Proposal for 2001 NUS Equipment Fund for Research

Title: **NUS Patent Database**
Date: **18 January 2002**
Contact: **Kwanghui Lim, Asst Professor, NUS Business School.**
http://kwanghui.com

(i) Name of equipment: NUS Patent Database.

This includes a computer, tape drive and media (containing US and European patents granted and bibliographic data on international patents).

I propose to start with the major patent databases (US, Europe), keeping in mind the possibility of adding country-specific databases in the future (e.g. Singapore, Japanese, German patents).[1] I already have the NBER citation database, which can be included free of charge.

(ii) Cost

(iii) Usage/Purpose

A patent database will allow NUS faculty and staff to search and analyze patents from a variety of sources. Patent data has been used in a number of ways by academic researchers:

- Patents are used to measure the innovation output of firms, research institutes, universities and countries.
- Patent-to-patent citations are used to trace flows of knowledge, or to measure a firm's technological intensity.
- Patent-to-science citations are used to trace the link between science and invention, as well as to measure a firm's science-intensity.
- Patent claims are used by legal scholars studying intellectual property issues.
- The number of patents awarded in various technology classes has been used to create technology profiles of companies (e.g. Narin, Noma and Perry, 1987).

A patent database is a worthwhile investment for NUS. It offers a valuable resource to scholars in a wide range of areas, including Management (strategy, management of technology, finance, technopreneurship), Law (e.g. intellectual property), Economics (R&D and productivity) and Public Policy (national innovation systems).

Just as with financial and macroeconomic data, the use of patent data has many limitations (see Griliches, 1990). Despite these imperfections, in many instances it is the best available public data on innovation. When used by a careful researcher, patent data can result in important studies, including:

- Work by Griliches and his colleagues to measure the impact of R&D on productivity (see review by Griliches, 1984)
- An important paper on the factors that affect the commercialization of new ideas by high-technology entrepreneurs (Shane, 2001).
- An influential study on national innovation systems (Stern, Furman and Porter, 2000)
- A well-cited paper showing that knowledge spillovers are geographically localized (Jaffe, Trajtenberg and Henderson, 1993).
- A path-breaking study on how social networks affect knowledge transfer (Ahuja, 2000)
- An interesting study showing how firms engage in “strategic patenting” (Hall and Ziedonis, 2001)
- A novel paper exploring the recombination of technologies as a complex adaptive system (Fleming and Sorenson, 2001)
- An exciting new study that relates patenting to the stock-price valuations (Hall, Jaffe, Trajetenberg 2000);
- An interesting study on the commercialization of national laboratory technology (Jaffe and Lerner, 2001)

---

[1] At present, it is too costly to purchase the Japanese and European Patent Application (EP-A) databases.

Patent data is being used by an increasing number of researchers across disciplines, many of whom do not focus exclusively on innovation. Moreover, scholars are beginning to complement the use of patents with other sources of data, thereby overcoming some of its limitations. For example, Fiona Murray (a sociologist at MIT) is using in-depth interviews to trace the social structures and networks around an important patent-publication pair in biotechnology. Scott Shane (2001) uses patent data in conjunction with extensive empirical fieldwork to study the commercialization of university technology. Lim (2000) combines patent data with data on publications in basic and applied journals to study the relationship between science and technology in two industries.

Synergies exist between the patent database and our existing financial database because researchers can combine data from both sources to explore interesting research questions. For example, Wong Poh Kam (Director for the Center for Entrepreneurship) is currently working on a project to combine data on patenting by Singapore-based companies with data on their financial performance from Datapool's "Singapore 1000" listing.

In addition to the research benefits, a patent database is also a useful resource for teaching courses on innovation management, new product development, intellectual property management and law, etc.

(iv) Immediate Research Projects

- Several projects I am working on require the use of patent data, including a study on Intel vs IBM with Hank Chesbrough at Harvard and a project on commercializing new technologies with Dave Hsu at MIT.
- Wong Poh Kam has studied patenting trends in Singapore and other NIEs. He plans to use this patent database in his future research. For example, he will be working with Scott Shane at U of Maryland on university patenting trend (see Shane, Management Science, 2001).
- Pasha and Chi Nien are already using patent data in their work. The existence of an NUS database would bring them many long-term benefits.
- Susanna Leong and I are at the initial stages of a project to explore how changes in patent regimes affects patenting behavior.

(v) Users

Some immediate users are listed above. However, there are potentially many other researchers in NUS who could benefit from the creation of a patent database. As described above, such a database could benefit faculty members and research students in the Business, Arts and Law faculties.

Just as with financial databases and the Science Citation Index in the past, the number of users will probably increase once people begin to realize how useful patent data can be.

(vi) Availability of alternatives at NUS

Preliminary inquiries indicate that a similar resource is not yet available at NUS.

The NUS library apparently has access to patent data that you can print out, but not a full-text database that is exportable into a form easily analyzed.

Our Center for Entrepreneurship doesn't have this resource, which is why Wong Poh Kam and I have long discussed it as an important investment to consider.

(vii) Other Justification

Patent searches can also be done via online services such as Delphion, Micropat and Derwent. The table below shows the advantages and disadvantages of online services vis-à-vis standalone databases. Essentially, online services provide convenience, ease of use and automatic updates. However, setting up our own database offers much higher speed, the opportunity to do far more sophisticated searches and analysis, and unlimited access to the data (including exporting data fields not available online).

While the annual subscription for online services is relatively low, additional costs are involved. For example, they charge for the number of full-text patents downloaded, and this can quickly add up for studies that require large datasets.[2]

Investing in a patent databases should be viewed as a complement rather than a substitute to having subscriptions to online services. Researchers planning to perform narrow, targeted searches on a specific firm or technology should definitely use online sources. However, the power and flexibility of owning a database is necessary for heavy-duty research that requires sophisticated search algorithms, large datasets, and custom programs to analyze the data. Papers such as Fleming-Sorenson (2001), Hall-Ziedonis (2001) and Jaffe (1986) would not be possible without high-speed access to large chunks of patents. Likewise, scholars that want to do country-level or industry-level studies on innovation would be hampered without easy access to large patent datasets (imagine the cost and effort needed to download all pharmaceutical industry patents from Delphion!) For this reason, many other leading academic institutions have purchased their own patent databases, in addition to relying on online services[3] (MIT/Harvard run the NBER Data Center; scholars at Wharton and U.C. Berkeley purchased the entire CDROM collection from Micropatent; Harvard has a copy of the entire USPTO database).

Standalone versus Online Patent Databases

	Standalone, in-house Patent Database	**Online Patent Search Services**
Search Engine	Unlimited search capabilities, since the data can be exported and custom programs written to search the database if needed. Fast searching on local machine.	Can only handle relatively simple search strategies. Also, some search engines timeout if the search strategy is too complex. Web-based search engines are a shared resource and can be very slow during busy periods.[4] Interface is "user-friendly".
Search Results	Unlimited number of search results can be obtained. No need to split a large search into many small, tedious ones.	The number of search "hits" is limited. For example, Delphion allows each search to return only 500 records. Micropat limits it to 2500. Thus, a search for all the patents by a large company like IBM will have to be broken down into many small searches to be performed manually
Exporting the data	Ability to export any/all of the data to other formats, regardless of size.	Only a limited number of datapoints may be downloaded per search. Download over the web is particularly slow especially for large datasets.
Database Coverage	At present, I am proposing to obtain only the fulltext US and European patents Note: the US database is text-only, which is great for researchers wanting to analyze textual data, but not for those needing to fully evaluate a particular invention (engineers, scientists, inventors). The European database contains full images.	Broader number of datasets included Patent images are available for download at a fee.
Value-added services	Not included. However, sophisticated users can create their own spreadsheets and programs. This gives rise to the ability to create *new* measures and constructs not	Many online sources offer value-added services, such as graphing capabilities and citation-based measures of technology intensity.

[2] For example, Delphion charges US$5 per 500 records. Derwent charges US$8 each time you perform a search, US$0.5 per matching record found, and US$4.5 per full record viewed. Micropatent charges US$4.95 per downloaded US/European patent, $8.95 per Japanese patent and US$1.5 per downloaded front page.

[3] This is based on personal contact with researchers at the institutions mentioned.

[4] I have tested the Micropat.com and Delphion.com services and found this to be true.

	offered by online services.	However, these services cost extra.[5]

References

- Ahuja, G., 2000, Collaboration networks, structural holes and innovation: a longitudinal study, *Administrative Science Quarterly* 45, 425-455.
- Fleming, L., Sorenson, O., 2001, Technology as a complex adaptive system: evidence from patent data, *Research Policy* 30, 1019-1039.
- Griliches, Zvi, 1984, R & D, patents, and productivity, Chicago: University of Chicago Press.
- Griliches, Z., 1990, Patent statistics as economic indicators: a survey, *Journal of Economic Literature* 18 (4), 1661-1707.
- Hall, B.H., Jaffe, A., Trajtenberg, M., 2000, Market value and patent citations: A first look, *NBER working paper* W7741, Cambridge, MA.
- Hall, B.H., Ziedonis, R.H., 2001, The patent paradox revisited: an empirical study of patenting in the US semiconductor industry, 1980-1995, *Rand Journal of Economics* 32(1), 101-128.
- Jaffe, A., 1986, Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value, *American Economic Review* 76(5): 984-1001.
- Jaffe, A., Henderson, R., Trajtenberg, M. (1993), Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics* 434:578-598.
- Jaffe, A., Lerner, J., 2001, Reinventing public R&D: patent policy and the commercialization of national laboratory technologies, *Rand Journal of Economics* 32(1): 167-198.
- Lim, K., 2000, Basic Research, Applied Research, and Innovation in the Semiconductor and Pharmaceutical Industries, Doctoral Dissertation, MIT Sloan School.
- Narin, F., Noma, E., Perry, R., 1987, Patents as indicators of corporate technological strength, *Research Policy* 16, 143-155.
- Shane, S., 2001, Technology regimes and new firm formation, *Management Science* 47(9): 1173-1190.
- Stern, S., Furman, J., and Porter, M. (2000), The determinants of national innovative capacity, NBER working paper 7876 (forthcoming in *Research Policy*).
- Trajtenberg, M., 1990, A penny for your quotes: patent citations and the value of innovations, *RAND Journal of Economics* 21 (1), 172-187.

Appendix: Description of databases

USPTO Full-Text/APS Retrospective 1971-2000 (EIP-1095P-DL)
The Retrospective File from 1971-1999 inclusive contains the full text of each patent issued. This file is formatted in accordance with the specification required by the text and retrieval software which resides on the PTO's automated patent search systems. Tables and "in-line" mathematical equations are present, where appropriate, and appear as text data. Chemical structures are not present, but their location is indicated by a structure call-out. Includes patent number, series code and application number, type of patent, filing date, title, assignee name at time of issue, attorney, agent or firm/legal representative, related U.S. Patent documents, classification information, field of search, U.S. and foreign references, priority data, abstract, specification, and claims. (source: USPTO catalog).

Espace EP-B European Patents Awarded (website)
This CD-ROM series contains bibliographic data, full text and embedded facsimile images of granted European patents. It includes searchable titles and claims in English, French and German. The complete description could be searched in the filed language of the application. References to the PCT publication number for the Euro-PCT documents are available. A pilot index gives reference to the documents published in the previous 5 years.

Searchable titles and claims in English, French and German. Searchable bibliographic data and description as filed.

Coverage of the overall series: From the first granted European patent to date.

Update frequency: Weekly. The discs are distributed every Wednesday (on the day of publication) and normally reach the user the following day.

[5] Micropatent Online offers citation reports by CHI Research, costing $6/report or $3920/user/year. A graphing feature costs $9/user/day or $540/user/year.

Searchable fields
Following 28 fields are available for searching, alone or combined:
AD Application date, AN Application number, BD B publication date, CI Patent citation, CLE English claims, CLF French claims , CLG German claims , DAN Divisional application data , DC Correction date , DEE English description , DEF French description , DEG German description , DP A publication date , DS Designated state , EP Publication number , ET English title , FT French title , GT German title , FULLTEXT Full text index , IC All classifications , IN Inventor , KI Document kind , MC Main classification , NP Priority number , PA Applicant , PAN Parent application data , PD Priority date , RP Representative

Special features
The data and images are fully downloadable. Bibliographic data in SGML format, images in TIFF format. Using the MIMOSA software version 4, the patent specifications can be downloaded or printed.
To reproduce the originally published paper document, the patent documents are also stored in PDF format, which could be printed by the MIMOSA software in using Acrobat® Reader.
Nucleotide or amino acid sequence listings are stored according to WIPO Standard ST.25 in separate ASCII textfiles with the extension TXT.

GlobalPat: Database of the First Pages of Worldwide Patents (website)

GLOBALPat is derived from the First Page Data Base (FPDB). The FPDB is an English-language collection representing nearly all of the world's patent literature. Documents are presented as bibliographic text, including title and abstract, and, where appropriate, the image of a representative drawing. The Trilateral Offices (the U.S. Patent and Trademark Office, the European Patent Office and the Japanese Patent Office) jointly financed the preparation of the database and the translation of all non-English abstracts.

The EP/US part of the FPDB contains a single, representative member of almost every patent family (96% coverage for the backfile) published by the United States (US), the European Patent Office (EP), France (FR), Germany (DE), United Kingdom (GB), Switzerland (CH), or the World Intellectual Property Organization (WO). Patent families occur when a patent applicant submits an application for the same invention in several countries, resulting in several patents based on the same priority application. Because the Trilateral Offices started using computer-based photo composition in the 1970s, the FPDB includes most documents from that time forward (varies with country and year).

Records in the FPDB include a bibliographic summary, abstract, and a representative drawing (where appropriate) as they appear on the first pages of the source documents. A small number of records in the FPDB may contain information from more than one family member. Drawings are combined with text into ISO-8879 Standard Generalized Markup Language (SGML) documents. The Document Type Definition (DTD) for GLOBALPat is based on WIPO Standard ST.32. GLOBALPat uses the MIMOSA software which was developed by the Trilateral Offices for SGML-encoded documents.

The Japanese Patent Office publish their part of the FPDB in Patent Abstracts of Japan (PAJ) on CD-ROM (available from the Japanese Patent Office), using the MIMOSA software and SGML. GLOBALPat contains data from the United States, World Intellectual Property Organization, European Patent Office, United Kingdom, German, France, and Switzerland. The GLOBALPat backfile covers 1971-1996 and is arranged in 69 technology groups based on the International Patent Classification (IPC) system. The front file (1997+) is arranged by issuing country and document number on monthly issues

Languages available: Searchable data in English throughout.

Coverage of the overall series: First pages of patent applcations from 1971 to date for the above-mentioned offices.
Volume 1-116: 1971 to 1996
Volume 117 - : 1997

Update frequency :The backfile (discs 1 to 116) have been prodcued in 1997. The frontfile starting on 1 January 1997 is produced on a monthly basis.

Searchable fields
Following 15 fields are available for searching, alone or combined:
AB: Abstract , AP: Application ID (Application number) , BI: Basic index , EC: EPO classification , FA: Family data, IC: International Patent Classification , ICO: ICO classification , IDT: IDT classification , IN: Inventor , PA: Applicant , PD: Publication date , PNR: Puplication ID (Publication number) , PR: Priority data , Presence: Presence of data , TI: Title