IFC-ECCBSO-Bank of Spain Workshop on "New insights from financial statements"

17 October 2024

Connecting financial data with intellectual property rights: relevancy and (data) complexity[1]

Julie Callaert, Phoenix Chen, Bart Van Looy, Steven Vanhaverbeke, Jan-Bart Vervenne and Jesse Wursten, KU Leuven

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the European Committee of Central Balance Sheet Data Offices (ECCBSO), the Bank of Spain, the BIS, the IFC or the other central banks and institutions represented at the event.

Connecting financial data with intellectual property rights: relevancy and (data) complexity

B. Van Looy, J-B Vervenne, J. Callaert, J. Wursten, P.-H Chen, S. Vanhaverbeke

KU Leuven, Department of Management, Strategy and Innovation (MSI), Faculty of Business and Economics (FEB), B-3000 Leuven, Belgium

Expert Centre R&D Monitoring (ECOOM)

Abstract

This paper presents the findings of several recent studies that integrate firm-level Intellectual Property Rights (IPR) and financial data, highlighting both advancements and ongoing challenges (Vervenne et al., 2020; BOIP/FOD, 2021). Integrating these data types holds substantial potential for generating valuable insights for policymakers and practitioners. However, aligning these data types at the firm level introduces unique challenges, particularly in harmonization, disambiguation and matching of corporate applicant names obtained from patent and financial databases.

Over recent years, the department of Management, Strategy and Innovation (MSI) and ECOOM at KU Leuven have developed specific algorithms to identify the type of (IP) applicant (Companies, Individuals, HEI's, Governmental agencies) and to harmonize, disambiguate (and regionalize) applicant names[1]. These algorithms have been adopted and used by a number of governmental agencies, both regionally (ECOOM, Flanders) and internationally (Eurostat, EPO, EIB).

In addition, we designed a specific methodology (see figure 1) to match these IP data with financial repositories (Amadeus and Belfirst (BvD/Moody's)). These methodological advancements not only allow for the identification of IP ownership at the firm level and the distillation of relevant indicators (at the firm level and beyond); they also enable analyses that focus on the relation between IP practices of firms (size, scope, nature of IP portfolios) and the financial performance of the implied companies, such as growth and value (e.g. Leten et al. 2007; Belderbos et al. 2014).

Despite these methodological improvements, significant limitations remain. The financial databases that are currently available in the market are limited in terms of comprehensiveness and historical coverage. Additional gaps include critical details such as firm location, ownership history (essential for consolidation), and name changes (critical for accurate longitudinal studies).

The aim of this paper is to demonstrate that while significant strides have been made in harmonizing patent counts and linking IP and financial data at the firm level,

[1] For an overview of relevant methodologies and algorithms, see Magerman et al. (2009), Du Plessis et al. (2009), Callaert et al. (2011)

ongoing data limitations must be addressed to fully realize the potential of these integrated datasets for researchers, practitioners, and policymakers. We strongly advocate for the development of enhanced financial databases that provide accurate longitudinal firm-level data.

Keywords: SME's – patents – EU technology landscape – corporate / IP database linkages

JEL classification: O32 – O34

Contents

Outlining a methodology to match patent assignees with firm level data, including characterizing the size of firms (SMEs or not)

Introduction

The methodology deployed to arrive at a reliable and comprehensive coupling between corporate patent assignees and financial information pertaining to the assignee firms has been developed over recent years at KU Leuven (ECOOM, MSI). It enables studies on the firm level (e.g. IP and valuation of firms) and it supports further development of indicators for assessing the performance of innovation systems (ECOOM). In the context of the latter application domain, one of the major studies, aimed to identify the contribution of SMEs to technology development (measured by patents) on a European level. This study was commissioned by Eurostat and has recently also been published (Vervenne et al. 2022).

As the referenced study is the most exhaustive and comprehensive one so far – in terms of matching and characterizing firm applicants by means of their financial accounts – we stay close to its findings in the elaboration of the current paper.

The Vervenne e al. study implied an exhaustive methodology, consisting of four steps. First, corporate (patent) applicant names were matched to firm names in financial directories. Next, a disambiguation procedure was applied to identify the right firm when multiple companies were matched to the same corporate applicant. Subsequently, to the extent that entity size indicators and dependency status information was available, a considerable portion of the matched applicants became classified according to firm size. However, a non-negligible portion of firm applicants remained unassigned after the first three automated steps, either because no match was found or because the necessary information for size evaluation was lacking. In a final step, this void was addressed by investigating stratified samples and extrapolating the findings to the population level.

The mere fact that this last step is indispensable to arrive at reliable estimates within reasonable boundaries clearly signals the deficiencies of current financial databases: a number of critical variables are lacking, and the absence of accurate longitudinal data complicates and even prevents a dynamic perspective (spanning decades) all together. As such, both academia and policy makers alike would benefit considerably from the presence of financial databases that are complete and historically correct. As current providers of such databases are clearly not interested in addressing these needs, we plea for an initiative that addresses this void.

Outline of the methodology

Matching

Patent documents report no information on the size of the applicant. Nor do they include a (unique) firm identifier facilitating an unambiguous match with (annual) financial accounts. To classify corporate patent applicants according to size, one has to merge patent and financial data and identify both primarily on the basis of name information. Several commercial databases are available pooling financial data from annual account filings with national business registries. We chose to extract annual account information from Bureau Van Dijk's Amadeus database (2012 edition), which provides data for European firms. Patent data was extracted from the EPO Worldwide Patent Statistical Database or PATSTAT (autumn 2011 edition). The approach adopted to create correspondence between PATSTAT applicants and Amadeus firms is schematized in Figure 1.

In its raw form, the PATSTAT database provides unprocessed (non-harmonised) applicant names, as well as country and address information. Similar to Frietsch et al. (2013), a sector allocation algorithm is applied to all applicant names in PATSTAT to reduce the odds of associating non-corporate applicants with companies. The algorithm uses a keyword logic to filter out non-corporate applicants (see Eurostat, 2011b).[2]

Next, European corporate applicant names in PATSTAT are matched to the names of firms included in the Amadeus database. Before matching, the same name harmonisation procedure reported in Eurostat (2011b) is applied to all company names in Amadeus – the latter with the intent of limiting the number of potential false negatives attributable to the use of different name variants across both databases.[3] In this study, attention is confined to corporate applicants filing for patent protection at the EPO or the USPTO, or relying on the PCT procedure. We considered patent applications filed over a relatively long-time window (1999-2011) to obtain an idea of the proportions prevailing in the long run.

The matching of applicants and firms is based on exact string correspondence, which prioritizes precision of the retained pairs over recall. In the first round of matching, corporate applicants are matched exclusively to companies from the same country (sharing the same country code in both Patstat and Amadeus). Harmonised corporate applicant names are compared consecutively with harmonised versions of current company names, former company names and company aliases. In a second round of matching, the same procedure is repeated whilst relaxing the country condition, assuming that subsidiaries may be established under names that resemble those of

[2] Other types of actor discerned by the sector allocation algorithm include government and non-profit bodies, individuals and universities. Corporates accounted for 66% of the patents filed in the countries in the reference period; individuals 29%; governments and non-profit bodies 3%; and universities 2%.

[3] The name-harmonization methodology removes discrepancies in company names caused, for example, by variation in punctuation, legal form indication abbreviations and spelling, notation of common words, character formatting, and the use of umlauts (Eurostat, 2011b).

the parent companies. Table 1 shows the recall rates of both rounds for the entire time frame (application years 1999-2011) expressed in applicants and applications.

TABLE 1: APPLICANTS AND APPLICATIONS MATCHED TO AT LEAST ONE COMPANY IN THE FINANCIAL DIRECTORY

Country	*Corporate applicants*			*Corporate applications*		
	Total	**Matched**	**%**	**Total**	**Matched**	**%**
EU-27 [4]	104166	64496	61.9	1316568	1094349	83.1
BE	2218	1542	69.5	26129	23220	88.9
BG	107	45	42.1	190	73	38.4
CZ	500	336	67.2	1450	967	66.7
DK	3593	2101	58.5	29487	24468	83.0
DE	30130	16320	54.2	537847	453746	84.4
EE	112	65	58.0	226	136	60.2
IE	1235	912	73.8	8767	6575	75.0
EL	209	59	28.2	676	196	29.0
ES	4234	2395	56.6	17019	11494	67.5
FR	10763	5587	51.9	179457	144112	80.3
IT	13104	8974	68.5	77186	60358	78.2
CY	245	62	25.3	932	323	34.7
LV	74	18	24.3	288	37	12.8
LT	16	8	50.0	27	13	48.1
LU	649	259	39.9	5399	3107	57.5
HU	513	181	35.3	1689	636	37.7
MT	82	53	64.6	426	363	85.2
NL	6891	4720	68.5	132865	121315	91.3
AT	3042	1632	53.6	25293	18588	73.5
PL	401	238	59.4	1179	796	67.5
PT	382	192	50.3	1065	738	69.3
RO	57	17	29.8	95	34	35.8
SI	265	135	50.9	1438	678	47.1
SK	124	76	61.3	305	225	73.8
FI	2683	1724	64.3	51052	44874	87.9
SE	6226	3452	55.4	84844	53081	62.6
UK	16311	13393	82.1	131237	124196	94.6

Source: PATSTAT autumn 2011 edition, Amadeus 2012.

Matching rates (aggregated across patent offices) range between 24.3 % (Latvia) and 82.1 % (United Kingdom). Overall, 61.9 % of the harmonised corporate applicant names are matched to the companies reported in Amadeus (57.9 % in the same country and 4.0 % in other Member States). These matched corporate applicants account for 83.1 % of patent applications filed by corporate applicants. 77.9 % can be assigned to corporate applicants matched to companies from the same country and 5.2 % to those matched to companies from other Member States. The comparison between applicant and application figures reveals that, on average, unmatched corporate applicants patent less than matched ones (38.1% of the unmatched names account for 16.9 % of the remaining corporate patent volume).

Disambiguation

[4] At the time of the study, the EU comprised 27 Member States.

The name-based matching algorithm resulted in corporate applicants being matched to multiple companies present in Amadeus: approximately one out of every eight harmonized corporate applicant names was linked to more than one of the legal entities included in Amadeus. The statistics reported in Table 2 reveal large cross-country differences in the share of matched corporate applicants allocated to multiple firms.

TABLE 2: SHARES OF MATCHED CORPORATE APPLICANTS MATCHED TO MULTIPLE FIRMS AND AVERAGE NUMBER OF AMADEUS FIRMS MATCHED TO THEM.

Country	% applicants matched to multiple firms	Average number of Amadeus firms linked to corporate applicants matched to more than 1 firm
IT	35%	12.3
FR	21%	7.1
RO	15%	2.0
PL	15%	3.7
DE	13%	4.4
BE	11%	3.7
AT	9%	2.4
NL	7%	3.7
CZ	4%	2.5
SI	4%	2.0
LV	3%	2.0
GR	3%	2.0
SK	2%	2.0
PT	2%	4.0
GB	2%	2.1
DK	2%	2.4
SE	2%	5.0
FI	1%	2.1
IE	0%	2.0
ES	0%	2.0
BG	0%	-
CY	0%	-
EE	0%	-
HU	0%	-
LT	0%	-
LU	0%	-
MT	0%	-

These cross-country differences can, among other factors, be attributed to the lack of strict name requirements prior to business registration in some countries (e.g. Italy) or the non-binding nature of such requirements in other countries (e.g. France) (World Bank, 2017). A number of selection rules are applied to disambiguate these associations. The full disambiguation process consists of four automated rounds applied consecutively. First, companies with addresses that do not correspond to the address of the corporate applicant are removed if at least one other company is matched to the corresponding address information. Next, when one of the matched companies holds the majority of shares in another company matched to the same corporate applicant, the latter is discarded. Thus, priority is given to companies at the top of the shareholder hierarchy. Subsequently, any liquidated, dissolved, bankrupt or inactive company matched to a corporate applicant that has filed patents after the year in which it published its last available annual accounts is also discarded. Finally, in line with Squicciarini and Dernis (2012), in the final disambiguation round, matched

companies showing maximum values for revenue, staff count and total assets – in that order – were retained. For the few corporate applicants that remained ambiguously matched after subjection to this decision rule, the most adequate match was then selected by means of human assessment until every matched corporate applicant was paired with a single firm. Table 3 reports the effects of each disambiguation step on the number of firms remaining matched to the corporate applicant in Amadeus. Clearly the first and last disambiguation rounds reduce the number of potentially corresponding firms the most extensively.

TABLE 3: IMPACT OF DISAMBIGUATION STEPS ON NUMBER OF AMADEUS FIRMS LINKED WITH THE MATCHED CORPORATE APPLICANTS

# firms matched to applt	*Initial Matches*	*Address*	*Shareholders*	*Inactivity vs 1st filing*	*Financial/manual*
1	55,397	58,943	59,006	59,249	64,496
2	3,963	2,706	2,652	2,480	-
3	1,306	800	795	774	-
4	714	420	418	402	-
5	496	304	303	293	-
6	342	206	205	204	-
7	281	157	157	151	-
8	198	107	107	107	-
9	180	91	91	86	-
10	162	84	84	87	-
11-15	493	255	255	253	-
16-20	251	130	130	122	-
21-30	298	139	139	140	-
31-50	234	94	94	90	-
51-100	153	51	51	49	-
101-200	29	9	9	9	-
201-250	2	0	0	0	-
	64,496	64,496	64,496	64,496	64,496

Classification

The classification of matched corporate applicants according to firm size is based on the European Commission's (EC) SME definition, which sets out three criteria for SME status: a staff headcount below 250 full-time equivalents (FTEs) and either an annual turnover below or equal to 50 million Euro or a balance sheet total below or equal to 43 million Euro.[5] Preferably, these criteria should be evaluated at the level of the business group to which the entity belongs. Entities belong to a group when a natural person controlling multiple firms or a corporate shareholder owns a majority of the voting rights in the entity's board. Ownership of at least 50% of the shares – be it directly, or indirectly through an intermediary subsidiary – signals such a controlling position.

[5] On the basis of European Commission Recommendation 2003/361/EC, a new SME definition was adopted on 1 January 2005, incorporating updated thresholds for companies applying for the European support programme for SMEs.

Applying these financial and ownership criteria to a financial database is not a straightforward exercise. Due to disclosure exemptions for smaller firms and the scale of the targeted firm population, the data provided for an average firm is less complete than the standard witnessed in databases that cover only large and/or listed firms. As the next section makes clear, Bureau Van Dijk's Amadeus 2012 database is no exception in this respect.

Availability of financial and shareholder data

Bureau Van Dijk publishes a new version of Amadeus every year since it became a commercial product in 1996. Each version is updated regularly throughout the year, incorporating newly available information. In this study, we based the firm size assessment of patenting companies on the most recent annual accounts available per firm for two principal reasons. While time series data for the past 10 years are available for revenues, staff counts and total assets at the entity level, only the last known ownership information is reported. This hampers a dynamic assessment of firm size at the business group level.[6] In addition, rather than providing a historical account of the firm's shareholder structure, the majority of supplementary information sources (cf. infra) that were consulted to ascertain the size distribution of the unallocated applicants only provide an insight into their last known ownership situation. Therefore, rather than classifying the patent based on financial information derived from the annual accounts covering the year in which it was applied for by the applicant, each corporate applicant and all its patents are attributed to one size category for the entire time frame. By exclusively relying on the last available size indicators per applicant for classification, an additional source of potential bias is introduced to the SME contribution estimates. To the extent that small firms grow fast, the number of SME patents could well be underestimated during the initial years of the time period under study.

To evaluate the extent to which the EC's SME definition can be used to differentiate between SMEs and large companies in Amadeus 2012, we assessed per-country data availability for the indicators of interest. Table 4 reports coverage rates per indicator in Amadeus 2012.

[6] A combination of ownership information from multiple Amadeus versions at fixed intervals within the investigated time window could provide a solution in which a more precise picture of the size of the matched corporate applicants at the precise time of filing can be obtained. However, on top of the other arguments mentioned, the even more fragmented nature of ownership information in pre-2012 versions of Amadeus exercised an equal influence on our decision to refrain from such efforts.

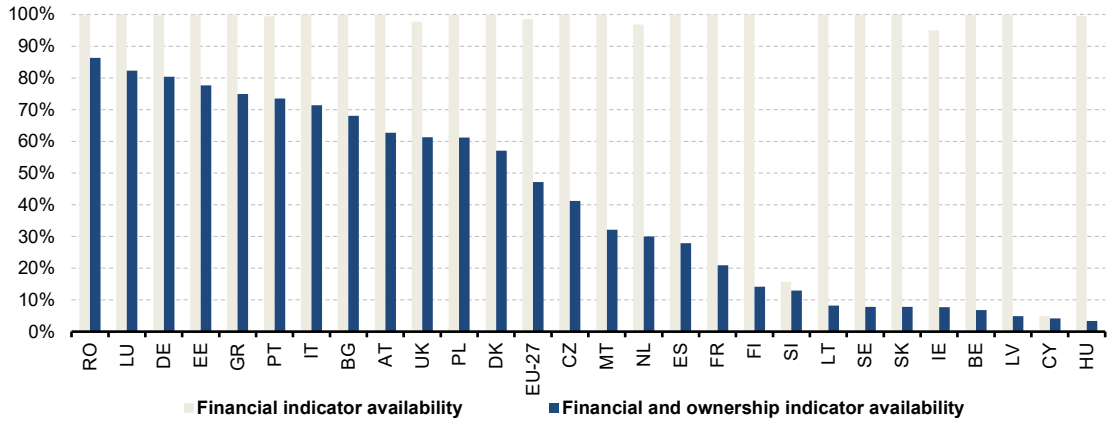
TABLE 4: AMADEUS – OVERALL COVERAGE OF FINANCIAL AND OWNERSHIP INDICATORS REQUIRED TO DETERMINE FIRM SIZE (PER INDICATOR)

	Total number of companies	*Companies reporting operational revenues*		*Companies reporting staff count*		*Companies reporting total assets*		*Companies reporting dependency status*	
		number	**%**	**number**	**%**	**number**	**%**	**number**	**%**
BE	609412	129201	21.20	257647	42.28	545187	89.46	41646	6.83
BG	494532	59231	11.98	488209	98.72	60490	12.23	336347	68.01
CZ	482679	469799	97.33	294733	61.06	184120	38.15	198742	41.17
DK	25230	42447	16.78	76887	30.40	252918	100.00	144339	57.07
DE	1456074	419446	28.81	437928	30.08	1101434	75.64	1170990	80.42
EE	108986	94667	86.86	52709	48.36	108936	99.95	84675	77.69
IE	211372	25951	12.28	21836	10.33	199798	94.52	16627	7.87
EL	28401	28401	100.00	23600	83.10	28401	100.00	21289	74.96
ES	1273351	1140063	89.53	843380	66.23	1273351	100.00	355419	27.91
FR	1291883	1291875	100.00	878954	68.04	1291882	100.00	269729	20.88
IT	119884	1188353	99.14	352781	29.43	1198684	100.00	856313	71.44
CY	41289	907	2.20	1878	4.55	1005	2.43	36963	89.52
LV	110292	85711	77.71	102382	92.83	7938	7.20	5406	4.90
LT	117370	26789	22.82	110033	93.75	3712	3.16	9636	8.21
LU	19240	5199	27.02	4040	21.00	16028	83.31	15838	82.32
HU	377912	316267	83.69	135538	35.86	375792	99.44	12683	3.36
MT	15259	15259	100.00	392	2.57	15259	100.00	4907	32.16
NL	895494	31108	3.47	643147	71.82	822005	91.79	272414	30.42
AT	224480	6385	2.84	167413	74.58	77880	34.69	140733	62.69
PL	960971	117796	12.26	902969	93.96	121316	12.62	589289	61.32
PT	434526	365782	84.18	343776	79.12	428069	98.51	319629	73.56
RO	571289	568039	99.43	566221	99.11	571038	99.96	493193	86.33
SI	76089	2512	3.30	10558	13.88	2574	3.38	15661	20.58
SK	230781	165399	71.67	197956	85.78	58557	25.37	17996	7.80
FI	170484	160798	94.32	47776	28.02	170484	100.00	24134	14.16
SE	866641	829664	95.73	848854	97.95	319732	36.89	67869	7.83
UK	3076136	447311	14.54	132348	4.30	2998120	97.46	1918280	62.36
Total	**15596557**	**8034360**	**51.51**	**7943945**	**50.93**	**12234710**	**78.44**	**7440747**	**47.71**

Source: PATSTAT autumn 2011 edition, Amadeus 2012.

The statistics presented in Table 4 reveal non-trivial data gaps, with considerable variation across Member States. This is partly caused by differences in legislation regarding registration/disclosure, whereby smaller entities are not bound to full disclosure or are exempted from financial information disclosure altogether.

FIGURE 2: AMADEUS – OVERALL COVERAGE OF (FINANCIAL AND FINANCIAL/OWNERSHIP) INDICATORS REQUIRED TO DETERMINE FIRM SIZE (%)



Source: Amadeus 2012

Figure 2 provides further insight into the availability of firm size indicators in Amadeus 2012. The Member States are ranked according to descending rates of joint availability of financial and ownership indicators. Overall, 99% of the firms in Amadeus 2012 report at least one of the three financial size indicators from which entity size can be derived. Cyprus and Slovenia appear to be the only 'problematic' cases in terms of coverage. The figures become more troublesome if one takes ownership information availability into account as well for evaluation of size at the enterprise or business group level. Only 47% of the firms provide at least one of the financial firm size indicators and offer sufficient shareholder information to distinguish dependent from independent firms.

Taking into account these data constraints, classification of the matched corporate applicants comprised two stages. In the first stage, entity-size indicators (revenues, employee count and total assets) are used to identify large and small corporate entities. Obviously, for large entities, one can already state that they represent a large enterprise. In the second stage, to filter out the *actual* SMEs among the small corporate entities, dependency information is introduced and consolidated financial size indicators are used in order to classify firms in terms of majority ownership.

Stage 1 of classification: using entity size indicators to differentiate between large and small entities

Entities with known information on all three entity size indicators allow for classification of the corporate applicants associated with them. taking into

consideration all criteria comprising the SME definition. The corporate applicants that are categorized using this baseline SME definition are labelled as entities that are 'certainly large' or 'certainly small'. However, the coverage gaps reported for the three indicators in Table 4 underscore the need for a complementary, looser version of the definition. To the extent that the thresholds for the three size indicators are determined according to proportions that are representative for the average industry, it seems plausible to presume that, if for instance one available indicator lies above the threshold, the other indicators are more likely to lie above than below their thresholds as well. Companies that are classified using this approach are labelled as entities that are 'most likely large'[7] or 'most likely small'[8].

Stage 2 of classification: using ownership information to identify actual SMEs

The shareholder information provided in Amadeus is essential to determine which small entities effectively qualify as independent SMEs and which are members of larger business groups or are backed by other types of shareholders such as governments, institutional investors and universities. Determining which of the matched companies required further exploration of their shareholder structure departed from the 'independence indicator' reported in Amadeus. Bureau Van Dijk designed this measure to facilitate quick evaluation of how dependent a company is vis-à-vis its shareholders. We rely on it to differentiate between independent SMEs, i.e. small businesses that have no majority shareholders, and a 'rest' category of non-independent small businesses. Further analysis of shareholder data is restricted to this 'rest' category of firms. More specifically, to the extent such data is reported, information about the nature of those firms' shareholders and the size of shareholder stakes in those and other firms (in the case of a natural person holding the majority of the shares) is designed to assess the size of the business group to which the entity potentially belongs. Insofar as available, this assessment is based on the sum of direct and indirect ownership percentages i.e. the percentage of the company directly owned by the shareholder and the percentage owned through subsidiaries.[9] The size

[7] The entities considered as 'most likely large' were those reporting: revenues over 50 million Euro, where staff count and total asset numbers are unavailable; total assets over 43 million Euro, where staff count and revenue numbers are unavailable; revenues over 50 million Euro and total assets of 43 million Euro or less, where staff count numbers are unavailable; revenues of 50 million Euro or less and total assets over 43 million Euro, where staff count numbers are unavailable.

[8] Companies that are 'most likely' small entities are those reporting: a staff count of less than 250 Full-Time Equivalents (FTE), where revenues and total assets are unavailable; revenues of EUR 50 million or less and total assets of 43 million Euro or less, where staff count numbers are unavailable; total assets of 43 million Euro or less, where staff count and revenue numbers are unavailable; revenues of 50 million Euro or less, where staff count and total assets numbers are unavailable; revenues of 50 million Euro or less and total assets of 43 million Euro or less, where staff count numbers are unavailable; revenues of over 50 million Euro and staff counts of less than 250 FTEs, where total asset numbers are unavailable; and total assets of over 50 million Euro and staff counts of less than 250 FTEs, where revenue numbers are unavailable.

[9] Where the ultimate owner controls the intermediate subsidiary by owning a majority of its shares, one can assume that its ownership stake in the company under consideration is a reflection of the share percentage its subsidiary holds in that company. Otherwise, Bureau Van Dijk calculates the indirect

of corporate applicants that are accordingly held to form part of a business group is evaluated at the group level (see below).

Results: small vs. large corporate entities and SMEs vs. large companies

The outcomes show that 8.9 % qualify as large corporate entities in the first round, whereas 52.5 % can be characterised as small corporate entities. A majority of the large entities are identified as 'certainly' large (8.5 %) and the remaining 0.4 % as 'most likely' large corporate entities (see above). 27.2 % of the small entities are characterised as 'certainly' small and 25.3 % as 'most likely' small (see above). The corresponding patent volumes are presented in Table 5, the first row of which contains the total number of distinct patents per size category for the EU overall.

TABLE 5: DISTRIBUTION OF CORPORATE PATENTS AFTER THE 1ST STAGE OF CLASSIFICATION BASED ON ENTITY-LEVEL SIZE INDICATORS ONLY

Country	*Large*				*Small*				*No financial size indicators*		*Not matched*		*Total*
	Certainly		Most likely		Certainly		Most likely						
	#	%	#	%	#	%	#	%	#	%	#	%	#
EU-27	696 716	52.9	25 896	2.0	174 998	13.3	192 480	14.6	4 259	0.3	222 219	16.9	1 316 568
BE	15 103	57.8	1 633	6.2	5 301	20.3	1 182	4.5	1	0	2 909	11.1	26 129
BG	37	19.5		0	19	10	17	8.9		0	117	61.6	190
CZ	362	25	5	0.3	518	35.7	82	5.7		0	483	33.3	1 450
DK	15 229	51.6	94	0.3	5 821	19.7	3 292	11.2	32	0.1	5 019	17	29 487
DE	293 215	54.5	7 412	1.4	35 956	6.7	116 945	21.7	218	0	84 101	15.6	537 847
EE	6	2.7	2	0.9	92	40.7	36	15.9		0	90	39.8	226
IE	1 225	14	315	3.6	1 472	16.8	3 434	39.2	129	1.5	2 192	25	8 767
EL	29	4.3		0	98	14.5	64	9.5	5	0.7	480	71	676
ES	4 838	28.4	18	0.1	5 900	34.7	738	4.3		0	5 525	32.5	17 019
FR	107 682	60	4 473	2.5	26 879	15	5 031	2.8	47	0	35 345	19.7	179 457
IT	30 752	39.8	343	0.4	23 066	29.9	6 192	8	5	0	16 828	21.8	77 186
CY	6	0.6	4	0.4	66	7.1	45	4.8	202	21.7	609	65.3	932
LV	9	3.1		0	18	6.3	10	3.5		0	251	87.2	288
LT		0		0	6	22.2	7	25.9		0	14	51.9	27
LU	1 199	22.2	103	1.9	249	4.6	1 556	28.8		0	2 292	42.5	5 399
HU	310	18.4		0	229	13.6	95	5.6	2	0.1	1 053	62.3	1 689
MT	56	13.1		0	18	4.2	289	67.8		0	63	14.8	426
NL	89 064	67	7 173	5.4	16 027	12.1	7 474	5.6	1 577	1.2	11 550	8.7	132 865
AT	11 072	43.8	26	0.1	1 447	5.7	6 042	23.9	1	0	6 705	26.5	25 293
PL	271	23	17	1.4	194	16.5	314	26.6		0	383	32.5	1 179
PT	223	20.9	9	0.8	434	40.8	72	6.8		0	327	30.7	1 065
RO	5	5.3		0	29	30.5		0		0	61	64.2	95
SI	116	8.1		0	35	2.4	102	7.1	425	29.6	760	52.9	1 438
SK	59	19.3		0	132	43.3	34	11.1		0	80	26.2	305
FI	35 366	69.3	1 756	3.4	4 260	8.3	3 452	6.8	40	0.1	6 178	12.1	51 052
SE	34 513	40.7	356	0.4	16 987	20	1 193	1.4	32	0	31 763	37.4	84 844
UK	55 969	42.6	2 157	1.6	29 745	22.7	34 782	26.5	1 543	1.2	7 041	5.4	131 237

Source: PATSTAT autumn 2011 edition, Amadeus 2012.

percentage by multiplying the ultimate owner's direct share in the intermediate subsidiary by the direct share that the subsidiary holds in the company under consideration. More specific information on Bureau Van Dijk's procedures for calculating total ownership percentages can be found in the Bureau Van Dijk Amadeus Ownership Guide (2008).

In the second stage of the classification process, shareholder information for the small entities is taken into consideration to identify actual SME activity. Further evaluation of shareholder information is deemed irrelevant for companies already identified as large corporate entities. Equally, for the small entities that are independent according to Bureau Van Dijk's 'independence indicator', shareholder data is not subjected to any further inquiry. The 'rest' category of non-independent small entities either qualifies as dependent (26.4 % of corporate applicants) or reports an unknown degree of dependence (15.0 %) according to the independence indicator.

In line with EU directives, the financials of companies that belong to multi-entity business groups are assessed on a consolidated basis. Companies in Amadeus can be members of two types of business group. Companies controlled by another company might be labelled as a more formal type of business group, while those controlled by a natural person who also holds majority stakes in other companies represent a more informal equivalent.

Of the 43,631 small entities that are non-independent, 16,755 have a corporate organisation[10] as majority shareholder (holding 50+ % of the shares). 13,920 of these are 'non-financial' companies – with 'financial institutions' or plain investment vehicles representing the counterpart – established in Europe. They belong to the population of companies covered by Amadeus and more detailed financial information is hence available.[11] For 5,030 of these 13,920 EU majority shareholders, this more detailed information indicates that the available size indicators pertain to the consolidated group rather than the entity and, as such, they can be assessed in isolation to determine business group size. Accordingly, we can determine that, among all corporate applicants, 1.9 % (accounting for 1.1 % of corporate applications) belong to a small business group and 2.8 % (3.9 % of corporate applications) to a large business group.

For the non-independent small entities controlled by companies only reported in Amadeus on an unconsolidated basis or from numbers for which the reporting basis is unknown (among others, the non-European corporate majority shareholders that are only reported in Amadeus as 'shareholder'), the consolidation methodology proposed by the European Commission's (2003) is applied.[12] Consolidated group-level figures are approximated by adding the revenue, staff count and total asset figures of the majority shareholder(s) with their equivalent for the non-independent small entity insofar as they are reported for both. On the basis of this

[10] That is, industrial companies, holding companies and private equity firms. Majority shareholders in the form of institutional investors such as pension and mutual funds/trusts, banks and insurance companies are treated separately.

[11] For non-European shareholders, the only relevant information provided by Amadeus in the shareholder section of the firm in which they hold shares is shareholder turnover, shareholder total assets and shareholder staff count.

[12] Given that the large majority of the parent company financials with a known reporting basis are unconsolidated, we assumed that this would also be the case for the available parent company financials with an unknown reporting basis.

cruder consolidation strategy, an additional 2.5 % of corporate applicants (2.4 % of corporate applications) can be assumed to belong to a large business group.

A similar approach was taken for companies reporting natural persons as majority shareholders. If Amadeus registered these natural persons as holding majority stakes in other companies, the financials across all majority-owned companies were aggregated and compared with large-company thresholds.[13] On the basis of this cruder consolidation strategy, the involvement of shareholder information in size assessment resulted in an additional 0.1 % of corporate applicants (0.1 % of corporate applications) being categorised as member of a large enterprise.[14]

Aggregating the results from all consolidation approaches, in the final analysis, 5.4 % from the 41.9 % non-independent small entities can be reclassified as large companies. This represents 6.4 % of the matched patent volume. The 1.9 % of non-independent small entities assigned to the small business group category account for 1.1 % of the matched patent volume. Obviously, both numbers can only be taken as lower bound estimates of the shares of non-independent small entities that can be reallocated to the large and small company categories respectively, since the Amadeus data covering the ownership structure and shareholder financials is highly incomplete.

Table 6 summarizes the outcomes by mapping the 10 categories of corporate applicants to four aggregate classes: 'large' companies, actual SMEs, companies for which insufficient information is available to classify them as either large or small, and corporate applicants not matched to any company in the Amadeus directory. Multiple categories are linked to the first three classes. The small entities that belong to large business groups and large entities are unquestionably 'large' companies. The small entities backed by (semi-)public actors or institutional investors are excluded from the category of actual SMEs and assigned to the 'large' company category. Questions can be raised about the decision to allocate both to the large business class: access to the resources of their majority shareholder may be less obvious than is the case for other types of majority shareholder. However, it should be noted that, given the low numbers that are reported for both groups, the effects of this decision on the final estimates are limited. The class of 'actual SMEs' consists of independent SMEs and SMEs linked to small business groups. The class of companies for which information is insufficient for reliable firm size determination comprises companies for which ownership (and financial) information is lacking but also small entities

[13] Note that, unlike the case of corporate shareholders, no unique identifier is provided in Amadeus 2012 for natural person shareholders. Therefore, the consolidation of informal business group numbers was preceded by a matching of natural person majority shareholders based on name and address information.

[14] The aggregate size indicators resulting from this cruder consolidation approach should be regarded as estimates of the minimal enterprise size because the consolidation is most likely to remain incomplete. Accordingly, these figures are only used to recognize the entities pertaining to business groups that are at least larger than the SME threshold.

controlled by natural persons. The incompleteness of Amadeus and the possibility that those natural persons hold majority stakes in other companies prevent us from directly classifying these applicants as actual SMEs. The production of reliable estimates of the SME contribution for all corporate applicants, including the non-matched ones and those matched to entities not reporting sufficient size information, requires additional efforts in the form of an extrapolation step.

TABLE 6: SHARES OF PATENTS ATTRIBUTABLE TO ACTUAL SMES AND LARGE FIRMS AFTER THE CLASSIFICATION STAGE (IN %).

Country	*'Large' company*	*SME*	*Matched but unknown*	*Not matched*	*Total*
EU-27	**61.7**	**5.4**	**16.0**	**16.9**	**100.0**
BE	71.4	4.6	12.9	11.1	100.0
BG	21.6	3.2	13.7	61.6	100.0
CZ	29.2	8.4	29.1	33.3	100.0
DK	56.5	13.7	12.7	17.0	100.0
DE	63.8	3.2	17.5	15.6	100.0
EE	8.0	24.3	27.9	39.8	100.0
IE	29.5	3.8	41.7	25.0	100.0
EL	5.3	6.1	17.6	71.0	100.0
ES	35.7	7.1	24.7	32.5	100.0
FR	67.7	3.1	9.5	19.7	100.0
IT	44.8	14.0	19.4	21.8	100.0
CY	3.6	1.0	30.0	65.3	100.0
LV	4.2	2.4	6.3	87.2	100.0
LT	7.4	7.4	33.3	51.9	100.0
LU	25.5	2.2	29.8	42.5	100.0
HU	18.4	0.9	18.4	62.3	100.0
MT	14.1	8.5	62.7	14.8	100.0
NL	77.9	2.0	11.4	8.7	100.0
AT	50.9	6.5	16.2	26.5	100.0
PL	32.5	18.1	17.0	32.5	100.0
PT	28.4	20.9	20.0	30.7	100.0
RO	7.4	1.1	27.4	64.2	100.0
SI	8.8	2.6	35.7	52.9	100.0
SK	25.2	8.2	40.3	26.2	100.0
FI	74.5	3.2	10.2	12.1	100.0
SE	45.0	6.1	11.5	37.4	100.0
UK	56.3	14.9	23.4	5.4	100.0

Source: PATSTAT autumn 2011 edition, Amadeus 2012.

It goes without saying that the levels of unknown cases (SME or not) in the previous table do not allow to arrive at a precise estimate of the share of SMEs in patenting across European countries. This only becomes feasible after an additional phase (based on exhaustive sampling) in which extrapolation (with a 95% confidence level) becomes feasible (see table 7). Such additional, time/resource consuming efforts stem directly from the incompleteness of current financial databases.

TABLE 7: SHARES OF PATENTS ATTRIBUTABLE TO ACTUAL SMES AND LARGE FIRMS AFTER THE EXTRAPOLATION STAGE (IN %)

Country	*Large (%)*	*Unknown - non-matched (%)*	*Unknown - matched (%)*	*SME (%)*
EU-27	**78.9**	**2.3**	**1.2**	**17.6**
BE	79.2	0.0	2.6	18.2
BG	36.8	9.5	0.0	53.8
CZ	60.1	2.8	0.0	37.1
DK	67.2	3.7	1.5	27.6
DE	84.9	2.8	2.0	10.3
EE	19.9	0.0	2.3	77.8
IE	50.4	2.6	2.8	44.1
EL	46.1	14.4	0.0	39.6
ES	61.3	3.8	0.0	34.8
FR	83.4	2.4	0.1	14.1
IT	60.8	2.0	0.2	37.1
CY	28.3	9.0	0.0	62.7
LV	33.7	9.5	0.0	56.8
LT	50.5	0.0	0.0	49.5
LU	49.4	11.5	0.0	39.1
HU	59.3	3.8	0.0	37.0
MT	23.4	2.3	0.0	74.3
NL	83.8	0.9	0.7	14.6
AT	77.2	1.9	0.0	20.9
PL	62.0	0.0	4.0	34.0
PT	42.7	6.0	2.8	48.5
RO	46.9	0.0	5.6	47.5
SI	62.8	3.0	0.0	34.2
SK	43.7	5.3	0.0	51.0
FI	83.6	2.2	0.9	13.2
SE	78.8	2.2	0.2	18.9
UK	62.1	1.4	1.2	35.3

Source: PATSTAT autumn 2011 edition, Amadeus 2012, internet searches based on applicant name.

Table 7 reveals that 3.5 % of patent volume remains assigned to corporate applicants of unknown size. 17.6 % of the applications originate from innovative activities in SMEs, whereas 78.9 % are filed by large companies. More established knowledge economies such as Germany, France, The Netherlands, Sweden, Belgium, Austria, and Finland seem to show lower proportions of patents filed by SMEs than in the EU overall. However, exceptions to this observed tendency are noted as well: equivalent to some of the more peripheral and more recent Member States, advanced economies such as Denmark, and especially the United Kingdom, show SME contributions above the overall European level.

Concluding notes and suggestions for future research

We report on a previous effort to create a comprehensive picture of the SME contribution in corporate patenting in Europe (Vervenne et al., 2021). To achieve this, we assessed the feasibility of constructing a European-wide indicator reflecting the contribution of SMEs to the development of (patented) technology, based on existing databases (PATSTAT and Amadeus). The attempted integration of IP and financial data from these two databases, as highlighted throughout this paper, presents both a significant methodological advancement and ongoing challenges.

The methodology that was presented combined automated matching, disambiguation, and classification stages with targeted additional search efforts in an extrapolation step to arrive at a final estimate of the SME contribution to the stock of patents filed by European companies. Several significant methodological challenges had to be addressed in arriving at comprehensive estimates of the share of patenting by SMEs. Given the heterogeneity in name variants across both databases, as well as incomplete information regarding firm size and independence, a series of data treatment processes—harmonization, matching, and disambiguation—were required. Additionally, further sampling was necessary to produce an indicator that represented an acceptable range of estimated shares.

It became clear that the extrapolation step remains critical to the entire exercise, primarily due to the fragmented nature of ownership information compared to scenarios where only entity size indicators are considered.[15] These additional, time- and resource-consuming efforts arise directly from the incompleteness of current financial databases.

These challenges underscore the pressing need for more comprehensive and historically consistent financial databases. Significant gaps in current data sources limit the full potential of integrated IP-financial datasets. Without more complete financial repositories, future research will continue to face limitations in fully realizing the potential of IP-financial data integration.

We strongly advocate for the development of enhanced financial databases that provide accurate longitudinal firm-level data. Doing so would not only benefit academic research but also equip policymakers and practitioners with the tools necessary to better assess the dynamics of IP practices and their financial impact on firms.

[15] It must be noted that later versions of Bureau Van Dijk products have shown improvement in terms of ownership information coverage.

References

- Belderbos R., D Faems, B Leten, B. Van Looy (2014) Technological Activities and Their Impact on the Financial Performance of the Firm: Exploitation and Exploration within and between Firms . Journal of Product Innovation Management 27 (6), 869-882
- BOIP/FOD Economy (2021) Lecluyse, L., Callaert, J., Pellens, M., Van Looy, B., De Voldere, I. (Dec 2021). Econometric analysis of the use of IPR in Belgium. Analytical report BOIP / FOD Economie, Belgium. Available online: https://economie.fgov.be/en/publication/econometric-analysis-use
- Du Plessis, M., Van Looy, B., Song, X., Magerman, T. (2009). Data production methods for harmonized patent statistics: Patentee sector allocation.
- Leten, B., Belderbos R. & Van Looy B. (2016) Entry and Performance in New Technology Domains: Internal Capabilities and Environmental Influences. Journal of Management Studies, 53, 8, 1257-1291.
- Leten B., Belderbos R. & Van Looy B. (2007) Technological diversification, coherence and performance of firms. Journal of Product Innovation Management, 24, 6, 567-579
- Magerman, T., Van Looy, B. & Song, X. (2006). Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization. EUROSTAT Working Paper and Studies, Luxemburg.
- Van Looy, B., du Plessis, M. & Magerman, T. (2006). Data production methods for Harmonized patent indicators: Assignee sector allocation.' EUROSTAT Working Paper and Studies, Luxemburg.
- Vervenne J-B, Callaert J. Hoskens M., Van Looy B.(2020) To what extent do SMEs contribute to Europe's patent stock? A methodological outline for creating a Europe-wide SME technology indicator. Scientometrics 127 (6), 3049-3082
- Callaert, J. M Du Plessis, J Grouwels, C Lecocq, T Magerman, B Peeters, Van Looy B. (2011) Patent statistics at Eurostat: Methods for regionalisation, sector allocation and name harmonisation. Eurostat Methodologies and Working Papers, Luxemburg

FIGURE 1: PROCESS FLOWCHART FOR MATCHING PROCEDURE