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 (Helmers & Rogers, 2011)
• Patents increase the likelihood of obtaining venture capital and securing liquidity (Hall & Harhoff, 2012)

→ Link patent data to firm level data
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
   
Patent applicants’ names are matched to company names listed in the ORBIS® database using a tailored-made algorithm – *Imalinkerp* (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

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

Patents

Trademarks

Source: OECD, 2013.
<table>
<thead>
<tr>
<th>Industry</th>
<th>Number of companies</th>
<th>Million € spent per IP5 patent family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmaceuticals</td>
<td>150</td>
<td>35</td>
</tr>
<tr>
<td>Scientific R&amp;D</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>Wholesale, retail, repairs</td>
<td>47</td>
<td>17</td>
</tr>
<tr>
<td>Publishing &amp; broadcasting</td>
<td>83</td>
<td>10</td>
</tr>
<tr>
<td>Food products</td>
<td>40</td>
<td>9</td>
</tr>
<tr>
<td>Finance &amp; insurance</td>
<td>66</td>
<td>7</td>
</tr>
<tr>
<td>IT services</td>
<td>63</td>
<td>5</td>
</tr>
<tr>
<td>Transport equipment</td>
<td>107</td>
<td>4</td>
</tr>
<tr>
<td>Other manufactures</td>
<td>146</td>
<td>3</td>
</tr>
<tr>
<td>All Industries</td>
<td>65</td>
<td>3</td>
</tr>
<tr>
<td>Chemicals</td>
<td>135</td>
<td>2</td>
</tr>
<tr>
<td>Basic metals</td>
<td>68</td>
<td>2</td>
</tr>
<tr>
<td>Rubber, plastics, minerals</td>
<td>54</td>
<td>1</td>
</tr>
<tr>
<td>Computers &amp; electronics</td>
<td>465</td>
<td>0</td>
</tr>
<tr>
<td>Machinery</td>
<td>157</td>
<td>0</td>
</tr>
<tr>
<td>Electrical equipment</td>
<td>55</td>
<td>0</td>
</tr>
</tbody>
</table>

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