The Knowledge Bases of the World's Largest Pharmaceuticals Groups: What do patent citations to non-patent literature reveal?

by

Stefano Brusoni,[*] Paola Criscuolo[*♠] and Aldo Geuna[*1]

** SPRU –University of Sussex, UK*
♠MERIT, University of Maastricht, NL

January 2003

Useful comments from Lionel Nesta are gratefully acknowledged. The authors are extremely grateful to Juan Mateos Garcia who provided research assistance during the project NewKInd; this paper is based on the results of that project. A early version of this paper has been presented at the Enterprise Knowledge Workshop: Defining and Measuring Knowledge, NESIS, Luxembourg: 16-17, January, 2003. Financial support from the Commission of the European Communities-EUROSTAT, IST project "NewKInd" is acknowledged.

[1] *corresponding author*: SPRU – University of Sussex, Mantell Building, Falmer, Brighton, BN1 9RF (UK), Tel: +44 (0)1273 877139, Fax: +44 (0)1273 685865, E-mail: a.geuna@sussex.ac.uk

Abstract

This paper examines the knowledge bases of the world's largest pharmaceuticals groups by sales. It puts forward the concepts of knowledge *breadth* and *depth* as the relevant dimensions along which knowledge bases can be mapped. Breadth is studied by analysing the evolution of specialisation by scientific field over time. It hints at the widening range of bodies of scientific and technological knowledge relevant to firms' innovative activities. Depth (or integration) is studied by analysing the evolution of specialisation across different typologies of research. It hints at the complex, non-linear interdependencies that link the scientific and technological domains. We develop the analyses on the strength of an original database of 33,127 EPO patents, and of 41,931 'non patent document' citations (of which 19,494 were identified as scientific articles included in the ISI databases). The groups studied seem to have incrementally increased the breadth of their knowledge bases, moving toward the fields proper of the new bio-pharmaceutical research trajectory. At the same time, some of the groups studied exhibit remarkable depth in knowledge specialisation in particular fields such as biotechnology, biochemical research and neurosciences. Finally, this paper also provides a first methodological test of possible problems deriving from the use of 'unidentified' patent citations. We compare a random sample of citations with a sample of citations explicitly added by the original inventor, and compare the results in terms of scientific specialisations.

Key words: Knowledge Breadth, Depth, Integration, Patent Citations, Scientific Publications, Pharmaceuticals.

JEL: O3, L2, L65

1. Introduction

This paper builds upon previous research into the knowledge boundaries of firms. In recent years, empirical and theoretical research has bestowed considerable attention on issues related to the changing characteristics of firms' knowledge bases. In particular, some authors argue that there is considerable support for the hypothesis that the knowledge bases relevant to firms' innovative activities are increasing in breadth, i.e. in their range of useful disciplines, and depth, i.e. in their analytical complexity (e.g. Granstrand, Patel and Pavitt, 1997; Wang and von Tunzelmann, 2001). This trend is often associated with the emergence of new ways of organising and coordinating innovative processes that rely not only on the activities of a few, large and integrated innovating firms, but more on networks of innovators that share the burden of developing new products and processes (e.g. Powell, 1990; Arora and Gambardella, 1994). This paper aims at developing indicators that capture some key characteristics of the knowledge bases of the firms that coordinate these networks of innovating organisations in the case of the world pharmaceutical industry.

Indeed, alongside the increasing importance of networks of innovators, recent research based on in-depth case studies of industries and firms highlights the pivotal role played by firms that maintain wide capabilities to be able to act as network coordinators. These firms have been termed 'systems integrators' (Prencipe, 1997; Miller et al., 1995). Systems integrator firms rely on wide science and technology (S&T) capabilities to co-ordinate, from an organisational and technological viewpoint, the manufacturing and research activities they have outsourced. On this basis, it has been argued that the 'knowledge boundaries' of (systems integrator) firms differ from their 'production boundaries' and that the 'knowledge' and 'product' domains evolve according to different principles (Brusoni and Prencipe, 2001). The integration and co-ordination of dispersed, decentralised knowledge-acquisition processes may require further knowledge-related investment that allows (some) firms to act as 'loci of integration.' This observation is particularly important for economic analysis because it governs the extent to which knowledge-related inputs, at the level of the firm, can be expected to be accurate indicators of the knowledge available to the firm. When a firm is able to draw on more extensive networks of knowledge and is able to effectively coordinate these dispersed sources of knowledge generation, traditional approaches to understanding the relationship between inputs and outputs are challenged.

This paper develops indicators and methodologies to tackle quantitatively the issue of what are the specific characteristics of the knowledge bases of those firms that rely on, and coordinate, networks of specialised suppliers of both equipment and knowledge. To answer this, the paper examines the knowledge bases of the world's largest pharmaceuticals groups by sales. In recent years the pharmaceutical industry has attracted considerable attention from scholars of industrial economics and technical change, and recent research has highlighted the key coordinating role played by large pharmaceutical firms with respect to their (increasingly) wide networks of suppliers of very specialised knowledge (Orsenigo, Pammolli and Riccaboni, 2001).

We develop the analysis on the strength of an original database of 33,127 EPO patents, and of 41,931 'non patent document' citations (of which 19,494 were identified as scientific articles included in the ISI databases). We analyse two key indicators. First, we assess firms' specialisation patterns (and their evolution during the 1990s) at the scientific level relying on the analysis of citations from patents to scientific literature. The analysis of specialisation by scientific discipline captures changes in the *breadth* of firms' knowledge bases. Second, we analyse firms' specialisation patterns in terms of the different types of research and development activities they perform. This analysis is made possible by the Computer Horizons Inc. (CHI) classification, which links scientific journals to specific types of research and development (e.g. clinical observation, a mix of clinical observation and clinical investigation, and clinical investigation and basic research). The analysis of specialisation by type of research (within each discipline) is an indication of the *depth* of firms' knowledge bases.

The paper is structured as follows. Section 2 describes the key dimensions of firms' knowledge bases that will be analysed and why the pharmaceutical industry was chosen for this study. Section 3 discusses the methodology of citation analysis and presents the data sources on which this paper builds. Section 4 develops the quantitative analysis of the breadth and depth of groups' knowledge bases and Section 5 discusses the robustness of the results. Finally, Section 6 concludes.

1. The changing structure of knowledge accumulation

The issue of integration and co-ordination of labour is of great relevance as it directly impinges upon the evolution of the key input to innovative activities: knowledge. As early as 1945, von Hayek considered that the key economic problem was 'the problem of the utilization of knowledge not given to anyone in its entirety'. In what follows, we maintain that the principle of specialisation applies to the process of developing new bodies of knowledge. Knowledge develops through the generation of branches and sub-branches of increasing depth and specificity (Loasby, 1999). The principle of specialisation applies at the level of: (a) disciplines, e.g. physics, chemical engineering, economics; (b) institutions, e.g. firms' R&D laboratories, government research centres, universities; and (c) functions, e.g. corporate R&D units, engineering units, business development units (Patel and Pavitt, 1998: 4). These areas intersect in complex ways. Knowledge generated through division and specialisation is both incomplete and dispersed. Such incompleteness and dispersion appear to be inherent properties of knowledge itself.

Incompleteness implies that learning processes cannot converge to any state of rest, as such a state does not exist: knowledge cannot be defined as a closed and well defined set of events that only some form of bounded rationality prevents from being fully discovered. Dispersion relates to one of the main functions performed by organisations: the co-ordination and integration of heterogeneous learning processes. Given the inherent limitations of human cognition, idiosyncratic learning processes need to be 'completed' by being embedded in a suitable institutional framework (Loasby, 1999). The (increasing) variety of the knowledge bases relevant to firms' innovative activities underlines the importance of the analysis of the mechanisms and processes that allow coordination of specialised learning processes to be achieved.

2.1. The dimensions of knowledge specialisation: breadth and depth

Within industry studies type research, Granstrand, Patel, and Pavitt (1997) found that large firms are more diversified in terms of the technologies that they master than in the products that they make, and that their technological diversity has been increasing while typically their product range has narrowed. Similar results have emerged from studies of highly innovative sectors, such as aero-engines (Prencipe, 1997), telecommunications infrastructure (Davies, 1997), and hard disk drives (Chesborough and Kusunoki, 2000). Other evidence is emerging from detailed studies of traditional sectors such as chemical engineering (Brusoni, 2001), tyres (Acha and Brusoni, 2002), oil exploration (Acha, 2002) and automotives (Takeishi, 2002).

Moreover, these studies confirm that (some) firms within networks of vertically related companies maintain S&T capabilities over a set of fields wider than can be justified by only their in house activities. In other words, some firms appear to be specialised in terms of the number of activities they carry out in house, while at the same time maintaining wide (and even widening) in-house scientific and technological capabilities. These firms have been labelled in existing research ‘systems integrators’ (Prencipe, 1997). These firms ‘know more than they make’ which allows them to act as problem solvers of last resort should, for example, contingencies emerge anywhere in the value chain (Brusoni, Prencipe and Pavitt, 2001).

Past research on systems integration activities have highlighted the dimensions along which the characteristics of the knowledge bases on which integrating firms rely should be evaluated (Prencipe, 2000). First, integrating firms need to maintain capabilities on most (if not all) the bodies of scientific and technological knowledge that impinge upon the development of the specific product market in which they compete. We call this dimension *breadth*. Breadth is studied by analysing the evolution of specialisation by scientific field over time. It hints at the widening range of bodies of scientific and technological knowledge relevant to firms’ innovative activities. Breadth of knowledge base is necessary for firms interested in maintaining a leading innovative role, playing the role of ‘intelligent customer’ for their suppliers, and being able to recognise at an early stage new and promising technological developments, and identifying and abandoning as quickly as possible unsuitable development paths.

Breadth is not the only dimension that matters, though. Firms’ investments in specific bodies of knowledge vary according to their perceived importance within firms’ strategies. For example, Granstrand, Patel, and Pavitt (1997) argued that firms’ capabilities could be classified as distinctive, background, niche or marginal. In other words, whether a specific body of knowledge is used to define the operating principle of any given artefact (e.g. combustion in automotive), or whether it is perceived as a possible development that might revolutionise that artefact in the distant future (e.g. cell fuel technology in automotive), or whether it is part of the firm strategy to remain an intelligent customer (e.g. organic chemicals with respect to tyres technology in automotive again) is of significance. That is to say, the knowledge base of firms with respect to specific disciplines may also vary in *depth*. Prencipe (2000) studied depth in the case of the evolution of the aero-engine control system. His study focused on two dimensions of the knowledge bases of aero-engine makers: the stages of the development process performed by engine makers, and the different types of knowledge related to either the combination of control system’s components, or specific components (p. 898). His study showed the importance of considering this additional dimension alongside breadth. However, the methodology developed in that paper is tailored for the specific industry on which the paper focused.

In this study, we endeavour to develop a methodology based on publicly available data, which is then replicable across sectors. Depth is studied by analysing the evolution of specialisation across different typologies of research (i.e. basic, applied and development oriented research). This refers to the complex, non-linear interdependencies that link the scientific and technological domains. Core disciplines demand specifically extensive investments that other disciplines do not because certain competence are needed to develop and expand the knowledge base while different competence are needed to understand and intelligently exploit the knowledge. In this sense, by interpreting depth as specialisation across different types of research we try to operationalise the definition of depth put forward in Wang and von Tunzelmann (2000), who talked about the degree of technical difficulty entailed by, in our case, specific disciplines: the more complex (i.e. deep) a discipline, the more it becomes necessary to focus on developments in that disciplines across all the stages of research.

2.2. Why pharmaceuticals?

The pharmaceutical industry is a fast growing, rapidly changing sector that in the past three decades has greatly increased both in its contribution to economic growth and its visibility in the public policy arena. From our viewpoint, two empirical phenomena related to its recent development are of particular interest. First, much of the growth of the industry is due to endogenous processes of innovation (Orsenigo, 1989; Gambardella, 1995). While changing regulatory frameworks and globalisation of the user markets certainly are playing a role, the driving force of the industry is the emergence of new fields of scientific and technological knowledge. Firms struggle to keep abreast of developments in a number of specialised bodies of scientific and technological knowledge that on the one side provide them with great innovative opportunities, while constantly challenging their established ways of doing things. Increasingly specialised knowledge, and the capability to continuously generate new scientific insights, have become the key to the competitive advantage of the world's leading pharmaceutical firms. Thus, the pharmaceutical industry is characterised by the co-existence of heterogeneous bodies of scientific and technological knowledge that need to be developed, coordinated and integrated. For this reason it is important to consider the issue of the heterogeneity of its knowledge base, and the changing breadth of firms' capabilities.

Also, the industry is growing thanks to the co-existence of two mechanisms of growth. First, the incumbents are growing larger and larger as a result of complex mergers and acquisition (M&A). At the same time new, small, highly specialised firms are entering the industry. Markets and hierarchies, the traditional ideal types of organisations considered by economists and industry analysts, are complemented by increasingly complex webs of relationships linking incumbents and new entrants involved in drug discovery. Networks of innovating organisations have become the organisational cornerstone of the modern pharmaceutical industry (Arora and Gambardella, 1994; Pisano, 1991). For example, Orsenigo, Pammolli and Riccaboni (2001) have analysed in depth the organisation of innovative activities, looking at the characteristics of pharmaceutical R&D networks, and their evolution over time. They pointed to the pivotal role played by large pharmaceutical groups (and a very few first generation biotechnology firms) in designing and coordinating these networks.

The heterogeneity of pharmaceutical firms' knowledge bases is reflected in the fact that development projects in this industry can last for over ten years, and involve a wide range of activities. Not only is taking compounds from discovery to market a highly uncertain process, but different tasks along the value chain also rely on very specific skills and competences. Increasingly, pharmaceutical firms rely on Contract Research Organisations (CROs) and Contract Manufacturing Organisations (CMOs) to undertake specific steps in the value chain. At the same time, the company in charge of the overall development project maintains sole responsibility as far as the regulatory authorities (e.g. the FDA) are concerned. Issues of coordination and control are paramount. Firms involved in these activities need to be able to perform basic research (to identify possible cures), clinical investigation (to test them through the various stages of the trials), and clinical observation (to finally test them in hospitals on patients, gather the data, feed them back to the development team); and also to manage relationships with the relevant regulatory authorities, to conduct negotiations with the various national health services, to maintain links with doctors and hospitals, to organise the marketing and distribution effort, and so on. All these activities must be carefully coordinated.

Does such coordination emerge out of arm's length market relationships, or are some firms in these networks 'more equal than others'? And do they maintain the capabilities needed to intervene should problems arise at any stage of this very deep and, nowadays, rather disintegrated value chain? One way of approaching this question is to consider whether large pharmaceutical firms are focusing their capabilities only on high-end development activities, or are maintaining capabilities also at the level of clinical investigation and clinical observation (stages in the value chain that are increasingly outsourced to contract organisations). Hence, we will develop indicators of the depth of firms' knowledge bases, on the assumption that firms need a deep understanding of their value chain in order to be able to act as integrators. Depth is measured by evaluating firms' capabilities across different types of research and development activities in any given discipline. Being able to assess the depth of a firm's knowledge base in a specific discipline would allow indirect assessment of the 'role' that this firm actually plays in the value chain, i.e. whether it could, in principle, act as an 'integrator'.

In the following section the concepts of breadth and depth will be operationalised on the basis of an original database of citations to non-patent references in the patent portfolio of the 30 largest pharmaceuticals corporate groups.

2. Methodology and data sources.

Patent documents must contain citations to other patents and references to a variety of other documents such as papers, abstracts, conference proceedings, books, etc. (non-patent references) as a legal requirement to supply a complete description of the state-of-the-art. Citations limit the scope of the inventor's claim to novelty and, in principle, they represent a link to previous innovations or existing knowledge upon which the inventor builds. On the basis of this observation, during the nineties, a large body of literature spanning the discipline of economics and science and technology studies has used information contained in citations to 'non patents documents'. Most of these works are based upon US Patent Office (USPTO) data (see, among others, Narin and Olivastro, 1992; Narin, Hamilton and Olivastro, 1997; Trajtenberg, Henderson and Jaffe, 1997), while very few use European Patent Office (EPO) data (Verspagen, 1999; Malo and Geuna, 2000; Tijssen, 2002; Verbeek et al., 2002a). The core issue addressed by these studies is the linkage between science and technology. According to their disciplinary background the studies attempt to trace the knowledge flow between science and innovation or to examine the science dependence of specific technologies. Usually, particular attention is paid to the localisation effect, i.e. is the linkage between innovation and scientific discovery affected by the location of the research is carried out (regional, national, etc)? Non-patent references are used as a proxy (a trace) for knowledge flow between different organisations. The inventor citing an article is seen as the receiver of the knowledge produced by the author of the article in another organisation (being it a university, a firm or a public research centre for example).

Although a fair amount of interest has been generated by the results of the literature referred to above, recent research warns about the risk of over-interpreting the existence of non-patent references as proof of a direct link between cited papers and citing patents (Meyer, 2000), and especially in the case of analysis at the micro level (Verbeek et al., 2002b). Also, most citation studies are not able to precisely identify those citations chosen by the inventor (Schmoch, 1993, briefly examines the citations made by inventors). The patent document reports the citations as chosen by the examiner. Such citations can include all, part or none of the citations originally chosen by the inventor. This problem may further limit the interpretation of citations as being a direct link between the article and the patents. Finally, it must be taken into account that, even if inventors' citations were available, most patent applications are filed with the help of large, and sophisticated, legal offices. Given the increasing importance of the legal aspects of a patent (patent litigation increased substantially in the 1990s), firms hire patent lawyers that are heavily involved in the process of application. Citations may be chosen strategically and thus reflect the legal strategy of the firm (aimed at minimising the risk of litigation, for example), rather than the key linkages and networks on which the inventor relied.

In line with these observations, this paper proposes a less demanding interpretation of non-patent references. In this work, citations are used as a proxy for some characteristics (i.e. breadth and depth) of the knowledge bases upon which firms build. They hint at the bodies of scientific and technological knowledge on which patents (and thus firms) rely, but, for the current purposes, are not meant to imply any direct linkages between the inventors and researchers active in those fields. They are not used to imply any specific network structure, or localisation effects. They point to bodies of knowledge that the inventors (or the examiners) thought relevant for the invention reported in the patent application.

In the context of our analysis the use of EPO patents might represent an advantage with respect to using USPTO patents, because of the different legal requirements concerning the inclusion of citations. While applicants to the USPTO are obliged to cite all the documents describing the state-of-the-art and failing to provide a comprehensive list might result in the patent application being refused, for applications to the EPO providing such a list is voluntary. As shown by Michel and Bettels (2001), this leads to a greater number of citations in US patents (three-and-a-half times as many references to non-patent literature compared to EPO patents) that may not be completely relevant to the patent being applied for. EPO patents do not suffer from this problem and citations in these patents might better reflect the company's knowledge base.

Specifically, this paper examines the knowledge bases of the world's largest pharmaceutical groups (in term of sales in 1997). The 30 groups considered (see Table 1) encompass some 3,500 subsidiaries in 1997. To derive indicators of their knowledge bases we started from the patent portfolios of the groups. More precisely, the database used includes the 33,127 patents filed with the EPO by the 30 largest corporate groups in the pharmaceuticals industry during the period 1990-1997 (the 322 subsidiary patenting units were consolidated with their parent companies in 1997 on the basis of the information in 'Who Owns Whom') as compiled by the Observatoire des Science and Technologie (OST).[1] Table 1 shows the number of patent applications by group for the period under consideration. Hoechst AG with 5,711 is the group with the largest number of patents, followed by another German group, Bayer AG, with 5,534 patents. Four other groups: Novartis Holding AG (3,373), Merck & Co. Inc. (1,769), Rhone-Poulenc SA (1,129) and Zeneca Group Plc (1,067) each had over a thousand patent applications during the period. The three Japanese groups Yamanouchi Pharmaceutical Co. Ltd (150), Sankyo Co. Ltd (201), Shionogi & Co. Ltd (212) and Astra AB (292) had less than three hundred patents each. Consistent with the evidence from the literature about the 'home advantage' (EC, 1997; EC, 2001), our database provides a good representation of EU group patenting profiles, while it is under-representative of the patenting profiles of US and Japanese corporate groups. This is a common bias in patent analysis: corporate groups tend to apply for patents to their home country patent office. Usually inventors apply first in their home country, then they apply for protection abroad for only a sub-set of the home country patents (costs, penetration in the foreign market and other reasons being the reason for this).

{TABLE 1 ABOUT HERE}

From the 33,127 patents 41,931 non-patent references were extracted. Out of these, we identified 25,996 citations (from 11,279 patents) to scientific articles included in the expanded *Science Citation Index* of ISI (SCI). 6,502 SCI citations were to abstracts included in the Chemical Abstracts and not to articles, thus they were excluded. On average each patent cites 1.3 non-patent references and 0.6 scientific articles; however, the 8,474 patents that cited scientific papers from the SCI had, on average, 2.3 citations to articles.[2] We decided to focus only on the scientific publications included in the SCI database for three reasons. First, the SCI database includes publications in peer-reviewed journals with

---

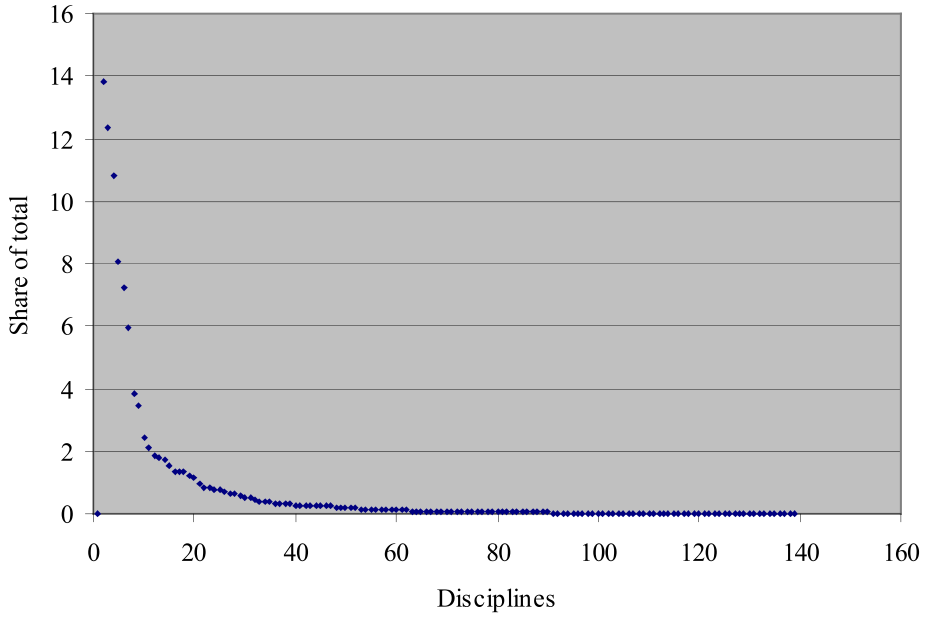
[1] See Filliatreau et al. 2002 for a description of the database. For access to the database see information on the NewKInd web page http://www.researchineurope.org/newkind/Proj_doc.htm.

[2] 2,805 patents were excluded from the final database because either they had only citations to Chemical Abstracts, or they had citations to journals not included in the CHI classification.

international recognition; second, journals in the SCI database are classified in scientific fields; and, third, journals in the SCI database can be linked via the CHI journal classification to four major categories of research -- e.g. applied technology, engineering technological science, applied research, and basic research.

The SCI classification of journals in scientific fields used in this paper includes 132 fields (see Appendix 1). We decided not to aggregate the 132 fields in broader scientific classes, as it has been done in other studies, because we want to capture the detailed specialisation process across groups. We also used a full count method, instead of a fractional one, in counting the number of citations to a particular scientific field. A journal may be classified under one or more scientific fields as it may publish research pertaining to one or several disciplines. The full count method implies that we include all the scientific fields relevant to a journal.[3] Given the focus of this paper, we assumed that if an article is published in a journal classified under several scientific fields this is an indication that the article it will be of as much interest to members of these fields as an article that is published in a single discipline journal. The final database with scientific field classification includes 33,349 entries (see Table 1). As in the case of patent numbers, Hoechst AG is the group with the largest number of scientific field citations while Yamanouchi Pharmaceutical Co. Ltd is the group with the least.

Figure 1: Distribution of citations in the different disciplines



---

[3] The size of our database did not allow us to develop a better classification on the basis of analysis of article titles.

Figure 1 shows the result of the analysis of concentration structures of scientific fields. In this figure, the number of scientific fields cited as a share of the total of citations is shown for the complete population. There is a high level of concentration with a reduced number of scientific fields accounting for a very significant share of the citations, and a long tail of scientific fields with almost negligible amounts. The top six scientific fields in number of citations account for about 58% of citations; they are, in order of importance: biochemistry and molecular biology, multidisciplinary chemistry, pharmacology and pharmacy, organic chemistry, medicinal chemistry and multidisciplinary sciences. The first 16 scientific fields account for 80%. The long tail of the skewed distribution is composed of a large number of scientific fields with a small number of citations; for example, more than 70 fields register less than 20 citations each. The field of biochemistry and molecular biology is the only scientific field present in the top six fields for all 30 groups. Multidisciplinary chemistry is present in 28, pharmacology and pharmacy in 26, organic chemistry and medical chemistry are both present in 22 groups and, finally, multidisciplinary sciences is within the top six fields in 18 groups (see Table 2).

The CHI classification categorises all the SCI journals according to their research orientation, more basic versus more applied, on four broad levels: applied technology, engineering technological science, applied research, basic research. In the context of biomedicine the four types of research levels are: clinical observation (Level 1), a mix of clinical observation and clinical investigation (Level 2), clinical investigation (Level 3) and basic research (Level 4) (Narin and Rozek, 1988). As Narin and Rozek (1988) state, Level 1 is typified by the *Journal of the American Medical Association*, Level 2, by *The New England Journal of Medicine*, Level 3, by the *Journal of Clinical Investigation*, and Level 4, by the *Journal of Biological Chemistry*.

Notwithstanding the limitations to relying on a classification of journals and not articles, this does allow the citations to scientific journals to be classified in research typologies. Using the 1999 CHI classification we categorised the 18,773 references to articles in the four CHI levels (we excluded from the analysis the 721 references to social science journals and to journals not included in the 1999 CHI classification). About 8% of the citations are to applied journals classified in the first two levels, 30% are to journals devoted to clinical investigation and the remaining 61% of references are to basic research journals. The strong reliance on basic research suggests a reclassification of the CHI levels into two major typologies of research: the first including both clinical observation and clinical investigation (levels 1, 2 and 3), the second accounting for the basic research component. This classification will be used in the next section for analysis of the depth of the knowledge bases of corporate groups.

3. The knowledge bases of the world's largest pharmaceutical groups

In this section we analyse a set of indicators of breadth and depth of the knowledge bases of the 30 groups. First, we start by developing indicators of the breadth of the knowledge base over the whole period, followed by an analysis of its evolution. Second, we develop two indicators of the depth of the knowledge base that try to capture the presence of integration capabilities in the group at field level.

3.1 Breadth of the knowledge base

A preliminary exploration of groups' specialisation portfolios was made. We found that, when focusing on the scientific fields in which corporate groups specialise, there is a greater variety than that present in the total population. The profiles of different groups are heterogeneous and depend on the activities in which they engage and the markets they target. Table 2 presents a summary of the individual specialisation profiles of different corporate groups; it includes the top six scientific fields by number of citations, the number of scientific fields with at least ten citations (over the whole period) and the scientific fields that have a Relative Specialisation Index (RSI) > 0.3 (those fields of *core specialisation* for which a given group has a share in the group's total citations which is at least the double of that of the field's aggregated share in the citations of the total population – see formula below). Finally, we define *distinctive scientific specialisation* as those fields among the top six for which the corporate group has a RSI > 0.3. On average, the groups have at least 10 citations across 18 scientific fields. Hoechst AG (43), Novartis (39) and Bayer AG (35) are active in more than 35 fields, while Teijin Ltd (3), Yamanouchi Pharmaceuticals Co. Ltd (4), Sankyo Co. (5) and Johnson & Johnson Inc. (5) have less than 6 fields with at least 10 citations over the whole period.

{TABLE 2 ABOUT HERE}

Although the top six fields across groups tend to be quite similar, some important differences appear both in terms of field concentration and field specialisation. A few corporate groups stand out in terms of their focus on certain scientific fields. For example, the top six scientific fields for Baxter International are biochemistry and molecular biology, cardiac and cardiovascular systems, haematology, immunology, medicine (research and experimental), and peripheral vascular diseases. Also, Baxter International has very strongly defined *distinctive scientific specialisations*, with only four fields with RSI > 0.3 all of which are in its top six fields. This specialisation profile seems to point to a specific market focus for Baxter's innovative activities. Although most groups do not show such a focused specialisation profile, nevertheless a few specific scientific fields, relatively important in their citation portfolios, can be detected. For example, Monsanto has a *distinctive scientific specialisation* in plant sciences and organic chemistry, these being among the top six fields with an RSI of respectively 0.77 and 0.36. Similarly, Zeneca, though with a large number of active scientific fields, has a clear specialisation profile with an important *distinctive scientific specialisation* in plant science (top six and RSI=0.8) and *core scientific specialisation* in agronomy and biology. At the other extreme, Bayer AG, Hoechst AG and Novartis, with large citations portfolios, show a more diversified pattern of scientific specialisation making it quite difficult to characterise in a clear-cut manner their scientific profile in term of specific competences.

From the results of this preliminary analysis, it is clear that the 30 groups rely, at least in part, on different knowledge bases (for example both Bayer AG and Novartis have very broad specialisation profiles yet they exhibit different *distinctive specialisations* in, respectively, polymer sciences and plant sciences). However, analysing the citations across the whole period, provides only limited information on the breadth of the knowledge bases of different corporate groups. In order to overcome this limitation, we focus on the evolution of the breadth of the knowledge bases over time.

To do this, we examined the citation portfolios of the groups in the two sub-periods: 1990-92 and 1995-97. We compared these two periods with the aim of examining trends in the evolution of the breadth of the knowledge bases. Two approaches were adopted. First, we looked at which fields each group was active in during both periods (persistency), and which fields each group cited at the end of the period only (entry), or at the beginning only (exit). Also, measures of similarity and concentration by sub-period were calculated for each group citation profile. Second, we calculated the Relative Specialisation Indexes for the two periods and examined their correlation across periods. This approach allowed us to characterise the breadth of the knowledge base of the groups both in term of the presence of citations in a specific field and of changes in scientific specialisation.

Table 3 presents the changes in the breadth of the knowledge bases of the corporate groups in terms of entry, exit and persistence in the citations to scientific fields. Table 3 presents the number of scientific fields in which each group was active in the second sub-period but not the first (entry: number of new fields in 1995-97);[4] the number of fields in which the group no longer has citations (exit: number of fields exiting); and finally the number of fields in which the group cites in both periods (number of persistent fields in the two sub-periods). To examine the changes in the breadth of the knowledge base of a group in terms of the number scientific fields in which it is active, we used a measure of similarity originally derived by Jaffe (1986). In our case the indicator (S_k) provides a measure of the scientific distance between the breadth of the specialisation profile of each group in the two sub-periods. The distance (i.e. variation in breadth) in scientific specialisation across time can be approximated by an un-centred correlation coefficient of the vectors (f_i and f_j) of citation share in each scientific field for each group in the two sub-periods.

$$S_k = \frac{f_i f_j^{'}}{\sqrt{(f_i f_i^{'})(f_j f_j^{'})}}$$

where $k = 1 \ldots 30$; corporate groups
and i = period 1990-1992; j = period 1995-1997

This similarity measure is bounded between 0 and 1, and the greater the degree of similarity between the breadth of the groups in the two periods the closer it is to unity.

---

[4] 'Active' indicates that a group has cited at least 10 times articles classified in a particular scientific field in a given sub-period. It is worth pointing out that this means that the number of active fields throughout the period is in general higher than the sum of the number of persistent, exiting, and new fields. There might be in fact fields with more than 10 citations between 1990-97 but those citations are uniformly spread across the whole period and they never reach the threshold level in any one particular sub-period.

{TABLE 3 ABOUT HERE}

In terms of similarity, most groups show a high level of similarity (average of 0.75). 21 groups have a similarity index above 0.5; of these 14 have an index above 0.8 and 7 above 0.9. However, four groups do not have any overlap among the fields cited at the beginning and at the end of the period (these are the same groups with small number of citations, that in Table 2 have less than 6 fields with at least 10 citations), and five groups have a similarity index lower than or equal to 0.5. The table shows a weak trend toward increasing the range of fields in groups' citation portfolios. Taking the arithmetic average of the differences between entry and exit, the result is an average 'net entry' for two fields (2.23). Adjusting this average by the number of fields in which groups are active throughout the period (i.e. calculating the ratio between net entry and total number of cited fields over the entire period, then taking the arithmetic average), one gets a positive increase in breadth (0.12).

Although not reported in the table, we measured the concentration of the scientific specialisation of the groups using the Herfindhal index. The index for each group was calculated for the two sub periods and for the whole period. Most groups show very low levels of concentration (< 1,000) of citations across scientific fields. Over time, no clear trend toward increasing or decreasing concentration emerges.

Given these results we proceeded to the examination of the changes in the specialisation profiles of the groups. To do this we calculated the symmetric Relative Specialisation Index (RSI) for all groups and for all fields for the initial period (1990-1992) and the final one (1995-1997).[5] The symmetric Relative Specialisation Index (RSI) is obtained standardising the activity index (AI), which is defined as the share of citations in a given scientific field in the citation portofolio of a given corporate group, relative to the share of citations in a given scientific field, for all corporate groups, in the overall sample of citations.[6] The RSI index indicates whether a firm has a higher-than-average activity in a scientific field (RSI > 0) or a lower-than-average activity (RSI < 0).

$$AI = \frac{\dfrac{p_{ij}}{\sum_i p_{ij}}}{\dfrac{\sum_j p_{ij}}{\sum_{ij} p_{ij}}} \qquad\qquad RSI = \frac{AI - 1}{AI + 1}$$

where p = number of cited publications, $i = 1....n$ = number of scientific fields = 132 and $j = 1...m$ = number of corporate groups = 30.

---

[5] Due to the problem of small numbers in cells, the analysis of the changes in the specialisation profile of corporate groups is based upon scientific fields with at least 10 citations in the considered periods at the group level.

[6] The two totals for the group and for the overall sample include citations in all scientific fields regardless of the specified threshold level of the 10 citations in the period under analysis.

The aim here is to determine whether we can see persistence in the structure of specialisation. We first focus on entry and exit into new scientific fields. Table 4 builds upon Table 3 focusing on the fields which groups have ceased or have begun citing, whose RSI is above 0.3. In other words, we look at changes in the groups' fields of *core specialisation*.

{TABLE 4 ABOUT HERE}

Again, cross-group heterogeneity emerges quite visibly. However, some trends are also observable. Of the 30 groups considered, 14 show no exit from the fields of *core specialisation* (i.e. those with an RSI above 0.3). Seven groups show the loss of only one field of *core specialisation*. Five groups (Bristol Myers Squibb, Glaxo Wellcome, Hoechst AG, Novartis and Sanofi-Synthelabo) exhibit a fairly big change in *core specialisation* with four or more fields exiting the core. Conversely, on the entry side, only five groups show no new entry in the areas of *core specialisation*, 8 groups have at least one new entry, and 7 groups exhibit at least four new fields with an RSI above 0.3. Bristol Myers Squibb, Hoechst AG and Novartis appear to be the most active groups in terms of both entry and exit. In some cases, we can distinguish a few clear patterns. For example, it seems clear that Abbott Laboratories is moving into the bio-pharma area: new fields of *core specialisation* include biochemistry research methods, biotechnology, genetics and heredity, microbiology, and virology. In other cases the patterns are less clear. For example, Bayer seems to be moving into traditional fields such as analytical chemistry, physical chemistry, and agronomy, but also optics and radiology. Overall then, pharmaceutical groups appear to be more 'active' on the entry than on the exit side. This is consistent with the argument that the emergence of new fields of useful research leads to an increase in the breadth of the knowledge base.

Table 4 focuses on the fields of specialisation that entered or exited the *core specialisation* of the groups studied. Table 5 focuses, instead, on changes in the fields that persistently appear as part of the specialisation profile of the group. To determine the relationships between specialisation structures during the period 1990-92 and the period 1995-97, we calculated the Pearson and Spearman correlation coefficients between RSIs in 1990-92 and 1995-97 by corporate groups. Correlation coefficients and their levels of significance are presented in Table 5.

{TABLE 5 ABOUT HERE}

First of all, only 15 corporate groups have at least 7 scientific fields of persistent specialisation –i.e. those fields in which the group was active in both sub-periods. For these 15 groups[7], the Pearson and Spearman coefficients deliver consistent results: the coefficients, while different in value, do not affect the ranking of groups. Of the 15 groups considered in Table 5, only eight show significant correlation across RSIs in the two sub-periods. Of the eight groups for which we have significant results, Bayer appears to be the one with the most stable specialisation profile (but this result should be read in conjunction

---

[7] The groups with less than 7 fields of persistent specialisation were not included in this analysis because the correlation coefficients tend to be less reliable with few observations.

with the data in Tables 3 and 4, which show the extent to which Bayer is entering new fields). Merck & Co. appears instead as the group characterised by the sharpest changes in specialisation profile. In particular, it can be seen from Table 4 that Merck has entered the fields of cellular biology and histology, biochemical research and virology. Unlike Merck, Abbot Laboratories appears to have added the bio-pharma fields highlighted above (see discussion of Table 4 above) to a more stable specialisation profile.

Overall the analysis of the knowledge breadth of the largest pharmaceutical groups reveals a high level of group-level heterogeneity in the field of active involvement and in the fields of specialisation when considering either the entire period or the changes from the starting to the end. Particularly interesting are the results of the analysis of change in the knowledge bases. Though groups have a high level of similarity in their portfolio of citations at the starting and end of the 1990s, we found some evidence of a weak but positive increase in their breadth, a significant entry in new fields of *core specialisation*, and a low level of persistent specialisation for the stable fields. The analysis allows one to highlight a few clear patterns at the level of the individual group. For example, while Abbott Laboratories is specialising in bio-pharma, Bayer seems to maintain, and actually reinforcing, its specialisation in traditional fields.

4.2 Depth of the knowledge base

Evaluating firm capabilities across different types of research and development activities gives an indicator of the depth of the capabilities maintained by an organisation in any given discipline. The intuition behind this analysis is that if a corporate group has a positive specialisation in a given scientific fields for each typology of research (e.g. it has a positive specialisation both in clinical observation/investigation and in basic research in the field of pharmacology and pharmacy) it has a knowledge base of greater depth in that scientific field compared to the other groups because its citation profile points to higher involvement in that discipline across typologies of research. In other words, the group is not only active in the more applied part of the research (clinical observation and investigation), but it also has competences in the area of basic research.

In order to analyse the integration of research across categories (depth of knowledge base), we calculated two indicators of integration (or depth) based on the symmetric RSI. The first indicator uses an RSI index calculated for each pair of scientific field and typology of research with at least 5 citations. The integration index is given by the ratio between the numbers of pairs of scientific fields and typology of research which show a positive specialisation in both CHI levels and the number of scientific fields of positive specialisation. So, for example, let us consider the case of 'neurosciences'. In order to be able to say that a company is 'integrated' in neurosciences (i.e. it has a 'deep' understanding of this discipline), we calculate the RSI for that company on the basis of the citations to neurosciences journals that focus on *basic research*. Then, we calculate the RSI in neurosciences for the same company on the basis of the citations to neurosciences journals that focus on *clinical observation and investigation*. If the company considered is relatively specialised (i.e. RSI > 0.3) in both types of research, within neurosciences, than we say that this company is 'integrated' in neurosciences, and we add 'one' to the numerator. The numerator of the indicator then says in how many disciplines a firm is relatively specialised in both categories of research. The denominator controls for the size of the groups' citation portfolios. This integration measure is derived using the formula below and it varies between 0 and 1. It is 0 when the group considered does not exhibit any overlap between the two types of research. It is 1 when the group considered is fully integrated across all types of research in all the fields in which it exhibits positive specialisation (Brusoni and Geuna, 2003).

$$x = \begin{cases} 1 \text{ if } RSI_{i,k} > 0 \quad \forall k = 1,2 \\ 0 \quad \text{otherwise} \end{cases}$$

$$y = \begin{cases} 1 \text{ if } RSI_i > 0 \\ 0 \quad \text{otherwise} \end{cases}$$

$$INT1_j = \frac{\sum_i x_i}{\sum_i y_i}$$

where

$i = 1...132$ number of fields

$j = 1...30$ number of corporate groups

$k = 1, 2$ number of research typologies

The second integration index provides a less restrictive measure of the depth of the knowledge base. Instead of requiring a positive specialisation for the pair of scientific field and typology of research, it accounts only for the fact that the corporate group is active (p_{ik} number of cited publications) in both levels of research in a given field (with at least 5 citations) with a positive specialisation. Let us use again the example of neurosciences. In order to consider a firm integrated in neurosciences, we first calculate the RSI for this filed. If the firm is relatively specialised in neurosciences (i.e. RSI > 0.3), we check whether this firm's patents cite journals focused on both types of research (i.e. basic research and clinical observation and investigation). If so, we consider this firms integrated in neurosciences. We do not require the firm to be relatively specialised in both types of research (within the field of neurosciences). We just require the firm to *cite* neurosciences publications that focus on both types of research. If the firm's patents cite journals in only one of the research categories, then it is not integrated in neurosciences. As shown in the following formula the denominator used is the one used in the INT1 measure. This indicator also varies from 0 to 1 (but it is just more likely to exhibit values closer to 1 than INT1).

$$x = \begin{cases} 1 \text{ if } p_{i,k} > 0 \text{ and } RSI_i > 0 \ \forall k = 1,2 \\ 0 \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

$$y = \begin{cases} 1 \text{ if } RSI_i > 0 \\ 0 \ \ \text{otherwise} \end{cases}$$

$$INT2_j = \frac{\sum_i x_i}{\sum_i y_i}$$

where

$i = 1...132$ number of fields

$j = 1...30$ number of corporate groups

$k = 1, 2$ number of research typologies

Table 6 presents the two integration indicators and the scientific fields in which each corporate group is integrated. Though INT2 is less restrictive than INT1 in 97% of the cases both indicators found the same scientific field to be integrated. INT2 has slightly higher values and it identifies more scientific fields as being integrated. The following analysis will be mainly based upon the indicator INT1. Only three corporate groups have INT1 ≥ 0.2, they are: Eli Lilly and Co., Rhone-Poulenc SA and SmithKline Beecham Plc. Abbott Laboratories and American Home Products also have high values. Eli Lilly and Co. (0.25) has a positive specialisation across research typologies in 5 fields: biotechnology and applied microbiology, medicinal chemistry, research and experimental medicine, neurosciences, pharmacology and pharmacy. Rhone-Poulenc SA is integrated in biochemistry and molecular biology, biotechnology and applied microbiology, genetics and heredity, immunology, microbiology and neurosciences. SmithKline Beecham Plc is integrated in biochemistry and molecular biology, microbiology, neurosciences, pharmacology and pharmacy. Abbott Laboratories exhibits a fairly clear profile in bio-pharmaceuticals. The integrated fields are biochemical research methods, biotechnology and applied microbiology, microbiology and virology. The extent of overlap with Table 4 is remarkable: the integrated fields are all fields in which this group was not active in the period 1990-92. Monsanto is integrated in plant sciences only, which reflects the strategic focus of this group over the period considered. Moreover, Hoechst AG and Bayer are integrated only in material sciences and multidisciplinary chemistry respectively, reflecting that these groups (in the period considered) were still behaving as traditional ‘chemical’ companies. Overall, the fields related to the new bio-pharma trajectory seem to be quite likely to be developed in an integrated manner. In other words, the data in Table 6 point to the fact that those groups that have entered on this research trajectory, have done so by quickly developing strong competencies across research typologies in the relevant fields. The data capture the delay of the German groups, and hints at the possibility that such delay is characteristic only of the German groups, rather than all the EU countries.

{TABLE 6 ABOUT HERE}

Finally, and from a methodological perspective, we should point out that the depth indicator employed in this section makes it possible to discriminate between groups in a more straightforward way than the breadth indicators calculated in Section 4.1. Specifically, the depth indicator is not biased by the size of the patenting activity of a group, and it provides a better proxy for the strategic research orientation of the group; for example, Bayer AG, Hoechst AG and Novartis have a very high number of patents and citations, but they have only very few integrated fields, with the German groups focused on traditional chemicals and the Swiss group integrated in the bio-tech, microbiology area.

4. Inventors' citations versus examiner's citations, a methodological test

In this section we investigate the robustness of the indicators of breadth and depth of groups' knowledge bases calculated above. We focus on the difference in the citation sources; as we pointed out before, the list of references to patent and non-patent literature is initially proposed by the applicant, but the final decision about which documents are cited in an application lies ultimately with the patent examiner, who might decide to accept those included by the applicant or add new references.

In this study we were able to discriminate between the citations listed by the examiner and those initially proposed by the applicant and accepted by the patent examiner, using the search report completed by the patent examiner during the screening of the technically relevant literature. The EPO's search reports contain various characters relating to each citation indicating the categories of the citation, which grade the cited document according to its relevance.[8] In addition, when the citation attached by the examiner is the same as that proposed by the applicant, the examiner will add the letter 'D' to the category of citation.

From our initial database we extracted all the citations to non-patent references proposed by the applicant and accepted by the examiner and a random sample of citations, for which the source is unknown (this sample reflects the data normally used in citation analysis). We were able to identify 2,843 citations to non-patent literature included by the applicant, which correspond to 4,436 observations with scientific field classification. The random sample contains 2,498 citations, which leads to 4,054 entries with scientific field classification. The two samples are also similar in terms of the number of patents: the sample containing applicant citations has 1,987 patents while the one with random citations has 2,114 patents. Using the two samples we can verify the validity of a non-patent references citations analysis that does not distinguish the source of the citations as we have done in the previous sections.[9]

---

[8] For example the category X citation indicates a document of particular relevance; the claimed invention cannot be considered novel and/or inventive in view of the citation.

[9] Due to the very small number of citations in the two samples of citations considered we could not develop a significant test for the integration measure.

Although not reported, we calculated the RSI index for each company and for each scientific field with at least 5 citations using data from both samples of citations over the period 1990-1997.[10] The two sets of RSI indexes have almost 60% of scientific fields in common, which implies that the portfolio of citations added by the applicant is to some extent different in its scientific focus from the random one. To test if the two sets of RSI values for the scientific fields in common are statistically different, we ran the following regression:

$$RSI_r = \alpha + \beta \; RSI_a + dummygroup + \varepsilon$$

where RSI_r is calculated using citations from the random sample, RSI_a is calculated using the applicant sample, and *dummygroup* are dummy variables for each corporate group to control for firms specificities.

The results of the regression show that there is strong positive and significant partial correlation between the two sets of RSI indexes (β equals to 0.87 significant at 1% level), once controlled for firms' specificities. We can therefore conclude that the common practice of using citations to non-patent literature without distinguishing their source can lead to some differences in terms of fields of specialisation, although for the common fields the value of the RSIs is not substantially different.

5. Conclusions

This paper has proposed indicators to capture the evolution of the knowledge base of the world's largest pharmaceutical groups by sales. The analysis of such evolution has focussed on two dimensions: *breadth* and *depth*. Overall the analysis of the knowledge breadth of the largest pharmaceutical firms reveals high group-level heterogeneity in terms of specialisation profiles. We found some evidence of a weak but positive increase in groups' breadth, a significant entry in new fields of *core specialisation*, and a low level of persistent specialisation for the stable fields. The analysis allows a few clear patterns at the level of individual groups to be highlighted. For example, while Abbott Laboratories is specialising in bio-pharma, Bayer seems to maintain, and actually be reinforcing, its specialisation in traditional fields.

Alongside traditional indicators of specialisation, we have proposed two new indicators of knowledge depth (or integration). We have argued that this additional dimension is fundamental to capture recent changes in the processes of knowledge production highlighted so far through case study research. In terms of depth, only three groups appear to be integrated in about a quarter of the fields in which they are positively specialised. Also, we found evidence of substantial group-level heterogeneity in terms of integration, with, in a few cases, the integration clearly reflecting the strategic focus of the group in question (e.g. Monsanto in plant sciences). Finally, it appears that those groups that have

---

[10] As in the first part of the analysis we use a threshold level of 5 citations per scientific field to avoid the problem of small numbers. In the formula for RSI we use the totals for the group and for the overall sample including citations in all scientific fields regardless of the specified threshold level.

entered the bio-pharma trajectory, seems to have done so by quickly developing strong competencies across research typologies in the relevant fields (e.g. Abbott Laboratories).

The empirical results of this paper seem to confirm the findings of recent research on the organisation of knowledge production in the firm, and raise a few more questions. First of all, it is important to observe the clear presence of group-level heterogeneity, in terms of both breadth and depth. We found some evidence of the fact that the results highlighted by the indicators of breadth and depth are consistent with what we know about group-level strategies and market focus (e.g. in the period studied, German firms were still specialised only in traditional chemical fields). Second, the indicator of depth seems to capture the increasing specialisation of both US and European firms (except the German groups) in fields related to bio-pharmaceuticals. The fact that these groups were developing highly integrated capabilities while at the same time entering a large number of alliances and joint R&D projects with specialised bio-technology firms, hints at the fact that these ‘knowledge integrated’ groups might play a co-ordinating role similar to that played by ‘systems integrators’ in other industries. This point seems to be consistent with the analysis of Orsenigo, Pammolli and Riccaboni (2001), which looked at the evolution of R&D networks in the bio-pharmaceutical area. This paper has provided a first look at complementary evidence focussing on one specific node of these networks.

This type of analysis needs to be pushed forward in a number of directions. Three would seem particularly promising. First, we need to study in more detail the relationship between the specialisation patterns identified (in terms of both breadth and depth) and the technology and product strategy of the groups. While some clear connection emerges (Abbot Laboratories entry in bio-pharma, Monsanto’s specialisation in plant sciences, the emphasis on traditional chemistry fields of the German groups) other groups do not show clear patterns. Second, the indicators developed here (and particularly that of integration) need to be probed against indicators of innovative performance. Can these indicators help to distinguish between innovative and non-innovative firms? Third, on a methodological note, as the use of patent citations increases, the reliability of such indicators must be questioned and, particularly in relation to what empirical phenomena they really capture. For example, the issue of identifying possible differences between the results achieved using the inventors’ original citations, as opposed to the citations as reported by the patent examiner need serious consideration in order to validate these, and other, indicators based upon the use of patent citations.

References

Acha, V. L. (2002) 'Framing the Past and Future: The Development and Deployment of Technological Capabilities in the Upstream Petroleum Industry', Unpublished Dphil Thesis, SPRU, University of Sussex at Brighton, UK.

Acha V. L. and S. Brusoni (2002) 'Knowledge on Wheels: Lessons from the tyre industry', paper prepared for the EAEPE Meeting, Aix-en-Provence, November.

Arora, A. and A. Gambardella (1994) 'The Changing Technology of Technical Change: General and abstract knowledge and the division of innovative labour', *Research Policy*, **23**: 523-532.

Baldwin C. and K. Clark (2000) *The Power of Modularity*, Cambridge MA: MIT Press.

Brusoni, S. (2001) 'The Division of Labour and the Division of Knowledge: The organisation of engineering design in the chemical industry." Unpublished Ph.D. thesis, SPRU, University of Sussex at Brighton, UK.

Brusoni, S. and A. Geuna (2003) 'The Key Characteristics of Sectoral Knowledge Bases: An international comparison', forthcoming in C. Antonelli, D. Foray, B. Hall and W. E. Steinmueller (eds), *Technical Choice, Innovation and Knowledge: Essays in Honour of Paul A. David*, Cheltenham: Edward Elgar.

Brusoni, S., A. Prencipe and K. Pavitt (2001) 'Knowledge Specialisation, Organisational Coupling, and the Boundaries of the Firm: Why do firms know more than they make?', A*dministrative Science Quarterly*, 46, 597-621.

Chesbrough, H., and K. Kusunoki (2001) 'The Modularity Trap: innovation, technology phase-shifts, and the resulting limits of virtual organizations', In I. Nonaka and D. Teece (eds), *Managing Industrial Knowledge: Creation, Transfer and Utilization,* Thousand Oaks, CA: Sage Publications, pp. 202-230.

Davies, A. (1997) 'The Life Cycle of a Complex Product System', *International Journal of Innovation Management*, 1(3): 229-256.

European Commission (2001) *Key Figures 2002 – Towards a European Research Area*, Bruxelles: DG Research.

European Commission (1997) *Second European Report on S&T Indicators* EUR 17639EN, Bruxelles: European Commission.

Filliatreau, G., F. Laville, E. Vachon, and S. Chapelle (2002) The construction of the NewKInd database, at http://www.researchineurope.org/newkind/Documentation/PDF/1v_D9Ann7_snt.pdf

Gambardella, A. (1995) *Science and Innovation in the US Pharmaceutical Industry*, Cambridge: Cambridge University Press.

Granstrand, O., P. Patel and K. Pavitt (1997) 'Multi-Technology Corporations: Why they have distributed rather than distinctive core competencies', *California ManagementReview,* 39: 8-25.

Howitt. P. (ed.) (1996). *The Implications of Knowledge-based Growth for Micro-economic Policies*, Calgary, Atla: The University of Calgary Press.

Jaffe, A. B. (1986) 'Technological Opportunity and Spillovers of R&D: Evidence from firms' patents, profits and market value', *American Economic Review* LXXVI: 984-1001.

Loasby B. J. (1999) *Knowledge, Institutions and Evolution in Economics*, London: Routledge.

Malo, S. and A. Geuna (2000) ‘Science-technology Linkages in an Emerging Research Platform: The case of combinatorial chemistry and biology’, *Scientometrics*, 47: 303-321.

Meyer, M. (2000) Does Science Push Technology? Patent citing scientific literature. *Research Policy*, 29(3): 409-434.

Michel, J. and B. Bettles (2001) ‘Patent Citation Analysis. A closer look at the basic input data from patent search reports’, *Scientometrics*, 51(1): 185–201.

Miller, R., M. Hobday, T. Leroux-Denvers and X. Olleros. (1995), ‘Innovation in Complex System Industries: the case of flight simulations’, *Industrial and Corporate Change*, **7** (2): 363-400.

Narin, F. and D. Olivastro, D. (1992) ‘Status Report: Linkage between technology and science, *Research Policy*, 21(3): 237-249.

Narin, F., K. S. Hamilton and D. Olivastro (1997) ‘The Increasing Linkage between US Technology and Public Science’, *Research Policy*, 26(3): 317-330.

Narin, F. and R. P. Rozek (1988) ‘Bibliometric Analysis of US Pharmaceutical Industry Research Performance’, *Research Policy*, 17(3): 139-154.

Nelson, R. R. (2003) ‘On the Uneven Evolution of Human Know-how’, forthcoming in *Research Policy*.

Nightingale P. (2000) ‘Economies of Scale in Experimentation: Knowledge and technology in pharmaceutical R&D’, *Industrial and Corporate Change*, 9(2): 315-359.

Orsenigo L. (1989) *The Emergence of Biotechnology*, London: Pinter.

Orsenigo, L., F. Pammolli, and M. Riccaboni (2001) ‘Technological Change and Network Dynamics: Lessons from the pharmaceutical industry’, *Research Policy*, 30(3): 485-508.

Patel, P. and K. Pavitt (1998) ‘National Systems of Innovation under Strain: the internationalisation of corporate R&D’. Brighton: SPRU, University of Sussex.

Pisano, G. P. (1990) ‘The R&D Boundaries of the Firm: An empirical analysis, *Administrative Science Quarterly*, 35 (1): 153-176.

Pisano, G. P. (1997) *The Development Factory: Unlocking the potential of process innovation,* Boston MA: Harvard Business School Press.

Prencipe, A. (1997) ‘Technological Competencies and Products Evolutionary Dynamics: A case study from the aero engine industry’, *Research Policy*, 25(8): 1261-1276.

Prencipe, A. (2000) ‘Breadth and Depth of Technological Capabilities in CoPS: The case of the aircraft engine control system’, *Research Policy*, 29: 895-911.

Schmoch, U. (1993) ‘Tracing the Knowledge Transfer from Science to Technology as Reflected in Patents?? Indicators. *Scientometrics*. 26(1): 193-211.

Takeishi A . (2002) ‘Knowledge Partitioning in the Interfirm Division of Labor: The case of automotive product development’, *Organization Science*, 13(3): 321-338.

Tijssen, R. J. W. (2002) ‘Science Dependence of Technologies: Evidence from inventions and their inventors’, *Research Policy*, 31(4): 509-526.

Trajtenberg, M., R. Henderson and A. Jaffe (1997) ‘University versus Corporate Patents: A window on the basicness of invention’, *Economics of Innovation and New Technology* 5: 19-50.

Ulrich, K. T. (1995) ‘The Role of Product Architecture in the Manufacturing Firm’, *Research Policy*, 24: 419-440.

Verspagen, B. (1999) ‘Large Firms and Knowledge Flows in the Dutch R&D System. A case study of Philips Electronics’, *Technology Analysis and Strategic Management.* 11: 211-233.

Verbeek, A., J. Callaer, P. Andries,K. Debackere,, M. Luwel and R. Veugelers (2002a) *Linking Science to Technology – Bibliographic references to patents. Vol 1: Science technology interplay: Policy relevant findings and interpretations*, EUR 20492/1, Bruxelles: European Commission.

Verbeek, A., K. Debackere, M. Luwel,P. Andries, E. Zimmermann and F. Deleus (2002b) Linking Science to Technology: Using bibliographic references in patents to built linkage schemes, *Scientometrics*. 54(3): 399-420.

von Hayek, F. (1989) ‘The Pretence of Knowledge’, *American Economic Review*, 79(6): 3-7.

Wang, Q. and G. N. von Tunzelmann (2000) ‘Complexity and the Functions of the Firm: Breadth and depth’, *Research Policy*, 29(7-8): 805-818.

Table 1: Patent applications and citations in the period 1990-1997, by corporate groups

Pharmaceuticals groups	Country	Patents	SCI-fields Citations
Abbott Laboratories Inc.	US	865	968
American Home Products Corporation	US	939	1661
Astra AB	Sweden	292	418
Baxter International Inc.	US	630	247
Bayer AG	Germany	5 534	1904
Bristol-Myers Squibb Co.	US	902	1335
Eli Lilly and Co Inc	US	944	1666
Glaxo Wellcome PLC	UK	353	1315
Hoechst AG	Germany	5 711	4112
Hoffmann-La Roche Inc.	Switzerland	990	1862
Johnson & Johnson Inc.	US	815	200
Kyowa Hakko Kogyo Co. Ltd	Japan	316	398
Merck & Co.	US	1 769	2483
Merck Patent GMBH	Germany	644	478
Monsanto	US	628	457
Novartis International AG	Switzerland	3 373	3329
Pfizer Inc.	US	954	940
Pharmacia	US	810	1578
Rhone-Poulenc SA	France	1 129	1535
Sankyo Co. Ltd	Japan	201	191
Sanofi-Synthelabo Inc.	France	465	754
Schering AG	Germany	589	682
Schering-Plough Corporation	US	400	653
Shionogi & Co. Ltd	Japan	212	302
SmithKline Beecham PLC	UK	862	1342
Takeda Chemical Industries Ltd	Japan	715	691
Teijin Ltd	Japan	353	137
Warner-Lambert Co. Inc.	US	515	549
Yamanouchi Pharmaceutical Co. Ltd	Japan	150	100
Zeneca Group PLC	UK	1 067	1062
Total		33,127	33,349

Table 2: Top six scientific fields and core scientific fields for 30 corporate groups.

Corporate Group	**Top 6 scientific fields**	**No. Fields**	**Core scientific specialisation, RSI>0.3**
Abbott Laboratories	CQ , DX, DY, NI, QA, RO	20	**QA**, CO, EA, JY, NN, PW, ZE
American Home Products	CQ, DR, DY, NI, RO, TU	26	**DR, NI**, AM, CU, FQ, MA, YA, YP, ZC, ZE
Astra AB	CQ, DX, DY, EE, NI, TU	10	NN, QU
Baxter International Inc.	CQ, DQ, MA, NI, QA, ZD	6	**DQ, MA, QA, ZD**
Bayer AG	CQ, DX, DY, EE, TU, UY	35	**EE, UY**, AM, DW, EA, EC, EI, II, IQ, IY, OA, PM, PW, SY, UB, UE, VY, YE
Bristol-Myers Squibb Co.	CQ, DX, DY, NI, RO, TU	23	DM, DQ, FQ, PY, VE, WE, ZD
Eli Lilly and Company	CQ, DX, DY, EE, RO, TU	22	CU, IA, PY, RU, SD, VE
Glaxo Wellcome PLC	CQ, DX, DY, NI, RO, TU	21	DM, NN, PY
Hoechst AG	CQ, DX, DY, EE, RO, TU	43	CU, EC, EI, II, IY, PK, PM, PW, UB, UK, UY, WE, YA, YP, ZD
Hoffmann-La Roche Inc.	CQ, DR, DY, EE, RO, TU	25	**RO**, CO, CU, FQ, MA,
Johnson & Johnson	CQ, DX, DY, EE, TU	5	**TU**
Kyowa Hakko Kogyo Co. Ltd	CQ, DX, DY, EE, QU, TU	11	**QU**, DB, DM
Merck & Co.	CQ, DX, DY, EE, RO, TU	27	IA, KI, TI, VY
Merck Patent GMBH	CQ, DB, DX, DY, EE, TU	9	**DB**, EA, FI
Monsanto	CQ, DE, DR, DY, EE, RO	12	**DE, EE**, AM, DW, UY
Novartis International AG	CQ, DE, DY, EE, RO, TU	39	**DE**, AM, DW, EA, EC, EI, II, IY, JY, KM, MU PM, UY, VE, YA, YP, ZD
Pfizer Inc.	CQ, DX, DY, EE, NI, TU	15	**DX, TU**, DX, NN, RT, TU, VE, ZC
Pharmacia & Upjohn Co.	CQ, DX, DY, EE, RO, TU	27	AD, CO, DW, EA, GA, IA, JY, PY, SA
Rhone-Poulenc	CQ, DY, EE, NI, RO, TU	27	DW, KM, MA, NN, RT, RU, ZD
Sankyo Co. Ltd	CQ, DX, DY, EE, TU	5	
Sanofi Synthelabo Inc.	CQ, DX, DY, EE, KM, TU	17	**KM**, DE, DQ, JY, RU
Schering AG	CQ, DX, DY, EE, NI, TU	15	IA, MA, PY, RT, RU, SD, VY
Schering-Plough Corporation	CQ, DY, EE, NI, RO, TU	14	**NI**, MA, QA, ZE
Shinogi & Co.	CQ, DA, DX, DY, EE, TU	9	**DA**
SmithKline Beecham PLC	CQ, DX, DY, QU, RO, TU	19	**QU**, KI, PY, RT, YO
Takeda Chemical Industries Ltd	CQ, DA, DX, DY, RO, TU	15	**DA**, DQ, IA
Teijin Ltd	CQ, MA, RO	3	**CQ, MA, RO**
Warner-Lambert Ltd	CQ, DX, DY, EE, RU, TU	8	**DX, RU, TU**, DQ
Yamanouchi Pharmaceutical Co. Ltd	CQ, DX, DY, TU	4	**DX, DY, TU**
Zeneca	CQ, DE, DX, DY, EE, TU	15	**DE**, AM, CU

See Appendix 1 for scientific fields codes.

No Fields: Number of scientific fields with at least 10 citations.

Bold Underlined are the fields of positive specialisation that are also within the top six fields, we consider these the fields of distinctive specialisation.

Table 3: Entry, exit and persistence in scientific fields, by corporate groups

Corporate Group	No. of active fields throughout period	Number of new fields in 1995-97	Number of fields exiting	Number of persistent fields in the two sub-periods	Similarity Measure
Abbott Laboratories	20	5	1	8	0.81
American Home Products	26	4	1	14	0.91
Astra AB	10	5	1	1	0.42
Baxter International Inc.	6	1	1	0	
Bayer AG	35	10	1	7	0.99
Bristol-Myers Squibb Co.	23	3	8	9	0.62
Eli Lilly and Company	22	8	2	9	0.83
Glaxo Wellcome PLC	21	2	4	8	0.83
Hoechst AG	43	5	7	21	0.91
Hoffmann-La Roche Inc.	25	4	2	13	0.82
Johnson & Johnson	5	4	0	0	
Kyowa Hakko Kogyo Co. Ltd	11	2	2	3	0.69
Merck & Co.	27	5	0	13	0.82
Merck Patent GMBH	9	1	1	5	0.87
Monsanto	12	2	0	4	0.42
Novartis International AG	39	11	5	18	0.94
Pfizer Inc.	15	5	0	5	0.94
Pharmacia & Upjohn Co.	27	4	3	12	0.98
Rhone-Poulenc SA	27	12	0	9	0.37
Sankyo Co. Ltd	5	0	2	1	0.50
Sanofi Synthelabo Inc.	17	2	5	5	0.75
Schering AG	15	4	1	5	0.96
Schering-Plough Corporation	14	3	0	7	0.77
Shinogi & Co.	9	2	0	0	0.41
SmithKline Beecham PLC	19	8	0	5	0.54
Takeda Chemical Industries Ltd	15	0	2	7	0.60
Teijin Ltd	3	0	1	0	
Warner-Lambert Ltd	8	2	0	3	0.88
Yamanouchi Pharmaceutical Co. Ltd	4	1	0	0	
Zeneca	15	2	0	8	0.89

No Active Fields: Number of scientific fields with at least 10 citations.

Table 4: Entry and exit in *core specialisation* fields, by corporate groups

Corporate Group	Fields exiting	New fields
Abbott Laboratories	PW	CO, DB, KM, QU ZE
American Home Products	IA	AM, ZE
Astra AB	NI	
Baxter International Inc.		MA
Bayer AG		AM, EA, EI, PW, SY, VY
Bristol-Myers Squibb Co.	DB, DQ, PY, QU, RU, VE, WE, ZD	DM, FQ
Eli Lilly and Co.	KM, QU	DM, IA, PY, RU, SD, VE
Glaxo Wellcome PLC	FQ, KM, NN	DM
Hoechst AG	II, UB, UK, ZD, ZE	CU, DW, EA, EI
Hoffmann-La Roche Inc.	CO	DM, MA
Johnson & Johnson		TU
Kyowa Hakko Kogyo Co. Ltd	NI, QU	
Merck & Co.		CO, FQ, ZE
Merck Patent GMBH	FI	DB
Monsanto		AM, DB
Novartis International AG	IA, PY, RT, RU, VE	CO, DQ, DW, EA, EI, II, MA, YP, ZD
Pfizer Inc.		DM, NN, RU
Pharmacia & Upjohn Co.	GA, QA, SA	EA, IA
Rhone-Poulenc		DB, DQ, FQ, MA, NN, QA, RU, ZD, ZE
Sankyo Co. Ltd	DX	
Sanofi Synthelabo Inc.	DB, DE, JY, KM, QU	QA, RU
Schering AG	RU	IA, SD, VY
Schering-Plough Corporation		QA, ZE
Shinogi & Co.		EE
SmithKline Beecham PLC		DA, KM, PY, QU
Takeda Chemical Industries Ltd	DR, KM	
Teijin Ltd		
Warner-Lambert Ltd		EE
Yamanouchi Pharmaceutical Co. Ltd		TU
Zeneca		DB

Only scientific fields with RSI>0.3

Table 5: Correlation between RSI-end of period on RSI-beginning of period, by corporate group.

Corporate Group	Pearson coefficient	Spearman rank coefficient
Abbott Laboratories Inc.	0.883**	0.810*
American Home Products	0.899**	0.745**
Bayer AG	0.972**	1.00**
Bristol-Myers Squibb Co.	0.513	0.517
Eli Lilly and Co.	0.042	0.183
Glaxo Wellcome PLC	0.657	0.548
Hoechst AG	0.621**	0.657**
Hoffmann-La Roche Inc.	0.557*	0.571*
Merck & Co.	0.575*	0.264*
Novartis International AG	0.652**	0.682**
Pharmacia & Upjohn Co.	0.897**	0.678*
Rhone-Poulenc	0.189	0.283
Schering-Plough Corporation	0.745	0.464
Takeda Chemical Industries Ltd	0.505	0.036
Zeneca	0.672	0.31

Only included corporate groups with at least 7 fields of persistent specialisation

Table 6: Integration indicators (INT1 & INT2) and code of fields integrated for each corporate group

Corporate Groups	INT1	Fields	INT2	Fields
Abbott Laboratories	0.18	CO, DB, QU, ZE	0.18	CO, DB, QU, ZE
American Home Products	0.16	CQ, IA, NI, QU	0.16	CQ, IA, NI, QU
Astra AB	0.09	QU	0.18	DY, QU
Baxter International Inc.	-		-	
Bayer AG	0.03	DY	0.03	DY
Bristol-Myers Squibb Co.	0.04	NI	0.08	NI, RU
Eli Lilly and Co.	0.25	DB, DX, QA, RU, TU	0.3	DB, DX, DY, QA, RU, TU
Glaxo Wellcome PLC	0.11	DB, QA	0.16	DY, QA, RU
Hoechst AG	0.03	PM	0.08	DY, IA, PM
Hoffmann-La Roche Inc.	0.08	CQ, IA	0.13	CQ, IA, KM
Johnson & Johnson	-		-	
Kyowa Hakko Kogyo Co. Ltd	0.09	DY	0.09	TU
Merck & Co.	0.2	CQ, IA, QU, TU	0.25	CO, IA, QU, RU, TU,
Merck Patent GMBH	0.12	DB, DY	0.12	DB, DY
Monsanto	0.11	DE	0.11	DE
Novartis International AG	0.05	DB, DE	0.11	CQ, DB, DE, QU
Pfizer Inc	0.13	DY, RU	0.13	DY, RU
Pharmacia & Upjohn Co.	0.08	CQ, DB	0.08	CQ, DB
Rhone-Poulenc SA	0.27	CQ, DB, KM, NI, QU, RU	0.27	CQ, DB, KM, NI, QU, RU
Sankyo Co. Ltd	0.14	DY	0.14	DY
Sanofi Synthelabo Inc.	0.05	RU	0.11	DY, RU
Schering AG	0.15	DY, IA	0.15	DY, IA
Schering–Plough Corporation	-		-	
Shinogi & Co.	-		0.1	DY
SmithKline Beecham PLC	0.23	CQ, QU, RU, TU	0.22	CQ, QU, RU, TU
Takeda Chemical Industries Ltd	-		0.05	DY
Teijin Ltd	-		-	
Warner-Lambert Ltd	0.1	DY	0.1	DY
Yamanouchi Pharmaceutical Co. Ltd	-		-	
Zeneca	0.07	DB	0.14	DB, Y

Appendix 1: Scientific cat codes

Code	Discipline
AA	acoustics
AC	automation and control systems
AD	agriculture, dairy and animal science
AH	agriculture, multidisciplinary
AI	engineering, aerospace
AM	agronomy
AQ	allergy
AY	anatomy and morphology
AZ	andrology
BA	anesthetology
CN	behavioral sciences
CO	biochemical research methods
CQ	biochemistry and molecular biology
CU	biology
CX	biology, miscellaneous
DA	biophysics
DB	biotech and applied microbiology
DE	plant sciences
DM	oncology
DQ	cardiac and cardiovascular systems
DR	cell biology
DS	critical care medicine
DT	thermodynamics
DW	chemistry, applied
DX	chemistry medicinal
DY	chemistry, multidiscinary
EA	chemistry, analytical
EC	chemistry, inorganic and nuclear
EE	chemistry, organic
EI	chemistry, physical
EP	computer science, artificial intelligence
ES	computer science, hardware and architecture
EV	computer science, interdisciplinary applications
EW	computer science, software, graphics, programming
EY	computer science, robotics
FF	emergency medicine
FI	crystallography
FQ	cellular biology and histology
FY	dentistry, oral surgery and medicine
GA	dermatology and venereal disease
GC	geochemistyr and geophysics
GM	substance abuse
GU	ecology
HB	education, scientific disciplines
HQ	electrochemistry
HY	developmental biology
IA	endocrinology and metabolism
ID	energy and fuels
IF	engineering, multidisciplinary
IG	engineering, biomedical
IH	engineering, environmental
II	engineering, chemical
IM	engineering, civil
IQ	engineering, electric and electronic
IY	entomology
JA	environmental sciences
JU	fisheries
JY	food science and technology
KA	forestry
KI	gastroenterology and hepatology
KM	genetics and heredity
KY	geology
LI	geriatrics and gerontology
MA	hematology
MU	horticulture
NE	public, environmental and occupational health
NI	immunology
NN	infectious disease
OA	instruments and instrumentation
OP	medicine,legal

PI	marine and freswhater biology
PJ	material science, paper and wood
PK	material science, ceramics
PM	material science, multidisciplinary
PO	mathematics, miscellaneous
PT	medical informatics
PU	mechanics
PW	medical laboratory technology
PY	medicine, general and internal
PZ	metallurgy and metallurgical engineering
QA	medicine, research and experimental
QB	medicine, miscellaneous
QE	materials science, biomaterials
QG	materials science, coatings and films
QJ	materials science, textile
QM	metallurgic and mine engineering
QQ	metereology and athmospheric sciences
QU	microbiology
RA	microscopy
RO	multidisciplinary sciences
RQ	mycology
RT	clinical neurology
RU	neurosciences
RX	neuroimaging
RY	nuclear science and technology
SA	nutrition and dietetics
SD	obstetrics and gynaecology
SU	ophthamology
SY	optics
TC	orthopedics
TI	parasitology
TM	pathology
TQ	pediatrics
TU	pharmacology and pharmacy
UB	physical, applied
UE	imaging science and photographic technology
UH	physics, atomic, molecular and chemical
UI	physics, multidisciplinary
UK	physics, condensed matter
UM	physiology
UY	polymer science
VE	psychiatry
VI	psychology
VY	radiology, nuclear medicine and medical imaging
WE	respiratory system
WF	reproductive biology
WH	rheumatology
XE	agriculture, soil science
XQ	spectroscopy
XW	sport sicences
XY	statistics and probability
YA	surgery
YE	telecommunications
YO	toxicology
YP	transplantation
YU	tropical medicine
ZA	urology and nephrology
ZC	veterinary sciences
ZD	peripheral vascular disease
ZE	virology
ZM	zoology
ZR	water resources