the current plan is to:

- Mitigating the complexities of data extraction from annual report PDFs by exploring advanced technological solutions and collaborative initiatives, in order to streamline the extraction process, reduce labour intensiveness, and enhance the overall accuracy and efficiency of data collection.
- Increase the number of MNEs covered by including new sources.
- Develop new indicators to detect forthcoming changes in MNEs and provide analysis on MNE developments more generally.
- Increase the timeliness of the product. This means investing in new data sources and data handling to allow for frequent updates, and further automating via AI tools, particularly with respect to webscraping.
- Provide a framework for others to perform their own profiling work. This means investing in new interfaces, to allow for simple access and use of our system.

References

Bundesanzeiger (2023), website providing data on German MNE subsidiaries, www.bundesanzeiger.de/, (accessed 29 November 2023).

Common Crawl Project (2023), website detailing the project, https://commoncrawl.org/, (accessed 29 November 2023).

Companies House (2023), website making UK business registry available, http://download.companieshouse.gov.uk/en_output.html, (accessed 29 November 2023).

Companies House (2023), website making UK persons of significant control available, http://download.companieshouse.gov.uk/en_pscdata.html, (accessed 29 November 2023).

Corpwatch (2023), website providing structured downloads of US 10-K Subsidiary filings, http://api.corpwatch.org/, (accessed 29 November 2023).

Crunchbase (2023), website detailing Crunchbase, www.crunchbase.com/, (accessed 29 November 2023).

GDELT (2023), website detailing the Global Database on Events, Language and Tone (GDELT), www.gdeltproject.org/, (accessed 29 November 2023).

GLEIF (2023), website detailing the Legal Entity Identifier (LEI), www.gleif.org/en, (accessed 29 November 2023).

OECD (2023), website detailing the Inclusive Framework on Base Erosion and Profit Shifting, www.oecd.org/tax/beps/, (accessed 29 November 2023).

OECD (2023), website detailing the MEIP initiative with interactive dashboards, https://bit.ly/mne-platform, (accessed 29 November 2023).

OECD (2020), "How to construct a global business register?", www.oecd.org/sdd/its/ADIMA-Methodology.pdf.

OECD (2009), *OECD Benchmark Definition of Foreign Direct Investment 2008: Fourth Edition*, OECD Publishing, Paris, https://doi.org/10.1787/9789264045743-en.

OpenCorporates (2023), website detailing OpenCorporates, https://opencorporates.com/, (accessed 29 November 2023).

OpenStreetMap (2023), website detailing OpenStreetMap, www.openstreetmap.org/, (accessed 29 November 2023).

PeopleDataLabs (2023), website detailing PeopleDataLabs, www.peopledatalabs.com/, (accessed 29 November 2023).

Refinitiv (2023), website detailing the PermID, https://permid.org/, (accessed 29 November 2023).

SEC (2022), *Apple Inc. Significant Subsidiaries from Annual Report 2022*, www.sec.gov/Archives/edgar/data/320193/000032019322000108/a10-kexhibit21109242022.htm, (accessed 29 November 2023).

SEC (2022), *Apple Inc. Annual Report 2022*, www.sec.gov/ix?doc=/Archives/edgar/data/320193/000032019322000108/aapl-20220924.htm, (accessed 29 November 2023).

Sectigo (2023), website detailing open data on X.509 Certificates, https://crt.sh/, (accessed 29 November 2023).

Tranco (2022), website detailing domain popularity on 31st December 2022, https://tranco-list.eu/list/X5ZKN/1000000, (accessed 29 November 2023).

UNECE (2023), website detailing the Task Force on Exchange and Sharing of Economic Data, https://unece.org/statistics/networks-of-experts/task-force-exchange-and-sharing-economic-data, (accessed 29 November 2023).

UNSD (2023), website detailing the Global Groups Register, https://unstats.un.org/unsd/business-stat/GGR/, (accessed 29 November 2023).

Web Data Commons (2023), website detailing the Web Data Commons project, http://webdatacommons.org/, (accessed 29 November 2023).

Wikimedia (2023), website detailing Wikidata, www.wikidata.org, (accessed 29 November 2023).