



OECD HAN DATABASE

A solution on the harmonisation
of applicant names for patent statistics

Hélène DERNIS

OECD, Directorate for Science, Technology and Innovation



OECD and IP Statistics

Patent statistics

- Unique insights into the **invention processes**
- Good proxies for **technology-related & product innovations**
- Good complement to existing data (R&D, innovation, trade, etc.).

Patents indicators to investigate:

Technological performance ;

Emerging technologies;

Knowledge diffusion and dynamics of technical change;

Geographical properties of inventive process;

Innovation & firm dynamics;

Economic value of inventions ;

Role of universities in technological development ;

Globalisation of R&D activities;

etc.



Explore the patenting behaviour of firms

Analysing the **relationship between patenting and firm dynamics**:

- Patents are **positively associated with the survival of firms** (*Wagner & Cockburn, 2010*)
- Patents have a significant **impact on firm-level productivity and market value** (*Bloom & van Reenen, 2002*)
- Decision to patent is **correlated with firm's subsequent growth** in high-tech start-ups (*Helmets & Rogers, 2011*)
- Patents increase the likelihood of **obtaining venture capital and securing liquidity** (*Hall & Harhoff, 2012*)

→ **Link patent data to firm level data**



Linking patents to firm level data 1/2

Harmonisation is key to matching:

- **No firm identification codes** in patent documents
- **No unique patent applicant name** in databases, due to:
 - Typing errors;
 - Use of acronyms or short names (e.g. I.B.M.);
 - Translations;
 - Indication of legal forms (A.G., Gmbh, S.A., Int, Inc. Ltd., Oy...);
 - Punctuation signs; ...

→ **Country-specific dictionaries** have been developed
(building on Eurostat/KUL - see Magerman et al (2009). Data Production Methods for Harmonized Patent Indicators: Patentee Name Harmonization. EUROSTAT Working Paper and Studies, Luxembourg)



Linking patent to firm level data 2/2

Patent applicants names are matched to company names listed in the **ORBIS©** database using a tailored-made algorithm – ***Imalinker*** (*Idener Multi Algorithm Linker*)

- Matching performed on a country by country basis
- Harmonisation of names with country specific dictionary
- String matching algorithms (token/string metric based)
- High precision (thresholds of high matching scores to minimise the false-positive and false-negative matches)
- Manual controls to adjust the precision

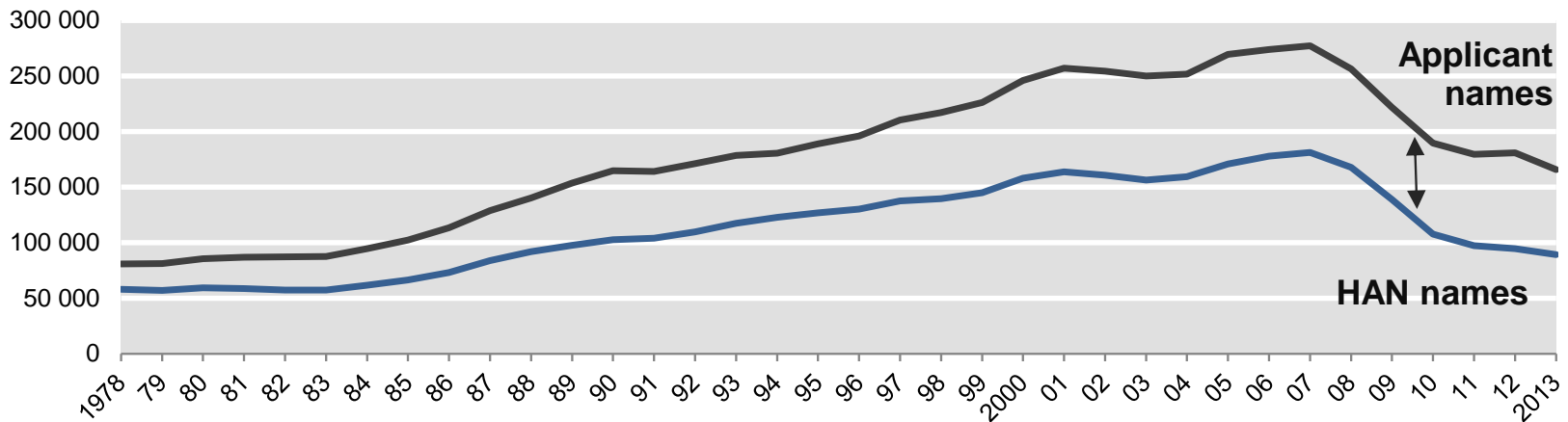
→ **OECD HAN database** (*Harmonised Applicant Names*)
derived from the name harmonisation procedure
and/or on the outcomes of the matching

→ *Raw data available upon request & part of PATSTAT.*

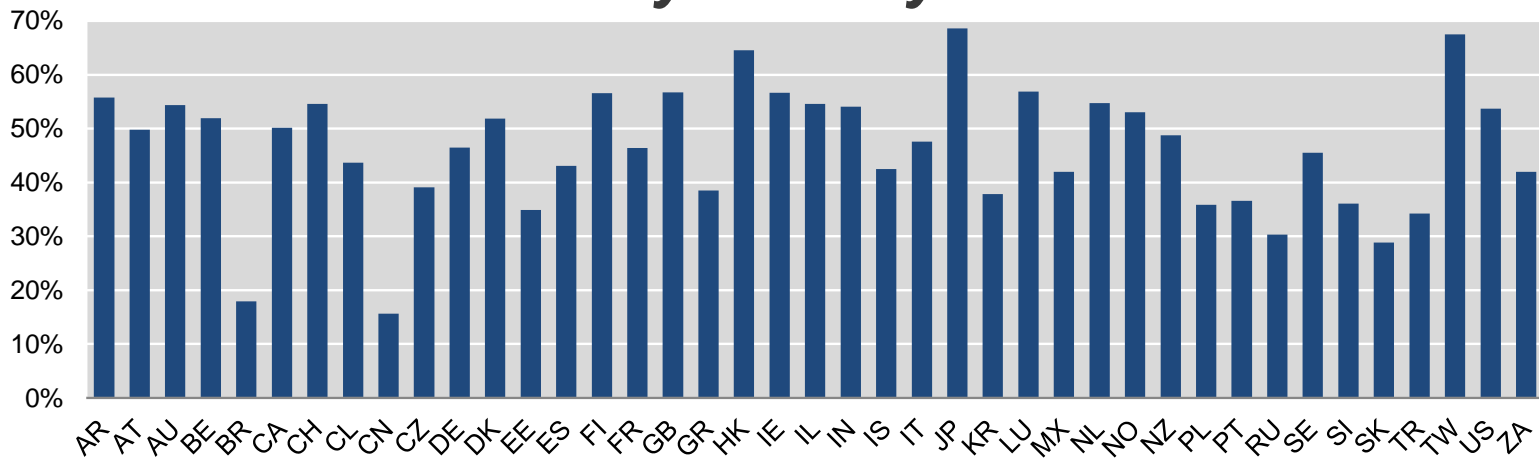


HAN Database

Reduction of name variation in PATSTAT, 1978-2013



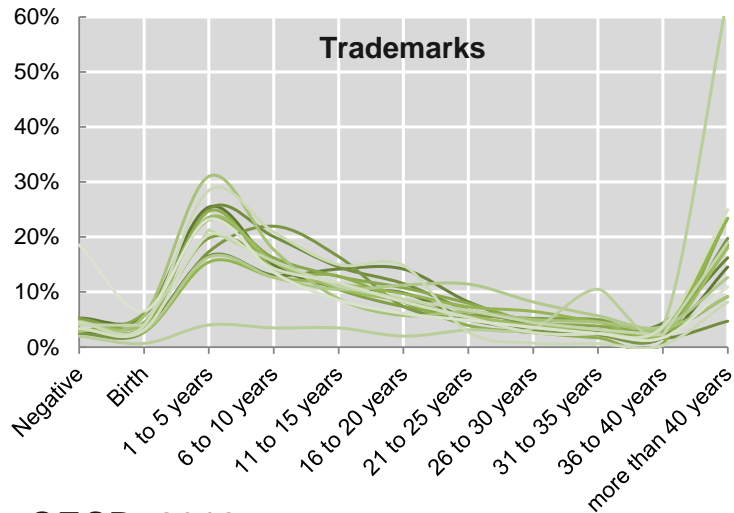
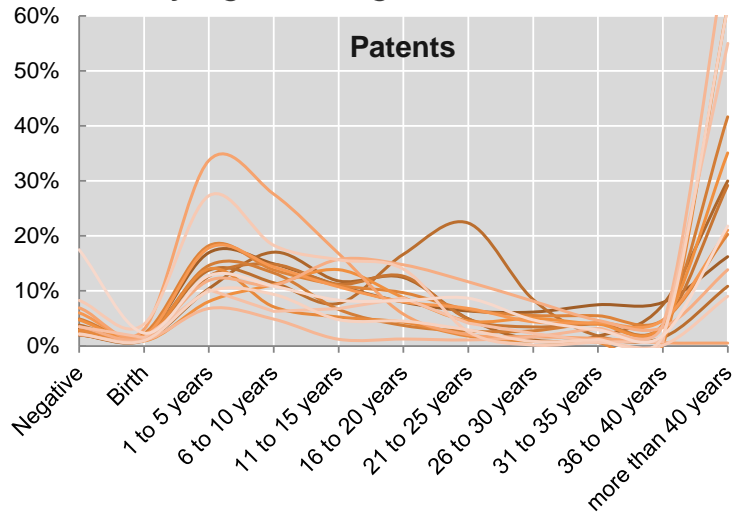
by country



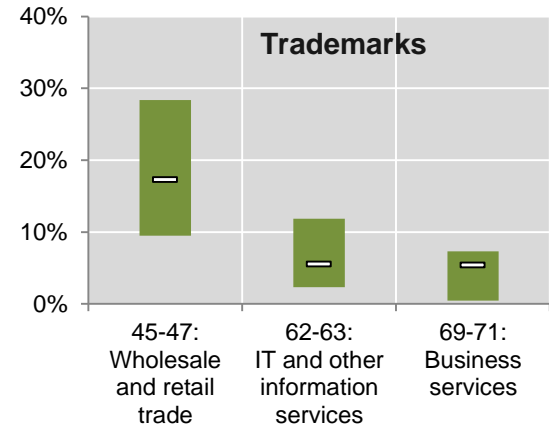
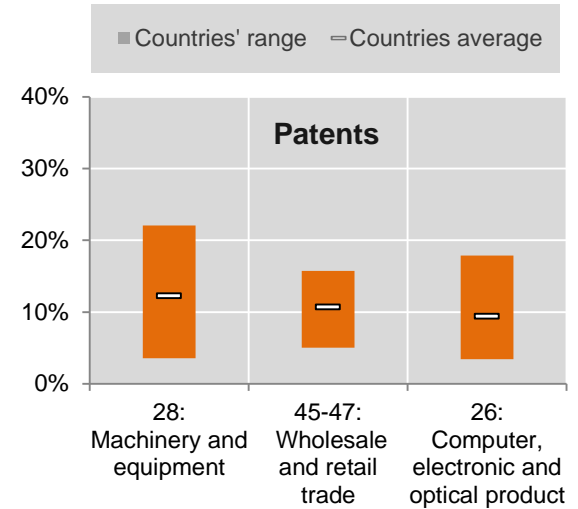


Patenting and trademarking firms by firm age; by industry

**Distribution of patents and trademarks
by age of filing firm, 2000-10**



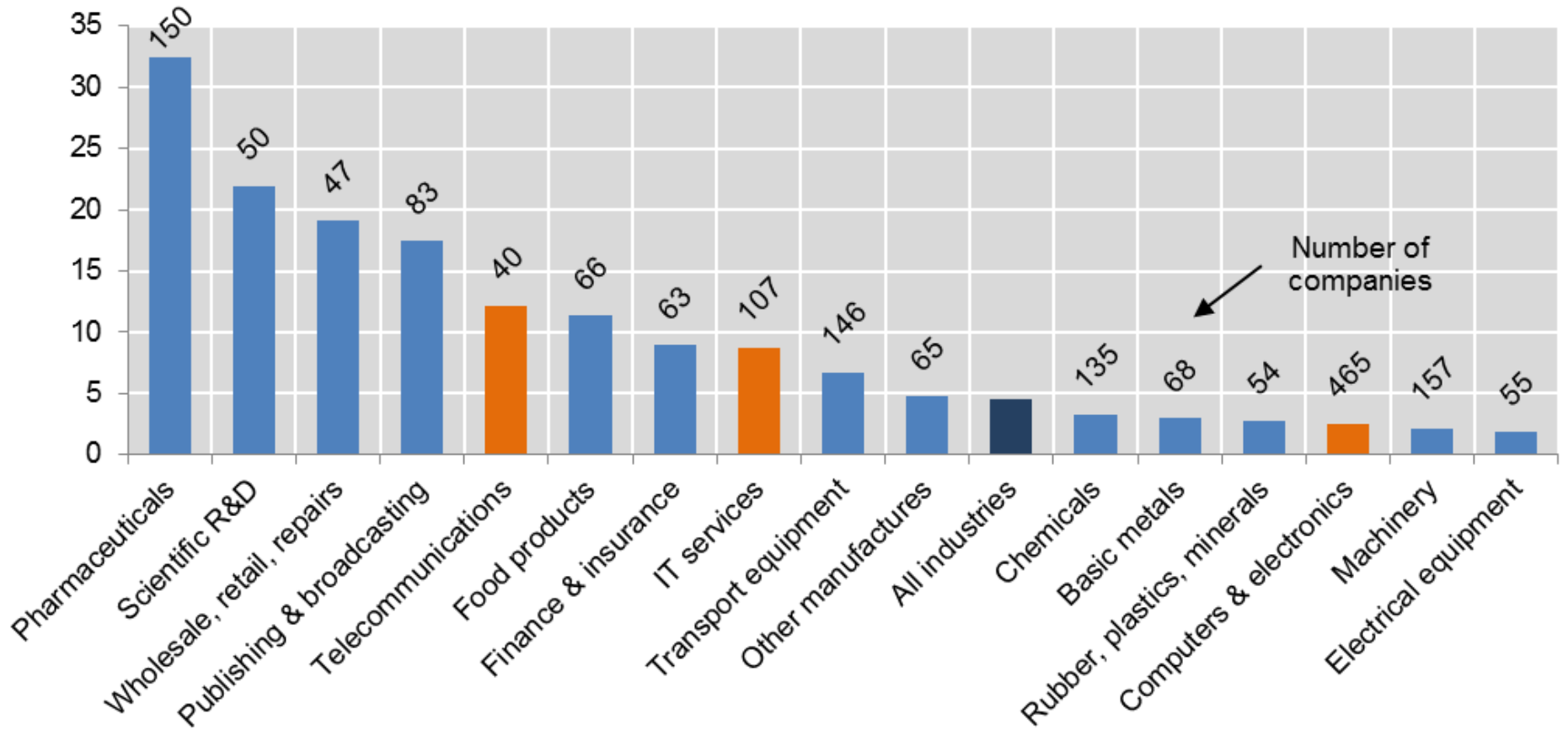
**Top 3 industry classes
trademarking and patenting firms, 2000-10**





R&D investment per patent

Million € spent per IP5 patent family





Conclusions

Standardisation / harmonisation of applicant names is key to:

- Better assess the patent portfolio of companies;
- Improve the relevance of data;
- Reduce the multiplication of harmonisation efforts undertaken by most IP offices, institutions and researchers worldwide.

To avoid harmonisation ex post, would IP offices be able to request companies to provide their company identification number when filing for IPR (e.g. official company number)?



Thank you

Contact: helene.dernis@oecd.org

Access to OECD HAN database: sti.microdatalab@oecd.org

OECD work on IP Statistics: oe.cd/ipstats

IP related statistics available in the OECD STI Scoreboard 2015 and on OECD.Stat portal

OECD.StatExtracts
Complete databases available via SourceOECD, OECD's Library

Patents by Main IPC¹ : Patents - total and technology domains

Change data selection: Patents Office & Trade Patent Families [4 / 4] Reference country [1 / 2] Country [90 / 196] Technology do

Data extracted on 07 May 2016 08:51 UTC (GMT) from OECD.Stat

Country	1999		2000		2001		2002		2003		2004	
	Time	ΔV	Time	ΔV	Time	ΔV	Time	ΔV	Time	ΔV	Time	ΔV
Australia	351.7	393.4	342.0	388.3	1 011.0	1 071						
Austria	1 069.2	1 178.1	1 201.7	1 269.8	1 333.2	1 416						
Belgium	1 330.7	1 298.0	1 203.6	1 296.0	1 323.2	1 445						
Canada	1 602.3	1 636.0	1 602.9	1 762.0	1 817.8	2 165						
Czech Republic	0.0	80.0	71.9	89.0	112.0	111						
Denmark	859.4	939.0	908.5	930.7	1 040.1	1 026						
Finland	1 421.8	1 413.0	1 387.5	1 355.3	1 287.8	1 301						
France	7 199.8	7 284.7	7 270.8	7 352.8	7 871.2	8 232						
Germany	20 509.8	22 021.3	21 774.8	21 629.3	21 608.7	22 094						
Greece	0.0	85.0	71.8	70.0	84.0	77						
Hungary	110.0	120.1	89.3	121.2	128.0	154						
Iceland	36.0	36.2	30.7	37.0	31.1	24						
Ireland	231.0	203.0	247.9	228.1	229.0	200.0	201.2	301.1				
Italy	3 713.3	3 962.1	3 963.1	4 199.9	4 505.4	4 628.4	4 901.9	5 040.3				
Japan	18 768.1	21 720.2	19 937.6	20 228.7	21 248.9	22 241.7	21 014.0	21 676.7				
Korea	1 941.8	1 285.1	1 813.0	2 293.4	3 261.2	4 361.2	4 910.2	5 301.2				
Lithuania	61.2	62.0	72.0	60.7	67.7	112.2	90.0	111.1				
Mexico	41.0	29.2	40.3	48.1	60.6	54.0	63.0	61.2				
Netherlands	2 972.3	3 443.0	3 888.1	4 492.0	4 437.9	5 076.0	5 402.0	5 502.0				
New Zealand	144.4	179.8	146.0	171.8	203.0	184.0	163.8	162.0				
Norway	372.8	400.2	364.6	385.4	337.8	381.7	481.0	487.0				
Poland	36.0	43.4	58.1	64.2	110.0	120.0	142.0	143.0				
Portugal	30.3	42.1	41.3	41.2	84.2	80.0	114.3	110.0				
Slovak Republic	18.4	11.2	12.2	24.0	30.1	29.0	30.7	30.7				
Spain	732.7	800.0	870.7	937.1	941.4	1 198.7	1 334.8	1 470.0				
Sweden	2 201.7	2 285.0	2 102.3	2 013.4	2 010.3	2 152.0	2 341.7	2 300.0				
Switzerland	2 492.0	2 713.4	2 781.7	2 841.0	2 740.7	2 899.0	3 000.0	3 128.0				
Taiwan	22.0	44.0	40.3	67.7	84.0	118.0	166.7	215.1				
United Kingdom	5 790.1	5 988.0	5 978.0	5 905.9	4 458.7	5 378.0	5 293.8	5 559.1	5 332.0			
United States	30 450.0	31 115.4	30 284.2	31 386.1	31 833.8	33 010.3	33 991.9	34 670.7	34 010.4			

