World Patent Information 39 (2014) 3–10

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

Patent landscape analysis: A methodology in need of harmonized standards of disclosure

Emily Grant [a], Megan Van den Hof [a], E. Richard Gold [a, b, *, 1]

[a] Faculty of Law, McGill University, Canada
[b] Department of Human Genetics, Faculty of Medicine, McGill University, Canada

ARTICLE INFO

Article history:
Available online 15 November 2014

Keywords:
Patent landscaping
Patent landscape analysis
Patent analysis
Patent informatics
Patent trends
Intellectual property analysis

ABSTRACT

Statistical analyses based on patent data appeal to researchers and policy-makers from diverse fields because patents can serve as quantitative indicators of phenomena that are difficult to measure, including innovation, knowledge spillovers, collaboration, and technological space. This article conducts a comprehensive literature review of studies employing patent landscape analysis—assessing research objectives, sources of patent data, dataset sizes, the structured or unstructured nature of data, and modes of data interpretation—in order to provide an overview of methods of patent analysis across research fields. This article concludes that, while studies in certain fields seem to follow consistent norms of methodological disclosure, these norms are not universal. It emphasizes a need for greater awareness of the limitations of patent data, for improved transparency through harmonized standards for the disclosure of patent methodology, and for cross-disciplinary sharing of best practices in order to develop methodologies tailored to specific research objectives.

1. Introduction

The increase in patent applications over the past decade has expanded the pool of patent data available for analysis [1]. Knowledge-intensive sectors such as information technology, nanotechnology [2,3], and molecular biotechnology [4] have seen extensive patenting. At the same time, economic and management analyses have embraced the use of patent data as a measure of innovation [5], knowledge spillovers [6], and collaboration [7,8], primarily through counts of patent applications and granted patents. Researchers and policy-makers from diverse intellectual fields have deployed patent analyses to pursue a variety of objectives. This article provides an overview of methods of patent analysis and finds that, while studies in certain fields seem to follow consistent norms of methodological disclosure, these norms are not universal.

New methods of searching patent documents have stimulated large-scale analyses of thousands, or even millions, of patents at a time. Computerization has also enabled researchers to precisely pinpoint particular patents, allowing for accurate, small-scale analyses. Patents appeal to researchers and policy-makers because they serve as a quantitative indicator of phenomena that are difficult to measure, including innovation, knowledge spillovers, collaboration, and technological space.

Despite the quantitative nature of patent data, there remain significant limitations to its use as an empirical measure. One limitation of patents as a measure of innovation, for example, is that patenting rates differ between countries [9], firms [10], and industries [11], thus engendering inter-firm and inter-sector bias.[2] The use of patent data as a measure of knowledge flows through citation analysis may be unreliable because patent examiners often add citations to earlier patents of which the inventor may have been unaware [12]. Finally, the accuracy of a technological landscape—an assessment of patent protection surrounding a given technology—will depend on the legal and scientific expertise of the person performing the analysis of the technological space.

Given these and other limitations on the use of patent data, it is important for patent analyses to conform to some methodological

* Corresponding author. Faculty of Law, McGill University, Quebec, Canada.
E-mail address: richard.gold2@mcgill.ca (E.R. Gold).

[1] The authors' research was supported by Value Addition through Genomics and GE3LS (ValGEN) a project sponsored by the Government of Canada through Genome Canada, Genome Prairie, and Genome Quebec. The authors also gratefully acknowledge Andrew Baker and Rachel Tonelli-Zasarsky for their initial data collection and analysis. The authors also thank Gregory Graff for his helpful comments on an earlier draft and Danielle Lewensohn for a preliminary study that assisted them in designing the present study.

[2] Elsewhere, one of us has noted that inventors in the United States and Japan are more likely than those in other countries to patent their inventions [9]. Bessen notes that some firms, such as Oracle, deliberately seek to patent less extensively as a business strategy [10]. Cohen, Nelson, and Walsh note that patenting levels differ between industries, with those focused on discrete technologies (e.g., chemicals) patenting more frequently than those focused on complex processes (e.g., telecommunications) [11].

http://dx.doi.org/10.1016/j.wpi.2014.09.005
0172-2190/

norms and common protocols. Patent analyses must be conducted in a consistent manner and with a full understanding of the restrictions implicit in patent data. Research elsewhere has suggested that patent analyses are in danger of being *ad hoc* [9]: studies may give little justification for their chosen methodology or may not justify the choice by reference to the needs and specifications of the precise research question that is being asked. There may be little attempt to follow common protocols so that results can be validated and compared. To suggest that patent analyses ought not to be *ad hoc* does not imply that methods of patent analysis cannot be flexible and responsive to researchers' needs. Rather, we argue that patent analyses could be made more valuable and robust through the development of methodological protocols and best practice guidelines that take into account the limitations of patent data in light of the particular research objective pursued.

A researcher's choice of methodology will depend on the purpose of the analysis, the availability of data (and its specificity), software and methods, and the tolerance for error. However, appropriate disclosure is important to ensure that quantitative patent observations can be validated and duplicated.

The present paper expands on the approach outlined by Lin et al. [13]. Our research moves beyond the narrow focus on the field of technology management and conducts a literature review of patent analysis research across a greater variety of disciplines. Although other research has drawn conclusions about various types of patent data (i.e., patent count, patent claims, patent licensing data, patent citations), few studies have examined all forms of patent analysis across all disciplines.

In Section 2, we set out the definition of patent analysis used in this article. In Section 3, we outline our methods. In Section 4, we present our results, including qualitative and quantitative findings from our literature review. Finally, in Section 5, we discuss the implications of our study regarding standards of disclosure of methodology in patent analysis.

2. Definition

For the purposes of this article, we adopt the following definition of patent landscape from Bubela et al. [14].

> A landscape is an analysis of the relationships between multiple sets of indicators or of those indicators measured against temporal, technical or spatial dimensions. Indicators might include scientific articles, patents, clinical or field trials, regulatory approvals, and actors or institutions. Additional analyses can represent network connections or the density of clusters of scientific or technological fields.

This definition captures the various types of patent analysis observed in the literature, which range from simple patent counts to complex assessments of a patent's scope or validity [15] or of patent citation networks [16].[3]

3. Methods

3.1. Search

Our literature search was conducted using the EBSCO meta-database. We selected EBSCO because it contains a full range of patent analysis contained in both peer-reviewed and non-peer-reviewed publications. We searched the 79 English-language databases available to us through EBSCO as of September 2012.

In order to obtain a qualitative sense of the many uses of patent analyses and associated methods across multiple disciplines, we developed a set of key search terms by drawing on the expertise of our research group, which includes experts in law, management, economics, and science policy. We selected broad and varied terms with the goal of ensuring our results did not favor a particular discipline. Our key terms consisted of the following: patent landscape, intellectual property landscape, intellectual property mapping, patent analysis, patent data analysis, freedom to operate analysis (within 5 words of the word patent),[4] patent mapping, patent trends, patent search, patent indicators, patent statistics, patent citations, patinformatics, and patentometrics.

The search required that each document include the word "patent" and be written in English. Finally, our search captured articles published in the five-year period between October 1, 2007, and September 30, 2012. Our search included any article that contained any of our key terms in its full text.

3.2. Relevancy screening process

The search generated 3360 results. Of those results, 524 articles were considered relevant. The process of screening for relevancy was undertaken in two steps.

In the first step, we reviewed the abstract of each article and eliminated 2167 articles through a conservative application of our relevancy criteria (see Section 3.3). The remaining dataset consisted of 1193 articles.

In the second step, which was conducted simultaneously with data collection, we reviewed the full content of each article and eliminated a further 669 articles, as described below.

We excluded all duplicate results ($N = 48$). Articles were considered duplicates if two articles by the same author conducted the same patent analysis. We excluded all results where the full patent analysis was inaccessible through the EBSCO meta-database or through the McGill University library holdings.[5] Specifically, we excluded press releases referencing commercial reports from various publishers, all of which required a substantial subscription fee ($N = 430$). Nevertheless, we contacted these publishers to confirm that excluding their reports would not affect our overall conclusions. We were informed that the reports deployed *ad hoc* methods that would have led us to exclude them as not being systematic, in accordance with our relevancy criteria, discussed below. Last, 191 articles were excluded using our content-based relevancy criteria. The final dataset therefore included 524 relevant articles.

3.3. Relevancy criteria

We included a broad variety of sources in our review. As our goal was to evaluate the full range of uses of patent analyses, we examined academic sources (e.g., articles and book chapters) and non-peer-reviewed sources (e.g., periodicals, conference

[3] See Heines [15] for an example of a complex landscape of a particular technological field. See Oxley and Wada [16] for an example of a complex patent citation analysis.

[4] Despite the inclusion of "freedom to operate" within the search algorithm to ensure the comprehensiveness of the dataset, any article actually focused on freedom to operate was excluded, following Bubela et al. [14].

[5] Note that we included 23 search results from *World Patent Information*, initially excluded from the dataset by EBSCO's "full-text" search filter, as the full text of these articles was nonetheless available through the McGill University Library Holdings by following a link provided on the EBSCO database. These supplementary search results were added because of the authors' actual knowledge that *World Patent Information* contained many important articles on the topic.