Identifying the technology profiles of R&D performing firms – A matching of R&D and patent data

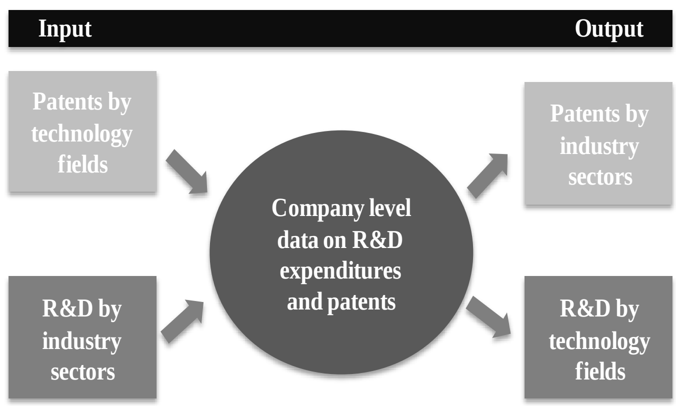
Peter Neuhäusler*, Rainer Frietsch, Carolin Michels and Verena Eckl

*peter.neuhaeusler@isi.fraunhofer.de
Fraunhofer Institute for Systems and Innovation Research ISI (Germany)

The aim of this analysis is to create a concordance list between industry sectors and technologies, enabling us to report the R&D expenditures of companies not only by industries (NACE) but also by technology fields (IPC classes). As especially large and multinational enterprises often are technologically heterogeneous, the statistical classification of economic activities is not able to adequately display companies' R&D expenditures. This implies that we are not aware of how much business R&D is spent for certain kinds of technologies. In addition, numerous other indicators and data at hand are only available at the level of technologies. Besides patents, these include data on support or funding programs or research-oriented indicators, such as employees in public institutions. In order to address this problem, surveying companies is not sufficient. Although they might be able to quantify R&D expenditures by individual projects or in total, they can only seldom provide information on their R&D expenditures by fields of technology.

A further advantage of the creation of a concordance between patents and R&D expenditures at the firm-level is that it also allows us to report patent filings by NACE sectors. Patents are classified via their technological implications within the International Patent Classification (IPC) and are one of the most important indicators for the output of R&D processes. They are often employed to assess the technological performance of companies, technology fields or entire economies (Freeman 1982; Griliches 1990; Grupp 1998).

Figure 1: Schematic representation of the analysis



A large number of patents indicates higher R&D activities and thus a higher innovative output. Reporting patents by NACE sectors therefore gives an overview of the innovative output of industries. This can be used to study which areas of the economy particularly resort on specific kinds of technologies, like ICT for instance.

In order to construct the concordance, we match data on R&D expenditures with patent data at the micro level, i.e. at the level of companies and patent applicants, respectively. The companies' R&D expenditures can in a further step be aggregated at the level of technology fields – for example with the technology field list of 35 WIPO classes (Schmoch 2008) – or vice versa, patents can be reported at the level of industry sectors.

The patent data for the analysis were extracted from the "EPO Worldwide Patent Statistical Database" (PATSTAT), which provides information about published patents collected from 83 patent authorities worldwide. The data on R&D expenditures of German companies are provided by the "Stifterverband for the German Science System", based on a biannual survey of R&D performing firms in Germany.

A first step towards the matching of the two data sources is the harmonization of patent applicant and company names. This includes the removal of special characters, umlauts and company legal forms etc. In a second step, the R & D and patent data were consolidated at the corporate level by applying a string matching algorithm based on the Levenshtein distance. The Levenshtein distance is a measure of similarity between two text strings. As soon as two text strings exceed a certain similarity threshold, the record is stored as a "match". In order to define the optimal similarity threshold, the measures "recall" (number of actual matches) and "precision" (number of exact matches), based on a manually created benchmark dataset of 1,000 randomly selected companies, were calculated. The arithmetic mean between the two measures defines the threshold value as an optimal compromise between accuracy and coverage. In a final step, the unmatched patent applicants with more than 100 patent filings between the years 2005 and 2009 were manually matched.

The results reveal that the matching covers about 40% of all German patent applicants and more than 80% of all patent filings at the EPO and the German Patent and Trademark Office (including transferred PCT-filings) in the year 2008. The coverage for large firms is by far higher than for smaller firms, which explains the difference between the number of matched applicants and filings.
By applying a weighting scheme of patent intensities by R&D expenditures from an external source (Schmoch, Gauch 2004) we are now able to provide a weighted and an unweighted matrix for a conversion from R&D expenditures by NACE sectors to technologies, and vice versa for patents, and are now able to assess R&D expenditures by technology fields. The weighting is hereby an important factor, since we cannot necessarily assume that the R&D expenditures per patent are equal across technology fields (a patent in pharmaceuticals, for example, is more expensive than in automobiles). The results show that the (weighted) R&D expenditures in Germany are highest in transport technologies, followed by pharmaceuticals, digital communication technologies, measurement technologies, basic communication processes and electrical machinery and apparatus.

In order to qualify our results we compared our findings with information from the German Ministry of Education and Research on nanotechnology, information from German industry associations on pharmaceuticals and biotechnology as well as findings from the "Future Filings Survey" of the European Patent Office (2010). It can be shown our results are similar to the numbers provided in these reports.

References
European Patent Office (EPO) (2010): Future Filings Survey 2009. Inventions of Applicants Regarding Patent Applications at the European Patent Office and Other Offices: European Patent Office.

Freeman, C. (1982): The Economics of Industrial Innovation, London: Pinter Publishers.

Griliches, Z. (1990): Patent Statistics as Economic Indicators: A Survey. In: Journal of economic literature, 28, S. 1661-1707.

Grupp, H. (1998): Foundations of the Economics of Innovation - Theory, Measurement and Practice, Cheltenham: Edward Elgar.

Schmoch, U. (2008): Concept of a Technology Classification for Country Comparisons. Final Report to the World Intellectual Property Office (WIPO), Karlsruhe: Fraunhofer ISI.

Schmoch, U.; Gauch, S. (2004): Innovationsstandort Schweiz, Studie für das Bundesamt für Berufsbildung und Technologie, Karlsruhe: Fraunhofer-ISI.