KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0123

LINKING SCIENCE TO TECHNOLOGY: USING BIBLIOGRAPHIC REFERENCES IN PATENTS TO BUILD LINKAGE SCHEMES

by
A. VERBEEK
K. DEBACKERE
M. LUWEL
P. ANDRIES
E. ZIMMERMANN
F. DELEUS

D/2001/2376/23

Linking Science to Technology: Using Bibliographic References in Patents to Build Linkage Schemes

ARNOLD VERBEEK[1], KOENRAAD DEBACKERE[1], MARC LUWEL[2], PETRA ANDRIES[1], EDWIN ZIMMERMANN[1], FILIP DELEUS[3,4]

1. Catholic University of Leuven, R&D division INCENTIM, Groot Begijnhof 59, B-3000 Leuven, Tel.: +32 16 326897, Fax: +32 16 326732, E-mail: arnold.verbeek@econ.kuleuven.ac.be
2. Ministry of the Flemish Community, Private Office of the Flemish Minister of Education and Training, Albert II-laan 15, B-1210 Brussels, E-mail: marc.luwel@vlaanderen.be
3. Catholic University of Leuven, Medical School, Department of Neuro- and Psychophysiology, Herenstraat 49, B-3000 Leuven, E-mail: filip.deleus@med.kuleuven.ac.be
4. Synes NV, Next Generation Data Mining, Technologielaan 11, B-3001 Heverlee, Belgium, E-mail: filip.deleus@synes.com

0. Abstract

In this paper, we develop and discuss a method to design a linkage scheme that links the systems of science and technology through the use of patent citation data. After conceptually embedding the linkage scheme in the current literature on science-technology interactions and associations, the methodology and algorithms used to develop the linkage scheme are discussed in detail. The method is subsequently tested on and applied to subsets of USPTO patents. The results point to highly skewed citation distributions, enabling us to discern between those fields of technology that are highly science-interactive and those fields where technology development is highly independent from the scientific literature base.

1. Introduction

The importance of the interaction between science and technology (S&T) for economic growth and progress is beyond any doubt (see for example Dosi and Fabiani 1994; Silverberg and Soete 1994; Nelson 1994). Technical progress and change are fundamental issues in economics (Grupp, 1998). A number of schools have developed and adopted complementary approaches for understanding the dynamics of technological progress and economic development. The interaction between those spheres is far from linear and straightforward; it is dynamic, heterogeneous, and increasingly complex. The systems of science and technology are assumed to be converging, an evolution presented and discussed already in 1963 by Toynbee who compared the S&T interaction with a 'pair of dancers'. The emergence of 'sciento-technologies', technologies increasingly depending on scientific discovery and progress is gaining importance, especially in policy oriented (research) circles.

Processes of 'knowledge creation' and the different possible modes of 'knowledge diffusion' are central themes, and even pillars, in the ongoing debate on science, technology, and innovation and their interaction (Gibbons et. al. 1994). The nature of knowledge itself is evolving to a more network-oriented structure, with greater emphasis on strategic alliances, knowledge demand and supply chains, a growing transdisciplinarity and heterogeneity. The social imbeddedness of knowledge creation and diffusion is becoming increasingly manifest. The linear model of knowledge transfer and diffusion does no longer represent the current complexity, despite the fact that some technological fields, such as biotechnology, are heavily based on scientific discovery in order to shift the boundaries of technological application. The S&T interaction triggered the theoretical development and empirical testing of knowledge production functions (Grilliches, 1990).

However, before being able to model an entire knowledge production function, a more detailed understanding of the interaction between science and technology is needed. A central issue hereby is the quantification and modelling of the complex web of linkages and interactions between S&T development, an issue that will be dealt with in this paper. Are there any implications in regard of science and technology policy? In the course of the past two decades, wide-ranging socio-economic and technological transformations have caused European governments to reformulate their policies concerning government-supported scientific activity. This reformulation has been accompanied by shifts and even complete turnarounds in research funding between fields and between the orientation of research (basic or applied). The present constraints on public expenditures, the enormous investments involved, and the actual debate on the effectiveness of government supported scientific research, also increases the need for greater accountability and effectiveness in all areas of the public domain, and more specifically the area of publicly funded basic research (Ziman, 1994; Moed, 1989). Indeed, on a policy level, the disentangling of the S&T interrelation may lead to considerable support in handling the above-mentioned challenges.

In this paper we shall present an S&T linkage methodology based on the analysis of non-patent references present in US patent documents (applied for between 1992 and 1996). The application of this methodology reveals a strongly concentrated science and technology interaction pattern with a limited number of related technology and science fields forming the backbone of the interaction. Before presenting the methodology and the results of its application, we shall first elaborate more in detail on the science and technology interaction, non-patent references, and the nature of the established S&T linkage. As we shall see, scepticism about the nature of the captured S&T interrelation is in place, just as the valuation of the obtained results is. The next four sections will be devoted to these issues. This paper is a direct result of a project funded by DG Research of the European Commission (situated in the Fifth Framework Programme), whose support in the establishment of this paper is greatly being acknowledged.

2. Science and Technology Interaction

For a long time it was believed that there existed a continuum stretching from very basic scientific research, through applied research and technology, to economic growth and subsequent national prosperity (Narin & Olivastro, 1992). However, reality has outdated this view. Throughout the years, different ways of approaching the S&T interaction have emerged. At first, the knowledge transfer from science to technology was considered to be linear– as expressed in the 'linear model'. Later on, this view evolved to a 'network model', where the relationship between science and technology was considered more reciprocal.

The traditional understanding of the contribution of basic research to industrial innovation, as investigated in the late 1960s when retrospective studies like TRACES (Illinois Institute of Technology Research Institute, 1968) and Project Hindsight (Sherwin and Isenson, 1967) were carried out in the US, is based on the 'linear model' of knowledge production and transfer. Science is viewed as a 'social instrument' that is expected to generate economic returns as the produced knowledge is commercially developed and exploited. Use of the linear approach ignores the evidence that technological change is often built upon experience and ingenuity divorced from scientific theory or method; the role or technological developments in motivating scientific explanation; and the sources of instrumentation for scientific investigation (Rosenberg, 1982; Gibbons et. al., 1994). Effective science-technology interfaces are human in character and hinge on person embodied 'tacit' knowledge and skills. This model overlooks the influence of technology on the scientific agenda (Tijssen, 2001; Steinmueller, 1994), as demonstrated in the early years of the industrial revolution where technological breakthroughs were followed much later by scientific explanations.

The 'network model' of knowledge production, transfer and use is likely to characterise more adequately the complex interactions between knowledge producers and users. The 'network' approach opens new and useful economic perspectives, like the increasing network value with the number of participants, decreasing rate of overlapping research projects through network centralisation, and complementary investments for information dissemination that may lead to economic benefits. Information flows within the network appear to be more easily accessible by governments and firms, increasing their choices about specialisation, co-operation and competition. The network model can be associated with a view on science as a 'social institution', whose norms and practices are distinct from, and only partially reconciliable with, the institutions of market (Steinmueller, 2000). This once more illustrates the increasing complexity of the S&T interaction.

3. Measurement of science and technology in general

3.1 Patent and publications as proxies of the respective science and technology system

Understanding the S&T interaction requires a separate understanding of the scientific and technological systems. A first step in that direction is measurement. One of the major concerns of analysts in this regard is to describe S&T activities in qualitative as well as in quantitative terms so that indicators can be used in the context of models, explicit or implicit. The general recognised problem is that S&T can only be measured indirectly, using input, output or impact indicators (OECD, 1994a). Patents and publications, as representatives of technology and science, are so-called proxy measures. Is it then possible to base S&T interaction analyses on these proxies? The answer is yes, simply because these indicators are the best available at the moment, and also because of their analytical possibilities.

Patents, as a detailed source of information on inventive activity, offer an interesting monitoring device to identify main lines and trends, and even, under specific conditions, the possibility to analyse R&D processes in more detail. But what do patents exactly measure? A patent at least represents a minimal amount of invention that has passed a thorough examination by the patent office on both the novelty of the claimed item and its potential utility (Grilliches, 1990). A patent usually follows successful R&D activity thereby offering detailed information on the activity itself. Scientific publications constitute an (imperfect) output-indicator of research activity. One of the objectives of a scientific publication is to spread scientific findings within and outside the scientific community. As such, publishing in scientific journals – 'serial literature' – plays a leading role in the dissemination of research findings. Patents as well as scientific publications allow for detailed analysis of the relational structures in both the technology and science sphere. Social and cognitive networks can be discovered and analysed due to the availability of information on authors, inventors and assignees, their addresses, references and citations etc. Cross citation analysis (patents citing scientific literature) between both spheres, as reported in this paper, provides similar opportunities for studying the S&T interaction patterns, under the condition of being aware of its limitations. More details on trivialities around patents and publications are given elsewhere.

3.2 Patents and publications for analysing the S&T systems interaction

With De Solla Price (1965) and Rosenberg & Birdzell (1990) there has been quite some qualitative understanding of the S&T interaction. Yet until the '90 there has been very little quantitative data to specifically characterise this relationship or to pinpoint the subject, national, international and temporal aspects of the coupling between science and technology. Recently, the quantitative analysis of the S&T interaction has been receiving more and more attention (e.g. Schmoch et. al. 1993).

Basically we can distinguish two approaches for studying the relationship between science and technology: the 'indirect' linkage approach and the 'direct' linkage approach. The direct linkage approach refers to the possibility of studying the S&T interaction through bibliographic references present in patent documents. Specifically, non-patent references 'relate' science and technology in a direct and straightforward way. The S&T relation however is not always direct and straightforward. The absence of bibliographic references does not necessary imply a lesser science dependency of the technology involved. On the contrary, it may indicate a different type of science interaction inherent to the technological nature and stage of evolution of the field involved. A weak S&T interaction, measured by the presence of non-patent references, may be in contrast with the present academia - industry co-operation. Despite the limitations of the direct linkage approach (such as database shortcomings, skewed distribution of non-patent references, complexity of the data involved), this approach offers substantial possibilities for analysing the S&T interplay thereby acknowledging its qualitatively controlled nature (the examination procedure). In the next section we shall elaborate further on non-patent references. How do they occur and how well do they represent the science interaction?

4. Non-Patent References

Let us review briefly the citation rationale in general. Referencing, as one of the widely accepted and utilised norms, confirms and illustrates the social character of the knowledge creation and diffusion process. Citations occur not only within the scientific community but also within the technologic community (provided by the actors involved in the invention and patenting process). Moreover, there is a profound 'cross' citation practice between both communities, mainly from technology to science. Within the academic system several citation motives apply (see elsewhere for details). This is also the case with patent citations, although they are primarily legal-based. Comparison between the motives for academic-, and the motives for patent citations may provide relevant insights in the differences in citation behaviour between the technology and the science system and also the relevance of patent citations (see the work of Meyer, 2000b). Patent references are less likely to be irrelevant or superfluous than references in journal papers (Collins & Wyatt, 1988) due to the controlled nature of the patenting process. In the near future we intend to analyse these issues more in detail.

Non-patent references (NPRs) result from the so-called 'search for prior-art', i.e. the search for state-of-the-art technical and/or scientific literature; they encompass references to a variety of non-patent documents, such as scientific articles, technical papers, conference proceedings, textbooks, disclosure bulletins, abstract services, etc. We can distinguish between examiner- and inventor-given references, as the source of the reference. In some occasions, when the patent examiner includes one or more inventor-given references, they can be found on the so-called 'front-page' of a patent. Narin et. al. (1989) indicated that there is much similarity as to the specificity between both sources of references, whereas Schmoch (1993) pointed out that about 8% of all examiner-given references originate from the inventor. Inventor originating citations have, until now, not been available in machine-readable form.

The presence of NPRs indicate that the technical invention is related to – or in some cases initiated and/or stimulated by – research activities performed in related fields. The average level of references to non-patent literature is an appropriate indicator for describing the relation of a technology field to science (Schmoch, 1997). Practice however shows that it is just a minority of patents that contains references to non-patent literature. A recent study of the Norwegian knowledge base by Iversen (1998), for example, shows that not more than 30% of Norwegian-originated US patents contain NPRs (Meyer, 2000a). Collins & Wyatt (1988) however found that patents in fields that are young, developing rapidly and with a strong scientific content generally cite a substantial number of scientific publications. Meyer-Krahmer and Schmoch (1997) observed a much higher citation frequency in pharmaceutical patents than in mechanical and automobile patents. Narin and Olivastro (1992) also found significant variations in the number of NPRs present in patents belonging to different technology fields. Science interaction is thus a field-specific phenomenon, much more as it is country-specific, which does however not imply the absence of national influences.

But why does an examiner specifically cite scientific literature (besides patent documents)? Grupp and Schmoch (1992a) have identified a number of reasons varying in their reflection of the science involvement in the invention process. The reasons reflecting possible science interaction are related to the limited availability of patents describing the prior-art, the examiner's intention to cite scientific literature, and the inaccessibility of patent documents due to the fast development of certain technology fields. Another reason for the occurrence of NPRs is the so-called 'hidden' patent references (usually Japanese language documents) that are retrieved via English abstract services, and as such end up as non-patent references. As such, the degree of science involvement reflection varies per NPR. Beyond these more rational features of the examination process there are also a number of other reasons for NPRs to be included in the prior-art description. They are mainly the result of the social character of the patenting process. Several actors (inventor, examiner, patent-attorney, colleagues, etc.) are involved in the patenting process and willingly or unwillingly influence the shaping of the patents. As such, NPRs originate from a highly mediated process, which certainly has interpretational implications.

Finally we would like to mention the role of the examination offices in the frequency of NPR availability. The higher citation frequencies present in US-covered patents, compared to EU-covered patents, is mainly caused by differences in the examination procedures. The main argument in this respect is the so-called "duty of disclosure" in the United States. Whereas in Europe the applicant can choose to introduce prior-art known to him in the examination procedure or to refrain from doing so, the US law stipulates that the applicant is obliged to refer to any prior documents known to him to the USPTO, for as long as the application is under examination. This may explain why citation frequencies in US-covered patents are higher than in Europe. Due to this multitude of references a wide and universal S&T modelling becomes possible.

5. The nature of the citation link

It has been argued in several studies that NPRs indicate the "science relatedness" of a technology field. In general, a higher number of NPRs is observed when a particular technology domain is more science-based. In line with the previous discussion however no direct or causal relation can be established as to inter-connected patents and publications (see also Meyer, 2000a). On this 'micro-level' no such relation can be established, except for perhaps a very limited number of cases, where indeed a scientific discovery directly led to a technological application. On the 'meso-' (S&T sub-fields) and 'macro-levels' (S&T fields), the analytical and interpretational possibilities increase. On these levels the technology fields involved touch upon related scientific areas of importance, and vice versa. As such, in line with the modern views on knowledge production and diffusion, scientific research constitutes relevant background knowledge playing an important indirect rather than direct role with respect to technological development.

The question is however whether science really pushes technology, and if so, whether and to what extent patent citations actually reflect this science-push idea. Traditionally, citation links between patents and publications are viewed as an indication of the contributions of science to technology (the linear approach). Recent findings do not contest the strong relationship between science and technology and their impact on economic progress, but they do question the assumed direction of the knowledge flow between science and technology or, between academia and industry. As Toynbee (1963) also pointed out it is becoming increasingly difficult to differentiate between S&T. Disappearing boundaries between research disciplines and even research organisations, introduction of multi-task research teams with, in many areas, a strong focus on application, makes it increasingly difficult to judge whether science pushes technology or the opposite.

The early discovery of food preservation in tin-coated steel by Nicholas Appert in 1810, and the explanation of this process much later in 1873 with the discovery of the role of micro-organisms in food spoilage, the birth of the science of bacteriology, is an early example of science lagging behind technological development (technology pull). Just as Meyer (2000a) concludes, the S&T interaction seems to be much more reciprocal than the linear model suggests. In this light, NPRs indicate much more the kind of closeness between science and technology and not so much a direct scientific contribution to technology. However, there seems to be no consensus as to the possible role of NPRs in S&T linkage studies and to the interpretational boundaries that this approach is subject to. The studies performed by the CHI (Computer Horizons Incorporated) concluded that the technology areas whose patents cited scientific papers, were in fact rated by their peers as far more science dependent than areas of technology which did not interact with science (Narin & Olivastro, 1998).

6. Science and Technology Linkage Methodology

The developed S&T linkage methodology is based on NPRs as units of analysis, and more specifically the citations to scientific journal publications. The identification of NPRs and the subsequent identification of the 'source' publication in the Science Citation Index (SCI [Thomson-ISI]) enable the broader interconnection of S&T fields, through respectively the IPC classification of the patent involved and the SCI-ISI (Institute for Scientific Information) journal classification system. Only the 'front-page' references will be taken into consideration. Besides the SCI, we also make use of the patent data of the United States Patent and Trademark Office (*US Patent Bibliographic Data; 1978 onwards*), and the European Patent Office (*1978 onwards*). In regard of the European patent data, the REFI file (*1978 onwards*), containing all patent- and non-patent citations of European patent documents, has been additionally acquired. Our in-house database (INCENTIM-database) contains the mentioned data in a fully normalised setting (relational environment). We shall continue our methodological review by presenting the broader methodological framework that has been designed in regard of the aforementioned EC-project (see figure 1). Please note that the scope of that framework is wider than what possibly can be reflected upon in this paper.

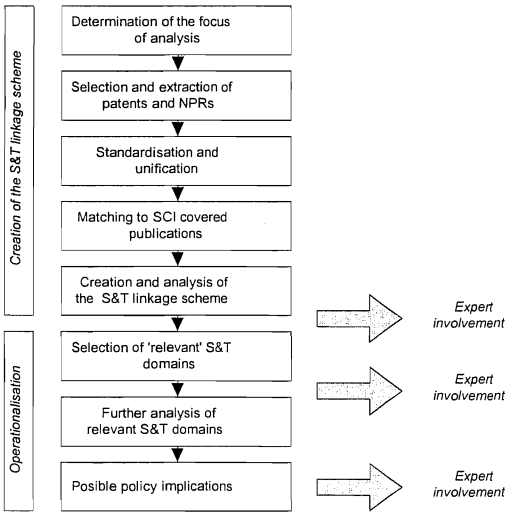


Figure 1 – Methodological Framework

In general, two major phases can be distinguished. The first phase consists of a number of steps leading to the creation of S&T interaction model. The developed methodology can be applied repeatedly in order to update the S&T interaction scheme regularly. The second phase includes a number of steps transferring the results of the first phase to a more policy-oriented setting. At this point we also supplement the up to now mainly quantitative approach, with more qualitative elements mainly through intensive expert-involvement.

(1) Determination of the focus of analysis

This first methodological step, mainly an awareness step, consists of making a choice in the coverage of the S&T modelling. The analysis can be time- and/or field-related. In the above mentioned EC-project the time span covers a period of 17 years (1980-1996), subdivided in 3 analytical benchmark periods of a different time-span (1984-1986, 1987-1991 and 1992-1996). By benchmarking the science and technology interaction the co-evolution of the S&T interaction can be analysed. In the present paper we shall focus our discussion on the period 1992-1996, a period that is consider to be the 'reference' period.

In other words, modelling the S&T interaction for the period 1992-1996 leads to an 'actual' linkage scheme that can be used as input for further analytical steps and policy issues. This period involves a sufficiently high number of NPRs (68% of all NPRs in the period 1980-1996). As the S&T interaction pattern remains rather stable over time, an actual linkage scheme can be derived.

(2) Selection and extraction of patents and NPRs

The objective is to select those patents, and the NPRs they contain, that comply with the criterion set in step 1. In case of a field-related S&T analysis, alternately IPC- and keyword-based search strategies may be applied. In the analysis reported here, we have selected all US covered utility patents that have been applied for between 1992-1996 (Application Filing Date). As a result the 'earliest' possible knowledge transfer stage is captured. We chose not to work with the 'Priority date' for reasons of availability of this type of date-field (827.861 priority dates on a total of 2.259.780 patents), and as such to prevent usage of mingling date-types. In total, 656.695 'inventory' patents with an application filing date lying between 1992-1996 were selected, after which all NPRs present in those selected patent documents were retrieved for further processing (in total 1.147.160 NPRs). Within the diverse collection of NPRs we specifically focussed on journal citations. Scientific journal publications form the primary communication medium within the scientific community, and as such they are a proxy measure of scientific activity. The final aim was to identify the 'source' publications covered in the SCI-data, through the application of a match-key based approach.

(3) Standardisation and unification of Journal references

A complex parsing algorithm, based on a textual analysis approach, has been designed in order to identify and parse the scientific journal references into a number of components such as {author name} and {publication year}. Grammatical deviations such as misspelling, misplaced points and/or commas, capital letters versus small letters made this operation complex and very time consuming. Several iterations proved to be necessary. From each journal reference we identified and extracted {lead author name}, {publication title}, {journal title}, {volume}, {number/issue}, {publication year}, and {starting page}. Each text fragment has been assigned to one of these data types, after which they underwent a number of standardisations. For example, a text fragment like "vol. 55" had to be transformed to "55", "12-05-1986" had to be transformed in "1986". For the period 1992-1996, 296.679 scientific journal references (26% of all NPRs) have been identified, successfully parsed, and subsequently standardised.

(4) Matching to SCI covered publications

The approach developed to trace the 'source' publication covered by the SCI, relies on a match-key based approach (based on the work of Luwel, 1999a). The match-key is composed of a combination of the following fields {lead author name}, {publication year}, {volume} and {starting page} – see figure 2. As such the use of the journal title for matching purposes, which displays misspellings, synonyms and even acronyms, could be avoided. Once the 'source' publication has been identified in the SCI, the related science field has been detected by tracking the SCI journal classification. The matching process was carried out in a number of iterations. Initially, all four fields were used in the composition of the match-key. In the subsequent three rounds we interchangeably used one 'free-floating' field, except for the field {lead author} that was a fixed element in all rounds. The lead author's name has been reduced to the first six letters so that discrepancies between citation and 'source' publication could be prevented (Luwel, 1999a). When all match-key fields squared with the corresponding fields in the SCI, the journal citation in question was assumed to be uniquely linked to an SCI covered source publication (this appeared to be the case in 106.636 journal citations).

(5) *Creation and analysis of the linkage scheme*

The last step in the first phase is the creation of the S&T interaction linkage scheme. Technology and science areas are being operationalised respectively by IPC 4-digit classes and

SCI-ISI journal classification (see figure 2). Departing from the match between a specific journal reference, given in a patent, and the 'source' publication covered by the SCI, we subsequently traced the IPC class of the patent involved and then the classification of the journal in which the 'source' publication appeared. On this higher level of aggregation, S&T interrelation patterns become visible. The linkage results are projected into a matrix consisting of technology classes, scientific sub-fields, and the cross-citation frequency between both spheres (cell-values). A 'normal' counting scheme has been applied. All inventory patents have been re-classified into an IPC 4-digit level. Besides the description and analysis of the S&T interplay, a number of additional indicators have been calculated. By the application of 'self organising' neural-net based clustering algorithms the basic matrix of the S&T interaction has been clustered in order to identify interrelated groups of S&T domains, based on the underlying referencing patterns and frequency. So far the 'static linkage analysis'. For a number of policy-relevant technology and science fields, a 'dynamic linkage analysis' will be performed. The co-evolution of patenting and publication activities just as the general evolution of the S&T interaction patterns over time will be analysed. The dynamic linkage analysis is part of phase 2. Steps 4 and 5 are illustrated in figure 2.

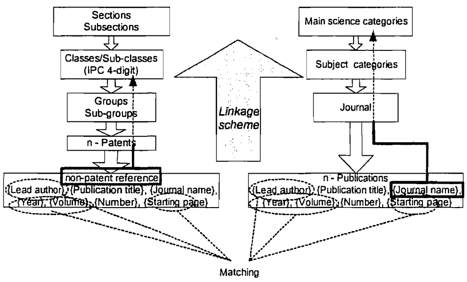


Figure 2 – Matching and S&T linkage procedure

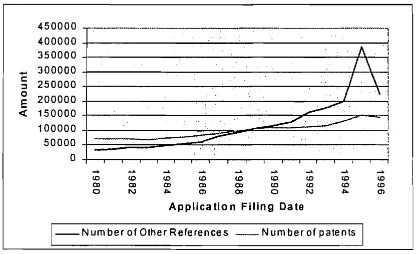
The methodology and the first results of its application, have already been validated by our expert review committee; a group of international experts closely related to this project. In phase 2 we shall zoom into a number of policy 'relevant' domains in which a detailed analysis of the S&T interaction will be performed. This selection process has been recently finalised. The EC 6th Framework Research Priorities will be considered as leading principles, thereby putting emphasis on Biotechnology; IT and Telecom; Nanotechnology, intelligent materials and new production processes; Aeronautics and Space; Food safety and health risks; Sustainable development and global change. A short digital E-mail Delphi questionnaire has been developed in order to obtain broader expert-validation on whether the identified S&T interrelations are valid according to their experience. In the remainder of this paper we shall present and discuss the results of phase 1.

7. Results

7.1 Pre-linkage statistics and findings

Looking at the general co-evolution of USPTO covered patent documents and NPRs (figure 3) it can be observed that as from 1988 the number of NPRs exceeds the number of patents. This is however not caused by a general increase in the number of patents that cite non-patent literature, but due to the intrinsic rise in the number of NPRs per patent in certain technological areas. The enormous rise in the average number of NPRs can be explained by a number of specificities related to the US patenting office. Due to a severe backlog in the US examination procedure the availability of granted patent documents has been marginal in a number of areas (e.g. biotechnology, agriculture). As a consequence, NPRs are being cited instead of patent documents, which normally are cited first in order to describe the prior-art. It is assumable that due to this backlog, patent examiners where stimulated much more than in the past to search for related research. To a certain extent the rise in NPRs, on a higher abstraction level, also is an indication of the increased role of scientific exploration for technological applications, at least in the research-intensive technological areas.

Figure 3 – Evolution in the number of patents and non-patent references (in absolute numbers)



The skewed distribution of NPR's (majority of patents containing no references, while only a fragment of all patents contains numerous references) is an important validity aspect in the direct S&T interaction analysis (Van Vianen et. al., 1990; Schmoch et. al. 1993). For the period 1992-1996 the following distribution is noted: 65% of all patents contained 0 NPRs; 8% only 1 NPR; 19% between 2 and 4 NPRs; 1% exactly 5 NPRs; and 7% of all patents displayed more than 5 NPRs. All together 35% of the patents contain 1 or more reference to a non-patent document. Patents with 5 or more NPRs are partly responsible for the strong increase in the S&T interaction between 1994-1995. Now, is this evolution in patents with 5 or more NPRs equally attributable to all technological patenting classes? The answer is no. A major share of the increase in patents with high levels of NPRs occurs in so-called science based areas such as Pharmaceuticals, Biotechnology, Organic Fine Chemistry, and to a lesser extent also in Instrumentation. These are also the areas that will prove to display a strong S&T interaction. The distribution of references over the number of patents is assumed to be a function of the specificities of the field involved.

The overall analysis of the science cycle time on the level of the identified source publications in the SCI, an indicator of the development speed of technical areas or of the possible presence of 'high-tech' fields, showed an average time lag, between patent application and paper publication year, of 3 years. As such the S&T interaction for the period 1992-1996 can be characterised as intense and dynamic. Relatively 'young' publications are cited thereby supporting the general idea that S&T are getting increasingly intertwined. On a field-specific level we see that the technological areas in which these short science cycles are profound, are related primarily to Biotechnology, Organic Fine Chemistry, Semiconductors, Control technology and IT. These are also the areas in which a steep intrinsic increase in the number of NPRs can be observed. A last intriguing finding is that in a number of cases we came across negative science cycle times, possibly indicating that the during the examination phase new scientific background material was included.

7.2 S&T interaction modelling

Step 5 of the methodological framework lead to the construction of the S&T interaction matrix. The matrix consists of 441 IPC 4-digit classes (rows) and 187 related science subfields (columns). The absolute number of cross-citations (linked scientific journal references) displays the intensity of the interaction; these are the cell values in the matrix. Each connected technology class is interacting with one or more areas of scientific research. By distributing the uniquely linked journal citations over the different IPC-classes and science sub-fields, inflation of the number of traced journal citations by a factor 1.75 (from 106.636 to 184.959) is unavoidable. The matrix also provides information as to the role of science fields in the interaction, the intensity of this interaction and the technology associates involved. The S&T

domains that are of importance for their mutual development are depicted and related to each other.

One of the expectations towards the results of this analysis was the discovery of a wide and varied map of S&T interactions. On the contrary, regarding the complete S&T interaction landscape as a point of reference, it appears that 7% of all technology fields (31 individual IPC sub-classes) account for more than 80% of the total science interaction. Moreover, these 7% represent a patent share of 40% of all patents in the interconnected sub-classes. Analogously, 20% of all interconnected technology fields (89 IPC sub-classes) account for 60% of the total patent population (see figure 4).

Figure 4 – Distribution of journal references and patents

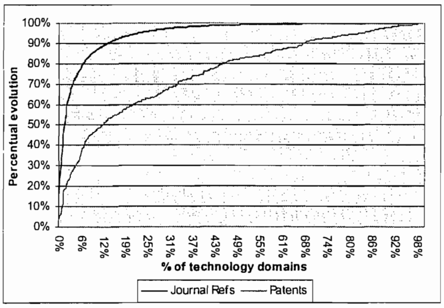
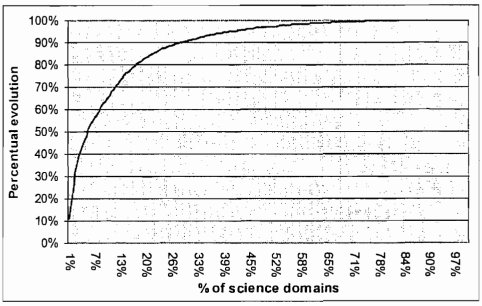


Figure 5 – Distribution of citation output over the science fields



When analysing the characteristics of the role of science subfield in the S&T interaction, a similar diagnosis can be established. Only 18% (33 domains) of all science domains account for more than 80% of all technology relevant scientific output (figure 5). The backbone of the S&T interplay is formed by a limited number of science and technology sub-fields. In regard of the further description of the S&T interaction pattern two new indicators have been developed. The first one concerns the 'science absorption ratio', the number of science fields a certain technology interacts with. The second indicator, 'science diffusion ratio', is defined as the number of technology areas touching upon one single science field. Both indicators reveal the multisciplinarity of the interaction between both spheres. A low 'science diffusion ratio' where the output of a science field interacting with only 1 or a limited number of technology areas implies a certain specialisation of the science area involved. A higher ratio on the contrary points towards broader oriented science sub-fields. The lower the ratio the higher the importance of the field for the development of the technologies in question. Similarly, the 'science absorption ratio' indicates whether a certain technology has a broad or a narrow science orientation scope. The multidisciplinarity of science and technology areas can be deduced by this approach. As to the 'science absorption ratio', 50% of all technology domains interact with less than 12 science sub-fields. As to the 'science diffusion ratio' we have established that 50% of all science fields interacts with less then 43 technological areas (31% with less than 20). As such, based on this macro-level analysis, it can be concluded that for the period 1992-1996, the science field interaction of technology is far more concentrated (50% of the technology fields interact with less than 12 science fields), than the technology interaction of most science fields (50% of all science fields are related to less than 43 technological areas). The orientation of science fields as well as technology fields is rather focused.

Analysis of the most intensive science interacting technology domains (threshold value set on 80% of all science interactions) and the intrinsic science intensity of the technologies measured by the citation propensity, results in the overview presented in table 1. Based on the share in the total S&T interaction the top-10 of most science related technologies is constituted by Pharmacology (A61K), Biotechnology (C12N,Q,P), Organic Fine Chemistry (C07K,D,H), Semiconductors (H01L), IT and specifically Electrical digital data processing (G06F), and finally Material analysis focused on chemical or physical properties (G01N). Biotechnology and Organic Fine Chemistry, broadly considered as science dependant areas, are indeed profoundly present. However, when looking at the highest intrinsic science interaction of these technologies (the propensity ratio), we observe that Optics (H01S), Medical technologies, and specifically Electrotherapy, Magnetotherapy, Radiation therapy, Ultrasound therapy (A61N), and another sub-area of Organic Fine Chemistry (Acyclic, carbocyclic, or heterocyclic compounds containing elements other than carbon, hydrogen, halogen, oxygen, nitrogen, sulfur, or tellurium - C07F) are included in the top. On the other hand, Semiconductors, Electrical Digital data processing, and Material analysis are pushed out of the top-10 but are still situated within the top-30. The absolute number of total science interactions of a certain technology area does not suffice for deciding upon the intensity of the science interaction. The science intensity of the field involved, for example through the propensity ratio, should also be considered.

As to the identified science associates that play a significant role in the technology interaction, Biochemistry & Molecular Biology, Biophysics, Chemistry, Electrical & Electronic engineering, Immunology, Pharmacology & Pharmacy, Cancer, Organic Chemistry, Applied Physics, Instruments & Instrumentation are most frequently present. In the S&T interaction, a substantial role is played by multidisciplinary research accounting for more than 7% of all technology interactions. The important role of multidisciplinary research in technological development may point towards the genesis of young rapidly growing areas of research, which already have found technological applications, but are not yet established in terms of maturity.

Table 1 – Overview of the most science intensive technologies based on the absolute number of citations

IPC class	*Description*	*# S-interactions*	*Impact*	*#Patents*	*T&S impact*	*Propensity ratio*
A61K	Preparations for medical, dental, or toilet purposes	29264	0,16	26644	0,04	1,10
C12N	Micro-organisms or enzymes and compositions thereof	15949	0,09	9908	0,01	1,61
C07K	Peptides	10657	0,06	5858	0,01	1,82
H01L	Semiconductor Devices and electric solid state devices	9903	0,05	22075	0,03	0,45
C07D	Heterocyclic compounds	8759	0,05	13060	0,02	0,67
G06F	Electrical digital data processing	7341	0,04	30813	0,05	0,24
G01N	Investigating or analysing materials by determining their chemical or physical properties	7303	0,04	12687	0,02	0,58
C07H	Sugars, derivatives thereof, nucleosides, nucleotides and nucleic acids	6702	0,04	4787	0,01	1,40
C12Q	Measuring or testing processes involving enzymes or micro-organisms	6246	0,03	4401	0,01	1,42
C12P	Fermentation or enzyme-using processes tot synthesise a desired chemical compound or composition or to separate optical isomers from a racemic mixture	5997	0,03	4423	0,01	1,36

A more detailed analysis of the journals cited provides additional information as to the role of science in the interaction with technology. In view of the previously mentioned policy reformulation that has taken place in Europe as an answer to the socio-economic evolutions, it is of great interest to be able to differentiate between the general research orientation of the science associates of the relevant technologies. Currently, this is in progress. Looking at the number of journals that account for the technology interaction, a similar skewed distribution is noted, which is in line with the above-discussed findings in regard of the distribution of the interactions. Less than 10% of all journals (355) account for almost 75% of all citations. Apparently not only a low 'science absorption ratio', but also a limited number of scientific journals empowering the S&T interaction, can be observed. Similar findings have also been reported by Van Vianen et. al. (1990) in Chemical technology and Schmoch et.al. (1993), in the field of Biotechnology. The multidisciplinary journals "Nature", "Science" and "Proceedings of the National Academy of Sciences of the United States of America" are frequently present. From a patenting point of view, this is in line with the normal practice of the patent offices to generally cite patents (of a more applied character) and only fall back on papers, if patents are less available.

The frequent presence of papers edited by IEEE (the Institute of Electrical and Electronic Engineers, an American association of engineers) is also of interest. Most of the few covered journals are IEEE journals whereas many often more important and more basic journals are not included. According to Schmoch et. al. (1993), this points towards a weak representation of electronics and information technology by the SCI database. Also strongly present over the years is the more general journal of "Applied Physical Letters", most cited journal over the years except for 1995.

Besides this high level impression of the 1992-1996 S&T interaction, it is equally important to identify the science associates on the level of the individual technology subfields. Each technology subfield involved in the S&T interaction has been analysed in terms of its science associates. For illustration purposes we shall discuss the science interaction in the case of Pharmaceuticals, the more application oriented side of Biotechnology. Pharmaceuticals (A61K), displays a science interaction intensity of 29464 citations. The science absorption ratio of Pharmaceuticals equals 148 science fields implying that this technology area touches upon the research of 148 science subfields. The strength of the interrelation varies between 1 and 4405 journal citations. In determining the major science associates of a specific technology, a threshold value of 5% of total science involvement to that technology has been applied (science absorption ratio amounts 4) in order to prevent inclusion of less relevant science areas with limited numbers of interactions. In table 2 we illustrate the 10 science fields that are of major importance for Pharmaceutical development.

Table 2 – Overview of the science associates of Pharmaceuticals

Science subfield	*% of science interactions*	*# number interactions*
Biochemistry & Molecular Biology	15.05	4405
Pharmacology & Pharmacy	12.19	3568
Multidisciplinary	9.49	2777
Immunology	6.07	1776
Chemistry, Organic	4.30	1259
Biophysics	3.73	1091
Cancer	3.34	976
Chemistry	2.88	842
Medicine General & Internal	2.86	836
Endocrinology & Metabolism	2.58	754

By modelling the S&T interaction it appeared that the science subfields of Biochemistry & Molecular Biology, Pharmacology and Pharmacy, Multidisciplinary science, and Immunology can be regarded as the science associates of Pharmaceuticals. As such, a major part of the development in this technology area depends on the research activity within these subfields, not in causal meaning but rather in reciprocal supportive one, thereby not wishing to understate the importance of scientific research in those areas for further technological development.

7.3 Discussion

From a socio-economic perspective, but also from an EU science and technology policy perspective, it is important to have a thorough understanding of the S&T interaction. A first step towards increasing our understanding of the dynamics in the science and technology interaction is measurement. It is in this context that we present and demonstrate a S&T linkage methodology enabling the modelling of the universal S&T interaction patterns. The developed methodology is based on non-patent references as units of analysis; the interpretational limitations however have to be taken thoroughly into consideration.

Especially because of the use of USPTO patent data, which according to many analysts offers the best possibilities for a varied and wide S&T modelling, we expected to come across diverse and dispersed interrelation patterns. The results of our analysis however pointed towards a more focused S&T interaction, the backbone of which is constituted by a limited number of S&T subfields. As to the issue of causality of the identified interrelations, even on this higher level of abstraction where technology areas are related to one or more science fields, nothing definite can be stated. However, there are many indications that the interrelation is of mutual importance for future development. As such the analysis illustrated here can be regarded as a starting point, and even a point of reference, for future disentangling of the S&T interrelation. Furthermore, the developed methodology offers the possibility of repetitive application. As such it is of major importance for the S&T interaction benchmark activities in Europe, thereby taking into account the specificities of each country's system of innovation. The more science-based technologies such as Chemistry, Life Sciences, and ICT also factually appear to be interacting intensely with several science fields.

Phase 2 of the methodological framework is already in progress. The challenge during this transition is to coincide with policy-related issues and mindsets, and especially to translate science and technology domains, as found in the analysis, to political-administration domains. That is also the reason why extensive expert involvement and EC-interaction has been foreseen. It is important to ally with policy-makers notions of research and technology areas. A first step in that direction has already been performed. Based on the EC Framework Priorities (2002-2006), IPC-sub-classes have been re-grouped into broader technological sub-fields. This is the subject of an ongoing Delphi survey involving several fields-specific across in Europe. On a field specific base, actor-related analyses (positioning of countries in the S&T model) will be performed. This paper touched upon a number of issues and results related to the developed

S&T linkage methodology and its application. In the course of this project several other publications are expected to appear.

References

Collins, P., and S. Wyatt (1988), "Citations in patents to the basic research literature", Research Policy, 17, pp. 65-7

De Solla Price, D. J. (1965), "Is Technology Historically Independent of Science? A study in Statistical Historiography", Journal of Technology and Culture, 6, pp. 553 – 568

Dosi, G. and S. Fabiani (1994), "Convergence and Divergence in the Long-term Growth of Open Economics", In: Silverberg & Soete (eds.

Gibbons, M., C. Limoges, H. Nowotny, S. Schwartzman, P. Scott and M. Trow (1994), "The New Production of Knowledge", Sage, London

Grilliches, Z. (1990), "Patent Statistics as Economic Indicators: A Survey", Journal of Economic Literature, 28, pp. 1661-1707

Grupp, H. (1998), "Foundations of the Economics of Innovation", Edward Elgar, Cheltenham, UK; Northampton, MA, US

Illinois Institute of Technology Research Institute (1969), "Report on project Traces", National Science Foundation, Washington DC

Luwel, M. (1999a), "Is The Science Citation Index US-Biased", Scientometrics, Vol. 46, No. 3, pp. 549-562

Meyer, M. (2000a), "Does science push technology? Patents citing scientific literature", Research Policy, 29 (3), pp. 409-434

Meyer, M. (2000b), 'What is special about patent citations? Differences between scientific and patent citations', Scientometrics, 49, 1, pp. 93-123

Meyer-Krahmer, F. & U. Schmoch (1998), "Science-based technologies: university-industry interactions in four fields", Research Policy, 27, pp. 835 – 851

Moed, H.F. (1989), "The Use of Bibliometric Indicators for the Assessment of Research Performance in the Natural and Life Sciences", Doctoraatsdissertatie Rijksuniversiteit Leiden, DSWO Press

Narin, F., and D. Olivastro (1992), "Status Report: Linkage Between Technology and Science", Research Policy, 21, pp. 237-330

Narin, F., and D. Olivastro (1998), "Linkage Between Patents and Papers: An Interim EPO/US Comparison", Scientometrics, 41, pp. 51-59

Narin, F., M. Rosen and D. Olivastro (1989), "Patent citation analysis: new validation studies and linkage statistics", In: Van Raan et. al. (1989)

Nelson, R.R. (1994), "What has been the Matter with Neoclassical Growth Theory?", In: Silverberg & Soete (eds.)

OECD (1994a), "The Measurement of scientific and technological activities. Using Patent Data as Science and Technology Indicators", Patent Manual 1994, Paris

Rosenberg, N. and L.E. Birdzell Jr. (1990), "Science, technology and the Western miracle", Scientific American, 263 (5), pp. 42-54

Rosenberg, N. (1982), "Inside the black box: technololgy and economics", Cambridge University Press, Cambridge

Schmoch, U. (1993), "Tracing the knowledge transfer from science to technology as reflected in patent indicators", Scientometrics, 26 (1), pp. 193-211

Schmoch, U., E. Strauss, H. Grupp and T. Reiss (1993), "Indicators of the Scientific Base of European Patents", Report to the European Commission, EUR 15 330 EN, Brussels/Luxemburg

Sherwin, C. and R. Isenson (1967), "Project Hindsight", Science, 156, pp. 571-577

Silverberg, G. and L. Soete (eds.), (1994), "The economic growth and technical change: technologies, nations, agents", Elgar Aldershot

Steinmueller, W.E. (1994), "Basic Research and Industrial Innovation", Chapter 5 in Dodgson, M. and Rothwell R. (eds.), The Handbook of Industrial Innovation, Edward Elgar, Aldershot, pp. 54-66

Steinmueller, W.E. (2000), "Economics of Science", In: Pavitt, K., E. Steinmueller and J. Calvert & B. Martin (2000), "International Encyclopedia of the Social and Behavioural Sciences: 3 papers", Paper no. 44 of SPRU Electronic Working Papers Series, Science Policy Research Unit, University of Sussex, UK

Toynbee, A. J. (1963), "Introduction: The Genesis of Civilisations, A Study of History", 12 vols, 1, 3, New York
Van Vianen, B.G., H.F. Moed and A.F.J. van Raan (1990), "An Exploration of the Science Base of Recent Technology", Research Policy, 19, pp. 61-81
Ziman, J. (1994), "Prometheus Bound. Science in a Dynamic Steady State", Cambridge University Press, Cambridge