Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 69, No. 1 (2006) 3–20

Traces of Prior Art: An analysis of non-patent references found in patent documents

JULIE CALLAERT,[a] BART VAN LOOY,[a,b] ARNOLD VERBEEK[b]
KOENRAAD DEBACKERE,[a,b] BART THIJS[b]

[a] *K.U.Leuven, Faculteit ETEW, Research Division INCENTIM, Leuven (Belgium)*
[b] *K.U.Leuven, Steunpunt O&O Statistieken, Leuven (Belgium)*

The recent developments towards more systemic conceptualizations of innovation dynamics and related policies highlight the need for indicators that mirror the dynamics involved. In this contribution, we assess the role that 'non-patent references', found in patent documents, can play in this respect. After examining the occurrence of these references in the USPTO and EPO patent systems, their precise nature is delineated by means of a content analysis of two samples of non-patent references (n=10,000). Our findings reveal that citations in patents allow developing non-trivial and robust indicators. The majority of all non-patent references are journal references, which provide ample possibilities for large-scale analyses focusing on the extent to which technological developments are situated within the vicinity of scientific knowledge. Application areas, limitations and directions for future research are discussed.

Introduction

Over recent decades, models of innovation have become increasingly systemic. Non-linear knowledge flows and interactions between multiple actors – mostly companies, knowledge-generating institutes and governmental agencies – are central to these models. The concept of (national) innovation systems (LUNDVALL, 1992; NELSON & ROSENBERG, 1993; FREEMAN, 1994; MOWERY & NELSON, 1999; DOSI, 2000),

---

Received June 20, 2005

Address for correspondence:
JULIE CALLAERT
K.U.Leuven, Faculteit ETEW, Research Division INCENTIM
Naamsestraat 69, 3000 Leuven, Belgium
E-mail: Julie.Callaert@econ.kuleuven.be

0138–9130/US $ 20.00

the Triple Helix metaphor (ETZKOWITZ & LEYDESDORFF, 2000) and network models as put forward by STEINMULLER (1994), PAVITT (1997), DAVID et al. (1997) bear witness to this transition in assumptions. 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 PRICE (1963), PRITCHARD (1969), VAN RAAN (1988), SCHUBERT et al. (1989), amongst others. Similar developments at the technological level took place, in which patent information played a central role (PRICE, 1963, 1965; SCHMOOKLER, 1966; CALLON et al., 1986; GRILICHES, 1990; NARIN & NOMA, 1985). Bibliometric indicators pertaining to both scientific and technological activity are nowadays gathered on a regular basis, 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 weight 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 period. This published information provides different possibilities for analyzing the specific environment in which development efforts take place (TRIPPE, 2003) and, hence, offers scope for identifying science-technology exchanges. Firstly, interpersonal inventor-author linkages can be assessed. By analyzing the extent to which patent inventors are also registered as authors of scientific publications, boundary crossing between the science and technology spheres can be monitored at the individual level of analysis. Examples of this approach are found in the work of NOYONS et al. (1994), PACKER & WEBSTER (1996), BALCONI et al. (2004) and MURRAY (2004). A second approach consists of looking at organizational entities that appear both as assignee in patents and as affiliate in publications, shifting the level of analysis towards the organizational level. Thirdly, analogous to co-publication analysis, co-patenting behaviour – involving companies and research organizations – can be mapped, reflecting a degree of cooperation between representatives of the technological and scientific activity spheres. Note here that relatively low levels of such boundary-crossing co-patenting efforts may limit wide applicability (HICKS & NARIN, 2004). Another way to use patent documents to reveal science-technology associations involves an examination of the nature of the so-called 'prior art' found in published patent documents – the sources that are provided by the patent examiner and/or the inventor, and that are considered relevant for assessing the invention and the claims it entails. In so far as these sources are scientific, they provide the opportunity to systematically examine relationships between science and technological development. This prior art information found in patent documents and its potential usefulness in revealing science-technology relationships are the focus of this contribution.

A closer look at prior art and non-patent references found in patent documents

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 its novelty, inventive activity and industrial applicability. During the granting process, patent examiners review the prior art that pertains to the invention. Based on information archives and databases, they decide which references are relevant in assessing the patent and its constituent claims. In this process, examiners do not limit themselves to the prior art signalled by inventors and/or applicants. The front pages of patent documents include *examiner-given* references and these do not necessarily coincide with references provided by the inventor: the latter may be omitted by the examiner and/or the examiner may add references that were not mentioned by the inventor. It can be noted here that the specific role of references in patent applications differs to some extent 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, following the argument on cumulativeness in knowledge production (FORAY, 2004), it can be argued that the cited references have influenced the genesis of the insights developed in the citing article. This is not necessarily the case for the front-page references in patent documents.

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. A majority of previous research has focused on the role of *patent* references and citations, used as an indication of patent value (OPPENHEIM, 2000; FLEMING & SORENSON, 2001; JAFFE et al., 2002; HARHOFF et al., 2003; REITZIG, 2004). The role and possible contribution of *non-patent* references was pioneered by Narin and his colleagues. Studies in this field have investigated the nature of science-technology relationships as implied by citation links (e.g. NARIN & 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., 2002a; ACOSTA & 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).

Proponents sometimes portray scientific references in patents as signalling a direct influence of science on technology (e.g. NARIN et al., 1997), while others advocate a more modest interpretation. MEYER (2000a, 2000b, 2001), having performed 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, 2002) and TIJSSEN et al. (2000), having surveyed inventors on scientific contributions to their patents, point 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). Similarly, several authors point to contextual elements that should be taken into account when interpreting such indicators. MICHEL & BETTELS (2001) argue that the comprehensiveness and the quality of citation lists appearing in patent documents vary significantly as a function of the patent office. Differences between the USPTO and EPO examination procedures may influence the number and type of references cited, as we illustrate in the empirical section. HARHOFF et al. (2003) and VAN LOOY et al. (2003a,b) point to field specific effects that need to be taken into account when using and interpreting patent related indicators.

Hence, one should be careful in depicting citations in patents as interactions or direct links of causation between two pieces 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 greater relevance or relatedness between the technology at hand and scientific activity.[*] 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 following section, we will provide a systematic view on the information that is observable, i.e., the amount and nature of non-patent references in patents. This can help put the current debates in perspective, and it allows assessing the feasibility and precise meaning of indicators based on non-patent references. More specifically, we want to contribute to 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 10,000 non-patent references retrieved from EPO and USPTO patent files.

[*] Note that Science-Technology relatedness as described here also has a counterpart, namely references towards patents found within scientific publications.

Comparison of occurrence of patent references and non-patent references

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.* 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 years 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 non-patent references observed in this period.

Table 1. Occurrence of patent and non-patent references (USPTO – EPO)

USPTO granted patents with application years between 1991 and 2001					
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 *With (1) as denominator:*	12.55 *11.33*
# patents containing non-patent references	445,466 (34%)	# non-patent references	3,018,943 (17%)	NPR intensity *With (1) as denominator:*	6.77 *2.2*
EPO granted patents with application years between 1991 and 2001					
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 *With (1) as denominator:*	4.20 *4.09*
# patents containing non-patent references	130,511 (38%)	# non-patent references	293,977 (17%)	NPR intensity *With (1) as denominator:*	2.25 *0.86*

Table 1 shows that the majority of patents contain patent references (90% and 98% for USPTO and EPO respectively). This is not the case for non-patent references: the proportion of patents containing such references amounts to 34% and 38% for USPTO and EPO. Moreover, patent references are more numerous resulting in a share of 83% compared to non-patent references. As a consequence, for both USPTO and EPO patents, the average number of patent references per patent ('technology intensity') doubles the average amount of non-patent references per patent ('NPR intensity'**).

* Patent references differ not only from 'other references' in terms of the nature of the cited documents (patents versus all other types of 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).

** The term 'science' intensity is mostly used for the average number of non-patent references per patent. This suggests that all other references would be references to the scientific literature. As we will see later on, this is not completely accurate. Therefore, at this stage we prefer to use the term 'NPR intensity'.

This observation may – at least partly – be related to alleged search impediments in the examiner procedures. As the president of the International Intellectual Property Institute pointed out (LEHMAN, 2001), the USPTO and other patent offices lack comprehensive and easily accessible databases of non-patent prior art. Effective examination today requires comparing claimed inventions with information disclosed in numerous journals and other publications, to which examiners have limited access and for which they lack effective search tools. In practice, the USPTO purchases access to electronic databases that contain many publications in electronic form (e.g. Derwent, Nexis, Dialogue). However, none of these are searchable across the entire database, and they may be searched using only the key words familiar to a given examiner. The fact that there is such a difference in terms of complexity and scope between patent database searches and those of non-patent literature (see also: SAMPAT, 2004) undoubtedly explains part of the relatively lower occurrence of non-patent references.

Figure 1 shows the development in NPR and technology intensity over time. For the EPO data, the development in NPR and technology intensity over the decade under review seems to have remained relatively flat. The figure shows a minor but steady decline of NPR intensity over time. 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 and at first glance, the 'closeness' to the technological and scientific spheres has remained rather constant over time (decade under review).

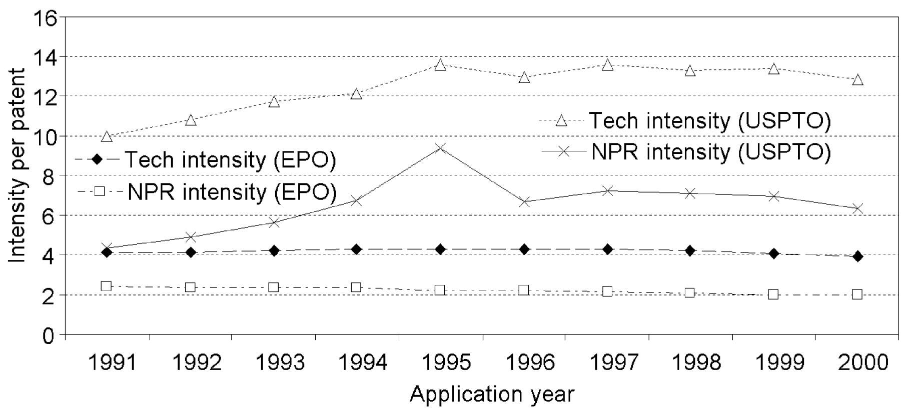

Figure 1. Evolution of NPR and technology intensity over time
(between 1991 and 2000,* EPO and USPTO patent data)

*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 at the time of data extraction and hence did not appear in the database

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% up to 1999, after which a decline is 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. When interpreting this figure, it should be noted that patents with pending periods extending three years (applied for before 2000, granted after 2003) are not included due to the 2003 extraction date. For the USPTO, average time between application and grant is about 35 months, but can extend to 44 months. For the EPO, this period is about 4 years (LEHMAN, 2001). To the extent that longer pending periods are associated with more complex patents or patents in newer fields, this may somewhat alter the picture that emerges, at least from 1998 onwards.

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.

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

	EPO patents		USPTO patents	
Technology field	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

For the EPO data, one observes the highest intensity of patent references per patent for Mechanical Engineering and Machinery (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 Pharmaceuticals show the highest number of NPRs (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., 2002a).

Whereas the proportions of the different reference types in the USPTO and EPO systems are comparable (see Table 1), it is evident that, in absolute terms, USPTO patents hold approximately 3 times more references on average than EPO patents. Such an observation could be directly related to the differences between the American and the European system in the rationale of citing prior art. In the USPTO system, the 'duty of candour' principle postulates that all prior art documents (including patents and other written 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 failure to disclose relevant prior art (AKERS, 2000). Such different disclosure obligations can be considered an important reason for the higher number of references in USPTO compared to EPO.

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, a further assessment of the nature of these other references seems appropriate. To the extent that these other references are of a scientific nature, developing indicators that depict science–technology proximity becomes relevant. 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.

A closer look at the nature of non-patent references

Systematic overviews of the nature of other references in patents are scarce, although some efforts in this direction have been made in the past. NARIN & NOMA (1985) reported – for the period 1978-1980 – on average 0.3 other references per patent, which is considerably lower than our observations. 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. Moreover, HARHOFF et al. (2003) briefly illustrated the fact that not all non-patent references refer to scientific sources. They evaluated 100 patent document records and found approximately 60% of non-patent references referring to scientific and technical journals. The remainder was largely made up 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 obtain a fresh insight into the nature of the other references, we extracted two samples of non-patent references from the USPTO and the EPO databases. In each database, 5,000 non-patent references were randomly drawn from granted patents with application years between 1996 and 2001. For USPTO patents, front-page references were extracted. For EPO, the REFI (Reference File) database was used. To ensure a representative sample, the group of patents from which references were drawn was stratified according to the overall distribution of patents over technology domains (International Patent Classification system, 3-digit level). Both samples of non-patent references (EPO and USPTO) were classified; each citation according to the document type referred to. The taxonomy of reference types used – as presented in Table 3 – was based on previous categorizations (NARIN & NOMA, 1985; VAN VIANEN, 1990; HARHOFF et al., 2003) and extended while conducting the content analysis of the extracted sample of references. Most of the non-patent references could be categorized in this scheme. Only for a limited number of references, did incomplete information preclude a precise categorization. These are referred to as ‘other’.

In the group of non-patent references, an initial distinction can be made between journal and non-journal references. In a narrow sense, only journal references refer to the actual scientific journal literature. Scientific journal references were most easily recognized when the journal was SCI-covered: a match of the journal title with an existing list of SCI-covered journals allowed for straightforward identification of these references. For other serial references, a case-by-case evaluation was needed to establish whether they were references to scientific literature or to a publication (e.g. newspaper or magazine) with a non-scientific orientation. The latter are classified as ‘non-journal references, newspapers/magazines’. This assessment was based on the cycle of appearance (e.g. weekly: more likely to point to non-journal category), references and descriptions in academic databases (e.g. EBSCO host, Academic Search Premier…), as well as a content analysis of a limited sample of issues.

Table 3. Taxonomy of reference types

Category	Sub-category	Description	Illustration
Serial journal references:	*SCI-covered*	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 of Lovastatin".
	Not SCI-covered	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.
Non-journal references:	*Conference Proceedings*	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).
	Reference Books/ Databases	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 with 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.
	Industry/ Company - related documents	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.
	Books	All books except those categorized as Reference Books	Burger, "Medical Chemistry", 2nd Ed., pp. 72-88 (1960).

Table 3 (continued)

Category	Sub-category	Description	Illustration
	Patent-related documents	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
	Research/ Technical Reports	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.
	Newspapers/ magazines	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.
	Unclear/ other	If source could not be identified in a straightforward way	*Protectoral Features * B.B. Sol 1974.

Table 4a shows the relative occurrence of journal and non-journal references in the USPTO and EPO samples of non-patent references. For both USTO and EPO references, more than half are journal references. An additional check reveals that the SCI database provides almost full coverage of these journal references, holding respectively 90% and 86% of journal references in our USPTO and EPO sample. A comparison of the USPTO and EPO sample using a chi square test reveals a significant difference between the two ($p<0.01$). The expected values are provided in Table 4b.

As is apparent, USPTO references contain less journal and more non-journal references than expected. For EPO references, the opposite holds. Therefore, in general, more than half of all non-patent references are journal articles; a somewhat smaller proportion 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 proportion, are not inconsiderable, representing 36% and 45% for EPO and USPTO respectively.

Table 4a. Occurrence of Journal and Non-journal References in USPTO and EPO: observed values
(row percentages between brackets)

	Journal	Non-journal	Total NPR's
USPTO	2,766 (55%)	2,242 (45%)	5,008
EPO	3,218 (64%)	1,803 (36%)	5,021
Total	5,984	4,045	10,029

Table 4b. Occurrence of Journal and Non-journal References in USPTO and EPO: expected values
(row percentages between brackets)

	Journal	Non-journal	Total
USPTO	2,988 (60%)	2,020 (40%)	5,008
EPO	2,996 (60%)	2,025 (40%)	5,021
Total	5,984	4,045	10,029

Therefore, as a further step, we 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 (see Table 3).

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

	USPTO	EPO	Total
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
Total	2,242 (100%)	1,803 (100%)	4,045

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

	USPTO	EPO	Total
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
Total	2,242 (100%)	1,803 (100%)	4,045

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 scientific. These include conference proceedings, which often precede or, in some cases, even equal journal articles. At the same time, the majority of reference books/databases (often referring to abstract services) and books can be considered scientific information sources. Likewise, some types of non-patent references can be considered technology-related. 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 and, therefore, represent 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 apart from journal references. 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 in 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 small portions of the non-journal references.

These observations demonstrate that at least 42% of all USPTO non-journal references refer to scientific knowledge. In addition to the actual patent citations (see Table 1), a further 40% of non-journal references relates to technological information. In 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 the major part of non-patent references is scientifically oriented for both USPTO and EPO, this phenomenon is again more pronounced 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, while the opposite holds for patent related documents.

Such differences can be traced back to 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 above-mentioned duty of candour, which is specific to the US system and which is the source of many patent-related legal disputes.

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 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. Additional efforts to define clear-cut criteria for recognizing and assessing scientific content of references, other than journal references, will be useful in the development of such indicators.

The most straightforward use of an indicator based on non-patent references consists of counting backward citations to scientific articles as a proxy for 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 national level, 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 subsequent 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., 2002a) as Table 6 illustrates.

When constructing such linkage matrices, the introduction of a time dimension can uncover the development of science intensity, allowing an assessment of the presence and nature of Science-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 TIJSSEN, 2001; 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 patterns.

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 / Technology fields	EARTH & SPACE SCIENCESS	AGR. BIOL. & ENVIRONMENTAL SCIENCES	CHEMISTRY	CLINICAL MEDICINE	ENG.TECH. & APPLIED SCIENCES	LIFE SCIENCES	MATHEMATICS	MULIDISCIPLINARY SCIENCES	OTHERS / SSCI / AHCI	PHYSICS
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., 2002b

While possibilities for analyzing the obtained indicators are numerous, some limitations should be taken into account when interpreting them. 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 allow for classification of the information obtained in a systematic and consistent way (e.g. scientific discipline, affiliation). The most widely adopted publication database for such purposes is the ISI Web of Science database, covering a large proportion 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 Web of Science database – are far less captured in such encompassing databases and are, therefore, much harder to include in large-scale quantitative analyses. Our analysis revealed that 50% to 55% of non-patent references are journal references covered by the Web of Science. Secondly, our findings reveal that the amount of scientific references found in patents differs among technological fields. This implies that the possibilities and 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 way – and a partial one, at that – to identify science-technology interactions. Therefore, part of the picture – e.g. informal, non-traceable flows of tacit knowledge – remains hidden. 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 systems were pointed out above. They can be attributed largely 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 for the development of 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.

References

ACOSTA, M., 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., 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., 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., RIP A. (1986), *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. Sheridan House.

DAVID, P.A., FORAY, D., 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.

DOSI, G. (2000), *Innovation, Organization and Economic Dynamics*. Edward Elgar Publishers, Cheltenham, UK.

ETZKOWITZ, H., LEYDESDORFF, L. (2000), The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations, *Research Policy*, 29 (2) : 109–123.

FLEMING, L., SORENSON, O. (2001), Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, 30 : 1019–1039.

FORAY, D. (2004), *The Economics of Knowledge*. The MIT Press.

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., 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., 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., VOPEL, K. (2003), Citations, family size, opposition and the value of patent rights, *Research Policy*, 32 (8) : 1343–1363.

HICKS, D., 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., 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., HENDERSON, R. (Eds) (2002), *Patents, Citations, and Innovations: A Window on the Knowledge Economy.* MIT Press.

KATZ, J. S., HICKS, D. (1998), *Indicators for Systems of Innovation. A Bibliometric-Based Approach*, IDEA Paper Series, n° 12, STEP Group.

LEHMAN, B. A. (2001), *Presentation CIBC World Markets,* New York, May 3, 2001.

LUNDVALL, B.-Å. (1992), *National Systems of Innovation.* Pinter, London.

MCMILLAN, S., NARIN, F., 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*, 48 (2) : 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., 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., 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., 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., NOMA, E. (1985), Is technology becoming science? *Scientometrics*, 7 : 369–381.

NARIN, F., HAMILTON, K., OLIVASTRO, D. (1997), The increasing linkage between US technology and public science, *Research Policy*, 26 (3) : 317–330.

NELSON, R. R. (Ed.) (1993), *National Innovation Systems*. Oxford University Press, New York.

NELSON, R. R., 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., SCHMOCH, U. (1994), Exploring the science and technology interface: Inventor-author relations in laser medicine research, *Research Policy*, 23 : 443–457.

OPPENHEIM, C. (2000), Do patent citations count? In: CRONIN, B., ATKINS, H. B. (Eds), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Metford, NJ: Information Today Inc. ASIST Monograph Series, pp. 405–432.

PACKER, A., 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.

PRICE, D. DE SOLLA (1963), *Little Science, Big Science*. Columbia University Press.

PRICE, D. DE SOLLA (1965), Is technology historically independent of Science? A study in statistical historiography, *Journal of Technology and Culture*, 6 : 553–568.

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.

SAMPAT, B. (2004), *Examining Patent Examination: An Analysis of Examiner and Applicant Generated Prior Art,* Working Paper, School of Public Policy, Georgia Institute of Technology.

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., 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., 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., 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., 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., 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., DELEUS, F. (2002a), Linking science to technology: Using bibliographic references in patents to build linkage schemes, *Scientometrics*, 54 (3) : 399–420.

VERBEEK, A., CALLAERT, J., ANDRIES, P., DEBACKERE, K., LUWEL, M., VEUGELERS, R. (2002b), *Science and Technology Interplay – A Modelling Approach on a Regional Level*, Final Report to the EC DG Research, Brussels.

VERBEEK, A., DEBACKERE, K., LUWEL, M. (2003), Science cited in patents: A geographic 'flow' analysis of bibliographic citation patterns in patents, Scientometrics, 58 (2) : 241–263.