

# **Traces of Prior Art**

## **An analysis of non-patent References found in patent documents**

**Julie Callaert<sup>1</sup>, Bart Van Looy<sup>1,2</sup>, Arnold Verbeek<sup>2</sup>,  
Koenraad Debackere<sup>1,2</sup> and Bart Thijs<sup>2</sup>**

<sup>1</sup> K.U.Leuven, Faculteit ETEW, Research Division INCENTIM, Naamsestraat 69, 3000 Leuven

<sup>2</sup> K.U.Leuven, Steunpunt O&O Statistieken, Dekenstraat 2, 3000 Leuven

## Abstract

Research in the area of innovation has pointed out the relevance of conceiving innovation as processes in which a multitude of actors and a variety of interactions play a role. Integrative notions like (national) innovations systems, (techno-scientific) networks, or the triple helix metaphor, have been widely accepted as relevant to grasp the complexities entailed. This development highlights the need for indicators that mirror the dynamics involved. This contribution presents an in-depth examination of the role of “Non-Patent references”, found in patents. After examining the occurrence of these references in the USPTO and EPO patent systems, the precise nature of these references is delineated by means of a systematic content analysis of two samples of non-patent references (n=10.000). Our observations reveal the relevance of ‘non-patent references’ for developing indicators to depict the proximity of technological and scientific developments. Application areas, limitations and directions for future research are discussed.

## 1. Introduction

Over the last decades, models on innovation and innovation policies became increasingly systemic; innovation processes are now looked upon as continuous flows of interactions between companies, knowledge generating institutes like universities and research organizations and governmental agencies and institutions. Notions capturing these innovation dynamics include the concept of (national) innovation system (Freeman, 1987, 1994; Lundvall, 1992; Nelson, 1993; Nelson and Rosenberg, 1993; Mowery and Nelson, 1999; Dosi, 2000), the Triple Helix metaphor (Leydesdorff and Etzkowitz, 1996, 1998; Etzkowitz and Leydesdorff, 1997, 1998) and the scientific network concept (Pavitt, 1997; Steinmuller, 1994; David et al., 1997). Central to these models is the multiple-actor perspective and the non-linear nature of knowledge flows.

For practitioners and science and technology (S&T) policy makers, addressing innovation-related concerns requires indicators that provide insight into the structure, effectiveness and dynamics of innovation systems. For these purposes, scientific indicators, based on publication output, have been pioneered by de Solla Price (1963), Pritchard (1969), van Raan (1988), Schubert, Glänzel and Braun (1989) amongst others. Similar developments at the level of technology took place whereby patent information plays a central role (de Solla Price, 1963, 1965; Schmookler, 1966; Callon et al., 1986; Griliches, 1990; Narin and Noma, 1985). Bibliometric indicators pertaining to both scientific and technological activity are nowadays gathered on a regular base, providing reference points for mapping and analyzing activities and their outcomes (e.g. the European Science and Technology Indicator Reports; the US NSF Reports on Science and Engineering Indicators).

At the same time, the majority of indicators focus exclusively on either scientific or technological activity. To quantitatively grasp the full body of what is referred to as “innovation system”, indicators reflecting relationships between the different activity realms seem highly relevant. In this contribution, we examine the usefulness of information that is provided in patent documents for mapping and assessing connections between technological activity development and scientific activities. Patents are documents issued and *published* by an authorized governmental agency, granting the exclusive right to the applicant to produce or use a specific new device, apparatus or process for a limited time period. This published information provides different possibilities for analyzing the specific environment in which development efforts take place (Trippe, 2003) and hence to arrive at indications of science-technology exchanges. Firstly, interpersonal inventor-author linkages can be assessed: to what extent are inventors also actively engaged in scientific activities, which makes them appear as authors within scientific publications? Examples of this approach, whereby individuals - crossing the boundaries of science and technology spheres- act as the unit of analysis, can for instance be found in the work of Noyons et al. (1994), Packer and Webster (1996), Balconi et al. (2004) and Murray (2004). A second approach consists of looking at organisational entities that appear both as assignee in patents and as affiliation in publications, shifting the level of analysis towards the organizational level. Thirdly, and situated at the same level of analysis, mapping and assessing co-patenting behaviour – involving companies and research organizations including universities – can be envisaged, be it that the low levels of occurrence of co-patenting might limit wide applicability (Hicks and Narin, 2004). A final way of

revealing science-technology associations, based on patent documents, consists of examining the nature of the so-called ‘prior art’ found in published patent documents; the focus of this article. Here, one engages in mapping and analyzing the nature and the sources of the citations provided by the patent examiner and/or the inventor when s/he considers them relevant for assessing the invention and the claims it entails.

In terms of content, several types of prior art can be distinguished. A distinction is generally made between patent references and other – mostly scientific - references<sup>1</sup>. Whereas the work of Jaffe et al. (2002) and others (Fleming and Sorenson, 2001, Harhoff et al., 2003; Reitzig, 2004) focused on the role of *patent* references and citations (e.g. as an indication of the value of patents), Narin and his colleagues pioneered the role and possible contribution of non-patent references. Studies in this field have investigated the nature of science-technology relationships as implied by citation links (e.g. Narin and Noma, 1985), the role of public science for developing technology (e.g. Narin et al. 1997), the frequency and nature of occurrence of such interactions in new emerging technology domains (Van Vianen et al. 1990; Meyer, 2000a; McMillan et al., 2000; Tijssen et al.; 2000; Verbeek et al. 2002; Acosta and Coronado, 2003), as well as the relationship between the science intensity of patents – as measured by the amount of other references – and technological productivity (Van Looy et al., 2003b).

At the same time, using non-patent references as indicators of interrelations and dynamics between scientific and technological activity fields has recently been the subject of some debate. Whereas some question the relevance of such citation analyses for indicating direct links (Tijssen, 2002; Meyer, 2000 a,b), others plea for carefulness in using these indicators and point to contextual elements to be taken into account when interpreting them (Michel and Bettels, 2001; Harhoff et al, 2003; Meyer et al., 2003; Van Looy et al., 2003a). A systematic view on the information that is observable, i.e. the amount and nature of non-patent references in patents, would help put these current debates in perspective and allows assessing the feasibility and precise meaning of indicators based on other or non-patent references. In this article, we contribute to the current developments by (1) providing an updated and exhaustive overview of the *amount* of patent and other references to be found in the USPTO and EPO patent systems (covering the period 1991-2001) and (2) examining more closely the *nature* of non-patent references by performing a content analysis on a sample of 10000 other references retrieved from EPO and USPTO patent files. Both analyses will allow us to assess more precisely the relevance of patent indicators based on other references and might add to a proper interpretation of such indicators. Before turning to the results of both empirical assessments, we briefly expand on the exact role that non-patent references play in the context of patent procedures. This should provide a proper background for interpreting the meaning of non-patent references and hence of indicators that build on this information.

---

<sup>1</sup> Patent references differ not only from ‘other references’ in terms of the nature or content of the cited documents. Extracting and assessing ‘other references’ is also a more complicated endeavour, due to the idiosyncrasies in terms of reporting such references as well as the multitude and variety of written documents being cited (for an extensive overview, including the outline of an adequate parsing method, see Verbeek et. al., 2002).

## 2. A closer look at the role of Non-Patent References

As was clear from the introduction, the analysis of non-patent references has lately received considerable attention. At the same time, some concerns have been expressed about the exact denotation of these references. Proponents of the use of non-patent references sometimes portray the scientific references found in this universe as signalling a direct influence of science on technology. (e.g. Narin et al. 1997). Meyer (2000b, 2001) however, after performing a number of detailed patent case studies, concludes that non-patent references should not be interpreted as indicating a direct and uni-directional link or influence from science to technology. Tijssen (2001) points in a similar direction: non-patent references should be considered a general indicator of interaction between science and technology, rather than as the reflection of scientific sources leading directly to the invention (Tijssen, 2001, p. 39). A closer look at the specific role of prior art references in the patent application is highly relevant when looking into the nature and the occurrence of such references, especially when formulating implications for indicator development.

As mentioned, patents are documents issued by an authorized agency, granting exclusive right to the applicant to produce or to use a specific new device, apparatus or process for a limited period. They are granted to the applicant after an examination that focuses on the novelty, inventive activity and industrial applicability. During the granting process, patent examiners review the prior art that pertains to the invention. The USPTO system obliges applicants/inventors to provide all relevant prior art that they know of (including patents and other written documents). Such ‘duty of candour’ does not exist in the EPO, where applicants/inventors provide only references that are considered relevant. It can be noted at this stage that, due to different disclosing obligations, the number and type of references can differ considerably between both patent systems. This will become clear in the empirical section. In both systems, however, patent examiners do not limit themselves to the prior art that was signalled by inventors and/or applicants. Based on information archives and databases, patent examiners decide which references are relevant for assessing the patent and its constituting claims. As such, the front page of patent documents include the *examiner-given* references, used to decide on granting and including restricting claims, besides the information pertaining to the invention, assignee(s) and inventor(s). It should be clear that these front-page references do not necessarily coincide with references that are provided by the inventor: the latter may be omitted by the examiner and/or the examiner may add references that were not mentioned at first instance.

The function of front-page references, introduced during the granting process, is to help evaluate the novelty and inventiveness of the claims and their applicability, and to contextualize the claims being made. As stated in the USPTO patent examination procedure manual: “the basic purpose for citing prior art in patent files is to inform the patent owner and the public in general that such patents or printed publications are in existence and should be considered when evaluating the validity of the patent claims. Placement of citations in the patent file along with copies of the cited prior art will also ensure consideration thereof during any subsequent reissue or re-examination proceeding.” (Manual of Patent Examining Procedure, USPTO; see also directives 35 U.S.C. 301 and 37 CFR 1.501). Notice also that in USPTO patent documents, the distinction between different rationales for including references (in terms of different reasons for citation) is not manifest. The EPO on the other

hand does explicitly mark the different categories of references. Distinctions are made between individual documents (e.g. publications, other patents) that show essential features of the invention under review (type 1), or that question – part of – the novelty or inventiveness of these features. This category is marked with the letter ‘X’. Y-documents have the same function, but they appear in combination with another document, implying that there are always two Y references in a search report. A third category of references is marked by the letter ‘A’. These relate to the technological background of the invention, showing the general state of the prior art to the current patent application. A-type references contain general information for the applicant; they are important for the assessment of the so-called ‘inventive step’, which forms the second basic criterion for the approval of a patent (see Schmoch, 1993; Michel and Bettels, 2001).

To some extent, the specific role of references in patent applications differs from the role that references or citations play in scientific publications. Article references indicate sources of influence or serve as reference points to delineate differences (novelty). They are introduced by the authors (sometimes with some support from reviewers), implying that the cited references are always known to the author(s). Therefore, it can be argued that the cited references have had an influence on the genesis of the ideas and insights that are developed in the citing article or paper. This is not necessarily the case for the front-page references in patent documents. These might be added by examiners, and in some cases, the inventors and/or applicants are not even aware of the existence of the cited documents. Therefore, as documented by Meyer (2000b) and Tijssen et al. (2000), some of the knowledge captured in references that appear in patent documents were not or only very indirectly influenced the creation of the invention.

Hence, one should be careful in depicting citations in patents as interactions or direct links of causation between two bits of information. These references are part of the context in which the patent and its claims are situated. The presence of scientific research in the ‘prior art’ description of a patented invention should be considered an indicator of the relevance of scientific findings for assessing and contextualizing technology development. At the same time, it is plausible to state that more scientific references signal more relevance or relatedness between the technology at hand and scientific activity<sup>2</sup>. As such, indicators based on these references might provide useful additional information on science-technology relatedness or vicinity, at least if their presence displays sufficient levels of occurrence. In the next section, we will map occurrence patterns and address the nature of the other references in more detail.

### **3. Comparison of occurrence of “Patent References” and “Non-Patent References”**

In this section, we report on the occurrence of patent and other references that are found in the EPO and USPTO patent systems. For this analysis, all granted patents were considered with application year between 1991 and 2001, with data extraction taking place during the summer of 2002. Table 1 provides an overview of the occurrence of both patent and other references observed in this period.

Insert Table 1 – Occurrence of Patent and Non-Patent References

---

<sup>2</sup> Notice that Science-Technology relatedness as described here also has a counterpart, namely references towards patents found within scientific publications.

From Table 1, it is clear that the majority of patents contain patent references (90% and 98% respectively). This is not the case for non-patent references where the number of patents containing such references amounts to 34% and 38%. Moreover, if references are observed, they are more numerous in the case of patent references, resulting in a share of 83% compared to non-patent or other references. As a consequence, for both USPTO and EPO patents, the average number of patent references per patent ('technology intensity') is about double the average amount of non-patent references per patent ('NPR-intensity'<sup>3</sup>).

Figure 1 shows the evolution in NPR and technology intensity over time. For the EPO data, the evolution in NPR- and technology-intensity over the decade under review seems to have remained relatively flat. The figure shows a small but steady decline of NPR-intensity over time. Also for technology intensity, it seems that the number of technological references needed to delineate and evaluate the claims within the EPO system has hardly changed over time. From a conceptual perspective, at first glance, the 'closeness' to respectively the technologic and scientific spheres has remained rather constant over time (decade under review).

Insert Figure 1 – Evolution of NPR- and technology-intensity over time

Examination of the USPTO patent data leads to a somewhat different conclusion (see Figure 1). First of all, we find that technology intensity increases over time, from 10 patent references per patent in 1991 to almost 13 patent references per patent in 2000 (an increase of 30%;  $p < 0.01$ ). The development in the average number of non-patent references appears to be rather turbulent with a strong increase during the first half of the 1990s (from 4 non-patent references per patent in 1991 to over 9 in 1995) followed by an equally strong decline and subsequent stabilization in the second half of the 1990s. The share of patents containing patent references in the USPTO system remains constant at the level of 90-91% until 1999, after which a decline becomes apparent. The share of patents containing non-patent references increases in the first half of the 1990s, a development that is reversed during the second half. Both developments result in a non-significant trend for the whole 10 years under consideration.

A second aspect that should be highlighted is the concentration and distribution of patent and non-patent references over the different fields of technology. By examining and regrouping the classification of a patent (International Patent Classification), a distribution of references over broader technological areas can be established (the nomenclature used here is based on the technology classification scheme designed by OST in France in collaboration with the Fraunhofer Institute and INPI). The findings are presented in Table 2.

Insert Table 2 – Breakdown of NPR- and technology-intensity per technology domain

For the EPO data, we see that mechanical engineering and machinery show the highest intensity of patent references per patent (technology-intensity of 4.64), followed by process engineering and special equipment (4.46). Electrical engineering fields display the lowest average number of references to patent documents. As for NPR-intensity, chemistry and

---

<sup>3</sup> The term 'science' intensity is mostly used for the average number of non-patent references per patent. This suggests that all other references are references to the scientific literature. As we will see further on, this is not completely accurate. Therefore, at this stage we prefer to use the term "NPR intensity" to refer to the number of non-patent references per patent.

pharmaceuticals show the highest number of NPR's (2.68 non-patent references per patent). Comparing these findings to the USPTO data, we find that process engineering and special equipment display the highest average number of patent references (14.17) followed by instruments and mechanical engineering and machinery (respectively 13.76 and 13.06 references per patent). Here too, chemistry and pharmaceuticals contain the highest average number of non-patent references (see also Verbeek et al., 2002).

Whereas for the USPTO and EPO systems, proportions of the different reference types are comparable in absolute terms (see Table 1), it can be seen that USPTO patents hold on average about 3 times more references than EPO patents. Such an observation could be directly related to the differences in the rationale of citing prior art between the American and the European system. As explained, in the USPTO system, the duty of candour principle postulates that all prior art documents that are in any way considered relevant to the patentability of the invention, must be disclosed. Failing to do so can result in patent litigation and severe penalties. The European system, on the other hand, postulates no such requirement. To this date, no obligation is placed on the applicant or his representative to inform the EPO of any prior art believed to be relevant and no penalties exist for not disclosing relevant prior art (Akers, 2000).

In general, the most important information source of technology development is technology itself (other patents). However, occurrence of other references can be considered non-trivial, especially in certain technological domains most notably in Chemistry & Pharmaceuticals. Therefore, an assessment of the science proximity of technology domains seems relevant, especially in domains that display higher levels of science intensity. If the nature of the references allows for it, a more elaborated set of indicators and analyses becomes feasible. A closer look at the nature of the references in patents, as provided in the next section, helps to uncover possibilities and limitations in this regard.

#### **4. A closer look at the nature of Non-Patent References**

Updated systematic overviews of the nature of other references in patents are scarce, although some efforts in this direction were made in the past. Narin and Noma (1985) reported – for the period 1978-1980 – on average 0.3 other references per patent, which is considerably lower than what we observed in the last decade. Thirty seven percent of these references related to SCI journals, 11% to other journals, 15% to books and 11% to abstracts. The final 26% related to miscellaneous sources. Van Vianen et al. (1990), in their exploration of the science base of technology, found that for a total of 2900 Dutch patents between 1982 and 1985 from all technological classes, 55.7% of the non-patent references were journal citations. Of these, 82% were SCI covered journals. Non-journal references appeared to cite mostly books and abstract services, and to a lesser extent meeting abstracts. Harhoff et al. (2003) also briefly illustrated the fact that not all non-patent references refer to scientific sources. They evaluated 100 patent document records, and found about 60% of non-patent references referring to scientific and technical journals. The remainder was made up largely of references to trade journals, to firm publications or to standard texts in the technical fields e.g. for the classification of chemical substances or specific mechanical designs.

In order to create an updated insight into the nature of the other references, we extracted two samples of approximately 5000 non-patent references from the USPTO and the



EPO databases and we classified them according to the document type that was being referred to. For this analysis, a stratified sample was created according to the distribution of patents over technology domains (International Patent Classification system, 3-digit level). Both samples of non-patent references (EPO and USPTO) were manually checked with regard to the type of documents they referred to. A first round of inspection of the sample of references served as the base for an ad hoc taxonomy of reference types, presented in Table 3. As we will see further on, most of the non-patent references could be categorized in this scheme. Some however remained unclear in the sense that incomplete information did not allow for a precise categorization. They are referred to as ‘other’.

#### Insert Table 3 – Taxonomy of Reference types

In the group of non-patent references, a first distinction can be made between journal and non-journal references. In a narrow sense, only journal references refer to the actual scientific journal literature. The availability of databases pertaining to scientific publications enables the development of indicators on scientific and technological field-interrelatedness as outlined above. Table 4a shows the relative occurrence of journal and non-journal references in the USPTO and EPO samples of non-patent references.

#### Insert Tables 4a and 4b – Occurrence of journal and non-journal references in USPTO and EPO (observed and expected values)

For both USPTO and EPO references, more than half are journal references. An additional check reveals that the SCI database provides an almost full coverage for these journal references, holding respectively 90% and 86% of journal references in our USPTO and EPO sample. Comparing the USPTO and EPO sample using a chi square test reveals that there is a significant difference between both ( $p < 0.01$ ). The expected values are provided in Table 4b.

As can be seen, USPTO references have less journal and more non-journal references than expected (observed values of 2766 and 2242 as opposed to expected values of 2988 and 2020 respectively). For EPO references, the opposite holds (observed values of 3218 and 1803 as opposed to expected values of 2996 and 2025 respectively). So in general, more than half of all non-patent references are journal articles, a somewhat smaller share refers to documents that are not purely scientific in the narrow sense. Journal references are more prominent in EPO patents. It is important to note that non-journal references, although smaller in share, are not inconsiderable, representing 36% and 45% for EPO and USPTO respectively.

In a next step, we therefore took a closer look at the type of documents that are cited and that are not scientific journal articles. Table 5a provides the USPTO and EPO occurrence of the non-journal sources distinguished in the taxonomy (cf. Table 3).

#### Insert Tables 5a and 5b – Occurrence of non-journal sources in USPTO and EPO (observed and expected values)

While the concept of ‘scientific references’ would, in its most narrow sense, relate only to journal references, an inspection of Table 5a reveals additional references that can be considered as scientific, ‘science at large’ so to speak. These include conference proceedings, which often precede or equal journal articles; at the same time, the majority of reference books/databases and books – often referring to abstract services – can be considered scientific information sources. Likewise, some types of non-patent references can be considered ‘technology at large’. Firstly, apart from the actual patent citations, non-patent references were retrieved that are, nevertheless, clearly patent related. Examples are provided in Table 3. In addition, industry-related documents often contain technological or product-related information, thus representing technologically oriented rather than scientific information.

As can be seen in Table 5a, ‘Conference Proceedings’, ‘Industry Related Documents’ and ‘Reference Books/Databases’ are the three most common types of prior art that are cited besides journal references. As stated earlier, conference proceedings, books, and reference books/databases can be considered ‘science at large’. Industry-related and patent-related documents can be seen as technology at large. The actual type of information of Newspapers, Research and technical reports and the unclear category is not as clear-cut in terms of representing scientifically or technology-oriented knowledge sources. At the same time, these categories represent only very marginal shares of the non-journal references.

These observations point out that at least 42% of all USPTO non-journal references refer to scientific knowledge. In addition to the actual patent citations (see Table 1), another 40% of non-journal references relates to technological information. For our EPO sample, 77% of non-journal references are nevertheless scientific in the broad meaning of the word. Technology at large is referred to in 20% of EPO non-journal citations. Although for both USPTO and EPO, the major part of non-patent references is thus scientifically oriented, this phenomenon is again more outspoken in the EPO sample of references. Specifically, a chi square test (expected values are shown in Table 5b, significance at 0.001 level) reveals that conference proceedings and reference books/databases are much more prominent in EPO references.

Such differences can be traced back to the different procedures followed in the USPTO and EPO. EPO examiners – who are responsible for the search report – have a broad range of standard electronic databases available when performing searches for related documents. Documents cited in USPTO references, on the other hand, must be included in hard copy. This could explain why in the EPO references, relatively ‘more’ references were to be found in the category ‘Reference Books/Databases’ (the latter being the standardised sources available to EPO examiners). In addition, it can be observed that in the USPTO sample ‘Patent-Related Documents’ are more prominent. This category includes legal documents that represent patent litigations. The stronger prominence of such documents in the USPTO references is most probably due to the duty of candour, which is specific to the US system. Inventors and their representatives are obliged to inform the USPTO of all prior art documents that are considered relevant to the patentability of the application.

## 5. Applications, Implications and Limitations

This analysis of the occurrence of references in patents clearly indicates the non-trivial nature of non-patent references. In addition, when looking at the nature of these references, a majority consists of references to the scientific literature. These observations allow for the conclusion that that developing recurrent, robust indicators based on these references is plausible. Such indicators can depict the extent to which technology development is situated within the vicinity of scientific findings, and they offer multiple possibilities for mapping and analyzing technological activity along this dimension.

The most straightforward use of an indicator based on non-patent references consists of counting backward citations towards scientific articles as a proxy of the ‘science intensity’ of patents. Mapping the variety of science intensity of technology development across fields, domains and even actors can be relevant when analyzing differences in terms of technological effectiveness (for an illustration on the level of nations, see Van Looy et al., 2003b).

It goes without saying that, once non-patent references have been identified, science intensity can be disentangled in a more substantive manner. Scientific disciplines, as well as affiliations of the authors and institutions involved can be introduced in consecutive analysis. Linking the technology domain of the citing patent to the science field of the cited publication for instance, results in matrices that represent the presence of specific scientific disciplines and that relate them to different technological domains (Schmoch, 1997; Verbeek et al, 2002b). Table 6 illustrates this approach by depicting technology domains in the columns and science domains in the rows. Cell values (e.g. cell “chemistry” - “optics”) represent the number of journal references that are classified in scientific discipline “chemistry” and that were found in patents situated in technology domain “optics”. As such, the cell values can be interpreted as an indication of the extent to which technology domains are situated close to research carried out in different scientific fields.

Insert Table 6 – Illustration of the interaction pattern and intensity between technology and 10 main science fields

When constructing such a linkage matrix, introducing time can uncover the development of science intensity, allowing us to assess the presence and nature of technology life cycle dynamics. Further analytical possibilities arise when drawing document affiliation information into the equation. For example, an analysis of contributing institutions from cited articles can uncover important knowledge ‘providers’ in one or more technology domains. Likewise, patent assignees from the citing patents can be linked to author information from the cited article and as such provide an insight into knowledge “flows” (see also Tijssen, 2001). Figure 2 provides an example of such an approach whereby the geographical origin of citations in patents is mapped, illustrating geographical flows of science towards technology (see also Verbeek et al., 2003). Note that, as with science intensity, introduction of the time dimension for all of these indicators and approaches can reveal interesting trends and developments.

Insert Figure 2 – Geographic distribution of citation flows to the published literature present in USPTO patents

While possibilities for indicator development are numerous, some limitations should be taken into account in accurately interpreting results from using such indicators. One limitation was clearly illustrated in this article. When conducting large-scale analysis for examining e.g. patent-to-patent citations and patent-to-paper citations, one is confined to database availability and limits. For large-scale analysis of non-patent references in which information is needed on characteristics of the cited documents, one has to rely on publication databases that provide the information needed in a systematic and consistent way. The most widely adopted publication database for such purposes is the ISI Web of Science database, covering a large share of the scientific journal literature in many scientific disciplines. The other types of non-patent references – as well as the journal articles not covered in the database – are not captured in such systematic databases and are, therefore, much harder to include in large-scale quantitative analyses. Our analysis on occurrence revealed that 50% to 55% of non-patent references are Web of Science - covered journal references that can be included in large-scale analysis. Secondly, our findings reveal that the amount of scientific references found in patents differs between technological fields. This implies that the relevance of using such references varies, depending on the technology domain under consideration. In addition, one should keep in mind that citation analysis is based solely on codified knowledge and information flows. Obviously, science-technology interactions are not limited to the citation of documents. This method is but one – and even a partial one – way to identify science-technology interactions. So part of the picture – e.g. informal, non-traceable flows of tacit knowledge – remains uncovered. Finally, a proper interpretation of indicators based on other references should take into account the context in which patent documents are situated. Several differences between the EPO and USPTO system were pointed out above. They can mostly be attributed to the fact that the EPO system does not have the USPTO's duty of candour, which can influence reference patterns. When interpreting any patent related indicator, one should always be aware of the procedure that has preceded the grant or application of the patent documents under consideration.

Notwithstanding these concerns, our findings reveal that citations in patents allow developing non-trivial and robust indicators. The majority of all non-patent references are journal references, providing ample possibilities for large-scale analyses focusing on the extent to which technological developments are situated within the vicinity of scientific knowledge.

**Table 1 - Occurrence of Patent and Non-Patent References (USPTO – EPO)**

<i>USPTO granted patents with application year between 1991 and 2001</i>					
Total # patents (1)	1,299,817	Total # references	17,757,797		
# patents containing patent references	1,173,593 (90%)	# patent references	14,738,854 (83%)	Technology-intensity <i>With (1) as denominator:</i>	12.55  11.33
# patents containing non-patent references	445,466 (34%)	# non-patent references	3,018,943 (17%)	NPR-intensity <i>With (1) as denominator:</i>	6.77  2.2
<i>EPO granted patents with application year between 1991 and 2001</i>					
Total # patents (1)	342,704	Total # references	1,698,218		
# patents containing patent references	334,413 (98%)	# patent references	1,404,241 (83%)	Technology-intensity <i>With (1) as denominator:</i>	4.20  4.09
# patents containing non-patent references	130,511 (38%)	# non-patent references	293,977 (17%)	NPR-intensity <i>With (1) as denominator:</i>	2.25  0.86

**Table 2 - Breakdown of NPR- and technology-intensity per technology domain**  
 (USPTO and EPO patents with application year between 1991 and 2001)

Technology field	EPO patents		USPTO patents	
	Technology intensity	NPR-intensity	Technology intensity	NPR-intensity
Electrical Engineering	3.74	2.24	11.25	4.83
Instruments	4.34	2.32	13.76	6.72
Chemistry, Pharmaceuticals	3.87	2.68	11.39	13.23
Process Engineering, Special Equipment	4.46	2.08	14.17	4.66
Mechanical Engineering, Machinery	4.64	1.74	13.06	3.27

**Table 3 - Taxonomy of reference types**

Category	Sub-category	Description	Illustration
<b>Serial journal references:</b>	<i>SCI covered</i>	References to scientific publications published in serial journal literature and covered by the scientific database, The Science Citation Index (a Thomson-ISI product)	*Schoentag et al. (1987), Cancer Research 47: 1695-1700 *MacDonald et al.; The American Journal of Cardiology; 62: 16J-27J (1988); "Preclinical Evaluation fo Lovastatin".
	<i>Not SCI covered</i>	References to scientific publications published in serial journal literature but NOT covered by the scientific database, The Science Citation Index (a Thomson-ISI product)	*Pharmazeutische Zeitung, 124 No. 20, May 17, 1979, pp. 946-957. *"Formation of Si-Si Bonds From Si-H Bonds in the Presence of Hydrosilation Catalysts", Organometallics 1987, 6, 1590-1591, Katherine A. Brown-Wesley.
<b>Non Journal references:</b>	<i>Conference Proceedings</i>	Proceedings from conferences, workshops, consortia,... except for those that are WoS covered serials (such as some IEEE proceedings and Proceedings of the National Academy of Sciences)	Kellner, R. and G. Jung., Proc. 20th Europ. Peptide Sym. 366-368 (1988).
	<i>Reference Books / Databases</i>	Encyclopaedia, Dictionary, Lexicon, Handbook, Manuals, Databases of genetic sequencing, protein information,... (e.g. GenBank, Swissprot, EMBL, PIR,...), but also Chemical Abstracts, Biological Abstracts. Manuals that are clearly associated to a company product are categorized as 'industry documents'	*Maniatis et al, In Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs. 1982, pp. 3-5, 24-27 and 31 *Suzuki et al., Chem. Abs., vol. 107, (1987), Abs. 87142x.
	<i>Industry / Company related documents</i>	Catalogues (e.g. Nike Footwear Catalog, Fall 1993, published Dec. 1, 1992 (pp. 10,16)); Brochures (e.g. "OCULUS-300", Product Brochure of Coreco Inc., date unknown.); IBM Technical Disclosure Bulletin; Advertisement (e.g. Passage St. Roch, Eyeglass Advertisement); Product information (e.g. Natuzzi Model 1207, Oct. 1994, at International Home Furnishings Market in High Point, North Carolina); Internal Company Project Reports	*USCU Sales Brochure 6-74/5070107. *Brochure entitled, "Danniflex CPM 500" by Danninger Medical Technology Inc. *Cross R.G. "Keyboard Overlay", IBM Technical Disclosure Bulletin, vol. 15 No. 1, Jun. 1972.
	<i>Books</i>	All books except for those categorized as Reference Books	Burger, "Medical Chemistry", 2nd Ed., pp. 72-88 (1960).
	<i>Patent related documents</i>	Patents (including the JAPIO abstract service); Legal documents (motion, declaration, letter); Duplicates (documents that represent a re-issued patent); Search Reports; License agreements	*Japanese Laid-open Patent Application No. 294848/86, dated Dec. 25, 1986. *European Search Report of European Application No. EP 94 30 6086. *License Agreement Between Dr. Albert M. Kligman and Johnson & Johnson, Jul. 18, 1984
	<i>Research / Technical Reports</i>	Patient Information Sheets; Reported Results of experiments/try outs; Technical or research reports of (public) research centres; PhD and master's thesis	*1987 and 1988 Tables of Data from Official Canadian Rapeseed CO-OP Trials. * 1982 ACM 0-89791-066-4, pp. 39-47, "The 801 Minicomputer", by G. Radin.
	<i>Newspapers / magazines</i>	Non-scientific, popular (e.g. PC Magazine, the Wall Street Journal, Dr Dobb's Journal,...)	*Grabowski, Ralph, "Z Mouse Gives CAD Designers 3-D Control," Infoworld, p. 93, Jul. 13, 1992. *Michael Segell, Sports Illustrated, 1985, 1 pg.
	<i>Unclear / other</i>	If source could not be identified in a straightforward way	*Protectoral Features * B.B. Sol 1974.

**Table 4a** – Occurrence of journal and non-journal references in USPTO and EPO: observed values  
(row percentages between brackets)

	<b>Journal</b>	<b>Non-journal</b>	<b>Total NPR's</b>
<b>USPTO</b>	2766 (55%)	2242 (45%)	5008
<b>EPO</b>	3218 (64%)	1803 (36%)	5021
<b>Total</b>	5984	4045	10029

**Table 4b** – Occurrence of journal and non-journal references in USPTO and EPO: expected values  
(row percentages between brackets)

	<b>Journal</b>	<b>Non-journal</b>	<b>Total</b>
<b>USPTO</b>	2988 (60%)	2020 (40%)	5008
<b>EPO</b>	2996 (60%)	2025 (40%)	5021
<b>Total</b>	5984	4045	10029



**Table 5a** – Occurrence of non-journal sources in USPTO and EPO: observed values  
(column percentages between brackets)

	<b>USPTO</b>	<b>EPO</b>	<b>Total</b>
Conference Proceedings	381 (17%)	612 (34%)	993
Industry related documents	560 (25%)	304 (17%)	864
Books	333 (15%)	186 (10%)	519
Reference books / Databases	234 (10%)	600 (33%)	834
Patent related documents	327 (15%)	46 (3%)	373
Research / Technical reports	138 (6%)	27 (2%)	165
Newspapers	106 (5%)	10 (0%)	116
Unclear / Other	163 (7%)	18 (1%)	181
<b>Total</b>	<b>2242 (100%)</b>	<b>1803 (100%)</b>	<b>4045</b>

**Table 5b** - Occurrence of non-journal sources in USPTO and EPO: expected values  
(column percentages between brackets)

	<b>USPTO</b>	<b>EPO</b>	<b>Total</b>
Conference Proceedings	550 (25%)	443 (25%)	993
Industry related documents	479 (21%)	385 (21%)	864
Books	288 (13%)	231 (13%)	519
Reference books / Databases	462 (21%)	372 (21%)	834
Patent related documents	207 (9%)	166 (9%)	373
Research / Technical reports	91 (4%)	74 (4%)	165
Newspapers	64 (3%)	52 (3%)	116
Unclear / Other	100 (4%)	81 (4%)	181
<b>Total</b>	<b>2242 (100%)</b>	<b>1803 (100%)</b>	<b>4045</b>

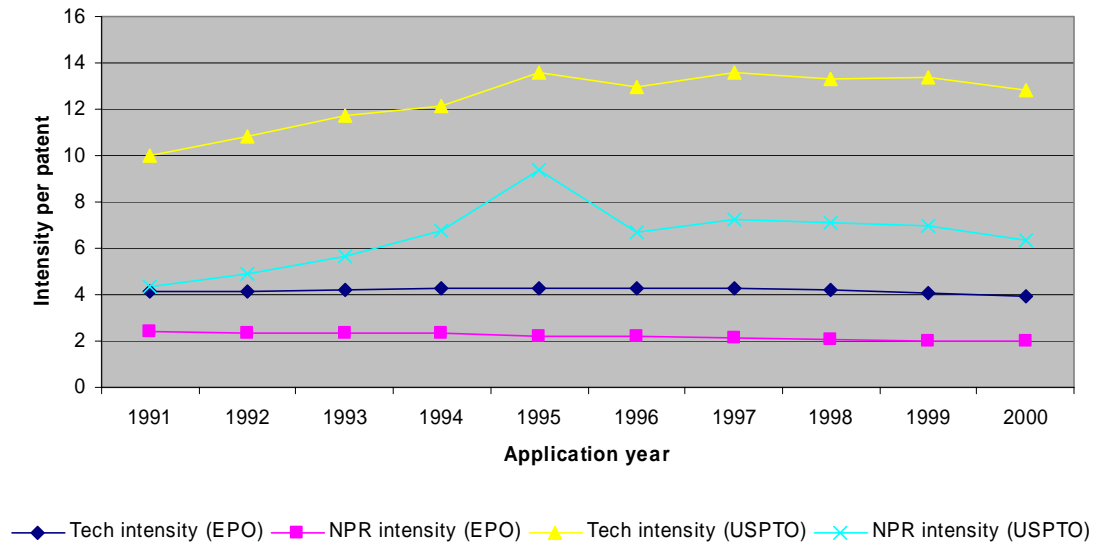
**Table 6** - Illustration of the interaction pattern and intensity between technology and 10 main science fields

(Illustration specifications: identified scientific journal publications, based on EPO-data)

Science fields	AERTH & SPACE SCIENCES	AGR. BIOL. & ENVIRONMENTAL SCIENCES	CHEMISTRY	CLINICAL MEDICINE	ENG. TECH. & APPLIED SCIENCES	LIFE SCIENCES	MATHEMATICS	MULDISCIPLINARY SCIENCES	OTHERS / SSCI / AHCI	PHYSICS
<b>Technology fields</b>										
Biotechnology	0.12%	4.37%	2.03%	21.25%	0.16%	52.78%	0.00%	19.18%	0.05%	0.07%
Organic fine chemistry	0.02%	2.82%	10.75%	28.78%	0.24%	40.62%	0.02%	16.54%	0.09%	0.11%
Pharmaceuticals, cosmetics	0.04%	2.47%	5.70%	42.36%	0.17%	34.01%	0.03%	14.92%	0.20%	0.11%
Analysis, measurement, technology	0.40%	1.52%	6.27%	29.05%	6.59%	32.36%	0.04%	13.34%	0.08%	10.35%
Chemical and petrol industry, basic materials chemistry	0.20%	8.06%	10.61%	32.02%	1.08%	33.30%	0.00%	13.56%	0.10%	1.08%
Telecommunications	0.57%	1.95%	1.03%	2.52%	76.98%	0.69%	0.23%	0.34%	0.00%	15.69%
Agriculture, food chemistry	0.14%	33.43%	2.08%	4.16%	0.83%	48.13%	0.00%	11.23%	0.00%	0.00%
Information technology	0.12%	4.37%	2.03%	21.25%	0.16%	52.78%	0.00%	19.18%	0.05%	0.07%
Macromolecular chemistry, polymers	0.02%	2.82%	10.75%	28.78%	0.24%	40.62%	0.02%	16.54%	0.09%	0.11%
Optics	0.04%	2.47%	5.70%	42.36%	0.17%	34.01%	0.03%	14.92%	0.20%	0.11%
Biotechnology	0.40%	1.52%	6.27%	29.05%	6.59%	32.36%	0.04%	13.34%	0.08%	10.35%
Organic fine chemistry	0.20%	8.06%	10.61%	32.02%	1.08%	33.30%	0.00%	13.56%	0.10%	1.08%
Pharmaceuticals, cosmetics	0.57%	1.95%	1.03%	2.52%	76.98%	0.69%	0.23%	0.34%	0.00%	15.69%
Information technology	1.24%	1.03%	1.24%	6.82%	71.07%	5.37%	0.21%	1.65%	0.00%	11.36%
Macromolecular chemistry, polymers	0.24%	4.29%	42.62%	13.33%	3.10%	25.95%	0.00%	9.29%	0.24%	0.95%
.....										

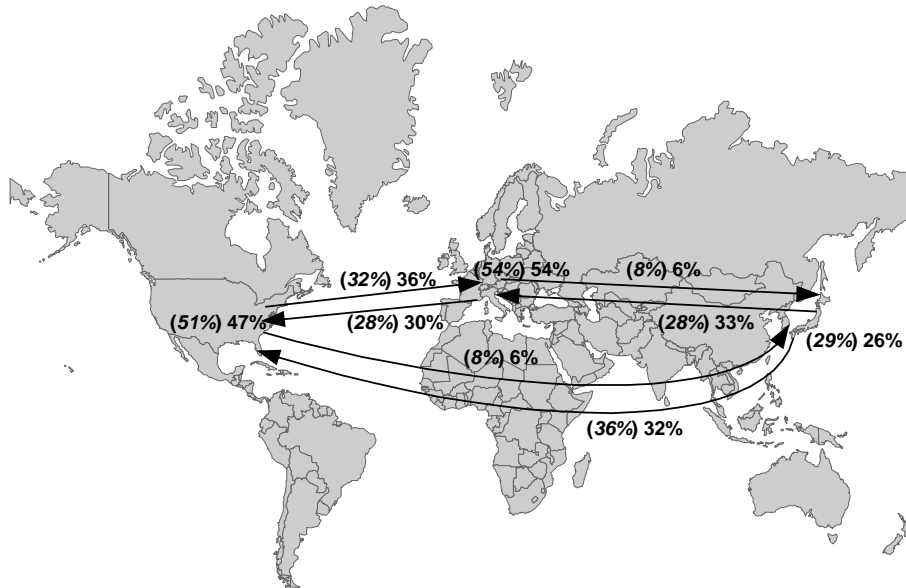
Source: Adapted from Verbeek et al., 2002

**Figure 1** - Evolution of NPR- and technology-intensity over time  
 (between 1991 and 2000<sup>4</sup>; EPO and USPTO patent data)



<sup>4</sup> Application year 2001 was eliminated from the figure, because frequency of patents in this year was too low for the intensities to be meaningful. Such low frequency is due to the time lag between application date and grant date: many patents that were applied for in 2001 had not been granted yet at the time of data extraction, hence did not yet appear in the database.

**Figure 2** – Geographic distribution of citation flows to the published literature present in USPTO patents  
 (Based on USPTO patent data; 1987-1991 and 1992-1996)



Annotations Figure 2:

The indicator developed depicts the proportion of scientific references to be found in patents, divided by region of origin. For example: “EU → US = 30%” implies that 30% of the scientific publications cited in EU-patents, in a specific technology field, originate from research conducted in the US. The proportions of local references are also depicted in the respective regions. E.g. 47% of scientific references in US patents are US-affiliated scientific references. In those instances where papers were co-authored by authors belonging to different geographical entities, the papers were assigned to each entity (integer count).

- Percentages between brackets ( ) represent the findings based on patents that have been applied for during the period 1987-1991; otherwise the period 1992-1996 is referred to.
- Arrows point towards the location of the institutional affiliation of the author mentioned on the cited publications.

## References

- Acosta, M., and Coronado, D. (2003). "Science-technology flows in Spanish regions: An analysis of scientific citations in patents", *Research Policy*, 32 (10), 1783-1803.
- Acs, Z.J., Anselin, L., and Varga, A. (2002). "Patents and innovation counts as measures of regional production of new knowledge", *Research Policy*, 31, 1069-1085.
- Akers, N.J. (2000). "The referencing of prior art documents in European patents and applications", *World Patent Information*, 22, 309-315.
- Balconi, M., Breschi, S., and Lissoni, F. (2004). "Networks of inventors and the role of academia: An exploration of Italian patent data", *Research Policy*, 33 (1), 127-145.
- Callon, M., Law J. and Rip A. (1986). *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. Sheridan House.
- David, P.A., Foray, D., and Steinmueller, W.E. (1997). *The research network and the new economics of science: From metaphors to organizational behaviour*. In: Gambardella, A., Malerba, F. (Eds.), *The Organisation of Innovative Activities in Europe*. Cambridge University Press.
- De Solla Price D. (1963). *Little Science, Big Science*. Columbia University Press.
- De Solla Price, D. (1965). "Is technology historically independent of Science? A study in statistical historiography", *Journal of Technology and Culture*, 6, 553-568.
- Dosi, G. (2000). *Innovation, Organization and Economic Dynamics*. Edward Elgar Publishers, Cheltenham, UK.
- Etzkowitz, H. and Leydesdorff, L. (1997). "Introduction to special issue on science policy dimensions of the Triple Helix of university-industry-government relations", *Science and Public Policy*, 24 (1), 2-5.
- Etzkowitz, H. and Leydesdorff, L. (1998). "The Endless Transition: A 'Triple Helix' of University-Industry-Government Relations", *Minerva*, 36, 203-208.
- Fleming, L., and Sorenson, O. (2001). "Technology as a complex adaptive system: Evidence from patent data." *Research Policy*, 30, 1019-1039.
- Freeman, C. (1987). *Technology Policy and Economic Performance*. Pinter, London.
- Freeman, C. (1994). "The economics of technical change", *Cambridge Journal of Economics*, 18, 463-514.
- Furman, J.L., Porter, E.P., and Stern, S. (2002). "The determinants of national innovative capacity", *Research Policy*, 31, 899-933.
- Griliches, Z. (1990). "Patent statistics as economic indicators: A survey." *Journal of Economic Literature*, 28, 1661-1707.
- Grupp, H., and Schmoch, U. (1999). "Patent statistics in the age of globalisation: New legal procedures, new analytical methods, new economic interpretation?" *Research Policy*, 28, 377-396.
- Harhoff, D., Scherer, F.M., and Vopel, K. (2003). "Citations, family size, opposition and the value of patent rights", *Research Policy*, 32 (8), 1343-1363.
- Hicks, D. and Narin, F. (2004). "Strategic Research Alliances and 360 Degree Bibliometric Indicators", CHI Research Inc., *Strategic research partnerships: proceedings from an NSF workshop*.
- Jaffe A.B., Trajtenberg, M, and Fogarty, M.S. (2000). "Knowledge spillovers and patent citations: Evidence from a survey of inventors", *American Economic Review*, 90 (2), 215-218.
- Jaffe A.B., Trajtenberg, M, and Henderson, R. (Eds.) (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. MIT Press.
- Katz, J.S., and Hicks, D. (1998). "Indicators for systems of innovation. A bibliometric-based approach", IDEA Paper Series, n° 12, STEP Group.
- Leydesdorff, L. and Etzkowitz, H. (1996). "Emergence of a Triple Helix of University-Industry-Government Relations", *Science and Public Policy*, 23 (5), 279-286.
- Leydesdorff, L. and Etzkowitz, H. (1998). "Triple Helix of innovation: introduction", *Science and Public Policy*, 25 (6), 358-364.
- Lundvall, B.-Å., (1992). *National Systems of Innovation*. Pinter, London.
- McMillan S., Narin F., and Deeds D. (2000). "An analysis of the critical role of public science in innovation: The case of biotechnology", *Research Policy*, 29, 1-8.
- Meyer, M. (2000a). "Does science push technology? Patents citing scientific literature," *Research Policy*, 29, 409-434.
- Meyer, M. (2000b). "Patent citations in a novel field of technology – What can they tell about interactions between emerging communities of science and technology?" *Scientometrics*, Vol. 48 (2), pp. 151-178.
- Meyer, M (2001). "Patent citation analysis in a novel field of technology: An exploration of nano-science and nano-technology", *Scientometrics*, 51 (1), 163–183.
- Meyer, M., Sinilainen, T., and Utecht, J.T. (2003). "Towards hybrid Triple Helix indicators: A study of university-related patents and a survey of academic inventors," *Scientometrics*, 58 (2), 321-350.
- Michel, J., and Bettels, B. (2001). "Patent citation analysis: A closer look at the basic input data from patent search reports", *Scientometrics*, 51, 185-201.
- Mowery, D.C., and Nelson, R.R. (1999). *Sources of Industrial Leadership*. Cambridge University Press, Cambridge.
- Murray, F. (2004). "The role of academic inventors in entrepreneurial firms: Sharing the laboratory life", *Research Policy*, 33 (4), 643-659.
- Narin F., and Noma E. (1985). "Is technology becoming science?" *Scientometrics*, 7, 369-381.

- Narin F., Hamilton K. and Olivastro D. (1997). "The increasing linkage between US technology and public science", *Research Policy*, 26 (3), 317-330.
- Nelson, R.R. (1993). *National Innovation Systems*. Oxford University Press, New York.
- Nelson, R.R., and Rosenberg, N. (1993). Technical innovation and national systems. In: Nelson, R.R. (Ed.), *National Innovation Systems*. Oxford University Press, Oxford, pp. 3-21.
- Noyons, E.C.M., Van Raan, A.F.J., Grupp, H., and Schmoch, U. (1994). "Exploring the science and technology interface: Inventor-author relations in laser medicine research", *Research Policy*, 23, 443-457.
- Packer, A., and Webster, K. (1996). "Patenting culture in Science: Reinventing the scientific wheel of credibility", *Science, Technology and Human Values*, 21 (4), 427-453.
- Pavitt, K. (1997). Academic research, technical change and government policy. In: Krige, J., Pestre, D. (Eds.), *Science in the Twentieth Century*. Harwood Academic Publishers, Amsterdam, pp. 143-158.
- Pritchard, A. (1969). "Statistical bibliography or bibliometrics?" *Journal of Documentation*, 24, 348-349.
- Reitzig, M. (2004). "Improving patent valuations for management purposes – validating new indicators by analyzing application rationales." *Research Policy*, 33, 939-957.
- Schmoch U. (1993). "Tracing the knowledge transfer from science to technology as reflected in patent indicators", *Scientometrics*, 26 (1), 193-211.
- Schmoch, U. (1997). "Indicators and the relations between science and technology", *Scientometrics*, 38 (1), 103-116.
- Schmookler, J. (1966). *Invention and Economic Growth*. Harvard University Press, Cambridge.
- Schubert, A., Glänzel, W., and Braun, T. (1989). "World flash on basic research: Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields, 1981-1985", *Scientometrics*, 16 (1-6), 3-478.
- Schwander, P. (2000). "An evaluation of patent searching resources: Comparing the professional and free online databases", *World Patent Information*, 22, 147-165.
- Steinmueller, W.E. (1994). Basic research and industrial innovation. In: Dodgson, M., Rothwell, R. (Eds.), *The Handbook of Industrial Innovation*. Edward Elgar, Aldershot, pp. 54-66.
- Tijssen, R.J.W. (2001). "Global and domestic utilization of industrial relevant science: Patent citation analysis of science-technology interactions and knowledge flows", *Research Policy*, 30, 35-54.
- Tijssen, R.J.W. (2002). "Science dependence of technologies: Evidence from inventors and their inventions", *Research Policy*, 31 (4), 509-526.
- Tijssen, R.J.W., Buter, R.K., and Van Leeuwen, T.N. (2000). "Technological relevance of science: Validation and analysis of citation linkages between patents and research papers", *Scientometrics*, 47, 389-412.
- Trippe, A.J. (2003). "Patinformatics: Tasks to tools", *World Patent Information*, 25, 211-221.
- Van Looy, B., Callaert, J., Debackere, K., and Verbeek, A. (2003a). "Patent-related indicators for assessing knowledge-generating institutions: Towards a contextualised approach", *Journal of Technology Transfer*, 28, 53-61.
- Van Looy, B., Zimmermann, E., Veugelers, R., Verbeek, A., Mello, J., and Debackere, K. (2003b). "Do science-technology interactions pay off when developing technology?" *Scientometrics*, 57 (3), 355-367.
- Van Raan, A.F.J. (Ed.) (1988). *Handbook of Quantitative Studies of Science and Technology*. North-Holland, Amsterdam.
- Van Vianen, B., Moed H. and Van Raan A. (1990). "An exploration of the science base of recent technology", *Research Policy*, 19, 61-81.
- Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., and Deleus, F. (2002). "Linking science to technology: Using bibliographic references in patents to build linkage schemes", *Scientometrics*, 54 (3), 399-420.