Guidelines for Preparing Patent Landscape Reports

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

2015

Guidelines for Preparing Patent Landscape Reports

Guidelines prepared
for the

World Intellectual Property Organization (WIPO)

by Anthony Trippe,
Patinformatics, LLC

With contributions from WIPO Secretariat

2015

Chapter 1: Executive Summary

Patent Landscape Reports (PLRs) support informed decision-making, and are designed to efficiently address the concerns associated with making high stakes decisions in various areas of technology, increasing the related degree of confidence. For many years decision-makers operated based on personal networks and intuition. With the institution of patent analytics, and PLRs, it is possible for these critical decisions to be made with data-driven, evidence-based approaches that deliver informed choices, and mitigate the associated to the decision risks.

The insight gained from the preparation of a patent landscape report can be applied to almost any organization engaged in the evaluation of technology, and its impact on society. Government agencies, as well as private enterprise can gain valuable perspective on a developing, or well-established field by generating a PLR.

As an example, PLRs can be used as instruments to inform public policy makers in strategic decisions to related to R&D investment, prioritization, technology transfer or local manufacturing. Patent information can and is increasingly being used as a tool to inform public policy: Policymakers who deal with innovation have increasingly focused on the patent system. They look for clearer, more accessible and geographically more representative information to support key policy processes. They seek a stronger empirical basis for their assessments on the role and impact of the patent system in relation to key areas.

While PLR are undoubtedly useful instruments for informed decision-making, producing one can be a time-intensive and expensive process. An organization willing to devote the resources necessary to generate a PLR often does so when they are preparing to make a significant monetary or headcount investment in developing or moving into a technology area. It is critically important to make certain that a PLR is prepared properly in order to ensure that the insight it provides is accurate, and directed towards the key issues associated with technological implementation.

This document provides details on the stages required for the preparation of a PLR. It is provided as a means for instructing new practitioners on the steps required in generating a PLR, but it is also useful for recipients of PLRs having such reports produced for them in-house or by third-party service providers. It is also hoped that providing a template for the preparations of PLRs will help continue the development of standards, and best practices that can be used, and built upon by the patent analysis community.

By understanding the processes involved in preparing a PLR all of the varied stakeholders, providers, and users of the information they contain will have a better understanding of what can be expected from a PLR, and how many resources will need to be involved in its creation.

The present Guidelines were developed in the framework of the Development Agenda project "Developing Tools for Access to Patent Information" (DA_19_30_31_02), in particular Phase II of the project which was approved by the Committee for Development and Intellectual Property at its tenth session in November 2012[1]. It is also intended to be used as resource for patent information users in general, and in particular for capacity building activities on patent analytics at WIPO's Technology and Innovation Support Centers (TISC)[2].

---

[1] For the project document see http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=219002
2 http://www.wipo.int/tisc/en/

Chapter 2: Table of Contents for Patent Landscaping Reports Guidelines

Chapter 3: Introduction

Patents provide a right to exclude others from operating in the area the invention is being claimed for. Thus, they have business and legal implications associated with them. Receiving a patent can be a reasonably expensive endeavor costing from $10,000 on the low-end to five to ten times that for more complicated applications. Since there are substantial costs associated with them, when organizations pursue a patent it is generally an indication of high interest and potentially significant investment by them in the subject.

Patents are also critical sources of information that may not be found anywhere else. A paper from 1986 citing a report from 1977[3] claims that 80% of the information found in patents is not found elsewhere. It is extremely difficult to quantify a value in this way, but it is generally accepted that due to the nature of novelty, associated with patents, and the general practice of most organizations to not publish their findings in scientific literature, that patent information is a source of unique content, complementary to non-patent literature.

While it can be difficult to work with, and misleading, if not handled correctly, patent data is critical to a thorough understanding of most technological areas. Jacob Schlumberger best encapsulated these feelings in 1966[4] when he wrote:

"We have the choice of using patent statistics cautiously and learning what we can from them, or not using them and learning nothing about what they alone can teach us."

Due to the critical nature of patent documents and the information associated with them, they are used in a variety of different business contexts. There are different reports affiliated with providing information on patent data in these different environments. These guidelines are focused on the use of patent information, i.e. the information included in patent documents and related to a patient's lifecycle, to generate Patent Landscape Reports (PLRs).

There is no single or universally accepted definition of a Patent Landscape Report; in general, one can say that it constitutes an overview of patenting activity in a field of technology, in a specific geographical area. A landscape normally seeks to answer specific policy or practical questions and to present complex information about this activity in a clear and accessible manner for audiences with different background. Industry has long used patent landscapes to make strategic decisions on investments, research and development (R&D) directions, and competitors' activity as well as on freedom to operate in introducing new products. Now, public policymakers are increasingly turning to landscaping to build a factual foundation before considering high-level policy matters, especially in fields such as health, agriculture and the environment.

These guidelines provide details on the stages involved in the preparation of a PLR. This is provided as a means for instructing new practitioners on the steps required in generating a PLR, but it is also useful for individuals who will have PLRs produced for them by third-party consultants. It is hoped that providing a template for the preparations of PLRs will help continue the development of standards, and best practices that can be used, and built upon by the patent analysis community.

The present document begins with a primer on patent information, what portions of it are generally used for the production of PLRs, and how PLRs are distinguished from other reports that use patent information. Patent information professionals are well acquainted with the majority of this information,

---

3 http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=7162811
4 Invention and Economic Growth, Jacob Schmookler. Cambridge, Mass., Harvard University Press, 1966.

but for those who are new to the profession, or who will be receiving PLRs generated by others this section provides a starting point for understanding the nuances of patent information and analysis.

The next section, covered in chapters 5-7 discusses the philosophy behind the preparation of patent landscape reports including a discussion of the motivations and objectives for creating them, the tasks generally associated with performing patent analytics, and frameworks, or concepts that practitioners should consider as they are beginning a PLR project. If the first section was a backgrounder for patent information, then these chapters can be considered as background information on performing patent analytics, as it will be applied to the creation of a PLR.

Chapter 8 provides step-by-step instructions for the preparation of PLRs. It begins with a section on the planning required before a project is initiated. A section follows this on the dynamics associated with conducting a search in association with a PLR. With an understanding of what the objectives will be for the report, and with search results in hand a discussion of the pre-processing steps associated with patent data is provided. A list of the various analytics normally included in PLRs is covered, as well as suggestions for writing, publishing and evaluating the report once the analysis work is complete and conclusions need to be shared.

The final section covers chapters 9-12, and includes lists of resources, providers and literature references associated with the subject of patent landscape reports. There is also a chapter on additional topics related to the strategic use of patent information other than PLRs. This Section was included to provide an introduction to additional applications of patent information that may be of interest to organizations that are also interested in PLRs for strategic technical planning.

After reading these guidelines interested practitioners will have the information they need to begin preparing their own PLRs in support of key decision-makers. During the development of this document two workshops on its contents were conducted with patent office practitioners from Latin America and South East Asia. The level of interest in developing patent analytics and PLRs for developing countries was extremely high, as reflected on the results of related surveys, and almost universally the workshop participants felt that the information in the guidelines was essential to help them begin preparing their own reports.

Chapter 4: Basics of Patent Information

The primary information source for Patent Landscape Reports (PLRs) is data coming from patent documents. Additional information is sometimes used from other sources, such as the non-patent, scientific literature, but patent data is used most frequently for the analysis that makes up the majority of the insights identified for the report.

A general understanding of patent information is critical to producing well-researched PLRs, since raw patent data is notoriously difficult to work with, for a variety of reasons. In particular, the variety of publication policies applied by different jurisdictions which partially derive from differences in patent prosecution. Understanding the nuances associated with patent information will help prevent an analyst from misinterpreting the data and come to incorrect conclusions.

This chapter provides background on patents as a type of intellectual property, as well as looking closely at the various types and parts of patent documents, especially those that are typically utilized in the generation of PLRs, supplementary information associated with each patent application and the sources of patent information (databases) that can be used to prepare a data collection to be analyzed[5].

4.1 – Why Analyze Patent Information?

Patents are intellectual property rights for the protection of an invention in the territories of individual jurisdictions which may be granted in exchange for disclosure of the invention[6]. Since a granted patent represents a right to exclude others from making, using or selling the invention in the specified jurisdiction, it has a business value associated with it. Patents are sometimes referred to as a "limited monopoly" based on their ability to prevent competitors from entering a market or making use of a patented technology. Due to the potential business and legal implications, understanding which organizations own patents, and what technological areas they cover, can have a significant impact on policymaking and corporate decision-making.

Obtaining a patent can be a reasonably expensive endeavor, costing from $10,000 on the low-end to five to ten times that for more complicated applications. Due to the substantial associated costs, when organizations pursue a patent - in particular, if in a plurality of jurisdictions - it is generally an indication of high interest and potentially significant investment by them in the subject.

---

[5] General information on patent information is included in the WIPO Handbook on Industrial Property Information and Documentation: http://www.wipo.int/standards/en/index.html (henceforth called the 'WIPO Handbook').

[6] The legislation of each jurisdiction usually defines the Intellectual Property rights available for the protection of inventions. They may include different instruments such as patents and utility models, and they may use varying designations for such instruments, such as "patent", "petit patent", "inventors' certificate", etc. Several international treaties deal with such Intellectual Property rights for the protection of inventions. They use the term patent as comprising all these rights irrespective of their designations in the legislation of a member states. Similarly, the term patent is used in these Guidelines as comprising all such instruments. For general information on the patent system, see http://www.wipo.int/patents/en/; or the WIPO Intellectual Property Handbook (not to be confused with the Handbook mentioned in the previous footnote): http://www.wipo.int/export/sites/www/freepublications/en/intproperty/489/wipo_pub_489.pdf

Patents are also critical sources of information that may not be found anywhere else. A paper from 1986 citing a report from 1977[7] claims that 80% of the information found in patents is not found elsewhere. It is extremely difficult to quantify a value like this, but it is generally agreed that due to the nature of novelty, associated with patents, and the general practice of most commercial organizations to not publish their findings in journal literature, that patent information is a source of unique content.

While it can be difficult to work with, and misleading, if not handled correctly, patent data is critical to a thorough understanding of most technological areas. Jacob Schlumberger best encapsulated these feelings in 1966[8] when he wrote:

"We have the choice of using patent statistics cautiously and learning what we can from them, or not using them and learning nothing about what they alone can teach us."

This statement crystalizes the essence of why patent information is analyzed. Section 4.6 below elaborates on different types of such analyses and their respective objectives.

4.2 – Types of Patent Documents and Publication Policies

The specific rules for applying for a patent and for processing patent applications, including their publishing, can be different and should be considered on a jurisdiction to jurisdiction basis. Patentable subject matter is also different between various jurisdictions. Most jurisdictions have a system in place with substantive examination, i.e. where the claimed technical subject matter is examined whether it meets certain conditions for patentability, such as novelty, inventive step and industrial applicability. In such systems it is customary to distinguish a pre- and a post-grant prosecution phase. The below distinction between different publication stages related to a single application, i.e. pre-grant, grant and post-grant publications, applies mostly to such systems.

Few jurisdictions have a mere patent registration system in place, i.e. without regular substantive examination. Such systems are similar to utility model systems. For such systems the below distinction between pre-grant and post-grant publications does not apply.

Depending on the jurisdiction, and in particular its publication policy, there are various types of patent documents published at various stages during the lifecycle of a patent application. All these publications associated with an individual application constitute a so-called domestic patent family (see section 4.4.5 below).

With most patenting authorities, patent applications are published for the first time 18 months after their priority or filing date, even if they have not yet been granted. If they are granted another publication follows the first publication which includes the claims granted by the patenting authority.

Some jurisdictions publish only granted patents. In such cases, pending applications may not be known to the public until the publication of the grant, and in fact may never become known if the application fails during examination or is withdrawn.

Some jurisdictions do not publish all parts of an application or a granted patent, but rather a notification in a gazette or bulletin. In such cases, the disclosure and claims become publicly

---

[7] http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=7162811
[8] Invention and Economic Growth, Jacob Schmookler. Cambridge, Mass., Harvard University Press, 1966

accessible after the publication of the notification, e.g. through file inspection (see below) or through ordering a (certified) copy.

It is important to understand the difference between the official publication of patent documents or of gazettes, and the making publicly available of at least parts of applications or other documents.

Any published patent document is identified by a unique publication number and its content is usually fixed with the publication on the particular publication date. Subsequent publications related to the same application, i.e. being members of the same domestic family, are usually distinguished by kind codes (see below) as parts of the publication number. For some jurisdictions, these subsequent publications related to the same application are only distinguished by using different kind codes (e.g. publications of the European Patent Office). In other jurisdictions, however, these publications belonging to the same domestic family have distinct publication numbers, while the publication stage is still identified by the respective kind code (e.g. publications of the Unites States Patent Office or the Japan Patent Office).

Understanding the difference in national publication policies may be important for certain analyses and the conclusions drawn, e.g. if data related to pending, withdrawn or lapsed applications cannot be researched, and if only publications of grants reflect the innovation activity.

4.2.1 – Pre-grant Publications

The process of generating a patent right starts with the first filing of an application with a national or regional patent office or with WIPO (namely, the International Bureau of the PCT). This office is sometimes referred to in patent analytics as the Office of First Filing (OFF).

Often the same invention (or an improvement thereof) is filed subsequently with other patent offices to obtain protection in further jurisdictions, usually by claiming the priority of the first filing. These offices are called Offices of Second Filing (OSF). Such second filings lead to the creation of patent families and associated relations between patent family members, which are further explained below in section 4.4.5.

Some patent authorities keep patent applications secret until they are granted, but most authorities publish patent applications 18 months after their filing date, or the priority date, if the office is an OSF[9]. These documents are called pre-grant applications and they don't represent a granted right in their present form, but may be granted in the future. They provide clues on investment and interest in a technological area, and how the environment around a technology may change, if the application goes on to grant.

Depending on the national publication policy, pre-grant publications may also comprise separate publications of search reports or corrections[10]. Different such pre-grant publications related to the same application can usually be distinguished by their publication kind codes as part of the publication number. For the preparation of statistical analyses such publication policies may have to be taken into account.

---

[9] In some jurisdictions the applicant can request earlier publication; this option is often chosen for defensive publications.

[10] For instance, that is the case of the European Patent Office (EPO). More information available at
http://www.epo.org/applying/european/Guide-for-applicants/html/e/ga_d_iii.html

Publications of OSFs are often fully equivalent to the publication of the OFF and represent mere translations. It should be noted however that this is only a rule of thumb, because the Paris Convention expressly permits additions to the disclosure of the first filing when claiming the priority rights of earlier filings.[11] In particular, if two or more priorities are claimed in a second filing, it is very likely that the claimed subject matter is somehow different from the individual priority applications.

It is important to recognize that pre-grant applications, while potentially important indicators, are not granted, and in fact, may never be granted. Applications can be abandoned or withdrawn during prosecution for a variety of different reasons; but the primary reason is that an examiner has stated objections in an office action. Once an application has been abandoned any subject matter disclosed within it is now part of the public domain of the jurisdiction where it was abandoned and can be used by others, assuming that other granted patents don't exist on the same subject. Abandoned applications still represent interest on the part of the applicant and can still provide valuable insights even if they don't represent a property right. On the other hand, large numbers of applications that do not make it to grant may also be an indicator that there are incentives in place for filing applications.

Understanding the difference between granted patents and pre-grant applications is critical for interpreting their impact on a field. In the development of analytics associated with PLRs it is good practice to separate pre-grant applications from granted patents when conducting an analysis, e.g. by using kind codes (see section 4.2.4 below). The implications of pre-grant applications are significantly different than what can be implied from a granted patent and they should be considered separately, or at the very least, identified as a different type of document when visualizing a result.

Some pre-grant publications can be considered as defensive publications since they were not filed with the intention to obtain patent protection but rather to prevent others, e.g. competitors, from obtaining patents on the technical subject matter disclosed in the application[12]. With other words, the intention of these filings is to place their technical disclosure in the public domain for free use by anybody. It is however not readily possible to distinguish such pre-grant publications from others where the applicant seeks protection.

4.2.1.1 – PCT Applications

While pre-grant publications of applications are usually associated with specific jurisdictions and therefore may represent an indicator where innovation takes place (in case of OFFs) or where patent protection is sought (in case of OSFs), there is also a special type of patent application that facilitates the filing of applications in many jurisdictions simultaneously.

---

[11] Article 4 (F) provides that "No country of the Union may refuse a priority or a patent application on the ground that the applicant claims multiple priorities, even if they originate in different countries, or on the ground that an application claiming one or more priorities contains one or more elements that were not included in the application or applications whose priority is claimed, provided that, in both cases, there is unity of invention within the meaning of the law of the country.With respect to the elements not included in the application or applications whose priority is claimed, the filing of the subsequent application shall give rise to a right of priority under ordinary conditions."

[12] Before the America Invents Act (AIA), which was signed into law on September 16, 2011, the US system knew the so-called **Defensive Publication (DEF)**, and the **Statutory Invention Registration (SIR)** which replaced the Defensive Publication in 1985-86. The AIA repealed these provisions because all pending applications are now published 18 after filing or priority date.

The Patent Cooperation Treaty (PCT), which has effect in 148 jurisdictions (as of July 2015), provides for a system that
- permits the applicant to lodge a single application with a Receiving Office of the PCT system; and
- to obtain a search report and written opinion from an International Searching Authority of the PCT system, and eventually a Supplementary International Search and/or an International Preliminary Examination which assesses the patentability, and thereby
- enables the applicant to take an informed decision if and for which countries he will seek protection;
- grants the applicant a period of 30 months (in most member jurisdictions) to seek patent protection in each of those jurisdictions (national phase entry); which
- allows for more time to assess the commercial viability of the invention; and delays the considerable expenses associated with pursuing the patent prosecution in those jurisdictions (such as translation, legal representative, national fees).

The Patent Cooperation Treaty and the related services are administered by the World Intellectual Property Organization (WIPO) which publishes respective PCT applications (also referred to as WO documents). WIPO is responsible for administering the PCT system and for ensuring that PCT applicants receive an initial prior art search report and a written opinion[13] regarding the potential patentability of the subject matter claimed in the application against prior-art from around the world.

Based on the benefits attributed to the PCT system, WIPO is often used as an OSF on many applications where the idea of protection on a more global scale is being considered. To a lesser extent it is used as an OFF. Due to their popularity PCT applications are an important source of patent information for PLRs.

If a PCT application enters a national phase, there may be subsequent pre-grant publications of these national phase entries (NPE), depending on the respective publication policy of each jurisdiction, however only with a certain delay because the NPE usually becomes effective only 30 months after the filing or priority date of the PCT application.

4.2.2 – Granted Patents

Publications of granted patents are of particular importance in comparison to publications of unexamined applications because the grant asserts that the invention disclosed in the application is indeed new and inventive over the known prior art. A grant can therefore be taken as a quality indicator for innovation activities. The time of grant of patents, i.e. the publication date of the grant, depends however very much on the pendency of patent examination and can differ considerably from jurisdiction to jurisdiction or also for certain areas of technology. Publications of granted patents may usually be identified by a specific kind code as part of the publication number (see below).

4.2.3 – Post-Grant Documents

---

[13] As from July 1, 2014, the written opinion is made publicly available on PATENTSCOPE in its original language as of the date of publication of the international application

There are a number of additional patent documents that may be published after the publication of the grant of a patent. The most important ones are publications following reexamination or opposition procedures that were initiated by third parties after the publication of the grant. If, as a result of these procedures, the scope of protection was restricted, a new publication would be made including the modified claims. That is the case, for instance, for the EPO, where if the patent is maintained in an amended form, a new patent specification is being published.

Similarly, patent documents would be reissued if the patent owner on his own initiative wishes to restrict the scope of protection of the patent in order to escape an imminent reexamination or opposition procedure. A third, although less important, reason for post-grant publications may be the correction of clerical or typographical errors.

These post-grant documents, while they may impact the scope or the term of a granted patent, are normally not considered when collecting a data set for PLR related analysis. If these documents are included in a corpus it is generally a good idea to filter them out before conducting the analysis. A notable exception to this is when an analysis of claim language is being conducted. In this case, reissue and reexamination documents can modify the original, granted claims and under these conditions they should replace the original patent.

4.2.4 – Kind Codes

It was already repeatedly mentioned that different publication stages of an application are usually distinguished by kind codes (e.g., A1, A2, A3, B1,..) which are part of the document publication numbers.

The WIPO Handbook provides the following definition for kind codes of patent document:

Several countries and organizations publish patent documents for various types of protection possible within their jurisdiction. Furthermore, according to certain laws or regulations, patent documents may be published at various stages of the procedure leading from the application for a given industrial property right to its final grant (or refusal). Thus, for certain countries and organizations, various "kinds of patent documents" exist, which may be characterized by the specific type of protection to which they refer and by the stage of the administrative procedure at which they were published.

For more details and a complete list of the kinds of patent documents issued by each patent authority see Part 7.3.1 of the WIPO Handbook, and for an inventory of kind codes per issuing patent authority see Part 7.3.2 of the WIPO Handbook.[14]

The WIPO Standard ST.16 provides for a basic standardization of kind codes. It should be noted, however, that patent authorities do not use kind codes in a fully standardized manner because of differences in their publication policies. For example, for the EPO and WIPO, the kind code 'A1' designates publications of patent applications with a search report; while for the USPTO, it designates publications of patent applications without a search report since the USPTO does not publish search reports. For such publications of patent applications without a search report, the EPO and WIPO would use the kind code 'A2'.

It should further be noted that the kind codes used by each issuing patent authority may have changed over time; for example the USPTO used kind code 'A' for the publication of granted patents

---

[14] http://www.wipo.int/standards/en/part_07.html#7.3

through December 2000, and kind codes 'B1' and 'B2' as from January 2001. These changes of the use of kind codes are also described for each patent authority in Section 7.3.2 of the aforementioned WIPO Handbook.

4.3 – Components of Patent Documents

While patents documents contain a good deal of raw text, they are referred to as semi-structured, since they have a number of sections that are found in almost every document, regardless of its country of origin. At a high level these sections of a patent document are represented by a Front Page with bibliographic data, a Description (Disclosure) and a Claims section. Within each of these high-level sections there are subsections that provide specific information about the particular document. These subsections are typically segmented into individual fields when the documents are processed for electronic delivery or the generation of databases.

An additional Drawings section is facultative, but often included to illustrate the description and facilitate the interpretation of the claims. In some jurisdictions, the publication of an application further includes a search report as a further section of a patent document when the search report is available at the time the publication is prepared. Else, the search report may be published as a separate document at a later time once it has been established.

4.3.1 – Front Page and Bibliographic Data (Metadata)

The WIPO Handbook provides the following definition of bibliographic data:

The term "bibliographic data" denotes the various data normally appearing on the first page of a patent or industrial design document or in a comprehensive entry in an official gazette concerning granted patents, industrial design or trademark registrations or the corresponding applications. Such data comprise document identification data, data on the domestic filing of the application, priority data, publication data, classification data and other concise data relating to the technical content of the document or of the entry in the official gazette.

The majority of the statistical analysis conducted on patent collections takes place using data collected from the bibliographic fields within them. Many of these fields contain categorized text or numbers and thus are readily applicable to statistical analysis (see also section 7.1 below).

To assist with working with this data across different jurisdictions and languages, an international standard for bibliographic data within patent documents, called INIDs has been developed by WIPO. The WIPO Handbook provides the following definition of INIDs:

INID is the acronym for Internationally agreed Numbers for the Identification of Data. The INID codes are numerical codes allotted to bibliographic data relating to industrial property documents and printed on their first page and in corresponding entries of Official Gazettes.

INID codes are standardized by the World Intellectual Property Organization (WIPO) in ST.9[15] which includes a complete list of the INID codes. A few of the bibliographic fields, used most frequently in the statistical analysis of patent collections, for PLRs are provided below.

---

[15] http://www.wipo.int/export/sites/www/standards/en/pdf/03-09-01.pdf

4.3.1.1 – Applicant/Assignee

The WIPO Handbook provides the following definition of applicants:

The applicant is the entity or person which or who presents ("files") an application for the grant of an industrial property right (e.g., a patent application, or an application for the registration of a trademark) in an industrial property office, or in whose name an agent (representative) files such an application.

In general, the applicant is the inventor, but it may also be the employee or the person to who the inventor assigned his/her right to the invention (assignee). Ordinarily, this will be a company or organization, but can be the inventors when the rights associated with the invention are not transferred, or assigned, to a different entity.

In the United States, an assignment is required because, the Constitution of the United States provides in Article 1, Section 8, that: the "Congress shall have power . . . to promote the progress of science and useful arts by securing, for limited times, to authors and inventors, the exclusive right to their respective writings and discoveries." In other words, the inventor, and not the organization that employs them receives a patent right. However, corporations can now apply for patents directly, without a formal re-assignment from the inventor under the American Invents Act[16]. Even with the new statutes most patenting in the United States is still done the traditional way; the rights are granted to an inventor, and are then assigned to the legal owner based on the stipulations of the employment agreement the inventor signed when joining the organization.

In the context of PLRs, the Applicant/Assignee represents the owner of a patent and with whom negotiations for the rights associated with the invention will have to be conducted. Studying them identifies investors within a technical area. Network analysis is frequently applied to identify collaborations, e.g. in certain technical fields.

Applicant and assignee names can change over the life cycle of a patent application whenever the rights in the invention are transferred. Marginal changes can also occur in case of clerical corrections of misspellings of names. A problem that frequently occurs with any names are variations of names that derive from transcriptions of names from other scriptures like Chinese if varying transcription rules are applied. One and the same person can thus be represented by slightly different spellings of his or her name.

Another frequently appearing problem in search and analysis is that subsidiaries of corporations often use varying names in different countries. Analyses that wish to cover complete paten portfolios need to take this into account and utilize various tools to tackle this, such as the so-called corporate trees. These gather various variations of an entity and their affiliations and group them together, in a more or less automated way.

4.3.1.2 – Inventor

The WIPO Handbook provides the following definition of an inventor:

---

[16] http://www.ladas.com/Patents/PatentPractice/AIA_Filing_Requirements.html

A person who is the author of an invention. According to Article 4ter of the Paris Convention, the inventor has the right to be mentioned as such in the patent.

The set of inventor names in the bibliographic data of an application should therefore be comprehensive, and unlike the names of applicants or assignees, the names of inventors usually don't change over the life cycle of a patent application (except for clerical corrections of misspelled names). In an application claiming the priority of an earlier application, inventor names may however be added if additional inventive subject matter is included in the later application which involved further inventors.

In the context of PLRs, the inventor represents the person or persons who are responsible for the intellectual effort associated with the invention. Studying information related to the inventors provides an idea about potential experts and leaders in an area of technology. Network analysis is frequently applied to identify collaborations between different inventors or groups of inventors, and institutions for which they work.

4.3.1.3 – Dates

Dates correspond to the timing of significant events in the lifecycle of a patent application. The three most significant patent related dates are the priority, the filing and the publication dates.

The filing or application date is determined by the patent authority that receives the application if certain minimum requirements are fulfilled, which actually differ from jurisdiction to jurisdiction. The filing date may therefore differ from the date the applicant lodges the application with the patenting authority.

The priority date (or dates if the priorities of several earlier applications are claimed) corresponds to the filing date of an earlier application if the applicant claims the priority of that earlier application. It is important because it may determine the relevant prior art if certain conditions are met.

Another important date is the publication date, which is the date when a patent document is published. Patent applications are published 18 months after the filing date or the earliest priority date in most patent issuing authorities.

In the case of granted patents the publication date is also referred to as the grant date. It is important because, in most jurisdictions, the protection provided by a patent enters into force with the publication of the grant.

In the context of PLRs, the dates represent the timing associated with the development or patenting of an invention and are used for analyzing trends. Studying filing or priority dates provides an indication of when inventions were developed and how long it took for improvements, and modifications to start occurring. Publication dates are less useful for this pupose. In particular, the grant date is rather an indicator for the pendency of applications until grant.

4.3.1.4 – Priority Data

The WIPO Handbook provides the following definition of priority data:

The part of the bibliographic data (normally published on the first page of a patent document) identifying the earlier patent application(s) on the basis of which a so-called priority right has been claimed (usually based on Article 4 of the Paris Convention). These identification data comprise three elements: the application number, the filing date and the identification of the country or organization where the respective earlier application was filed. Priority data belong to the basic bibliographic data of a patent document and may serve, inter alia, for identifying patent documents published in different countries and languages but referring to the same invention ("Patent Family").

4.3.1.5 – Classifications

The WIPO Handbook provides the following definition for patent classifications:

In patent information and documentation matters "classification" means a specific system which subdivides technology into distinct units. A classification symbol is defined for each of those units. The classification symbol designating the unit into which the invention falls is usually printed on the first page of the corresponding patent document and recorded in databases as part of the bibliographic data.

To "classify" a patent document means to determine that subdivision of a classification system to which, because of its technical nature, the invention claimed in the said document belongs and to allot a classification symbol to it. Sometimes, the classification relates not only to the claimed invention but also to other disclosures contained in the patent document.

In the past, different national classification systems were developed and applied to the patent publications of each respective country. In a first attempt to harmonize these systems, the International Patent Classification (IPC)[17] was created in 1968 which is nowadays applied to patent publications of almost all jurisdictions worldwide. Each patenting authority is obliged to classify the applications filed in its jurisdiction. The term bibliographic IPC was coined to address these classifications provided by the publishing authority and presented as part of the bibliographic data on the front page of the official patent publications. The classification by the publishing authority does however not prevent other patenting authorities from reclassifying these publications when they add them to their search file.

The IPC is regularly revised to include new technologies, or to divide existing classification places into several subunits with a more narrowly defined scope. Classification symbols are therefore usually accompanied by version indicators. With each revision, all patent publications belonging to the PCT Minimum Documentation are reclassified according to the new classification, and the updated classifications of documents are made available to database hosts. It is their responsibility to update the database accordingly.

In October 2010, the EPO and USPTO launched a joint project to create the Cooperative Patent Classification[18] (CPC) in order to harmonize their proprietary patent classifications systems, the United States Patent Classification (USPC) and the European Classification (ECLA). Like the former ECLA, the CPC is based on the IPC and provides a more detailed classification scheme in order to meet classification requirements of the EPO and USPTO. With the entry into force of the CPC all

---

[17] For more information on the IPC as well as revision and reclassification procedures see http://www.wipo.int/classifications/ipc/en/; the 2015 version of the Guide to the IPC is available at http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf .

[18] http://www.cooperativepatentclassification.org/index.html

patent publications of the EPO and the USPTO previously classified according to the former ECLA and USPC were reclassified according to the CPC. Hence, there is no need to search these older publications by using the former classifications, although many databases still permit such searches.

The CPC is also applied to patent publications of other jurisdictions either because these jurisdictions have opted for the CPC in addition to the IPC, or because of reclassification efforts of the EPO and the USPTO which complement the bibliographic IPC assigned by the publishing patent authority with additional relevant CPC classification codes in order to enhance the search efficiency of their examiners. In such cases, it should be noted that only a certain fraction of the publications of these jurisdictions is classified according to the CPC while all publications are classified according to the IPC. Since the EPO shares this CPC reclassification data with other patent database hosts, it can be used for searches in other database provided it was included by the host.

In addition to the IPC and the CPC, further classifications still exist and may be useful: the classification system of the Japan Patent Office (JPO), comprising the File Index[19] (FI), which is based on the IPC, and the F-Terms, which represent a multi-dimensional keyword system complementing the FI; and the Derwent classification system which was developed by a commercial database provider.

In the context of PLRs, the classification codes represent predefined concepts for describing the technical features or attributes associated with an invention. Often these concepts have a very narrow scope to facilitate focused prior art searches. For a broader analysis, e.g. of trends in wider areas of technology, documents classified by a range of classifications may need to be aggregated. For statistical analyses it should be borne in mind that only the IPC is applied to all patent publications of almost all jurisdictions, while the CPC is applied to some of them.

It should also be noted that patent documents often have multiple classification codes assigned. In some jurisdictions one of them is considered as main classification that most closely describes the claimed subject matter and also determines the unit in charge of examination. Nevertheless, these main classifications are not necessarily recorded as such in databases. They are often only highlighted as main classification on front pages of patent publications. In databases, multiple classifications are often ordered alphabetically. The first code should therefore not necessarily be interpreted as main classification. This may also bias certain classification-based analyses if only the first code is selected for the analysis.

4.2.1.6 – Citations

During the prosecution of a patent application, an examiner will look for prior-art related to the novelty, obviousness, or an inventive step, associated with an invention. When references of this nature are discovered they are cited within the document during different publication stages. Usually within a search report that accompanies the document.

In the United States, there is also a duty of "candor and good faith" that requires applicants to share prior-art with the USPTO during the examination of an application.[20] These documents are also

---

[19] http://www.jpo.go.jp/torikumi_e/searchportal_e/pdf/classification/fi_f-term.pdf
[20] 37 C.F.R. 1.56 "Duty to disclose information material to patentability, of the Manual of patent examining procedure" see http://www.uspto.gov/web/offices/pac/mpep/s2001.html for the text of the provision, http://www.patinformatics.com/blog/all-citations-are-not-the-same-exploring-examiner-citations-from-us-patent-documents-part-1-an-introduction/

citations, and they appear on the front-page of granted US patents along with the prior-art identified by the examiner.

Since they are associated with prior-art, or references that are potentially covering a similar topic as the proposed application, a citation implies a shared technological relationship between two documents.

Within this context, there is the concept of forward or backward citations. Any discussion of patent citations begins with a root document. This is the application being applied for in the discussion above. The references that the root document cites or references itself are referred to as backward citations, since they are references, which preceded or were published before the root document. Conversely, going forward in time from the root document, any more recent document which references the root document are referred to as a forward citation for the root document.

In the context of PLRs, the Citations represent a potential relationship between two inventions. Studying them provides a means for identifying seminal documents that could have had a high impact on the development of a technology.

4.2.2 – Description (Disclosure)

The WIPO Handbook provides the following definition:

The description of the invention is one of the essential parts of certain kinds of patent documents, e.g., patent applications or patents. It usually specifies the technical field to which the invention relates, includes a brief summary of the technical background of the invention and describes the essential features of the invention with reference to any accompanying drawings.

The patent system is built on the principle that protection of an invention is granted in exchange for disclosure of the invention in order to spur further innovation. The disclosure of the invention has to be clear and sufficient enough to enable experts in the field to carry out the invention. In this respect, it is the complement to the claims section of a patent application (see below) that defines the scope of protection. The description section of a patent document is therefore sometimes simply referred to as the disclosure. Specification is a further synonym for this part of a patent document.

From a patent landscaping perspective, the description section is one of the most difficult portions to interpret, since it contains information on the invention itself, as well as information on other inventions that are similar but were developed previously. From a text-mining and searching perspective this dichotomy within the disclosure can be misleading since it is difficult to determine whether the terms being searched for, or analyzed against, are referring to the invention, or the background information. Generally, it is not a good idea to conduct text-mining or analytics on the full-text of a patent document because of the ambiguity present in the disclosure. If possible, the disclosure is normally excluded in analysis of this type.

4.2.3 – Claims

The WIPO Handbook provides the following definition for claims:

The part of a patent document which defines the matter for which protection is sought or granted.

Each patent application has to include at least one claim. The first claim, the so-called main claim, is supposed to include all technical features of the invention that are essential to solve the technical problem that led to the invention and which is supposed to be solved by the invention.

The main claim is supposed to include only these essential features. Additional features or details that are not essential but provide certain benefits or additional advantages can be included in so-called dependent claims that refer to the main claim or any other claim.

In most jurisdictions the application or the granted patent may also include further so-called independent claims in addition to the main claim, i.e. claims that do not refer to other claims. That is possible if the invention, for example, not only covers a device or product but also a method or process that are based on the same inventive concept. Further independent claims may also be admissible if there are alternative ways of implementing or carrying out the inventive concept and if they cannot be described by a single independent claim. Such further independent claims are however only admissible as long as the principle of unity of invention is observed.

Claims determine the scope of each prior art search of the examiner since an examiner has to determine to what extent the claimed technical subject matter, i.e. the technical features of the invention as defined by the claims belong to the prior art. Technical features not included in claims but only in the description are usually not searched by the examiner. Claims may however evolve over the examination process, for example if an applicant adds or replaces features disclosed in the description part of the application to overcome objections by the examiner. The claims granted at the end of the examination procedure are usually much narrower than the originally filed claims that are included in the pre-grant publications.

There are specific rules that attorneys need to follow, when writing claims and, as such, they are not written in conversational English, and can be confusing to non-practitioners who are not familiar with how to interpret them. The uniqueness of claim language can also pose a challenge when performing text-mining or analytics, since most systems are developed, or trained using standard, or journalistic English and not the specialized, legal language of patent claims.

Regardless of the difficulties, understanding the scope of the claims associated with a patent document is an essential requirement for understanding what the patent covers and how valuable it could potentially be. If it can be said that, "the devil is in the details", then the details can be found in the claims.

With respect to PLRs, claim analysis is normally conducted as a follow on step, since it is sometimes done on a case by case basis, and only on documents that are identified as being in force, or of high interest based on the interpretation of the other analytics associated with the PLR.

4.4 – Publicly Accessible Supplementary Information Associated with Patent Applications

Besides looking at the structure of individual patent documents, it is also important to understand that patents exist within an infrastructure of additional information associated with the development of the document, what occurs with it as it matures, and how it relates to other documents that are associated with it. Some of this information can be incorporated into PLRs, while other details are only explored if additional details on a particular asset are of interest.

A great deal of this supplementary data is found in the National Registers associated with the prosecution of a patent document in a particular jurisdiction. As an example, the European Patent Office (EPO) describes the European Patent Register as follows[21]:

The European Patent Register contains all the publicly available information on European patent applications as they pass through the grant procedure, including oppositions, patent attorney/EPO correspondence and more. This service also provides for public file inspection.

Most patent issuing authorities keep all of this information in one place, but the United States provides three different databases for finding this data:

- Public PAIR – US Case Histories – http://portal.uspto.gov/pair/PublicPair
- USPTO Assignments – Re-assignments - http://assignments.uspto.gov/assignments/q?db=pat
- US Patent Maintenance Fees – https://ramps.uspto.gov/eram/patentMaintFees.do

4.4.1 – File Wrappers and Prosecution History

Patents, in the process of being examined, are prosecuted at a patent issuing authority. During the process, Office Actions and other procedural items take place between the patent office, or examiner, and the applicant with their attorneys. The documentation associated with the interaction between the applicant and the patent office is referred to as the prosecution history and the paper trail associated with it contained in a file wrapper or dossier. The case history contains details on rejections from the examiners, the responses from the applicants, any changes that are made to the language of the claims, and disclaimers and amendments, filed by the applicant, among other details.

An example of understanding patent file histories in the United States can be found at http://www.tms.org/pubs/journals/JOM/matters/matters-0302.html.

The file wrapper (dossier) becomes publicly accessible only after the patent application has been published (i.e. in most jurisdictions 18 months after the filing date oder the priority date. Not all jurisdictions however provide public file inspection.

4.4.2 – Maintenance Information

Maintenance fees or renewal fees are fees that are paid to maintain a granted patent in-force. Some patent laws require the payment of maintenance fees for pending patent applications. Not all patent laws require the payment of maintenance fees and different laws provide different regulations concerning not only the amount payable but also the regularity of the payments. In countries where maintenance fees are to be paid annually, they are sometimes called patent annuities. [22]

When conducting an analysis, for inclusion in a PLR, which examines the status of a patent document it is usually important to sub-divide granted patents in to those that are currently in-force versus those that have been allowed to go abandoned, or found to be invalid after re-examination. The documents in the latter two categories are no longer in-force, are effectively in the public domain, and available for use by others. While status can be determined by looking at maintenance data, calculating term

---

[21] http://www.epo.org/searching/free/register.html

[22] https://en.wikipedia.org/wiki/Maintenance_fee_(patent)

and searching for re-exams, it can be a complicated item to determine, and a patent attorney should be consulted in matters where significant investment may be involved.

4.4.3 – Assignment Information

An assignment involves the sale and transfer of ownership of a patent by the assignor to the assignee.

The assignee is the entity that is the recipient of a transfer of a patent application, patent, trademark application or trademark registration from its owner on record (assignor).[23]

As discussed in section 4.2.1.2 on Inventors, in the United States, one of the first assignments that takes place is between the inventor and the organization they are employed by, assigning the inventor's rights to an invention to the organization that paid for its development.

Since patents are a property right, they can also be sold or licensed to other entities. While disclosure of patent sales is not formally required in all countries, unless a patent is going to be used in litigation, some organizations will file a re-assignment to ensure there is a record of the change in ownership. Patent licenses, on the other hand, can be more difficult to keep track of, since there is often not a record of the license being negotiated and between which parties. In addition, licenses are often considered as confidential or part of business intelligence information; as a result of that, licensing data is difficult to retrieve and is usually not included in a PLR, unless it refers to specific competitors and more limited number of patents. There are even some databases including some licensing information, based on some publicly available information and M&As, nevertheless most of the times is such information incomplete.

Patents can also be used as collateral against a loan. This type of assignment is referred to as a Security Agreement, and while not a formal change in ownership, this type of agreement will show up in databases that cover assignment data.

Many electronic patent databases will incorporate assignment data by providing separate fields for the original, and current assignee, where the current assignee will reflect the impact of any changes in the ownership of a patent right since it was first applied for.

PLRs will typically incorporate the current owner when an analysis of the assignees prevalent in a particular area is being studied.

4.4.4 – Patent Infringement and Litigation

Patents, by definition, are a right to exclude others from making, using, offering for sale, or selling an invention in the jurisdiction covered by an in-force document. After a patent has been granted and when the patent owner believes that an organization is performing one of these acts, with an invention covered by one of their patents, they can initiate litigation in the form of a patent infringement lawsuit[24].

---

[23] http://inventors.about.com/od/definations/g/Assignment.htm

[24] Additional details can be found at WIPO Handbook - Chapter 4 - Enforcement of Intellectual Property Rights http://www.wipo.int/export/sites/www/about-ip/en/iprm/pdf/ch4.pdf, WIPO IP Panorama 3 – Learning Point 3 http://www.wipo.int/export/sites/www/sme/en/documents/pdf/ip_panorama_3_learning_points.pdf

Organizations who believe they may be sued for patent infringement, and who believe that the patents involved are invalid, or that their organization does not infringe them, may also initiate legal action, e.g. in the US in the form of a Declaratory Judgment or DJ action. The US Declaratory Judgment Act[25] provides US federal courts with the authority to "declare the rights and other legal relations of any interested party" where an "actual controversy" exists.

Patent infringement and invalidation law suits are the actions most often associated with litigation involving patent assets.

Litigation issues are not normally covered in the course of developing a PLR, but for organizations entering a new market or technological field, understanding the litigious nature of the current players can be a valuable competitive and strategic tool.

Many patent databases have begun including litigation data on individual patents, as well as the organizations that own them. Details on the motions involved during the court proceedings can be downloaded from the individual courts associated with the cases.

4.4.5 – Patent Families

Due to the territorial character of the patent system worldwide, patents protection is sought in individual jurisdictions. The Paris Convention of 1883 facilitates the filing in different jurisdictions by claiming priority rights derived from earlier filings (at the offices of first filing - OFF). These priority claims lead to relations between different national patent applications, so called patent family relations. Since the Paris convention expressly permits the claiming of more than one priority rather complex family relations may exist depending on whether two applications share priorities in full, partially or only indirectly, i.e. through other ones.

There is also the opportunity to file an international patent application, referred to as a PCT application, as discussed in section 4.2.1.1. Nevertheless, that does not lead to a patent grant, unless and until they enter the national phase of the individual jurisdictions of interest in order to be examined by them at national level. PCT applications can be filed with or without claiming priority rights of earlier filings. This creates a situation where a single invention might have many individual patent documents associated with it, depending on the number of countries the applicant sought protection in, which are linked to each other through a PCT application number and not necessarily through one or several Paris Convention priorities.

The family becomes even more enlarged when applications, which are normally published separately from granted patents, and thus are discrete documents themselves, are also added to the collection.

In order to simplify some of the dichotomy between inventions and the various patent documents that can be associated with them, the concept of a patent family was created. There are a variety of different definitions provided for patent families depending on how tightly linked the documents are based on priority filings. According to the WIPO Handbook, these are defined as:

- Domestic patent family - a patent family consisting solely of a single office's different procedural publications for the same originating application.

---

[25] Act of June 14, 1934, Pub. L. No. 73-343, 48 Stat. 955 (1934) (current version at 28 U.S.C. §§ 2201-2202 (2000)).

- Simple patent family - a patent family relating to the same invention, each member of which has for the basis of its "priority right" exactly the same originating application or applications.

- Extended patent family - a patent family relating to one or more inventions, each member of which has for the basis of its priority right at least one originating application in common with at least one other member of the family.

A detailed example of the different definitions of patent families can be found at the URL below:

http://www.epo.org/searching/essentials/patent-families/definitions.html

In addition to the general concepts of simple and extended patent families, various database producers have created their own definitions of patent families for organizing patent documents within their collections, such as FamPat. The Intellogist wiki provides definitions for the major providers as well as additional general definitions for patent families besides simple and extended.

http://www.intellogist.com/wiki/Patent_Families

Organizing patent collections by some form of family is an essential activity, which will have a significant impact on how statistics are generated for a PLR. Determining which method will be used and consistently applying it across an entire project will ensure that accurate comparisons can be made between different entities being studied. Generally speaking, using simple families will create larger numbers of narrowly defined collections to analyze while extended families will produce smaller, broader collections. Analysts must determine whether the use of an extended family will severely underrepresent the amount of investment made by an organization for instance when deciding to use that method.

4.5 – Sources of Patent Information

The decision on what patent information source will be used is an important one for an analyst to consider as they are initiating a project. The cost of acquiring data to analyze should be weighed against the time it will take to work with the data, and its comprehensiveness. For convenience, a list of database providers who offer patent information, organized by these three categories, can be found in section 9.2 of these guidelines.

While only the patenting authorities themselves generate **authoritative** patent data (**primary** sources), there are a number of different **secondary** sources for patent documents and information that are usually used to generate PLRs because they include patent information of more than just one jurisdiction which they have obtained from different primary sources.

When comparing patent information sources or databases one needs to distinguish between which data are searchable (search fields) and which are retrievable. Namely not all data which are included in a database and which can be viewed or downloaded are also searchable. For example, many databases do not permit keyword searches in the fulltext including claims and description. Sometimes, keywords can only be searched in title and abstract. Once a relevant publication is identified, e.g. through a search, claims can however be read.

Due to their nature, secondary sources usually include information regarding patent family relations. In some databases this family information is used to perform a family reduction on the results list, i.e.

the search result list would include only one document per family that represents the patent family. The reduction method varies from database to database leading to different representative patent family members. Primary sources may also sometimes include information on national patent family relations, such as continuations or continuations in part.

Since analysis is usually performed after search and by a separate set of tools, it is particularly important whether sources permit the download of structured data (see Section 4.5.1).

While primary sources are usually free, secondary sources follow a continuum from free sources that provide basic bibliographic, text and/or image data, to for-fee sources that offer additional enhancements and features associated with the data, or even integrated analysis tools.

4.5.1 – Primary Sources: Patent Authorities

Each patent jurisdiction defines its publication policies and the authority in charge of producing the official patent related publications and providing access to other information like legal status data or the public part of the file wrapper. Many patent authorities around the world have websites with data services that allow the general public to search and retrieve the respective patent documents. These sources may be addressed as primary sources since they are the authoritative sources in comparison to other (secondary) databases that gather such information from many different primary sources and make them searchable through a single search interface. Some of the features associated with these primary sites include:

- These collections are generally available to search for no cost; very few jurisdictions permit access to full publications only for a fee. Basic bibliographic data are generally accessible for free.
- Some of the offices (e.g. the United States Patent Office) separate the searching of applications from granted patents but most allow the user to search both simultaneously.
- Some of the authorities maintain separate patent register databases that provide information on the most recent legal status of pending applications or granted patents, or permit file inspection, in addition to the official document publication services.
- Many of the primary sources allow searching in an English interface, regardless of the native language of the country, although data (e.g. legal status) or documents retrieved are only in the national language.
- Search syntax and functionality varies from site to site so individual search strategies need to be developed for each such source.
- Some primary sources allow for bulk downloading of patent documents discovered during a search, while others only allow small numbers or single documents to be downloaded.
- While entire documents can be downloaded, most sites do not allow individual patent data fields to be exported (structured data) or, if they do, the number of fields available is limited.

Due to the limitations imposed by the National Patent Office sites, especially the inability to export individual data fields (structured data) and the limitation to the authority's own publications, these collections are not normally used for generating datasets associated with PLRs, unless the geographic scope of a report is only the one jurisdiction, or the national data are not included in any secondary source, or if certain data like legal status need to be verified. They may provide an

inexpensive means for exploring a topic area, but once that is accomplished most analysts will shift to other patent sources to generate the data used for analysis.

4.5.2 – Free Secondary Sources

A few patent authorities maintain secondary patent databases which allow searching for patents from several countries together. These are mainly PATENTSCOPE from WIPO, Esp@cenet from the European Patent Office, or DEPATISNET from the German Patent and Trademark Office.

There are several further patent searching services online that are available for free. Their characteristics are similar to the offerings from the patent offices although their country coverage may be smaller. They offer sometimes advantages over the patent office sites since, their user interface is often a little more polished, and end-user friendly. They also occasionally offer additional features that are not normally found on the patent office sites.

Rudimentary analytics tools can be found on a few of the sites but this functionality is normally left to the commercial sources. For instance, TheLens from Cambia, and PatentInspiration from CREAX, and PATENTSCOPE offer some statistical analysis and visualization features.

Some of these sites have been used to generate PLRs[26] and their free nature makes them an attractive source for collecting data. The balance an analyst has to strike is between the low cost of the data versus their ability to manipulate data during subsequent analysis. In some cases, the features and functionality available from commercial tools justify the cost of access since they save time in the subsequent analysis stages.

4.5.3 – Commercial Sources

Commercial sources of patent information have been available for over a century. What started as abstracting and indexing services covering patents from a handful of countries and, on a small variety of topics, has developed into a large business with many significant players. Some of the characteristics associated with commercial patent database providers include:

- **Enhanced content** – editorial staffs create titles, abstracts and indexes that "translate" the legal language used in patents into standard terms familiar to practitioners. When searching, the addition of enhanced content has a significant impact on the comprehensiveness of a patent collection.
- **A "one-stop-shop" for searching, analysis and dissemination** – several of the major commercial providers allow an analyst to search, refine, review, analyze and share collections and output within the same system. Having most of the functionality on one place can be a significant time saver.
- **Flexibility in exporting data** – Commercial sources generally have a higher limit on the number of records available for export. They also, generally, have a greater variety of fields to choose from, providing more options for an analyst to explore.
- **Additional tools for refining data collections** – as will be discussed in subsequent sections of the guidelines, patent data can contain errors, such as typos in patent assignee names, or

---

[26] http://www.patentlens.net/daisy/patentlens/landscapes-tools.html

redundancies, such as the same invention being represented in different countries. Many services have mechanisms in place to assist users in dealing with these items, as opposed to having to do them by manually, providing significant time savings in preparing data to be analyzed.

4.6 – Reports Associated with Patent Information

Due to the critical nature of patent documents and the information associated with them, reports related to patent information are used in a variety of different business contexts. There are different reports affiliated with providing information on patent data in these different environments. These guidelines are focused on the use of patent information to generate Landscape Reports but the following definitions of additional reports that incorporate patent information are provided for reference.

4.6.1 – Patent Landscape

There is no single definition or common understanding for a patent landscape report. There are various approaches, some of which broader, covering even Freedom to Operate elements and other, non-patent related data, such as market analysis, while others much narrower, with certain understandings being that a patent landscape is identical to a patent map (for more information about that, please consult section 4.6.2). One could state that a patent landscape report provides an overview of the patenting activity and trends in a field of technology. A patent landscape normally seeks to answer specific policy or practical questions and to present complex information about this activity in a clear and accessible manner. Industry has long used patent landscapes to make strategic decisions on investments, research and development (R&D) directions, competitors' activity as well as on freedom to operate in introducing new products. Now, public policymakers are increasingly turning to landscaping to build a factual foundation before considering high-level policy matters, especially in fields such as health, food security and the environment.

4.6.2 – Patent Map

While the name patent map sounds similar to a patent landscape, a patent map generally represents a graphical representation of a data collection that borrows characteristics of cartography. Maps are usually focused on a single attribute associated with a data collection such as the classification of documents based on the topics covered within them. A map paradigm is used to represent similarity between documents or concepts since the human mind is used to and can readily understand the use of maps to correlate distance between two items.

4.6.3 – Watch or Alerts

A patent watch is a process for monitoring newly issued patents, as well as possibly pending patent applications, to assess whether any of these patent documents might be of interest.[27] Patent alerts are also performed in order to determine if patent documents of interest undergo a change in status. For instance, a patent application of interest may be monitored to determine if it goes on to grant. Organizations also set up patent watches to monitor new patent applications coming from competing organizations in high interest technologies.

4.6.4 – Freedom-to-Operate / Clearance

---

[27] http://en.wikipedia.org/wiki/Patent_watch

In this type of report, which involves an organization asking for a legal opinion on whether a product they are planning on shipping will infringe any existing patents before they launch. The search involved is very specific since it is country specific and usually only applies to in-force granted patents and their claims. There is nothing offensive about this type of report since the interested party is not going to assert patents against anyone else, they are simply looking to make sure that they are not going to be infringing someone else's patents. An analyst in this case needs to identify the critical components of the product in question and search jurisdiction specific claims of in-force patents to see if any of them cover the product components in question. In most cases a great deal of money has gone into a product launch or can be involved with a successful product which is generating a great deal of revenue so it is important for companies to know that they will be reasonably safe from future litigation before they make an even larger investment.

4.6.5 – Patentability / Prior-Art

This type of report is usually performed in the legal context of determining if a new invention is eligible for patent protection and determining how broadly the claims for the new invention can be written. This type of report can cover both patent and non-patent literature and is typically looking for references that were published before the filing date of the invention in question. In the United States inventors have up to a year from first public disclosure of an invention to file a patent so some searchers will go back an additional year with their searching to make sure they have found the best references.

This report helps identify the boundaries of the known references and will help the attorneys drafting the claims to ask for the broadest coverage possible. Without knowing the scope of the known references it is difficult for the attorney to know how broadly they can write the claims and still expect the examiner to grant a patent.

4.6.6 – Validity

Validity reports provide the results of the largest and most comprehensive of all patent searches. These reports are almost always associated with large sums of money and critical business decisions and as such need to be as comprehensive as possible. This report shares similar characteristics to Patentability but is normally far more comprehensive since there is typically much more at stake when this sort of report is asked for.

The object of the search involved with this report is to identify prior art references, which will allow a granted patent to be made invalid or revoked during a particular proceeding before the particular patent office of interest or during a court proceeding. Sometimes organizations will also initiate validity challenges for patents that they are thinking of acquiring especially if they believe these patents will later be used in some type of litigation or another. On the flip side of this an organization who is provided with a cease and desist notice will often want to make the patents in question go away by finding invalidating prior art and then entering into re-examination. The prior art references in question can come from the patent or non-patent literature must be available in the public domain and have to have been published prior to the priority filing date of the patent in question. In the United States there is a one-year grace period on patents filings so some analysts will look back an additional year when they search so they can be sure to avoid this type of situation.

4.6.7 – General Statistics

Reports of this nature are generated by patent offices[28] and other organizations to provide metrics on the performance and output associated with an area of interest. PriceWaterhouseCoopers for instance, publishes a yearly litigation study that looks at the statistics associated with patent litigation in the United States[29].

WIPO cooperates with intellectual property (IP) offices from around the world to provide its stakeholders with up-to-date IP statistics[30]. Generally, these statistics are provided as raw data that can be used by analysts to draw conclusions based on their own interests and experimentation. WIPO also publishes statistical reports on worldwide IP activity and on the use of WIPO-administered treaties in the protection of IP rights internationally, such as the PCT Yearly Review[31] and the World Intellectual Property Indicators[32]. In addition, WIPO IP Statistics Data Center is an on-line service enabling access to WIPO's statistical data on intellectual property (IP) activity worldwide[33]. Users can select from a wide range of indicators and view or download the latest available as well as historical data according to their needs, based on the Worldwide Patent Statistical Database (PATSTAT[34]) data, which is administered by the European Patent Office. Moreover, patent statistics are often paired with other data and indicators to provide a more holistic approach of innovation. An example of this is the Global Innovation Index (GII)[35] which ranks the innovation performance of 143 countries and economies around the world, based on 81 indicators. The GII is co-published by WIPO, Cornell University and INSEAD.

---

[28] http://www.wipo.int/ipstats/en/resources/office_stats_reports.html
[29] http://www.pwc.com/en_US/us/forensic-services/publications/assets/2014-patent-litigation-study.pdf
[30] http://www.wipo.int/ipstats/en
[31] http://www.wipo.int/edocs/pubdocs/en/wipo_pub_901_2015.pdf
[32] http://www.wipo.int/ipstats/en/wipi/
[33] http://ipstats.wipo.int/ipstatv2/?lang=en
[34] http://www.epo.org/searching/subscription/patstat-online.html
[35] http://www.wipo.int/econ_stat/en/economics/gii/

Chapter 5: Objectives and Motivations for Generating Patent Landscape Reports

Producing a Patent Landscape Report (PLR) can be a time intensive and expensive process. Devoting the resources necessary to generate a PLR is often tied to a business objective, e.g., where an organization is preparing to make a significant monetary or headcount investment in developing or moving into a technology area. Various types of organizations have different objectives that need to be explored in order to make an informed decision about the allocation of resources to a new project or area. For the purposes of these guidelines, the types of organizations will be either governmental and inter-governmental, or corporate.

The approach taken to developing a PLR will differ depending on the business objectives that necessitated the ordering of the report for an individual decision cycle. Generally speaking, PLRs support informed decision-making. Regardless of the business objective, PLRs have developed a specific format, and are designed to efficiently address the concerns associated with making high stakes decisions in technologically advanced areas, with a maximum degree of confidence. For many years decision-makers operated based on personal networks and intuition. With the institution of patent analytics, and PLRs, it is possible for these critical decisions to be made with data-driven approaches that deliver informed choices, and lower risk profiles.

5.1 – Objectives behind Patent Landscape Reports

The issues associated with public policy decisions, initiated by government agencies, are usually different from the decisions that are important to corporate entities, and their stakeholders. The analyses of patent information, and the generation of PLRs, are increasingly required by both types of organizations, in order to understand a technological area. Understanding how the decisions differ between these two types of entities allows the analyst to tailor their report in order to most efficiently meet the needs of the respective audiences. In most cases, there is not much overlap between the objectives associated with each entity, but in the cases of using PLRs to explore technology transfer, and research and development questions there is substantial similarity in what both groups are attempting to discover for informed decision-making.

5.1.1 – To Support Governmental Policy Discussions

At the beginning of April 2008, the World Intellectual Property Organization (WIPO) in cooperation with the Food and Agricultural Organization (FAO) organized a Symposium on Public Policy Patent Landscaping in the Life Sciences[36]. The stated goals for this symposium provide a succinct explanation of how PLRs can be used as instruments to inform public policy makers as they look to tackle technological issues.

The Symposium draws together two important trends:

- *Patent information as a tool of public policy: Policymakers who deal with innovation and access in the life sciences – concerned with agriculture and food security; public health and pharmaceuticals; and environmental issues – have increasingly focused on the patent system. They look for clearer, more accessible and geographically more representative information to support key policy processes. They seek a stronger empirical basis for their assessments on the role and impact of the patent system in relation to key areas of life sciences technology.*
- *Improved analytical tools and access to patent information: Rapid growth in the use of the patent system, and in the diversity of users, has led to an explosion of raw data on patenting*

---

[36] http://www.wipo.int/meetings/en/2008/lifesciences/patent_landscaping

activities in the life sciences. This data is progressively being turned into useful information. Availability and quality of patent information have increased. Analytical tools and methodologies are better understood and are more widely available. And greater practical experience has been harvested from a range of recent patent landscaping initiatives. This trend opens up enormous practical potential for improved patent information resources for public policymakers addressing the life sciences.

This Symposium aims to take a first step towards more systematically matching the policy needs – the international policy agenda on public policy issues of concern in the life sciences – with the practical capacities – the diverse resources that are now increasingly available to gather, analyze and extract key trends and findings from patent information.

PLRs are designed to provide efficient access to a large collection of technologically focused data and to answer key questions about what technologies are covered, which organizations own the patents and in which countries they are held. Where previously, this data might only be available to the technologically savvy, now it can be made available to individuals at all technological knowledge levels. Making technological topics available to policy makers leads to better decision-making and additional resources devoted to critical issues. Within the context of Governmental Policy discussions, an examination of the activities associated with various jurisdictions can help identify the elements required for preparing PLRs for these agencies.

5.1.1.1 – Global Efforts

WIPO is involved with several global efforts to enhance the availability of information on patent activity. A WIPO Magazine article, **Shedding Light on the Life Sciences: Patent Landscaping for Public Policymakers** provides excellent reasons for why these international efforts are necessary for informed public policy discussions[37].

Good quality information about patenting activity is an essential input for some of the most critical international policy debates today. Yet patent information is unavoidably complex, constantly evolving, and difficult to capture in readily accessible form for a non-specialized audience. There are real risks associated with making judgments on the basis of limited patent landscapes without considering the full technical and legal context. Thus the demand for reliable patent landscaping in the life sciences is strong, and there are no shortcuts to meeting that demand.

A positive feedback loop is developing: patent informatics are delivering increasingly focused and accessible information products for policymakers, who in turn can sharpen and distil their demands for patent information, leading in turn to increasingly more relevant and useful support. Patent landscaping is not a substitute for the policy debates and deliberations on the key life sciences issues of the day. But it can inform, support and strengthen the factual basis for discussions, so assisting the policymakers in those fields to set future directions on health, the environment and food security.

WHO's Global Strategy and Plan of Action also identified the need to improve access to patent information to facilitate the determination of the patent status of health products. It urges stakeholders to:

---

[37] http://www.wipo.int/wipo_magazine/en/2008/04/article_0005.html

- Facilitate access to user-friendly global databases which contain public information on the administrative status of health-related patents. This includes supporting existing efforts for determining the patent status of health products, and to
- Promote further development of such global databases including, if necessary, compiling, maintaining and updating such global databases.

To date, there was no comprehensive overview of patenting activity and trends in the area of vaccines. In the framework of policy discussions at the World Health Assemblies on vaccine local manufacturing, WHO and WIPO jointly developed a patent landscape report[38] that provides an overview on what is being patented in terms of selected disease targets, who is doing the patenting, where patents are filed and on how patent policies change over time. This provided the factual evidence and the background to support the related policy discussions.

WIPO has also worked with the World Health Organization (WHO) on understanding the patent environment associated with essential medicines from around the world. In a summary of this work[39], the following details were shared:

For more than 30 years, the WHO has published a Model List of Essential Medicines, which is updated every two years. Most countries have adopted the concept and have developed their own national lists of essential medicines. One important question is to what extent patents protect the essential medicines on the WHO Model List. One of the projects presented at the symposium focused on assessing the patent status of the medicines that have been added to the WHO Model List in recent years. The study, based on data from the US Federal Drug Administration's Orange Book, identified relevant patent families for these medicines in countries where patent data were available.

Access to affordable generic medicines can be achieved through licensing agreements. A new approach to increase access this way is the creation of a patent pool for antiretroviral medicines, undertaken by the Medicines Patent Pool, a United Nations-backed organization established in 2010[40]. This requires reliable patent information, including:

1. *Knowing what patents cover the products to be used;*
2. *What the patents exactly cover for these products;*
3. *Who holds the patents;*
4. *The countries where the patent applications have been filed and where they have B been granted; and*
5. *The current legal status of those patents.*

These are complex tasks. Many national and regional patent collections can only be consulted on-site. Information is often not updated or incomplete, especially on the legal status. With the support of WIPO (among others through the preparation of two patent landscape reports on Ritonavir[41] and Atazanavir[42], and information collected by WIPO with the support of national IP Offices), and a wide range of national and regional patent offices, the Medicines Patent Pool has identified the legal status

---

[38] http://www.wipo.int/patentscope/en/programs/patent_landscapes/reports/vaccines.html
[39] http://www.wipo.int/edocs/mdocs/mdocs/en/who_wipo_wto_ip_med_ge_11/who_wipo_wto_ip_med_ge_11_www_169578.pdf
[40] http://www.medicinespatentpool.org
[41] http://www.wipo.int/edocs/pubdocs/en/patents/946/wipo_pub_946.pdf
[42] http://www.wipo.int/edocs/pubdocs/en/patents/946/wipo_pub_946_2.pdf

of key patents for selected antiretroviral medicines in low- and middle-income countries[43]. A database has been launched in the meantime to allow open access to this resource and the legal status information related to these medicines[44]. The discussion raised the question of whether patent pooling could be a general solution in cases of patent thickets, i.e. situations involving overlapping patents, which prevent competition.

In some cases, PLRs have also an awareness raising role on the importance of the IP aspect and patent information in policy discussions of various subject matters involving technology; and they can also have an impact. After WIPO's collaboration with IRENA[45], the International Renewable Energy Agency, and the production of a PLR on Desalination Technologies and the use of Renewable Energies for Desalination[46], the importance of IP and patent information became more clear to IRENA and its stakeholders and some years later it even lead to the launch beginning of 2015 of IRENA's Standards and Patent Information Platform[47].

WIPO has generated several PLRs associated with these on-going, global efforts. Additional details on these efforts and a list of available PLRs can be found at: http://www.wipo.int/patentscope/en/programs/patent_landscapes/index.html.

5.1.1.2 – Regional Efforts

The World Health Organization works regionally, specifically in developing countries, to ensure that access to vital medicines is available to individuals of all economic and social backgrounds. In order to understand the technology, and IP rights, associated with providing access to critical care, WHO has worked with WIPO and the World Trade Organization (WTO), and looked at the patent activity around vaccines. On February 18th 2011, the three organizations organized a joint Technical Symposium on Access to Medicines, Patent Information and Freedom-to-Operate[48] that provides details on this regional effort.

In the field of vaccines, WHO is monitoring the patenting activity to identify the extent to which vaccines and production technologies are protected by intellectual property. When patents apply, in some cases WHO supports research on alternative technologies or negotiates licenses with the right holders on behalf of developing country manufacturers. For most existing vaccines, patents do not generally prevent production by competitors, but there are some notable exceptions, including reverse genetic engineering, a key technology for the production of pandemic influenza vaccines and the human papilloma- virus vaccine.

---

[43] The database coverage as of June 1, 2015 can be found at
http://www.medicinespatentpool.org/wp-content/uploads/Patent-Status-Table-1June2015.xls
[44] http://www.medicinespatentpool.org/patent-data/patent-status-of-arvs/
[45] http://www.irena.org/
[46] http://www.wipo.int/patentscope/en/programs/patent_landscapes/reports/desalination.html
[47] http://community.irena.org/t5/Innovation-for-Energy-Transition/LAUNCH-OF-IRENA-S-STANDARDS-amp-PATENTS-INFORMATION-PLATFORM/gpm-p/2257
[48] http://www.wipo.int/meetings/en/2011/who_wipo_wto_ip_med_ge_11/program.html

A major barrier to increasing the uptake of vaccine manufacturing in developing countries is the lack of know-how. Thus WHO also focuses on the transfer of vaccine production technology to these countries.

The Dengue Vaccine Initiative of the International Vaccine Institute presented a global freedom to operate analysis with different candidates for dengue vaccines. The goal was to understand how intellectual property may affect access to future vaccines in developing countries and to evaluate how free developing country developers are to market their vaccines internationally.
The analysis revealed that the sponsors of vaccine candidates seem to have intellectual property required to regulatory agency approval for their candidates and to market them. However, in the future, problems might still arise from patent applications, which cover certain delivery mechanisms.

WIPO is also working closely with various national institutions to support them in addressing issues of interest for various regions. For instance, WIPO is working in collaboration with the Malaysian (MyIPO), the Malaysian Palm Oil Board and the Phillipines IP Office (IPOPHIL) on a Patent Landscape Report on Palm Oil Production and Waste Exploitation Technologies, providing a general overview of related technologies globally, but also with a focus on the region whose economy is strongly active in the field.

5.1.1.3 – National Efforts

Several Intellectual Property Offices have become engaged in the field of patent analytics, e.g., the Intellectual Property Office of the United Kingdom (http://www.ipo.gov.uk/) initiated an informatics team in 2009 with the stated goal of "using patent data to mine, reveal, and inform, for government and for industry"[49]. To date they have twelve PLRs listed on their reports page[50] covering topics from stem cells to 3DTVs. They list the following national benefits to the use of patent analytics and PLRs to both governmental and business interests in the United Kingdom:

- **Innovation Policy –** Providing evidence of the emerging trends in technology
- **Investment Opportunity –** Identifying the technologies that could create a whole new market
- **Competitor Intelligence –** Profiling your competitors using their patent portfolios
- **Knowledge Transfer –** Analyzing the flow of knowledge and collaborations
- **Geographical Profiling –** Comparing markets between countries and regions

Looking at the specific objectives listed in the IPO study on the generation of energy from waste materials, the following rationales are given:

The objectives as defined in the original project proposal are as follows:

- *Provide an overall patent landscape analysis in the technology area of Energy from Waste.*
- *Provide analysis of the level of UK research in comparison to the rest of Europe and rest of the world.*
- *Identify key active companies and key patent applications.*

The objectives for Phase II were designed to focus on the UK energy from waste patent landscape, covering:

---

[49] http://www.ipo.gov.uk/informatics.htm
[50] http://www.ipo.gov.uk/informatics-reports

- *Specific technology fields: biogas / biohydrogen from waste*
- *Emergent technologies*
- *UK patent applicant types: commercial, academic, or government*
- *The activities of individual patent applicants and the extent to which they influence the overall patterns in the UK*
- *Consolidated IPC classifications to form larger groups and produce more focused results*
- *Producing and interrogating a UK patent landscape map*

The objective for Phase II, in particular demonstrates how the creation of PLRs can assist the development of high impact technologies within national jurisdictions. Investing in high impact technologies helps to establish both academic and industrial centers of excellence within a country that often leads to an increase in the number of organizations that will establish research and manufacturing facilities in the country.

5.1.1.4 – Technology transfer and licensing

In order to assist industries residing in their countries, several national governments have recently started purchasing patent assets on behalf of the organizations that manufacture products in their jurisdictions. A news story on this practice from Reuters[51] provides details on this practice:

France Brevets was launched in 2011 with 100 million euros, half from the state and half from the Caisse des Depots, a publicly managed investor in French economic development.

Pascal Asselot, licensing director for France Brevets, said that by assembling patent pools with intellectual property bought from French and foreign businesses, France Brevets aims to convince other companies to sign licensing deals and pay royalties. If France Brevets can show a healthy revenue stream, the hope is to attract sustainable private investment, Asselot said.

Korea's Intellectual Discovery, which was started in 2010 amid government fears that domestic companies were losing key patents that could be used against them by foreign companies, has a $140 million government commitment.

Prominent South Korean companies like Samsung Electronics Co Ltd and LG Electronics have signed up as "shareholders," providing Intellectual Discovery with additional revenue in exchange for a license to its patent portfolio.

Intellectual Discovery chief general manager Chant Kim compares the company to San Francisco-based RPX Corp, which acquires patents to protect its members but doesn't initiate lawsuits.

PLRs are often used by organizations exploring technology transfer, and licensing in order to understand what other organizations have invested in a particular area. If another organization has invested in a technology, especially if the investment was made a few years earlier, there is a higher likelihood that they will be receptive to hearing about new developments, and potentially acquiring or licensing the technology.

While government agencies have shown recent interest in this area, it has been established practice for many years for technology intensive corporate entities. Many examples abound, including Texas

---

[51] http://www.reuters.com/article/2013/03/20/patents-nations-idUSL1N0BZ10C20130320

Instruments and Tessera, of companies that originally received most of their revenue from the sales of physical products, but began to shift their business models to rely more heavily on revenue generated from licensing.

5.1.1.5 – Research and development (R&D) decision-making

As was seen in the UK IPO national efforts example, the use of PLRs can influence decisions around investments in academic and non-profit funding for the creation of economically favorable technologies, which can increase the Gross Domestic Product of a country. Governments use PLRs to ensure that investments in R&D will be directed to technologies and industries that will ensure their future competitiveness in high impact areas.

Understanding patent rights also have a major impact on R&D since patents, by their nature, are a right to exclude, and this can have a significant impact on the ability to conduct further research in an area covered by them. Different countries allow for different Research Exemptions involving patent rights. The Wikipedia page on Research Exemptions[52] provides the following details on the International framework around this exemption:

Article 30 of the WTO's TRIPs Agreement permits this type of exception:

"Members may provide limited exceptions to the exclusive rights conferred by a patent, provided that such exceptions do not unreasonably conflict with a normal exploitation of the patent and do not unreasonably prejudice the legitimate interests of the patent owner, taking account of the legitimate interests of third parties."

PLRs are an excellent method of determining which technologies have patents associated with them, and in which countries those rights are in-force. If the initiation of R&D, in a particular region, is one of the objectives of the report, then additional discussion of the Research Exemption law for that region should be included.

Corporate entities are also subject to the same dynamics, when it comes to R&D decision-making and PLRs are often generated in this sector as well to support management. A particularly good example of this involves the so-called Safe Harbor exemptions that allow for R&D activities in association with the generation of generic drugs. The Wikipedia page on Research Exemptions[53] discusses this situation as well:

In patent law, the research exemption or safe harbor exemption is an exemption to the rights conferred by patents, which is especially relevant to drugs. According to this exemption, despite the patent rights, performing research and tests for preparing regulatory approval, for instance by the FDA in the United States, does not constitute infringement for a limited term before the end of patent term. This exemption allows generic manufacturers to prepare generic drugs in advance of the patent expiration.

In the United States, this exemption is also technically called § 271(e)(1) exemption or Hatch-Waxman exemption. The U.S. Supreme Court recently considered the scope of the Hatch-Waxman exemption in Merck v. Integra. The Supreme Court held that the statute exempts from infringement all uses of compounds that are reasonably related to submission of information to the government under any law regulating the manufacture, use or distribution of drugs.

---

[52] http://en.wikipedia.org/wiki/Research_exemption
[53] Ibid

In Canada, this exemption is known as the Bolar provision or Roche-Bolar provision, named after the case Roche Products v. Bolar Pharmaceutical.

In the European Union, equivalent exemptions are allowed under the terms of EC Directives 2001/82/EC (as amended by Directive 2004/28/EC) and 2001/83/EC (as amended by Directives 2002/98/EC, 2003/63/EC, 2004/24/EC and 2004/27/EC).

In both governmental and corporate environments PLRs are an essential tool for understanding the competitive environment around research areas of interest, and discovering whether groups interested in pursuing research initiatives have the freedom to do so.

5.1.2 – Business or corporate uses

While there are some overlaps between the uses of patent analytics, and PLRs for governmental policy decision-making, the circumstances under which these tools are used for business, or corporate situations can be somewhat different. The methods used to generate a PLR, under both circumstances, are similar, but the implications, both in the short and long term, can vary significantly.

5.1.2.1 – Competitor monitoring

Competition is an inherently business related concept and it is nearly impossible to find a successful industry in which there is not some form of competition between organizations looking to gain an advantage over other companies in a space. Since businesses compete against one another, understanding the capabilities, resources and expertise associated with a competitor becomes a key component of corporate strategy.

These ideas are often associated with Sun Tzu, in The Art of War, when he said the following:

"If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle"

This thinking is often applied to business strategy and is especially the case when looking at technologically focused industries. PLRs usually answer the first question in a strategy session; does my competitor have patent rights in any of the areas of interest? They also address the next, follow-up question; how many do they have and in what aspects of the technology do they have coverage?

Patents can provide knowledge on levels of competitive expertise, timing, and investment, in addition to providing a right to exclude. They are even more important in technology intensive industries since many of the insights contained in them are only published in patents and are not described in other types of publications.

5.1.2.2 – Technology monitoring

Corporate entities usually associate themselves with specific technological areas. In the Pharmaceutical industry, for instance, no single company can excel in all therapeutic areas. Most companies tend to focus on a few areas, and concentrate their efforts on those items to the exclusion of others. There may be specific competitors, which the organization will monitor, but it is recognized that innovation can come from unexpected sources, and as long as it covers an area of interest, the company will be aware of these new developments regardless of their source.

Occasionally, organizations will look to move into new markets, or enter new technological areas that they did not have direct experience in previously. When this happens a PLR can provide key intelligence on the top players, inventors, and technology sub-categories associated with the area of interest. Armed with this data companies can determine if they will develop new technologies or seek to acquire technology from others.

PLRs tend to be technologically focused in scope, as opposed to competitor focused, and are well suited for providing a background collection of insights associated with a particular area. Once an initial understanding is gained, ongoing monitoring can also be established using the PLR as a foundation.

5.1.2.3 – Mergers and acquisitions

Technology savvy organizations will look at innovative benefit and fit, in addition to the more traditional economic and market driven factors, when deciding on whether to acquire, or merge with another company. A technology based due diligence assessment is necessary to ensure that two groups will be compatible with one another, and is often used to determine how much technological overlap there is between them.

It is not unusual to find that, while two organizations are working in the same technological area, they are taking different approaches to solving key issues. In this case, the companies are complimentary to one another, and a stronger argument can be made for justifying a merger, especially when a combination will cover all major methods for deploying a technology.

Under different circumstance, it is sometimes discovered that an acquisition would provide redundant capabilities that are already owned by the acquiring company. This scenario can occur with large, multi-national and divisional companies, or ones that become involved with several mergers, or acquisitions in a short time frame.

In any of these scenarios PLRs are often generated to provide focused insight on the organizations involved and project how the combined resources will compare to the other companies involved in a technology area.

5.2 – Motivations for generating Patent Landscape Reports

Having looked at the reasons why organizations are interested in the analysis of patent information, and its association with strategic decision-making, it is also important to look at the aspects of PLRs, in particular, and how they can be maximized to achieve this benefit.

In chapter 8, individual sections and specific analytics will be addressed, but in this section the value associated with the reports, especially to the client, typically a technologically based, decision maker, are addressed.

5.2.1 – Who is the report intended for?

In order for a PLR to be impactful it has to be read by a decision maker in the first place. The two primary considerations in determining whether a PLR will be read, and better yet acted upon, are based on the position your client holds within the organization, and their personal information intake profile.

As an example, consider the following advice on presenting to executives within an organization[54]:

The biggest mistake I see people make is they present to Senior Management the same way they would present to their peers or their team. Remember the higher up a person is in an organization, the larger the picture they have of problems and issues. Think of moving up through the ranks of management like a helicopter that climbs higher and higher in the air. On ground level you only see things right in front of you, but when you get very high in the air, you can see for miles and miles.

So this means that when you are presenting to upper management you need to THINK about the problems and issues from THEIR perspective, not yours. It is your job to show how what you are talking about fits in to the corporate GOALS, VISION & MISSION.

It is also important to recognize that there is an inverse relationship between the importance of your client's role in an organization, and the amount of time an analyst is likely going to be given to present to them. Generally, as discussed in the next section, the most important feature of a PLR should be about turning data into intelligence and insight, thus saving your client the time of having to do this themselves. In this section a different implication of time, the time available to share the most critical aspects of the research, is being discussed. PLRs should include an Executive Summary, and if this is the only section the client reads then it needs to deliver the message, or insight the analyst wants to portray. Consider generating different versions of the PLR that are appropriate to the different roles within an organization that are likely to be interacting with it.

It is also important to consider the audience for a PLR, and think about the ways in which they process and analyze information. On one level people can be consider right-brained or left-brained as discussed below[55]:

According to the theory of left-brain or right-brain dominance, each side of the brain controls different types of thinking. Additionally, people are said to prefer one type of thinking over the other. For example, a person who is "left-brained" is often said to be more logical, analytical and objective, while a person who is "right-brained" is said to be more intuitive, thoughtful and subjective.

More recent research indicates that this dichotomy is overly simplistic. A more detailed theory involving Socionics, provides a more granular approach to methods of learning based on Myers-Briggs personality types. Additional information on Socionics can be found at the URL provided:

https://en.wikipedia.org/wiki/Socionics

When generating a PLR consider whether the primary client is more of an analytical thinker, someone who will want to see a logical progression leading to a conclusion, or an intuitive thinker, who is leaping ahead to the implications of the research and what impacts they have on their pre-conceived notions. Ideally, try to address both types of thinkers but give priority to the personality type of the primary client.

5.2.2 – How does it save the client time?

The most valuable commodity a PLR can provide a decision-maker is to save them time. PLRs take large amounts of raw, unprocessed data and generate insight, and intelligence based on the analysis

---

[54] http://www.impressionmanagement.com/blog/09-24-2010/presentation-skills-37
[55] http://psychology.about.com/od/cognitivepsychology/a/left-brain-right-brain.htm

conducted and, perhaps more importantly, the interpretation of the results by a skilled analyst. The analyst is responsible for applying the intelligence to the specific business context associated with the decision being made, and provides recommendations that can be acted upon by the client. Clients are looking to analysts to provide an expert opinion, and perspectives based on their knowledge of the topic in question, not simply data.

This can sometimes be a daunting task and many PLRs simply provide summaries of data, without taking the extra step of interpreting the results, in view of the objectives that led to the ordering of the report in the first place.

It is also important to recognize that the client does not necessarily understand the nuances of patent information, and may likely misinterpret the data if they are not provided with conclusions generated from the analyst. Having an analyst expertly interpret this critical data source also saves the decision-maker from having to develop these skills themselves.

5.2.3 – How does it add value to the decision making process?

Many decisions are made without proper intelligence, or analysis, leading to higher risks associated with them. Senior-level managers, in particular, tend to rely on a network of familiar advisors, as opposed to data driven decision-making, for developing organizational strategy. In technology-driven industries, the volume of information available, and the complexity of the issues surrounding the decision-making process, is such that relying on personal networks increases the risk that a suboptimal result will take place.

Data analysis allows for all available options to be explored, including allowing for historical information that might not otherwise be considered, to approach decision-making in a systematic fashion. Since more options, and information, are being considered the risks associated with the development of a strategy, or a specific decision, are lowered. Risk cannot be eliminated completely, but when all available data is analyzed, using a variety of complementary methods, and procedures, it provides confidence that the proposed recommendations will reflect the desired outcome for the business.

From the analyst's perspective the most valuable results are achieved when a plan for presenting the recommendations of the research is developed, and the likely implications, and impacts to the organization, are clearly stated. Practically speaking, this requires forming an opinion, supporting it through interpretation of data analysis, and providing business perspective for acting on it.

5.2.4 – How will the user evaluate the effectiveness of the report?

As stated at the beginning of the chapter, the generation of a PLR is a time intensive and expensive endeavor, and it is embarked upon since an organization is preparing to make a significant monetary or headcount investment. Understanding how the user will evaluate the PLR is a critical component in ensuring that the end product will be satisfactory to them.

Generally speaking, in regards to PLRs, evaluation looks at whether the right questions were asked and whether they were analyzed correctly. It examines the rationale, and the justification of activities, in relationship to the decision, associated with the questions. Finally, it eexamines the efficiency of the results, or how the time and effort associated with the PLR was converted into results, and provided value to the decision-making process.

Understanding these standards leads to the production of high value PLRs that provide excellent return on the investment, and lower the risk associated with technological decision-making. Once a business leader is introduced to reports of this nature, and discovers how they provide key intelligence, they are likely to want to incorporate them frequently into critical objectives, as a matter of practice.

Chapter 6: Tasks Associated with Patent Analytics and Landscaping

When performing data analysis, using patent information, for a Patent Landscape Report (PLR), or for most patent analytics projects, there are a finite number of functions or tasks that are available for the analyst to consider using. In addition to providing analysis capabilities, some of the tasks are associated with preparing data for subsequent analysis, while others provide means for visualizing the output from an analysis task.

This chapter defines and explores the various tasks associated with patent analytics and PLRs, provides explanations of why they are used for patent analysis, and suggests a few tools that can be used to perform each task. By systematically looking at all of the tasks that are typically associated with PLRs an analyst can consider the range of options available to them, and decide which elements they will include in their study.

6.1 – Data Cleanup and Grouping

Data cleanup and grouping are processes for the manual, or automatic standardization of terms or items, within a data field, to correct errors or inconsistencies, or to group synonymous entries. It is required by patent analysts in order to produce statistically relevant results. It is necessary since raw patent data is notoriously "messy" and requires cleanup or standardization to produce accurate results. Misspellings, for instance, are a common occurrence within certain fields, and require correction. There are also many terms with the same or similar meanings, within the English language, and these should be grouped together when analyzing concepts.

Using a simple data cleanup example, 3M is listed in most patent assignee fields in a number of different ways including: 3M, 3M Inc., 3M Inc, and by its full name, Minnesota Mining and Manufacturing. All of these represent the single entity, and have to be grouped together, or standardized, in order to perform accurate statistics reflecting the total impact of the organization. Grouping these terms together into a single entry is the essence of the data cleanup task.

Often, it takes more time to prepare patent data for analytics than it does to actually perform the analysis. That is certainly true with the data cleanup task, which depending on the method used, can take many hours if the user is working with a large data collection. In particular, data cleanup is applied to the following fields within patent analytics:

- **Patent Assignee Cleanup** – Perhaps the most often used cleanup task due to misspellings and alternative representations of company names. The cleaning of this particular field is accomplished by a variety of different methods including bootstrapping, fuzzy logic algorithms, and manual methods such as PivotTables.

- **Inventor Name Cleanup** – Misspellings are also common in this field, but just as frequently there are issues with whether an individual uses their full first name, their middle name, or initial, and in the case of the last name, whether they have changed it due to marriage. Asian names, and the order in which they appear also cause problems when cleaning this field.

Besides cleaning inaccurate data, grouping is also performed in order to aggregate data in various patent analysis applications including the following:

- **Technology Categories** – synonymous terms from text collections within patent data fields, such as the abstract, claims or examples, are grouped together to represent a single concept. Cancer, for instance, can be described using additional terms such as neoplasm. Grouping terms when creating technology categories are required to ensure that all relevant occurrences of a concept are captured during analysis.
- **Up Posting Classifications** – Technical subject matter is often classified by hierarchical lists of concepts but analysts may not always want to use them at their most granular, or detailed level. When this is the case, more detailed classification can be grouped together, and collected as part of a higher-level entry. Using IPC classifications as an example, an analyst may group several subgroups together, provide them with a name that represents a meaningful concept, and then processes them as a single entity.

Generally speaking, methods for cleaning up data lists, or grouping concepts, in patents can essentially be divided into two categories: manual and automatic.

Manual Methods

These methods require the analysts to work with each individual entry in a data field. It requires that the analyst knows the relationship between one entry and another, and to be able to decide, if they are in fact, the same entry. These methods can be very time consuming, and can be accomplished using Pivot Tables or by manipulating lists, one object at a time.

- **Pivot Tables –** Pivot Table functionality allows a user to select one or more rows in a spreadsheet and group them together. Once they are grouped the new group can be renamed to reflect the desired name for the combined entry. See Figure 1 for an example of using a Pivot Table for patent assignee cleanup.

- **Drag and Drop or Manual Grouping –** Some tools provide a list of entries within a field and allow the user to either drop one entry on to the top of another to create a grouping, or allow multiple entries to be selected, and then grouped by right-clicking or pushing a button. Once the entries are grouped the new group can be renamed to reflect the desired content of the combined items.

	A	B
20	▼ BRUSHTIME PRODUCTS, INC.;BOOKER, Winifred, J.	
21	BRUSHTIME PRODUCTS, INC.;BOOKER, Winifred, J.	1
22	▼ C&C. LTD.	
23	C&C. LTD.	1
24	▼ CHEN, Chunmei	
25	CHEN, Chunmei	1
26	▼ CHIN CHUZAN;沈 仲山;CHIN SHUNRYO;沈 俊 良;CHIN CHUN-CHIUNG;沈 俊 ⑦	
27	CHIN CHUZAN;沈 仲山;CHIN SHUNRYO;沈 俊 良;CHIN CHUN-CHIUNG;沈 俊 ⑦	1
28	▼ CHOI, Byeong Gap	
29	CHOI, Byeong Gap	2
30	▼ CJLION CORP.;C&C. LTD;LEE, Eul Kyou;CHOI, Il Gyu	
31	CJLION CORP.;C&C. LTD;LEE, Eul Kyou;CHOI, Il Gyu	1
32	▼ Colgate	
33	COLGATE PALMOLIVE CO	2
34	COLGATE-PALMOLIVE COMPANY	3
35	COLGATE-PALMOLIVE COMPANY;FONTANA, Jose Eder;LEMOS, Edilberto;PERNA, Fernando;FOCASSIO, Paulo	1
36	COLGATE-PALMOLIVE COMPANY;PATEL, Madhusudan;GATZEMEYER, John J.;JIMENEZ, Eduardo J.;KENNEDY, Sharon	1
37	COLGATE-PALMOLIVE COMPANY;WONG, Chi Shing;FONTANA, Jose Eder;FOCASSIO, Paulo	1
38	▼ CRISP, Jackson	
39	CRISP, Jackson	1
40	▼ DELTA OF SCIENCE APS;LYSTLUND, Thomas	
41	DELTA OF SCIENCE APS;LYSTLUND, Thomas	1
42	▼ DENTALPOINT AG	
43	DENTALPOINT AG	1
44	▼ DENTEK ORAL CARE INC	
45	DENTEK ORAL CARE INC	1
46	▼ Dentsoll	
47	DENTSOLL KOREA CO., LTD.;KIM, Yun soon	1
48	DENTSOLL KOREA CO.,LTD.;KIM, Yun Soon	1
49	▼ GC CORP;株式会社ジ⑦シ⑦	
50	GC CORP;株式会社ジ⑦シ⑦	1

Figure 1 - Patent Assignee Cleanup Using Pivot Table Function in Microsoft Excel

Automated Methods

These methods allow the processing of an entire field of data based on the use of an algorithm, or an agreed upon collection of facts. While not as time consuming as manual methods the accuracy of the method is only as good as the algorithm used, or the knowledge of the individuals who built the collection of facts. Automated methods include the use of Fuzzy Logic, and Bootstrapping.

- **Fuzzy Logic –** An algorithm looks at the string of characters associated with an entry and determines the likelihood that two items represent the same entry based on how similar they are to one another. There are a variety of different algorithms that take this approach and many can be tuned based on the amount of similarity required to form a grouping.

- **Bootstrapping –** This method involves the collection of lists of standardized items that can be used to group together entries consistently. The organizations that build databases will often use these lists to make certain that they consistently apply the same name to an entry regardless of how it may have been used in the source material. Lookup tables can be used to transform raw input into standardized entries with large collections.

The following tools provide for some form of data cleanup or grouping. This is not an exhaustive list but provides some suggestions for starting with this task. Contact information for these tools can be found in section 9.1 of these guidelines.

- STN AnaVist – provides manual grouping which can be saved by user
- VantagePoint – allows for the creation of user defined thesaurus and fuzzy logic grouping

- Orbit.com – provides manual grouping which can be saved by user, online
- Microsoft Excel – includes Pivot Table functionality for manual grouping

A recent blog post, covering the use of Open Refine to assist with patent assignee cleanup can be found at: http://www.patinformatics.com/blog/patent-assignee-cleanup-using-google-refine-open-refine-text-facets-and-clustering/

An additional resource available for grouping assignee names is provided by the Organization for Economic Co-operation and Development (OECD), which has developed a collection of cleaned patent assignee names that can be used with bootstrapping methods. OECD provides the following description of the HAN database[56].

The OECD "Harmonized Applicants' Names" database provides a dictionary of applicants' names which have been elaborated with business register data, so that it can easily be matched by all users. The data is based on applicants for patents filed to the EPO and through PCT. The dataset is complementary to Eurostat's method for harmonizing applicant's names.

6.2 – List Generation

List generation is a statistical method that provides counts of various patent related metrics within individual data fields. Identifying the top ten inventors, for instance, is an example of generating a list of the most frequently occurring inventors in a patent collection. Figure 2 provides an example of a list generated in Microsoft Excel. The preferred method for visualizing a list is by generating a bar chart or histogram. This is a two-dimensional analysis with list entries typically displayed along the x-axis and the number of occurrences of the entry presented on the y-axis. Lists, and bar charts are used frequently in patent analytics, and within PLRs, since they allow for the statistical comparison of two or more entities in the same data field. In order to accurately account for the entries in a field, it may require standardization, or cleaning, as discussed in section 6.1.

A great deal of list manipulation is accomplished in a spreadsheet application, like Microsoft Excel, where each data field being analyzed is found in a column and the discrete items being analyzed are contained in the rows. Columns, or data fields can be sorted in a variety of different ways, but they are principally either in descending or ascending order. In a column containing alphabetical data, descending order would start with entries containing the letter Z, and would work down the column to the letter A. Ascending order would do the opposite, and rows containing the letter A would be at the top and would proceed downward towards Z. In a column containing numeric data, descending order would begin with the highest number and progress towards the smallest. Ascending order in a numeric column would begin with the smallest number and move towards the largest. In some cases, the same number may appear in multiple rows, in these circumstances there is normally a secondary sort column, which is used to order the entries that share the same value within the primary sort column.

Be careful when working with numerical data columns since sometimes they are entered as alphabetic characters as opposed to numbers. When this occurs the sorting will place 11 and 12 next to 1 as opposed to 2, as it should be. In most spreadsheet programs it is simple to select a column and convert it from alphabetic to numeric entries by changing the format of the entries in the column.

When sorting within a single column, it is also important to make certain that all of the other columns in the data collection get sorted along with the one of interest. Most spreadsheet programs

---

[56] http://www.oecd.org/sti/inno/oecdpatentdatabases.htm

automatically ask the user if they want to expand to all columns when sorting, or if they want to sort the column individually and leave the other columns as they were. It is almost always the case, when generating lists for patent analytics that the user will want to expand the sorting for all of the columns.

The number of times an entry appears in a list can be generated in a few different ways. A filter can be used to remove all of the rows in a spreadsheet except for the ones that contain a specified value. After the filtering is complete the analyst can count the number of rows that remain and assign the number to that entry in a separate worksheet. This process is repeated for the other unique entries in the column, or field. A more efficient way of accomplishing this process is by selecting the rows and columns of interest and generating a Pivot Table from them. Once generated, usually in a separate worksheet, the application will allow the analysts to decide which data field should represent the rows in the table, and which should populate the columns. In a standard list, no additional columns are added to the table, and the sum of the field being used for the rows is used in the values portion of the table builder. This sequence will be re-visited in section 6.3, but an additional field is added as a column for comparing the fields.

Company	Patents
P&G	11
Colgate	8
Kim Hye Kyung	5
Bosch Cerda	2
CHOI, Byeong Gap	2
Dentsoll	2
LEE, KYUE HYOO	2
LEE, KYUNG HO	2
MCNEIL PPC INC	2
MURPHY, PAUL	2
Profimed	2
Ranir	2
TIPHONNET JOEL	2

Figure 2 - List of Patent Assignees Using MS Excel

Once the list is generated, the rows and columns representing the list can be selected, and a bar chart can be created by pushing a chart button, or by selecting the appropriate menu item within a spreadsheet program. In Microsoft Excel, once a collection of rows and columns is selected, a bar, or column chart is created by choosing the option in the Charts portion of the Ribbon or under the Charts Menu Bar. At the beginning of this section, a bar chart was described where the field entries were listed on the x-axis. In Excel, this sort of horizontal chart is referred to as a column chart. If the field entries are listed on the y-axis instead, this is referred to as a vertical chart, and called a bar chart in Excel. Figure 3 shows a column chart visualizing the data contained in the list from Figure 2.
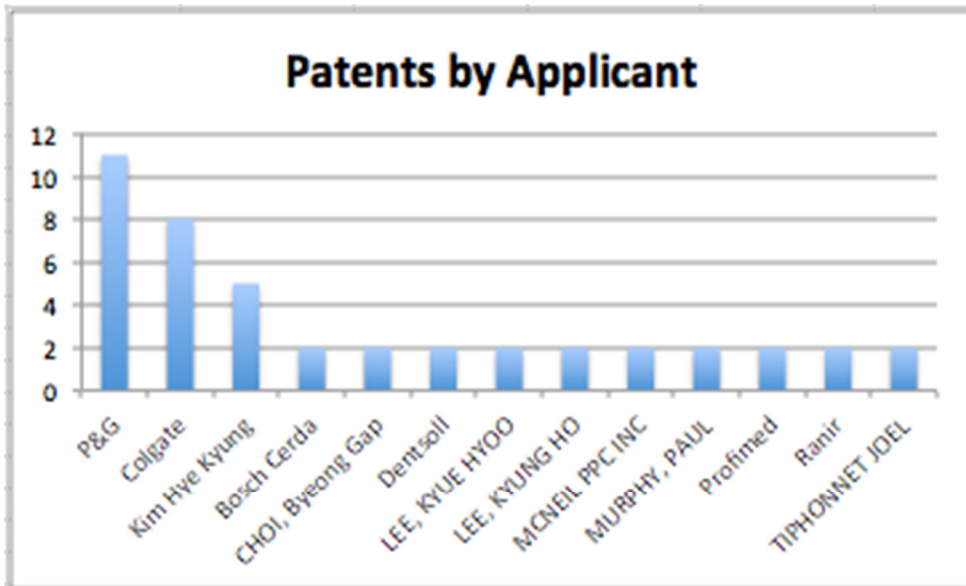
Figure 3 - Column Chart Visualization of Data from Figure 2

Most of the analytics tools provided in section 9.1 include functionality for generating simple charts of most patent information fields associated with the systems.

6.3 – Co-Occurrence Matrices

Counting or comparing entries within a single data field is accomplished by generating a list, as covered in the previous section. Frequently, an analyst will want to provide additional context to a comparison by incorporating an additional variable or field into their analysis. For instance, it is useful to understand which organizations have the highest number of patent documents in a technical field, but more insight is gained by studying when those organizations filed the patent documents in question. It is often the case that two organizations will have a similar total number of documents, but one of them will have filed them much earlier than the other. Using a co-occurrence matrix, or table, this additional context can be readily represented.

Co-occurrence matrices are also referred to as tables, and are generated by placing data fields on an X and Y-axis, or within a table. The number of overlapping occurrences of shared X and Y can be seen as numbers within the matrix. This representation allows connections to be made between two or more fields of information and provides a measure of how strong the connection is. This is interpreted by looking at the value provided at the intersection, the higher the value, the stronger the correlation between the two elements.

Most of the analytics tools provided in section 9.1 include functionality for generating co-occurrence matrices using patent information fields associated with the systems, but many analysts use Pivot Tables in Microsoft Excel to generate them as well. As discussed in section 6.3, the rows and columns of interest are selected, and a Pivot Table is generated using them. Once generated, usually in a separate worksheet, the application will allow the analysts to decide which data field should represent the rows in the table, and which should populate the columns. In a co-occurrence matrix, at least one data field is added as column to the table, as well as a field for the rows and the sum of the field being used for the rows is used in the Values portion of the table builder.

The same issues of sorting, cleaning and filtering apply to a co-occurrence matrix as they do to a list. Figure 4 shows an example of a simple matrix comparing citing assignee by publication year.

Count of Publication Number	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2004	2005	2006	2010	(blank)	Grand Total
P&G	14	19	3	5	6	7	8	21		13	1	5				102
McNeil	10	8	8	6		5	7	14	3			2				63
Intl Tape Partners	2	4		2	4	9	6	9								36
Colgate Palmolive	5	2		1	1	6	9	7	1		1					33
Braun		4			1			11	5							21
BERGMAN MARK	2	4		2	2		8	2								20
Warner Lambert	1					15										16
Kerrhawe				6		2		6								14
LG							6			6						12
Johnson & Johnson	2	1	4				2				1					10
Whitehill Oral Tech	1	5		2		1			1							10
WL Gore	4	3		1		1										9
Grand Total	41	50	15	25	14	46	46	70	4	25	2	8				346

Figure 4 - Building a Co-occurrence Matrix Using Microsoft Excel

6.4 – Clustering and Classification

These methods are often used interchangeably but are actually quite different from one another. Clustering is normally associated with unsupervised methods of organizing document collections based on a similarity comparison between documents. With a fixed number of clusters identified at the outset, document collections that meet a threshold similarity component are grouped together. Ideally, the documents within a cluster should be similar to one another but dissimilar to documents in the other clusters. Classification, on the other hand, is usually accomplished using a supervised machine learning method that uses learning sets to identify key attributes of documents in a class. New documents are compared to the learning collections and assigned to a class based on their similarity to the documents that have already been assigned to the class.

The following stackoverflow.com webpages provide explanations of clustering vs. classification and supervised vs. unsupervised machine learning methods:

http://stackoverflow.com/questions/5064928/difference-between-classification-and-clustering-in-data-mining

http://stackoverflow.com/questions/1832076/what-is-the-difference-between-supervised-learning-and-unsupervised-learning

When it comes to clustering, the two most often used algorithms are k-means and force-directed placement:

- **K-means** – a method of cluster analysis, which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.[57]
- **Force Directed Placement** – At the most basic level the algorithm tries to place similar objects close together and dissimilar objects far apart. The process is achieved by moving the objects randomly around the solution space via a technique similar to 'simulated annealing'. The criterion for moving a node is the minimization of energy[58].

Looking at classification, two frequently applied algorithms are Artificial Neural Networks and Support Vector Machines:

- **Artificial Neural Networks –** In computer science and related fields, artificial neural networks are models inspired by animal central nervous systems (in particular the brain) that are capable of machine learning and pattern recognition. They are usually presented as systems of interconnected "neurons" that can compute values from inputs by feeding information through the network[59].
- **Support Vector Machines** – supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories; an SVM training algorithm builds a model that assigns new examples into one category or the other[60].

As applied to PLRs, and patent analytics, the most frequently used sources of text for both clustering and classification exercises come from patent classification codes, or from raw, or standardized text coming from the source document:

- **Classification Codes** – Intellectually assigned classification systems produce standardized codes that can be used as a means of categorizing documents that share similar subject matter.
- **Raw Text** – processed to identify concepts and phrases contained within specific sections of the source document, such as the abstract or claims. As with the clustering of structured data, concepts, instead of codes, are used to group documents that share a high degree of overlap.
- **Indexing Terms** – producers of "abstract and indexing" databases normally produce hierarchical lists of indexing terms that are used to classify documents based on standardized terms and phrases. Since these lists are standardized and intellectually assigned they can be used for clustering exercises.

For additional discussion on the use of machine learning methods in patent analytics please see the following blog posts on the subject:

---

[57] http://en.wikipedia.org/wiki/K-means_clustering
[58] http://citeseerx.ist.psu.edu/viewdoc/download/doi:10.1.1.14.1822
[59] http://en.wikipedia.org/wiki/Artificial_neural_network
[60] https://en.wikipedia.org/wiki/Support_vector_machine

http://www.patinformatics.com/blog/machine-learning-in-patent-analytics-part-2-binary-classification-for-prioritizing-search-results/

The following tools provide clustering or classification functionality. This is not an exhaustive list but provides some suggestions for starting with this task. Contact information for these tools can be found in section 9.1 of these guidelines.

- Thomson Innovation – provides text clustering based on enhanced titles and abstracts, using K-means, see Figure 5 for an example
- Relecura – clusters concepts, extracted and standardized from text, in Topic Map functionality
- Intellixir – provides clustering functions, using K-Means
- Treperal – KMX product includes classification function based on Support Vector Machine
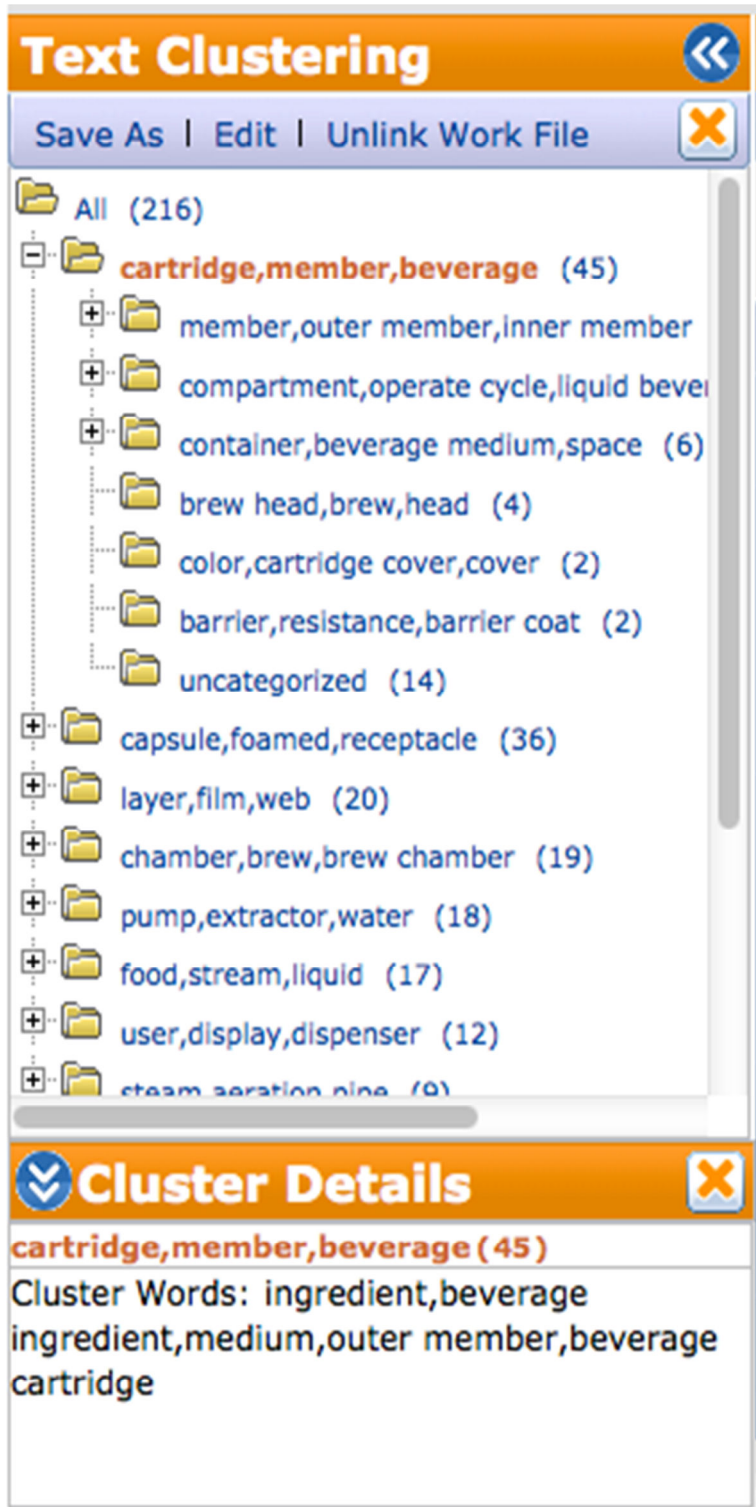
Figure 5 - Text Clustering Example from Thomson Innovation

6.5 – Spatial Concept Mapping

Mapping is related to clustering or classification exercises, where the systems involved take the document clusters or classes and arrange them in 2-dimensional space by considering the similarity of the documents relative to one another over the entire collection. Documents that share elements in common are placed closer together spatially, while ones with less similarity are placed further away.

The FAQ section on the IN-SPIRE tool[61], a related cousin of the ThemeScape tool, both originally developed at Pacific Northwest National Laboratories, provides the following explanation of the process used for creating spatial maps:

In brief, IN-SPIRE™ creates mathematical representations of the documents, which are then organized into clusters and visualized into "maps" that can be interrogated for analysis.

More specifically, IN-SPIRE™ performs the following steps:

- *The text engine scans through the document collection and automatically determines the distinguishing words or "topics" within the collection, based upon statistical measurements of word distribution, frequency, and co-occurrence with other words. Distinguishing words are those that help describe how each document in the dataset is different from any other document. For example, the word "and" would not be considered a distinguishing word, because it is expected to occur frequently in every document. In a dataset where every document mentions nanotech, "nanotech" wouldn't be a distinguishing word either.*
- *The text engine uses these distinguishing words to create a mathematical signature for each document in the collection. Then it does a rough similarity comparison of all the signatures to create cluster groupings.*
- *IN-SPIRE™ compares the clusters against each other for similarity, and arranges them in high-dimensional space (about 200 axes) so that similar clusters are located close together. The clusters can be thought of as a mass of bubbles, but in 200-dimensional space instead of just 3.*
- *That high-dimensional arrangement of clusters is then flattened down to a comprehensible 2-dimensions, trying to preserve a picture where similar clusters are located close to each other, and dissimilar clusters are located far apart. Finally, the documents are added to the picture by arranging each within the invisible "bubble" of their respective cluster.*

Spatial concept maps can also be made using classification methods. Arguably, the most famous of these is the Kohonen Self Organizing Map (SOM):

Kohonen Self Organizing Maps – a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a low dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space[62].

For additional discussion on the use of spatial concept maps in patent analytics please see the following blog post on the subject:

http://www.patinformatics.com/blog/machine-learning-in-patent-analytics-part-3-spatial-concept-maps-for-exploring-large-domains/

---

[61] http://in-spire.pnnl.gov/faq_7.stm

[62] http://en.wikipedia.org/wiki/Self-organizing_map

Additional discussion on spatial concept maps can also be found in section 8.6.2 of these guidelines.

The following tools provide spatial mapping functionality. This is not an exhaustive list but provides some suggestions for starting with this task. Contact information for these tools can be found in section 9.1 of these guidelines.

- Thomson Innovation – ThemeScape, shown in Figure 6 is a K-means clustering, with a mapping of n-dimensions onto two dimensions
- STN AnaVist – uses Force Directed Placement to generate Concept Maps
- Orbit.com – provides concept map using similar principles as ThemeScape
- Treperal – KMX product generates maps using K-means clustering

Figure 6 - ThemeScape Concept Mapping for Wearable Fitness Bands

6.6 – Layering or Stacking Information

Analyses looking at a single variable, or field, can be inefficient and lack context as discussed in section 6.3. Positioning two types of visualizations next to one another, or adding overlays to an analysis, allows the analyst to reference several attributes of a data set simultaneously without asking the client to refer back to previous illustrations. Providing additional context within a single visualization also allows for easier, richer comparisons to be made between different entries. Three examples of this technique are provided to demonstrate the value of this approach.

A stacked chart can be used to enhance a standard bar, or column chart when there is a need to explore a second variable with a small number of entries. If the second variable had a large number of values then a co-occurrence matrix would be used to explore these items. Figure 7 shows a stacked chart, where patent document type has been stacked within the patent document count by patent assignee. This information could have been provided with separate charts, but it is much more interesting to combine the variables, in a single chart, for means of comparison. In this case, the analyst can demonstrate whether an organization's patent documents are primarily of the utility or design type.
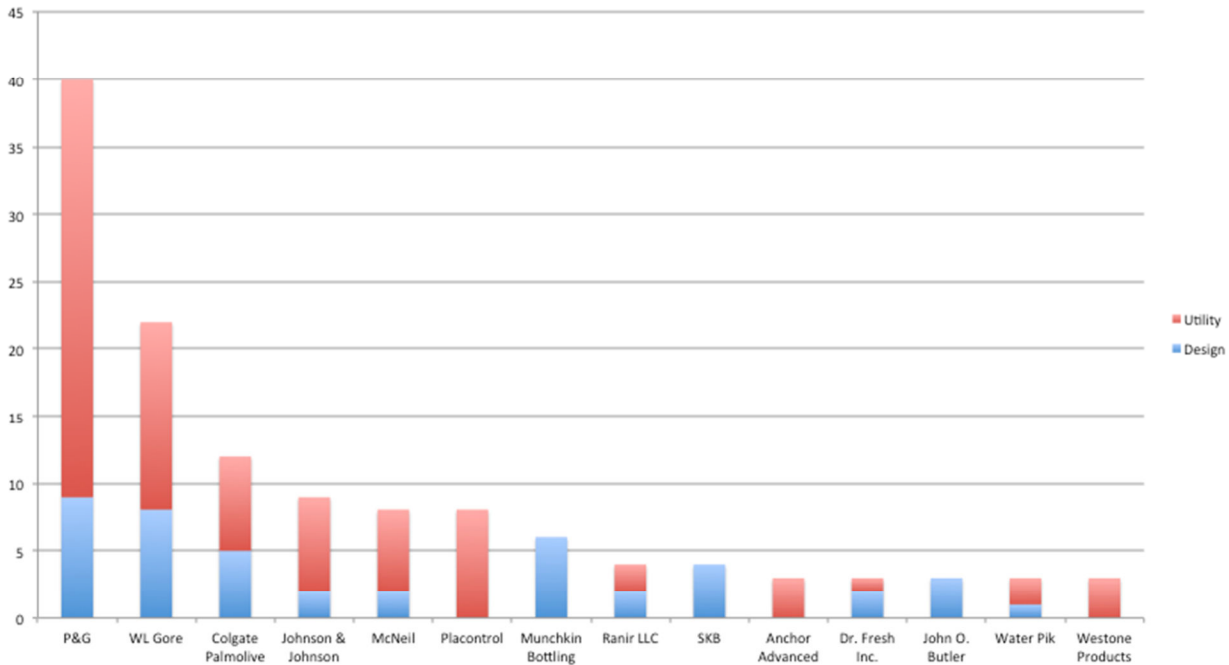
Figure 7 - Using a Stacked Chart to Visualize Two Variables

In section 6.5, the mapping task was introduced, taking document clustering, and adding a similarity metric between documents to produce a graphic representation of how documents relate to one another based on shared concepts. As stated, these maps offer a nice way to identify technology segments that are related to one another. When additional information is layered on top of the maps they can be used to add an extra dimension to an analysis involving technological concepts. Most mapping systems provide a means to highlight, with the use of different colors, two or more patent assignee or periods of time within the collection used to generate the map. These groupings are then laid over the existing map and can be used to provide context on when technology subsections were investigated, or which organizations were investing in different areas compared to others. Figure 8 shows an example of a Thomson Innovation ThemeScape map where a few of the patent assignees of interest have been called out with different colored dots.
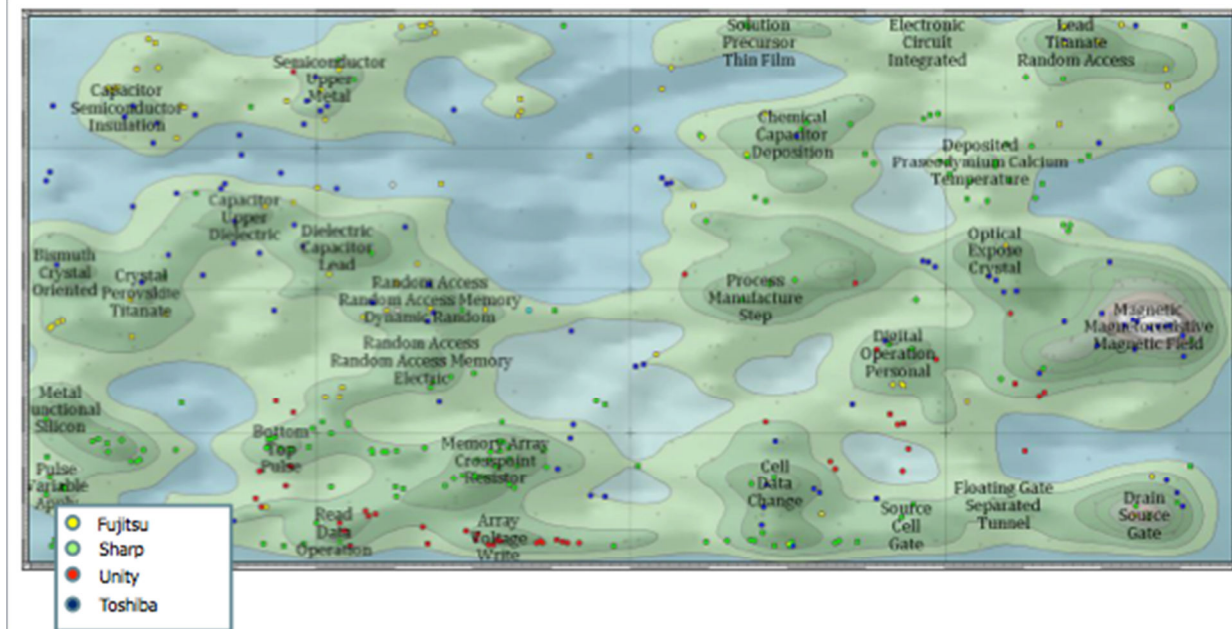
Figure 8 - ThemeScape Map with Patent Assignee Call Outs

Charts, as discussed in section 6.2 are used to visualize total interest, or number of entries in a single, or at most two patent information fields. Tables, covered in section 6.3 allow the examination of two fields, both with a significant number of entries. The two can be combined into a single visualization that expands the number of variables being studied simultaneously, and allows more complicated questions to be answered without needing to resort to multiple visualizations. In the example provided in Figure 9 the chart showing top assignees by their patenting type is turned 90-degrees, and placed next to a table showing the top assignees, and when they filed applications for the patents in question. In one visual the analyst can now make comparisons between which organizations are the most active patentees, when they invested in the technology, and what type of protection they sought. Again, this could be provided in three or four individual illustrations but it is easier to make detailed comparisons when the data is stacked or layered in a single illustration.

Figure 9 - Visualization Encompassing Chart and Table to Study Multiple Variables

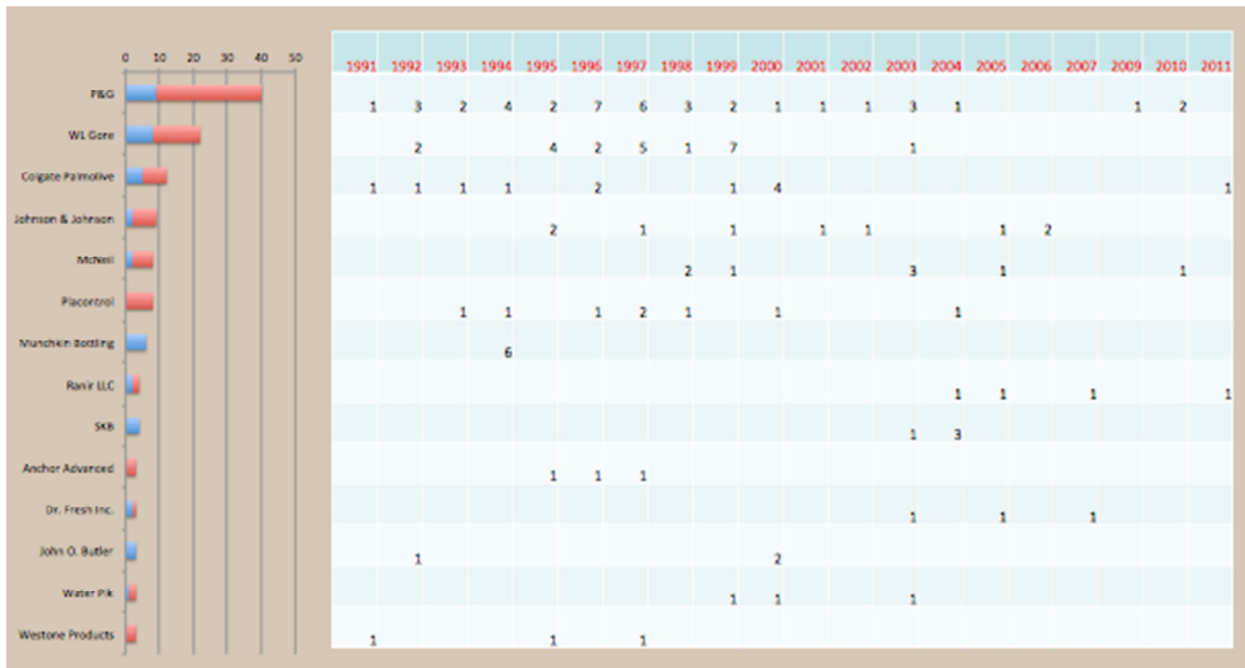|	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2009	2010	2011
P&G	1	3	2	4	2	7	6	3	2	1	1	1	3	1				1	2	
WL Gore			2		4	2	5	1	7				1							
Colgate Palmolive	1	1	1	1		2			1	4										1
Johnson & Johnson					2		1		1		1	1			1	2				
McNeil								2	1				3		1			1		
Placontrol			1	1		1	2	1		1				1						
Munchkin Bottling				6																
Ranir LLC														1	1		1			1
SKB														1	3					
Anchor Advanced					1	1	1													
Dr. Fresh Inc.														1		1	1			
John O. Butler		1							2											
Water Pik								1	1					1						
Westone Products	1				1		1													

6.7 – Geographic Representation

On most patent documents the physical addresses of both the applicants, and the inventors associated with them are given. Geographic representations of this data provide the information overlaid on a city, country or world map along with relevant geographic placeholders, such as educational infrastructure, like major universities, or per capita income for the region. This process is also referred to as Georeferencing and is defined in Wikipedia as:

To georeference something means to define its existence in physical space. That is, establishing its location in terms of map projections or coordinate systems.[63]

The process is carried out by geocoding individual patent documents so they can be placed within an established geographic framework, typically by using zip, or postal codes, but also by using street addresses. Various software packages are available that can take address data from individual patents, and supply a set of longitude and latitude coordinates for the document. A collection of geocoding packages can be found at: http://en.wikipedia.org/wiki/Geocoding.

Once the coordinates for a patent document are identified, various Geographic Information Systems[64] can be used to create the actual map with the patent documents placed on it. Google Maps[65] makes it application programing interface (API)[66] available for free thus making it a useful tool for distributing patents over a geographic area.

---

[63] http://en.wikipedia.org/wiki/Georeference

[64] http://en.wikipedia.org/wiki/Geographic_information_systems

[65] https://en.wikipedia.org/wiki/Google_Maps

[66] http://en.wikipedia.org/wiki/Application_programming_interface

Recently, two services, UK Patents on a Map[67] and World Patents, Mapped[68], were launched providing examples of geographic representations of patent data. The home page of UK Patents on a Map provides the following rational for providing the service:

Launched on 3 July 2013, this site maps many of the UK's innovators, specifically UK based holders of UK patents. The data includes links to the relevant page of the IPSUM database operated by the UK Intellectual Property Office (IPO) where the latest and most accurate data concerning any patent can be found.

This is a service that I very much hope will be of help to inventors, patent applicants and owners as well as their advisors. My aim in placing this data into a map format (with the help of the lovely people at eSpatial) is that it brings alive the potential for collaboration when you see how close you are as a patent holder to others, whether in your sphere of innovation or otherwise.

The World Patents, Mapped service was described during an introductory post from the Patent Information Users Groups (PIUG) wiki:

World Patents, Mapped, is an experimental service that plots the locations of applicants on WIPO patent publications.

Universities have long been recognized as key players in innovation. For this reason, I am adding an overlay of university locations to World Patents, Mapped.

The overlayed map is at http://w.pat.tc/maptop.htm. This map shows locations of applicants on PCT patent applications published during 2012 and 2013. It also shows the locations of about 600 universities.

Figure 10 provides an example of WO applications filed for by inventors residing in Dublin, OH, USA.
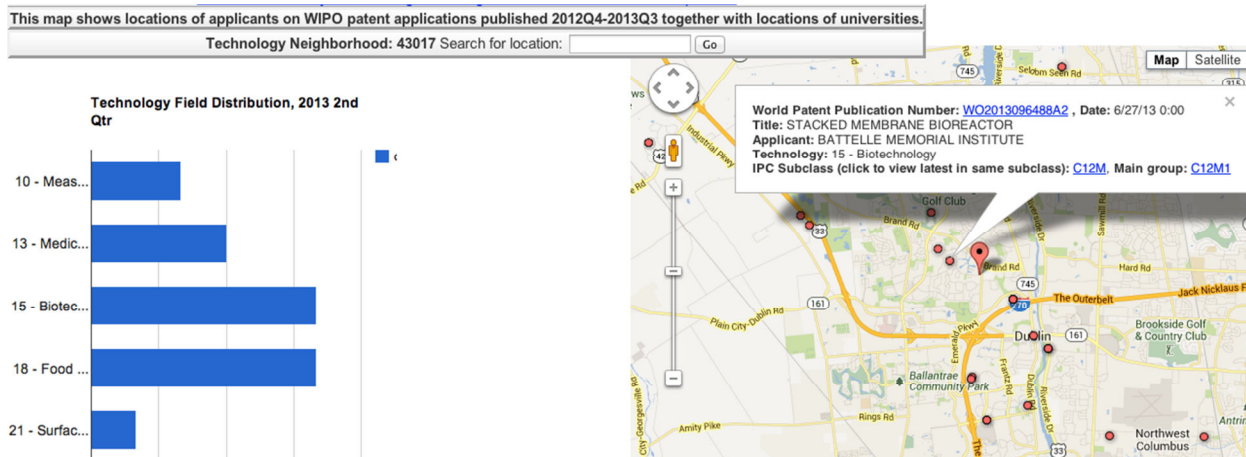

Figure 10 - WO Applications from Inventors in Dublin, Ohio, USA Using World Patents, Mapped

6.8 – Network Analysis

---

[67] http://www.patentsonamap.co.uk
[68] http://w.pat.tc/maptop.htm

Borrowing from Wikipedia's discussion on Social Network Analysis[69] and applying this idea to the analysis of networks within patent document collections, the following definition is proposed: Network analysis is the viewing of relationships in terms of network theory[70], consisting of nodes, representing individual actors within the network, and ties, which represent relationships between the individuals, such as co-inventorship, co-assignment and co-citation. These networks are often depicted in a network diagram, where nodes are represented as points and ties are represented as lines.

The two most common uses of network analysis, when looking at patent documents, are inventor and citation networks. Inventor networks, when also associated with patent assignees, allows an analyst to discover which organization's key researchers and thought leaders have worked with in a technological area. University Professors, for instance, often collaborate on patents with several different organizations and these relationships can be discovered with a network diagram. Their students can also be traced as they leave school and potentially embark upon their own academic career, or transition into industry themselves. This is a useful technique for identifying potential collaborators, or for looking for new employees, with experience in a technological area.

The concept of citations within patent documents was introduced in section 4.2.1.5. When patent documents are examined, relevant prior art is mentioned on the search report, or on the front page of the documents. These citations provide a link between the documents that can be analyzed and visualized in a network analysis.

The simplest citation network diagram involves the use of a single patent document as a root node and displays the direct citations, both backwards and forward, that are associated with that documents and its descendants. Hyperbolic trees are used, in this case, to show relationships between patents that directly cite one another. Figure 11 provides a sample hyperbolic tree analysis.

---

[69] http://en.wikipedia.org/wiki/Social_network_analysis
[70] http://en.wikipedia.org/wiki/Network_theory
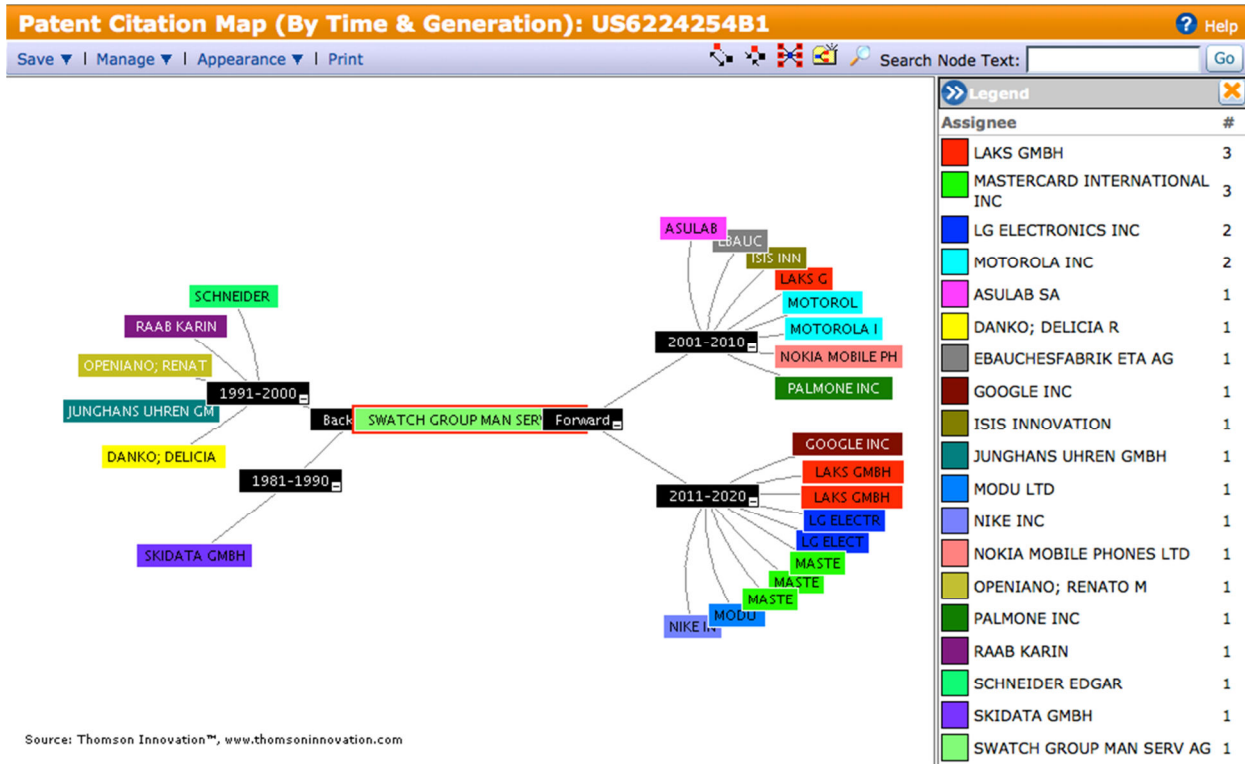
Figure 11 - Swatch Patent as Root in Hyperbolic Tree Citation Analysis

When an entire collection of patents, and their citations to one another are studied, a full network analysis can be graphed that identifies seminal patents, cited by many, over a period of time. A network approach has advantages over a hyperbolic tree since trees only show linear relationships between documents, from one generation to another, but do not show relationships that skip a generation. They are also advantageous since there is no single root, which represents a one-to-many relationship, but all of the documents, in the collection, are roots, providing for the viewing of many-to-many relationships.

While citation links exist between discrete patent documents, it is often useful to aggregate the patents from an assignee together, and create a citation network that shows the relationships that organizations have to one another via their citation of one another's patents. Figure 12 provides an example of a network graph of assignee citation from Orbit.com.
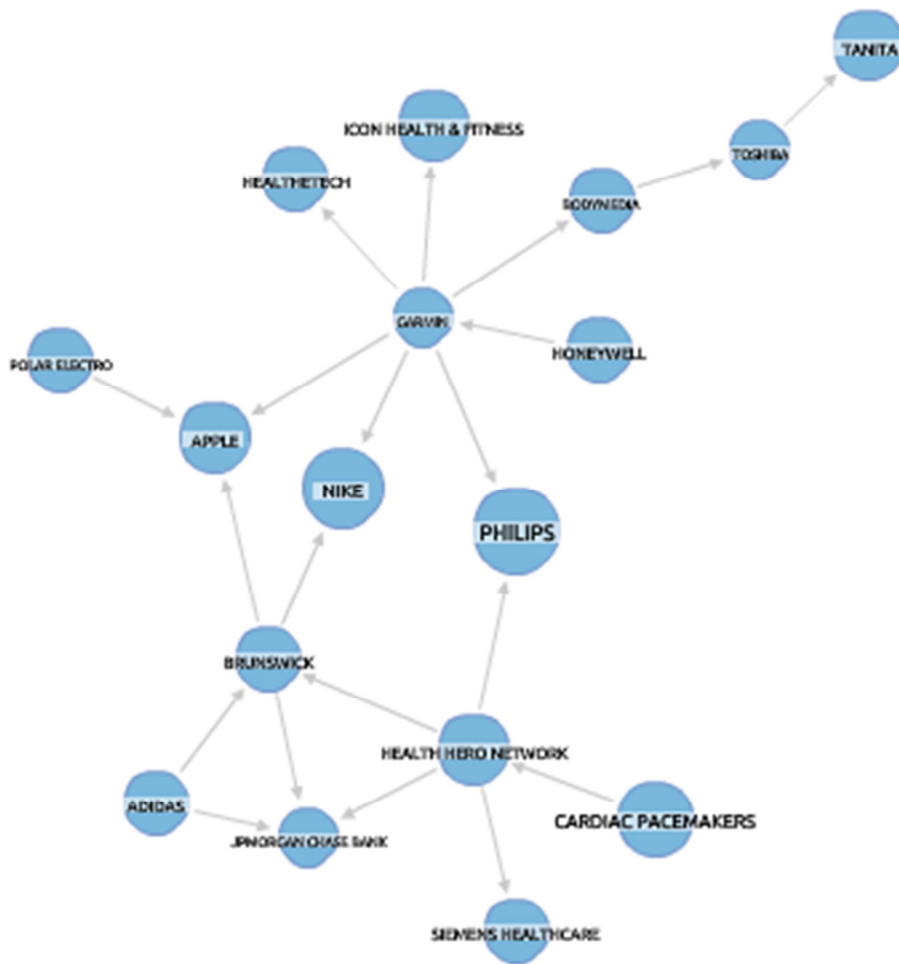
Figure 12 - Network Graph of Assignee Citation

Co-assignment, or organizations collaborating together on developing an invention, can also be a useful and insightful exercise, but since most inventions are developed by a single entity, there are simply not many examples of this type of relationship to study. This type of analysis can become more interesting when the ownership of a patent changes and the relationships between organizations buying and selling patents is added to the ones associated with co-development. The purchasing of patents between organizations is arguably an even more interesting relationship then when two groups decide to develop an invention together.

It is more common for patent analysis tools to make use of hyperbolic tress, rather than a full network analysis, but recently additional effort has been applied to the development of network analysis tools for patents. The following tools provide for some form of hyperbolic tree or network analysis. This is not an exhaustive list but provides some suggestions for starting with this task. Contact information for these tools can be found in section 9.1 of these guidelines.

- Thomson Innovation – provides hyperbolic trees starting with a single patent
- Orbit.com – provides a network graph of assignee citation within its business intelligence module

- Intellixir – provides network analysis for assignees and inventors
- AmberScope – generates citation networks based on Network Patent Analysis

6.9 – Semantic Analysis

While there are several semantic techniques, such as latent semantic analysis, that are useful in the context of information retrieval, and patent searching, the primary form of analyzing patent documents using semantic information is the study of subject, action, object (SAO) triplets. SAO triplets utilize parts of language, such as nouns and verbs that are used to describe the teachings that the applicant wants to share. Key SAOs, within a patent document encapsulate the technical learnings contained in it.

When extracting SAO triplets from a document, objects and actions can be grouped together and represented as a problem to be addressed. An example of this would be the treatment of asthma, where treatment is the action and asthma is the object. A potential solution to this problem would be the use of an allergy shot, where the allergy shot is the subject. This triplet would be extracted from a sentence like; an allergy shot was used to treat the asthma patient. Figure 13 provides a different example of the SAO triplet principle.
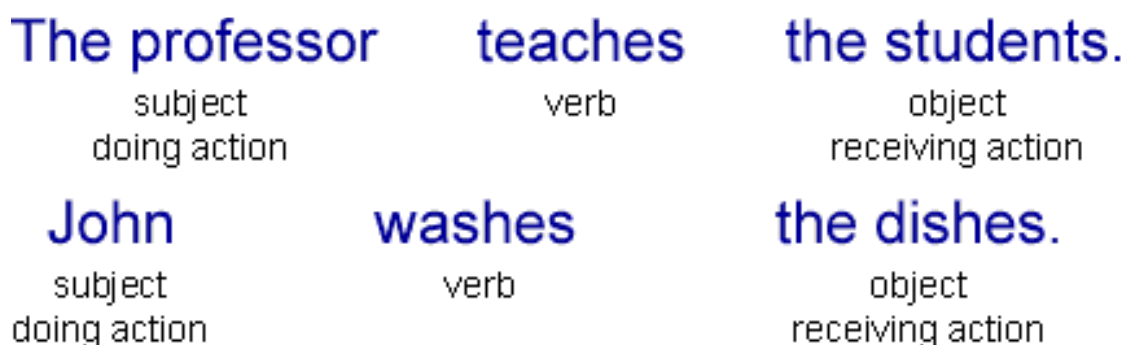
Figure 13 - Subject Action Object Triplet Example from English Text

The aggregation of a collection of problems gathered from a set of disparate documents can be represented as a knowledge base, and can provide a variety of potential solutions even if they are not found within a single document. In the asthma example, a sentence in another document may refer to the treatment of asthma, but the solution, or subject, in this case, might be immunotherapy.

Additional information on information extraction using SAO triples can be found at – https://www.ibm.com/developerworks/community/blogs/nlp/entry/triple_extraction_from_unstructured_text5?lang=en

In the realm of patent analytics and PLRs SAO triplets can be used to summarize the key technical challenges in a specified area and provide an index of the approaches used to address those challenges. Portions of this activity are sometimes done manually, by an analyst, or can be accomplished using classification techniques, as discussed in section 6.4. SAO triplets provide the additional benefit of providing solutions to the technical challenges as well as identifying them.

The I2E system from Linguamatics[71] and Goldfire from IHS[72] are two commercial systems that provide SAO triplet functionality specifically for working with patent documents. Information on open source approaches for extracting them can also be found at – http://stackoverflow.com/questions/8063334/extract-triplet-subject-predicate-and-object-sentence.

---

[71] http://www.linguamatics.com/welcome/software/I2E.html
[72] http://inventionmachine.com/products-and-services/innovation-software/

Chapter 7: Frameworks Associated with Performing Patent Analytics and Patent Landscape Reports

When developing a patent analytics output, there are certain fundamental ideas, or philosophies, associated with conducting the work. These thoughts apply generally to almost any task associated with patent analysis and are provided at a higher level of abstraction than the tips and instructions associated with specific analysis. They are referred to as frameworks since, as with a building scaffold, they provide a foundation on which more detailed and specific analysis tasks can be attached. In this chapter various frameworks, or general principles associated with patent analytics, and landscapes are discussed.

7.1 – Content Types for Conducting Analysis

The vast majority of analytics projects can be broken into two categories; those that work with data in the form of exact strings in structured fields, and those that work with text or semantic content that is unstructured. Analytics utilizing these different sources of content are referred to as data mining or analytics, and text mining or analytics respectively. Some data mining purists consider text as simply a more complicated form of data, but for the purposes of this discussion they will be separated since the methods used to work with exact strings in data mining is different than how text is analyzed in text mining.

7.1.1 – Data Mining

When analysts are working with numbers, or strings of letters divorced from context or semantic meaning, they are performing data mining or analysis. The practitioner is generally conducting a statistical analysis of a list of items to identify patterns within a collection.

In the most common form of this type of analysis, discrete items such as Patent Assignees or Application Filing Years, are counted, and potentially ranked within a set. Tools, such as a Pivot Table in Microsoft Excel, are looking for exact string matches to determine how frequently a specific value is found. All analyses in this category are based on exact string matches so misspellings or inconsistencies within a set will be counted as separate values. For instance, when working with Patent Assignee Names, if there are entries in a data set for Vertex Pharmaceuticals and Vertex Pharmaceuticals Inc., these will be treated as discrete items, and will be counted separately. Since the collections of alphabetic strings do not match exactly the analysis will count these items separately. Clearly, these items belong together, and as discussed in chapter 6.1, the process of cleaning up or grouping these items together is required to ensure that an accurate count of the appropriate values takes place.

Another important component of data mining or analysis is that the material being analyzed is structured or fielded. When items are placed in a specific field, within a database, or when they are found in a specific column, in a spreadsheet, they are considered structured or fielded. This implies that they belong to a category of one type or another. Inventor Names, for instance, are recognized as being data of a particular type and are segregated and organized so they are collected together in one place for analysis. Additional names might be found within a document but the analyst can distinguish these other names from the Inventor Names since the later has been structured by being collected into a specific field.

Data mining is thus characterized as the analysis of exact string matches contained in structured or fielded databases or collections.

7.1.2 – Text Mining

Linguistic content is distinct from data strings for a variety of reasons. To start with, there is the concept of language, where a word can mean the same thing but is spelled differently. There is also the concept of context where the same word is used but it has a different definition depending on how it is used. Similarly, parts-of-speech can be considered when analyzing text data, where a word can be used as a verb in some cases and as a noun in others.

Raw text is generally considered to be unstructured or semi-structured since the content is not organized into categories. According to Wikipedia, unstructured text[73] refers to information that either does not have a pre-defined data model and/or does not fit well into relational tables. A patent abstract would be an example of a semi-structured item since the patent abstract is a field in most databases, and there is an expectation on what type of content will be found there. The claims, on the other hand, while representing the legally binding portions of the text, can be very long and deal with a variety of concepts, and by nature is not structured into discrete items.

The methods and means for analyzing linguistic content, due to the complicated nature of the source, is very different than working with data, so it is important to consider, and understand them individually. As an example, the following series of steps might be conducted in order to prepare a collection of unstructured text for analysis:

- Tokenization – explaining to the computer where one word ends and another begins
- Stemming – removing common suffixes and prefixes from words to generate the root of a word for subsequent use
- Part-of-Speech Tagging – identifying words as nouns, verbs or adjectives
- Entity Tagging – using lists of items or linguistic rules to identify a token as a type or person, organization or other type of object
- Term Filtering – reducing the number of terms or objects to be analyzed by removing stopwords (non-content bearing terms), or frequently or infrequently applied terms, in a corpus or collection

In general, when conducting analytics associated with generating PLRs, the analyst needs to understand whether a data or text-based method is being performed. Since the methods involved are quite different, optimal results will depend on the analyst understanding the different approaches and applying them properly.

7.2 – Data Scale for Conducting Analysis

In addition to thinking about data collections based on their content, exact strings vs. raw text, data scientists also tend to think about data in terms of the size of the collection they are working with. Generally, this is done since different methods are used depending on how large the collection of data being worked with will be. Most analysis that end up in PLRs are concerned with larger data sets, these are being conducted on a macro-level, but occasionally it is necessary to provide more detailed examinations of small subsets, a micro-level analysis.

7.2.1 – Macro-Level Analysis

---

[73] http://en.wikipedia.org/wiki/Unstructured_data

Also referred to as a global-level view, analysis at this scale if being performed for health care or other socially related data collections, would be done on a population-wide level. In the area of patent analytics, macro-level data sets contain greater than 10,000 records being studied. Since PLRs are generally broad overviews of a topic area, most of the analytics that go into them are conducted on the macro scale. When working with macro-level collections there is a greater reliance on computational methods due to the amount of time and effort it would take to analyze sets of this size manually.

7.2.2 – Meso-Level Analysis

Sometimes referred to as a local-level review, analysis at this scale if being performed for social data, would be done on a group-wide level. Thinking about patent analysis projects, meso-level data sets contain between 1,000 and 10,000 records. Many of the same methods used for macro-level analysis will also be used with meso-level collections since sets this large are difficult to manage when records are looked at individually. The computing resources and time required to perform these analysis are going to be less than what is done on the macro-level. Many PLR analytics are conducted on this level when sub-collections within a broader topic area are explored. The practice of working with subsets of a larger whole is sometimes referred to as "drilling into" a data set.

7.2.3 – Micro-Level Analysis

Sometimes referred to as the individual level, analysis is generally conducted on a one-on-one basis. Thinking about patent analysis projects, micro-level data sets contain less than 1,000 records and frequently is done on collections of less than 100 documents. Many of the analysis done on this level are done manually and in circumstances where a high degree of precision and human ingenuity is needed to ensure a meaningful result. In work associated with PLRs, detailed analysis of this type is performed in order to confirm trends and associations discovered while conducting macro or meso-level analysis. This is especially the case when counter-intuitive results are obtained during larger scale analytics, and the analysis wants to better understand the cause of these trends. Certain activities related to PLRs, such as patent valuation, is often best done on a case-by-case basis.

7.3 – The Linear Law of Patent Analysis

The Linear Law of Patent Analysis was proposed as a framework for performing patent analytics projects in 2002[74]. It was originally developed to assist practitioners in understanding the importance of starting an analysis by investigating the needs of the customer for the analytics, as opposed to simply jumping in with an analysis tool. It has since been used as a general method for planning analysis projects. The steps in the process are:

- *Create a toolkit of analysis tools*
- *Understand the business need and the need behind the need*
- *The need drives the question*
- *The question drives the data*
- *The data drives the tool*

This is referred to as a linear law since in this framework the steps have to be followed in order to provide the best results. Often companies or analysts would start with the purchase of the tool and once that was accomplished, since a significant investment had been made in the tool, they would use it exclusively to conduct all of their analysis projects. In the suggested framework the choice of

---

[74] http://www.infotoday.com/searcher/oct02/trippe.htm

which tool to use is left as the last step once all of the other parameters associated with the analysis have been worked out.

The law starts with gathering a collection of tools or a toolkit. There is no one tool that can work with all sorts of data, and can conduct all types of analysis, so it is important for the analyst to have options. Some projects require semantic or linguistic analysis of text, others require the study of citation patterns and networks, and others still require studying the changes that take place within the text of a patent as it progresses through its life cycle. So within reason, given budget constraints, a suite of tools should be collected.

The next step speaks to understanding the business requirements that will be satisfied by conducting the analysis. Under ideal circumstances, the analyst should know precisely what decision a business leader would be making with the analysis provided. They should also have a good idea about the situation the organization finds itself in, why there is an issue with it, and have some idea how a preferred path forward might look. Analytical results should be told as a narrative to have the greatest impact with the decision maker, and understanding all of the context will allow the analyst to craft their results into a compelling story that drives decision-making.

Only after the needs are thoroughly understood can the analyst start suggesting questions, and potentially hypothesis that should be explored during the course of the project. The questions at least can be confirmed with the decision maker, and provides confidence that the analyst understands the needs and is thinking about ways to address them. Depending on the needs either one or several questions can be addressed.

Now that the questions have been established, the experiments can be developed that will either confirm or discredit the hypothesis associated with them. In the case of patent analytics, experiments are designed by considering the data that will be analyzed.

Finally, now that all of the other details have been worked out a decision can be made on which tool will provide the proper insight into the appropriate data to either support or dismiss the hypothesis. The use of the right tool is often critical to success as an analyst but their application must be applied under the proper circumstances to provide critical insight.

Additional background on the history of the Linear Law of Patent Analysis can be found at http://www.patinformatics.com/blog/the-linear-law-of-patent-analysis-revisited/.

7.4 – Precision and Recall

Information retrieval or searching effectiveness is traditionally described in terms of two measures, recall and precision. These items are defined as:

- *Recall – how much of the useful information has my search retrieved?*
- *Precision – how much of the information that I have retrieved is useful?*

There is also a useful probabilistic interpretation of recall and precision: recall estimating the probability that a relevant document will be retrieved in response to a query and precision estimating the probability that a retrieved document will be relevant.

Thinking about the issues in searching during the preparation of a PLR, information retrieval methods usually look at precision and recall simultaneously and measure their methods by how techniques

stack up against both elements. Even though this is the case, precision and recall are normally opposed to one another such that with an increase in recall there is usually a subsequent drop in the level of precision. Generally speaking, as searches are designed to maximize recall, the results can suffer since more off-topic references get included in the collection.

In generating collections for PLRs it might be more productive to begin with creating sets using methods that produce high recall exclusive of precision. When statistical analysis is performed on large or macro-level sets only major trends or items that appear frequently are going to be seen. Precision, in this instance, can suffer to some degree with these types of searches, since minor occurrences within these sets will not be seen in the larger context. This can often be evaluated by examining several of the significant trends to ensure that they are coming from reasonably precise references. If this is the case then it is general accepted to sacrifice some precision for the sake of recall.

7.5 – General Skill Set Requirements for Analysts

Proficiency as a patent analyst requires a collection of skill sets on the part of the individual performing the task. At a minimum a patent analyst should have experience in the following areas:

- Patent information – due to the idiosyncrasies and nuances of patent data it is critical that people who understand this collection intimately be the ones conducting the analysis. Patent information, perhaps more than almost any other data source, can be misinterpreted if the analyst is not familiar with the history and details of it.
- Data analysis and statistics – while most analysis tools and methods are semi-automated and don't require adjustments on the part of the analyst, optimal results are obtained when the practitioner thoroughly understands the variables and parameters associated with an analysis and can modify them as needed. The results of an analysis are also easier to understand and explain to a client when the analyst knows the method and how they manipulated it.
- Legal knowledge – while formal accreditation, such as passing a patent bar, is not required, a general understanding of the legal aspects of the patent system, especially in a worldwide context is certainly helpful. This is especially the case if interpretation of claim language is requires to conduct a PLR. Legal perspective is also useful for understanding patent families and how they relate to various national patenting systems.
- Presentation skills – one of the key features of a PLR is its ability to collect a large amount of information and provide a concise report of the key trends and observations in the area being studied. The ability to organize large amounts of data into a compelling story and present the results in an engaging fashion tailored to the learning style of the potential readers is essential to obtain maximum impact.
- Deductive ability – the launch of each PLR is a blank page with an open question that needs to be investigated. Looking at each project as a new mystery to solve, with its own unique challenges and outcomes, is required. Individuals who enjoy intellectual puzzles and discovering and exploring new topic area typically enjoy the deductive reasoning aspects of the analyst position.

While it is not necessary for a beginning analyst to have all of these skills as they get started, since many can be developed as they train, there should at least be an aptitude and interest in gaining all of them.

7.6 – General Thoughts on the Use of Visualizations

Visualizations are sexy; they grab the eye of the reader and can easily become the focus of a presentation or a PLR. They are an essential component of telling a story during the presentation of ideas and learnings discovered while conducting research for a project. While there is a place for informed and well-designed visualizations it is important to keep them in context as a means to an end, in the sharing of insights, and not as an end to themselves.

In the early days of using thematic maps, as an example, there was a tendency for users to focus on the maps themselves, as opposed to the process of generating the insights that the visualization is designed to share and highlight. This was caused by sloppiness in the creation of collections to be analyzed and subsequently, confusion on what insights the visualization was providing. Just as the Linear Law of Patent Analysis provides a means for conducting a well-reasoned analysis, a similar set of steps should be used to create visualizations that are developed once analysis has led to knowledge.

In a sense, visualizations can be thought of as the dessert portion of a meal. They make for a nice treat, and are something to look forward to, but they should not be providing the substance of the meal. Understanding the business needs of the project, generating relevant data collections and using the appropriate methods to analyze the data are the protein and vegetable components of a balanced meal. These items may seem boring or pedestrian, but they are essential to provide valuable knowledge on trends and perspectives. Once they are provided for then the client can enjoy a treat in the form of a visualization that concisely summarizes the insights discovered.

A recent Harvard Business Review article entitled, The Three Elements of Successful Data Visualizations[75] also provides useful advice on the generation of powerful visualizations, including the fact that they need to tell a story, which is covered in the next section. The article also suggests two additional pieces of advice:

1. Understand the audience – this was covered in chapter 5 of these guidelines.
2. Set up a clear framework – frameworks have been covered in this chapter but in this case the article is referring to the ensuring that the analyst makes sure that their data is clean and that they understand its nuances. Pre-processing of data before analysis is covered in section 8.3.

7.7 – The Story Telling Method

Humans have been telling stories as a means of communicating with one another for thousands of years. It is an established and well-engrained means for sharing thoughts and ideas with others in an engaging fashion. One of the outcomes of a PLR is to influence decision making by the application of relevant data and trends to the issue at hand. One means to share insights, in the context of addressing decision making, is to use the classic story telling method when considering how to structure a PLR, and especially a presentation that might be generated to further summarize the key takeaways from a PLR.

The story telling method involves the creation of a story arc that embodies the following elements:

---

[75] http://blogs.hbr.org/cs/2013/04/the_three_elements_of_successf.html

- Begin by talking about how things were in the past
- Move to how things are today
- Paint a picture of the likely future

Structuring a PLR in this way creates a compelling narrative that can increase the effectiveness of the report. Looking at what has occurred previous, using the analysis of patent information, and historical contexts, helps to frame the question being pursued and provides motivation for why it is important to pursue the project. There is also something to be said for the old adage, "Those who cannot remember the past are condemned to repeat it", especially during strategic discussions.

Moving to how things are currently lets the reader know where things stand in the present and allows for a comparison to what happened previously. When pursuing decision making options it is important for the reader to recognize the cause and effects that have transpired as an area has developed. Comparing what happened twenty years ago to the current situation could provide key insights to help predict what possible outcomes of decisions made in the present will have.

Finally, due to the inherent time lag associated with the filing of patents, it is also possible to begin to predict the future by suggesting the direction and scale of effort being applied by existing participants or new entrants to a technical field. In many fields, it may take three to five years, or more, from the time that an idea is developed, for it to be commercialized and put into production. Patent applications are normally published after 18 months so looking at very recently filed applications can provide a window into future plans. Building on the insights generated by looking at the historical data this additional knowledge can provide options for decision makers to consider and help identify the risks associated with the developing strategy.

Using the story telling method is a good way to keep clients engaged since the analyst is building to a conclusion that will be useful to them. It also provides a way to organize the information developed during the course of the project since it is structured to support decision-making and the exploration of various scenarios.

Chapter 8: Preparing Patent Landscaping Reports

To this point, the foundation has been laid for understanding all of the attributes and resources that are required to develop a patent landscape report (PLR). Previous chapters have covered the basics of patent information, the rationales and tasks associated with patent analysis and PLRs, and some thoughts on how to properly perform patent analytics. Generating an impactful PLR takes more than a few charts and graphs, regardless of how well they have been done; they also require careful planning and adherence to the idea that the component pieces are aligned to explore, and lead a decision-maker through a topic area for an explicit objective. This exploration, if done well, will lead to an informed decision, resulting in a change in organizational strategy, or direction.

This chapter walks through the entire process of generating a PLR, starting with the planning stage, where the audience, scope, and objectives for the report are agreed on. Followed by the performance of searches, to provide data, the preparation of the data, so it is in a state appropriate for analysis, and the subsequent analysis of the data that will be interpreted by the analyst, in order to provide insights, appropriate to the stated objectives of the report. The writing of the report, and the subsequent publication of it along with the appropriate, corresponding data, and potentially, interactive visualizations, will be covered. Finally, thoughts on the evaluation of completed reports will be discussed.

8.1 – Planning for the production of a report

Before any searching or analysis work can be done an understanding between the analysts, and their client needs to be made on why the work is being conducted, who will be influencing and ultimately, utilizing the report, and what it will be covering, including the time period that it will reflect. These items, since they are of such critical importance, should potentially be explicitly stated in a Terms of Reference, or another appropriate document that captures the understanding, between analyst and client, of what the expectations are.

Many organizations have internal resources available for conducting patent landscaping, but others require the hiring of third parties, who will perform the work on their behalf. The process of identifying third party analysts, bidding on projects, and agreeing on timing and deliverables needs to be part of the planning process, and agreed to before the work can be initiated. While a bidding process generally doesn't occur when internal resources are employed, there is still, typically, an agreement about timing, and deliverables, which is associated with in-house projects as well.

8.1.1 – Selecting a Topic

The selection of the topic for a PLR is arguably the most important aspect of the planning process since it will be used as a means to guide the remaining planning activities, and the execution of the project. As discussed in section 5.1, PLRs are normally requested in order to generate intelligence in support of an organizational decision. Since this is the case, the topic will typically be based on an industry, or technological focus area that is critical to the decision. The topic needs to be expansive enough to encompass all of the options associated with the decision, but narrow enough so as to provide focused information that will be actionable.

Another way of looking at the determination of the topic was covered in section 7.3 while discussing the Linear Law of Patent Analysis. In this case, understanding the needs of the business was required before relevant questions could be generated. Deciding on a topic for a PLR is an analogous process, since an understanding of the needs of the requester is required in order to understand the questions

they have. Answering the most pressing or critical questions, generally, provides the topic for the PLR.

An excellent example of the process, which led to the selection of the topic of alternate energy technology, can be found in the Introduction section of the PLR produced by WIPO[76] on this subject:

Industrial development has led to increased prosperity for many people around the world but has also led to a depletion of natural resources and environmental damage. The consumption of fossil fuels, on which industrial development has been largely based, has been recognized as a major cause of climate change. The impacts on the global ecosystem resulting from climate change are in turn expected to lead to substantial economic losses. As a result, it is clear that new means of fueling industrial development must be found in order to avoid compromising the gains in human welfare that have been achieved over the past decades. The continued dependence of most countries on fossil fuels – primarily oil and natural gas – from a small number of often politically unstable regions is also troubling from a political and security perspective. The potential for resource conflicts and other political and social problems will only increase as oil and natural gas resources become increasingly concentrated in these regions and global demand for fossil fuels rises.

People have turned increasingly to alternative energy sources as an answer to the environmental, political, and social problems linked to fossil fuel use. Alternative energy sources are broadly defined as energy sources that do not cause or limit net emissions of carbon dioxide and thus largely avoid the environmental impacts associated with the combustion of fossil fuels. Furthermore, they are generally defined as being renewable sources of energy not requiring the input of fossil fuels, which also speaks to their political and social advantages.

This particular topic embodies both corporate, and governmental policy matters, and provides intelligence that will be used to drive decisions associated with solving the environmental, political and social problems connected with the use of fossil fuels. The importance of this topic, and the questions it looks to address are also clearly stated, and relevant to the stakeholders of the report.

8.1.2 – Identifying Collaborators and Partners

There are three primary purposes for identifying collaborators and partners to assist with the creation of a PLR; these individuals, and groups provide subject matter expertise, their endorsement of the project adds credibility to the importance of the topic, and their inclusion provides confidence that the appropriate resources are being applied to it.

Collaborators and partners should be recognized authorities in the field covered by the topic of interest. In section 5.1.1.2 an example was provided of a cooperative effort between WIPO, the World Health Organization (WHO), and the World Trade Organization (WTO) on worldwide access to medicines. WHO and WTO are world-renowned organizations specializing in health and economic concerns. Both organizations bring a tremendous amount of knowledge, expertise and authority to the topic being explored. By partnering with the organizations WIPO was able to bring attention and credibility to a topic of great interest and importance.

---

[76]

http://www.wipo.int/export/sites/www/patentscope/en/technology_focus/pdf/landscape_alternative_energy.pdf

Generally speaking, a partner should bring subject-matter expertise to the topic being explored. Analysts should have a technical background and feel comfortable working in the technology associated with a topic area, but they normally require subject matter experts to assist them with gaining a deeper understanding of the nuances associated with a technological area. A collaborator can provide access to professionals that have spent many years understanding the details associated with a topic area. These individuals can help validate the findings and ensure that misconceptions are not propagated during the generation of a PLR.

An endorsement by a collaborator will generate interest in a PLR for an extended collection of parties outside of the original requestors. This is especially critical for topics that are intended to influence public policy decisions since an endorsement by a well thought of partner will provide credibility that is borrowed from the reputation of the partner. A PLR generated in collaboration with a respected partner will generally be received more openly from groups that were not associated with the initiation of the project.

Finally, recognized subject-matter expertise, in addition to assisting the analyst in researching and analyzing the project, also adds confidence that the appropriate people have worked together to produce a high-quality product. Working with recognized experts leads to an initial impression that the work conducted is superior, and held to the highest standards.

8.1.3 – Defining the Scope

There is an excess of 60 million patent documents, from more than 100 patent issuing authorities, around the world. On any given topic there are likely to be tens of, if not hundreds of thousands of, patent documents that could be relevant to a technical area. Trying to look at all of these documents would be like "trying to boil the ocean" and can lead to a project that lacks direction and focus. The process of defining the scope for a PLR provides a means for identifying, and stating, the focus and direction of the work to be done.

While patent landscapes provide an overview of a topic, it is also important that they stay focused on the business need. Providing the scope will ensure that the analyses performed will concentrate on the critical issues that need to be addressed in the PLR. If the scope is clearly stated, and agreed upon by all the stakeholders than it can also be used to ensure that the readers of the PLR will understand the constraints under which the analysis was conducted.

In any given technical area there are likely to be many different approaches to addressing a particular issue. Looking at all of them, in a single PLR, would potentially lead to superficial research that would not produce sufficient insight. With large, complex topics it is good practice to limit a PLR to a handful of the most important technical approaches to dealing with the stated issue.

An example of defining the scope of the technology to be analyzed in a PLR can be found in a report generated by the Franklin Pierce Law Center on Protein/Peptide Vaccines for HIV[77]:

Many strategies have been employed to search for a vaccine to combat the rampant spread of HIV worldwide. As research has progressed towards a better understanding of the virology, pathogenesis and immunological properties of HIV, vaccine designs that incorporate subunit proteins or epitope-

---

[77]

http://www.wipo.int/export/sites/www/patentscope/en/programs/patent_landscapes/documents/itti_patent_ls_hiv_protein_vaccines.pdf

based peptides have emerged as viable candidates for developing effective therapeutic and preventative treatments for HIV. Protein subunits and peptides in a vaccine elicit humeral immune responses by stimulating antibodies to neutralize the native virus. Though a high specificity related to HLA alleles decreases the universal effectiveness of a peptide vaccine approach, many protein subunit and peptide vaccine designs incorporate conjugates or adjuvants to increase their immunogenicity. The purpose of this patent landscape study was to search, identify and categorize patent documents that are relevant to the research, development and distribution of a subunit protein or peptide based HIV vaccine.

While the scope typically identifies what will be covered in a PLR, it is also sometimes necessary to explicitly state what will not be covered, or what is out of scope, in a particular project. Providing this information at the beginning of the PLR will set the expectations of the readers and prevent them from looking for information and analysis that was not included since it was tangential to the stated objectives.

In addition to looking at the technological aspects of scope there are a few analysis related items that should be clearly stated, and agreed upon before the work is started as well. These include the countries and time period that will be covered, if a patent family reduction will be conducted, and if non-patent literature (NPL) will be included in the analysis.

8.1.3.1 – Country Coverage

Country coverage refers to the patent documents issued by a particular country that are going to be analyzed during the preparation of a PLR. The key item to consider, when deciding on the country coverage, is which countries are going to be impacted by the conclusions made in the report, and which countries will have an impact on the issues associated with the project. As stated previously, with so many countries to choose from, it is important to focus on the ones that are directly effected in the agreed upon objectives for the work.

For example, in a PLR for a corporate customer, looking to launch a product based on a new technology, the country coverage should include the places around the world where the company is planning on selling their product, as well as the countries in which they are planning on manufacturing it. The countries where manufacturing will take place represent areas that will have an impact on the generation of the product, in question. Patent clearance in the countries where the product will be sold will have an impact on the business, and so the patent environment in those areas is critical as well.

The same idea can be applied to policy decisions where freedom to practice is required in the areas to be served by a technology, as well as in the regions where the technology will be manufactured by the organizations assisting in the effort.

On a practical note, it is generally a good idea to also include PCT documents in analyses conducted in association with a PLR. As covered in section 4.1.1.1.1, PCT documents are a special type of patent application that can be applied to many countries simultaneously. The country of interest maybe included as a designated state in a PCT application, implying that a future national stage filing could take place, providing patent rights for a technology in that jurisdiction.

8.1.3.2 – Time Period Coverage

In almost all jurisdictions, the natural term of a patent document is 20 years from the date of priority filing. As discussed in section 4.2.1.3, the priority date, sometimes called the "effective filing date", is the date used to establish the novelty and/or obviousness of a particular invention relative to other art.

It represents the earliest date from which an assignee claims priority for their application. After this 20-year term has elapsed, the patent is said to have expired, and the technology enters the public domain.

This is germane to a discussion on time period coverage since the objectives of many PLRs revolve around the application, or consequences, of in-force patents, that is patents, which are up to date on their maintenance fees, and have not expired naturally. Under these circumstances, analysts tend to use a twenty-year time period, from the present date, when generating a data collection to explore.

For technologies that evolve quickly, or that have developed only recently, a shorter, more recent time period may be used. Shorter time periods are also used as a method for segmenting extremely large data collections, where analyzing the complete set would be too onerous to accomplish.

Analysts, who want to provide an historical perspective, may decide to not impose a time frame on their analysis at all, and will generate one or more analyses, in order to highlight how an area has evolved over time.

8.1.3.3 – Patent Family Reduction Method

The concept of a patent family was covered in section 4.3.5 where the definition of a basic patent family was listed as, comprising all documents having exactly the same priority or combination of priorities. Priorities in this case refer to the priority-filing document.

Due to the nature of the patent system worldwide, patents must be applied for in individual countries. This creates a situation where a single idea might have many individual patents associated with it depending on the number of countries the applicant sought protection in. In some cases, an assignee will seek protection in dozens of countries, and thus there will be many equivalent documents associated with their invention.

The idea behind counting patent documents when conducting some patent analysis projects is to ensure that a fair representation of the number of inventions, or the amount of work performed, in a particular subject, or by a specific assignee, is adequately accounted for. For the most part, analysts want to ensure that if the same invention is covered by patent documents in different counties they are not counted individually. While country coverage can be an important indicator of the value an assignee places on an invention, it can also introduce a bias since organizations with more resources might decide to seek coverage in more countries, while a lone inventor or small company might only seek protection in a single country. When comparing different methods for solving a technical problem, if you have one approach with companies that seek protection in a large number of countries, it might look like it has more activity than another approach where the companies involved don't seek patents in as many countries. The process of eliminating redundant country filings is referred to as a patent family reduction.

Also covered in section 4.3.5 were alternate methods of generating patent families beyond the basic family described at the beginning of this section. One of the most popular methods of generating a patent family is referred to as an extended family, most often associated with the INPADOC database, and thus referred to frequently as an INPADOC family. The definition of an extended family[78] is a broader definition of a patent family that takes domestic application numbers as additional connecting elements, and includes documents having the same scope, but lacking a common priority.

---

[78] http://www.epo.org/searching/essentials/patent-families.html

Organizing patent collections by some form of family is an essential activity, which will have a significant impact on how statistics are generated for a PLR. Determining which method will be used and consistently applying it across an entire project will ensure that accurate comparisons can be made between different entities being studied. Generally speaking, using basic families will create larger numbers of narrowly defined collections to analyze while extended families will produce smaller, broader representations. Analysts must determine which method best suits the objectives of the PLR being generated.

As an example, the use of an extended family may severely underrepresent the amount of investment made by an organization, since many related, but distinct patent filings maybe correlated into a single family. If one of the objectives of the PLR is to identify which organizations have invested the most resources into a technological approach than a basic family, or alternative approach, would be better suited than the use of an extended family. A discussion on the potential dangers of using extended families for PLR studies can be found at:

http://www.patinformatics.com/blog/why-i-dont-use-extended-families-when-counting-patents/

One alternative to the use of extended families is the One Document per Invention (ODPI) method. This involves the selection of a primary country where all potential inventions are identified based on eliminating redundant applications that have progressed into granted patents. Sorting the documents by shared application number can usually accomplish this.

Once the primary country has been processed for inventions the original corpus is reduced using an extended family making certain that the most recent document from the primary country is kept as the representative document. Once this is completed the extended family set is combined with the primary country, all potential inventions, set to produce a simple ODPI corpus. Additional details on this method can be found at:

http://www.patinformatics.com/blog/counting-documents-when-conducting-a-patent-analysis-project/

There are additional family member reduction methods available including the use of simple families, which is used by the European Patent Office (EPO), and is becoming increasingly popular. The EPO website provides the following definition of a simple patent family[79]:

All documents having exactly the same priority or combination of priorities belong to one patent family.

Generally speaking, it is absolutely essential to perform some type of family reduction, when putting together a PLR, but the decision on which method to use should be made by the analyst based on the message they are trying to deliver about the inventive output, or level of investment made in association with a technology. A more comprehensive look at the types of patent family reduction methods and the impacts of using them can be found in the blog post provided below:

http://www.patinformatics.com/blog/why-i-dont-use-simple-families-when-counting-patents-either/

While the title of the post seems to downplay the use of simple families it represents a thorough examination of the alternative methods, and explains the requirements associated with a methods to enable its effective use.

---

[79] http://www.epo.org/searching/essentials/patent-families/definitions.html

8.1.3.4 – Inclusion of Non-Patent Literature (NPL)

The word patent is included in the name Patent Landscape so the expectation is that patents will be the sole source of data being explored in these reports. Sometimes, however, the inclusion of NPL is required in order to meet the objectives of the report. For instance, due to the 18-month delay in the publication of patent documents, in most jurisdictions, truly cutting-edge developments can only be discovered when looking at NPL.

University professors tend to favor NPL, as opposed to patenting, although this has changed somewhat with recent legislation providing additional incentives for the commercialization of patent technology from universities. Having said this, in technologies where the primary innovators come from universities, it is still often necessary to include NPL to ensure that a complete picture is provided.

While producing NPL can be potentially time consuming, and expensive, it is generally less so than generating patent documents. This is important to keep in mind when looking at the relative number of documents being produced of each type. From a practical perspective, if it is necessary to include NPL analytics in a PLR, then they should be performed separately from one another. In a manner of speaking, they are like apples and oranges, similar, but different, so most analysts do not try to combine them to produce a single analysis.

Again, include NPL analyses when they are required to achieve the objectives of a PLR, and they provided a richer view of a topic area, but keep them separate from the analytical work being done on the patent documents.

8.1.4 – Preparing a Terms of Reference (TOR)

A Terms of Reference (ToR) document is not a standard practice in the private sector before preparing a Patent Landscape Report. That is comprehensible; Terms of Reference may seem to be too time-consuming or unnecessary for a task that may appear to be clear and straight-forward. Nevertheless, as the experience from the WIPO Patent Landscape Reports project has shown, drafting Terms of Reference before initiating a Patent Landscaping exercise is highly recommendable. And that,for a number of reasons.

First of all, the drafting of a Terms of Reference allows the commissioning organization and recipient of the Report to have a clearer idea about what exactly he is interested in and phrasing it in a concise manner. Often, the needs assessment for the scope of the report is difficult to be made and the specific deliverables not always easy to predefine. The ToR may be time-consuming, but they help in the clear definition of the specific conditions and requirements of the work, namely the scope of the report, the expected search and analysis to be carried out, the geographical and historical coverage, the timelines, the deliverables and the payment procedure. The clear description of all these factors assists in the creation of a common understanding of the expectations and the deliverables both for the recipient and the provider/contractor of the report, is increasing the efficiency of the work and is ensuring a smooth delivery of the report, minimizing the risk of unsatisfactory deliverables and unpleasant surprises.

The drafting of the Terms of Reference for a PLR should ideally reflect and follow discussions between recipient and provider of the report. That allows the assessment of the feasibility of certain requirements related to delivery timelines, pricing and the content of the report before the initiation of the projects, as often the lack of technical knowledge on patent information from the side of the commissioning party paired with possible lack of understanding about the contractors limitations and

options in his deliverables cause misunderstandings and false expectations. As visualization plays an important role for decision makers, the profile of each report recipient and the questions each report is addressing are different, part of the discussion and the ToR should be which types of analysis and visualization are preferred and appropriate for the specific recipient's needs. Moreover, both sides should know for a better planning the duration of the preparation of the report, and also the human and time resources that need to be dedicated for it, along with the required technical infrastructure and access to specific databases.

Depending on the questions addressed, and the level of understanding of the recipient of the report on IP and patent information, even more technical issues may be included in the Terms of Reference, such as the methodology to be followed, or the specific search or analysis tools to be used.

An example of Terms of Reference can be found on Annex of the present Guidelines and are taken from a WIPO Patent Landscape Report on Vaccines for Selected Neglected Diseases. In general, the Terms of Reference may include the following points:

- Introduction/Background information: description of the technical problem and contextualization
- Scope and Objective of the Report: the description of this part will facilitate the provider of the report in the approach it is going to take and the perspective it should interpret the results and make its recommendations
- Content of the report:

o An executive summary may be required for the senior management/patent portfolio management
o A description of the technical area and/or problem may be necessary should the report be addressed to non-technical experts
o The types of patent analysis that are required should be described to ensure a minimum content an analysis
o Additional, non-patent analysis which is needed should be requested and described accordingly
o Conclusions and recommendations based on the analysis and the background for the preparation of the report, as described in the "scope and objective of the report" section
- Delivery description:
o Various expected deliverables and delivery stages
o Timeline and deadline for each of the deliverables
- Payment modalities:
o Timelines for payment
o Penalties in case of delays

8.1.5 – Deciding on How the Work will be Conducted

Patent Landscaping is an activity which can either be commissioned to an external service provider or which can be conducted in-house. The decision depends on a number of conditions and may vary case by case, based on the specific needs of a project.

8.1.5.1 - Using external providers

Usually smaller entities without specialized patent information resources and/or without a need for patent landscaping exercises on regular basis outsource the preparation of a Patent Landscape Report to an external service provider/contractor. An external provider may also be deemed necessary if the in-house expertise does not cover the specific topic that should be researched and

analyzed or complementary analysis is needed with the use of visualization or analysis tools that are not available in-house.

The selection of the contractor is sometimes simply based on recommendations, a web search with comparison of prices and description of services, or on previous cooperation with certain service providers. The same Patent Landscape Report can look completely different and follow a different approach, depending on the service provider and in some cases it may even make sense to commission the same report to two different providers in order to cover various approaches, but also as an additional quality check, foremost when the decision that the Report will support is important and the budget allows so. When also time allows, a way to ensure a wider choice among service providers for selection, based on comparable specifications is to tender the provision of the report and make a call for tender or issue a request for proposals (RFP).

Request for proposals

In the framework of WIPO Patent Landscape Report Project the preparation of the reports has been commissioned to external providers due to the broad spectrum of topics of the reports, requiring expertise in patent search in the relevant fields. In order to ensure a broad range of offers from various service providers using different methodological approaches and tools, a Request for Proposals (RFP) is issued every time that the preparation of a Report has been decided and the Terms of Reference document has been prepared. The RFP includes the finalized Terms of Reference (ToR), along with other general terms and conditions and forms the basis for the submission of technical and financial offers for the provision of a Patent Landscape Report.

Within a given deadline, the various suppliers provide a technical and financial offer, based on the ToR. That allows for comparability of the offers based on specific requirements and deliverables, and the inclusion of certain added-value deliverables.

Selecting a candidate

The selection of a candidate for the preparation of a Patent Landscape Report is not always a straight-forward decision. In the framework of a public procurement procedure, the most common approaches are the lowest price or the most economically advantageous offer. Since the provision of the report is a service where quality is a high priority, the price as such may be important, but is or should not be the decisive criterion. Through the evaluation and selection process it should be ensured that the successful candidate has a correct understanding of the commissioning party's needs and requested deliverables, fulfills the technical requirements, and submits a convincing offer that lies within the estimated budget.

During the WIPO Patent Landscape Reports the need for the definition of specific evaluation criteria has been identified. The approach followed aimed at a balanced evaluation of the financial and technical offer:

Patent Landscape Report on …...

Technical Evaluation Criteria

#	Main Criteria			Candidate A	Candidate B	Candidate C	Candidate D
1	The Bidder's general understanding of the project			3	4	4	4
2	Specific experience relevant to project			4	3	3	4
3	Qualification and expertise of the project team			3	3	3	5
4	Understanding of the project deliverables and proposed approach for the preparation of the PLR			4	2	2	5
5	Responsiveness and compliance with legal requirements			3	3	3	3

Technical Evaluation Criteria - Weighted scores

#	Main Criteria	Weight	Max Score	Candidate A	Candidate B	Candidate C	Candidate D
1	The Bidder's general understanding of the project	3	15	9	12	12	12
2	Specific experience relevant to project	3	15	12	9	9	12
3	Qualification and expertise of the project team	6	30	18	18	18	30
4	Understanding of the project deliverables and proposed approach for the preparation of the PLR	6	30	24	12	12	30
5	Responsiveness and compliance with legal requirements	2	10	6	6	6	6
		Total	100	69	57	57	90

Financial Component

		Candidate A	Candidate B	Candidate C	Candidate D (
	Total cost in CHF	50000	33000	48895	64000
	Score	22.9	34.7	23.4	17.9

Overall Weighted Scores and Ranking

#	Overall Scores	Weight	Candidate A	Candidate B	Candidate C	Candidate D
1	Technical Evaluation - Weighted Score	65%	44.9	37.1	37.1	58.5
2	Commercial Evaluation - Weighted Score	35%	8.0	12.1	8.2	6.3
		Overall Score	52.9	49.2	45.2	64.8
		Overall Rank	6	0	0	2

NOTE: Scoring shall be as follows: 5-Excellent; 4-Good; 3-Average; 2-Below average; 1-Poor; 0-not acceptab

Disqualified in Technical Evaluation (below minimum acceptable threshold X/5
Qualified in Technical Evaluation

Financial Evaluation
Lowest offer score 100 points and the rest are calculated as follows Lowest Offer/The Offer in Evaluation

8.1.5.2 - Conducting the work in-house

A number of bigger corporation have an in-house team with patent information professionals which can carry various names, such as Business or Competitive Intelligence, Patent Information, Patent or IP Analytics. This team usually works in cooperation with the IP or Patent Law area and provide shorter or longer landscape reports to the senior management to facilitate their decision making.

8.2 – Performing the Search

While discussing the Linear Law of Patent Analysis, in section 7.3, it was stated that the question drives the data. The data, in the case of PLRs, is patent information, and it is collected by means of various types of searches conducted in a variety of different databases. Using the Linear Law of Patent Analysis as a backdrop for thinking about searching, there are two elements that need to be considered, the first involves generating a query that will provide the right data for addressing the business question behind the PLR, and the second involves producing the proper output and formats for ensuring that the data can be analyzed in the next step of this process.

All of the planning behind a PLR is directed towards providing insight on a particular organizational issue that the client is looking to explore. As already described, the topic and scope are generated based on this principle, and once they are decided upon they frame the questions that will be

addressed with the search and subsequent analysis. If the topic, for instance, is access to essential medicines, and the scope includes developing countries, then the type of searching should included methods that will find patents on the agreed upon medicines, and the databases searched should cover the countries determined to be part of the definition of developing nations. Based on our earlier discussion on scope, the databases searched should also include countries where the medicines will likely be manufactured. This may seem obvious or intuitive, but it reinforces the reason why the Linear Law of Patent Analysis is referred to as a linear law, the reason being that each subsequent step builds upon the previous one. Initial agreement on the topic, and the scope makes the development of the search parameters straightforward, and when it comes time to explain it to the client, easier for them to understand in the context of fulfilling the objectives of the PLR.

It is also important to ensure that the items to be analyzed, in the subsequent steps, are included in the data to be exported from the database of interest, once the search has been conducted. Different databases include different exporting options so an analyst will need to ensure that the requirements for analysis are met before time, and effort is put into the searching. At a minimum, as covered in section 8.4, the following fields should be available for export: Patent Assignees, Inventors, Application Date, Priority Date, International Patent Classifications, Forward Citations, Backward Citations and Patent Status.

While it is not the intention of these guidelines to provide a comprehensive discussion on patent searching, some thoughts on an approach for doing so will be provided. Conducting an Internet search for the phrase "patent searching tutorial" can discover additional perspectives on patent searching. One of the better examples comes from Stanford University:

http://www.stanford.edu/group/biodesign/patentsearch/howto.html

Another patent searching approach is referred to as "reverse engineering". Using this method an analyst will conduct a quick search for the main topic in the title field of the database of interest. This will generate a few references, which are likely to be highly on target. These references are then "reverse-engineered" to identify relevant patent classification codes, database specific indexing, and other terms used by the inventor to describe the concept involved. As these searching artifacts are collected, another search is performed, generating a large collection of documents that are also examined for new synonyms, and additional codes. The process is repeated until no additional search items are identified, and the analyst feels they have conducted a thorough search.

Referring back to section 7.4, on precision and recall, it was discussed how these principles are normally in conflict with one another. Precision is sacrificed normally in pursuit of higher recall. PLRs are overviews, and statistical measures are conducted in association with them, so recall is generally more critical than precision. This applies to work done in conjunction with a PLR where high recall searching methods should be employed. From a practical perspective, if the level of recall can be established at higher than 90%, while the precision kept above 70%, then the likelihood of finding statistically relevant items appearing in the subsequent analysis steps that are significantly off-topic is reasonably small. Major statistical finding should always be validated, to ensure that they are not coming from imprecise searching, but the majority of the time records that are off-topic will not be associated with major trends, and thus not seen in the analysis.

8.2.1 – Determining the Type of Searching to be Done

The following explanation of an approach to searching pharmaceutically relevant compounds comes from the WIPO PLR on Atazanavir[80]:

Of course, to find patents about a particular drug, the best way is to search for it by name. This is easier said than done with chemical compounds, and in this section of the report, the "pre-clinical" and "clinical" names will be discussed, along with the strategies for assembling a collection that covers the development of the compound from beginning to end. Having confidence in the quality of the collection is crucial to the subsequent analysis.

When the initial patent application was made (1995), the compound was not known as Atazanavir, which is its current non-proprietary generic name. Generic names are only applied to compounds after they have been approved for clinical investigation. It is usual also for the claims of the initial composition patent to cover not just one composition, but a range of compositions with different substitutions in various sites on the basic backbone structure. The backbone with variable substituents (shown as R groups in the variable positions) is called a Markush structure.

The report goes on to give examples of the various ways to identify all of the chemical names associated with Atazanavir, as well as the various additional methods, such as Markush and substructure searching, for finding documents on chemical compounds. It provides a nice example of the need to structure an approach to the type of searching needed, depending on the topic of the PLR.

Searching in some technological fields, especially the life sciences, requires subject-matter expertise from scientists who understand the field, and perhaps more importantly, from professional searchers, who know how to find patent information in these areas. If the analyst is not a trained patent searcher than assistance should probably be pursued in this practice.

8.2.2 – Determining which Databases to Use

There are many databases containing patent information available. A list of some of the more widely used ones can be found in section 9.2 of these guidelines. Some databases contain information that can subsequently be used to identify data in other databases that are configured to allow for the exporting of information that will be used in the analysis steps. The example below, from the Initiative for Medicines Access & Knowledge (I-MAK) PLR on HIV Drug Patents in China[81] demonstrates the use of this approach:

The ARVs covered in this patent landscape are based on the generic names (International Nonproprietary Names – INN) of marketed ARV treatments listed by the US Food and Drug Administration (USFDA). In addition, patent information for some ARVs showing potential in Phase III of clinical trials are also provided.

The first step was to search for patents covering INNs as listed in the USFDA Electronic Orange Book.4 As the Orange Book only provides US patent data, this information was used to conduct

---

[80] http://www.wipo.int/patentscope/en/programs/patent_landscapes/reports/atazanavir.html
[81]
http://www.wipo.int/export/sites/www/patentscope/en/programs/patent_landscapes/documents/i_mak_hiv_drug_patents_in_china_tahir_amin_may_2010.pdf

patent family searches that would capture patents claiming the same priority and which have been filed or have entered the national phase in China under the PCT. It should be noted that a number of PCT applications designating China were located in the search, but have not yet entered the national phase there. These applications are not included in this patent landscape.

The European Patent Office (esp@cenet) and Thomson Innovation databases were used for conducting patent family searches. Once the Chinese patent numbers were identified, the status of the patents was checked using Thomson Innovation, the Chinese Patent Office (SIPO) and Chinese Patent Information Centre (CPIC) databases.

However, relying solely on patents listed on the Orange Book will only yield some of the patents relating to ARVs. First the Orange Book only lists granted US patents. Therefore, pending US patents that may relate to an HIV drug would not be listed. For example, the relevant patents covering the heat- stable tablet formulation for ritonavir, ritonavir and lopinavir (Kaltra/Aluvia) are currently not patented in the US and, therefore, do not feature in the Orange Book. Second, the Orange Book only covers HIV drugs marketed in the US. As a result, many patents covering fixed dose combinations of ARVs will not be listed. An example of this is the combination lamivudine and zidovudine (Combivir). Also, patents covering drugs that are going through clinical trials do not feature in the Orange Book. Finally, patents on intermediate compounds and processes or methods of manufacturing a drug are not permitted for listing on the Orange Book.

Taking into account the limited patent information available in the Orange book, further searches were conducted using keywords including chemical names for INNs and brand names for ARVs. In addition, citation searches of to earlier patents identified in the Orange Book were also conducted. Only patents held by originator companies were included in the landscape. Again, the patent status was checked using Thomson Innovation, SIPO and CPIC databases.

Once the relevant Chinese patents were identified, copies of the published or granted patent specifications and their claims were downloaded from Thomson Innovation and Esp@cenet. Where possible, the key claims of the Chinese patent documents were translated using machine translation to obtain a basic understanding of the subject matter covered. Additionally, claims of the Chinese patents were compared with their equivalent PCT, European and US patents.

The USFDA Orange Book was used to identify some of the relevant patents, while Thomson Innovation, and Espacenet were used to identify Chinese family members. The translated claims were also extracted from these databases since they were required for subsequent analyses within the report.

Once again, if the analyst is not comfortable with the art of patent searching, then appropriate, professional assistance should be sought.

8.2.3 – Sharing the Strategy

When sharing the search strategy used for collecting data for a PLR, there is a fine line between providing enough details to inspire confidence in the search while not overwhelming the readers of the report with a large amount of searching syntax, and details. A practical approach is to provide a written overview of the methods used, which will be included at the beginning of the PLR, and providing details in an appendix. The Atazanavir PLR[4], discussed in section 8.2.1, also provides a good example of this practice:

A number of approaches are useful to prepare a collection of pre-clinical patents covering the same or similar compounds, even though the Atazanavir name was not yet available. These approaches are applicable even when access to enhanced content such as the Derwent World Patent Index is not available, and all have been used in this report.
List of search approaches:

- *Identifying founder compositions by using the SPC filing registration found in the Legal B Status field. In later documents this field is also likely to contain the generic or brand B name of the drug.*
- *Searching for key terms selected from the unique parts of the chemical name (± B classification codes).*
- *Searching for CAS codes in the description field.*
- *Searching for founding inventor names together with their developer company names to B select potentially related documents from the company's holdings.*
- *Searching target key terms together with developer company names.*
- *Searching for documents that cite patents owned by the companies involved in B development.*

The results from such searches do overlap with one another, and may also overlap with the clinical collection. After cross-deduplication, they will often need further review to determine if they are sufficiently on-topic. It is expected that this list will contain many fewer documents than are present in the clinical collection, because the volume of patents is typically lower before the discovery phase moves into clinical development. However, these documents should help fill in the "gap" period mentioned earlier, the period between the first composition patent, and the beginnings of clinical testing. The details for these search strategies are provided in Appendix A.

The search methodology section of the PLR contains a written description of the various methods used. The breadth and depth of the approach can be clearly seen from the outlined steps, in a fashion that can be understood by the average reader of the document, but without overwhelming them with jargon and scripting. The details, important to a patent information professional who might be tasked with updating or replicating the work, can be found in the appendix.

8.3 – Preparing the Data to be Analyzed

With data from the search in hand, it is time to put it into a form that is appropriate for analysis. Under most circumstances, especially with large data collections, more time will be spent preparing the data than required to conduct the analysis. Almost all analysis and visualization tasks require one form of data preparation, or another, to be performed, but it is especially important for analytics involving statistical measures. The methods used and the results of the preprocessing will significantly impact the values obtained in the statistical analyses. This requirement is so prevalent that it has an acronym associated with it, GIGO, Garbage In Garbage Out.

The following subsections examine, in greater detail, some of the more common preprocessing steps used in conjunction with patent data sets for PLRs. These operations take place before any analyses are conducted, and don't generally require an iterative series of vetting steps as will be seen when performing an analysis.

8.3.1 – Field Cleanup and Grouping

On a number of occasions, particularly in section 6.1, patent data has been referred to as "messy". In particular, there are frequent misspellings that need to be addressed in the Patent Assignee/Applicant and Inventor fields, before they can be used in statistical analyses. General methods for manually and

semi-automatically cleaning up misspellings in these fields were also briefly covered in section 6.1. A semi-automated method for correcting misspellings that is gaining in popularity is found in the use of Open Refine for this task.

Open Refine was once called Google Refine and was developed by Google as a power tool for dealing with messy data[82]. There are six different algorithms provided for cleaning up data using the methods found in Refine. In particular, the Metaphone3 algorithm does a nice job of automatically cleaning many of the common spelling errors found in Patent Assignee and Inventor fields. Once the algorithm is launched it will provide the results along with some statistics about the entries in the cells. The results of the cleanup can be reviewed and if the analyst agrees with the way the algorithm performed they can check the Merge box. A New Cell Value can also be added so the organization name looks the way the user wants it to as opposed to what is provided by the tool.

A full description, and example, for cleaning fields using this method can be found at:

http://www.patinformatics.com/blog/patent-assignee-cleanup-using-google-refine-open-refine-text-facets-and-clustering/

Besides dealing with misspellings, it is also necessary to group items together under a single name for accurate statistics. Accounting for mergers and acquisitions is an example of a case where grouping is done to reflect all of the documents associated with a single organization. In a similar fashion, when they are known, it is also important to standardize inventor names based on a change of name after marital status, or when middle names and initials, or multiple generations of individuals with the same name, need to be account for. These items are groupings, as opposed to cleanups since they require expert knowledge of external events that may have taken place, which caused the change in status. Discovering these events often requires additional research into the details associated with the documents in question, but once know the actual mechanism for conducting the grouping is performed using the same tools that are associated with cleanup.

Looking at Open Refine again, groupings can be managed using a feature called text facets. Within the text facet window there is the capacity to manually edit any of the values, so grouping can be performed by scanning the list while in alphabetic order. If a change is required, pass the cursor over an entry, and an edit link will appear that allows the user to change the value to something different. Remember to stay consistent if repeating an organization or inventor name that was used in a different part of the list.

8.3.2 – Family or Invention Reduction

Patent family, or invention reductions were covered extensively in section 8.1.3.3. As stated in that section, many analysts use an extended family, of which the most popular is the INPADOC family, to eliminate the same invention covered by applications in multiple countries. As dicussed, the use of extended families can dramatically underrepresent the amount of investment an organization has made in a technology, especially in the United States and Japan.

It is critical, however, that some family or invention reduction takes place, since over representation, based on a multi-country filing bias, can occur if a reduction of the set to be analyzed is not

---

[82] https://github.com/OpenRefine/OpenRefine

performed. It has been suggested that a One Document per Invention (ODPI) approach[83], where all inventions from a primary country, the United States for instance, are retained, provides a middle ground for eliminating multi-country biases while ensuring that investment is properly represented. The use of simple families or vendor specific families is also encouraged.

Regardless of the method used, the choice of reduction method should be clearly stated in the Methodology or Issues and Limitations section of the PLR, and it should be consistently applied to all of the analytics associated with the project.

8.3.3 – Is Manual Review Required for Precision

In section 8.2 it was stated that greater than 90% recall, and 70% precision, were required before accurate analyses could be generated on a data collection. Even with advanced searching tools it is sometimes impossible to produce a query that will provide at least 70% precision. There are also occasions where 70% precision is not high enough and levels closer to 90% are required for trustworthy results. This is especially true with smaller data collections, were errors will have a larger impact than they would with sets numbering in the tens of thousands.

In either event, if the analyst reviews the data generated from a search and finds that an uncomfortable number of irrelevant documents, or "false drops" have made their way into the collection, a manual review of the documents will be required. This step should be conducted after cleanup, grouping or family reduction has taken place since these processes will either lower the number of documents that need to be reviewed or will organize them in a fashion that will make review more efficient.

Manual review, under these circumstances, normally involves the scanning of titles, and perhaps abstracts, to determine if a document should remain in the corpus. The source titles and abstracts, of patent documents from most issuing authorities are notoriously vague and ambiguous, and many times are not particularly helpful. Enhanced titles and abstracts, as provided by several value-added database producers, are generally very helpful when performing a scanning process. Using enhanced titles and abstracts it is often possible to scan several hundred documents within an hour or two. Document collections with less than 500 members should be reviewed quickly for relevance to achieve maximum results.

8.3.4 – Determining Which Year Field will be Used

When conducting analytics, they are typically done on a year-to-year basis. Dates are provided for a number of milestones in the on-going lifecycle of a patent application, but working with specific dates is normally too granular, so years are used instead. There are three primary year choices associated with patent data, the application or filing, the grant or publication and the priority filing. These were discussed in some detail in section 4.2.1.3.

Depending on the objectives and questions being asked, in association with a PLR, a choice should be made about the year being used. As with the family reduction method employed, once a decision is made on which year will be used, it should be consistently applied across the entire PLR. Deviation from this principle should not occur unless there is an important reason for representing some analyses differently than the others. Under those circumstances, it should be made clear to the client that a different year type is being used.

---

[83] http://www.patinformatics.com/blog/counting-documents-when-conducting-a-patent-analysis-project/

Application year is used when an analyst wants to provide a closer approximation to when the research associated with an invention was performed, or when additional investments on behalf of an organization were being made. This is a popular year format for use in analytics, but there are issues associated with its use, since the analyses performed with this field generates a dip in the most recent years. This decline has to be explained to clients since it can be mistaken for a decrease in recent interest by a company, or in an area, but is really based on the 18-month delay in the publication of patent applications. Many analysts either cut out the last two years of their analysis, which is not recommended, or they place a red line on their visualization, 18 months from the end of it, and add a note to the bottom suggesting that new patents applied for in the last 18-months will not have published yet.

Publication year, on the other hand, does not generate a dip in recent years, since this event occurs when the patent grants, or when the application publishes in real time. The advantage, in this case, is that sudden downward trends don't have to be explained to the clients. The disadvantage to the use of publication year comes from issues surrounding the length of time required for patents to grant in some art units, between patent offices, and other factors. If only applications are being studied than this is not an issue, since almost all of them publish 18-months after filing. With granted patents though, the path to investment can be difficult to gauge when some of the patents took three years to grant while others took five or more. The difficulty in interpreting publication year analytics are compounded when data collections contain both granted patents and pending applications.

Analysts frequently also use the priority year when performing analyses. This practice should only be used if an expanded family reduction was used on the collection before an analysis was performed. As was the case with underrepresentation with extended families, the use of priority year can create analytics where it appears that all of the work was done in a single year, since a large, extended family might all claim priority to a single, early year. Normally, large families develop over time, and if a reduction method such as ODPI is used than the application, or publication year, is a better choice since they provide details on the timing associated with the growth of the individual inventions.

Ultimately, the choice of which year to use will depend on whether an extended family reduction is used, or if the analyst is more interested in highlighting time of investment, or accurately portraying recent trends in the publication of patent documents of interest.

8.3.5 – Generating Technology Categories

Most PLRs provide a broad overview of a technological topic area. There are often many approaches in which the issues associated with any given research area can be addressed. One of the key objectives of many PLRs is to identify the sub-categories, or approaches pursued within a topic area and provide statistics on relative interest, and timing for these. Before analyses are conducted the appropriate technology categories should be identified, and the data collection tagged, or grouped into them. The general concept of grouping was discussed in section 6.1, along with data cleanup.

The WIPO PLR on Solar Cooking[84] provides an example of how technology categories are used:

The obtained patent family members were broadly classified into two major segments:

84

http://www.wipo.int/patentscope/en/programs/patent_landscapes/reports/solar_cooking.html

- *Segment I – Complete solar cooking systems/solutions*
- *Segment II – Components only/also for solar cooking applications*

"Segment I" mainly comprises patent family members that were grouped primarily based on heating techniques like reflected concentration, trapped heating, indirect heating (using steam/vapor cooking or using heat transfer medium) and direct solar absorption. Furthermore, they were sub grouped under different components like concentrator/reflector, solar tracking device, heat absorber, heat storage, heat trap, insulation and other accessories.

"Segment II" comprises patent family members that were grouped primarily based on different types of components only/also used for solar cooking applications, whereas the components include concentrator/reflector, collector, solar tracking device, heat absorber and other accessories.

Technology categories are sometimes identified using the patent data itself, for instance, with classification codes, but ideally they should be generated based on input from a subject-matter expert based on an industry standard view on how approaches are categorized. Using a market or industry-based approach to creating categories will make it easier for the clients of the PLR to identify with the technology, and apply it to the environment that they are already comfortable with. Once consensus has been reached on the appropriate categories, the analyst has a variety of methods for grouping the documents. At this point, the analyst should also decide whether a single document can only appear in a single category or, if it deals with several potential approaches, be a part of multiple categories. If documents can only reside in a single category than a judgment will have to be made on which category best represents the main focus of the document. Under these circumstances, the analyst should probably refer to what is covered in the independent claims to make this determination. Five methods for grouping documents into categories will be discussed in this section. Regardless of which approach the analyst takes, their choice on document inclusion in multiple categories, and the method for grouping the documents, should be clearly stated in the Methodology section of the report.

As covered in section 4.2.1.4, patent documents are assigned classification codes when patenting authorities process them. Classification codes represent a hierarchical means for sorting documents into technology subcategories, and since an examiner vets them they would appear to be a good choice for identifying, and creating technology categories. Under ideal circumstances, the classification codes would conform to the market or industry thoughts on technological categories, but, unfortunately, this is rarely the case. Examiners, to help segment work in a patent office, create classification codes, and it is a happy accident when they align with the business needs of a PLR client. Classification codes also present difficulties since different patent offices sometimes use different systems, and even when the same system is used, can apply the codes at different levels of granularity. Codes are often the first place an analyst will look for technology categories, but frequently other methods will need to be used to meet the expectations of a client.

Analysts can also populate categories by creating individual search strategies for each of them. This method allows for large collections of documents to be grouped, but works most effectively when there are clear distinction between the categories and a small number of documents that would fall into multiple categories. Once again, ideally, text from the claims would be used for conducting these searches.

Many analysts will generate categories on the fly while they are manually reviewing documents for precision, as discussed earlier, or in conjunction with an effort to group them manually. An electronic spreadsheet, pre-loaded with titles, abstracts, and the independent claims, is often used to collect this

information. The analyst will add a new column onto the end of the spreadsheet, and as documents are reviewed an entry will be made in the new column for technology category, and potentially, sub-category. When the analyst comes across a new approach they can create a name for it, and continue to re-use the category as they come across additional documents of this type. The analyst will typically have spent time researching a topic, or speaking with a subject matter expert so the categories they create will closely reflect these learnings. This approach is precise, but time consuming, especially for large data collections, and the analyst also needs to attempt to be consistent with their assignments, especially as they learn new things while conducting the review. Sometimes it is necessary to make a first pass through a set, and then revisit the collection a second time, to fine-tune and reclassify some of the documents.

Finally, semi, or fully automated methods, based on machine learning or semantic approaches can be applied to populate technology categories. Supervised machine learning approaches to classification were covered in section 6.4, and semantic analysis was covered in section 6.9. As discussed earlier in this section, the choice of text used to perform the grouping into categories is critical to the success of the process. This choice is also impacted by which family reduction method has been used on the set. Using an extended family will reduce the number of documents to one per family and the analyst will be at the mercy of which document is selected, usually the most recent, and the classification will be based on what is covered in that particular document. If several aspects of a broader idea are covered in different filings, but claim the same priority than all of this detail is lost. Alternatively, many family members have identical specifications and it is only the claims that change from document to document. If a one document per invention approach is used then the claims should be analyzed to identify the differences between the family members for the purposes of populating categories.

8.3.6 – Reconciling Forward Citations

The concept of citations was introduced in section 4.2.1.5, and it is important to recognize the impact of redundant applications, and patent families on citation counts. Citations are based on the referencing of discrete documents, so a recent granted patent may not have any forward citations associated with it, but the corresponding, redundant, pre-grant application may well have several. An analyst can also look out over the entire extended patent family, especially if this method was used to reduce the collection, and find that there are forward citations. These citations needed to be aggregated in some fashion so that the document being discussed in a PLR is reconciled against all of the family members it represents.

At the very least, forward citations associated with a redundant pre-grant application should be aggregated with its subsequent granted patent. While these are discrete documents and yes, there are often times differences between them, they should be equivalent. They are, after all, the same application. If possible, it is even better to count all the non-redundant citations between all of the documents in the same basic (as opposed to extended) family. If there is a WO and a series of EP documents along with two US documents for instance which all share the same priority application number, and essentially all have the same set of claims (country specific modifications not withstanding) then all of these citations should be aggregated together.

A series of blog posts on counting forward citations was produced looking more closely at this issue and its impact depending on the patent's originating country. The URL for these posts are below:

http://www.patinformatics.com/blog/issues-with-counting-citations-how-many-forward-citations-does-us8341981-have/

http://www.patinformatics.com/blog/us-pre-grant-applications-have-significant-impact-on-citations-associated-with-granted-equivalents-issues-with-counting-citations-part-2/

http://www.patinformatics.com/blog/citation-trends-with-european-patent-documents-are-dramatically-different-than-in-the-us-issues-with-counting-citations-part-3/

While looking at forward citations, it is also important to distinguish citations coming from the assignee themselves versus those coming from other organizations. There are varying schools of thought on what is represented by self-citations compared to citations from others, but an analyst should segregate the two, and be prepared to address this metric if asked.

8.4 – Statistical Analyses to Include

Having gotten to this stage, it is now time to perform the analyses, which will provide the insight that will be shared in the PLR. Everything to this point has been done to ensure that the analytics are done correctly, and is as unbiased as possible. The majority of the analytics found in PLRs revolve around counting items in certain patent information fields. These are referred to here as statistical measures. The primary methods for generating these statistics, and visualizing the results, include the use of lists, for looking at one primary field at a time, and tables, or co-occurrence matrices, for working with two primary fields. Thoughts on generating lists were covered in section 6.2, co-occurrence matrices were covered in section 6.3, and layering or stacking information in relation to these items was covered in section 6.6.

While the ultimate decision on which analyses to include in a PLR will depend primarily on the business objectives, and needs associated with the requisition of it, there are a number of "standard" items that can be found in almost every PLR produced. This section looks at the statistical analyses that are generally "required" in a PLR. Once the statistical analysis is conducted, often in a simple spreadsheet program, the output is visualized and added to the report.

Visualizations are appealing and provide insight on their own but the analyst should not be shy about annotating and discussing the implications of what is seen in them. Analyst John Paul Nettles suggests the following when thinking about how charts, and tables are interspersed within a research report[85]:

I recommend using no more than one graph, chart, or bulleted list for every 1.5 pages of words. Visual appeal should not be the only thing keeping the reader awake. At the end of the day, the paper should leave the audience's need for information on the topic pretty damn well satisfied.

A good example of this approach is seen in the WIPO PLR on Membrane Filtration and UV Water Treatment[86]. In this case, the analyst is providing a potential explanation for why a trend might be occurring, in addition to showing the trend as well:

We performed a range of additional analyses around the patent datasets to examine the trends underpinning the patent datasets (see Table 6). For instance, we analyzed the average patent family size24 with more than 1 member - thus stripping out the 'noise' from patent families with a single publication. We found that for the membrane and UV datasets the average patent family has 9 or 10

---

[85] http://repcapitalmedia.com/three-reasons-to-ditch-the-charts-in-white-papers/
[86]
http://www.wipo.int/patentscope/en/programs/patent_landscapes/reports/water_treatment.html

members in both the membrane and UV water treatment datasets. However for the membrane-UV 'combination dataset the average family size was much smaller at 4 and 2 patents for the overall and desalination-focused datasets. We also saw some very large extended patent family sizes (in the 'largest patent family' category).

Possible interpretations of these findings could be that:

- *More dynamic technology areas have a higher number of SMEs, who due to resource constraints may only file one patent or abandon experimental technology,*
- *A 'younger' technology or product would have had less time to develop 'mature' patent families*
- *Presence of several core product technology of very high importance for a corporate player justifying significant investment in a large extended patent families*

One of the motivations, covered in section 5.2, for producing PLRs, involved their use to enhance organizational decision-making. PLRs need to generate insight and this is done by interpreting the analyses provided and putting them in context with the issues being investigated. Providing statistical analyses is better than providing raw data but the value of the analyst is really seen in the performance of sound analytics, well-documented visualizations and reasoned, topical interpretations of what is being observed, and why it's important.

In section 8.1.4 the preparation of a Terms of Reference (TOR) was covered, and one of the key items included in that document is an agreement on which analyses will be contained within the PLR. In the remainder of this section the mandatory fields that should be included in a PLR are defined, and a brief description of the insights they provide listed.

All of the listed analyses, with the exception of Highly Cited Patents, looks at a single variable and counts the number of times each entry appears. In addition to lists based on this data, tables can also be created with them. In all of these cases, an appropriate year type can be added as a series of columns, generating a table that buts the total number of items in context by identifying if their occurrence took place recently or sometime in the past. Some of these fields can also be combined with one another to provide additional context, for instance, technology categories can be combined with top assignees, or inventors, to demonstrate what aspects of a topic are of interest to different organizations, or researchers, working in a field.

8.4.1 – Number of Families or Inventions

After a data set is reduced by families, or inventions, as covered earlier in this section, the total investment, in patenting, for the topic can be found. This would be a single number without much context; so many analysts add granularity to this value by representing it using a year type. So while the analysis is referred to as the number of families, or inventions, it should be called the number of families by year. The consequences of using the various year types was discussed in section 8.3.4 but regardless of which one is used this analysis demonstrates whether interest in a topic, based on the number of patents that have been applied for, is either growing, declining, or static. Technologies can also exhibit an ebb and flow where initial interest wanes but then accelerates at a later date when a new application is found, or greater efficiencies are gained through improvements.

8.4.2 – Number of National Phase Entries

The general idea behind this analysis can also be represented by the number of countries covered, or the average number of family members per invention. In all cases, the analyst is looking at how

pervasive the technology is, for instance, has interest only occurred in developing countries, or has protection been sought worldwide. Filing a single PCT application and designating many countries is relatively inexpensive and straightforward, but moving from a PCT application to a National Phase filing in other countries is a signal of increased investment, and interest. When talking about this concept in terms of country coverage the analyst must also distinguish whether they are talking about only granted patents, pending applications that are actively being examined, or potential coverage that could be acquired if a National Stage filing is conducted before the deadline expires.

This information can be presented by using a bar chart with countries of interest listed on the x-axis while the number of entries, or filings for that country is plotted on the y-axis. Alternatively, if additional context, involving a year type, is required than a line chart can be used where the years are on the x-axis and individual lines on the chart represent the countries of interest.

These analyses can also signal the maturity, and uptake of a technology since emerging technologies will not be extensively covered by larger families or around the world.

8.4.3 – Number of Granted Patents

As seen with the number of families analysis, the number of granted patents would be a single number unless some additional context, once again, usually by year, is added to this. This analysis is particularly powerful when a stacked bar chart is used to compare the number of granted patents to the number of pending applications that are in the course of being examined. If total number of documents, regardless of kind is looked at, as in the case of number of families then it is difficult to determine whether a topic area has patent protection in place, as opposed to the potential of having rights granted. Only in-force, granted patents provide the right to exclude so understanding the status of the documents being analyzed provides perspective on whether the topic at hand is extensively protected currently, or if it has the potential to become a minefield in the future. Adding a year type to this analysis also informs the analyst if patent protection took place ten or fifteen years ago, in which case, the patents may be expiring soon, opening a possibility of utilizing them in the near future.

8.4.4 – Top Technology Categories and/or IPC Classifications

Generating technology categories, and their value was covered earlier in this chapter. A chart of these can be used to compare relative interest between different technological approaches to dealing with issues in a topic area. A simple chart of this item involves the listing of the categories along the x-axis, and the document count tallied on the y-axis.

As stated, analyses using this field can be even more valuable when a table is created looking at top assignees, or inventors by technology categories. Analysts can also examine interest in the categories by year by using a line chart, with each category represented by a different line and the years provided on the x-axis. The document count by year is reflected in the numbers on the y-axis.

Classification codes have been covered in sections 4.2.1.4 and 8.3.5. They can be used in addition to, or in place of the Technology Categories and generally represent the same principles of understanding relative interest between different approaches within a topic area.

8.4.5 – Office of First and Second Filing

This analysis is conducted by looking at the priority filing country associated with an invention. The priority filing country is referred to as the office of first filing. The office of first filing approximates where the invention was likely created, and provides an indicator of which countries are leading in the

development of a technology topic. Since it is country based, it can be analyzed, and visualized, in a similar fashion to how country coverage, or National Stage entries, is represented. This analysis is sometimes tempered by normalizing the output based on the Gross Domestic Product of the countries involved. This way, output from more economically developed countries is kept in perspective compared to countries with fewer resources to call upon.

The office of second filing analysis is conducted by determining the country, other than the priority filing country, with the earliest application date, in a patent family. While office of first filing indicates where a technology was developed, the office of second filing can provide insight into which country is likely to represent a good market, or location for manufacturing, of the products generated from the technology. This field is visualized using the same methods used for the office of first filing analysis.

8.4.6 – Top Applicants/Assignees

Applicants, or assignees, as they relate to patent documents were introduced in section 4.2.1.1. They represent the owner of a patent, and with whom negotiations for the rights associated with the invention will have to be conducted. They also represent the organizations who have made a significant investment, or who have a potentially high stake in a topic area of interest. Studying the top applicants ranks the organizations, associated with a technology, by the ones that have devoted the most resources in researching and applying for patents. There are additional measures for looking at investment in a technology, but the number of patents applied for is a commonly accepted means for measuring this. The applicant, and inventor fields almost always need to be cleaned, or grouped, before statistical analysis can be carried out and this was covered in section 8.3.1.

Output from this field is normally visualized using charts, especially bar, and stacked bar charts where the organization name is placed on the x or y-axis and the number of documents associated with the organization is presented on the opposite axis. This field is also used in conjunction with other fields, such as a year type, or Technology Categories, or Classification codes to provide context about the different interests one organization, versus another, or when each organization was doing the investing, to produce tables or co-occurrence matrices. The applicant field is a popular one, and is used frequently when stacking, or layering information, as was discussed in section 6.6.

8.4.7 – Top Inventors

These individuals represent the key thought leaders, and innovators within a topic area. When these people patent, in conjunction with a University, they can be considered to be a potential source of collaborators, for organizations looking to enter an industry. When they are unaffiliated, and appear to be a sole inventor, or part of a small, non-competitive group, they can potentially be a source for acquiring rights or expertise in a technology. Inventors who are associated with a competitor can also be a source of intelligence on a topic, if they are no longer employed by the company they filed their patents with.

An inventor analysis is visualized using the same types of charts and tables that are used to represent data on applicants/assignees.

8.4.8 – Highly Cited Patents

As discussed in section 4.2.1.5, citations represent a relationship between two inventions. Studying them provides a means for identifying seminal documents that could have had a high impact on the development of a technology. When discussing highly cited patents the analyst is referring to patents

within a topic area that have the highest number of forward citations. This refers to documents that have been cited by applications filed later that refer to the original document.

The importance of aggregating, where citations associated with a redundant patent application are reconciled with the corresponding granted patent was discussed in section 8.3.6. This aggregation needs to take place before any list of highly cited patents can be generated.

Since highly cited patents refers to individual inventions they are normally summarized in a text-based table, not a co-occurrence matrix, as opposed to a chart. Information on the assignee, inventors, publication year, expiration date, number of forward citations, and the companies that have cited the patent, should be included in the table.

8.5 – Additional Statistical Analyses to Consider Including

Having covered the "essential" statistical analyses that should be included in almost any PLR, it is also important to mention additional analyses that are frequently used, and can provide valuable insight, but are not on the "must have" list. This section covers these frequently used analyses, and describes the organizational value of including them.

8.5.1 – Type of Applicants/Assignee

Throughout these guidelines, the word organization has been used when discussing companies and governmental entities. This has been done since these entities have different ways of approaching and utilizing patents, in terms of their approach to research and development, for instance. Since the organizational types are different, it can be valuable to create groups based on them, and examine patenting activities based on these groupings. The most popular categories used to group entities are:

- Industry – both for profit, and non-profit companies and businesses
- Governments – research conducted by labs associated with a particular country
- Universities – while normally tax payer funded universities behave differently than governments
- Individual Inventors – people who develop technologies without being associated with a large firm

Additional sub-categories can also be created, if they are meaningful for the business objectives associated with the PLR, but these are the four major ones. The objectives and motivations for most of these entity types were discussed in chapter 5, and it can be valuable to segregate and compare the output from these different sectors as applied to a particular technology, or sub-category area. Some technologies might feature university, or governmental organizations predominantly, for instance, which could indicate that an area is still in the basic research stage and may not be ready for commercialization or application.

Type of applicant is normally generated by manual grouping, but can also be accomplished by using automated methods that look for keywords, such as Univ, that can be used to group organizations into one of the categories used. Once the groupings have been made the output is visualized using the same methods used for the Applicant/Assignee field.

8.5.2 – Percentage of Foreign Born Inventors

Working with assignees can sometimes be misleading since an organization may be headquartered in one location while the research conducted is performed in a different location not normally associated with them. Looking at foreign-born inventors can provide data on the country where the research was actually generated as opposed to where it was filed for. This statistic can also be used to speculate on the amount of partnering, and collaboration that takes place within an organization, and their research centers in different parts of the world.

Taken in total, this data can be used, in conjunction with the Office of First Filing data to provide a more accurate picture of what countries are the real drivers for the development of different technologies or approaches to a topic.

8.5.3 – Patent Quality Indicators

There are many methods, and arguments to assess whether a patent is of high quality. These methods form part of the so-called patent valuation. It is beyond the scope of these guidelines to evaluate the patent valuation methods. Neverthess, there is a brief discussion on the practice in section 10.1. The issue of patent quality and the ways to assess it has been discussed for a long period of time and it is the issue on whether it can be assessed in an accurate way is disputed. It remains a fact that various methods are being used in practice to assess the quality and related value of a patent, as the latter remains an asset with economic value that needs to be assessed on various occasions and for various reasons. Regardless of the method used, it can be insightful to compare the number of high quality patents coming from different countries, organizations, technological sub-categories, or time periods.

It is important to provide some details on the method used to generate the quality scores in the methodology section of the report, and ensure that the method is applied consistently over all of the documents being studied.

8.5.4 – Number of Patent by R&D Spending

As suggested in section 8.4.6, there is more than one way to measure investment as it pertains to research and development (R&D). Analysts will often look at actual R&D currency spent as another measure of interest. These two metrics can be combined to evaluate the effectiveness of a research program, or country by looking at the number of patent filings generated based on the amount of money spent on R&D. The theory goes that a higher number of filings relative to spending makes for an effective use of currency, while a lower number of filings suggest inefficiencies, or difficulties in translating money spent into defined inventions.

This measure, of course, is nuanced and should be used selectively, and in context. Often a comparable with another organization in the same technological category, or a comparison between closely associated sub-categories will be useful for providing insight.

8.5.5 – Percentage of Triadic Families

Triadic families have traditionally included a granted patent in the United States, Europe and Japan. It can be argued that there are other countries that should be added to the list in general, or countries selected should be based on relative strengths associated with a particular technology. In any case, looking at the percentage of families that have multiple, relevant countries included indicates maturity and likely commercialization potential within a technology area.

When looking at organizations, it is also speculated that the ones with a high percentage of "triadic" families are likely to have long-term interests in a technological area.

8.6 – Additional Analysis Types to Consider Including

While most analyses, include in PLRs are statistical in nature, they are not the only methods used to provide insight. Other analysis tasks were covered in Chapter 6 and many of these methods find their way into PLRs. These methods are considered to be a little more advanced and typically require the use of additional tools outside of a spreadsheet application. Tools for conducting these types of analyses are shown in section 9.1 of these guidelines.

8.6.1 – Citation and Co-Inventor Networks

Network analysis was discussed in section 6.8; the most relevant analyses pertaining to the creation of PLRs involves citations between assignees and, co-invention networks.

Citation networks involve the visualization of forward and backward citations between a large collection of documents within a topic. Traditional citation analysis would start with a single patent, as a root, and show forward and backward citations only from the root document. Subsequent generations could be added, but they were treated as new roots as opposed to showing shared connections from one generation to another with a document that was common to both.

With a network diagram, each patent document is a node and all of the connections, both forward, and back, are represented as edges, regardless of when they occurred and whether there was a linear, or direct connection between the nodes. In this fashion, connections are shown even if the citing skips a generation or two. With this type of visualization, seminal or lynch pin documents, one that get cited frequently over time, can be identified regardless of whether there is a direct connection or not.

With a big collection, there can be a large number of nodes, and direct citation network visualizations can become very busy and difficult to interpret. To address this issue many tool providers will correlate the individual patents into groups based on their assignment and then use the organization name to label the node. This approach reduces the number of nodes and provides insight into which organizations are most well connected and regarded within a technology area.

Co-inventor networks share the same qualities, nodes, and edges, and connections that don't rely on direct linkages, but individual inventors are used as nodes, as opposed to patent numbers, or assignees. The patent documents are still the source of the inventor data but the analyst, in this case, is looking for relationships, such as a student, and a professor, between individuals. This type of analysis can also be combined with nodes for assignees to visualize the movement of key people from one organization to another. Consulting agreements, and partnerships can also be identified this way by looking for connections between individuals who are associated with multiple assignees.

One of the biggest concerns, for analysts, working with network analysis tools, is dealing with networks that contain a large number of nodes. The key to making diagrams that can be understood by clients is to cut down on the number of nodes that are visible to the user. This is typically done by filtering based on the number of documents associated with the node representative. An analyst may decide, for instance, to restrict nodes to only those data points that have at least three occurrences associated with them. This will eliminate nodes for minor contributors in an area. Filtering can also be done on the number of occurrences that define the edges as well.

Network analysis can be a very powerful visualization technique but it has not fully caught on with the majority of vendors who make tools in the patent analysis space. Hopefully, more examples of this type of functionality will be available in the future.

8.6.2 – Spatial Concept Maps

Spatial concept mapping, is related to clustering, or classification, since it generally begins with one of these methods, but adds an extra component, identification of relative similarity between the categories created, to the task. The tools involved take the document clusters, or classes, and arrange them in 2-dimensional space by considering the similarity of the documents, or clusters, relative to one another, over the entire collection. Documents that share elements in common are placed closer together spatially, while ones with less similarity, are placed further away This analysis task was introduced in section 6.5. Using layers in conjunction with spatial concept maps was covered in section 6.6. Most spatial concept maps begin with a clustering, or unsupervised machine learning step, which was covered in section 6.4.

Since there seems to be an X and Y-axis on most maps, many users think these visualizations behave like a scatterplot, where extrapolating between the X and Y can identify empty spaces on the map. In reality, there are no X and Y-axis associated with the maps and the distance between documents, usually represented by dots, are based on similarity of the documents to one another and compared to all of the other documents in the collection. Distance therefore, is relative, based on the document collection and guesses cannot generally be made about what sort of document might occupy an empty space on the map.

While the maps, and document organization, is provided in two-dimensions a third-dimension is often implied by incorporating document density. The tightness of the clustering, in a group, or the number of documents, found in the group, will be used to demonstrate which groups have the highest number of documents in them. On a topographical version of spatial maps this is represented by an implied increase in peak heights on the map, visualized using a change in color. Many of the spatial maps, especially the ones based on clustering methods, also provide contour lines on the diagrams. Generally, these lines are drawn based on the distance between the document dots. The distance between a dot and its nearest neighbor determines the boundaries of the lines. Once the threshold is exceeded the line is drawn between the two dots. It has been speculated that contour lines encompassing multiple groups on a map implies a relationship between these groups, but generally, this is not the case and the lines are simply based on the spread of the documents.

There are a few keys to creating good spatial concept maps that will be more easily interpreted by clients. The first involves the choice of words used to generate the vector that will be compared between documents. When working with full-text patent documents an analysis of this type should be restricted to certain sections of the document, such as the claims, or the titles and abstracts. Working with the entire body of text can confuse the system since there are sections, such as the background of the invention that are talking about other inventions, as opposed to the one covered by the patent. In addition, when working with full-text, the words chosen by the algorithm creating the vector will likely be sub-optimized since there are so many words to choose from.

Users can selectively add stopwords to their map settings. Stopwords are also referred to as non-content bearing words, and they can adversely impact clustering results if they are included in the vector since they do not impart knowledge of the topic area. Almost all mapping tools come with a list of standard stopwords, such as "the", "and", "a", and other non-content bearing terms, but users can also look at initial results and identify other words that do not add meaning to the technology being

examined. New words can be added to stopword lists within tools on a map-to-map basis, or permanently. Modifying stopwords provides an analyst with a means for influencing the placement of documents on a spatial concept map.

Finally, once the analyst feels comfortable that the system has done a reasonable job clustering documents, they can change the labels on the map so they reflect the terminology used by the stakeholders of the PLR. Most systems generate labels on these maps by looking at frequently used words, or terms, especially if they are unique to a particular cluster. Sometimes this works well, but often the label terms are too generic and don't really reflect the contents of the cluster. The clustering, in fact, may have been quite good, but a poor label may be the first, and only, thing that a client sees. If the labels are poor, and don't reflect meaningful categories, the client can lose interest or believe that the map is not meaningful. Labels can be changed within most mapping tools and should be done on a cluster-by-cluster basis by examining the titles of the individual documents with them.

8.6.3 – Problem/Solution Semantic Examination

Using semantic analysis based on subject action object triplets to build a knowledge base based on a collection of patents was discussed in section 6.9. Using semantic analysis to assist with the building and population of Technology Categories was also covered in section 8.3.5. The aggregation of a collection of problems gathered from a set of disparate documents can be represented as a knowledge base, and can provide a variety of potential solutions even if they are not found within a single document. Once a semantic analysis has generated categories, within a topic area of interest, it is possible to count the number of documents associated with each of them. In this sense it is the same as Technology Category visualization.

Alternately, the knowledge base can also be represented as problems, and their corresponding solutions, for representation within a PLR. A nice means for visualizing a problem/solution can be seen with the use of a mind map. Wikipedia contains the following definition for mind maps[87]:

A mind map is a diagram used to visually outline information. A mind map is often created around a single word or text, placed in the center, to which associated ideas, words and concepts are added. Major categories radiate from a central node, and lesser categories are sub-branches of larger branches. Categories can represent words, ideas, tasks, or other items related to a central key word or idea.

With this method, the topic can be used as the central node, problems listed as major categories radiating, or branching from the central node, and the solutions used as sub-branches of the larger, major category branches.

8.6.4 – Top Patents for Immediate Consideration

While most analytics are conducted on a macro-level, as described in section 7.2, frequently analysts will come across patents that appear to be particularly relevant to the organizational needs associated with the tendering of a PLR. When this occurs, they will often call these patents out so the clients will immediately have some information on these documents without having to refer back to the raw data associated with the project.

Documents of this type are determined to be important for a variety of reasons, especially if they are coming from a major competitor, or represent potentially valuable patents from an organization of

---

[87] http://en.wikipedia.org/wiki/Mind_map

interest to the PLR stakeholders, for one reason or another. These patents should be summarized in a table that includes information on the assignee, inventors, publication year, expiration date, number of forward citations, the companies that have cited the patent, and the technology category or sub-categories associated with the document. If applicable, the first claim, or some form or claims summary, may also be included in the synopsis of these patents.

8.7 – Writing and Publishing the Report and Accompanying Data

All of the component pieces have now been assembled, it is time for the analyst to interpret their results, and decide on the most critical insights that will be shared in the deliverables to the decision-makers, and stakeholders in the PLR project.

8.7.1 – Writing the Report

A written report should be part of the deliverables associated with the completion of a PLR project, and ideally include the following sections:

- Executive Summary – where key learnings and relevant observations that reflect the objectives of the report should be shared. This maybe the only section that is read in any detail, so it should include all of the major findings.
- Introduction – should provide a legend, or key to how the PLR is organized and demonstrate the major sections of the report to the reader.
- Background on the Technology – a brief description of the technological topic area being studied. This section should also include a definition of the various subcategories used in the PLR, if any have been identified.
- Background on Patents – most readers, especially executives, will have limited experience with patents and patent information, a brief description of this area, similar to the backgrounder provided with these guidelines, should be included.
- Justifications for Creating PLR – The objectives and goals for creating the PLR should be clearly stated at the beginning of the documents since they will provide the lens through which the reader should consider the remainder of the document.
- Economics Associated with the Topic – regardless of the potential humanitarian purposes behind a PLR, there is always a financial component associated with any key organizational decision. Some basic information about the economics behind the industry, or topic area provides important context for the report.
- Methodology – There are four issues which should be covered here:
    - Search Strategy – as discussed in section 8.2
    - Data Preprocessing – as discussed in section 8.3
    - Analysis, or Task Methods Used – as discussed in sections 8.4-8.6, and introduced in chapter 6
    - Issues and Limitations – there are always assumptions made, and disclosures that should be shared, about the availability, and use of patent data, and how it's interpreted. These should be covered here.
- Analysis section – relevant charts, tables and assorted analytics, along with detailed commentary, on the insight and implications of the data.

- Additional Resources – other sources that the reader can use to discover additional information, and gain insight from others on the topic.
- Conclusions – a summary of the major findings and insights, along with recommendations for action, associated with the objectives, and topic of the report.

Most PLRs are 20-40 pages long, when provided in long form, which should include all critical charts and tables. As discussed in section 5.2, the main purpose of a PLR should be to assist with organizational decision-making and should help an executive or manager come to an informed decision quickly. A balance must be established between a report that has too much detail, and may not be read, and one that does not provide enough insight, and thus is not useful for decision-making.

8.7.2 – Publishing the Report and Accompanying Data

The manner in which the PLR is delivered to the stakeholders is one of the items that should be agreed to while the project is being planned, and before any of the work is initiated. In section 5.2.1 there was also some discussion on the learning styles associated with the intended audience for the PLR, and this should be taken into consideration when deciding whether a long form written document or a presentation style deliverable is appropriate. Detail oriented decision makers will almost always want to see a long form written document, while an intuitive thinker will appreciate the abbreviated style of a presentation document.

These are guidelines for Patent Landscape Reports, and with report in the name the expectation is that there will be a report generated in association with the project. In some circumstances, the report will be the sole deliverable, but in many instances analysts will also provide the data used to generate the analyses conducted. If available, analysts may also decide to load the data into an interactive visualization tool, which will allow the readers to explore the environment as well. While the analyst is the key driver in providing insight some users, especially the analytical thinkers, will appreciate having access to tools and data to ponder on their own. Many of the analysis tools mentioned in section 9.1 provide "reader" or simplified versions of their tool that can be configured for use by an end-user.

8.7.2.1 – Published Reports

PLRs can be produced in a variety of different written formats. For the purposes of these guidelines, the written formats have been divided into long form, and presentation styles.

The overwhelming favorite for long form documents is the use of the Portable Document Format or PDF. Originally developed by Adobe the PDF format allows the creation of stylized long form documents that can be viewed on almost any computer platform. They are the preferred long form output since they will look, and print the same way, regardless of what system the eventual reader is using. Wikipedia describes the PDF[88] document as follows:

Portable Document Format (PDF) is a file format used to represent documents in a manner independent of application software, hardware, and operating systems. Each PDF file encapsulates a complete description of a fixed-layout flat document, including the text, fonts, graphics, and other information needed to display it. In 1991, Adobe Systems co-founder John Warnock outlined a system called "Camelot" that evolved into PDF.

---

[88] https://en.wikipedia.org/wiki/Portable_Document_Format

Word processing programs, including MS Word are often used to generate PLRs, since it is relatively easy to review, and track changes with these applications, but they are not used as frequently for sharing the final long form document. Often different word processing programs, and even different versions of the same application, will display results differently from one another. Word processing programs can be used, but the analyst should be aware that formatting and breaks might not transfer as anticipated to other readers of the report.

Some analysts also decide to create web-based versions of their long form documents. An example of this type of PLR can be found at:

http://www.patentlens.net/daisy/adjuvants/Introduction.html

Creating a web-based version of a PLR is supposed to offer some of the same benefits as using a PDF document, with regards to retained formatting, but there can be issues with compatibility and interpretation between different browser applications. The primary advantage of a web-based report is that extensive reports can be broken into sections, as opposed to downloading a single large document, to speed viewing and sharing. The drawback though is that a reader will need to have continuous Internet access to keep accessing the report since there is not, ordinarily, an off-line method for reading if the web-based report is the only one provided.

Presentation style deliverables of PLRs are considered to be a little less formal, and appeal more to intuitive thinkers, but they can be used effectively to produce directed outputs that speak specifically to the objectives of the project. In many circumstances, analysts will also be asked to present on the key results of a PLR, regardless of whether a long form document is generated, and thus will also be producing a presentation style document.

While there are many presentation applications available, MS PowerPoint is the overwhelming standard, and will likely be used for sharing with others. Some analysts may feel more comfortable generating slides in other programs, but eventually these will probably need to be exported, and vetted in PowerPoint format.

Creating effective PowerPoint presentations is an art in itself, and will not be covered extensively in these guidelines. There are many who feel that slide presentations should follow the 10/20/30 rule[89], or ten slides, twenty minutes and 30 point font, when producing content, but this rule is not particularly effective unless a true executive summary is being given. When a presentation style document is the primary deliverable for a PLR then many more slides and an appendix will often be used. The primary points are made using representative slides, while background and supporting data slides are placed in the appendix where they can be quickly, and easily be referred to, if needed.

8.7.2.2 – Published Data

Analytical thinking stakeholders in a PLR will appreciate being given access to the data used to create the analyses associated with a project. This is especially the case when the PLR was produced by a third-party doing work on behalf of another organization. Under these circumstances, the analyst should provide only the post-processed, value added data for delivery to the client. Raw data should never be provided, since it will not lead to results that are reproducible. While data is generally provided to allow users to address different questions, and explorer unanticipated circumstances, it is also used to evaluate the analytic methods of the analyst. As such, a different analyst should be able to reproduce the results if provided with the same data.

---

[89] http://blog.guykawasaki.com/2005/12/the_102030_rule.html

Data can be shared in several ways but is almost always provided as a file from a commercial spreadsheet application, such as MS Excel. Database files, including MS Access, or SQL are also provided on occasion. A good example of a WIPO document that includes the appropriate data access in a MS Excel file can be found at:

http://www.wipo.int/patentscope/en/programs/patent_landscapes/reports/vaccines.html

8.7.2.3 – Published Visualizations

While there are charts and tables included in the long form, or presentation style representations of the written report, an analyst can provide additional functionality to the data associated with a topic if they use static, or interactive visualizations in conjunction with it. Interactive visualizations, in particular, allow the reader to explore the data themselves, and explore new questions, which may not have been identified when the objectives for the PLR were first established. Static visualizations, normally large, high-resolution images of very detailed maps and network diagrams can be posted to a data repository, or website associated with the project, especially when the primary deliverable is a long form report, since these images may not be able to read easily in the document.

A good example of a WIPO document that provides interactive visualizations associated with a PLR can be found at:

http://www.wipo.int/patentscope/en/programs/patent_landscapes/reports/atazanavir.html

8.8 – Evaluating the Report

Once a PLR is delivered, the clients will evaluate it. The ultimate validation for an analyst is to learn that the organization implemented the recommendations suggested by them, and this resulted in a successful outcome that followed the reasoning proposed. An analyst will also know if their project was successful if the clients, or perhaps more importantly, their associates, return to them with additional work. Word travels quickly when a critical decision-making resource is discovered.

Some organizations, especially governmental ones, will put more formal evaluation criteria in place. WIPO, for instances uses the following criteria for evaluating the PLR projects that they have contracted out for:

The key purpose of the evaluation was to assess whether the Project provided the right type of support in the right way based on four main evaluation criteria:

- *Relevance: The extent to which project objectives were consistent with beneficiaries' requirements, member countries' needs, global priorities and policies.*
- *Efficiency: How economically inputs (e.g. funds, expertise, and time) were converted into results ("value for money").*
- *Effectiveness: The extent to which objectives were achieved, or are expected to be achieved, taking into account their relative importance.*
- *Sustainability: assesses the likelihood of continuation of project benefits after the assistance has been completed.*

Formal evaluation documents may include the following sections as well:

- Findings and Assessments – detailed discussion on the evaluation criteria and how they were applied to the particular project
- Recommendations – key learnings that the organization can apply to future work conducted
- Conclusions – discussion on whether the project was successful and what impact it had on the organizational objectives

Once again, the most critical evaluation criteria are whether the PLR was read, and acted on by the organization.

Chapter 9: Third-Party Analysis Tool, Patent Database and Report Production Providers

Lists of analysis tool providers, patent databases, and contractors, available to conduct work, or provide services associated with PLRs, can be found in the sections below. These lists are not meant to be comprehensive, but provide some additional sources of information for individuals who would like to explore this topic further.

9.1 – List of Analysis Tool Providers

Search Technology – VantagePoint – Text-mining tool for discovering knowledge in search results from patent and literature databases – http://www.thevantagepoint.com

Search Technology / Thomson Reuters – Thomson Data Analyzer – Version of VantagePoint designed to work with Thomson Reuter's data – http://ip.thomsonreuters.com/product/thomson-data-analyzer

Thomson Reuters – Thomson Innovation – Integrated patent and non-patent literature searching and analysis system, featuring ThemeScape thematic maps – http://info.thomsoninnovation.com

Chemical Abstracts Service / FIZ – STN AnaVist – Online analysis tool designed to work with the STN International online system, contains dynamic charts and thematic maps – http://www.cas.org/products/stn/anavist

Questel – Orbit.com – Integrated patent searching and analysis system, contains network and thematic mapping tools – http://orbit.com/#WelcomePage

Innography – Correlates patent data with financial, litigation and other key business information to instantly generate a variety of unique visualizations – http://www.innography.com

Treperal – KMX Patent Analytics – Patent text mining and visualization software solutions for various stages of R&D and IP management – http://treparel.com

Intellixir – Intuitive navigation through thousands of documents using powerful visualization and patent mapping tools – http://www.intellixir.com

Gridlogics – Patent iNSIGHT Pro – Patent research and analysis platform that includes specialized text mining algorithms for patent and scientific literature, powerful charting and mapping capabilities – http://www.patentinsightpro.com

BizInt Solutions – Smart Chart for Patents – Create and customize tabular reports from patents databases including Derwent World Patents Index, CLAIMS, and CA/CAPLUS – http://www.bizcharts.com

LexisNexis – TotalPatent – Research and analyze patent data and protect intellectual property, contains user-driven text mining tools – http://www.lexisnexis.com/en-us/products/total-patent.page

Matheo Patent – Allows the integration of all patent database formats, use for creating statistical lists, charts, matrices, networks and clustering – http://www.matheo-software.com/en/products/matheo-patent.html

Microsoft – Excel – Industry standard spreadsheet tool, also contains PivotTable, charting, analysis and visualization capabilities – http://www.microsoftstore.com/store/msusa/en_US/pdp/productID.259321400?siteID=SRi0yYDlqd0-reDw70O5LZZIvfBHNuVkZA

Minesoft – PatBase – Integrated patent search and analysis tool, contains various charts, graphs and mapping tools – http://www.patbase.com/login.asp

Open Refine – formerly Google Refine, a free, open source, power tool for working with messy data – http://openrefine.org

TIBCO – Spotfire – Visualize, interact with, and share data with data discovery and visualization – http://spotfire.tibco.com/en/discover-spotfire/what-does-spotfire-do/data-discovery-and-visualization.aspx

Linguamatics – Text mining software that can be used to analyze scientific literature, patents and other sources, features subject, action, object triplets – http://www.linguamatics.com

AcclaimIP – Patent search and analysis software for attorneys, analysts and patent owners – https://www.acclaimip.com

Relecura – Patent and portfolio analysis platform that simplifies tasks related to IP creation, prior-art searches, IP landscaping, and IP licensing – http://www.relecura.com

AmberCite – AmberScope – Patent analysis and mapping tool based on citation-based network patent analysis – http://www.ambercite.com

IPVision – See-the-Forest – Patent mapping and analytics based on backward and forward citations – http://www.see-the-forest.com/G4/Main.act

TEMIS – Identifies and extracts targeted information to semantically enrich content with domain-specific metadata – http://www.temis.com/home

Pantros IP – Provides patent quality metrics based upon the commercial value, legal strength, and technology value of a patent, also includes Natural Language Processing based searching – http://www.pantrosip.com/solutions/patent-analytics

Instem – OmniViz – off-the-shelf data mining and visual analytics tool that allows users to analyze and explore data sets through interactive visualizations – http://www.instem.com/solutions/omniviz.html

Landon IP – Patent Workbench – Track and compare claims, quickly identify subtle, but potentially critical, changes in claim language – http://www.patentworkbench.com/Home/Features

PatentInspiration – Free online patent search and analysis tool with more than 40 different analyses – http://www.patentinspiration.com

Cambia – The Lens – A worldwide, open-access, free full-text patent informatics resource – http://www.lens.org/lens/

Wisdomain – Worldwide online patent searching and analysis system – http://www.wisdomain.com/wis_html/en/index.htm

9.2 – Non-exhaustive List of Database Providers

Patent Office Sources

World Intellectual Property Organization – Patentscope – http://patentscope.wipo.int/search/en/search.jsf

European Patent Office – Esp@cenet – http://worldwide.espacenet.com/?locale=en_EP

European Patent Office – PatStat – http://www.epo.org/searching/subscription/raw/product-14-24.html

United States Patent and Trademark Office – Patent Search page – http://www.uspto.gov/patents/process/search/index.jsp

Korean Institute of Patent Information – KIPRIS – http://www.kipris.or.kr/enghome/main.jsp

Japan Patent Office – IPDL – http://www.ipdl.inpit.go.jp/homepg_e.ipdl

Chinese Patent Office – Patent Search – http://59.151.93.237/sipo_EN/search/tabSearch.do?method=init

Free Online Sources

Google Patents – http://www.google.com/?tbm=pts

United States Patent and Trademark Office Bulk Data Downloads – http://patents.reedtech.com

FreePatents Online – http://www.freepatentsonline.com

TheLens – http://www.lens.org/lens/

PatentInspiration – http://www.patentinspiration.com

SureChem – SureChemOpen – https://www.surechem.com/products/open/

Commercial Sources

STN International – STN – http://www.stn-international.de/index.php?id=123

Questel – Questel-Orbit – http://www.questel.com/index.htm

Thomson Reuters – Thomson Innovation – http://info.thomsoninnovation.com

Minesoft – PatBase – http://www.patbase.com/login.asp

LexisNexis – TotalPatent – http://www.lexisnexis.com/en-us/products/total-patent.page

Chemical Abstracts Service – SciFinder – https://scifinder.cas.org/scifinder/login

ProQuest – DIALOG – http://www.dialog.com/proquestdialog/

IFI Claims – CLAIMS Direct – http://ificlaims.com/index.php?page=products_data_global

Gridlogics – PatSeer – http://patseer.com

Pantros IP – http://www.pantrosip.com/solutions/patent-search

ArchPatent – http://www.archpatent.com

AcclaimIP – https://www.acclaimip.com

GenomeQuest – http://www.genomequest.com

Worldwide Intellectual Property Service – WIPS – http://www.wipsglobal.com/ZZ0000/

9.3 – List of Landscape Report Service Providers

Landon IP – http://landon-ip.com/IPAnalytics.aspx

Thomson Reuters – http://ip.thomsonreuters.com/product/patent-analytics-services

Cambridge IP – http://www.cambridgeip.com/services/landscape/iplandscape/

Griffith Hack – http://www.griffithhack.com.au/informationservices-LandscapeSearches

Patinformatics – http://www.patinformatics.com/about/about-our-services/

Fist & Frinnov – http://www.fist.fr/en/ip-overview/index.html

IP Checkups – http://www.ipcheckups.com/custom-consulting/

Dolcera – http://www.dolcera.com/website_prod/services/ip-patent-analytics-services/patent-landscaping

Perception Partners – http://www.perceptionpartners.com/tools.htm

Totaro & Associates – http://www.totaro-associates.com

Cardinal IP – http://www.cardinal-ip.com/products-and-services/patent-searches/search-types/landscape-search/

Cantor Colburn – http://www.cantorcolburn.com/practices-Patent-Landscaping-Risk-Analysis.html

IPCalculus – http://www.ipcalculus.com/patent_landscape.html

The Small Patent Law Group – http://www.splglaw.com/Specialties/PatentLandscapingServices.aspx

IP Pragmatics – http://www.ip-pragmatics.com/landscaping

TechInsights – http://www.techinsights.com/services-and-solutions/technology-market-assessment/

Chapter 10: Additional Topics Related to the Strategic Use of Patent Information but Not Covered in Guidelines

As seen in section 4.4, there are reports, other than Patent Landscape Reports (PLRs), which take advantage of patent information as their primary data source. In a similar fashion, while PLRs are key instruments for the formation of organizational strategy, they are not the only tools used to develop one. This is especially the case in industries that are technologically focused, and require an alignment of a patent strategy with the corporate goals of the participating players.

This chapter provides some background, and references associated with additional topics that involve the analysis of patent information for the development of organizational strategy. These guidelines are focused on the use of patent analytics to generate PLRs, but the following coverage of additional topics that incorporate patent analytics are provided for reference.

10.1 – Patent Valuation

In essence, there are two philosophies associated with the valuation of patent documents. One is based primarily on economic and market driven principles, while the other focuses more on the scope of the claims associated with the patent of interest. In a blog post on the Seeking Alpha site[90], an organization called Markman Advisors provided the following assessment of the two approaches.

The economic or market driven approaches include the following:

- *One is a **cost-based approach**, which seeks to capture the value of a patent by estimating the cost of replacing the patented technology with another technology. By calculating the total costs of developing the patented technology economists usually arrive at a very limited value, one that is exclusively based on a single factor - cost.*
- *Second, some valuation experts use a **market-based approach** that values a patent by comparing it with other transactions that involve similar patents. However, if no similar transactions are available, this approach is of little use.*
- *In a third model, use of an **income-based approach** values a patent on the basis of the future income derived from utilizing the patented technology. Essentially the present value of a patent is derived from future income, taking into account the net cash flow (extra cash earned due to patented technology), the duration of income, and the discount rate, factoring in inflation, risk, interest rates, etc.*
- *Finally, the **option-based approach** is based on the option pricing theory developed for use in pricing financial options. It takes into account the value of the options involved in the R&D projects that lead to patents, the choices made in the prosecution of the patent, as well as in the post-grant phase of patent commercialization. Essentially you can think of the patent filing process or commercialization of the underlying invention as a series of call options (the right to buy at a future date) and abandoning the patent or terminating the R&D project as a put option. Needless to say, this is an extremely complicated method for valuing patents and in many ways is unrealistic.*

---

[90] http://seekingalpha.com/article/1540362-investor-friendly-patent-monetization-a-marathon-not-a-sprint

They contrast these methods with the prosecution history based methods by saying:

The takeaway is that such valuation methods are based on the success of the underlying protected technology. They don't take into account the fundamental meaning of a patent - not the affirmative right to practice (e.g., make, use or sell) the protected invention, but the right (derived from Article 1, Section 8, Clause 8 of the U.S. Constitution) to exclude others from practicing the patented invention. This exclusion concept of patents is what differentiates patents from a frequently compared-to real property right, where the right to exclude ensures the right to use.

More can be read on Markman Advisors' thoughts on the exclusionary nature of patents and it's impact on their potential value at the site reference provided. The next section also provides examples of the use of file wrapper analysis, including a discussion of potential claim scope, and their application to patent valuation.

10.2 – Prosecution history or file wrapper analysis

The topic of prosecution histories, and file wrappers was introduced in section 4.3.1. While the primary use of a file history is to understand how the claims of a patent document developed during the prosecution process, there is also additional information contained in the file that can assist with the determination of potential value associated with a patent document, and its associated portfolio family. While the US is used as an example below, the same principles can be applied to analyzing the prosecution history of almost any document from any patenting authority that makes it patent register publically available.

Public PAIR[91] is the USPTO system for allowing any interested party to look at the prosecution history associated with a granted United States Patent or a published pre-grant application. The information in PAIR is also referred to as a file wrapper (or file history) for a particular patent application. Prior to 2002 (or thereabouts) this information could only be obtained by going to the Public Search Room at the USPTO, or ordering a copy of the file history from a variety of different agencies. This was a reasonably expensive process since the various agencies (including the USPTO) charged for file wrappers by the page, and there was traditionally no away of knowing how many pages the document might be in the first place or deciding that you really only wanted to see some of the more critical documents, as opposed to the entire wrapper.

This all changed when the USPTO began making the file wrappers available electronically through Public PAIR. With this information more readily available it is easy for anyone interested in a patent to look at the course it took during prosecution. Looking at this information in detail can also help analysts decide how valuable a patent could be based on a number of different criteria that can be investigated in Public PAIR.

To help new analysts familiarize themselves with US Public PAIR an infographic was developed, which looks at the various tabs in PAIR, and points out the portions of the system that can be used to help identify the potential value of an individual US patent application. The infographic can be downloaded from the URL below:

http://www.patinformatics.com/wp-content/uploads/2013/01/Look_For_In_Public_PAIR.png

Public PAIR is organized using a tab interface, and the infographic goes through the following tabs in some detail:

---

[91] http://portal.uspto.gov/pair/PublicPair/

- Application Data
- Continuity Data
- Patent Term Adjustment
- Display References
- Image File Wrapper

By going through each of these sections an analyst can provide a prosecution level analysis of the potential value of a patent.

Additional discussion on the details on this process can be found at the following URLs:

http://www.patinformatics.com/blog/file-wrapper-analysis-of-us6932368-apparatus-for-harnessing-wind-to-drive-a-bicycle/

http://www.patinformatics.com/blog/google-patent-us-file-wrapper-download-of-us6932368-apparatus-for-harnessing-wind-to-drive-a-bicycle/

http://www.patinformatics.com/blog/a-first-look-at-patent-document-search-a-new-feature-on-reed-tech-patent-advisor/

10.3 – Mapping patents to products

It can often be useful to identify the patents associated with specific products produced by an organization. A recent survey[92] suggests that 60% of the triadic patents owned by an organization are commercialized in some fashion. The percentage of patents commercialized decreases as the size of the organization increases, so very large companies only use about 20% of their portfolios, but understanding which products are covered, and which ones are not is a critical corporate exercise. The following blog post from Innography[93] further explains this practice, and provides some rationale for why an organization would go through this exercise.

The business world runs on products. Profits and losses, revenue forecasts, and product offerings are all the lifeblood of a company and they're all driven by products. It should come as no surprise, then, that you would want to protect your products with patented technology — but it turns out there is more to it than just protection.
With a patent to product mapping, you could start assigning a true value to patents. With that improved capability you have a greater capacity to maximize the return on your intellectual property investment.

That's because you would have a better understanding of which patents are your most valuable and which have very little value. If you could tie your patents to your products the actual assessment could be performed with greater accuracy and precision, which in turn could help you:

- *Understand whether or how much to prune your portfolio because you could tie it more directly to your balance sheet*
- *Gain insight into new licensing opportunities*
- *Determine how to direct future R&D investments*

---

[92] http://www.rieti.go.jp/jp/publications/dp/09e011.pdf
[93] http://www.innography.com/blog/?p=18

- *Enhance your existing capability to manage protect and exploit your patents*

An additional application of this practice occurs when organizations file US Continuation Applications in order to create new claims, which expressly cover new products, from competitors, as they enter the market that may incorporate a patented technology.

It is common practice, with many inventions; to always have at least one application pending in the United States so additional claims can be written by filing Continuation, or Divisional applications. Once the last pending application has issued (with the fees being paid) the door closes for the filing of new claims by these routes.

It is also a sign of good patent strategy when a company follows this practice since it allows them to shape their claims to cover new entrants to their field. This is possible since in the definition of a Continuation application[94] it expressly says, "… where an applicant may not have exhausted all useful ways of claiming different embodiments of the invention".

In this case the applicant is writing claims on different embodiments based on what they see being developed or marketed by a competitor, as they enter the space. Additional discussion on the process of claim shaping and the mapping of new claims to cover a competitor's products can be found at the URL below:

http://www.patinformatics.com/blog/patent-strategy-lesson-shaping-patent-claims-to-match-changing-markets/

10.4 – Litigation analysis

The reality of patent monetization is that in order to be taken seriously it is sometimes necessary to engage in litigation over patent infringement. The threat of legal action is sometimes required even when the goal of the patent owner is to engage in a negotiation for a reasonable license to the technology developed. Understanding the patent litigation environment, and how it operates is a key requirement for organizations who intend to operate in technology driven industries.

Both PriceWaterhouseCoopers, and RPX provide litigation reports that provide data on the filing of patent related lawsuits in the United States. The landing page for the 2013 PWC report[95] provides the following rationale for generating this data:

Prior to 2012, only three patent cases eclipsed the $1 billion mark in damages awarded. But last year alone, three cases, tried before juries in separate districts, resulted in awards of $1 billion or greater. Additionally, NPEs continued to play a significant and growing role in patent litigation in 2012. Our analysis shows a significant disparity in median damages awarded to NPEs versus practicing entities. Over the last 12 years, the median damages award for NPEs has averaged twice the median award for practicing entities.

The AIA also made an impact in 2012. The 'anti-joinder' provision of the AIA, which constrained the number of defendants that could be named in a single lawsuit, resulted in an increase in the overall number of suits, particularly those filed by NPEs. The AIA also largely ended the phenomenon of false marking 'qui tam' actions that had become so prevalent in 2010 and early 2011.

---

[94] http://en.wikipedia.org/wiki/Continuing_patent_application
[95] http://www.pwc.com/us/patentlitigation2013

Recognizing these developments and business and civic leaders' continuing deep interest in intellectual property matters, PwC maintains a database of damages awards and other case information related to identified patent infringement decisions. We collect data related to patent holder success rates, time-to-trial statistics, jury versus bench comparisons, and practicing versus non-practicing entity (NPE) statistics from 1995 through 2012. This year's study also includes statistics by judge. Our analysis yields a number of observations that can help executives, legislators, and litigators assess their patent enforcement or defense strategies, as well as the impact of NPEs.

RPX Corp[96] is a defensive aggregator, which in the world of patent monetization is a company that acquires patent assets so they can't be used against their member companies. They describe themselves by saying:

Any company that uses technology in its products or services today faces a steadily increasing threat of patent litigation. That threat is already costing operating companies more than $10.9 billion per year in legal costs and lost productivity.

RPX is changing this equation. Our market-based solution dramatically reduces patent-related costs for client companies by sharing risk across our network. We aggregate capital from annual subscription fees to acquire dangerous patents and patent rights, with each RPX client receiving a license to every asset we own. As the network continues to grow and our service offerings expand, we are removing progressively more high-threat patents – and more high-cost risk – from the operating ecosystem.

The result: strong, broad-based defense against wasteful patent litigation and dramatically lower operating costs and financial risk for our clients.

Beginning in 2012, RPX began compiling statistics on patent litigation in the United States[97]. The introduction of the 2012 report provides the reasoning for why they believe this is an important endeavor:

In 2012, patent monetization, including that by non-practicing entities (NPEs), once again made significant headlines. Despite the increasing prominence of patent monetization and the role NPEs play, limited information regarding the industry exists. With this in mind, RPX has decided to produce an annual report (this is the first) that includes comprehensive data on cases filed by NPEs. RPX hopes that an annual report will provide much needed transparency into significant economic activities that have long fallen under the radar.

Studying and following the litigious nature of a particular competitor, or industry should be a key component when developing a patent strategy.

10.5 – Predictive models

Various models have been generated that use patent information to predict future behavior of one type or another. One of the most famous uses US maintenance data to predict how valuable patents are, compared to one another, and how valuable a recently granted patent might become. This

---

[96] http://www.rpxcorp.com
[97]

http://www.rpxcorp.com/siteFiles/SiteManager/0BF995E82CFF591EE80EFE8AC69259E7.pdf

method is patented itself, and the mechanism for building the model can be found in the abstract of US 7,657,476:

A method and system for valuing patent assets based on statistical survival analysis. An estimated value probability distribution curve is calculated for an identified group of patent assets using statistical analysis of PTO maintenance fee records. Expected valuations for individual patent assets are calculated based on a the value distribution curve and a comparative ranking or rating of individual patent assets relative to other patents in the group of identified patents.

Patents having the highest percentile rankings would be correlated to the high end of the value distribution curve. Conversely, patents having the lowest percentile rankings would be correlated to the low end of the value distribution curve. Advantageously, such approach brings an added level of discipline to the overall valuation process in that the sum of individual patent valuations for a given patent population cannot exceed the total aggregate estimated value of all such patents.

In this manner, fair and informative valuations can be provided based on the relative quality of the patent asset in question without need for comparative market data of other patents or patent portfolios, and without need for a demonstrated (or hypothetical) income streams for the patent in question. Estimated valuations are based simply on the allocation of a corresponding portion of the overall patent value "pie" as represented by each patents' relative ranking or position along a value distribution curve.

Theoretically, a model can also be built based on the likelihood that an assignee will allow a patent to be abandoned by not paying a maintenance fee at some point during the life of the document. By studying the past history of an organization it is possible to predict what their future behavior will likely be. Equipped with this information predictions can be made about when a technology of interest might enter the public domain and be used by others.

Additional examples of using patent information to generate predictive models for new patent documents can be found at:

Predictive Modeling of Patent Quality by Using Text Mining –
http://www.geocities.co.jp/Technopolis/5893/publication/IAMOT2010.pdf

Latent Graphical Models for Quantifying and Predicting Patent Quality –
http://users.cis.fiu.edu/~lzhen001/activities/KDD2011Program/docs/p1145.pdf

A Predictive Model for Patent Registration Time Using Survival Analysis –
http://www.naturalspublishing.com/files/published/988u9y98k5r14f.pdf

Chapter 11: Web Resources on Patent Landscaping Reports (PLRs)

A list of resources available on the web, on methods, LinkedIn Groups, and collections of PLRs can be found in the sections below. This list is not meant to be comprehensive, but provides some additional sources of information for individuals who would like to explore this topic further.

11.1 – Patent Analysis and Landscaping Methods

WIPO Patent Landscape Reports Website
http://www.wipo.int/patentscope/en/programs/patent_landscapes/

WIPO Patent Landscape Reports Compilation
http://www.wipo.int/patentscope/en/programs/patent_landscapes/published_reports.html

Bizint Cookbook of Reports and Visualizations (2015)
http://www.bizcharts.com/pdfs/BizInt_Cookbook_June2015.pdf

Guide Book for Practical Use of "Patent Map for Each Technology Field", Invention Research Institute, Japan Institute of Invention and Innovation, Japan Patent Office, Asia-Pacific Industrial Property Center, JIII (2000)

Patent statistics and patent mapping FAQ at the European Patent Office

OECD Patent Statistics Manual - http://www.oecd.org/science/inno/oecdpatentstatisticsmanual.htm

Patinformatics Blog - http://www.patinformatics.com/category/blog/

Patent mapping - Charles Boulakia - http://sciencecareers.sciencemag.org/career_development/previous_issues/articles/1190/patent_mapping

Patent Analysis, Mapping, and Visualization Tools - PIUG Wiki - http://wiki.piug.org/display/PIUG/Patent+Analysis,+Mapping,+and+Visualization+Tools

Patent visualisation - Wikipedia - http://en.wikipedia.org/wiki/Patent_visualisation

Analystology - http://analystology.com/wp/- a blog with several posts on patent landscaping

What Is Patent Landscaping? - eHow - http://www.ehow.com/facts_6199916_patent-landscaping_.html

How to prepare a Patent Landscape Report? - Steps for Patent Landscaping Analysis - YouTube - http://www.youtube.com/watch?v=Y74xZhV7UGI

Patent Landscape Analysis IP-Search, The Untapped Potential of Patent Data, IP-Search, Switzerland – https://www.ip-search.ch/fileadmin/user_upload/ip-search/e/umfeldanalyse_e.pdf

Intellogist Interview with Matt Luby, how to define a patent landscape – http://intellogist.wordpress.com/2011/08/23/interview-with-matthew-luby-how-to-define-a-patent-landscape/

What exactly is a patent landscape report and why is It useful? Mark Lloyd, Ambercite – http://www.ambercite.com/index.php/amber/entry/what-exactly-is-a-patent-landscape-and-why-is-it-useful

Patent Landscaping Studies: Their Use in Strategic Research Planning, Mark Pohl – http://www.licensinglaw.net/files/Patent_Landscaping_Study.pdf

Advanced Patent Landscaping with Multi-Series Charting - AcclaimIP - https://www.acclaimip.com/node/121

How To Do Patent Landscaping Using Free Databases (e.g. USPTO & ESP@CENET) - http://www.slideshare.net/MedicineAndHealth/how-to-do-patent-landscapingusing-free-databases-eg-uspto-espcenet-how-to-do-patent-landscapingusing-free-databases-eg-uspto-espcenet

What Is Patent Landscaping? - LegalZoom - http://info.legalzoom.com/patent-landscaping-20459.html

Database and tool reports - Intellogist Wiki - http://www.intellogist.com/wiki/Category:Intellogist_Reports

Patent analysis guide - https://sites.google.com/site/analyzingpatenttrends/using-patent-data

What is Network Patent Analysis? - Ambercite - http://www.ambercite.com/index.php/ambermap/what-is-network-patent-analysis

Patent Analysis - Nick Peters, Pinterest - http://pinterest.com/nbpeters/patent-analysis/

Statistical Patent Analysis Indicators as a Means of Determining Country Technological Specialisation - http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2247936

Advanced Patent Analysis Workbook - Crafitti Consulting - http://www.slideshare.net/crafitticonsulting/advanced-patent-analysis-work-book

Understanding your patent landscape – The Patent Lawyer – http://www.patentlawyermagazine.com/understanding-your-patent-landscape/?goback=%2Egde_44433_member_247448288%2Egde_44433_member_247736241

The Patent Analyst Blog - Patent analysis and patent program processes and best practices - http://thepatentanalyst.wordpress.com

11.2 – LinkedIn Groups on Patent Landscaping and Analysis

Patent Searching and Landscaping - http://www.linkedin.com/groups?home=&gid=110874&trk=anet_ug_hm

IP Intelligence - http://www.linkedin.com/groups?gid=124885&trk=myg_ugrp_ovr

Patinformatics - http://www.linkedin.com/groups?gid=2391676&trk=myg_ugrp_ovr

Patent Landscaping Innovations - http://www.linkedin.com/groups/Patent-Landscaping-Innovations-1883637?trk=myg_ugrp_ovr

Patent Analysis & Landscaping - http://www.linkedin.com/groups/Patent-Analysis-Landscaping-1794428?trk=myg_ugrp_ovr

Patent Informatics Group - http://www.linkedin.com/groups?gid=76745&trk=myg_ugrp_ovr

Patent Landscapes - http://www.linkedin.com/groups?gid=2846540&trk=myg_ugrp_ovr

Competitive Intelligence from Patent Analytics - http://www.linkedin.com/groups/Competitive-intelligence-from-patent-analytics-3891330?trk=myg_ugrp_ovr

11.3 – Available Collections of Landscape Reports

WIPO compilation of published Patent Landscape Reports – http://www.wipo.int/patentscope/en/programs/patent_landscapes/reports/

Patent Landscapes Page - Cambia - http://www.patentlens.net/daisy/patentlens/landscapes-tools.html

Patent Analysis of RFID Technology - Univ. of Arizona - http://ai.arizona.edu/mis480/sample_projects/2008%20Spring/Patent%20Analysis%20of%20RFID%20Technology/KM%20RFID%20Report.pdf

Patent Informatics - UK Patent Office - http://www.ipo.gov.uk/informatics-reports

White Papers – Griffith Hack – http://www.griffithhack.com.au/mediacentre-WhitePapersArchive

Technology Insight Reports – Gridlogics - http://www.patentinsightpro.com/index.html

ITTI Landscape Reports – Franklin Pierce Law Center – http://law.unh.edu/franklin-pierce-ip-center/international-technology-transfer-institute/projects

Alertas Tecnológicos (Portugese only) - INPI Brazil - http://www.inpi.gov.br/portal/artigo/alerta_tecnologico

Alertas Tecnológicas (Spanish only) - INAPI Chile - http://www.inapiproyecta.cl/605/w3-propertyvalue-1363.html

Boletines Tecnológicos (Spanish only) - SIC Colombia - http://www.sic.gov.co/boletines-tecnologicos.#tab1

Technology Patent Maps (Japanese only) – JPO Japan - http://www.jpo.go.jp/shiryou/s_sonota/tokumap.htm

Boletines Tecnológicos (Spanish only) - OEPM Spain - http://www.oepm.es/en/informacion_tecnologica/informacion_gratuita/boletines_de_vigilancia_tecnologica

Chapter 12: List of Literature Publications on Patent Analysis and Landscaping

A list of literature publications, including scholarly papers, and books, on the topic of PLRs can be found in the sections below. This list is not meant to be comprehensive, but provides some additional sources of information for individuals who would like to explore this topic further.

12.1 – Papers

UK Intellectual Property Office, The Patent Guide - A handbook for analysing and interpreting patent data, 2015

Gwilym Roberts, Landscaping – a practitioner view, *Queen Mary Journal of Intellectual Property*, Vol. 3 No. 4, pp. 313–317, 2013

Tania Bubela, E Richard Gold, Gregory D Graff, Daniel R Cahoy, Dianne Nicol, David Castle, Patent landscaping for life sciences innovation: toward consistent and transparent practices, *Nature Biotechnology* 31, 202–206, 2013

ER Gold, AM Baker, Evidence Based Policy: Understanding the Technology Landscape, *Journal of Law, Information and Science*, 2012

Damm, A., Technology and competitor mapping designed to support strategic business decisions, *World Patent Information*, Volume 34, Issue 2, June 2012, Pages 124-127

Daim, T.; Iskin, I.; Li, X.; Zielsdorff, C.; Bayraktaroglu, A.E.; Dereli, T.; Durmusoglu, A., Patent analysis of wind energy technology using the patent alert system, *World Patent Information*, Volume 34, Issue 1, March 2012, Pages 37-47

Nele Berthels, Gert Matthijs, Geertrui Van Overwalle, Impact of gene patents on diagnostic testing: a new patent landscaping method applied to spinocerebellar ataxia, *European Journal of Human Genetics* 19, 1114–1121, 2011

Ernst, H.; Omland, N., The Patent Asset Index - A new approach to benchmark patent portfolios, *World Patent Information*, Volume 33, Issue 1, March 2011, Pages 34-41

Bilyana P Georgieva, Jane M Love, Human induced pluripotent stem cells: a review of the US patent landscape, *Regenerative Medicine*, Vol. 5, No. 4, Pages 581-591, 2010

Moehrle, M.G., Walter, L., Bergmann, I., Bobe, S., Skrzipale, S., Patinformatics as a business process: A guideline through patent research tasks and tools, *World Patent Information*, Volume 32, Issue 4, December 2010, Pages 291-299

Bonino, D.; Ciaramella, A.; Corno, F., Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics, *World Patent Information*, Volume 32, Issue 1, March 2010, Pages 30-38

Yang, Y.Y., Akers, L., Yang, C.B., Klose, T., Pavlek, S., Enhancing patent landscape analysis with visualization output, *World Patent Information*, Volume 32, Issue 3, September 2010, Pages 203-220

Yang, Y.; Akers, L.; Klose, T.; Barcelon Yang, C., Text mining and visualization tools - Impressions of emerging capabilities, *World Patent Information*, Volume 30, Issue 4, December 2008, Pages 280-293

Karl Bergman, Gregory D. Graff, The global stem cell patent landscape: implications for efficient technology transfer and commercial development, *Nature Biotechnology* 25, 419 - 424, 2007

Scheu, M.; Veefkind, V.; Verbandt, Y.; Galan, E.M.; Absalom, R.; Forster, W., Mapping nanotechnology patents: The EPO approach, *World Patent Information*, Volume 28, Issue 3, September 2006, Pages 204-211

Kyle Jensen, Fiona Murray, Intellectual Property Landscape of the Human Genome, *Science*, 310, 239-40, 2005

Ernst, Holger, Patent information for strategic technology management, *World Patent Information*, 2003, 25 (3), 233-242.

Fattori M., Pedrazzi G., Turra R. Text mining applied to patent mapping: A practical business case, *World Patent Information*, 2003, 25 (4), 335-342

Trippe, Anthony J., Patinformatics: Task to Tools, *World Patent Information*, 2003, 25(3), 211-221

Trippe, Anthony J., Patinformatics: Identifying Haystacks from Space, *Searcher*, 2002, 10(9), 28

Ernst, Holger, Patent Applications and Subsequent Changes of Performance: Evidence from Time-Series Cross-Section Analysis on the Firm Level, *Research Policy*, 2001, vol. 30, 143

M Pohl, Patent Landscaping Studies: Their Use in Strategic Research Planning, *Pharmaceutical News*, 2002

Mogee, Mary, Breitzman, Anthony, Recent Applications for Patent Analysis, *Journal of Information Science*, 2002, 28(3), 187

Breitzman, Anthony, Thomas, Patrick, Using Patent Citation Analysis to Target/Value M&A Candidates, *Research Technology Management*, 2002, 45(5), 28

 Adams, Stephen, Pharmacia Corp.: Analysis of Patenting 1998-2002, *Expert Opinion in Therapeutic Patents*, 2002, 13(2), 223

Awaya, Kohei, Analysis Method for Patent Documents Utilizing References, *Proceedings of the IASTED International Conference: Applied Informatics*, 2002, pg 15

Trippe, Anthony J., A Comparison of Ideologies: Intellectually Assigned Co-Coding Clustering vs. ThemeScape Automatic Themematic Mapping, *Proceedings of the 2001 International Chemical Information Conference*, 2001, pg. 61

Aumann Y., Feldman R., Yehuda Y.B., Landau D., Liphstat O., & Schler Y. Circle Graphs: New Visualization Tools for Text-Mining, *Proceedings of the Principles of Data Mining and Knowledge Discovering Conference*, 1999, pg. 277

David F. Beck, Kevin W. Boyack, Oline H. Bray & Warren D. Diemens, "Landscapes, Games and Maps for Technology Planning", *ChemTech*, June 1999, p. 8-16

James A. Wise, "The Ecological Approach to Text Visualization", *Journal of the American Society for Information Science*, Vol. 50, No. 13, Nov. 1999, p. 1224

Marc Glodkowski, "Patent Monitoring for SMEs: the Lorraine Experience. Prospects for the Internet", *World Patent Information*, Vol.21, 1999, p. 31

Henry Small, "Visualizing Science by Citation Mapping", *Journal of the American Society for Information Science*, Vol. 50, No. 9, July 1999, p. 799

Qin He, "Knowledge Discovery Through Co-Word Analysis", *Library Trends*, Vol. 48, No. 1, Summer 1999, p. 133

Holger Ernst, "Patent Portfolios for Strategic R&D Planning", *Journal of Engineering and Technology Management*, Vol. 15, 1998, p. 279

Ronald N. Kostoff, Henry J. Eberhart and Darrell Ray Toothman, "Database Tomography for Technical Intelligence: A Roadmap of the Near-Earth Space Science and Technology Literature", *Information Processing & Management*, Vol. 34, No. 1, p. 69

Klaus K. Brockhoff, Holger Ernst & Eckhard Hundhausen, "Gains and Pains from Licensing – Patent-Portfolios as Strategic Weapons in the Cardiac Rhythm Management Industry", *Technovation*, Vol. 19, 1999, p. 605

Ronald N. Kostoff, Henry J. Eberhart and Darrell Ray Toothman, "Hypersonic and Supersonic Flow Roadmaps Using Bibliometrics and Database Tomography", *Journal of the American Society for Information Science*, Vol. 50, No. 5, April 1999, p. 427

Gert T. Preez & Carl W. I. Pistorius, "Technology Threat and Opportunity Assessment", *Technological Forecasting and Social Change*, Vol. 61, 1999, p. 215

Glen Hoetker, "Patterns in Patents", *Econtent*, Oct./Nov. 1999, p. 37

Karki M.M.S. Patent citation analysis: A policy analysis tool, *World Patent Information*, 1997, 19 (4), 269-272

Campbell R.S. Patent trends as a technological forecasting tool, *World Patent Information*, 1983, 5 (3), 137-143

Carpenter M.P., Narin F., Woolf P. Citation rates to technologically important patents, *World Patent Information*, 1981, 3 (4), 160-163

Carpenter M.P., Narin F. Validation study: Patent citations as indicators of science and foreign dependence, *World Patent Information*, 1983, 5 (3), 180-185

12.2 – Books

Mihai Lupu, Katja Mayer, John Tait, Anthony Trippe, Editors, "Current Challenges in Patent Information Retrieval", Pub. Springer, Berlin Heidelberg, 2011

Alan Porter, Scott Cunningham, "Tech Mining: Exploiting New Technologies for Competitive Advantage", Pub. Wiley, Hoboken, 2005

John Wiley & Sons Inc. Technical Insights, "Competitive Technical Intelligence: Easy Steps to Track Your Business Rivals' R&D Efforts", *Report 251*, Pub. John Wiley & Sons Inc. New York, 1999

Mathias Coburn, "Competitive Technical Intelligence: A Guide to Design, Analysis and Action", Pub. American Chemical Society Washington, DC & Oxford University Press, New York & Oxford, 1999

W. Bradford Ashton, Richard A. Klavans, Editors, "Keeping Abreast of Science and Technology: Technical Intelligence for Business", Pub. Battelle Press, Columbus, 1997

Chapter 13: Annex – Example of Terms of Reference for the Preparation of a Patent Landscape Report

Terms of Reference
-
PROVISION OF A
PATENT LANDSCAPE REPORT ON TECHNOLOGIES RELATED TO
PALM OIL PRODUCTION AND WASTE TREATMENT AND EXPLOITATION

Project Background

The present Patent Landscape Report (PLR) is prepared in the context of WIPO's patent landscaping activities. These activities involve the preparation of PLRs on topics which are relevant for developing countries, the promotion of patent analytics and the advanced exploitation of patent information in developing countries.

The present PLR on palm oil related technologies is prepared in collaboration with the Intellectual Property Organization of Malaysia (MyIPO) and the Palm Oil Board of Malaysia (MPOB). The palm oil industry is an essential part of Malaysia's economy and plays an important role also for other economies in South-East Asia, Latin America and Africa. The global patent landscape report in the area of palm oil technologies will be paired with the patenting applications filed in Malaysia, with the support of the Malaysian Intellectual Property Office (MyIPO).

Following the completion of the patent landscape report, a regional seminar is envisaged to be organized in Malaysia with the participation of public and private sector institutions from Malaysia and other countries of the region (e.g. Philippines, Indonesia) working in the area of palm oil to inform them about the IP perspective of palm oil, share the results of the report with them, but also explore potential research and technology transfer possibilities. The national stakeholders in Malaysia which should be involved in the seminar will be identified with the help of MyIPO and the Malaysian Palm Oil Board. The report is intended to be used for awareness-raising on and use of patent information and patent analytics as tool for business intelligence and strategic planning of development and investment in research.

A separate patent landscape report on palm oil covering the national patenting activity in the Philippines is planned to be carried out by the Philippines Intellectual Property Office (IPOPHL). This data will be compared to the global patent landscape report and can also serve stakeholders in the field of palm oil in Malaysia to identify potential partners.

1. Purpose of the RFQ/RFP

I.   SCOPE OF THE PROJECT

Palm oil is the highest-yielding vegetable oil crop, and palm oil produced from its seeds is an important and sustainable source both for food and biofuel, and other products. This report aims to investigate the patenting activity related to technologies in the following areas: a) Production of palm oil, i.e. growing and harvesting the fruit, processing of the fruit, crude palm oil, refining of the oils. Both, technologies related to palm oil derived from the mesocarp of the fruit and palm kernel oil derived from the fruit kernel will be covered. b) Treatment of waste from palm oil production, extraction of phytonutrients, utilization of palm bio-mass and/or by-products generated by the palm oil for food and beverage, cosmetics, biofuel and other areas of industry.

The landscape report will research inventions disclosed in patent publications (for the purpose of the present ToR, the term "patent" includes both patents and utility models). Some non-patent literature should also be researched and analyzed.

The report should describe patterns and trends of patenting activities by including a standard statistical analysis of the search results, e.g. the key commercial and institutional innovators and patent right owners, their collaborations, the geography of the origin of their innovations (priority countries, offices of first filing) and the geographical extent of protection they seek (offices of second filing), the patent activity over time, distribution of patenting activity by category of technology etc. Patenting activity of research institutes and universities, as well as collaborations among them and between them and private sector entities should be analyzed and highlighted separately. In particular, the terminology used for writing the report and categorizing the technologies should be consistent with the respective terminology used by the institutions active in the field.

The final patent dataset will be made available in Excel format, while each patent family will be hyperlinked to the full publication on ESPACENET, in order to provide access of the readers to the related family information and available legal status of family members.

The report will not focus on aspects of validity of protection or freedom-to-operate, i.e. it will not investigate whether a patent that has been granted for a particular patent application has entered into force or is still valid. Claims need only be used as general guidance as to what types of subject matter is claimed as invention. However, in order to assess coarsely the level of innovation of applications, it will be researched, for each patent family, whether the family comprises at least one publication of a granted patent (based on the publication kind codes of patent family members).

A further, important objective of the report is to serve as an instructive example of how specific technical subject matter can be researched in existing patent databases. A comprehensive explanation of the applied search strategy, including the challenges and limitations of the search, along with well documented and thereby repeatable search queries, will be essential components of the report (see 3.4).

2.   List of services

- Preparation and delivery of a high-quality patent landscape report in Word format, in English language, following the formatting guidelines to be provided by WIPO to the contractor
- Inclusion of a detailed description of methodology with mention of all used tools and search

strings used, along with the advantages of each of each of the approaches used

- Provision of the intermediary and final datasets, including the search results in Excel format (or a different one, should it be considered more appropriate), hyperlinked to the full patent documents on ESPACENET and including technology tagging and further categorizations, as deemed necessary and discussed with the WIPO Secretariat
- Provision of all supporting data of the provided graphs and statistics
- Use of appropriate and as far as possible and sensible, different for each analysis visualization tools, showcasing the various patent analysis tools to the patent information users
- Provision of support to the Malaysian Patent Office for the required patent searches in their national patent database and coordination for the use of comparable search strings
- Inclusion of the Malaysian search results in the patent landscape report and comparison to the global landscape report
- Provision of inputs (mainly on key findings of the report) for the preparation of an Infographic

3. <u>Deliverables</u>

The search should identify all relevant patent families originating from jurisdictions whose publications are part of the PCT minimum documentation[98], and irrespective of whether a patent was granted at any patent office for any family member.

The successful candidate will be expected to deliver a patent landscape report in English language, in a format to be defined by WIPO, taking into account the above mentioned objectives and including at least the following sections:

3.1   Extended executive summary

This section of the report will consist of an executive summary of 3-4 pages, phrased in a manner to be addressed to policy and decision makers. This section should highlight the findings, possibly correlate them with other findings or additional data in a comprehensive manner to facilitate the relevant policy discussions and the use of the data provided through the report. The successful candidate will also provide the data and necessary support to WIPO Secretariat and the Infographic provider to feed into an Infographic serving dissemination purposes.

3.2   Introduction section

This section of the report should briefly explain the objective and scope of the report, and the framework within which it is established.

3.3   Technology section

This section of the report should describe the different categories of technologies related to the production of palm oil, treatment of waste from palm oil production and use of such waste for

---

[98] http://www.wipo.int/export/sites/www/standards/en/pdf/04-01-01.pdf

cosmetic, pharmaceutical, biofuel and other industry. Each category should be illustrated with selected patent applications. This section should therefore elaborate on the indexing/tagging that is applied to the patent families in the attached database (Excel sheet).

3.4   Description of search methodology and limitations

This section of the report is important in view of the objectives of the project to develop tools for access to and exploitation of patent information. The report should explain the search methodology in detail, e.g. how the search queries were developed and refined (search strategy narrative should be part of the report, while search queries are to be included in an Annex), how classification symbols were identified and to what extent they were useful for patent search in this area of technology, how citations were exploited, which databases were used, etc.

This section should also discuss difficulties, limitations and acceptable tolerances with respect to recall and precision. In particular, the effect of including different patent classification systems like CPC and FI in order to improve recall in comparison to purely IPC based searches should be presented. Other strategies to improve recall should also be explained, e.g. by supplementary or combined keyword searches or citation analysis. Numerical evidence should be included.

The report should include explanations on limitations of availability of patent information from developing countries and countries with economies in transition, e.g., of patent family data coverage of the databases used for search. Furthermore, particularities or difficulties related to searching this particular subject matter should be illustrated along with the approach followed to overcome the related challenges.

The report should include in this part a description on the family reduction method. Both that and the statistical analysis should be based on the concept of simple patent families, or similar family concepts like Fampat or Thomson DWPI families. The family reduction should not be based on the concept of extended (Inpadoc) families. It would however be advantageous if the subscriber would include information or analysis regarding the extended families, e.g. whether two simple families belong to the same extended family, or how different the size of the respective extended family is in comparison of the single family.

3.5   Analysis of patenting activity

The report should include a statistical analysis of the patenting activity, including appropriate visualization, according to at least the following aspects:

(1)   Number of simple patent families (additional information on extended or INPADOC families would be advantageous) and patent publications (i.e. including all family members), in total and per (earliest) priority year; average patent family size; size of the largest patent family.

(2)   Percentage of families comprising at least one publication of a patent grant (to be determined only according to the kind codes of publications; i.e. no legal status data need to be researched, e.g. in order to determine if the grant was revoked after an opposition), in total and per (earliest) priority year.

(3)   Percentage of patent families with at least one PCT family member, in total and per (earliest) priority year.

(4)   Distribution over "priority countries", i.e. the number of families filed per earliest priority filing office, i.e. Office of First Filing (OFF), including the International Bureau of WIPO as a PCT receiving office, in total and per (earliest) priority year. The report is expected to include an analysis of patent families originating from Malaysia, Philippines, Indonesia, Brasil, Colombia, but also other countries from Asia, Latin America and Africa due to the importance for these countries and regions for palm oil industry/market and the potential for technology transfer and partnerships.

The report will be complemented by national patent filing data related to palm oil in Malaysia, to be provided by MyIPO, and is expected to compare the national Malaysian activity with the PCT applications with OFF or OSF in Malaysia. The contractor is expected to assist MyIPO in assessing this information and work with them in the definition of the appropriate search strategy and the use of specific search queries which will ensure comparable search results to the provided by the contractor global and regional patent landscape analysis. The contractor is also expected to incorporate the Malaysian national data results into the report, analyze them and compare them to the global patenting activity results.

(5)   Geographical distribution of extensions, i.e. of patent family members filed with any Office of Second Filing (OSF) after the priority filing with the OFF, in total and per (earliest) priority year. The distribution of OSF should be determined such that the OFF is excluded, i.e. second filings in the country of OFF that are derived from the priority document should not be counted. Each OSF should be counted only once.

(6)   Geographical distribution of applicant nationalities (for families having several applicants each nationality is counted separately, i.e. the statistics should not be limited to the nationality of the applicant listed first).

(7)   Geographical distribution of inventor nationalities (for families having several inventors, each nationality is counted separately, i.e. the statistics should not be limited to the nationality of the inventor listed first).

(8)   Most active applicants (top 20); breakdown by the sector they represent (industry, academia/research/public sector, individuals); their geographical distribution; their preferred OFF; their use of OSF for extensions; their activity over time; relevant corporate trees, i.e. close links between distinct applicants; the categories of technologies for which they have filed patent applications.

(9) Most active inventors (top 20); their affiliation with industry, academia/research/public sector; their nationality, their activity over time; the categories of e-waste recycling technologies for which they have filed patent applications.

(10) Collaboration networks among applicants and among inventors.

(11) Citation frequency, i.e. patents that have been cited frequently and therefore possibly cover key innovations.

The analysis should be done in total (for all relevant patents) and separately, if significant, for the different categories of components or technologies.

The successful candidate will be invited to include further analysis and visualization that he discovers as interesting or instructive during the analysis stage. Suggestions in the proposals submitted in

response to the RFP are welcome.

3.6   Annexes

(1)   A table (Excel sheet) which sorting (e.g. for dates) and filtering functionalities that includes all relevant simple patent families. Each simple family should be identified by the publication number of a single published family member, preferably the earliest published PCT publication number, if available. If the family has no PCT family member, the family member given in the table should be one that is published in English language, preferably the earliest priority document defining the family.

Each of these publication numbers identifying a family should be hyperlinked to ESPACENET in order to readily allow the viewing of the full publication and the related INPADOC family and available legal status of family members. If ESPACENET does not include the publication, the publication number should be linked to another free-of-charge patent database that includes the publication (preferably the database hosted by the publishing office, if available).

The table should further include in separate additional columns for each family (derived from the publication representing the family):

- title;
- at least the first claim;
- applicant name(s);
- an indication as to whether at least one applicant is a private sector and/or a public sector entity (names of individuals are considered neither private nor public sector entity);
- inventor name(s);
- priority information:
        - earliest priority date (or application date in case no priority is claimed),
        - priority country of earliest priority (i.e. OFF)
        - further priority dates with respective priority country;
- the size of the simple family and the size of the related Inpadoc family;
- publication numbers of all members of the simple family (not to be linked);
- the country codes of all offices of second filing (OSF, each OSF indicated only once, OFF excluded);
- number of OSF (each OSF counted once);
- an indication as to whether the family comprises at least one publication of a patent grant (to be determined only according the kind codes of publications; i.e. no legal status data need to be researched, e.g. in order to determine if the grant was revoked after an opposition).
- all IPC symbols attributed to the document selected for identifying the family.

Each family should additionally be indexed/tagged according to the technology and use categories to which they belong (categories as described in the report body; see Section 3.3 above).

Where possible, columns should be sortable in alphabetical manner or according to date or number.

All dates have to be formatted as YYYY.MM.DD.

(2) All search queries and all cleaning methods that were used to establish the final set of results; they should be presented in such a way that they can be reused in the respective databases for updating the search results.

4. Input/support from WIPO; WIPO support to the contractor

The contractor will be expected to undertake the work autonomously, while working closely with selected staff focal points in WIPO. Whenever possible, WIPO will undertake to assist in providing some already existing studies, data and information which is required by the contractor. The contractor will report to the IP Information Officer of the Project. WIPO will provide the service with copies of its existing materials directly related to the topic, including both final and draft documents, and coordinate in the communication with the Malaysian IP Office. WIPO will provide feedback throughout the project delivery. In situations where neither WIPO nor the contractor own any original material or compiled data on a particular subject, external public or proprietary databases need to be researched and accessed by the contractor.

5. Intellectual Property

The patent landscape report, as well as all intermediate drafts and data the Contractor has developed under the contract shall be the sole property of WIPO. The Contractor acknowledges and agrees that such products and other materials constitute 'works made for hire' for WIPO.

6. Travel requirements

Travel may be required for presentation of the report's findings, in Geneva or in Malaysia. In that case, travel costs will be covered by WIPO.

II. TIME FRAMES/DURATION

7. The delivery of the patent landscape report will be according the following timeline:

(1) **Phase I**: within **four (4) weeks** upon receipt of notification on contract award.

(2) **Phase II:** within **four (4) weeks** upon receipt of WIPO comments of Phase I.

(3) **Phase III:**   within **two (2) weeks** upon receipt of WIPO comments of Phase II.

The specific time frame will be finalized together with the contractor.

III. KEY PERFORMANCE INDICATORS (KPI) AND QUALITY CONTROL

8.

- Quality of and response time to the communication with WIPO
- Responsiveness and level of assistance provided to the Malaysian Patent Office for conducting the patent search using their national patent database
- Timeliness of the submission of the deliverables
- Quality of the provided deliverables in comparison to previous WIPO Patent Landscape Reports (http://www.wipo.int/patentscope/en/programs/patent_landscapes/) and efforts to achieve the best reliable results
- Degree of implementation of WIPO's feedback in the deliverables
- Flexibility shown throughout the project in cases of need of adjustment of certain requirements/analysis types
- Quality of formatting and visualizations provided
- Project Stakeholder Satisfaction