

Title: **Evidence-Based Policy: Understanding the Technology Landscape**
Author: **E Richard Gold and Andrew M Baker**
EAP Date (approved for print): 02/11/2012
DOI: 10.5778/JLIS.2012.22.Gold.1



Journal of Law, Information & Science

Note to users: Articles in the 'Epubs ahead of print' (EAP) section are peer reviewed accepted articles to be published in this journal. Please be aware that although EAPs do not have all bibliographic details available yet, they can be cited using the year of online publication and the Digital Object Identifier (DOI) as follows: Author(s), 'Article Title', Journal (Year), DOI, EAP (pg #).

The EAP page number will be retained in the bottom margin of the printed version of this article when it is collated in a print issue. Collated print versions of the article will contain an additional volumetric page number. Both page citations will be relevant, but any EAP reference must continue to be preceded by the letters EAP.

ISSN-0729-1485

Copyright © 2012 University of Tasmania

All rights reserved. Subject to the law of copyright no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise, without the permission of the owner of the copyright. All enquiries seeking permission to reproduce any part of this publication should be addressed in the first instance to:

The Editor, Journal of Law, Information and Science, Private Bag 89, Hobart, Tasmania 7001, Australia.

editor@jlisjournal.org

<http://www.jlisjournal.org/>

Evidence-Based Policy: Understanding the Technology Landscape

E RICHARD GOLD* AND ANDREW M BAKER**

1 Introduction

Debates over gene patenting, genomics, pharmaceuticals and other fields have largely been conducted on the basis of assumption — that patents are essential to innovation¹ or that they stifle innovation² — rather than on evidence. It is thus refreshing that, in her article, *Implications of DNA Patenting: Reviewing the Evidence*,³ Dianne Nicol pointed to the centrality of evidence in resolving longstanding arguments over human gene patents. This article argues that it is not just any evidence, but evidence that is collected and analysed following standard practices in a transparent manner and with an acknowledgement of the limitations of the analysis, that counts.⁴

This article will illustrate the need for greater care in collecting and communicating evidence through a discussion of a increasingly important tool used to understand and analyse innovation policy and business strategy:

* James McGill Professor, Faculty of Law, McGill University. The authors wish to thank Ben James of the UK Intellectual Property Office, Dan Cahoy, Smeal College of Business, Pennsylvania State University, Greg Graff, Department of Agricultural and Resource Economics, Colorado State University and participants at the *Managing Knowledge in Synthetic Biology* conference held on 20-21 June 2012 in Edinburgh, Scotland for their insights and comments. This research was supported by VALGEN (Value Addition through Genomics and GE3LS), a project sponsored by the Government of Canada through Genome Canada, Genome Prairie and Genome Quebec.

** LL.B-B.C.L. student, Faculty of Law, McGill University.

¹ See, for example, *Association of Molecular Pathology v US Patent and Trademark Office*, 689 F 3d 1303, 1333 (2012) per Lourie J. (“Patents encourage innovation and even encourage inventing around”).

² D Resnik, “DNA Patents and Scientific Discovery and Innovation: Assessing Benefits and Risks” (2001) 7 *Science and Engineering Ethics* 29, 39 (“Allowing people to patent this basic information about DNA would hinder the free flow of information and stifle scientific innovation and progress. Allowing people to patent naturally occurring DNA sequences would also have detrimental effects on innovation and discovery.”)

³ D Nicol, “Implications of DNA Patenting: Reviewing the Evidence” (2011) 21 *Journal of Law, Information and Science* 7.

⁴ For a discussion of the introduction of this evidence into court proceedings in the Canadian context, see R Gold and R Carbone, “(Mis)Reliance on Social Science Evidence in Intellectual Property Litigation: A Case Study” (2012) 28 *Canadian Intellectual Property Review* 179.

patent landscaping. Patent landscaping is a method of understanding the interrelationship of data extracted from patent documents across a certain dimension, be that technological, geographical, or company, etc. It is used variously to track technological development,⁵ answer innovation policy questions⁶ and to develop business strategy.⁷

While a potentially powerful tool, patent landscaping has, to date, been conducted largely on an *ad hoc* basis.⁸ This can largely be explained by the emerging nature of patent landscapes as an analytical tool, greater access to primary patent documents through the Internet and growing interest in empirical analysis of innovation and innovation policy. Nevertheless, the *ad hoc* nature of its employment limits both the short- and long-term utility of the method. Notably, the lack of standards makes validation and comparison of results difficult if not impossible and an identification of errors opaque.

This article begins a conversation aimed at harmonising and making more transparent patent landscape studies. To accomplish this, it is important to standardise the steps involved in the landscape and to ensure that methodological decisions are made in light of the objectives being sought. After describing some of the problems with current methods, this article turns to a discussion of the steps required in a landscape and some of the pragmatic constraints that limit methodological choices before providing heuristic tables to assist in selecting the most appropriate methodology.

2 Common Problems in Patent Landscaping

A review of the literature reveals two common problems in patent landscaping:

- (1) landscapers fail to properly disclose justifications for methodological choices; and
- (2) landscapers often draw conclusions from analytical methods that do not meet the norms for evidence-based policy making.

⁵ See for example, UK Intellectual Property Office, “Regenerative Medicine: the Landscape in 2011” (June 2011) <<http://www.ipo.gov.uk/informatic-regenmed.pdf>>.

⁶ See for example, UK Intellectual Property Office, “Patent Thickets: an Overview” (November 2011) <<http://www.ipo.gov.uk/informatic-thickets.pdf>>.

⁷ B Fabry, H Ernst, J Langholz and M Köster, “Patent Portfolio Analysis as a Useful Tool for Identifying R&D and Business Opportunities — An Empirical Application in the Nutrition and Health Industry” (September 2006) 28(3) *World Patent Information* 215.

⁸ In an unpublished, preliminary study conducted under the supervision of one of the authors, the great majority of articles either failed to disclose their methodology or use their own in-house method.

In many cases, these problems could be addressed if landscapers were aware of the full range of methodological choices available to them as well as the relative strengths and shortcomings of landscaping methods.

2.1 Failure to sufficiently disclose methodological choices

When methods are insufficiently disclosed, it is not possible to determine the breadth of application of the results, if there are any other limitations to the conclusions, and what assumptions were made in conducting the study. This is particularly difficult for landscapes presented in academic journals and business publications where formatting requirements can be constraining and where authors wish to focus attention on conclusions rather than method. Full patent reports and white papers lend themselves to better disclosure as they often exceed one hundred pages in length and contain numerous appendices where methodological choices can be documented.⁹ Nevertheless, there are numerous disclosure pitfalls that do not require extensive appendices. The following examples are not exhaustive, but demonstrate some potential issues.

Many landscapes simply provide visualisations of results without fully disclosing the specifics of the underlying information.¹⁰ For example, in a plotted patent map, landscapers should disclose whether patent applications or families are plotted. Patent maps should also disclose whether pre-grant publications or post-grant publications, or both are plotted. Patents may be obtained through various routes and all of these choices should be considered and acknowledged. When conducting analysis over a time interval, landscapers should also disclose any pertinent changes to legal or institutional regimes or other historical artefacts that may influence the data. For example, the 2001 US switch between publishing only granted patents to publishing both patent applications and granted patents would have had a significant effect on results.

Landscapers should also disclose the scope and limitations of the analysis conducted. For reasons of economy, landscapers may only assess a particular geographical area or filings at a particular patent office. If landscapers draw conclusions about the global patent landscape or the efficacy of their analytical techniques on a broader scale than studied, they must justify the universality of their conclusions.

⁹ Many common examples can be found in WIPO's *Patent Landscape Reports* database:
<http://www.wipo.int/patentscope/en/programs/patent_landscapes/published_reports.html>.

¹⁰ Many definitions in the literature still limit "patent landscaping" or "patent mapping" to the visualisations themselves. The authors employ a broader definition of landscaping because many types of analysis do not create traditional visualisations such as clusters or themescapes, but nevertheless assess the interrelationship of data extracted from patent documents across a certain dimension.

2.2 Analytical rigour: meeting norms of evidence-based policy making

It is important that patent counts truly reflect the variable that the landscaper is attempting to assess and avoid duplicates. Patents in any particular data set will have multiple applicants, inventors and priorities that can lead to double counting. For example, a single US patent document with inventors from Canada, Japan and Denmark could appear three times in a simple plot of patents by inventor country. This may or may not be a problem, depending on the purpose of the analysis, but it should be considered when analysing data and reporting results.

One example of a corrective tool for counting errors is the Relative Specialization Index (RSI). Many simple landscapes present counts by country and assert that higher counts for one applicant country relative to another indicate greater innovation in that particular technology area. But this count fails to take into account existing variance due to propensity to file. For example, applicants from the US and Japan are prolific patentees and will almost always dominate a comparative analysis. An RSI takes into account pre-existing variance to see who is patenting at a greater rate relative to expected trends.¹¹

When using patent counts it is also important to remember that patents represent only one facet of the innovation environment. Some technology fields are more patent dependent than are others, which may rely on trade secrets, trade-marks, other forms of intellectual property protection or being first to market. Thickets in some technology areas may be caused by a small set of enabling technologies whereas thickets in other technology areas may be more diffuse.

Differing definitional boundaries of a technology area may also affect the final analysis. For example, a 2005 patent landscape concluded that 20 per cent of the human genome had been patented.¹² The authors' method for defining exactly what constituted a human genome patent was subsequently criticised.¹³ In some cases, categorisation by existing structured data such as International Patent Classification (IPC) code is not sufficiently accurate and some landscapes employ subjective hand-coding in order to define and categorise the technology space.¹⁴ Being clear and transparent about

¹¹ See for example, UK Intellectual Property Office, "Agrifoods: a Brief Overview of the UK Agrifood Landscape" (June 2012) 3, <<http://www.ipo.gov.uk/informatic-agrifood.pdf>> for an explanation of the RSI formula.

¹² K Jensen and F Murray, "Intellectual property landscape of the human genome" (October 2005) 310(5746) *Science* 239.

¹³ L Pressman, "DNA Patent Licensing Under Two Policy Frameworks: Implications for Patient Access to Clinical Diagnostic Genomic Tests and Licensing Practice in the Not-for-Profit Sector" (March 23, 2012) 6 *Bloomberg BNA Life Sciences Law & Industry Report* 329, 330-32.

¹⁴ See for example, M Lloyd and J Blows, "Clean Coal Technologies: Where does Australia Stand?" (April 2009) 4,

definitions is thus critical to being able to use and assess the validity of a patent landscape.

Keeping these common problems in mind, we turn, in the next section, to guidelines for the conduct of a patent landscape.

3 Elements of a Patent Landscape Methodology

The essential elements in conducting a patent landscape include the following: (1) developing an overall research strategy, (2) building the dataset, and (3) conducting the analysis. In addition, landscapers also perform a number of (4) post-analysis tasks that may inform methodological choices. In reality, the process itself rarely follows these components in a structured, step-by-step manner. Rather, landscapers often iterate between steps. For example, the overall research strategy may be left undefined until researchers have an opportunity to assess the breadth and robustness of preliminary search results.¹⁵ Often, keyword lists need to be revised, and useful methods of analysis are discovered during the carrying out of the landscape.¹⁶ Thus, the following breakdown is not intended as a step-by-step taxonomy but, rather, an explanation of the essential elements of a patent landscape methodology.

3.1 Identifying the key research objectives

Patent landscapers begin with an objective and develop research questions that aid in achieving that objective. A landscaper seeking to improve efficiency of intellectual property agreements and technology transfer may ask, for example: what are the trends in patenting in the defined technology area?; who is patenting in this area?; and which are the most significant patents?¹⁷ There are often multiple different analyses that could help to answer a particular research question. For example, assessing the most relevant patents in a technology field could require a claims analysis by a legal expert or could utilise a proxy such as citation counts in the scientific

<http://www.griffithhack.com.au/Assets/1672/1/GH_CleanCoal_April_2009.PDF>.

¹⁵ This is particularly the case when the objective of the landscapes seeks to define the contours of the technology area or forecast emerging technology trends. For example, regulators may wish to pre-emptively adapt existing frameworks to emerging fields such as synthetic biology and nanotechnology. See Table 5 in Section 5 for a further discussion of this type of landscape analysis.

¹⁶ See for example, J Cavicchi and S Kowalski, "Patent Landscape of Adjuvant for HIV Vaccines" (Fall 2009) <http://www.wipo.int/patentscope/en/programs/patent_landscapes/published_reports.html> for a reiteration due to an upgrade to higher-order analytics (Thomson Innovation).

¹⁷ Adapted from the patent landscape conducted for: K Bergman and G D Graff, "The Global Stem Cell Patent Landscape: Implications For Efficient Technology Transfer and Commercial Development" (2007) 25(4) *Nature Biotechnology* 419.

literature. As a result, most patent landscapes generally ask multiple research questions and conduct multiple different analyses in order to make sound evidence-based conclusions and effectively manage risk.

While the landscaper may not have an exact conception of which analyses will be conducted until after the data set has been compiled, it is still necessary to consider which types of analysis will best respond to the research question prior to performing the data search. Higher-order analytics increasingly allow for consolidation of search and analysis functions into a one-step automated process. While researchers can always go back and augment the dataset or perform novel analyses, the process of conducting a research strategy allows a landscaper to consider the broadest range of methodological choices and then match question to method. Moreover, it allows for hypothesis testing allowing for the validation of methods rather than ex-post rationalisation of observed trends in the data.

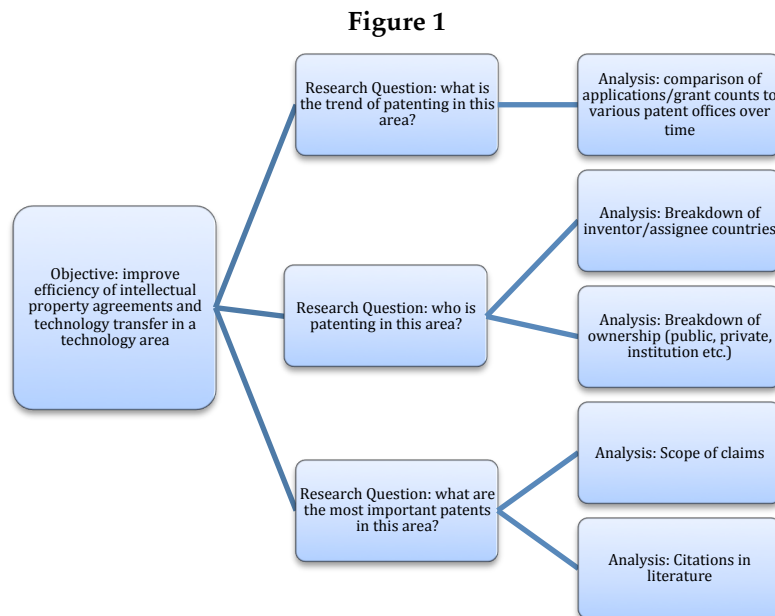


Figure 1: Example of a research strategy seeking to improve the effectiveness of technology transfer agreements in a given technology area.¹⁸

3.2 Building the dataset

Once the landscaper has developed the set of overall research objectives, it is necessary to build an appropriate dataset to respond to those objectives. Generally, landscapers begin with the broadest informational need at the lowest possible scaling keeping in mind time and cost constraints.

¹⁸ Ibid.

To build a dataset, the landscaper will need to undertake several steps: (1) set the search approach; (2) choose the search method; (3) select the dataset; and (4) curate the data. Figure 2 illustrates these choices.

3.2.1 Search approach

One can approach the search in one of two ways: top-down or bottom-up. A top-down approach is the more common. It entails first defining the field of interest and then designing a search strategy that captures as complete a set of patent documents falling within that field as possible. A bottom-up approach brings no preconception of the field of study, commencing with only a few specific documents and leveraging those to build a complete data set by seeking linkages to other patent documents through patent citation or through the collection of patent technology codes. This approach is more effective when the research question focuses on a foundational or enabling technology of which the landscaper is already aware. Landscapers may combine both methods. For example, one may begin with a standard top-down approach, but then conduct a bottom-up search to determine which patents are the most relevant or to determine if any relevant patents were missed by the initial search.

3.2.2 Search method

Following the selection of the general approach to the search, the landscaper will develop query strings to achieve the greatest capture of relevant documents while minimising irrelevant hits.¹⁹ Search methods range from searching a single variable in standard fields, such as a technology code, to simple Boolean searches, to advanced information retrieval approaches such as semantic searches. Landscapers can forgo developing query strings when the dataset is available as a secondary source or when dealing with a very small dataset.²⁰ Algorithms can mine for both structured data (eg, IPC code, assignee, year) and unstructured data (eg, keywords within the description or claims). Semantic solutions, in which word meaning rather than mere word occurrence or frequency is what is important, are being increasingly developed to address problems with traditional algorithms.²¹

It is crucial to consider the technology area one is examining. Some technology fields are easily captured through algorithms either because they have a specific vocabulary or because the landscaper can use structured data (standardised meta-data) searches such as title, IPC code, date, inventor etc.

¹⁹ By patent family, technology area etc.

²⁰ See for example, B P Georgieva and J M Love, "Human Induced Pluripotent Stem Cells: A Review of the US Patent Landscape" (July 2010) 5(4) *Regenerative Medicine* 581, which conducts a legal analysis of the first granted pluripotent stem cell patents.

²¹ D Bonino, A Ciaramella and F Corno, "Review of the State-of-the-Art in Patent Information and Forthcoming Evolutions in Intelligent Patent Informatics" (March 2010) 32(1) *World Patent Information* 30, 35.

Other technology fields that are not easily captured through algorithms and landscapers will require careful construction of keyword lists, perhaps resorting to Delphi methods or other forms of expert keyword identification. More complex qualitative research questions often cannot be captured in an algorithm and require a subjective coding framework.

3.2.3 Data source

Patents and patent applications are retrievable from international databases (eg, WIPO), national databases (eg, USPTO), publicly available tools that search international databases (eg, GooglePatents), and proprietary commercial databases (eg, Derwent). Free public databases generally have limited capabilities for complex algorithms, provide only basic informational fields within specific countries or regions, are subject to transcription errors, and offer no higher order analytics. Proprietary databases enable, on the other hand, complex algorithms across multiple jurisdictions, provide self-defined fields, and higher-order analytics.²²

Patent landscapers may also use secondary data sets. There are an increasing number of patent landscape data sets in the public domain that provide free data sets that can be analysed in novel ways. These are, however, often limited by technology field and may become quickly outdated requiring updating.²³

In many cases, searches will not be limited to patents but include other relevant data such as the scientific literature, business news and reports, law articles and Web 2.0 searches.²⁴ This is especially the case when the goal behind the landscape is to characterise emerging technology or when firms attempt to assess competitor intellectual property positions prior to applications being submitted.

3.2.4 Curation method

Data curation includes the merging of datasets from different databases into consistent fields and a reassessment of the degree of Type I (falsely including a patent document that ought not to be included) and Type II (falsely excluding a patent document that ought to have been included) error involved in the search. In the usual case, curation is conducted automatically.

²² Ibid 32.

²³ See for example the UK Intellectual Property Office Patent Data Reports database <<http://www.ipo.gov.uk/informatics-reports>> or the WIPO patent landscape report database <http://www.wipo.int/patentscope/en/programs/patent_landscapes/published_reports.html>.

²⁴ Bonino, Ciaramella and Corno, above n 21, 34-35. The authors discuss the advantage integration of heterogeneous information sources into patent analytics.

Subjective curation may be necessary, in certain cases, such as when the boundary of the technology is vague.²⁵

Figure 2

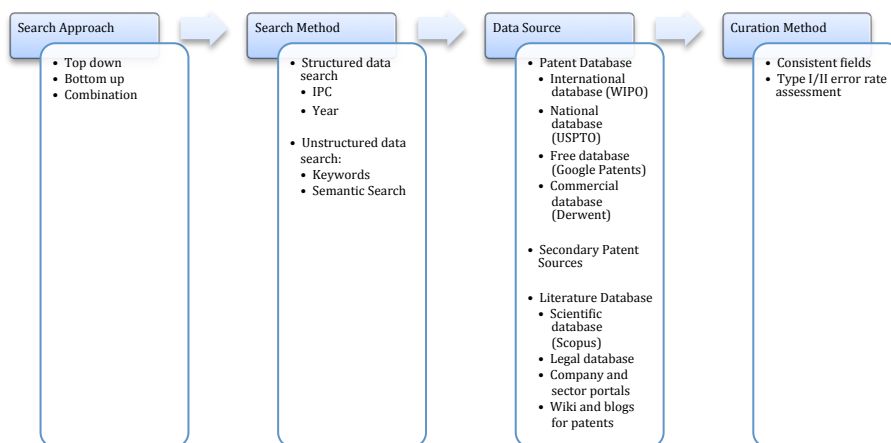


Figure 2: Summary of methodological choices involved in building the dataset

3.3 Data analysis

Landscapers have an increasingly wide number of analytical tools from which to choose in analysing datasets. These range from the basic categorisation of lexicographic data to assess patenting patterns to complex qualitative assessment of claims.

Automation minimises the marginal cost of the data analysis. Many basic analyses can be conducted through available computer applications. Lexicographic data (eg patent counts, filing date, granted date) is easily manipulated in a variety of ways. Examples of this type of analysis include technology and temporal comparisons. A wide variety of automated visualisation tools such as technology cluster and owner/assignee correlation maps have been developed to aid in recognising patterns in large datasets.²⁶

²⁵ Lloyd and Blows, above n 14. The authors assess both relevancy and technology area in a subjective coding framework. In the report, clean coal technologies were classified into three major tech areas in order to compare Australia's IP position in each area. The authors also chose to use curation to limit the dataset to technologies that are commercially proven, a distinction that is difficult to capture without subjective coding.

²⁶ See for example, Y Y Yang, L Akers, C B Yang, T Klose and S Pavlek, "Enhancing Patent Landscape Analysis With Visualization Output" (2010) 32 *World Patent Information* 203, for an overview of various contemporary visualisation applications.

Increasing volumes of applications, complexity of research, emerging markets with differing institutions and languages, non-roman character content, and longer pendency periods are all issues that pose challenges to state of the art patent landscaping.²⁷ In cases where traditional graphing, mapping, and other visualisation techniques are insufficient, landscapers need to consider higher-order analytics.²⁸

Cahoy's hierarchy of the most common analytical tools used in patent landscaping demonstrates the increasing complexity of various analyses commonly conducted by patent landscapers.²⁹ This is illustrated in Figure 3.

Figure 3

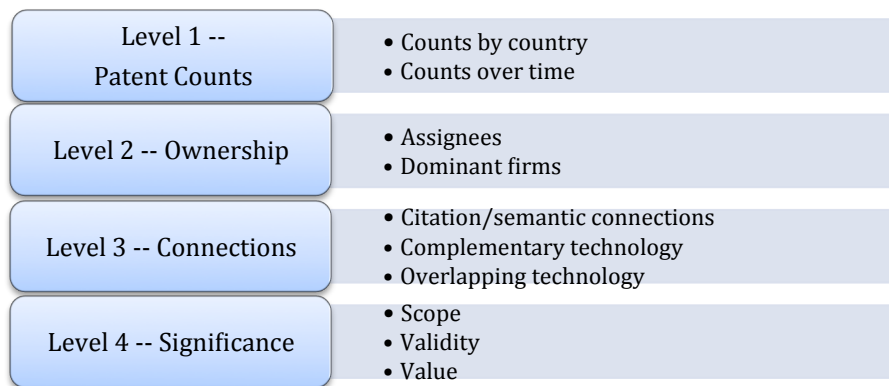


Figure 3: Hierarchy of common analytical tools; each subsequent level requires more complex and generally more costly analytical techniques to assess.

Landscapers have developed other methods to analyse data when no appropriate computer applications are available or prove too costly.

3.3.1 Subjective analysis by technical or legal experts

Highly qualitative determinations are sometimes best captured by a patent expert (eg, patent agent or examiner) or by a non-expert researcher using a pre-determined coding framework applied to individual patent documents. For example, precise legal questions involving the strength, scope, and

²⁷ V Caraher, "The Evolution of the Patent Information World Over the Next 10 Years: A Thomson Scientific Perspective" (June 2008) 30(2) *World Patent Information* 150, 150-1.

²⁸ Proprietary databases offer many solutions to landscapers. Also, see C H Wu, Y Ken and T Huang, "Patent Classification System Using a New Hybrid Genetic Algorithm Support Vector Machine" (September 2010) 10(4) *Applied Soft Computing* 1164, for an overview of some contemporary semantic search methods.

²⁹ D Cahoy, "Effective Patent Landscaping: Making Sense of the Information Overload" (January 20-21, 2011) Presented at VALGEN Workshop.

validity of claims can rarely be assessed by a computer application and have been traditionally the purview of formal freedom-to-operate analyses involving expert assessment of patent claims in the context of the entire patent document and prior art. Technical questions may also require the opinion of technical experts when keyword or semantic solutions are insufficient, as is the case in technology fields with a great deal of conceptual and vocabulary overlap with irrelevant technical fields.

3.3.2 Using proxy data

For some of the most complex research questions, there exists no direct variable that captures the phenomenon under analysis. Examples include the value of a set of patents, levels of innovation in a particular region or the quantum of knowledge transfer. In these cases, researchers must identify proxy variables that, as closely as possible, follow the pattern of the underlying phenomenon. There are lively scholarly debates about which analyses are best suited to answer particular research questions and about the relative effectiveness and limitations of competing types of analyses. The field of novel statistical analysis is rapidly advancing in an attempt to create standardised methods for assessing highly complex questions such as patent valuation and likelihood of litigation.³⁰

3.4 Post analysis and landscape users

While technically distinct from the landscape methodology itself, landscapers engage in a number of post-analytic activities that may influence methodological choices such as the ability to later extend the study in time, by country or by technology.

Institutions will often maintain landscapes over time in order to monitor changes in the innovation environment as a part of overall patent portfolio management or in the case of government, policy-making. For example, WIPO maintains several longitudinal studies, particularly in the fields of public health, food security, climate change, and environment in developing

³⁰ Much research has focused on finding proxies that correlate with patent value/quality such as citation analysis and litigation analysis. See for example, J Allison, J H Walker and M A Lemley, "Patent Quality and Settlement among Repeat Patent Litigants" (September 16, 2010) *Stanford Law and Economics Olin Working Paper* No 398; M Kramer, "Valuation and Assessment of Patents and Patent Portfolios through Analytical Techniques" (2007) 6(3) *John Marshall Review of Intellectual Property Law* 436; Y S Chen and K C Chang, "The Relationship Between a Firm's Patent Quality and its Market Value – The Case of the US Pharmaceutical Industry" (January 2010) 77(1) *Technological Forecasting and Social Change* 20; D Harhoff, F M Scherer and K Vopel, "Citations, Family Size, Opposition and the Value of Patent Rights" (September 2003) 32(8) *Research Policy* 1343.

countries.³¹ Firms may do the same in order to conduct long-term management of IP portfolios.

4 Pragmatic Considerations Regarding Selection of Methodology

In addition to methodological concerns associated with the selection of data sets, search algorithms, computer applications and so on, significant pragmatic concerns also shape the selection of methodology. These relate to issues external to the methodology itself but reflect the constraints and needs of both the landscaper and the landscape user. Three constraints come quickly to mind: cost, time and tolerance for error.

4.1 Cost

One of the main constraints on undertaking a landscape analysis is cost. There are two principal elements to consider in terms of cost.

First, the cost of accessing the patent database may be substantial, at least to private databases. As discussed earlier, commercial patent databases, such as Delphion and Derwent, offer superior search tools and other advantages over their public counterparts. Access to these databases start in the low thousands of US dollars per year.

Second is the cost of searching and analysing the patent documents in the data set. Depending on the number of patent documents involved, cost can range from the low thousands of dollars to hundreds of thousands or more.

Having patent agents or lawyers conduct an analysis of patent documents is expensive. This is because the expert must review the claims in light of the specification and determine the scope, application and validity of the reviewed patent. For a relatively small sample of patents, such as in a freedom-to-operate analysis in which one aims to identify the subset of patents that are relevant to the sale of a particular product, costs can run from several thousand US dollars to US\$40 000 or more.³² As the number of patents involved in the landscape increases, so does the cost of the associated expert analysis.

There are ways to reduce costs when using patent experts, such as only analysing issued patents and leaving aside patent applications or limiting the time period being studied. Similarly, the analysis could focus simply on patent scope and not on validity, again reducing the cost of the analysis.

³¹ WIPO has many ongoing studies, particularly in the fields of public health, food security, climate change, and environment in developing countries. See WIPO, *Patent Landscape Reports - On-going Work at WIPO* (2012) <http://www.wipo.int/patentscope/en/programs/patent_landscapes/ongoing_work.html>.

³² See, for example, IAM Consultants, "Freedom to Operate Analysis" (April 6 2012) <<http://ideaprotection.co.uk/freedom-to-operate-analysis/>>.

Nevertheless, when the landscape involves a large number of patents, use of patent agents or lawyers can become prohibitively expensive.

On the other end of the cost spectrum, an algorithmic search followed by curation offers a landscape at a relatively low cost. Further, this cost will not substantially increase with the number of patents involved. The largest component of cost for these searches is in developing the algorithm and the curation tools. Once developed, the only difference between identifying one and millions of patents is computer time.

In between these methods, an analysis drawing on university or student researchers will allow for the analysis of a significant number of patents — at least several hundred — at a reasonable cost. In addition to the researchers' time, which will cost much less than that of a patent expert, there will be costs involved in developing the coding frame that will instruct the researchers how to analyse the patent documents under examination. There will also be some costs involved in duplication of analyses to control for coding errors.³³

4.2 Time

Another important constraint is time. Depending on the method used, one can conduct a patent search in a matter of days or weeks to many months.

Algorithmic searches can be implemented relatively rapidly once the search algorithm is established and the curation tools are available. What can be time consuming is, however, the development of the algorithm itself, especially if in doing so one draws on external experts through a formal process, such as a Delphi.

The time involved in conducting expert and researcher analyses of patent documents will be proportional to the number of patent documents involved, the depth of the analysis and the skill level of those conducting the analysis. All things being equal, one would expect that analyses conducted by researchers will be the most time consuming given the need for training, validation and familiarity with the contents of the patent documents.

4.3 Accuracy

No method of patent landscape will always be accurate. All are subject to Type I and Type II errors. Type I errors tend to find more patents or patent claims in the landscape than exist in reality. Such errors tend to overestimate problems in bringing products to market, find too many blocking patents, and identify thickets when they do not exist. Type II errors underestimate the number of relevant patents, leading to increased but unrecognised liability for patent infringement, failing to recognise potential partners and underrating innovative strength. Both types of errors can lead to the development of poor business and policy strategy as well as inappropriate conclusions about the effect of patents on innovation.

³³ See Cavicchi and Kowalski, above n 16, for a good example of a transparent coding framework utilising teams of student researchers.

Patent landscape analysis techniques can be conducted by (1) automated computer applications; (2) subjective coding by technical experts; or (3) by patent experts with knowledge of claims construction and legal language. As research questions, tolerance for error and cost constraints can vary greatly, there is no one best method for conducting a landscape. There are, however, better and less good methodologies. Each of the three broad methodological approaches present various strengths and shortcomings that make them more appropriate for particular research questions and the pragmatic realities facing the landscaper.

Algorithmic searches will be subject to both Type I and II errors. No matter the sophistication of the search terms used, the search will inevitably catch patent documents that are not relevant to the subject of the landscape and will omit others that ought properly to have been included. These errors can, however, be substantially reduced by working with experts in the scientific field and in patents to identify all possible search terms and to exclude overly vague terms. This may, as noted above, involve more time, but the increase in accuracy will usually be justified.

The selection of database will also have an effect on accuracy. Some national patent databases may be incomplete or make searching of certain terms difficult. The databases themselves may contain incomplete or inaccurate information that will affect the results of an algorithmic search. Private databases, while more expensive, generally provide for better searching, for example, the ability to conduct multi-jurisdictional searches and patent family analysis (although Patent Lens, a free Internet service, offers these tools). Where accuracy is important, this extra expense may be justified.

The use of experts or researchers to analyse the results of an algorithmic search can both reduce Type I errors and introduce Type II errors.

Experts and researchers can identify those patent documents that were included in the search results by error by analysing each document in its entirety. In doing so, they can reduce the number of Type I errors. One would expect that patent experts will be in the best position to identify those errors since their increased understanding of patent claim language, key terms and patent strategy provide them with a more subtle and careful appreciation of patent documents. University and student researchers who do not share this level of knowledge are more likely to retain documents that ought to have been excluded.

Both experts and researchers can introduce Type II errors by falsely excluding patent documents that had been identified by the search algorithm. Once again, one would expect that patent experts would be less likely to commit these errors.

The above analysis suggests that researchers seeking to reduce costs by analysing patent documents themselves or with the assistance of their students risk increasing the prevalence of both Type I and II errors. If not acknowledged, this could lead to significant mistake in the formulation of policy.

As different business and policy questions demonstrate different levels of tolerance to Type I and Type II errors, selecting methodologies subject to either type of error may not present any difficulty. For example, in large-scale landscapes, missing some relevant patents and including others that are not relevant will be unlikely to alter the identification and analysis of overall trends. In smaller scale studies — up to a few hundred — in which one of the goals is to identify key or blocking patents, such errors may be more serious.

5 Reconciling Objective/Question to Method

Taking into account both the ideal scope and nature of the landscape derived from the analysis presented in Part 3 and the pragmatic constraints on methodological choice discussed in Part 4, the landscaper must choose the most apt methodology to achieve her or his overall goals.

The discussions below illustrate some of the most common research questions facing business, government, and academics and the relative strengths and weaknesses of the three broad methodological approaches.

5.1 Landscaping During the Technology Cycle

Patent landscapes are commonly used at various stages along the technology development cycle to assist in making strategic operational and planning decisions before proceeding towards the final implementation or commercialisation of an invention.³⁴ This is illustrated in Figure 4.

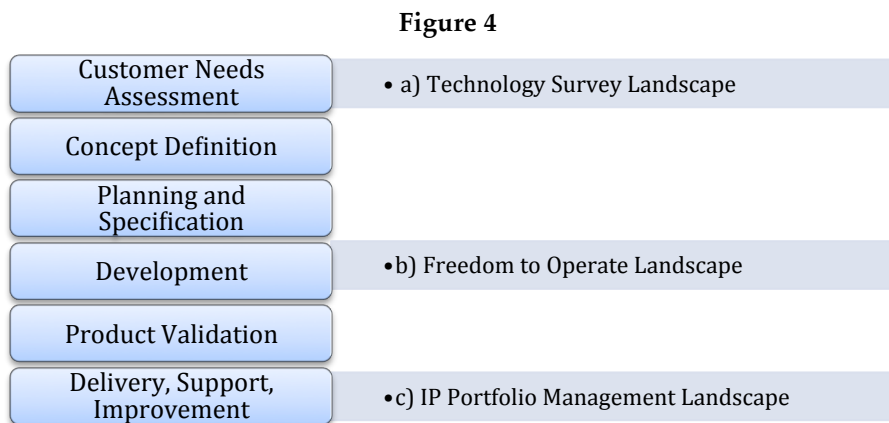


Figure 4: Technology development cycle³⁵ with three types of common landscapes and at what stage they are conducted.

³⁴ See J Burdon, “IP Portfolio Management: Negotiating the Information Labyrinth” (2010) Chapter no 12.4, PIPRA IP Handbook (2012) <<http://www.iphandbook.org/handbook/ch12/p04/>> for a discussion of patent analytic strategies used throughout the technology cycle.

³⁵ Ibid, adapted from Figure 3.

5.1.1 Technology survey

A technology survey is an overview of patent data within a technology field in order to broadly characterise the state of the art and to assess ownership. It is used to identify markets for an invention as well as to assess competitor strength and the potential for collaboration.

Table 1: Technology Survey

Computer-Based	Research Team	Patent Expert
Feasibility: • Useful	Feasibility: • Potentially useful	Feasibility: • Unnecessary
When Useful: Computer-based applications such as visualisations and comparative jurisdictional reports can give a great deal of information about a market. These landscapes can be done inexpensively to identify competitive strength before making strategic commercialisation decisions.	When Useful: Can provide some additional clarity, especially when limited to the most pertinent subset of the initial dataset. Due to the importance of unpublished patent data, literature reviews and competitor surveys can be utilised in a subjective coding framework to gather pertinent information unavailable in patent databases.	When Useful: Can provide clarity on a subset of the data, but this is usually unnecessary for a broad technology survey
Limitations: It is important that the landscaper understand the role of patents in the applicable technological space in order to be able to identify the limitation of these analyses. For example, certain technology fields rely more on trade secrets than on patents so clustering/white space analysis may be misleading. Literature reviews should complement patent data analysis.	Limitations: Firms, especially SMEs, may not possess the requisite human resources to conduct these analyses. Since the goal of the search is to only identify competitors and what technologies they are using, the marginal benefit of a complex coding framework in clarifying the landscape is low.	Limitations: Cost. The expense of hiring a patent expert is only justified once the enterprise is more advanced in the commercialisation process.
Example: Various automated analysis techniques such as text-mining, network analysis, citation analysis are used to conduct technology surveys. ³⁶	Example: Used when a subset of the landscape is of particular importance such an enabling technology or an invention required for market entry. ³⁷	Example? We found none.

³⁶ S Lee, B Yoon, C Lee and J Park, "Business Planning Based on Technological Capabilities: Patent Analysis for Technology-Driven Roadmapping" (July 2009) 76(6) *Technological Forecasting and Social Change* 769.

5.1.2 Freedom to operate

Once a product or process has been conceived, a freedom to operate analysis is conducted to determine potential infringement of any of claims or to assess what art is expired or in the public domain.

Table 2: Freedom to Operate

Computer-Based	Research Team	Patent Expert
Feasibility: • Not feasible	Feasibility: • Insufficient	Feasibility: • Sufficient
When Useful: Very low marginal cost of processing large datasets. Could mitigate a major barrier to entry for SMEs.	When Useful: Less expensive than patent experts; can leverage student researchers. Leverages technical knowledge of experts through a Delphi process; can be particularly useful when highly technical questions are key to assessing the strength of claims.	When Useful: The experience and expertise of patent experts, complimented by in-house resources not possessed by other landscapers permits the care of analysis required.
Limitations: It is generally impossible to assess claim strength, scope, and validity with sufficient accuracy.	Limitations: Technical experts lack knowledge of claims construction, legal language, and the current state of patent law; cannot guarantee the accuracy of an expert.	Limitations: Cost. SMEs, recognise that little else can be done for FTO than hiring a legal expert to conduct a formal FTO report. This may pose a barrier to entry for otherwise innovative firms.
Example: Some new analysis technologies are attempting to model the work of patent experts through artificial intelligence. ³⁸ Nevertheless, these tools are not yet sufficient to determine the strength, scope, and validity of patent on a broad scale. The same can be said for proxies. It is unlikely that any proxy meta-data could substitute for a patent expert at present.	Example: Businesses often conduct preliminary in-house FTO analysis, but more formal analysis is necessary before commercialisation. Academics often assess patent relevancy using subjective coding. Nevertheless, almost all reports conducted in this way carry a disclaimer warning users that the dataset and analytics are insufficient to meet FTO standards of accuracy. ³⁹	Example: FTO analyses abound, but methods are generally proprietary. ⁴⁰

³⁷ See for example, K OuYang and C Weng, “A New Comprehensive Patent Analysis Approach for New Product Design in Mechanical Engineering” (September 2011) *78(7) Technological Forecasting and Social Change* 1183, 1192. Subjective analysis by a research team is used in a patent landscape of armored vehicles for entry into Asian markets. The authors determine that suspension systems are the key to market entry and narrow their more detailed subjective analysis to this field, while conducting automated patent analysis for a broader dataset.

³⁸ Wu, Ken and Huang, above n 28.

³⁹ See for example, I Huys, N Berthels, G Matthijs and G Van Overwalle, “Legal Uncertainty in the Area of Genetic Diagnostic Testing” (2009) *27 Nature*

5.1.3 Portfolio management

Once an actor holds IP assets, landscaping can be used to manage its portfolio and monitor the portfolios of competitors and collaborators. Portfolio management includes a variety of tasks including valuation of IP assets and management of expiries.

Table 3: Portfolio Management

Computer-Based	Research Team	Patent Expert
Feasibility: <ul style="list-style-type: none"> Potentially Useful 	Feasibility: <ul style="list-style-type: none"> Potentially Useful 	Feasibility: <ul style="list-style-type: none"> Potentially Useful
When Useful: Cost effective. Can use the same analytical tools on newly acquired or newly identified patents with little additional cost.	When Useful: Only necessary when valuation hinges on a precise technical issue or detailed assessment of competitor IP position is required.	When Useful: Only necessary when valuing single or small groups of patents or when the value of a portfolio hinges on a particular legal interpretation.
Limitations: There are limitations to automated techniques for complex questions such as assessing valuation. The more specific the valuation sought, the more likely subjective analysis will be required. Some technology areas have more advanced techniques utilizing proxy data to assess portfolio value.	Limitations: May be unable to give certain answers when valuation could be affected by validity or likelihood of litigation.	Limitations: Generally unnecessary to conduct detailed claims analysis; especially when most of the dataset is maintained over time.
Example: Many indicators of portfolio strength can be automated. Some examples include: technology share, cooperation intensity, and share of granted patents. ⁴¹	Example: In many cases, landscapers will conduct as much automated analysis as possible and then identify a discrete set of data from the original dataset and perform subjective analysis to gauge more difficult indicators of portfolio value such as patent quality / strength. ⁴²	Example: Patent experts could also assess a discrete set of data if the indicator sought would benefit from a legal interpretation.

Biotechnology 903. The landscape is not a formal freedom to operate analysis, but the authors attempt to assess the possibility of circumventing claims in much the same way an FTO analysis would.

⁴⁰ FTO analyses are generally conducted by counsel or in-house and rarely released publicly. See S P Kowalski, "Freedom to Operate: the Preparations" (2010) Chapter 14 in PIPRA ipHandbook of Best Practices, available at <<http://www.iphandbook.org/handbook/ch14/p02/>> for a discussion of FTO best practices.

⁴¹ Fabry, Ernst, Langholz and Köster, above n 7, 218. See Table 2 which provides and overview of patent indicators for competitor monitoring.

5.2 Meta-analytical Landscaping

Patent landscapes are also conducted to assess broader, macro-level questions outside the technology cycle of a particular invention. These types of landscapes examine the workings of markets, the efficacy of regulation and public policies and the functioning of legal institutions.

5.2.1 Comparative analysis of patenting behaviour

One of the most common types of landscapes conducted assesses the comparative patenting behaviour of different actors (companies, universities, countries etc) to determine their relative innovative position and inform policy-making.

Table 4: Comparative Analysis of Patenting Behaviour

Computer-Based	Research Team	Patent Expert
Feasibility: <ul style="list-style-type: none"> • Sufficient 	Feasibility: <ul style="list-style-type: none"> • Potentially Useful 	Feasibility: <ul style="list-style-type: none"> • Unnecessary
When Useful: Can conduct a variety of comparative analyses quickly and cost-effectively.	When Useful: When researchers are sceptical that patent counts are not indicative of innovative strength and no good proxy data exists. This is especially the case when a qualitative technical variable is indicative of strength.	When Useful: When researchers are sceptical that patent counts are not indicative of innovative strength and no good proxy data exists. This is especially the case when a qualitative legal issue is indicative of strength.
Limitations: Less useful in technology fields in which patent counts are poor indicators of innovation strength.	Limitations: Cost. Only practical with smaller datasets.	Limitations: Datasets are far too large. Any benefit from a patent expert technique would be offset by huge costs. If the technology field is large, it would likely be impossible to process enough data to make meaningful conclusions with patent experts.
Example: Commonly done to assess the comparative strength of countries. ⁴³	Example: May be necessary to subjectively categorise data; for example, creating sub-categories in a technology field. ⁴⁴	Example: None found

⁴² Ibid.

⁴³ See for example, VGR C Govinjaraju and C-Y Wong, "Patenting Activities by Developing Countries: The Case of Malaysia" (March 2011) 33(1) *World Patent Information* 51.

⁴⁴ Lloyd and Blows, above n 14.

5.2.2 Defining the contours of a technology area

Landscapes are used not only to characterise a technology space across some dimension, but also to categorise the boundaries of the space itself or categorise technologies within the space. For example, law and policy-makers utilize landscaping to map out the contours of a new or emerging technology for the purpose of applying existing regulatory regimes or establish funding rules.

Table 5: Defining the Contours of a Technology Area

Computer-Based	Research Team	Patent Expert
Feasibility: • Insufficient	Feasibility: • Sufficient	Feasibility: • Sufficient
When Useful: N/A	When Useful: Technical experts are necessary. Delphi methods help to ensure that there is a consensus within the scientific community to whom the definitions are most pertinent.	When Useful: Particularly useful when legal and policy considerations come into play.
Limitations: Only deals with undefined parts of the technology landscape so it is unlikely the benefits of large processing datasets exist. The difficulty of the landscaper in defining the boundaries of the landscape subjectively renders any automated approach insufficient.	Limitations: Only problematic when the landscape requires a legal/policy dimension that the technical experts do not possess.	Limitations: None, as long as the patent experts are sufficiently trained in the state of the art and their opinions are reflective of the scientific consensus.
Example: Some automated techniques may be used, but subjective analysis will always be required to determine final definitions.	Example: None found, but this is an emerging area that is not as commonly conducted. For instance, policy-makers are beginning to recognise the problem of an ill-defined scope of patents granted in a new technology field that can hinder future innovation and commercialisation of new products. ⁴⁵	Example: The indeterminacy of the legal status of embryonic stem cell patents and the funding regimes for such innovation led to use of pluripotent stem cells which also gave rise to legal definitional issues. ⁴⁶

⁴⁵ T Bubela, G Hagen and E Einsiedel, "Synthetic Biology Confronts Public and Policy Makers: Challenges for Communication, Regulation, and Commercialization" (March 2012) 30(3) *Trends in Biotechnology* 3, 134-136. The authors discuss the legal and regulatory challenges facing a new technology, synthetic biology. While the authors do not conduct a formal patent analysis, it is foreseeable that technical expert patent analyses would be useful in defining the

5.2.3 Assessing the patent system

Academics and policy-makers utilise patent landscape analysis to assess the functioning of the patent system itself in order to see if the patent system encourages or hinders innovation within a particular technology field.⁴⁷

Table 6: Assessing the Patent System

Computer-Based	Research Team	Patent Expert
Feasibility: • Potentially Useful	Feasibility: • Potentially Useful	Feasibility: • Potentially Useful
When Useful: Cost-effective. Effective when strong correlative effects have been proven in the technology field.	When Useful: Less expensive than patent experts: can leverage student researchers; especially useful since datasets need to be large to make accurate conclusions. Useful when the technical experts can assess the ability of innovators to circumvent patents.	When Useful: Particularly useful because patent experts are best suited to determine what the appropriate scope of protection that the patent system ought to grant.
Limitations: Severely limited because most automated techniques such as patent counts and clustering do not give any information about the technical ability to circumvent. The analysis will always require some degree of technical expertise that may be best captured in a subjective coding framework.	Limitations: Technical experts lack knowledge of claims construction, legal language, and the current state of patent law; cannot guarantee the accuracy of an expert.	Limitations: High costs, especially difficult because the datasets will often need to be large.
Example: Many numerical indicators and visualization techniques applied broadly to a technology area have been used to assess the	Example: No formal examples found, but numerous expert commentaries attempt to use patent data to assess problems of patent thickets. ⁴⁹	Example: The role of patent thickets in human genome patenting has been a particularly contentious issue. A landscape of litigated

scope of technology areas, particularly when there is a rapid increase in applications and granted patents.

⁴⁶ Georgieva and Love, above n 20. While pluripotent stem cells are a well-defined concept from a scientific standpoint, the legal status of pluripotent stem cells is unclear and will have potential impacts on funding and regulatory policies. The authors analyse the first issued patents in this field in an attempt to predict the characteristics of a future landscape.

⁴⁷ See I Hargreaves, "Digital Opportunity: A Review of Intellectual Property and Growth" (May 2011) *Review of Intellectual Property and Growth* <<http://www.ipo.gov.uk/ipreview-finalreport.pdf>> for an assessment of the problem of patent thickets in the UK.

incidence of thickets and white space. ⁴⁸		patents was used to create a dataset that contained highly pertinent patents and was limited enough to be analysed by a patent expert. ⁵⁰ Legal commentaries without a formal patent-by-patent assessment are also common. ⁵¹
--	--	---

6 Conclusion

Patent landscaping has come of age with the availability of large patent databases and the increasing sophistication of search and analytical tools. While initial efforts were necessarily *ad hoc* in nature, the field has sufficiently developed to undergo normalisation to ensure the validity and utility of the resulting landscapes.

Evidence used to support policy formation, whether in respect of policy questions relating to human gene patents, to innovation policy more generally or business strategy needs to be robust, transparent and replicable. While most of the patent landscapes conducted to date may not meet these standards — although most still contribute useful insights — they provide the basis for harmonising methods in the future.

⁴⁹ G Clarkson and D DeKorte, "The Problem of Patent Thickets in Convergent Technologies" (February 2007) 1094 *Annals of the New York Academy of Sciences* 180.

⁴⁸ See for example, UK Intellectual Property Office, "Patent Thickets: An Overview" (November 2011) <<http://www.ipo.gov.uk/informatic-thickets.pdf>>. This landscape was released as a response to the Hargreaves Report to assess the comparative impact of patent thickets across of a variety of technology areas.

⁵⁰ C M Holman, "Impact of Human Gene Patents on Innovation and Access: A Survey of Human Gene Patent Litigation" (2007) 76 *University of Missouri Kansas City Law Review* 295.

⁵¹ See for example, D Harris and R Bawa, "The Carbon Nanotube Patent Landscape in Nanomedicine: An Expert Opinion" (September 2007) 17(9) *Nanotechnology Law & Business* 1165.