

26<sup>th</sup> Meeting of the Wiesbaden Group on Business Registers  
- International Roundtable on Business Survey Frames

Neuchatel, 24-27 September 2018

*Fabienne FORTANIER, Diana DOYLE, Graham PILGRIM and Cecilia CALIANDRO*  
*OECD Statistics and Data Directorate*  
*Session No. 5*

*New Data Sources, especially on Enterprise Groups*

**Measuring MNEs using Big Data: The OECD Analytical Database on Individual Multinationals  
and their Affiliates (ADIMA)**

## **1. Introduction**

Multinational enterprises (MNEs) have been key drivers of the increasing international fragmentation of production. MNEs account for nearly 30% of value added and trade, and about one-fifth of employment in OECD countries on average, and global Foreign Direct Investment (FDI) stocks now account for around 40% of world GDP (up from 25% in 2005). In developing countries, inward FDI has become the prime source of external funding, surpassing flows of official development assistance, and have provided important mechanisms for these countries to integrate into global value chains.

At the same time, the measurement of flows related to multinationals and indeed where they are registered is presenting significant challenges for the measurement of GDP. This has led to growing demands for improved statistics that provide insights on the scale and complexity of international MNE activity, and indeed for timely information on any restructurings they may undertake.

However, despite their significant and growing importance, with implications across a range of policy areas, information on Multinational Enterprises (MNEs) remains at best patchy. This is partly a function of complexity: by their very nature, MNEs are large, with a multitude of activities across a number of jurisdictions. However, for firms engaging in fiscal optimisation at least, it is also partly a function of design: some firms for example create elaborate chains of affiliates, holding companies and special purpose entities, designed to minimise taxes, but the consequence is also to obfuscate.

Another factor that complicates the measurement of MNEs is the limited possibility for National Statistical Institutes (NSIs) to obtain a holistic view of their activities, reflecting legislation that typically restricts data collections to activities within their economy or (and only very rarely) to the global activities of firms headquartered in the economy (and even in these cases it is not clear that the coverage of the MNE's activities is exhaustive).

The sharing of data across countries could provide a window to provide this holistic view but legal constraints aimed at preserving confidentiality and privacy of respondents within national borders in most countries mean that this is not, at least for now, possible.

To begin to address these challenges, the OECD has begun to develop an analytical database of individual MNEs and their affiliates (ADIMA), by compiling publicly available statistics on the scale and scope of the international activities of MNEs, thus providing a unique 'whole of the MNE' view.

## What is ADIMA?

ADIMA aims to cover 100 of the largest global MNEs (see Annex A) by the end of 2018, to be extended towards 500 by 2020. These sample sizes<sup>1</sup> were selected to achieve a good balance between relevance (for example, the top 100 or 500 MNEs account for a significant share of FDI<sup>2</sup>) and feasibility (in terms of resources), with MNEs selected according to the following criteria:

- **Revenues:** ranked by total revenues for (currently) 2016.
- **Stock listing:** As shown below, publicly available Financial Reports and Accounts are a key data source for ADIMA, hence a listing on a stock exchange is currently a pre-requisite for enterprises to be included; although future efforts will explore the possibility of including MNEs that may not be listed, including state-owned enterprises, given their growing importance.
- **Multinational.** Only enterprises that met the internationally accepted definition of being a multinational were selected – i.e. to have an operating affiliate in at least one country outside the enterprise's home country.
- **Online presence.** Since webscraping and text analytics of company webpages are an important tool for ADIMA, a significant online presence is required to allow for a meaningful analysis.

It is important to note that ADIMA has two primary functions. The first is to develop a view of MNEs that transcend national borders, which is essentially how MNEs operate: in other words to develop macro-economic statistics of MNEs that present a global view.

This, to a large extent, reflects the emerging view, (mainly because of the ability of the MNE to provide intellectual property and, more broadly, knowledge based services across affiliates), that the production function of an MNE can only meaningfully be assessed through a global perspective.

But, that is not to say that the 'national' within the 'global' is not important. Despite the measurement challenges of identifying the use and ownership of knowledge based services within an MME, there is considerable interest, from an analytical perspective, in understanding how MNEs are organised across borders; for example, for the analysis of global value chains or to have a firm level view of the direct impact of FDI in the invested economy.

But, obviously, there is also significant interest from an NSO perspective too. Whilst NSOs in many countries make significant efforts to identify foreign owned affiliates in their economies using a variety of (and increasing) sources, (such as the Euro-Group Register) there, to date, does not exist a single international reference point that allows NSOs to identify affiliates in their countries in an exhaustive nor consistent manner (for example, country A may ascribe ownership to country B in their inward FATS but the Country B may not recognise ownership in its outward FATS). This is of particular relevance for timely statistics, where restructurings may have a significant impact on macro-economic statistics (as illustrated in the recent Irish case).

Acting as this single reference point is the second primary objective of ADIMA.

ADIMA attempts to meet these goals through the development of three distinct (but related) outputs:

---

<sup>1</sup> Such sample sizes also align well with those currently chosen by large case units (LCUs) in Statistical Offices (the long-established LCU in Statistics Netherlands for example covers ~300 enterprises, while the Italian and Irish LCUs cover 140 and 60 enterprises respectively).

<sup>2</sup> Rugman (2005) for example showed that a relatively small set of MNEs (~500) was responsible for the overall majority (80 percent) of global FDI and little seems to have changed since his analysis, as the sum of foreign assets of the 100 largest MNEs equals was estimated at 31% of global FDI in 2016.

## A Register of parent and affiliate relationships

This dataset includes, for each parent MNE and all its affiliates, a variety of often-used enterprise identifiers, as well as information on the address (city, country) and industry of all affiliates. For the parent enterprise only, additional information is included on its year of incorporation (to establish MNE age), and consolidated industry classification (following ISIC Rev4), to allow for comparisons between MNEs within the same industry and align with industry classifications used in e.g. outward FATS statistics.

## Economic indicators

The second component of ADIMA is a set of indicators that describe the economic structure and performance of MNEs via a selected set of variables reported in Balance Sheets and Income Statement at the level of the MNE (see Table 1).

To capture the extent of MNEs' international activities, ADIMA also includes geographical segment information (for e.g. sales, assets, profits, employment) as reported by MNEs, using estimations derived from Big Data sources.

## Monitoring tool

The Monitoring tool is intended provide a timely flow of information on MNEs' restructuring to aid the work of national compilers, again by exploiting the various innovative options that Big Data analytics provides, and collaborative efforts with Eurostat are on-going to integrate ADIMA into Eurostat's *Early Warning System*: an information and data sharing system to identify organisational changes that may impact on macro-economic statistics.

Table 1 below provides an overview of the key variables and indicators included in each of the three components of ADIMA.

**Table 1. Variables included in ADIMA: Register, Indicators and Monitor**

Register	Indicators	Monitor
All enterprises (parents + affiliates):	Consolidated balance sheet	Changes in register variables, based on text analytics of unstructured big data sources (e.g. Open Street Map, GDELT, MNE IR and Jobs websites), classified by MNE, country and industry
Enterprise Identifiers (ISIN, LEI, CIK)	Assets (current; PPE; intangibles; ...)	
Address/country of operation	Liabilities (current; long-term)	
Industry (ISIC rev 4)	Equity	
Immediate and ultimate parent IDs	Consolidated income statement	
Demographic events + dates	Revenues/turnover	
Ultimate Parents only:	Expenses (COGS, SG&A, subcomponents)	
Consolidated industry	Profitability (EBIT, EBITDA, net income)	
Year of incorporation	Employment	
Country of global decision centre	Share of int'l sales, assets, employment, ... Country-level sales, assets, employment, ... Register-derived indicators (# affiliates, # countries, spread) Entropy measures of internationalisation	

## How is ADIMA constructed?

### *Data sources and methods*

To date, there is no publicly available database of corporate ownership linkages at a global level. At the national level, several countries have started to share firm-level records with a view to further their open data agenda. For example, INSEE (France) has opened its Sirene database<sup>3</sup> and in the United Kingdom, individuals and legal entities with significant control of UK companies should be identified on the register of Persons with Significant Control.<sup>4</sup> It is expected that more countries will follow, particularly in Europe, in response to EU regulation<sup>5</sup>, and, so, the pool of open data sources will progressively expand. However, at the time of drafting, the available datasets represented only a fraction of the global networks that Multinationals have built over the years.

ADIMA combines traditional data sources, including commercial databases and company Annual Reports, with newly emerging sources (such as the Legal Entity Identifier (LEI)), and innovative data collection methods and Big Data analytics (such as XBRL, web-scraping and text analytics). The key principle is to leverage all possible data sources in order to effectively overcome limitations in terms of data quality, granularity, and restrictions on dissemination; while at the same time carefully validating results and ensuring maximum alignment with national statistical concepts.

The register of multinationals, their affiliates and relationships includes all domestic and foreign controlled affiliates of the MNE Ultimate Controlling Parent, using an ownership threshold of 50% (defined as control, and following FATS) to decide whether to include an affiliate in the register.

### *Company reports and regulatory submissions*

The primary source of information on the consolidated activities of companies is Annual Reports. The International Financial Reporting Standards (IFRS) provide the most widely used internationally comparable accounting standard; however companies may also use the Generally Accepted Accounting Principles (GAAP) (for example, US listed enterprises are required to submit their accounts to the SEC using US GAAP). ADIMA classifies the reporting regime used, making users aware of the differences, however has not yet attempted to harmonise between accounting regimes.

While clearly a rich source of information, the extraction of data from Annual Reports in an automated fashion is complicated, as companies may publish their reports in a variety of formats (PDF, websites) that are not optimised for this purpose. However, securities regulators in several countries, including the United States, require that companies submit their annual reports using XBRL (which entails that data points are tagged with a harmonised coding), which facilitates the extraction of key data items. Some manual data entry remained necessary however for this pilot study, for example for variables on employment and certain parts of the data on geographical segments.

---

<sup>3</sup> The database “Base Sirene” from INSEE is an open register on enterprises and establishments. The information on the headquarter location is detailed for French locations, but is limited to being classified as ‘Foreign’ otherwise. See also: <https://www.sirene.fr/sirene/public/variable/rpen>

<sup>4</sup> Persons with Significant Control (Companies House): since mid-2016, UK companies are required to file information over individuals who have direct or indirect influence on them; in addition, UK companies and certain non-UK listed companies may also be registered as Relevant Legal Entities. See also: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/621566/170623\\_NON-STAT\\_Guidance\\_for\\_PSCs\\_4MLD.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/621566/170623_NON-STAT_Guidance_for_PSCs_4MLD.pdf)

<sup>5</sup> Article 30 of the 4th EU Anti-Money Laundering Directive (4AML), for example, requiring all EU Member States to put into national law provisions around beneficial ownership information for corporate and legal entities.

### *LEI relationship records*

The Legal Entity Identifier (LEI) is a 20-character reference code to uniquely identify entities engaged in financial transactions. This identifier is supported by the ‘Global Legal Entity Identifier Foundation’ (GLEIF), an initiative launched in 2011 by the Financial Stability Board (FSB), mandated by the G20. In addition to providing firm-level identification information (‘level 1 information’), the entities are required to declare the immediate and ultimate parent upon registration (‘level 2 information’), and are expected to confirm or update this information at least on an annual basis. Within the LEI relationship information, the “ultimate accounting consolidating parent” is defined as the highest level legal entity preparing consolidated financial statements (LEI ROC, 2016), which in principle aligns with the concepts and definitions in BMD4.

The LEI relationship data has been made public, gradually, since 2017 Q2. As of 12 February 2018, almost 1.1 million LEIs have been issued to legal entities globally, of which around 60 percent also reported information on direct and ultimate parents (up from 26 percent only a quarter prior).<sup>6</sup> The growth in uptake of LEI was particularly marked in 2017, as the LEI population nearly doubled in response to EU regulation.

As the LEI and the LEI relationship datasets have only recently started to be collected, their coverage is still insufficient for the construction of company ownership hierarchies. It remains unclear at present how many affiliates will acquire an LEI, and to what extent the data can be used as the *sole* data source for affiliate hierarchies. For this reason, validation work incorporating the LEI data should be seen as an assessment of progress to date, as opposed to the overall quality of the source. Importantly, however, further disclosure requirements are being considered, including for example the inclusion of the LEI of subsidiaries in SEC submission, which would support further international adoption of the standards (SEC, 2016).

### *Commercial sources (ORBIS)*

While the drawbacks of ORBIS are well-known it remains one of the largest cross-country databases of enterprise information compiled from a large variety of private and public data providers, with near-global coverage. One disadvantage particularly pertinent to the construction of enterprise registers is the lack of harmonisation of the unit of observations. For example, data for some countries such as France are organised around enterprise units, which in turn manage establishments; data for other countries such as the United States do not allow for straightforward identification of the decision making units and their respective branches/establishments<sup>7</sup>. To mitigate this issue only those affiliates where the ultimate parent owned more than 50% were retained.

### *Assumptions and reconciling sources*

To fully leverage all available information, the affiliates identified by each of three sources used are combined into a single database, using a fuzzy matching<sup>8</sup> procedure. The OECD MNE-affiliate structures were created by using company names and the countries and cities of registration in each of the three sources.

A key objective of ADIMA is to develop more granular (country-level) data regarding the activities of MNEs but Annual Reports are rarely able to provide this information beyond broad regions. Therefore,

---

<sup>6</sup> Information extracted from LEI website

<sup>7</sup> For example, for Walmart, the number of US subsidiaries dropped from 12,000 to 500 once the “branches” were removed.

<sup>8</sup> One of the main challenges of fuzzy matching is to fine-tune the algorithm by deciding on the degree of record similarity that is required to consider the records matched.

ADIMA captures the actuals and using these as constraints, explores the use of alternative data sources and the use of Big Data analytics to develop estimates of the geographical breakdown of MNEs activities.

This area remains a work-in-progress and three analytical approaches have currently been developed (further information can be found [here](#) in the full version of this paper):

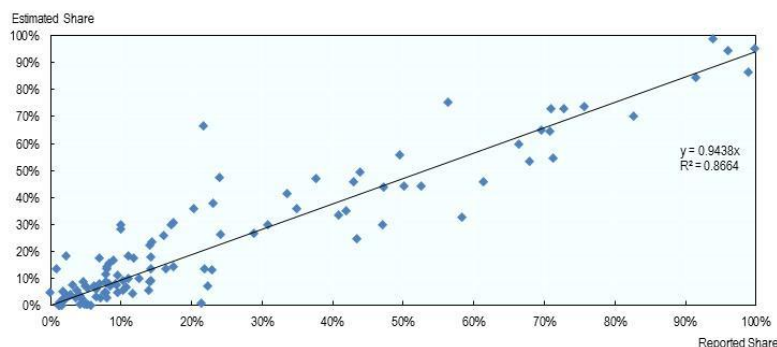
- **Text Analytics:** A significant section of the MNE website can be identified which describes the overall global strategy and organisation of the MNE. Advanced text analytics is used to develop frequency tables of country mentions in these webpages, which are used to disaggregate reported segment sales. In general, this methodology tends to be most valid for companies with significant Business-to-Business (B2B) transactions, (see Box 1),
- **Page Rank:** A company's business is segmented into location specific websites. An example of this is Wal-mart, whose websites include: walmart.com (United States), walmart.ca (Canada), asda.com (United Kingdom) and seiyu.co.jp (Japan). In this case, the popularity, or Page Rank, of each underlying site corresponding to a given location is used to disaggregate the reported segment information,
- **Link Analysis:** A company's business is operated primarily via a website with global product-oriented presence. Examples of this include: microsoft.com and ibm.com. In this case, the outward links from the website are used to determine the distribution of sales by country.

### Provisional ADIMA results

As noted above, especially given its relative novelty, ADIMA remains a work-in-progress and further work will be undertaken to continue to review data sources and estimation methods. However provisional results produced for 37 large US MNEs have proven promising.

Figure 1 presents the results of the geographical estimation methods used in ADIMA, grouped, and compared with the corresponding results, by geographical region available in annual statements for each of the 37 MNEs. Overall, the estimated values align well (R-squared, 0.87). However, it is also clear that in a few cases, the estimation could be further improved.

**Figure 1. Comparison of reported geographic segments and estimated geographic segments**

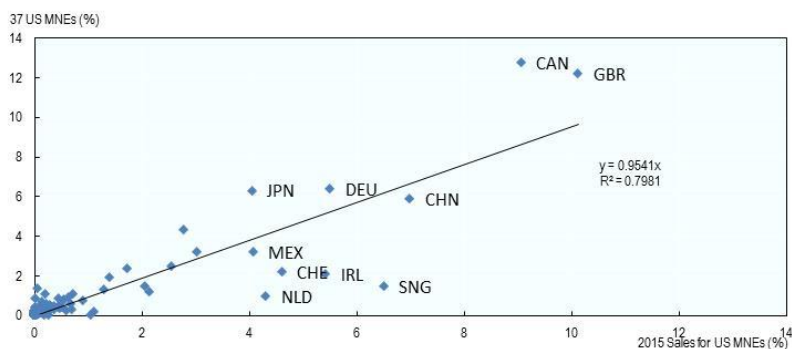


Source: OECD ADIMA.

Another comparison (Figure 2) is with official outwards FATS statistics of the United States. Although imperfect, as the total foreign turnover of the 37 US MNEs currently in ADIMA accounted for only one-fifth of US Outward FATS in 2015, the results also look promising ( $r = 0.8$ , Figure 2). However, for several countries, important differences are visible, notably for countries often used for fiscal optimisation

purposes, like the Netherlands, Ireland and Singapore, where official FATS data record (much) higher sales than in ADIMA, and further work will also be conducted here.

**Figure 2. Sales of Foreign US Affiliates compared to estimated foreign sales of 37 US MNEs**



*Note:* When 2015 Sales data has been suppressed in provisional 2015 data, 2014 values are utilised.  
*Source:* OECD ADIMA, US BEA: <https://www.bea.gov/international/usdia2015p.htm>

## Conclusions and next steps

Consistently and comparatively measuring the international activities of MNEs has been a longstanding and increasingly pertinent challenge in economic statistics. Given that national statistical institutes are typically limited in their (legal) ability to capture activities outside their jurisdiction, an international and ‘whole of the MNE’ approach is required to better understand the global scale and scope of MNEs, but also to support the consistent treatment of MNEs in national statistics. ADIMA, for which this paper presented a proof-of-concept, aims to provide such an approach, using both traditional data sources and innovative Big Data and analytical techniques.

Current results suggest that the process for delivering meaningful results on MNEs are feasible, certainly at the level of the MNE, where company accounts provide a robust source of information that can be used as benchmarks. Estimation methods currently being explored to develop breakdowns by the countries of operation have also proved promising, although further refinements will be introduced, such as improvements to the fuzzy matching algorithm and to the estimated geographical breakdown of sales, as ADIMA matures from a ‘proof of concept’ activity to a database by the end of 2018. Likewise, the data sources and analytical techniques currently used in ADIMA are merely scratching the surface of what is possible. One of the first priorities will be to use additional Big Data sources to develop breakdowns by country of MNEs’ total assets and employment. Additional verification of the data with official sources, in particular in collaboration with NSOs and Eurostat, should further validate the results and point to areas for improvement.

Collaboration with Eurostat has started with respect to the Monitoring tool, to ensure alignment and complementarity with the Early Warning System, in order to provide very timely information on changes in MNE activities, including for example investments and divestments, mergers and acquisitions, or relocations. Pilot tests with interested countries are envisaged to ensure that the specifications meet compilers’ needs as much as possible.





## Annex A. 100 MNEs in initial sample

**Table A.1. Largest 100 MNEs ordered by revenue**

#	Company Name	#	Company Name	#	Company Name	#	Company Name
1	<b>Wal-Mart Stores</b>	26	<b>Cardinal Health</b>	51	<b>Alphabet</b>	76	<b>Johnson &amp; Johnson</b>
2	China Petroleum & Chemical	27	Total	52	<b>Microsoft</b>	77	Uniper
3	Toyota Motor	28	<b>Verizon Communications</b>	53	Siemens	78	Engie
4	Royal Dutch Shell	29	Honda Motor	54	Assicurazioni Generali	79	Airbus
5	PetroChina	30	<b>General Electric</b>	55	Prudential	80	Indian Oil
6	Volkswagen	31	Japan Post Holdings	56	Nestle	81	Tesco
7	Berkshire Hathaway	32	<b>Costco Wholesale</b>	57	Petrobras	82	<b>Target</b>
8	<b>Exxon Mobil</b>	33	<b>Walgreens Boots Alliance</b>	58	NK Lukoil	83	SK Holdings
9	<b>Apple</b>	34	Fiat Chrysler	59	<b>Anthem</b>	84	Sony
10	<b>McKesson</b>	35	<b>Kroger</b>	60	<b>Phillips 66</b>	85	Aviva
11	<b>UnitedHealth Group</b>	36	Allianz	61	Carrefour	86	Panasonic
12	BP	37	<b>Chevron</b>	62	Hitachi	87	Ceconomy
13	<b>CVS Health</b>	38	SAIC Motor	63	NK Rosneft	88	<b>Procter &amp; Gamble</b>
14	Samsung Electronics	39	Nissan Motor	64	<b>Comcast</b>	89	<b>Lowe's</b>
15	<b>General Motors</b>	40	Ping An Insurance Group Co of China	65	<b>IBM</b>	90	Muenchener Ruckversicherungs
16	<b>AT&amp;T</b>	41	Nippon Telegraph and Telephone	66	SoftBank	91	Zurich Insurance Group
17	Daimler	42	China Mobile	67	China Life Insurance	92	Marubeni
18	Glencore	43	<b>Express Scripts</b>	68	Hyundai Motor	93	People's Insurance Group of China
19	<b>Ford Motor</b>	44	Gazprom	69	Deutsche Telekom	94	<b>MetLife</b>
20	Exor	45	BMW	70	Japan Post Insurance	95	<b>Marathon Petroleum</b>
21	<b>AmerisourceBergen</b>	46	Legal & General Group	71	<b>Valero Energy</b>	96	<b>Aetna</b>
22	China State Construction Engineering	47	<b>Home Depot</b>	72	Electricité de France	97	JXTG
23	<b>Amazon.com</b>	48	<b>Boeing</b>	73	DowDuPont	98	<b>PepsiCo</b>
24	AXA	49	China Railway Group	74	Aeon	99	Audi
25	Hon Hai Precision Industry	50	China Railway Construction	75	Enel	100	<b>Archer Daniels Midland</b>

*Note:* Of the 39 US companies in the Top 100 by revenues, 37 of 39 were selected for the Pilot (bold in the above list). DowDuPont and Berkshire Hathaway were the two MNEs excluded. DowDuPont was excluded because of its recent reorganisation (merger of DuPont and Dow Chemical) which has not yet been reflected in commercial, annual reports or website data sources. Berkshire Hathaway was excluded from the pilot sample because its economic variables do not relate directly to the economic performance of MNEs under control of Berkshire but rather the performance of Berkshire Hathaway investment services (i.e. sales of Berkshire not consolidated sales of Dairy Queen, Kraft, etc.).

*Source:* OECD ADIMA.