Helping Small Firms Map the Future Commercial Landscape – the Role of Patent Analysis

Mick McLean
Head of Economics and Public Policy
Scientific Generics

Introduction

Scientific Generics is a technology and management consulting company which in addition to normal client work also develops its own 'breakthrough product' concepts which are licensed to generate income. Intellectual property in general and patenting in particular is therefore a core business activity for the company. As a result the company has built up a wealth of experience in patenting and in IP management and analysis.

In our experience the patent system does not serve the smaller innovative company well. Our experience is supported by more systematic surveys[1] - recent research by the University of Sheffield Management School suggests both that the protection offered by the patent system is of little importance for innovation in small UK firms and that information from the patent system is of even less importance to them.

Many of the difficulties faced by the smaller innovative company stem from the ever increasing volume of patents. Figure 1 overleaf illustrates the growth in patenting in the last two decades as indicated by statistics from the two largest patent offices[2].

In 2000 some 826,572 new patents were filed, and although growth halted during the recent recession in high technology, the underlying growth trend has recently resumed and more than a million fresh patents are likely to be filed in 2006. The consequent size of the cumulative stock of valid patents is, of course, even more dramatic.

Faced by this complexity, it is only the very large organisations – those which can afford to invest in dedicated specialist resources to create and manage IP - which can 'play the system' effectively. In our experience it is hard for a small firm lacking in-house specialists to get a clear view of what is already protected and to find relevant information from patent databases.

Government policy should include support for providing smaller companies with access to patent analysis tools of the sort we describe below, to enable UK inventors better to extract valuable information from patent databases and maximise patent protection (and therefore the value of IP) by helping them identify and target 'white space' for innovation in the competitive patent landscape.

---

[1] See for example - Macdonald, S. (2004) 'When Means Become Ends: Considering the Impact of Patent Strategy on Innovation', Information Economics and Policy. vol 16, no 1, pp.135-58.

[2] 'Using Patent Indicators to Understand Innovation and Growth', Dominique Guellec, OECD/DSTI, WIPO Conference on Patents Geneva, September 17, 2003

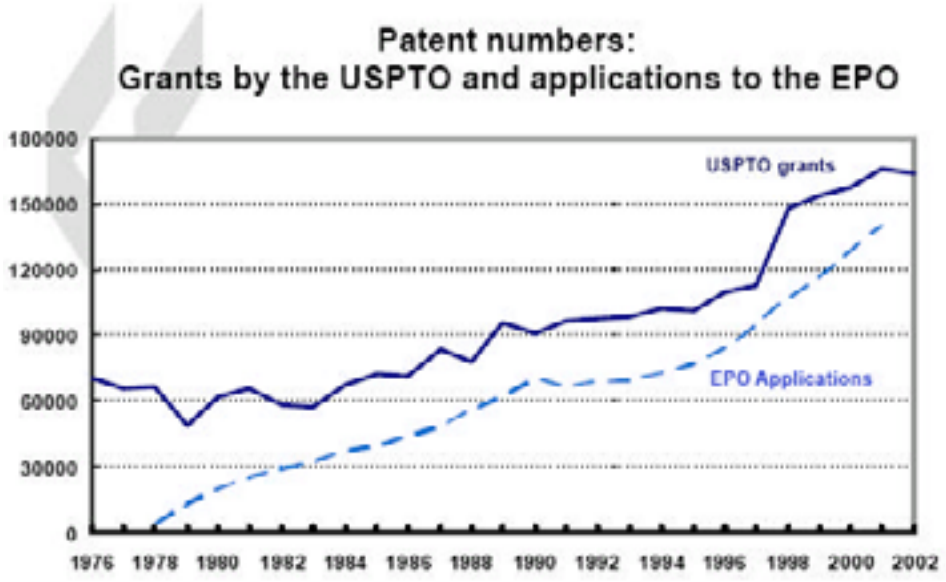


Figure 1 – Recent growth in patent grants and applications

Failure to understand the competitive IP landscape can often lead firms to take out patents which appear sound and will pass examination but which in practice prove weak because of their similarity to and overlap with a plethora of competing patents.

While patents are potentially a rich source of technological data their sheer abundance can make this valuable information hard to extract. We have found that it is most productive for inventors to read ten to twenty key patents relevant to their inventions before filing patents of their own – but how can they find the key patents and be sure they have made the right selection?

Our response to these challenges has been to develop our own analysis tools to help us cope with the growing scale and complexity of the body of granted patents - which we believe is a major contributor to the rather sad position revealed by the Sheffield surveys.

While the main motivation for developing these tools was our need to analyse large numbers of patents to clarify 'freedom of action' questions and identify white space for innovation for ourselves and our clients , we had come to realise that patents are an excellent leading indicator of commercial activity. Patent analysis has therefore also prove useful for forecasting - by looking at current patterns of patenting we are able to gain a view of the products and technologies which will be in common use in 10 to 15 years time.

The analysis tools, which we describe below, help extract valuable information from patent databases and maximise patent protection (and therefore the value of IP) by helping us identify rival patents and target 'white space' for innovation in the increasingly crowded patent landscape.

We always intended to recover the costs of developing these tools. Our initial plan was to continue to develop them for in-house use for our own inventors and for our clients but to invest in making mature versions more 'user friendly' and licensing them for use by selected third parties. The drawback of this approach is that we are not well organised to produce and support third-party software.

Last summer we had some discussions with the Search & Advisory Service of the UK Patent Office. These discussions convinced us that a viable alternative to limited third party licensing of our tools would be to work with UKPO on a commercial basis to make versions of them universally available to UK inventors. But, while the UK Patent Office appreciates the potential value of the provision of such tools as a stimulus to innovative activity, there are many practical difficulties to be surmounted and we believe this course of action would require policy intervention and governmental backing. It is hoped that this submission will make the case for such support.

Background

Scientific Generics has built up a lot of experience in patenting (over the last decade the company has filed over 1,700 patents and patent applications) and in managing its IP. Even with careful management of our IP portfolio the maintenance of our current collection of about 50 patent families costs the company in the order of £250,000 annually. In the process of attempting to maximise the value of its patenting and minimise the costs the company has acquired deep knowledge in the exploitation and use of commercially available IP search and analysis tools - Delphion and Derwent, for example - across a broad range of research sectors. The company has also developed proprietary tools to help manage its patent portfolio and a 'second generation' knowledge mapping toolkit for the analysis of competitive IP 'landscapes' – it is the latter which we describe in this submission.

The most common use of such patent analysis tools in support of innovation is in gauging the strength of actual or potential IP in relation to the competitive 'landscape'. But perhaps the most direct way in which knowledge mapping can stimulate innovation is the through the identification of 'white space' for innovation within an already crowded IP space. In our experience, this use of patent mapping as an innovation heuristic is greatly enhanced by the Cartesian framework of well defined dimensions provided by our IQIP toolkit (see section on 'Clustering' below).

Generics has recently - in addition to work for private sector clients which, by its nature, must remain confidential - completed a number of assignments whose results are in the public domain. These have involved using patent statistics to reveal trends and key players in particular technologies and markets. The results of these studies illustrate some ways these kinds of tools can be used.

While there is still considerable academic debate about what exactly patent statistics measure, there is an almost unanimous agreement about the importance of these data. Patent data – on granted patents and published applications - are considered nowadays as a unique, broadly available and reliable source of statistical material. Patents are accepted as a useful leading indicator of economic activity – since successful inventions emerge from patents over a long time scale with the delay depending strongly on the sector. Patents are a topical indicator of levels of R&D effort - being one of the principal outputs of such activity - and patent data are available with a delay of at most 18 months.

Furthermore patents have been classified into detailed technological categories that facilitate aggregation of the data; each patent grant identifies the owner, inventor, geographic origin, technology, issue date, application date, and other information associated with the patent to assist in describing and characterising the technological activity associated with that patent. Information on patents is available in electronic format for all patents granted and published applications in recent years - which enables the sort of computer-aided quantitative research studies we describe.

On a micro level knowledge of the competitive IP 'landscape' can stimulate innovation through the identification of 'white space' within an crowded IP market and technology space. On a macrolevel patterns of patenting can reveal broad changes in R&D activities over a long time-span[3], covering a large sample of firms and sectors.

Case study - NMSD

For the National Measurement System Directorate of the UK Department of Trade and industry (DTI) in 2004 we constructed 'clean' worldwide patent portfolios relating to flow measurement and to 26 'emerging technology' areas (which had been targeted by the DTI) for the previous five years. The results of these exercises were presented in poster form for use with stakeholder consultation exercises (a thumbnail of one of the posters is shown as Figure 2 overleaf).

For each portfolio of several thousands of patents, Generics:

- identified international profiles of patenting;
- analysed time trends;
- identified UK key players (academic and commercial);
- calculated metrics of importance of 'measurement';
- informally identified possible areas of NMS relevance;
- constructed detailed 'maps' of each area highlighting emerging technologies and applications using the Aurigin package;
- documented trends and relevant measurement issues.

[3] A recent OECD report – 'Forum on the Internationalisation of R&D, Egmont Palace, Brussels, Belgium, 29-30th March 2005. Background Report' - describes, for example, how the internationalisation of R&D by firms has been analysed using patent data. This report discusses the internationalisation of technology creation as measured through international (co)patenting and (co)inventing – it pays particular attention to the behaviour of multinational firms (MNEs) as major driving forces in R&D-internationalisation.

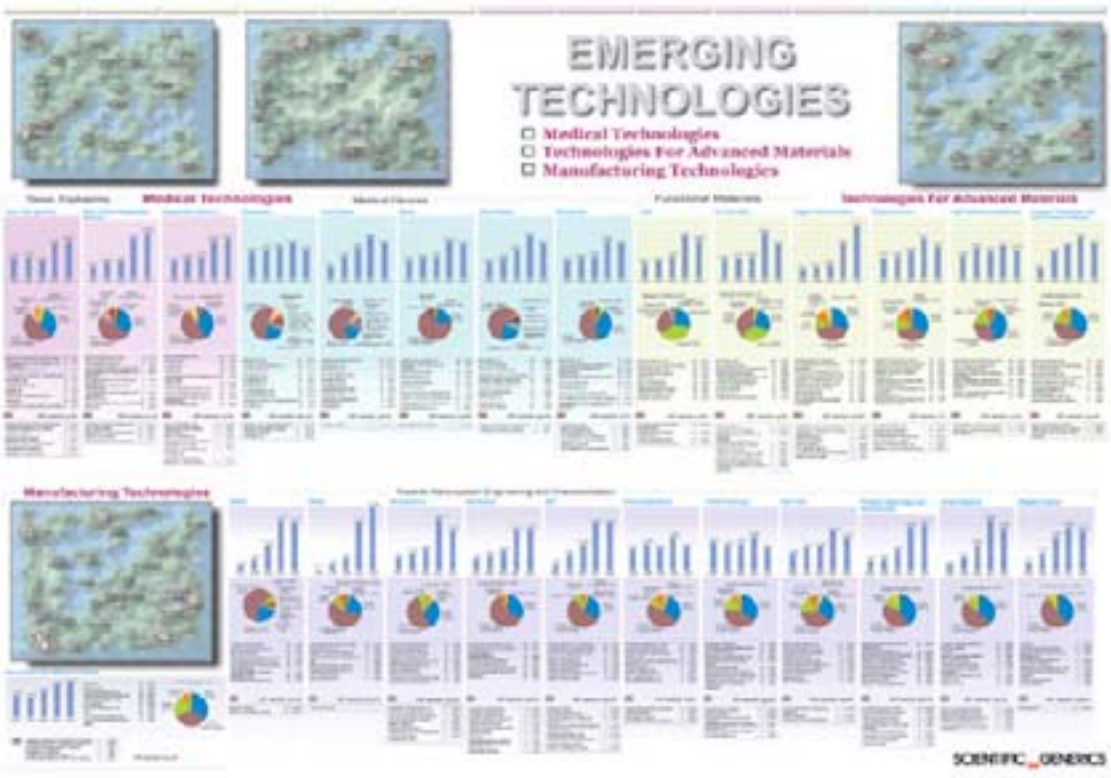


Figure 2 - Poster of IP analysis for 26 'Emerging Technology' clusters

A particular focus of the analysis was the extent to which 'measurement technologies' were associated with each area and sub-theme. We defined an indicator of the relevance of measurement in each market/technology sub-set using the International Patent Classification (IPC) for each individual patent. This 'IPC metric' was worked out for a given technology area by calculating the percentage of patents for which IPC codes relating to measurement codes (G01 Measuring, Testing) had been assigned by the patent examiner.

The resulting patent analysis was used:

- to confirm and reinforce the choice of broad technical areas by revealing UK patent activity as an indicator of future measurement infrastructure needs;
- to give guidance on the most fruitful area for research within these still quite broad fields, where:
    - the UK was relatively strong;
    - the need for standardisation/measurement infrastructure was most evident.

In addition the analysis revealed detailed information about geographical differentiation in innovation, and the names of the key commercial players and applied academic researchers in the UK and globally.

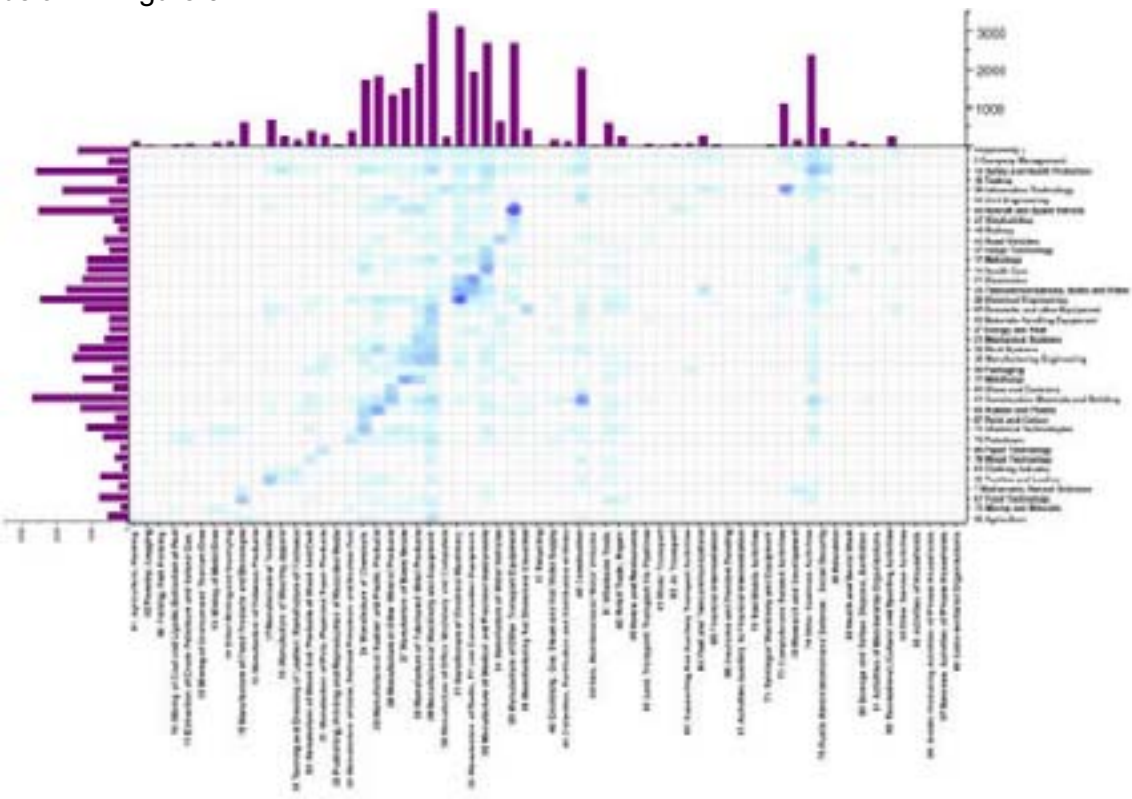
Case study – British Standards

In 2005 we mapped the contents of the database of then current British Standards – which contained some 35,000 documents. We also analysed the 13,000 or so standards at that time under development by working committees.

The results of the analysis proved useful for the body which administers the standards system in the UK - BSi – providing valuable 'market intelligence' about changing patterns of standards, the organisation and structuring of standards committees and the evolution of international $_{vs}$ national standards within sectors.

An important part of the exercise was the mapping of the standards to the user base using the Standard Industrial Classification system using a concordance as described later in this paper. The SIC standards maps were compared with other SIC-mapped data, for example, economic activity (gross value-added) by sector and UK trade by sector. The latter enabling a useful comparison with the pace of adoption of international standards in various sectors of the UK economy.

An example of the standards mapping output showing the correspondence between SIC codes and the International Classification for Standards[4] (ICS) coding is shown below in Figure 3.



[4] The International Classification for Standards (ICS) is managed by the International Organization for Standardization (ISO) to unify the classification of data about standards throughout the world.

Figure 3 - Correspondence between SIC and ICS codes

Elements of the analysis toolkit

As the name implies, we have developed the IQIP toolkit primarily for patent analysis and mapping but, as we have already pointed out, the tools can be, and have been, used with many other types of data (standards and scientific publications for example).

Key elements include:

- assistance for expert insight and judgement in building portfolios for analysis:
- use of forward and backward citation of documents;
- 'many-to-many' mappings of commonly used coding systems;
- semantic analysis:
- help with definition of categorisation systems (taxonomies)
- automatic or semi-automatic allocation of documents to predefined taxonomy
- clustering algorithms to reveal data structures
- graphical display options with fixed user-defined dimensions
- animated displays to highlight changes and exceptions

We describe the original features of some of these elements in more detail below.

Semantic analysis

We are building on the work of UK research groups which have long been active in the area of semantic analysis[5]. Advanced semantic analysis goes beyond 'frequency counting' to use contextual information to build classification systems based on textual analysis.

We are pioneering the use of this approach on the 'claims' and 'abstracts' of patents to complement the classification systems provided by patent codes (IPC and Derwent) but we believe the approach is equally applicable to other document types such as scientific papers and standards

Genetic algorithms

We are working on a document classification system that is based on the use of genetic programming techniques. The system uses a small set of example training documents as the basis for creation of a fitness test which is then used as the cost function to optimise a genetic algorithm search to discover rules that optimise classification performance against a sample data set. This mode of operation is quite distinct from that of vector models based on word frequency analysis in a number of ways:

[5] See, for example, 'Text classification using string kernels', Lodhi, Saunders Shawe-Taylor, Cristianini and Watkins, Department of Computer Science, Royal Holloway, University of London The Journal of Machine Learning Research Vol. 2, March 2002

- a vector model will typically generate a classifier profile that is effectively a keyword weight vectoring;
- a genetic algorithm approach characterises documents based on the occurrences of character sequences called 'N-grams';
- an N-gram representation structure is recursive and supports sequences of N-grams such as a word sequence;
- reliance on sequences is not word bounded and therefore word stems are automatically discovered.

The search space of N-grams is vast, so the system uses a genetic search algorithm based on initial discovery of simple N-grams followed by random combination - i.e. breeding - to enable discovery of combinations of N-grams that provide better classifiers (searches will run over a number of iterations until an optimal state has been located). The system generates rules which resonate with the kind of rule a human would use for fast pattern based recognition of documents.

Code mapping

A general knowledge mapping problem involves a 'many to many' mapping form one system of classification to another. We have developed a range of translation tools:

For example it has long been recognised that it would be useful to be able to map patterns of patent activity to the sectors of the economy in which the patent is intended to be exploited. The mapping of specific patent portfolios to such classifications as the SIC have been undertaken. These have been highly labour and skill intensive – involving the use of large teams of researchers reading and classifying individual patents.

There are, however, more automatic translation mechanisms – the best known being a concordance provided by the US patent office [6] between the U.S. Patent Classification System (USPC) and the US Standard Industrial Classification System (SIC). This was developed in 1974 with the support of the National Science Foundation (NSF) to attempt to bridge the gap between the industry and patent based classification systems associated, respectively, with economic and patent data.

This USPC to SIC Concordance is maintained by USPTO and is updated on a regular basis, generally annually, to accommodate any changes and revisions that are made to the USPC.

.

There is also a concordance between International Patent Codes (IPC) and the US version of the SIC – this was generated largely automatically and empirically by Silverman[7]. He relied on the fact that for a period between 1990 and 1993 some

[6] See: 'Revisiting the USPTO Concordance Between the U.S. Patent Classification and the Standard Industrial Classification Systems', Jim Hirabayashi, U.S. Patent and Trademark Office, WIPO-OECD Workshop on Statistics in the Patent Field, 18-19 September 2003, Geneva, Switzerland

[7] Brian S. Silverman, "Documentation for International Patent Classification - U.S. SIC

Canadian patents were both coded using IPC codes and assigned to (Canadian) SICs. Each patent in the 1990-1993 population has a primary SIC of 'Use' and a primary SIC of 'Manufacture' that was assigned by the Canadian patent examiners. For each IPC Silverman constructed a frequency distribution reflecting the proportion of patents assigned to IPC that were assigned to a Canadian four-digit SIC class.

The next step was to link these Canadian SICs to the US SICs. Fortunately there is a Canada SIC – U.S. SIC concordance published jointly by Industry Canada and the US Department of Commerce. Unfortunately many of the Canadian SICs map into multiple US SICs and so Silverman had to use somewhat arbitrary assignments to the US codes.

We produced a similar concordance for translating International Patent Codes (IPC) to the UK version of industrial codes (UK SIC) using a similar approach to that used by the USPTO. One methodological difference is that we used a 'voting' algorithm to weight the contribution of all the IPC codes cited. Most weight was assigned to the SIC codes signalled simultaneously by the main IPC code and other IPC codes of a particular patent.

In Figure 4 overleaf we show an example of a patent portfolio mapped to SIC sectors using our concordance. This portfolio represents all patents filed in the field of materials from 1995 to the present day. Only patents filed in either the US or the European Patent Office or under PCT were considered as relevant – since in our experience patents taken only to the national stage (GB patents, for example) do not usually represent significant inventions.

These materials patents were downloaded from the Derwent database. The portfolio of ~150,000 patents was built up using our search aid tools and processed to avoid double counting of patents from the same patent family.

In this example we distinguish and highlight three groups: patents originating from academia, from 'active' patenting commercial organisations and from small enterprises ('patent active' organisations are defined as those with more than two international patent applications per year).

The colour coding shows eight bands – each representing 12.5% of the range of values. The colours run from black (lowest) through the spectrum to red with red representing the highest numbers. The figures in each cell (A_{ij}) are doubly normalised on both axes (m,n) according to the formula set out below so that exceptionally low and high levels of activity stand out – we call this a HotSpot© plot.

concordance," web page at
http://www.rotman.utoronto.ca/~silverman/ipcsic/documentation_IPCSIC_concordance.htm

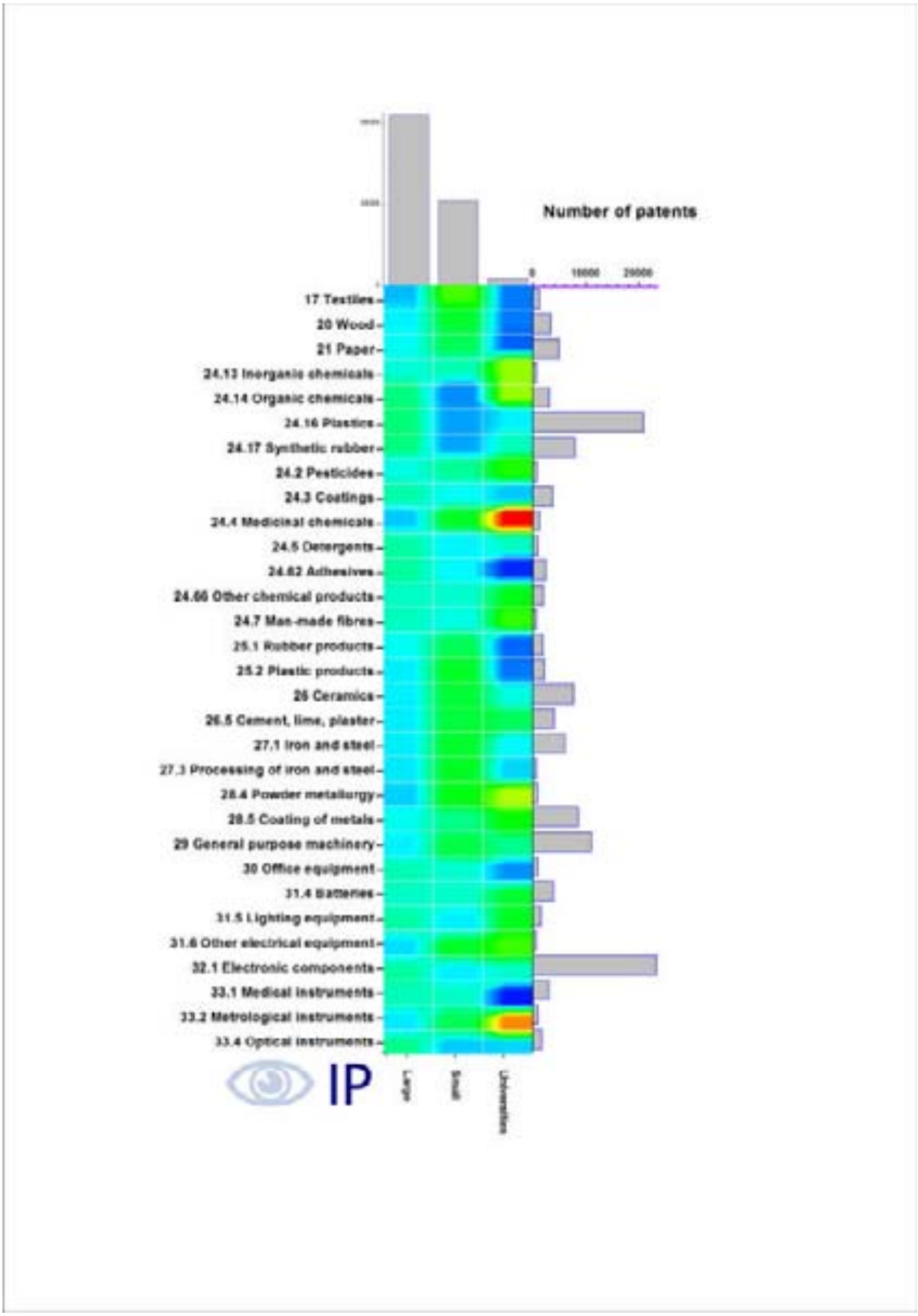


Figure 4 - Materials patents worldwide 1995-present

Relevance weighting

To ensure that the patent portfolio we identify properly represents the field of study, we first map the general concepts that make up the subject onto a hierarchical set of categories. We subsequently associate the elements of this hierarchy with one or more patent classification codes, and this assembly of codes is used to extract patents relevant to the enquiry from the public databases (principally Delphion) using searches.

We use an inclusive searching method based on patent classification codes to ensure that all relevant patents are considered. Although this method runs the risk of recovering irrelevant patents in addition, as well as patents that have been incorrectly coded, assignment of relevancy scores (see below) helps exclude these from further consideration.

Patents, like scientific papers, typically reference prior patents of particular relevance. Thus we use a method similar to that adopted for the bibliometric citation analysis of scientific publications to weight patents by an impact measure dependent on the extent to which they are cited by subsequent patents.

Our impact measure adjusts the raw citation counts to allow for the field of the invention and the age of the patent. Citations per patent vary between fields and will clearly be fewer for patents filed in more recent years than for older patents. Impact data is rebased against the world average for similar data, analysed by field and year, to provide an informative reference point for comparison between years and, with caution, between fields of invention.

It is interesting to note that the algorithm[8] which lies behind the highly successful search engine Google is also based on an analogous impact concept derived from bibliometrics. PageRank is the algorithm used by the Google search engine, originally formulated by Sergey Brin and Larry Page[9]. It was inspired by the observation that the importance of a research paper is often judged by the number of citations the paper has from other research papers. Brin and Page simply transferred this principle to its web equivalent: the importance of a web page can be judged by the number of hyperlinks pointing to it from other web pages.

Clustering

Perhaps the most direct way in which patent mapping can be used to stimulate innovation is through the identification of 'white space' for innovation within an already crowded IP space. So called 'clustering' techniques can be used to highlight such white space in patent landscapes and the analysis of this can be employed as an innovation heuristic and as a guide to possible merger and acquisition behaviour.

---

[8] See: http://www.webeventseurope.com/Common/StandModules/StandArticleDisplay.aspx?intStandID=311&intArticleID=4041

[9] The Anatomy of a Large-Scale Hypertextual Web Search Engine, L Page, S Brin - Proceedings of the 7th World-Wide Web Conference, 1998

Essentially mapping techniques like IQIP produce data in the form of sets of two dimensional matrices. Clustering relies on the fact that the ordering of the variables along each axis of the matrix is largely arbitrary. So unless it is important to preserve a particular variable order for comparative purposes (say in a coding system) then it is possible to reorder the matrix - rearrange the rows and columns - to produce a pattern of data that resembles a block structure. Our clustering technique which reduces a statistic which reflects the weighted spread of elements and we use an algorithm which is both practical in operation and guarantees convergence on a minimal bandwidth.

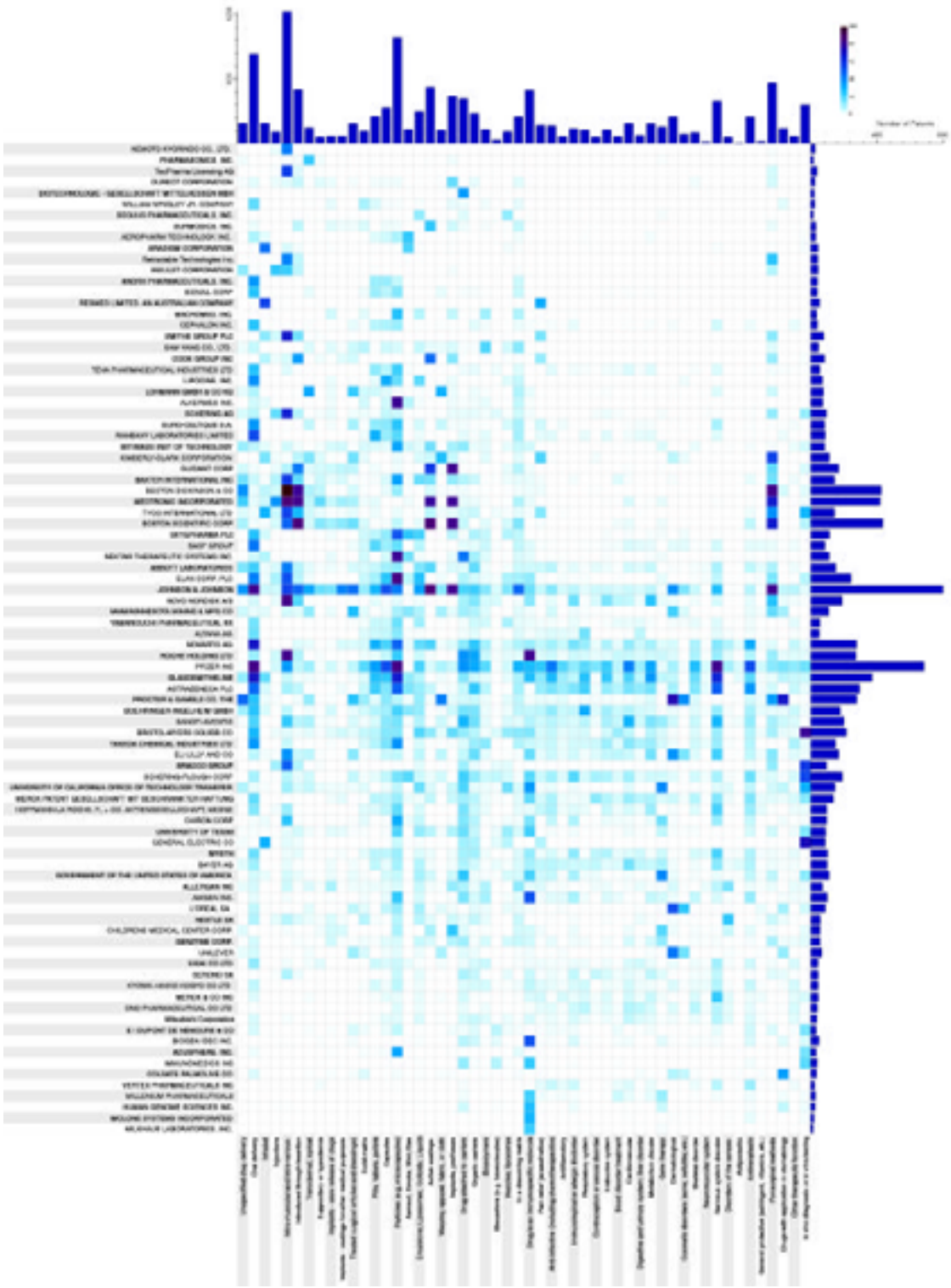
The two examples overleaf illustrate the power of the clustering technique. Both pictures present identical data – in this case a macro-level analysis of key players in patents related to drug delivery over the last ten years.

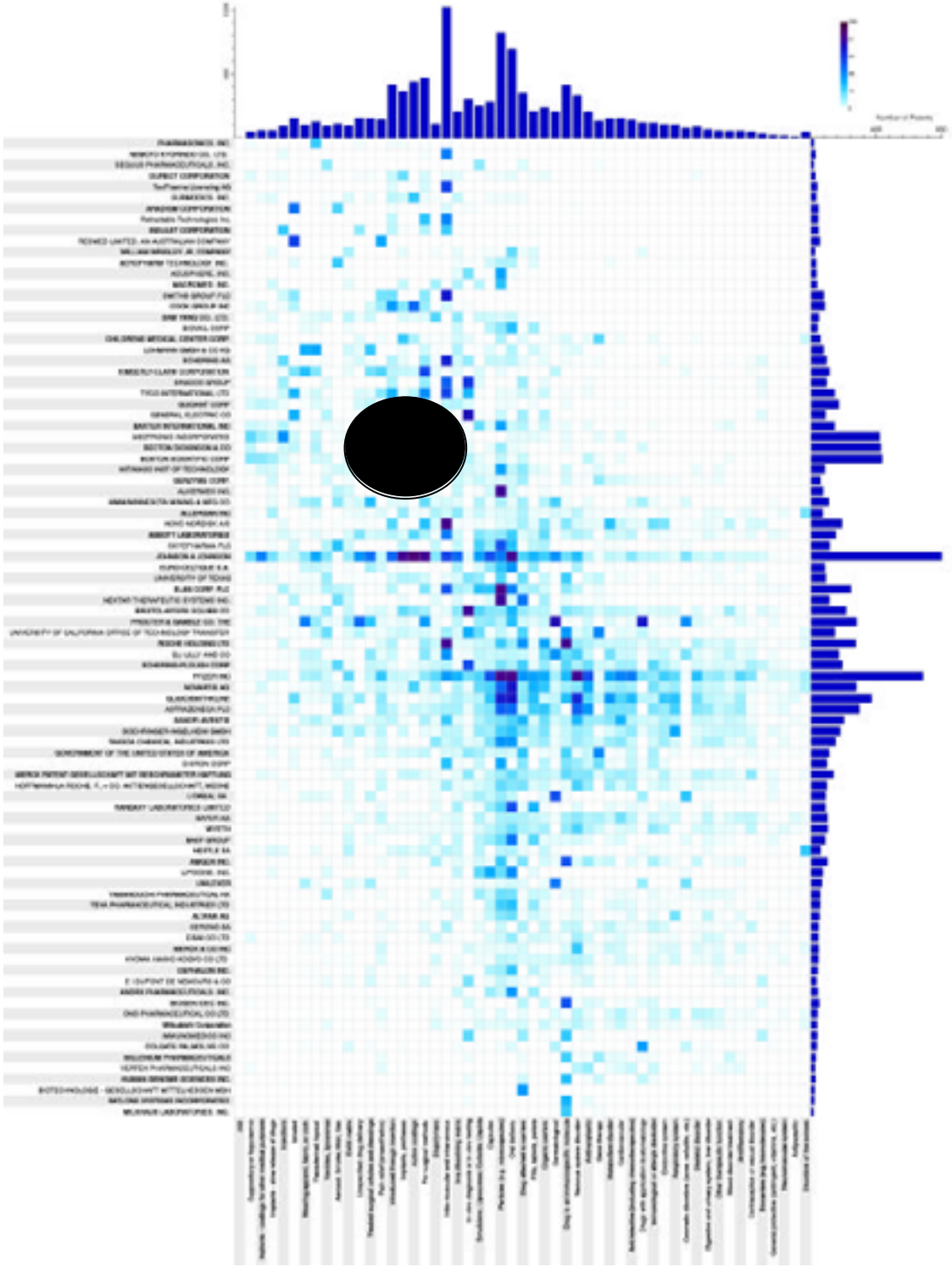
The first picture plots the 50 key players – the global organisations most active in patenting in this field (vertical axis) – against a set of technology categories (covering the delivery method, form of the drug and its therapeutic application)[10]. In this left-hand picture the ordering of technologies is fixed and arbitrary and only the players are clustered (according to the similarity of the composition of their patent portfolios).

In the second picture both axes are reordered using the algorithm described above. One feature of interest is highlighted – a tight cluster of players and inter-related technologies with 'white space' at the centre. Subsequent to this analysis (undertaken in December 2005 for the UK Department of Trade and Industry) one of the firms highlighted was subject to a takeover bid from another. We believe that this demonstrates dramatically the potential of this kind of analysis as a predictor of significant commercial activity.

In our experience clustering of micro-level data (very detailed information about a patent portfolio of several hundred patents) can also reveal fruitful white space for innovations – but the examples we have generated have been exclusively for our private sector clients and are therefore confidential.

[10] It is interesting to note that the pattern of patenting revealed by this analysis of these large corporations differs considerably from that shown by a similar analysis of smaller companies' patenting behaviour.





Case study – Stem cell patenting

The use of IP mapping on a relatively specialist field of interest but with international scope - is demonstrated by a recent IP mapping project on stem cells we undertook for the UK Department of Health. This represents a typical example of the work we undertake for private clients –and therefore a convenient illustration of the sort of results we obtain which are usually client confidential.

An expert body - the Stem Cell Initiative (UKSCI) - recommended[11] that Government spending on the research should be increased by between £350 million and £520 million over the next decade. Strong support for these recommendations was provided by our analysis of stem cell patenting.

Using the latest version of the knowledge mapping tool, IQIP, described above we identified relevant stem cell patents filed anywhere in the world since 1993. This analysis was designed to identify:

- prominent UK and global organisations in stem cell research, as judged by the citation levels of their patents in other patents;
- patents with significant impact on stem cell research, as judged by the by the citation levels of those patients in other patents;
- international trends in patenting activity over the last decade.

Our analysis identified sixteen key stem cell patents - the most influential when ranked by their number of citations in other patents. Twelve of these belonged to organisations from the USA, three to organisations from the UK and one to a Canadian organisation.

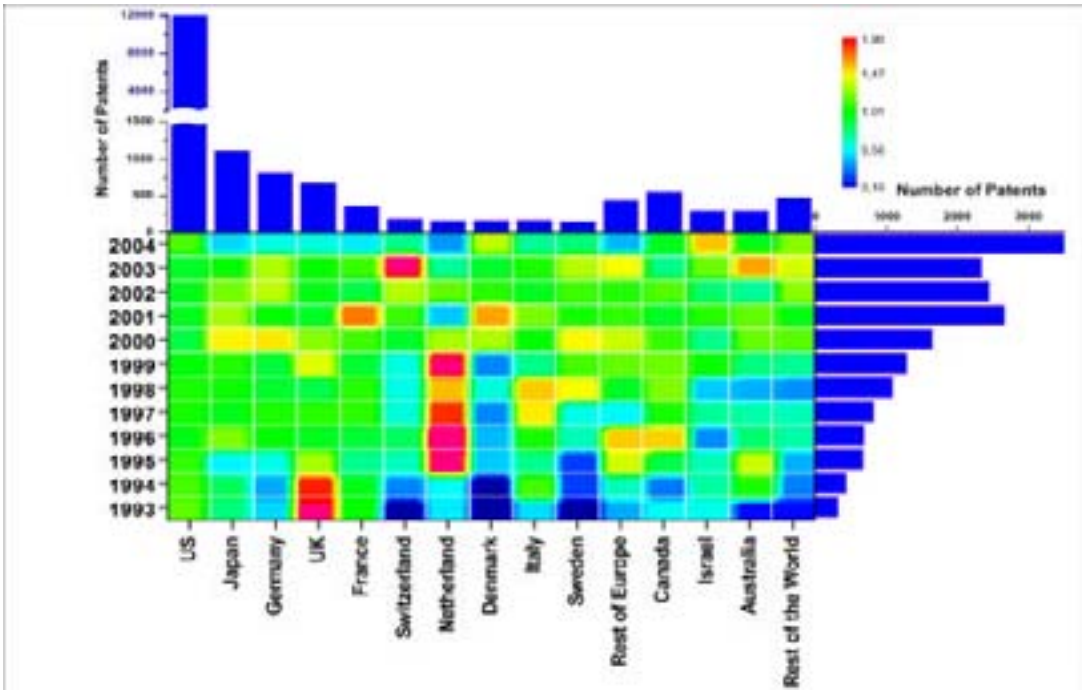
Lastly, we examined international trends in stem cell patenting activity over the last decade. As with other areas of technology, the US has been consistently dominant in this area.

Our analysis showed that UK inventors had a reasonable share, 3.8%, of the total 17,800 stem cell patents filed since 1994. However, since 2004 there had been a significant upturn in the level of patenting activities in the US. In addition, several other countries are patenting in the stem cell field more aggressively than before, including Israel, Australia, Canada, Denmark, Switzerland, Italy and Sweden.

These results indicated that while the rest of the world had increased activity in this strategic field the UK had not. These findings helped persuade the UK Government to follow the recommendations of the report and increase stem cell research funding substantially.

The graph below illustrates the recent trends in international activity in stem cell patenting we identified.

[11] UK Stem Cell Initiative Report & Recommendations, November 2005. The full report can be seen at http://www.advisorybodies.doh.gov.uk/uksci/uksci-reportnov05.pdf.



In this graph the relative proportion of patenting activity, ranging from 0.1-1.9, was calculated for each country over each of the years 1993-2004. In the Hotspot© diagram colour coding as usual illustrates this ratio of patenting activity using 1,000 bands, each representing 0.1% of the range of values.

The full report can be seen at http://www.advisorybodies.doh.gov.uk/uksci/uksci-reportnov05.pdf.

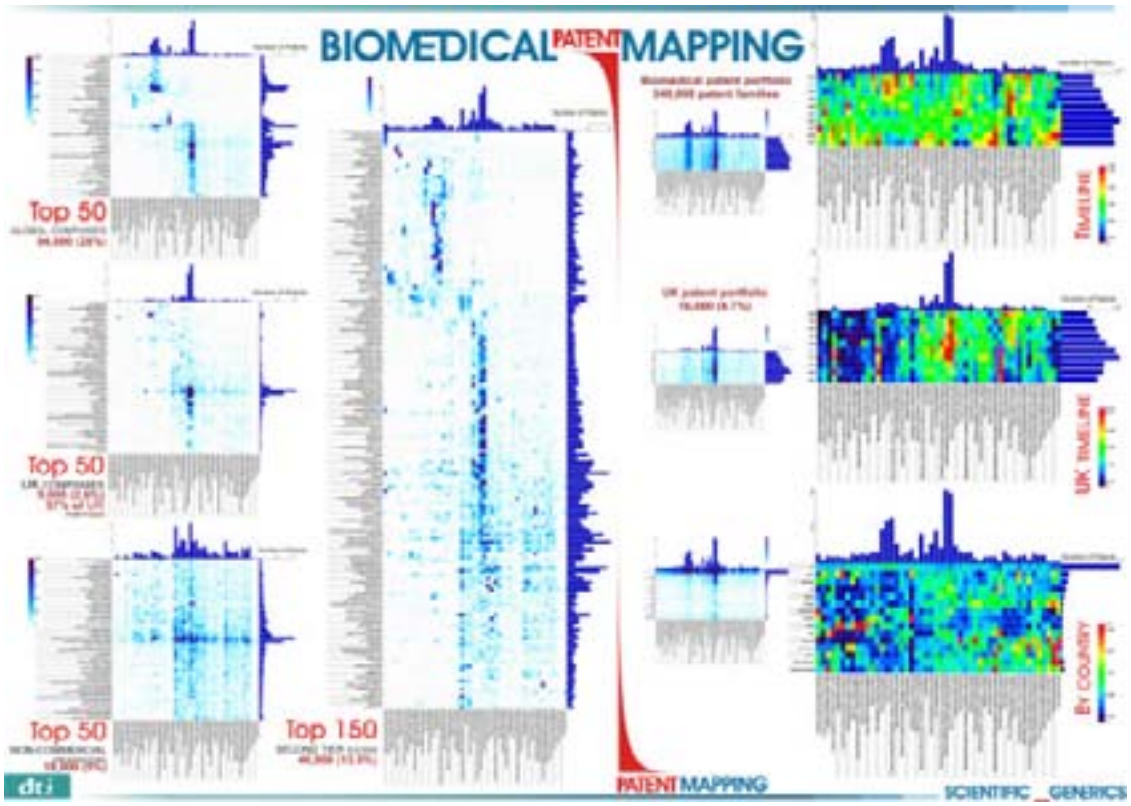
Case study – Biomedical patents

The Technology Strategy Directorate of the DTI needed to understand the world scene, assess relative UK status and identify key UK players in a diverse range of sectors broadly defined as pharmaceuticals; biotechnology and medical devices. We were asked to provide and analyse objective data of patenting and analyses to help meet this need.

The key tasks were:

- devise a practicable working taxonomy
- build a database of patents
- analyse the patent database to derive citation-weighted numbers of patents and used these data to identify global trends: over time and by geography and type of inventor;
- identify and list key UK and world players (big corporate, SME and academic).

The poster – which will shortly be published in A0 format – of which a thumbnail is shown below as Figure 6 summarises our findings



Provision of patent analysis tools as a public good

As part of an essentially international system there are few ways which national patent offices can or should legitimately favour local inventors. But the provision of high quality value adding services to all applicants by the UK Patent Office would inevitably worked to the advantage of UK inventors since the vast majority of inventors will naturally seek to prosecute patents through their home patent office.

The sort of analysis tools which we outline above would enable UK inventors better to extract valuable information from patent databases and maximise patent protection (and therefore the value of IP) by helping them identify and target 'white space' for innovation in the competitive patent landscape.

As a company we are not organised to support the use of third party software tools by an extensive user base. The UK Patent Office is well placed however to offer and support a patent analysis toolkit and we have reason to believe that the Search & Advisory Service is aware of the potential of such tools.

We would be prepared to work with UKPO on a commercial basis to make patent analysis tools universally available to UK inventors. It is hoped that this submission to the Gowers Review will help initiate government support for such a course of action.