

Committee on Development and Intellectual Property (CDIP)

Sixth Session
Geneva, November 22-26, 2010

Study Paper regarding Recommendation 8

prepared by the Secretariat of the World Intellectual Property Organization

TABLE OF CONTENTS

1.	Introduction	5
	(a) Scope of the Study Paper	5
	(b) Format of the Study Paper	6
2.	Review of patent database services	7
	(a) Overview.....	7
	(b) Database content sets.....	9
	(i) INPADOCDB bibliographic data	9
	(ii) INPADOC legal status data	9
	(iii) Derwent World Patent Index.....	9
	(c) Fee-based commercial databases	10
	(i) Orbit.com	10
	(ii) PatBase/PatBase Express.....	11
	(iii) STN.....	12
	(iv) Thomson Innovation	13
	(v) TotalPatent.....	14
	(vi) WIPS Global	15
	(d) Free-of-charge commercial databases.....	16
	(i) Google Patents	16
	(ii) Patent Lens.....	17
	(e) Databases provided by WIPO and national and regional offices	17
	(i) World Intellectual Property Organization	17
	(ii) Australia	18
	(iii) Canada	19
	(iv) China.....	19
	(v) Germany	19
	(vi) Hong Kong, Special Administrative Region of China	20
	(vii) India	20
	(viii) Israel	21
	(ix) Japan	21
	(x) New Zealand.....	22
	(xi) Republic of Korea	22
	(xii) United Kingdom	23
	(xiii) United States of America	23
	(xiv) Eurasian Patent Organization.....	24
	(xv) European Patent Office	25
3.	Review of non-patent literature database services	27
	(a) Introduction	27
	(b) Structure and methodology	28

(c)	Database content sets.....	29
(i)	Inspec	29
(ii)	IP.com Prior Art Database	29
(iii)	MEDLINE	30
(iv)	Embase.....	30
(d)	Free-of-charge search services.....	30
(i)	Google Scholar	30
(ii)	Scirus	32
(iii)	Entrez.....	32
(iv)	PubChem	33
(v)	PubMed.....	33
(vi)	PubMed Central.....	34
(e)	Fee-based search services	34
(i)	Dialog.....	34
(ii)	Scopus	35
(iii)	STN.....	36
(iv)	Thomson Innovation	38
(f)	Free-of-charge journal databases	39
(i)	Directory of Open Access Journals	39
(ii)	SciELO.....	39
(g)	Fee-based journal databases	40
(i)	American Chemical Society.....	40
(ii)	IEEE Xplore	41
(iii)	ScienceDirect.....	41
(iv)	SpringerLink.....	42
(v)	Wiley InterScience	42
(h)	Access to Research for Development and Innovation (aRD _i) and related services	43
4.	Analysis.....	45
(a)	Introduction	45
(b)	Needs analysis	45
(c)	User profiles and needs.....	45
(i)	Definition and methodology	45
(ii)	Conclusions from statistics on patent activity	45
(iii)	Conclusions from responses to questionnaire by IPOs of WIPO Member States ..	45
(iv)	User needs.....	50
(d)	Comparison between free-of-charge and commercial databases	51
(i)	General	51
(ii)	Coverage	51
(iii)	Tools	51

(e)	Issues and Recommendations	52
(i)	Database selection	52
(ii)	Access to commercial patent databases	53
(iii)	Enhancing PATENTSCOPE®	53
(iv)	WIPO Patent Information Services for Developing Countries (WPIS)	53
(v)	Analysis of Training Gaps	54

1. INTRODUCTION

(a) Scope of the Study Paper

1. During the first session of the Committee on Development and Intellectual Property (CDIP), held from March 3 to 7, 2008, Member States of the World Intellectual Property Organization (WIPO) approved the working document CDIP/1/3¹ containing 45 recommendations for the implementation of WIPO's Development Agenda. These recommendations were subsequently adopted by the General Assemblies of WIPO.²
2. The present Study Paper constitutes part of the implementation of Recommendation 8 of the Development Agenda under which the Member States:

“Request WIPO to develop agreements with research institutions and with private enterprises with a view to facilitating the national offices of developing countries, especially LDCs, as well as their regional and sub-regional intellectual property organizations to access specialized databases for the purposes of patent searches.”³
3. During its second session in July 2008, the CDIP approved the document CDIP/2/INF/3 on “Terms of Reference for a Study Paper regarding Recommendation 8” defining the scope of the paper and, in particular, focusing on: a needs analysis; a review of specialized patent databases; a review of specialized non-patent literature (NPL) databases; a comparative analysis between the added value of commercial databases with respect to free-of-charge databases; and further possible issues and recommendations.
4. The project and financing to implement Recommendation 8, which included the drafting of the Study Paper, were approved at the CDIP's third session in April 2009.
5. The present Study Paper aims to provide a broad overview of available online services offering the possibility to search and access patent and non-patent data. Given the vast and growing number of patent and non-patent database services offered by different providers, a complete review of all existing services is simply not feasible. In analyzing needs for access to patent databases, it is necessary to cover also databases that may not necessarily be “specialized”. An overview of all patent databases, whether specialized or not, should be useful so as to provide a complete understanding and assessment as to whether, and to what extent, certain features and functions of specialized databases are special and necessary for the particular needs of users. Consequently, the Study Paper examines a selection of commercial and non-commercial database services considered representative of the broader population of existing services in order to illustrate types and combinations of features available through these services.
6. Moreover, the Study Paper examines only databases available online to the public. Internal search systems such as the European Patent Office's EPOQUE system and the United States Patent and Trademark Office's Examiner's Automated Search Tool (EAST), Web-based Examiner Search Tool (WEST), and Foreign Patent Access System (FPAS), as well as national registries containing additional information such as legal status, are excluded from consideration as they are designed and developed for internal use.
7. Within the implementation of Recommendation 8, discussions with commercial patent database providers are foreseen, with the aim of negotiating preferential rates for access to specialized patent databases. The information contained in this Study Paper may provide a substantive basis for these discussions and assist beneficiaries of any program

¹ http://www.wipo.int/edocs/mdocs/mdocs/en/cdip_1/cdip_1_3.doc

² http://www.wipo.int/edocs/mdocs/govbody/en/a_43/a_43_16-main1.pdf

³ Annex A of A/43/16. See: http://www.wipo.int/edocs/mdocs/govbody/en/a_43/a_43_16-main1.pdf

that emerges from the negotiations to determine the database resources best suited to meet their specific needs.

(b) Format of the Study Paper

8. The Study Paper basically follows the format proposed by the “Terms of Reference” (CDIP/2/INF/3) with the exception that the needs analysis for each country follows the reviews of the patent and non-patent databases services so as to form two major sections: a first database reviewing section followed by a second analytical section. The latter provides an analysis of the intellectual property offices’ and users’ patent information needs, based among other things on the responses received to Circular C.N 3024 (attached as Annex I to this Study Paper), and a comparison of free-of-charge and fee-based database services, as well as general recommendations and conclusions.
9. In August 2009, WIPO sent Circular C.N 3024 to the intellectual property offices (IPOs) of its Member States comprising a patent information needs analysis questionnaire. The results of the questionnaire, which were returned by 72 IPOs, contributed to establishing a picture of the current state of patent activity and the needs of developing countries and least developed countries (LDCs), to selecting commercial and non-commercial patent and non-patent databases to be reviewed, and to providing recommendations on how to facilitate the access of these countries to patent and non-patent information so as to enhance their level of innovation and development. A table summarizing all the responses is attached as Annex II to this Study Paper.
10. A more in-depth review of the patent and non-patent database functionalities can be found in the document “Guide to Technology Databases” found on the WIPO’s website at www.wipo.int/patentscope. This document also provides summary tables containing additional, more detailed technical information as a quick reference guide so as to facilitate a broad comparison between the different database services. It can be used as a reference guide to assist intellectual property offices (IPOs) and users in general in selecting the most appropriate databases to search from, as well as in how to carry out a more effective search within these databases.

2. REVIEW OF PATENT DATABASE SERVICES

(a) Overview

11. Patent database services are commonly designed to enable various types of patent search, depending on the user's specific needs, including novelty or inventive step (non-obviousness), validity, infringement, freedom-to-operate or clearance, and state of the art searches. Database services differ in terms of geographic and historical coverage, as well as in the type of documents available (whether patent applications, granted patents or utility models) and the elements of these documents that are accessible and searchable (whether titles, abstracts, description, claims, or other elements). In addition, many databases offer various tools designed to facilitate the retrieval of relevant search results as well as the visualization and analysis of these results.
12. With regard to completeness and a better overall assessment of user needs, the present Study Paper provides an overview of all patent databases which are available to the public but exclude patent databases available only for internal use by IP Offices. The Study Paper distinguishes between three types of patent databases, namely public sector databases (i.e. those provided by national and regional patent offices), free-of-charge private sector databases, and fee-based private sector databases. Public sector databases have been selected from those provided by offices with the highest patenting activity, since these databases can be expected to represent the largest patent collections. Private sector databases are drawn from a group of the most widely known and commonly used databases, as indicated, e.g., by the responses received to Circular C.N 3024.
13. Certain fields of technology are especially relevant to the research and development activities of developing countries, notably pharmaceuticals and organic chemistry, as evidenced by patent filing activity in these countries. As a result, special weight was given in the selection of database services to be reviewed to those services featuring tools designed with these fields in mind.

14. Specifically, the Study Paper reviews the following databases:

Public sector database services provided by WIPO and national and regional Offices⁴	
World Intellectual Property Organization	http://www.wipo.int/patentscope/search/en (<i>PCT + national</i>)
Australia	http://www.ipaustralia.gov.au/patents/search_index.htm
Canada	http://patents1.ic.gc.ca
China	http://www.sipo.gov.cn/sipo2008/zljs (<i>Chinese language</i>) http://218.240.13.210/sipo_EN (<i>English language</i>)
Germany	http://depatisnet.dpma.de/DepatisNet
United Kingdom	http://www.ipo.gov.uk/types/patent/p-os/p-find.htm
Hong Kong, SAR	http://ipsearch.ipd.gov.hk/patent
India	http://www.patentoffice.nic.in/PatentSearch/ipirs_index.htm
Israel	http://www.ilpatsearch.justice.gov.il (<i>Hebrew language</i>) http://www.ilpatsearch.justice.gov.il/UI (<i>English language</i>)
Japan	http://www.ipdl.inpit.go.jp (<i>Japanese language</i>) http://www.ipdl.inpit.go.jp/homepg_e.ipdl (<i>English language</i>)
Republic of Korea	http://patent2.kipris.or.kr/pat (<i>Korean language</i>) http://patent2.kipris.or.kr/pateng (<i>English language</i>)
New Zealand	http://www.iponz.govt.nz/cms/banner_template/IPPATENT
United States of America	http://patft.uspto.gov
Eurasian Patent Office	http://www.eapatis.com (<i>Russian language</i>) http://www.eapatis.com/ensearch (<i>English language</i>)
European Patent Office	http://www.espacenet.com/access
Free-of-charge private sector database services	
Google Patents (Google)	http://www.google.com/patents
PatentLens (Cambia)	http://www.patentlens.net
Fee-based private sector database services	
Orbit.com (Questel)	http://www.orbit.com
PatBase/PatBaseXpress (Minesoft)	http://www.patbase.com (<i>PatBase</i>) http://www.patbaseexpress.com (<i>PatBaseXpress</i>)
STN (CAS/FIZ Karlsruhe)	http://www.stn-international.de
Thomson Innovation (Thomson Reuters)	http://www.thomsoninnovation.com
Total Patent (LexisNexis)	http://www.lexisnexis.com/totalpatent
WIPS Global (WIPS)	http://www.wipsglobal.com

15. For the purposes of the present Study Paper, commercial database services are organized alphabetically according to the name of the service, while database services provided by national and regional offices are organized alphabetically according to their recommended two-letter code under WIPO Standard ST.3, with national offices and regional offices indicated separately.
16. The landscape of patent and non-patent databases is constantly changing, with new services coming onto the market and existing services being expanded, merged or discontinued. Notable examples of this transformation include Delphion and Micropatent, provided by Thomson Reuters, whose features have been incorporated into Thomson Innovation, the company's flagship product, and which are therefore in the process of being phased out. Other examples include PatAnalyst from Jouve, which is being maintained for existing subscribers but for which no new subscriptions are accepted, and Questel's QPAT, which will be integrated into the Orbit portal in the future. A further

⁴ For further databases made available by national and regional offices, see: http://www.wipo.int/patentscope/en/search/national_databases.html

dimension for change lies in the pricing of services: Free-of-charge services frequently transition to paid services, as in the case of Delphion, PatentCafe, and Surechem.

17. The most basic features of patent database services include data coverage and search tools and functionalities. These features are a key to determining the most appropriate databases for a given task and can vary widely from one service to the next. For each database service, an overview is provided of key features, highlighting particular strengths and weaknesses of the service.
18. As mentioned in the Introduction above, a more detailed review of the database functionalities as regards coverage, classifications used, search options such as search criteria, Boolean, wildcard and proximity operators, language features and tools, as well as display/visualization formats can be found in the document "Guide to Searching Technology Databases" found on the WIPO's website at www.wipo.int/patentscope.
19. The following sections will mainly focus on patent databases which are generally regarded as specialized databases.

(b) Database content sets

20. The following content sets are identified as important collections of actual data as opposed to database services within which they can be included, e.g. INPADOC legal status data can be found in many database services.

(i) INPADOCDB bibliographic data

21. The INPADOCDB dataset (commonly referred to simply as INPADOC) is compiled by the European Patent Office (EPO) and included in the content of numerous database services, notably the EPO's esp@cenet search service and German Patent and Trademark Office's DepatisNet search service. It comprises a broad range of bibliographic data from patent offices around the world, though completeness of coverage varies greatly among offices.⁵ The dataset also forms the basis for the identification of INPADOC patent families, groups of patent documents related by one or more priority claims.

(ii) INPADOC legal status data

22. The INPADOC legal status data set is compiled by the European Patent Office (EPO) and is the most common source of legal status information available through patent database services. A wide range of patent authorities (currently 56 in total, including the former German Democratic Republic) are covered by the data set, with different types of legal status information provided for each specific authority. This information includes, for certain authorities, national phase entry data for European patent applications and PCT applications, corrections, invalidations, lapses and expiries, and assignments.⁶

(iii) Derwent World Patent Index

23. The Derwent World Patent Index (DWPI) is maintained by Thomson Reuters and includes value-added bibliographic data, notably additional classification (DWPI Classification) as well as rewritten titles and abstracts.

⁵ For complete INPADOCDB bibliographic data coverage information, see:
[http://documents.epo.org/projects/babylon/rawdata.nsf/0/71300EF4F56AF961C125770E0046F3DE/\\$File/PFS_1017.xls](http://documents.epo.org/projects/babylon/rawdata.nsf/0/71300EF4F56AF961C125770E0046F3DE/$File/PFS_1017.xls)

⁶ For complete INPADOC legal status coverage information, see:
[http://documents.epo.org/projects/babylon/rawdata.nsf/0/81D7A52B8CEC65C1C12577230040808D/\\$File/P RS_1020.xls](http://documents.epo.org/projects/babylon/rawdata.nsf/0/81D7A52B8CEC65C1C12577230040808D/$File/P RS_1020.xls)

(c) **Fee-based commercial databases**

(i) **Orbit.com**

24. **Orbit.com** is a web-based commercial searchable patent database made available by Questel, with full-text coverage of the PCT, Chinese, European (EP), Japanese, and US collections as well as a number of other, mainly European, collections.⁷ Bibliographic data is available for the Korean collection. Patent documents are grouped into database records according to the proprietary FamPat patent family scheme. Thus, while records represent a single invention, they can be accessed using the data contained in any of the patent documents belonging to the associated FamPat patent family. For example, if a record includes both a European and US patent filing, it could be retrieved using either ECLA (classification used for European patent documents) or USPC (classification used for US patent documents). While original patent classifications are retained for all patent documents, application and priority numbers for all patent documents are standardized using the Questel Orbit format. To assist users in entering patent numbers in the correct format, a **patent number formatting assistant** and **automatic patent number identification tool** (“Patent Number Wizard”) is provided. Legal status information is available for the Japanese collection as well as for certain other collections through the INPADOC legal status dataset.
25. Several different search interfaces are provided for Orbit.com, including basic Express and Advanced, classification (“Similarities Search”), Family and Extended Family, and Citation search interfaces. The structured **Express Search interface** allows the use of basic Boolean and wildcard operators and searching by field in the full-text, assignee/inventor, and number fields only, while the **Advanced Search interface**, which includes both command-line and structured search elements, permits searching by field in the full range of available fields as well as the use of proximity operators. Orbit.com supports simultaneous left and right truncation (SLART) as well as internal truncation. The Advanced Search interface also includes the possibility of **cross-language semantic search** for English, French, and German (i.e., search terms entered in any one of these languages can be searched in multiple variations in all three languages), though use of Boolean, wildcard, and proximity operators is only possible on the query generated when this option is selected. Results from both Express and Advanced searches can include **highlighting of search terms** according to user-definable color schemes (“keyword-in-context”, or “KWIC”).
26. The **Similarity Search interface** permits searching inventions similar to a given patent document according to ECLA patent classification, or IPC patent classification if ECLA is not present, and includes cited and citing documents. Patent collections to be included in the search can also be selected. The **Family and Extended Family search interfaces** allow the identification of patent families related to a particular patent document based on the proprietary FamPat and extended INPADOC datasets, respectively. FamPat patent families are smaller in size than extended INPADOC patent families, since they group only those documents sharing all priority claims rather than all documents sharing at least a single priority claim.
27. Any search, including Family and Extended Family searches may include **legal status information** as well as citations, the latter of which may be presented in a **graphical citation display** with priority data and timelines. Graphical views may be exported in PDF or GIF format. The Citations search interface permits the identification of forward citations

⁷ For more detailed information with regards to QPAT bibliographic and full text coverage please consult http://www.questel.com/customersupport/Coverage_and_Updates_FullText.htm and http://www.questel.com/customersupport/Coverage_and_Updates.htm

(i.e., citing documents) for patent documents using application, priority, or publication numbers or keywords (with the option of cross-language semantic search) as well as graphical presentation of citations. Search results for this interface include both **patent and non-patent literature citations**.

28. **Search histories** may be viewed and stored to user profiles, and **search alerts** set up to track new records meeting criteria for a search previously carried out as well as for changes in legal status.
29. Tools available for presenting and analyzing search results include the Family Citation and PatCitation tools and the ANALYZE tool. The Family Citation and PatCitation tools allow users to create **graphical representation of the various family/citation relationships** (by date, inventorship or ECLA/IPC classification) – with citation relevance indicators in international search reports (X, Y, A). The ANALYZE tool identifies the **top assignees, ECLA, US, and IPC classes** among 500 randomly chosen search records (if more than 500 search results are available). It also allows the expansion of a results set to cited and/or citing documents. Lastly, the Statistics module allows the statistical analysis and graphical display of up to 15 000 records.
30. A number of additional tools are offered within Orbit.com that are designed to enable **workfile management** (“Your Workfiles”/“PatentExaminer”) and provide **search alerts and legal status alerts** (“PatLegalMonitor”).
31. Related services: **PatCitation** provides the option of carrying out graphical citation analysis using data generated through Orbit.com.

(ii) PatBase/PatBase Express

32. PatBase is a search service developed by Minesoft Ltd and RWS Group. In contrast to many other patent search services, PatBase organizes searches and presents search results on the basis of patent families rather than individual patent documents. Thus, among other things, PatBase displays unified classification and assignee data for each patent family record. Within this framework, PatBase covers the Chinese, European, Japanese, Korean, PCT and US collections, as well as the British, French and German collections with historical coverage to 1900 for certain collections. In total, PatBase contains over 38 million patent documents, of which around 25 million are searchable in their full-text. Legal status information is available for certain records, based on the INPADOC dataset
33. PatBase offers a Quick Search and an extended structured search interface, non-Latin character and a command-line search interface. All interfaces permit the use of Boolean and proximity operators, as well as wildcard operators for simultaneous left and right truncation (SLART). The **Quick Search interface** and **structured search interface** differ in that the Quick Search interface only allows searching in a single field (or composite field) at a time, with the option of filtering by date range or patent office, while the structured search interface supports simultaneous searching in a number of fields at once. No option exists to use Boolean operators across fields in the structured search interface, so all fields are effectively included in the search (i.e., linked by the AND operator). The **non-Latin character search interface** permits searching in Japanese, Chinese, Korean and Russian. Results from the non-Latin character search interface may be merged into results from other search interfaces through the user’s search history. The **command-line search interface** permits the flexible use of the full range of operators and field codes supported by PatBase.
34. The tools available to facilitate searching through PatBase include a **country code search feature**, an **assignee/inventor index lookup tool** to assist users in identifying the spelling of names, and a **“Number Wizard”** designed to assist users in identifying the correct

format for priority, application, and publication numbers. PatBase also provides a chemical synonym lookup tool, which includes a limited number of CAS registry numbers.

35. PatBase offers two options for **reviewing search results**, namely a “view” and a “browse” option. The “view” option shows information for each individual record in a single window. The “browse” option provides information on individual records in a central pane, while other members of the related patent family and displaying tools are accessible through separate panes. The “browse” option also allows users to access the “smart claims viewer”, which allows users to click through to independent claims for US and GB documents.
36. The **display tools** accessible from the search results view include customized search term highlighting (“keyword in context”, or “KWIC”) and a “Family Explorer” tool. The **Family Explorer** tool facilitates the display of family members and text bibliographic/full-text data, allows the exporting of selected patent family members, and permits users to view the respective **priority map**, a graphical display of the relationship between the various family members based on their priority number. INPADOC legal status may also be reached using the Family Explorer tool.
37. **Keyword analysis** can be carried out using the VizPat Keyword Analysis tool to identify the most frequently appearing keywords in title, abstract and claims, in title and abstract only, in claims only, or in French, German, or US claims only. PatBase also provides **statistical analysis** of top assignees/applicants, most frequently appearing IPC, ECLA, and PCL classes and F-terms, and top countries of filing through the VizPat Statistical Analysis tool as well as the option of preparing **3D graphs** of results data.
38. **Search histories** can be viewed and stored in user profiles, and **search alerts** can be set up to obtain information on newly added records meeting specific search criteria.
39. **PatBaseExpress** is a simplified version of PatBase for novice users and is available in four interface languages, namely English, French, German, and Japanese. PatBaseExpress offers nearly identical search interfaces to those found in PatBase, though with more limited search options (less filtering options). PatBase Express also provides a Number Search interface to allow quick access to a specific patent document based on its publication, application, or priority number.
40. Related services: PatBase permits exporting of results to the **Minesoft Insight Pro** or **VantagePoint** analysis tool and the **BizInt Smart Charts** report preparation tool. Changes to patent families can be tracked through email alerts using **PatentTracker**.

(iii) **STN**

41. STN on the Web and STN Easy are two web-based system interfaces for the STN online database service, operated by the Chemical Abstracts Service (CAS) of the American Chemical Society and FIZ Karlsruhe. Through its various interfaces, STN provides full-text coverage of the European (EP), PCT and US collections as well as the British, French, and German collections. Bibliographic data and claims in English translation are also available for the Chinese, Japanese, and Korean collections. Other searchable content includes the Derwent World Patent Index (see Section 2. (b) (iii)), the MARPAT Markush structure database and, uniquely, the CAPIus content set, which permits searching of patent documents from the European, German, PCT and US collections from 1997 onwards (as well as non-patent documents) by CAS Registry Number, unique identifiers used to designate chemical substances, compounds, and mixtures as well as biological sequences.
42. STN on the Web provides a single **command-line search interface**. This interface permits the selection of the databases to be included in a search, which in turn determines which fields can be searched as well as the operators that can be used and their function. The exact characteristics of each database are available through database summary

sheets. Generally speaking, the command-line interface permits the use of basic Boolean operators and a wide variety of proximity operators (whose function and availability depends on the databases searched) as well as wildcard operators that can be used for simultaneous left and right truncation (SLART) as well as internal truncation.

43. The tools and features available through STN on the Web, which must be activated using command-line language, include a **search index viewer** ("EXPAND"), which allows users to examine part of the search index for a particular data field, i.e., a list of terms actually present in that data field, and obtain information on the number of records containing these index terms. This tool can assist users in identifying relevant search terms. STN on the Web also permits the **grouping of patent families** and **removal of duplicate records**. It also permits the **uploading and searching of chemical structures and substructures** in the CAPIus, MARPAT, and CAS Registry databases. **Session histories** can be viewed and exported in a variety of formats.
44. STN Easy offers four search interfaces, namely Easy, Advanced, Patent Lookup, and CAS Number search interfaces. The **Easy search interface** only enables searching in the "General Science" category of databases and permits the use of Boolean operators and wildcard operators, though only for right truncation. The **Advanced search interface** provides the option of selecting from a broader range of different database categories, including the patents category and several categories that combine patent and non-patent databases such as the "Pharmaceuticals" category. This interface also permits searching by field and the use of Boolean operators (accessed through drop-down menus) and wildcard operators, though also for right truncation only. The **Patent Lookup search interface** is adapted more specifically to searching for patents and provides a pre-defined set of search fields, unlike the user-definable search fields available through the Advanced search interface, and allows only a single Boolean operator to be selected for all search fields (i.e., either AND, OR, or NOT). The **CAS Number search interface** allow the retrieval of information on chemical compounds based on CAS Registry numbers or chemical names, e.g., various synonyms for a particular chemical substance, compound, or mixture. However, this interface does not in itself serve to identify patent documents in itself.
45. Both the Patent Lookup and Advanced search interfaces for STN Easy provide access to the **search index viewer** described above. Search results for all interfaces can be set to include **customized highlighting of search terms**. STN Easy also provides the option of viewing and exporting search histories in a variety of formats.

(iv) Thomson Innovation

46. Thomson Innovation is an integrated search service launched by Thomson Reuters that offers cross-content search across patent data, scientific and technical content, and business information. The patent collections covered by Thomson Innovation include the British, Chinese, European, French, German, Japanese, Korean, PCT and US collections as well as bibliographic data from other collections. Notable among the content included in Thomson Innovation is the Derwent World Patent Index (DWPI), an additional source of classifications and abstracts for the patent documents searchable through Thomson Innovation (see Section 2. (c) (iii)), as well as legal and procedural information for certain collections (European, INPADOC and US). Machine-assisted English translations of titles, abstracts, and claims for more recent Chinese patent documents and machine translations into English of the full-text of recent Japanese and Korean patent documents are also available (and searchable) through Thomson Innovation.
47. The user can carry out a search using Quick Search, structured Fielded, Expert, Publication Number, and Non Patent Literature search interfaces. The **Quick Search interface** allows users to select only whether to search within a single content set (patent, scientific and technical, or business content) or across all content. However, this

interface does support the use of Boolean and wildcard operators. The **Fielded search interface** already offers a greater range of options, including the possibility to select the specific patent collections to be searched (collections by authority) and to search in patent full-text, bibliographic data or Asian translated data or in DWPI data fields. Fields to be searched and Boolean operators can be selected from drop-down menus. The **Expert search interface** adds the possibility of using proximity operators and to more flexibly structure searches. All three search interfaces offer simultaneous left and right truncation (SLART) and internal truncation.

48. To assist users in formulating their queries in the Expert search interface, Thomson Innovation offers a set of tools including a **command-line code generator** ("fields and tags menu"), a **query syntax checker** ("text syntax"), and a **hierarchical corporate entity index** ("corporate tree"), which groups assignee names within a corporate hierarchy.
 49. The **Publication Number search interface** provides direct access to data associated with a particular patent document or set of patent documents, with four output options ("Result Set", "Work File", "Document Copies", "File Histories") and an expandable "specialized search" panel, allowing the limitation of the search to either patent citations (forward, backward, or both), patent families (INPADOC or Derwent), or records for which family, legal status, publication stage, citations or assignment have changed since a specified date. The Publication Number search interface also permits the option of automatically including the patent family in the search results ("Family Lookup").
 50. Non-patent literature search options are discussed in greater details below (see Section 3. (e) (iv)).
 51. Thomson Innovation offers a number of reviewing and analysis tools and features. Search results may be **grouped by INPADOC or Derwent patent family** ("collapse by family") and **highlighting of search terms** ("keyword in content", or "KWIC") is possible in the text of individual records, where the number of hits for a particular search term is given for specific data fields. Through the Analyze menu, the user can create **charts and graphs** based on pre-defined templates or user-definable options (selecting the fields to be analyzed, the number of items to be included, and various display options), perform **keyword clustering** and **graphical keyword analysis** ("ThemeScape") as well as **graphical patent and non-patent citation analysis**, based on generation, time and generation, and backward or forward citations or both.
 52. **Search histories** may be viewed and saved to user profiles, and **search alerts and record change alerts** ("Watched Records") may be set up to track new records meeting given search criteria or changes in the legal or procedural status or in applicant/assignee as well as the addition of new patent family members and new citations for specific records. Results may also be exported in a number of different formats, including PDF, XML, comma and tab-separated text, and proprietary RIS and BizSoft formats.
- (v) **TotalPatent**
53. TotalPatent is a patent research, retrieval and analysis product from LexisNexis, a member of the Reed Elsevier Group. It offers access to 27 full-text collections including the Chinese, European, Japanese and US patent offices, as well as Australia, Brazil, Canada, Russia and a host of European patent issuing authorities. In addition to the full-text collections, TotalPatent also permits searching in bibliographic data (including abstracts) from numerous other offices around the world. In total, the service includes more than 40 million records searchable in their full-text and approximately 75 million bibliographic records, 50% of which include an abstract. The majority of collections are searchable in their original language of publication as well as in English through the use of machine translations. TotalPatent also provides access to 16 million citations, and over 33 million INPADOC legal status data plus information on US re-assignments and litigation.

54. TotalPatent offers Guided Search and Advanced Search interfaces as well as Semantic Search and Notes Search interfaces. The **Guided Search interface**, designed for novice searchers, enables searching by assignee, inventor, IPC or USPC symbol, publication date, and in full-text, title, abstract, and claims simultaneously, in title and abstract, or in title or claims only, as well as other fields that can be added by the user. It also permits filtering by patent collection and the use of Boolean, wildcard and proximity operators within each given field. The **Advanced Search interface**, designed for more expert users, further permits filtering by document type, by patent collection (distinguishing between collections available in full-text or with bibliographic data only), and type of document (granted patent or application). Though both interfaces are laid out as structured search interfaces, they permit the use of field codes, thus effectively permitting command-line queries to be established by searching in the combined full-text, title, abstract, descriptions and claims fields. Overall, TotalPatent indexes approximately 200 data fields.
 55. To assist users accustomed to using Delphion and Micropatent in correctly entering queries, the Guided Search and Advanced Search interfaces provide a **“syntax converter”** from the syntax used by these search services to TotalPatent syntax. A further search tool made available through TotalPatent is its searchable **hierarchical corporate entity index** (“Directory of Corporate Affiliations”), and Assignee and Inventor name indices, which can assist users in identifying different names for companies, their subsidiaries, and inventors to be included in searches.
 56. The **Semantic Search** interface allows users to generate semantic search queries based on specific search terms entered by the user or any text of up to 32,000 characters describing the concept to be searched. TotalPatent identifies the keywords in the entered text and proposes related terms, weighted by relevance from mandatory terms to terms to be excluded in the search. Users have the option to remove suggested search terms or change their weighting before carrying out a search, with the option of including English translations in the searched content and the option to remove family member duplicates. The **Notes Search interface** allows searching of **tags given to specific patent documents** by the user himself or by other users within a particular user’s group (company or organization), which can be searched by the email address of the note’s author, its date of creation, title or text.
 57. Among the reviewing and analysis features offered by TotalPatent is the option to **view “main family” data** (equivalent to the Espacenet family) or **“extended family” data** (equivalent to the INPADOC family) for a particular patent document, where main families are composed of documents sharing all priority data and extended families consist of documents sharing at least one priority claim. TotalPatent also offers the possibility of carrying out **graphical results and citation analysis** according to user-defined criteria. Graphs, patent family data, and other data may be exported as customized reports in a variety of formats including text, PDF, and HTML formats.
 58. Search histories may be saved to user profiles, and **search alerts** and **record change alerts** established to track new documents meeting the criteria for a particular search query or changes to a particular patent document, including changes in legal status, new patent family members, new forward citations, and changes in applicant/assignee.
- (vi) **WIPS Global**
59. The World Intellectual Property Search (WIPS Global) is a searchable patent database that provides full-text coverage of the PCT, European (EP), and US collections. Bibliographic data and English translations of abstracts are available for the Chinese, Japanese, and Korean patent collections. Bibliographic data is available for other collections, while original abstracts are also present for several European collections. Legal status information is provided through the INPADOC content set.

60. WIPS Global offers a number of different search interfaces, including structured General, Advanced, and Step interface and a Number search interfaces. It also offers an **Integrated search interface**, which permits cross-collection searching (unlike the other interfaces which permit searching only in one collection at a time). The **General search interface** offers the possibility to search in up to two fields, selecting collection, patent type (application/granted patent), document field (front page and exemplary claim or full text), date range to be searched, and the information to be displayed in the search results. The General search interface supports the use of Boolean, proximity, and wildcard operators. The Advanced search interface offers a greater number of search fields, all of which are linked by the AND operator. The **Step search interface** enables users to establish more complex queries by combining individual queries (“steps”) set up using a structured search interface. The **Step search interface** supports Boolean operators within each individual query and for linking different queries as well as wildcard operators. The **Number search interface** permits direct access to specific patent documents by way of their application or publication number.
61. To facilitate searching by applicant name, WIPS Global provides a **searchable corporate name index**, accessible through the Advanced search interface, in which different names for specific applicants are grouped under a single code. **Results can be clustered** according to IPC class, applicant, and application date. **Graphical and text-based citation analysis** can be performed for specific patent documents, graphically illustrating the relationships between forward and backward citations and filtering citations according to generation and other user-definable criteria, respectively. Searches can be saved to user profiles (“My Folder”), while **search results can be batch downloaded** in image format or a variety of text-based formats.
62. Related services: **PM Manager** and **ThinkKlear** provide additional statistical analysis and report creation tools.
- (d) Free-of-charge commercial databases**
- (i) Google Patents**
63. Google Patents provides full-text search of the complete US patent collection, including the period prior to 1976, the earliest date for which full-text search is available through the US Patent and Trademark Office’s own PatFT database service. Full-text data for the period prior to 1976 is generated through an optical character recognition process and assigned to fields based on a similarly automatic process, which may be a source of errors in the searchable text and in its assignment to various fields.
64. Google Patents offers basic and advanced patent search interfaces. The **basic search** only provides one field to enter any search key word. The **advanced search**⁸ provides the possibility to limit the results in those which contain all the words, the exact phrase, one of the words or exclude these words from the search. Further search options are patent number, title, name of inventor or assignee, US/IPC classification, search in issued patents or applications, the type of patents (plant, utility, X patent, reissue etc), and issue and/or filing date. The keywords are highlighted, even within an image/drawing and can be viewed along with the rest of the patent/application information search results through the PNG format in one page.
65. A list of maximum 600 results is displayed, ranked upon relevance, with title of the invention/application, filing/application number and date, publication number, as well as name of assignee, a snippet of text and a hyperlink to retrieve the patent. The results can

⁸ To access the Google Patents advanced search interface, see:
http://www.google.com/advanced_patent_search?

be viewed also as thumbnails and the user can thus get a graphical overview of the applications/issued patents. Clicking on the hyperlink, the user has the option to read the patent (or application) or download it as PDF. He can directly be taken to the abstract, drawing, description and claims part of the patent (or application) and search within it for keywords, which then are highlighted. Google Patent Search offers a **clip tool** which enables the selection of a part of an image or text in order to paste the clip into a blog or a webpage.

66. Moreover, in a recent agreement with the USPTO, Google will provide bulk electronic patent and trademark data to the public.

(ii) Patent Lens

67. Patent Lens is a multi-country search service provided by Cambia. It offers full-text search of PCT applications, US patent applications and granted patents, and European (EP) granted patents as well as full-text search (excluding claims) of Australian patent applications and granted patents, a feature not currently offered by IP Australia. INPADOC legal status information is available for many patent documents included in the service.

68. The following search forms are offered: Number Search; Quick Search; Structured Search; Expert Search; and US Sequence Search. **Number Search** allows search by patent or publication number, where patent collections and document types to be included in the search can be selected. **Quick Search** enables keyword or patent/publication number search, with a stemming option and the possibility to limit the search to certain patent collections. In addition to the options available through the Quick Search, **Structured Search** provides the possibility to search in specific fields, including applicant/assignee, front page, title, abstracts, inventor, agent, references, description, and claims (not for Australian documents). It further permits the possibility to filter results by publication/filing date and by predicted expiry date or lapsed date (for US patents only). Expert Search offers the option of a command-line search in the fields covered by the Structured Search. In addition, PatLens also offers a DNA, RNA and Protein **Sequence Search** which can be carried out either by patent number and Sequence/Genbank ID or by INPC DNA, RNA and Protein. **Similarities search of biological sequences** can also be carried using the National Center for Biotechnology Information's BLAST software.⁹

69. The results can be ranked by relevance, patent number and application or publication date.

(e) Databases provided by WIPO and national and regional offices

(i) World Intellectual Property Organization

70. The World Intellectual Property Organization (WIPO) offers full-text search of international patent applications, filed within the framework of the Patent Cooperation Treaty (PCT), as well as access to certain national and regional collections through its PATENTSCOPE[®] search service. At present, the offices covered include ARIPO, Cuba, Israel, Mexico, the Republic of Korea, Singapore, South Africa and Viet Nam. For PCT applications and patent documents from Cuba, Israel, and Mexico, full document images are available. Additionally, related documents including search reports, forms, and declarations can also be retrieved for the PCT collection as well as sequence listings, where available.¹⁰ PCT national phase entry data can also be accessed for over 40 offices around the world as can the most recent status information available for PCT applications, through the International Application Status Report (IASR).

⁹ For more information on BLAST, see: <http://www.ncbi.nlm.nih.gov/blast/download.shtml>

¹⁰ For complete information on document coverage, see: <http://www.wipo.int/pctdb/en/content.jsp>

71. The PATENTSCOPE[®] search service provides basic, structured (fielded), advanced, and cross-lingual search interfaces. The **basic search interface** offers the option of searching in full-text, by ID/number, international classification, names, or publication dates across one or more collections. The **structured (fielded) search interface** provides a wider range of searchable data fields, which can be linked using Boolean operators. Additional search fields may be added as necessary to allow greater flexibility in searching without requiring knowledge of field codes. The **advanced search interface** permits the use of command-line query syntax, including field codes and a full range of Boolean, proximity, range, and special operators. All three interfaces allow the use of proximity and wildcard operators as well as the **fuzzy search** operator. They also permit the selection of the query language used and automatic word stemming.
72. The **cross-lingual search interface** of the Cross Language Information Retrieval (CLIR) tool allows the simultaneous searching of a given term in one language in its variants in multiple languages (currently: English, French, German, Japanese and Spanish). Automatic and manual (“supervised”) term expansion modes are available, where the manual mode permits the user to select the technical field to which the entered term belongs and adjust how closely the variants must be related to this term. Finally, the fields in which the term and its variants are to be located can be selected.
73. Search results can be ordered by relevancy, publication date, or application date. **Statistical analysis** is automatically performed on search results to show number of records for top offices of filing, top fields of technology (top IPC classes), top applicants, top inventors, and publication dates. **Graphical representations** of these top groups can be prepared, and **additional filtering** of search results can be carried out for these groups.
- (ii) **Australia**
74. Through its AUSPAT search service,¹¹ IP Australia offers search of Australian patent applications and PCT applications filed from 2004 that have entered the national phase in Australia and all PCT applications designating AU filed from 1980 to 2003. Australian applications filed between 1920 and 1979 are generally searchable only by application number and IPC symbol, while later applications are searchable according to additional bibliographic data. Descriptions (specifications) are available through the system but are currently not searchable.¹² Document images are provided for patent documents published from late 1998. Legal and procedural status information, including information on oppositions, is available and generally searchable for applications covered by AUSPAT.
75. AUSPAT offers three interfaces for searching patent applications, namely Quick, Structured, and Advanced Search interfaces. **Quick Search** allows simultaneous search in a limited number of bibliographic data fields, namely patent or application number, title, applicant/inventor/agent name and PCT/WO number. **Structured Search** permits the selection of one or more specific data fields in which to carry out a search, including additional fields not available for search using the Quick Search such as pharmaceutical name, priority date and number, and expiry date. **Advanced Search** further allows the use of Boolean operators.
76. Through the Advanced Search interface, users have access to **searchable applicant and inventor name indexes** (“Name Selector”) as well as a **procedural action search wizard** (“Publication Action Selector”). Search histories and selected applications may be saved to

¹¹ AUSPAT supersedes the PATADMIN and PatSearch services, which are expected to be fully integrated into AUSPAT in 2010.

¹² Coverage of searchable full-text from 1904 to present, including abstracts, descriptions, and claims (separately searchable from 1998), is under production.

user profiles (“MySearches” and “MyList”). Saved search histories may be combined to create composite queries, while saved applications can be further evaluated and exported.

(iii) Canada

77. The Canadian Intellectual Property Office (CIPO) offers full-text search (excluding descriptions) of its collection from August 15, 1978 onwards and search in bibliographic data from 1869 onwards. Full document images (including descriptions) can be accessed for applications registered as far back as 1920. Procedural status information, including regarding the payment of fees, as well as information regarding withdrawals is also accessible through CIPO’s database.
78. Canadian patent documents are published in the language in which the document was filed (English or French, Canada’s official languages). However, the Canadian Patents Database display both English and French titles except for applications filed between 1960 and 1978 where titles are available only in the language used at the time of filing. For Canadian patent documents originating from PCT applications, abstracts received since 2000 are also available in both official languages.
79. The database can be searched using Basic Search, Boolean Search, Advanced Search, and Number Search interfaces. **Basic Search** carries out a simultaneous search in multiple fields and offers the option of carrying out a search simultaneously in French and English data available through the databases (“bilingual search”). **Boolean Search** allows selection of up to three data fields to be searched including title, abstract, and claims, IPC, and CPC (Canadian patent classification) as well as the use of Boolean operators (AND and OR only). This interface also permits filtering of results by date range, direct or PCT filing, and availability for licensing. **Advanced Search** permits the use of additional Boolean, proximity, and special operators, however only within a particular search field, and extends the available search fields to include fields such as inventor country. **Number Search** can be used to navigate directly to the application or procedural status for a particular patent document. Though not available directly within a standard interface, full command-line search can be carried out from the results screen for the Basic, Boolean, and Advanced Search interfaces.

(iv) China

80. The State Intellectual Property Office of the People’s Republic of China (SIPO) offers search in bibliographic data (including abstracts) of Chinese patent application and utility models in both the Chinese and English languages.
81. Search interfaces are available in both English and Chinese, offering the possibility to search by application and publication numbers and dates, IPC classification, applicant, inventor, and agent names, priority number, as well as country and province codes using pre-defined search fields. The English language interface also permits command-line searching, including the use of Boolean operators. Results can be further refined by searching within the results set and can be filtered according to patent applications or utility models.
82. English translations are available of bibliographic data and abstracts for patent applications, while bibliographic data (including titles) for utility models are also available in English translation. Machine translations into English of descriptions and claims can be prepared on the fly using the China Patent Machine Translation (CPMT).

(v) Germany

83. The German Patent and Trademark Office (DPMA) offers search and access for the German collection in full-text and for numerous other collections (bibliographic data only)

through its DEPATISnet search service. Additionally, full document images may be downloaded in PDF format for a limited number of collections.

84. Search interfaces are fully available in German and English and include a Beginner, Expert, IKOFAX, and Family search interface. The **Beginner interface** provides a number of standard search fields including publication date and number, applicant and inventor, classifications, and full-text (though description and claims are only searchable for the German collection). The **Expert and IKOFAX interfaces** both permit command-line searching using the IKOFAX search language, while the Expert interface provides assistance in formulating queries by allowing users to select data fields and operators to be included in the search from a drop-down menu or using clickable buttons, respectively. The **Family interface** permits the identification of INPADOC patent families by entering the publication number of a specific patent document. The Assistant interface allows to forward requests to professionals at the German Patent Information Centres.
85. Related services: DEPATISnet Premium is offered by the DPMA to patent information centers and other select institutions and includes additional functionalities such as search histories and batch downloading of patent documents in PDF format.

(vi) Hong Kong, Special Administrative Region of China

86. The Intellectual Property Division of the Government of the Hong Kong Special Administrative Region offers the possibility to search published standard patent applications, granted standard patents and short-term patents with two search schemes, namely Simple and Advanced search. The **Simple search** allows the user to search within patent applications, granted or short-term patents, to carry out search by Hong Kong application/publication/patent number or even within a range of these numbers, applicant's name, title of invention, address for service or Hong Kong filing date. The user may use the operators "contain word", "exact word" or "part word".
87. The **Advanced Search** interface allows a more complex search allowing the user to search besides other search criteria also the publication/request to record date, Hong Kong Granting Patent or Designated Patent (DP) Filing date/publication number, as well as search in Abstracts (limited to short-term patents) and to priority date or inventor's name.
88. The results are displayed either in "**basic results**" or "**main results**" view, with publication number and invention title and hyperlink to the patent document specification in the first case and application/publication/DP publication number, applicant/proprietor name, title, status, maintenance due date and specification link in the second case.

(vii) India

89. Through its Indian Patent Information Retrieval System (IPAIRS, formerly IPIRS) search system, the Office of the Controller General of Patents, Designs and Trademarks of the Government of India provides separate search and access facilities for published Indian patent applications and granted patents. Full patent specifications can be accessed in text format for recently granted patents, while older granted patents can be retrieved in PDF image format for older granted patents.
90. For both patent applications and granted patents, Quick Search and Advanced Search interfaces are provided. The **Quick Search interface for patent applications** enables searching for a single term in one or more data fields selected by the user, including application and journal number, date of publication, applicant and inventor name, title, abstract, and IPC classification. The **Quick Search interface for granted patents** further allows the addition to the searched fields of patent number, date of grant, patentee name, and date and country of priority. The **Advanced Search interfaces** for both patent applications and granted patents allow the combining of two search terms using Boolean operators (selected together with search fields from drop-down menus). The Advanced Search interface for granted patents currently does not permit the searching of the

additional fields regularly permitted using the Quick Search interface. By contrast, the Advanced Search interface for patent applications additionally allows filtering of search results according to the office of filing (Delhi, Mumbai, Kolkata, and Chennai).

(viii) Israel

91. The Israel Patent Office enables searching of bibliographic data (including abstracts from 1970 to 2000) and access to full document images in PDF format for Israeli patent documents through its online patent database service. Legal and administrative status information can also be retrieved using the database service.
92. The service provides basic and Advanced Search interfaces both in English and Hebrew. Searches can be carried out in both English and Hebrew in all available bibliographic data fields. The **basic search interface** allows searching in a single data field, including application, international application, and publication number, reference, related patent or priority data. The **Advanced Search interface** further permits searching by title, applicant and inventor names, classification, application and priority dates, primary and secondary status (“new application”, “application abandoned”, “application rejected”, “patent cancelled”, “patent expired”, and others). Limited use of Boolean operators is possible using the Advanced Search interface.
93. Documents can be selected from the results list for a particular search, and marked documents can be opened for viewing in a **custom results list** or prepared for **batch printing**. For specific patent applications, the **related patent documents** can be retrieved through the record view.

(ix) Japan

94. The patent collection of the Japan Patent Office can be searched and accessed using the Industrial Property Digital Library (IPDL), operated by the National Center for Industrial Property Information and Training (INPIT). The collection comprises granted patents from 1996, published patent applications and utility models from 1971, examined patent applications from 1922 to 1966, and patent specifications from 1885 to 1956. Legal status information is also available for applications filed from 1990.
95. A range of different interfaces are available for the IPDL, including a Quick Search, Front-Page Search, Gazette Text Search, PAJ Search, FI/F-Term Search and number search interface. The latter three interfaces are available both in English and Japanese, while the other interfaces are available in Japanese only. A **FI/F-term search tool** (“Patent Map Guidance”) is also provided with both English and Japanese interfaces to allow the identification of FI/F-term symbols (Japanese classification) using IPC concordance or keyword search.
96. **Quick Search** enables simultaneous search in title, abstract, claims, etc, with the option to retrieve only those documents containing all entered search terms or to retrieve documents containing any entered search term. **Front-Page Search** includes the possibility to search up to three terms (all of which must be found in retrieved documents) and limit the results to a specific publication date range and according to a given set of IPC symbols. **Gazette Text Search** allows users to select the type of document to be searched and to flexibly combine a wide range of different search fields including classification (IPC and FI/F-terms), title, abstract, description, and claims using Boolean operators. Automatic word stemming is carried out, adding various word combinations sharing a common word stem with the entered search term to the search. Using **FI/F-Term Search**, document type, F-term theme, and publication date range can be selected as well as FI/F-terms/facets to be included in the search, the latter of which may be combined using Boolean operators. Legal and procedural status information can be retrieved from the record view for any of the four interfaces.

97. **PAJ Search** is carried out on the basis of the Patent Abstracts of Japan (PAJ), translations into English of bibliographic data and abstracts for patent applications filed after 1976. Patent Abstracts of Japan can be searched according to applicant name, title, and abstract and limited to a specific range of publication dates and set of IPC symbols. Legal status information can be retrieved in English for patent applications included in the Patent Abstracts of Japan.
98. **Number Search** facilitates direct access to specific patent documents by way of their publication numbers. A **concordance tool** is provided to assist users in identifying publication numbers required to retrieve patent documents using the number search interface by entering the appropriate patent application number, examined or unexamined document number, or registration number.
99. A **machine translation tool** is made available to translate patent document text available only in Japanese into English, including description and claims of applications for documents included in the Patent Abstracts of Japan, for which human translations are only available for bibliographic data and abstracts.

(x) New Zealand

100. The Intellectual Property Office of New Zealand offers search of bibliographic data (including abstracts) and access to full document images of New Zealand patent documents published from 1900 to the present through its online database service. Legal and procedural status information, including data regarding renewals, objections, and licensing, can also be retrieved for certain documents.
101. The service offers a single structured search interface to unregistered users. Through the interface, all fields can be searched simultaneously, or individual fields can be searched separately. Searchable fields include various names, ID numbers, dates (including expiry information), classifications (IPC and NZ), keyword fields, and document types. **Case-sensitivity** can be optionally selected for all fields. **Legal and procedural status information is also searchable.** A wide variety of Boolean, proximity, wildcard, and special operators can be used, however only within a given search fields and not across different fields, which are therefore all included in the search by default. To assist users in structuring their queries, **query builders** are provided for several fields. **Searchable applicant/assignee, inventor, and agent/contact name indices** including address information in certain cases are also provided to help users locate the desired person or corporate entity by entering at least three letters from the beginning or elsewhere in the name. Results can be selected for individual or **batch viewing and delivery.** Delivery can be made by email or fax free of charge.
102. Related services: Registered users have access to an additional **advanced search interface**, allowing them to make command-line queries using field names and a wider set of search operators. They may also set preferences, save queries, and access search histories.

(xi) Republic of Korea

103. The Korea Intellectual Property Rights Information Service (KIPRIS) provides two search options of its Patent and utility model database: General and Advanced search. Through the **General Search**, the user can perform a keyword search choosing if he wishes to search for patents and/or utility models, limit the results to a certain search period or administrative status and also use the option of Multilanguage search (English/Korean), while an extra section in the General search enables a number search through selection of the patent/utility model type (application / publication / unexamined publication / registration of patent or utility model) and date and relevant number. In case that a multilingual search is enabled, the system simply translates the English search keyword and then searches the Korean translation within the Korean full-text.

104. The **Advanced Search** on the other hand enables search in the above-mentioned fields, along with priority date, applicant/inventor/agent/patentee name and combine the search fields using Boolean operators “and” and “or”.

(xii) United Kingdom

105. In addition to fee-based “expert search” services, the Intellectual Property Office of the United Kingdom also offers a free-of-charge database service, permitting searching of patent applications and granted patents filed in the United Kingdom by: patent data, publication data, and supplementary protection certification (SPC) number. Patent documents for the United Kingdom are also accessible through the esp@cenet search service (see Section 2. (e) (xv)), though SPC number searches can only be carried out using the UK IPO database service.
106. Through the **Patent Number Search** the user can carry out a search by application or publication number (for application filed after 2007) and have information on the **patent status** delivered, either via the “full details” button for a complete status history or the “Current Details” button for current status. The status of European patents designating the United Kingdom may not be current, though, and can only be accurately determined based on information provided through the EPO Register.¹³ The **Publication Number Search** enables search by publication number or date and allows a search limitation to specific document type (application, specifications etc). It should be noted that this search only covers publications from January 3, 2007 to date. Furthermore, the **SPC Search** allows search by Patent or SPC number.

(xiii) United States of America

107. The United States Patent and Trademark Office offers access to granted patents through its PatFT database service and to published patent applications from March 15, 2001 through its AppFT database service. Patents granted prior between 1790 and 1975, available through the PatFT service, are only searchable by patent number, issue date, and current US patent classification (USPC), while all other documents in the PatFT and AppFT services are searchable in full-text. Text records do not include drawings; however, full document images in TIFF format including drawings can also be retrieved through the PatFT and AppFT services.
108. Both services offer Quick and Advanced Search interfaces, while the PatFT service additionally offers a **patent number search interface** and the AppFT service a **publication number search interface**. The **Quick Search interface** allows searching in two separate data fields, which can be selected by the user and combined using Boolean operators. Right truncation is available using wildcard operators. The **Advanced Search interface** allows command-line search syntax (a list of the allowed field codes is displayed underneath the query fields) along with a keyword search. For both the Quick and Advanced Search interfaces, the PatFT service also permits the selection of the period to be searched (from 1976 or from 1790). Search terms are highlighted in the search results.
109. Related services: **Legal and procedural status information** can be retrieved through the Patent Application Information Retrieval (PAIR) service using application, control, patent, PCT, and publication numbers. **Assignment information** can also be retrieved through a separate Patent Assignment Database service. An **IPC-USPC concordance tool** is provided to assist in the identification of USPC symbols based on IPC symbols and vice versa. The USPTO also offers a **sequence search** scheme through the Publication Site for Issued and Published Sequences (PSIPS) providing sequence Listings, tables, and other mega items for granted US patents or published US patent applications, available for

¹³ <https://register.epoline.org/espacenet/regviewer>

viewing and download. The users may access this information through the proper document detail page and then submitting a SEQ ID NO or a mega table ID number in the relevant search field.

(xiv) Eurasian Patent Organization

110. The Eurasian Patent Organization (EAPO) offers access to: (i) the **EAPO database** of published Eurasian applications and patents with bibliographic data and abstracts for patents in English, bibliographic data and abstracts for applications in Russian, full description of applications and patents in Russian and (ii) **EAPATIS**, with records from EAPO, WIPO, EPO, USPTO, Soviet Union, Russia and CIS countries, and countries included in the "PCT minimum documentation". Over 40 local databases (DB) are supported with access to world, regional and national collections of patent documentation. The Russian language collection consists of patent documentation of EAPO, Russia, national patent offices of the Eurasian region including documentation from Eurasian Patent Convention Contacting States. The following DBs are presented in English: Eurasian patents (since 1997), patent documentation from the CISPATENT data¹⁴ (Armenia, Belarus, Georgia, Kyrgyzstan, Moldova, Russia, Tajikistan, Ukraine and Uzbekistan, and the Eurasian Patent Office), patent documentation of countries included to the PCT-minimum documentation (WIPO, EPO, USA, Japan, Canada etc.), while the Eurasian version of esp@cenet provides access to the esp@cenet database with English and Russian as possible search languages.
111. The user may have access to published Eurasian applications and patents via the EAPO database only by using the application or patent number search option. The results are often displayed in Russian with a link to esp@cenet, where the abstract in English can be viewed.
112. The EAPATIS system on the other hand can be accessed in two ways: **free-of-charge** or on a **fee base**. In both cases the user may carry out an **advanced**, a **number** (quick) or a **professional search**, while the fee-based access provides the possibility of a metasearch, allowing search in esp@cenet, epoline, Patentscope, USPTO and JOPAL databases. The **quick search** allows only search by patent number with country codes but without truncation. The **advanced search** on the other hand allows the user to choose EAPO-EN and/or CISPATENT databases to be searched and includes publication/application number, IPC code, date, person/firm name and/or keyword as possible search fields. The result includes link to the front page and/or description PDF file, as well as links to Delphion (for users who are Delphion subscribers), esp@cenet and the Register.
113. The free-of-charge access covers the Eurasian patents and the CISPATENT databases. The Eurasian patent database contains bibliographic data and abstracts in English and the front pages (in Russian), full texts of the descriptions of inventions in Eurasian patents (in Russian). The CISPATENT database is a joint product of the Commonwealth of Independent States and the Eurasian Patent Office (EA) and contains bibliographic data and abstracts in English and full descriptions of patent documents in original languages.
114. The "**requests analyzer**" offers assistance to the users as it provides them with messages elaborating possible query mistakes and making proposals in order to optimize the search results, while the preview button allows seeing in advance the search query and altering it if necessary before proceeding to the search results. The query results contain

¹⁴ Including data from the Eurasian Patent Office(EA) and Commonwealth of Independent States (CIS): Armenia (AM), Belarus (BY), Georgia (GE), Kazakhstan (KZ), Kyrgyzstan (KG), Moldova (MD), Russia (RU),Tajikistan (TJ), Ukraine (UA), Uzbekistan (UZ).

abstracts and bibliographic descriptions of the patents and further data when IPC data is searched.

115. Access on a paid basis is open to all databases of the system in advanced, number and professional modes, as well as metasearch. Front pages and full descriptions of Eurasian patents and published Eurasian applications are presented. The features of zero results analysis and the request analyzer should be mentioned, providing help messages and recommendations in case an error is found in the query or zero results are displayed.

(xv) European Patent Office

116. Esp@cenet is a free patent search service offered by the European Patent Office. Its content is based on the INPADOCDB bibliographic data set and INPADOC legal status data set, currently covering over 80 and over 50 patent authorities, respectively.¹⁵ In addition to allowing searching in bibliographic data for the aforementioned collections, esp@cenet also permit full-text search of European and PCT applications in English, German, and French.
117. With regard to the search schemes four search options are offered, allowing the user to choose among Worldwide, WIPO or EP databases:
118. The **Quick Search** enables search by keyword, choosing whether to search in title/abstract or for persons/organizations. Full-text search is available for the PCT and European patent collections.
119. **Advanced Search** allows keyword search in title/abstract, search by publication/application/priority number, publication date, applicant/inventor's name and/or ECLA or IPC classification. Full-text search is available for the PCT and European patent collections. Searches must be carried out in English, so that only results with titles or abstracts in English will be retrieved (excluding patent documents from the results for which the abstracts have not been translated, such as patents before 1970). A maximum of four search terms may be entered in each field, and a total of 21 search terms and 20 logical operators within the search mask as a whole. With regards to the publication date, no wildcard characters can be accepted.
120. **Number Search** carries out search by application, publication or priority number with or without country code prefix. The two first characters of the search inquiry should be the country code. Non-patent literature (NPL) reference number search is also possible for the XP documents, namely the NPL documents classified by the ECLA classification system and/or cited in the EPO search report. It should be noted that the NPL reference number is the only way to retrieve such information.¹⁶
121. **Classification Search:** this search option allows keyword search in particular fields of technology showing the relevant ECLA classes/subclasses and allowing thus a refined search by ECLA classes, as the selected results may be copied and used for search in the Advanced Search form. The classification search results may include a number of embedded images.
122. **Smart Search:** The smart search enables simple/multi-word as well as complex searches and it also allows command line searches.
123. After choosing which type of search to carry out, the results – the first 500 - are **displayed** with the keyword(s) highlighted. 20 of them can be saved in “my patents list” and be

¹⁵ For full information on data coverage, see:

[http://documents.epo.org/projects/babylon/eponet.nsf/0/2464E1CD907399E0C12572D50031B5DD/\\$File/global_patent_data_coverage_0110.pdf](http://documents.epo.org/projects/babylon/eponet.nsf/0/2464E1CD907399E0C12572D50031B5DD/$File/global_patent_data_coverage_0110.pdf)

¹⁶ http://ep.espacenet.com/help?locale=en_EP&method=handleHelpTopic&topic=npl

retrieved –if cookies are enabled- for a year. The claims tab may include a “**claims tree**” which provides a graphical presentation and hierarchy of dependent and independent claims.

124. The document contents can be translated into German, French, Swedish, Portuguese, Italian and Spanish using the EPO’s own machine translation system, while non-Latin character text can be translated into English. An INPADOC legal status tab is also available while more information can be retrieved through the provided link to the European Patent Register. **Register Plus** allows access to all publicly available data related to European patent applications until the grant procedure, namely administrative and legal data on EP, Euro-PCT and PCT filings.
125. **Drawing mosaics** are offered allowing quick review scanning of all drawings, while batch document download and printing are possible via the EPO’s Open Publication Server (OPS).

3. REVIEW OF NON-PATENT LITERATURE DATABASE SERVICES

(a) Introduction

126. Non-patent literature (NPL), as the term suggests, consists of any documents that are not patents. In the context of the international patent system, it generally refers to scientific and technical literature and represents an important element of the prior art used to determine the novelty of an invention. The European Patent Office has found that, on average, 15 percent of citations in search reports are derived from NPL, while 3 percent of search reports contain only NPL citations. In fact, in certain fields of technology, NPL encompasses a larger part of the prior art than patent documents. For example, in biochemistry, the EPO has found that over 60 percent of examiner citations are derived from NPL sources.¹⁷
127. NPL describes a wide range of different publications, a few examples of which are depicted in Figure 1.

Figure 1. Non-patent literature



Source: CGIAR, "CGIAR centre publications as prior art",
<http://www.cas-ip.org/projects/cgiar-centre-publications-as-prior-art/>

128. Like patent documents, books, journals, conference proceedings, and other types of publications that constitute NPL have become far more easily accessible as a result of their increasing availability in electronic form through the Internet. Nonetheless, locating and accessing relevant NPL prior art remains a challenge as a result of the diversity of channels through which these scientific and technical publications originate and are distributed. Indeed, whole new sources of NPL prior art such as webpages and blogs have emerged with the advent of the Internet as a means of publication.

¹⁷ Yves Verbandt and Emmeline Marttin, "Search in Non-Patent Literature (NPL)", Presentation at Search Matters 2009 (30-31 March 2009).

129. Sources of NPL include:
- Publisher databases;
 - Library databases;
 - Government databases;
 - Trade association databases;
 - College and university databases; and
 - Personal websites of researchers.
130. Across these many sources of NPL, documents also lack the large degree of uniformity in structure present in patent documents, further compounding the challenges faced in searching scientific and technical literature.
131. Furthermore, in spite of an emerging trend towards free access and open access in scientific and technical publishing, most NPL publications can only be accessed on a fee-paying basis. Thus, regardless of the accessibility of NPL documents, organizations and individuals involved in the patenting process, from inventors and applicants to patent examiners, must be selective in their use of NPL as a source of information. To address this issue, which disproportionately affects patent offices, individual inventors, universities, and SMEs in developing countries, the World Intellectual Property Organization has launched the Access to Research for Development and Innovation (aRD_i) program in cooperation with several major scientific and technical publishers. This program, discussed in greater detail in Section 3 (h) is designed to facilitate access to scholarly journals for patent offices and academic and research institutions in developing countries in particular least developed countries (LDCs).

(b) Structure and methodology

132. To provide an overview of content types and features available across different services for searching and accessing NPL, this Study Paper reviews a selected group of online search services and journal databases. The services summarized here include some of the most widely used search services and journal databases, as identified through Circular C.N 3024. Nonetheless, they are not intended to serve as an authoritative list of NPL services but rather to illustrate characteristics common to various services. The Study Paper also outlines features of selected content sets searchable or accessible through a number of the services reviewed here such as the MEDLINE content set.
133. For the purposes of the present Study Paper, NPL is defined to exclude business methods and traditional knowledge. Though publications in these fields undoubtedly contribute to the prior art, business methods are patentable only in a limited number of jurisdictions, while the complexity of traditional knowledge as a source of data suggests a separate, dedicated study of the subject.

134. Though substantial overlap exists between the features available through search services and journal databases, these types of services are distinguished here for purposes of clarity and comparability. Generally speaking, journal databases (which may also contain other types of publications such as books or conference proceedings) lack the sophisticated search and analysis features available in search services, while search services frequently do not provide the possibility to access documents identified through searches in their full-text version unlike journal databases.

Search services (free)	
Google Scholar	http://scholar.google.com
Scirus	http://www.scirus.com
Entrez	http://www.ncbi.nlm.nih.gov/Entrez
PubChem	http://pubchem.ncbi.nlm.nih.gov
PubMed	http://www.ncbi.nlm.nih.gov/pubmed
PubMed Central	http://www.ncbi.nlm.nih.gov/pmc
Search services (commercial)	
Dialog	http://www.dialog.com
Scopus	http://www.scopus.com
STN	http://www.stn-international.de
Thomson Innovation	http://www.thomsoninnovation.com
Journal databases (free)	
Directory of Open Access Journals	http://www.doaj.org
SciELO	http://www.scielo.org
Journal databases (commercial)	
American Chemical Society	http://pubs.acs.org
IEEE Xplore	http://ieeexplore.ieee.org
ScienceDirect	http://www.sciencedirect.com
SpringerLink	http://www.springerlink.com
Wiley InterScience	http://www.interscience.wiley.com

(c) Database content sets

135. Certain content sets are common to a number of different search services. These content sets include: Inspec, the IP.com Prior Art Database, MEDLINE, and Embase.

(i) Inspec

136. Inspec is a bibliographic database published by the Institution of Engineering and Technology (IET), with a particular focus on physics, computer science, and engineering. Inspec contains more than 11 million records covering over 4,000 journals, around 2,200 conference proceedings, as well as books, reports and dissertations from 1969. An archive file extends the historical coverage to the period from 1898 to 1968. Records in Inspec are classified according to a proprietary hierarchical classification scheme but also according to the IPC (through a concordance with the Inspec classification scheme). Controlled index terms are also used to represent inorganic substances and materials, astronomical objects, and numbers found within records and thus address the difficulties posed by synonymous representations of identical concepts (e.g. the number one thousand being represented as 1000, 10³, 1.0E+3, etc.). Inspec content is incorporated into database services such as Questel Orbit and Thomson Innovation.

(ii) IP.com Prior Art Database

137. The IP.com Prior Art Database (often referred to simply as "IP.com") is a full-text database that incorporates a number of corporate disclosure publications such as the IBM Technical Disclosure Bulletin, the Siemens Disclosure Journal, and the Xerox Disclosure Journal as well as similar publications from Microsoft, Motorola, and Sony. Its content also includes disclosures from the US National Institute of Standards and Technology as well as

disclosures from various parties made through the US Patent and Trademark Office (“statutory invention registrations”) and the Internet Society (“requests for comments”). The historical coverage varies according to the specific dataset. The Xerox Disclosure Journal constitutes part of the PCT NPL Minimum Documentation. Content from the IP.com Prior Database is included in such services as Orbit.com.

(iii) MEDLINE

138. MEDLINE is a bibliographic database compiled by the National Institutes of Health (NIH) National Library of Medicine. Its content encompasses over 16 million records and covers over 5,200 journals in the fields of “life sciences, behavioral sciences, chemical sciences, and bioengineering”,¹⁸ the large majority of which are published in English. The journals indexed in MEDLINE date from 1963 to the present, though content from 1950 to 1963 is also available through the associated OLDMEDLINE database. Over 80 percent of MEDLINE records include English-language abstracts. MEDLINE incorporates the MeSH (Medical Subject Headings) hierarchical classification scheme, which encompasses over 25,000 symbols (“descriptors”). MEDLINE content is primarily disseminated by the NIH Center for Biotechnology Information through its PubMed database service but also through other sources such as Thomson Innovation and the STN database services.

(iv) Embase

139. Embase is a bibliographic database provided by Elsevier focusing on biomedicine and pharmacology. Its content includes more than 20 million records covering over 7,300 journals and book series from 1974, published in a variety of languages, notably English and Chinese. Content from 1943 to 1973 is available separately through the Embase Classic database. Around 80 percent of Embase records (90 percent of Embase Classic records) include abstracts. There is significant overlap between the content set of Embase and MEDLINE (approximately 60 percent of journals indexed in Embase are also contained in MEDLINE), particularly in certain subject areas.^{19,20} Records contained in Embase are classified according to the Emtree hierarchical classification scheme, which includes 56,000 symbols (“preferred terms”) and fully incorporates the MeSH classification scheme. Embase content is incorporated, e.g., into the Dialog and STN database services

(d) Free-of-charge search services

(i) Google Scholar

140. Google Scholar offers a **basic search interface** and an **advanced search interface**, both of which allow a range of common search operators. The **advanced search interface** distinguishes itself from the basic search interface by the option to search in specific fields and filter results according to language and subject area. The coverage of the search carried out by Google Scholar varies according to subject area²¹ and is thought to have a bias towards English language content.^{22,23} Generally speaking, the **coverage** includes

¹⁸ United States National Library of Medicine, “Fact Sheet: MEDLINE®”, <http://www.nlm.nih.gov/pubs/factsheets/medline.html> (Accessed 12 January 2010).

¹⁹ Elsevier, “What are the differences between MeSH and Emtree?”,

http://www.info.embase.com/UserFiles/Files/Embase_ementree_mesh_.pdf (Accessed 12 January 2010).

²⁰ P. Royle, L. Bain, L. Payne, M. Hill, and N. Waugh, “The value of literature searching beyond medline and embase for systematic reviews of interventions and epidemiology in diabetes”, Health Technology Assessment International. Meeting (2004: Krakow, Poland)

²¹ Virginia Wilson, “A Content Analysis of Google Scholar: Coverage Varies by Discipline and by Database” Evidence Based Library and Information Practice, Vol. 2, No. 1 (2007)

²² Chris Neuhaus et al., “The Depth and Breadth of Google Scholar: An Empirical Study” Libraries and the Academy Vol. 6, No. 2 (April 2006): 214-219.

²³ Philipp Mayr and Anne-Katrin Walter, “Studying Journal Coverage in Google Scholar” Journal of Library Administration Vol. 47, No. 1&2 (January 2008): 81-99.

“peer-reviewed papers, theses, books, abstracts and articles, from academic publishers, professional societies, preprint repositories, universities and other scholarly organizations.”²⁴

141. **Searching by field** is possible according to title, author name, publication (source title), and date. Since the fields are defined probabilistically, field searching is not error-proof (i.e., data may be considered erroneously to belong to a particular field, e.g., the page number 1945 may be interpreted as the date of publication).
142. **Filtering by language** is possible in the following languages: Chinese (traditional), Chinese (simplified); German, English, French, Japanese, Korean, Portuguese, and Spanish. **Filtering by subject area** is possible for the following areas: Biology, Life Sciences, and Environmental Science; Business, Administration, Finance, and Economics; Chemistry and Materials Science; Engineering, Computer Science, and Mathematics; Medicine, Pharmacology, and Veterinary Science; Physics, Astronomy, and Planetary Science; Social Sciences, Arts, and Humanities. Evidence points to the fact that these subject areas are not entirely distinct from one another in retrieved results (i.e., results may be found in multiple subject areas).
143. Both the basic and advanced search interfaces are available in **multiple languages**, including Arabic, Chinese (traditional), Chinese (simplified), English, French, German, Japanese, Korean, Portuguese, and Spanish (not Russian). The interfaces permit the use of Boolean operators and phrase truncation (effectively a type of unordered proximity operation where an asterisk is used as a wildcard operator to represent each word separating the given search terms). They also support **semantic searching** through the addition of a tilde (~) before each term for which synonyms and other related terms are to be searched.
144. Results are sorted by relevance, which is determined “weighing the full text of each article, the author, the publication in which the article appears, and how often the piece has been cited in other scholarly literature.”²⁵ From the results list, users can find additional content relevant to their search by following links below each title to: documents in which a particular result is cited (“**cited by**”); and similar articles (“**related articles**”). Citations can be imported using the “**Bibliography Manager**” in a number of formats, including BibTex, EndNote, RefMan, RefWorks, and WenXianWang.
145. Google Scholar only displays an extract from relevant documents in its results list and does not itself provide access to full-text content. However, for certain participating institutions, “**Library Links**” can be displayed in the results list, which highlight those documents to which users have authorized access through their institution. This feature can also highlight documents available in hard copy at local libraries, provided that the libraries list their holdings in the Online Computer Library Center’s Open WorldCat, or available for purchase in electronic copy through the British Library (“BL Direct”). Through the results list, **additional versions** of documents can also be accessed, which may include versions freely available in full-text (e.g., versions published on authors’ websites). However, these versions may differ from the version cited in the results list (e.g., with the peer-reviewed published version appearing in the results list and pre-print manuscripts appearing as additional versions).

²⁴ Google, “About Google Scholar”, <http://scholar.google.ch/intl/en/scholar/about.html> (Accessed 12 October 2009)

²⁵ Ibid.

(ii) Scirus

146. Scirus is a free search service provided by Elsevier. It covers over 350 million webpages determined to be “science-related” as well as a selection of additional sources including journal databases and content repositories. It also covers a range of patent data through Lexis-Nexis. The webpages searched include: “university pages (domains ending with .edu, .ac.uk, and educational sites within other countries); scientific organizations (.org) and conference pages; company pages with scientific R&D information or information relevant to researchers (.com); government pages dealing with science, including health and sciences like law (.gov); scientist and author home pages.”²⁶
147. It offers a **basic search interface** and an **advanced search interface**, both of which permit the use of common search operators and field searching (using field codes in the basic search interface). The advanced search interface additionally provides for the possibility to filter results according to date, information type, file format, content source, and subject area. **Filtering by information type** allows users to include or exclude abstracts, articles, books, company homepages, conferences, patents, preprints, scientist homepages, and theses and dissertations from their search results. **Filtering by file format** permits users to select whether the results are to be in any format, PDF, HTML, Word, Powerpoint, TeX, or PostScript format. By **filtering by content source**, users can choose to include the entire range of content searched by Scirus or pick from a pre-defined list of selected sources. The areas for which **filtering by subject area** is available include: Agricultural and Biological Sciences; Astronomy; Chemistry and Chemical Engineering; Computer Science; Earth and Planetary Sciences; Economics, Business and Management; Engineering, Energy and Technology; Environmental Sciences; Languages and Linguistics; Law; Life Sciences; Materials Science; Mathematics; Medicine; Neuroscience; Pharmacology; Physics; Psychology; Social and Behavioral Sciences; and Sociology.
148. Results can be sorted according to relevance (determined on the basis of keyword and citation frequency) or date. Each result is linked to **similar results**, determined on the basis of keywords from the search as well as keywords extracted from the search result itself. **Additional filtering** can be carried out from the results list, according to content source and file type. Furthermore, results can be refined by selecting additional terms proposed on the basis of the initial search. Finally, citations, abstracts and keywords from selected results can be emailed, saved, or exported in RIS or text format.

(iii) Entrez

149. Entrez is the portal to the database services made available by the National Institutes of Health (NIH) through its National Center for Biotechnology Information (NCBI). These database services include PubMed, PubMed Central, and PubChem as well as protein, gene and genome sequence and other types of databases. Entrez also provides links to books and journals indexed by NCBI and to the MeSH classification system.
150. Entrez offers a **command-line search across all NCBI databases**. The cross-database search permits the use of basic **Boolean operators, phrases, and nesting** as well as **right truncation** and **searching by field** in a wide range of data fields. Numbers of results for each database are shown, while results can be directly accessed from the individual databases. From the results for individual databases, users can **refine search results**. Additional search terms can be identified and added directly from the **database index**. The individual databases also generally allow users to **view and save the session history** and

²⁶ Scirus, “About Us”, <http://www.scirus.com/srsapp/aboutus/> (Accessed 13 October 2009)

combine results from multiple searches in their session history. Selected **records can also be exported in XML format**.

151. NCBI allows users to access its databases using their own **customized applications** through a set of server-side programs (“Entrez Utilities”, or “eUtils”). The programs permit retrieval of summary data or document IDs related to a particular search and batch downloads of documents, among other things.

(iv) PubChem

152. PubChem encompasses three interlinked databases, namely the **PubChem Substance**, **PubChem Compound**, and **PubChem BioAssay** databases. The PubChem Substance database contains chemical structure information submitted by numerous sources including government agencies, universities and research centers, and private companies. The PubChem Compound database clusters the information submitted to the PubChem Substance database according to structural similarity, grouping information related to identical chemical structures into single records. The PubChem BioAssay database contains biological activity information submitted for chemical substances included in the PubChem Substance database as well as information on the experimental conditions under which the biological activity data was collected. All three databases also include links to the respective data sources, from which additional substance or biological activity data can be obtained.
153. All three databases offer basic and advanced command-line search interfaces as well as a chemical structure search interface and biological activity (“BioAssay”) search interface. Both the basic and advanced command-line search interfaces permit the use of basic Boolean operators and searching by field (where the available fields are determined the database in which the search is carried out). The advanced command-line search interface provides assistance in structuring more complex queries by providing structured search fields (“limits”) as well as a query builder. The chemical structure search interface enables searching by synonyms, MeSH classification, and chemical properties and identifiers (“descriptors”) as well as molecular formula. It also allows users to draw chemical structures or enter CID, SMILES, or InChI text-based representations of substances and carry out identity or similarity searches using these structures or text representations. The biological activity search interface permits searching by single or multiple substances and targets using synonyms, text representations, and other identifiers as well as by assay ID (AID) number. Search histories are automatically generated and can be accessed for up to eight hours after the search has been carried out or they can be saved to user profiles.
154. Results can be displayed in summary format or other formats such as abstracts view and can be sorted according to data submission or modification date, data source, and substance, compound, and assay ID numbers as well as chemical properties such as molecular weight (Substance and Compound) or bioassay method (BioAssay). Results **can be filtered** according to presence of bioassay information, pharmacological activity or “BioSystem”, and depositor category (Substance and Compound) or according to target, bioassay method, related bioassay, and active substance (BioAssay). They can also be filtered for druglikeness according to Lipinski’s Rule of Five (Substance and Compound) or type of bioassay, e.g. primary or confirmatory (BioAssay). For all records, links are given to information available from the other PubChem databases (e.g. from Substance to BioAssay data) as well as to external sources such as PubMed Central or organizations from which the data was originally obtained. **Results can be exported** in a variety of formats including XML format from all three databases.

(v) PubMed

155. PubMed incorporates bibliographic data from 5,200 journals in health-related fields including biochemistry, biomedical engineering, botany, nanotechnology, and various fields of medicine. Its content is largely derived from the MEDLINE content set and covers all

articles contained in the PubMed Central full-text database. In total, it contains over 19 million records, of which 16 million records belong to the MEDLINE content set.

156. PubMed offers a basic and advanced command-line search interface, of which both permit the use of basic Boolean operators and searching by field. The advanced command-line search interface allows the use of structured search fields and search filters (“limits”). The filters enable users to limit their results, for example, to records with links to the full-text of the associated article or to those with links to free full-text. The advanced command-line interface also permits **direct access to the search indexes** for each data field, thus facilitating the identification of relevant search terms. Searches carried out using both the basic and advanced command-line search interfaces automatically are expanded as **semantic searches**. Additionally, PubMed makes available a structured search interface permitting searching of bibliographic data (“citation matcher”) and several prestructured searches (“topic-specific queries”).
157. Results are sorted by default by date added but can also be sorted by publication date, last name of the first or last author, journal title, and article title. Results also indicate the **availability of full articles** at no cost through PubMed Central. From the results display, **additional filtering** can be carried out using search criteria suggested on the basis of the initial search and related documents available in other NCBI databases including PubChem can be identified. Users can establish **profiles** (“MyNCBI”) to save session histories or customized search filters.

(vi) PubMed Central

158. PubMed Central is a document repository focusing on health and life sciences. Through PubMed Central, the NIH National Library of Medicine provides free full-text access to over 2 million articles from around 1000 scholarly journals including a significant number of journals belonging to the PCT NPL Minimum Documentation. Not all journals include their full archives back to their earliest issues. Articles from certain journals are also subject to an embargo period of up to 24 months before they become available through PubMed Central. However, once they appear in PubMed Central, many journals are made available under limited or full open-access conditions, that is to say, that they “are made available under a Creative Commons or similar license that generally allows more liberal redistribution and reuse than a traditional copyrighted work.”²⁷
159. PubMed Central provides basic and advanced command-line search interfaces similar to those found in PubMed, with automatic semantic search, though the searchable fields differ between the two services. Searches previously carried out can be combined to provide a single results set.
160. Results can be sorted by print or electronic publication date, journal title, or their date of addition to PubMed Central. From the record view, users **download full-text articles in PDF format** and view and access **citing documents** and **related articles**. Search sessions can also be tracked and stored to user profiles (“My NCBI”).

(e) Fee-based search services

(i) Dialog

161. Dialog is an online database service operated by ProQuest, which encompasses around 600 databases covering subject areas including “Agriculture & Nutrition, Chemistry, Energy& Environment, Medicine & Biosciences, Pharmaceuticals [and] Engineering &

²⁷ PubMed Central, “PMC Open Access Subset”, <http://www.ncbi.nlm.nih.gov/pmc/about/openftlist.html> (Accessed 23 January 2010).

Technology.”²⁸ Notable databases include **Beilstein Abstracts and Reactions (chemistry)**, **EMBASE (life sciences)**, and **Inspec (engineering)**. These databases contain bibliographic data and, in certain cases, abstracts and full-text from scholarly journals, technical reports, trade publications, newspapers and other sources as well as patent data. In total, the database contains over 470 million documents.

162. **DialogWeb** provides a menu-based (“one-click”) and a command line-based interface for searching the Dialog database content. Both interfaces support basic Boolean operators, proximity operators, and wildcard operators, which can be used for non-simultaneous right and internal truncation. They also permit searching by field in a limited range of fields across all databases and a more extensive set of fields for specific databases. Details on the characteristics of each database including searchable data fields are recorded in so-called “blue sheets”.²⁹ The menu-based interface allows users to select the databases in which to search either by entering the numbers assigned to these databases by Dialog or by narrowing the range of databases according to their subject area. Once the range of databases to be searched has been defined, users can either carry out a search using a number of pre-defined structured search interfaces with fixed search fields (“**targeted search**”) or using a structured search interface with a user-selectable search field and date range (“**dynamic search**”). Unlike the targeted search interfaces, the dynamic search interface permits the use of Boolean, proximity and wildcard operators in the defined search field. Additionally, the dynamic search interface enables a similarities search for author names, company names, journal names, or product names indexed by Dialog and, in specific databases, similarities search by subject term (“thesaurus”). Results from a dynamic search can be refined by adding further search criteria to the initial search. Searches carried out using either the targeted and dynamic search interfaces can be saved to the user’s profile or as user-customizable alerts. The command-line interface requires that users select the databases to be searched according to database code and supports the use of Boolean operators, proximity operators, and wildcards as well as field codes. Similarly to the menu-based interface, it allows searches to be saved or exported as alerts.
163. Results can be sorted according to a number of criteria including author name, document title, and document source. Results can be exported as full records or condensed lists in a variety of formats or sent to users by customizable email or by fax or US mail.
164. Other Dialog interfaces include Dialog1, a menu-based search interface, and DialogClassic and DialogClassic Web, command-line based interfaces (which allow chemical structure searches). DialogNewsroom and DialogSelect provide access to a limited selection of Dialog databases.

(ii) **Scopus**

165. Scopus is an abstract and citation database service provided by Elsevier covering a wide range of subject areas including “Agricultural and Biological Sciences; ... Biochemistry, Genetics and Molecular Biology; Business, Management and Accounting; Chemical Engineering; Chemistry; Computer Science; Decision Sciences; Dentistry; Earth and Planetary Science; ... Engineering; Environmental Science; Health Professions; Immunology and Microbiology; Materials Science; ... Medicine; ... Neuroscience; Nursing; Pharmacology, Toxicology and Pharmaceuticals; Physics and Astronomy; ... Veterinary.”³⁰

²⁸ Dialog, “Subject Category Index [All Databases]”, <http://library.dialog.com/bluesheets/html/bls.html> (Accessed 11 November 2009).

²⁹ Dialog, “Dialog Online Documentation”, <http://library.dialog.com/bluesheets/> (Accessed 11 November 2009)

³⁰ Scopus, “Content Coverage”, http://www.info.scopus.com/docs/content_coverage.pdf (Accessed 17 November 2009).

Regardless of the original language of the document, all records included in Scopus must also have an English version of the title and abstract, where available. The journal in which documents appear must also be published at least once yearly and must be subject to some type of peer-review process. In total, Scopus covers around 38 million documents, of which around half date from the period between 1823 and 1996 and the other half from 1996 to the present. Around 80 percent of these records include abstracts. Scopus also indexes around 435 million “scientific” web pages and 23 million patent documents from different patent offices.

166. Scopus offers four search interfaces: a basic (structured) search interface, an advanced search interface, an author search interface, and an affiliation search interface. The basic search interface supports basic Boolean operators as well as searching by field in a variety of fields including: combined article title/abstract/keywords or article title/abstract/keywords/author, authors, first author, source title, article title, abstract, keywords, affiliation, language, ISSN, CODEN publication identifier, digital object identifier (DOI), reference data, conference data, chemical name, and CAS number. It also allows results to be limited to a specified publication date range or date added as well as to a specified document type and set of subject areas. The advanced search interface supports Boolean operators, proximity operators, and wildcards, for right and internal truncation (though not simultaneously), as well as searching by field in the full range of searchable fields covered by Scopus. Searches can be saved to subscribers’ accounts or used to create email alerts or RSS feeds.
167. Search results are displayed on separate tabs for results from Scopus, the Web, patents, and “selected sources” (additional content repositories searched). They can be sorted according to date, relevance, authors, source title and number of citation. Results can be refined by adding further search criteria, and additional filtering can be carried out on the basis of source title, author name, year, document type, subject area, language, affiliation, keyword, or source type. Options for this additional filtering are extracted from the results set.
168. A major strength of Scopus is its citation tools. These tools include links from particular Scopus records to citing documents from Scopus, the Web, and US, European and PCT patent collections. Scopus also permits users to establish citation alerts to indicate when new citations are made for a particular record or author. Scopus also maintains author and affiliation indexes, integrated into its search interfaces, to allow authors or organizations to be unequivocally identified, in particular in cases of name variations for a single author or organization or in cases where a particular name or set of initials is shared among multiple authors.

(iii) STN

169. STN is an online database service jointly operated by FIZ Karlsruhe and the Chemical Abstracts Service (CAS). It provides access to a broad set of databases, each covering distinct subject areas including “Bioscience, Biotechnology, Business, Chemistry, Energy and Environment, Engineering, Health and Safety, Materials Science, Medicine, Nanotechnology, [...] Pharmacology [and] Toxicology”.³¹ The different databases contain “published research, journal literature, [...] structures, stereochemistry, trade names, physical properties, sequences, and other data.”³²

³¹ STN International, “STN Content”, http://www.stn-international.de/stn_content.html (Accessed 23 October 2009).

³² Ibid.

170. **STN Express** permits **command-line searching** in the full range of over 600 databases covered by STN, including Beilstein Abstracts, Chemical Abstracts Plus, and the Chemical Abstracts Registry (chemistry), BIOSIS (life sciences), and COMPENDEX and INSPEC (engineering). Users can opt to **include only certain databases** in their search, thus narrowing their search to particular subject areas or types of documents. STN Express supports basic **Boolean operators** and wildcard operators, which can be used for **simultaneous left, right, and internal truncation**, as well as a wide range of **proximity operators**. For certain databases an even greater range of operators is supported, including additional proximity operators and **comparison operators** (smaller than, greater than, etc.). **Searching by field** is generally possible using author name, journal title, article title, and article full-text. However, as with operators, individual databases support a much larger set of fields in which searches can be carried out, which must be determined by referring to the documentation for each database. BLAST **similarities search** is possible for nucleotide and protein sequences in the CAS Registry and gene sequence databases. For certain databases, **chemical structure search** and **chemical reaction search** is possible, using a proprietary structure and reaction drawing tool. Specific characteristics of each database including types of searches available, searchable fields, and supported operators are described on individual summary sheets.³³ Pre-defined **search query files can be imported** and **search transcripts can be exported** in order to facilitate automation of searches.
171. Results are displayed in bibliographic data format sorted by date within each database. **Results can be grouped** according to various criteria such as publication year, author name, and company name, which includes the possibility of similarities grouping. **Customized tables and reports** can be created from results and exported in Excel, RTF, Text, HTML, or proprietary STN Express format, which may include search transcripts, highlighting of search terms or other specified terms, and graphical representations of results data. **Alerts** can be set up to track new records relating to particular search queries.
172. **STN Easy** allows searching in a more limited set of databases than STN Express. It provides two general search interfaces, namely an Easy Search and Advanced Search interface, as well as a CAS Registry Number search interface, which are available in several languages. The **Easy Search interface** allows basic searching by keywords within specified database categories. It supports basic Boolean operators and wildcard operators, which can only be used for right truncation. The **Advanced Search interface** also allows users to select the category of databases within which to search but additionally permits **searching by field** according to a set of fields determined by the particular database category selected. The **CAS Registry Number search interface** permits search through chemical name or CAS Number to locate results in the CAS Registry. **Search history** allows users to repeat queries already carried out previously.
173. Results are displayed in a manner similar to STN Express, though none of the post-processing options available in STN Express exist in STN Easy.
174. Other STN interfaces and tools offered by CAS and FIZ Karlsruhe include **STN Easy for Intranets** and **STN on the Web**, a web-based interface similar to STN Express, and **STN AnaVist** and **STN Viewer**, both of which are designed to allow further analysis and processing of results from STN Express.

³³ See: STN International, "Database Summary Sheets", http://www.stn-international.de/sum_sheets.html (Accessed 9 November 2009).

(iv) **Thomson Innovation**

175. Thomson Innovation provides subscribers with tools to search, access, and analyze patent and non-patent data. Thomson Innovation's non-patent "Literature Collection" provides the user with access to several science and engineering related databases, specifically Thomson Scientific's proprietary Web of Science, Current Contents Connect, Inspec and Institute for Scientific Information (ISI) Conference Proceedings databases. In total, Thomson Innovation covers over 50 million bibliographic data records, many of which include specialized data such as chemical structures and reactions. Each of the content sets covers different though not necessarily discrete subject areas. Though certain data elements such as document title and author are shared among all content sets, others are distinct to particular content sets, e.g. CODEN publication identifier, serial item and contribution identifier (SICI), and other classifications (available only in Inspec). Furthermore, not all data fields included in particular data sets are searchable, notably abstracts. Finally, though links are provided to document full texts, access to these full texts must be obtained through the respective journal publishers.
176. For its non-patent literature search, Thomson Innovation offers quick search, fielded search, "cross-search", cited references search and expert search interfaces. All interfaces support the use of basic Boolean operators and most allow wildcard operators, which can be used for simultaneous left, right, and internal truncation. The **quick search interface** allows searching in one content set (e.g. Web of Science) or in all content within available data fields. The **fielded search interface** additionally permits users to search by fields such as author, title, date range, language, and document type, with the possibility of creating additional search fields as required. The **cross-search interface** allows searching for patents, scientific literature and business information simultaneously, with the possibility to search by fields including keyword, applicant/assignee, author, abstract, title, person, and date range. A search using the **cited references interface** will search for articles by cited author, cited work, or cited year. This interface does not permit the use of wildcard operators and is only available for certain content sets, such as Web of Science. The main element of the **expert search form** is the query box where users can enter search strings via command language. Above the query box, collapsible menus display the 'Collections to Search' menu (i.e. Web of Science, Current Contents Connect, ISI Proceedings etc). All the operators available in fielded search are available here, plus several additions, including numerical comparison operators (greater than or equal to, less than or equal to, and equal to). The expert search interface also allows the use of ordered and unordered user-definable proximity operators including SAME (terms must be in the same data field), NEAR, and ADJ. The database also provides an option to **save search history**. Saved queries may be organized into user-created folders.
177. Users can **customize the displaying of results** by selecting the data fields to be shown. Results can be sorted by author, source, and publication date. **Results can be refined** by adding search criteria to the original query or by filtering according to source, document type, author, and language, with a list of options being given for each filtering type. **Results lists or selections of results lists can be saved or downloaded and emailed** in a variety of formats including CSV, HTML, PDF and text formats. **Customized email alerts** can also be established to monitor results lists associated with a search previously carried out or changes to specific records ("watched records"). Users can also choose the **text clustering** option which uses linguistic analysis to create keyword clusters from the text of selected fields in the patent or literature records or **topographical content maps of search results** ("ThemeSpace").

(f) Free-of-charge journal databases**(i) Directory of Open Access Journals**

178. The Directory of Open Access Journals provides a list of journals that meet criteria for open access laid out in the Budapest Open Access Initiative of allowing individuals and institutions to “read, download, copy, distribute, print, search, or link to the full texts of ... articles” at no charge and without any embargo period. Additional criteria include the presence of a peer-review or editorial review process, that the content of the journal be “primary results of research or overviews of research results” and that the journal appears periodically over an indefinite period and have an ISSN number assigned to it.³⁴
179. The Directory covers a wide range of subject areas in over 4300 journals, which are available in wide variety of languages including Chinese, English, French, Japanese, Portuguese and Spanish. In certain cases, journals are available in multiple language versions. Though the Directory is primarily a gateway to the journal titles it lists, it is in the process of establishing a repository of the journals it covers through the e-Depot digital archive.
180. DOAJ offers a **basic search interface** and an **advanced search interface**. The basic search interface permits searching in the journal title field only, where multiple search terms are automatically treated as phrases. Users can also browse by journal title and subject area. The advanced search option allows the use of basic Boolean operators and searching by journal title, ISSN, journal and article keyword, article title, author name, and abstract. However, only a limited number of journals support article-level searching, i.e. by article title, author name, abstract, and article keyword. Journal and article bibliographic data and metadata, including article title, author name, publisher name, keywords, and subject area data, can also be accessed in a standardized format using the Open Access Initiative (OAI) protocol.
181. DOAJ offers an option of browsing “New Titles”, where users can find the most recently added journals. The list is automatically updated and shows by default the journals which have been added during the last 30 days.

(ii) SciELO

182. The Scientific Electronic Library Online (SciELO) is an electronic library covering a selected collection of scientific journals from Brazil and other Latin American and Caribbean countries, Portugal and Spain. It is a “virtual library, providing full access to a collection of serial titles, a collection of issues from individual serial titles, as well as to the full text of articles.”³⁵ The criteria for inclusion of journal titles take into consideration the previous evaluations that these journals have taken part in including their rating in selected international indexes as well as other factors such as the presence of a peer-review process and an editorial board, a specific frequency of publication and minimum lifetime, and availability of English versions of titles, abstracts, and keywords.³⁶ As a result of the latter criterion, all titles, abstracts, and keywords are searchable in the original article language or in English.

³⁴ Directory of Open Access Journals, “FAQ”, <http://www.doaj.org/doaj?func=loadTempl&templ=faq> (Accessed 28 October 2009).

³⁵ SciELO Spain, “SciELO – Scientific Electronic Library Online”, <http://scielo.isciii.es/scielo.php?lng=en> (Accessed 22 October 2009).

³⁶ See: SciELO, “SciELO Criteria: criteria, policy and procedures for admission and permanence of scientific journals in the SciELO collection, October 2004 Version”, <http://www.scielo.org/php/level.php?lang=en&component=42&item=2> (Accessed 22 October 2009).

183. SciELO is **fully available in English, Spanish, and Portuguese** and offers **two general search interfaces**. Both general search interfaces permit searching using basic Boolean operators across all SciELO content or within a particular collection, national or thematic (i.e. in the public health or social sciences collection). Furthermore, SciELO allows **searching with specific journals**. The first general search interface offers the option to search for articles by word or phrase in the country, author, keyword, title, organization, publication year, document type, and original language fields (either simultaneously or separately). It also permits **similarities searching** (searching by “lexical proximity”). Users can also **browse journals by title, subject area, or publisher**. The second general search interface allows searching by “all indexes” or by article title, author, or abstract. It also permits **filtering by subject area, journal, publication year, original language, and collection**. Within specific journals, users can search articles by author or subject, or they can search using basic Boolean operators by article title, abstract, author name, subject area, and publication year and, for certain journals, author affiliation and author country.
184. By default search results are sorted according to relevance. But results may also be sorted by year and SJR index (a citation index³⁷). Results lists can be emailed, exported in XML format, or tracked as RSS feeds. Individual results can be previewed in HTML format or viewed and downloaded in text or PDF format. RSS feeds may also be established to track new articles published in specific journals or new articles or journals added to SciELO as a whole. Users who have registered for free may also establish article and links lists (“My Collection” and “My Links”), track new articles in specific subject areas and subfields (“My Profiles”), or monitor RSS feeds (“My News” and “My Alerts”).
- (g) Fee-based journal databases**
- (i) American Chemical Society**
185. Through its publications site, the American Chemical Society (ACS) allows searching of and, for subscribers, access to articles from more than 30 of its journals and magazines in the field of chemistry as well as titles from its ACS Symposium Series and Advances in Chemistry series. This content covers a significant proportion of the titles in the field of chemistry in the PCT NPL Minimum Documentation.
186. ACS provides a basic search interface and an advanced search interface as well as options to **directly access specific articles or serials** through its citation search interface and DOI (digital object identifier) interface. The **basic search interface** supports basic Boolean operators, phrases, and wildcard operators (though wildcards cannot be used with phrases). It permits **searching by field** in full-text or in document and publication title, author name, or abstract. The advanced search interface allows combined searching in all these fields as well as in figure/table captions, within specified web or print publication dates, and in specified journals or book series. It also permits search term stemming. For both the basic and advanced search interfaces, ACS offers an option whereby users can track and save their past searches through their **session history** accessed through the results view.
187. Search results are shown on separate tabs for all results, research, news, features, and commentary, and general information and can be sorted by relevance, date of publication, manuscript type, publication title or first author’s last name. Results can be **filtered by content type** (book chapter or journal article), **publication, manuscript type, author name, or date range** (last week/3 months/6 months/year). Furthermore, results can be filtered to show only “Just Accepted” manuscripts that have not yet undergone technical

³⁷ See: Scilmago, “SJR: Journal Indicators”, <http://www.scimagojr.com/journalrank.php> (Accessed 29 October 2009).

editing or formatting for publication, which are made available to the public within 3 days of their acceptance for publication. Users can also **export results as RSS feeds** or **download bibliographic data (citations)** from selected search results

(ii) **IEEE Xplore**

188. IEEE Xplore provides access to journals, conference proceedings, books, and standards published by the Institute of Electrical and Electronics Engineers (IEEE) and the Institution of Engineering and Technology (IEE). The content covers a significant proportion of the titles in the field of electrical engineering in the PCT NPL Minimum Documentation. In total, the database contains over 2 million documents.
189. IEEE Xplore provides a basic search interface and a citation search interface, which are accessible to the public, and a structured advanced search interface, a free-text advanced search interface, **author search**, and a CrossRef search, which are accessible only to subscribers. The **basic search interface** supports basic Boolean operators (AND, OR, NOT). For non-subscribers, the basic search interface allows combined searching only in the document title, publication title, author name, abstract, index, and affiliation fields, while subscribers can additionally search simultaneously in the full-text. The **citation search interface** permits users to carry out searches by field in the publication title or ISSN/ISBN fields, the document title, author name, publication year, volume, issue, and start and end page fields. The **structured advanced search interface** also supports only basic Boolean operators but additionally permits field searching in a limited number of fields (full-text, document title, author name, publication title, abstract, index terms, and affiliation). The structured advanced search interface permits filtering according to publication type and date of publication. The **free-text advanced search interface** supports a wide range of operators, including diverse proximity operators (NEAR, PARAGRAPH, SENTENCE, PHRASE), relational operators (ORDER), as well as truncation and similarities searches, in addition to allowing field searching in a broad range of fields. These fields include metadata fields, namely index terms and CODEN designations (the latter being used to classify serial and non-serial titles). Both advanced search options permit search results to be refined by adding additional keywords to the search. The **CrossRef search** allows users to simultaneous search content from IEEE but also from other publishers participating in the CrossRef Search Pilot program.³⁸ Subscribers can track and save their past searches through their **session history**.
190. Search results can be sorted by relevance, date of publication, or document title. Results include **backwards citations** ("References") and **forward citations** ("Citing Documents") that are accessible only to subscribers. Subscribers also have the possibility to **download bibliographic data** from their search results.

(iii) **ScienceDirect**

191. ScienceDirect provides access to full-text articles of a large amount of scientific journals and books. It covers a substantial number of titles from the PCT Minimum Documentation in particular in the fields of biomedicine and chemistry, including Tetrahedron, Biochimica et Biophysica Acta, Cell, and Gene. In total, ScienceDirect provides online access to over 9 million articles.
192. ScienceDirect provides a basic search interface, an advanced search interface, expert search interface. The **basic search interface** allows users to search by all fields, journal/book title, author name, volume, issue, and page number but only supports the use of the AND operator and phrases. The **advanced search interface** supports basic Boolean operators (AND, OR, AND NOT), phrases, and filtering by publication type, date,

³⁸ See: CrossRef, "CrossRef search", <http://www.crossref.org/crossrefsearch.html>

and subject area and permits subscribers to search in a limited range of fields (all fields; abstract, title, and keywords; authors; “specific author”, source title, title; keywords; abstracts; references; ISSN, ISBN, affiliation, and full-text). Subscribers can also restrict search to their list of subscribed and favorite sources. The **expert search interface** permits the use of Boolean operators and phrases as well as proximity and wildcard operators. Users can search by a wide range of fields using field codes. The advanced and expert search interfaces also allow subscribers to determine whether they want to search only in the titles to which they have subscriptions or in all titles.

193. Search results can be sorted by relevance or date of publication. Results can also be refined to ‘**limit to**’ or ‘**exclude**’ particular content, filtering by content type, journal/book title, or year or by carrying out a search in the results. Subscribers also have the possibility to **download bibliographic data** from their search results. They also have the option to **open previews of all articles** which displays the abstracts, figures/tables or references of each result.
194. Subscribers can track and save their past searches through their **session history**. It also allows subscribers to set up **email alerts** based on specific search queries. Any user can also search through a list of the most downloaded articles (“Top-25 Hottest Articles”).

(iv) SpringerLink

195. SpringerLink provides the possibility to search and obtain paid access to the full text of journal articles, books, and other publication types covering a wide range of subject areas. These subject areas include: Architecture and Design; Behavioral Science; Biomedical and Life Sciences; Business and Economics; Chemistry and Materials Science; Computer Science; Earth and Environmental Science; Engineering; Humanities, Social Sciences and Law; Mathematics and Statistics; Medicine; Physics and Astronomy; and Professional and Applied Computing. SpringerLink includes a substantial number of titles in the fields of chemistry, electrical engineering, and instruments from the PCT NPL Minimum Documentation. In total, the database contains more than 4 million documents.
196. SpringerLink provides a structured search interface and a free-text search interface. Both search interfaces support basic Boolean operators (AND, OR, NOT) as well as wildcard operators, phrases, and nesting (to clarify ambiguous search syntax). The **structured search interface** allows combined searching in all text or field searching by document title, document abstract, author name, editor name, ISSN, ISBN and DOI and by date range. The **free-text search interface** supports a similar range of search fields (with the exception of editor name) but does not permit date range search. From the free-text search interface, it is also possible to browse publications by content type, subject area, or “featured collection”.
197. Search results can be sorted by date of publication, document title, type of document and author name. From the results view, it is possible to **narrow results** by adding additional search criteria or by applying **filtering by content type, language, date of online publication (“SpringerLink date”), copyright date, author, and content status** (i.e. by accepted manuscript or published document). Subscribers can track and save their past searches by marking items in their **session history**. It also allows for **emailing, exporting and saving search results** in a variety of formats (including Excel spreadsheets with customized content) as well as for the creation of **subscriber alerts**.

(v) Wiley InterScience

198. The Wiley InterScience website provides access to full-text journals, reference works, books, and databases in the fields of “Chemistry, Earth and Environmental Science, ...,

Information Science and Computing, ... Life Sciences, Mathematics and Statistics, Medical, Veterinary and Health Sciences, Physics and Astronomy, [and] Polymers and Materials Science"³⁹. The content covers a significant proportion of the titles in the field of physics in the PCT NPL Minimum Documentation. It contains around 3 million documents in total.

199. Wiley InterScience provides a basic search interface, an advanced search interface, and a CrossRef/Google Search interface to locate requisite journals and articles. Users can also **browse journals by subject area**. The **basic search interface** supports **basic Boolean operators** (AND, OR, NOT) as well as **wildcard operators**, which can be used for simultaneous left, right, and internal truncation. Search can be conducted by publication title or in all content of the database. The **advanced search interface** also supports **basic Boolean operators** and **wildcard operators** but also **proximity operators**. Users can specify up to three search expressions for different data fields including article title, publication title, full-text/abstract, author name, and keyword, which may be combined by selecting "AND," "OR," and "NOT" from accompanying drop-down menus. Searches can further be limited by publication type and date range. When viewing any journal, online book, reference work, database or Current Protocols technical manual, the user can choose to search content only within that specific title in all fields or by article title, author, full-text/abstract, author affiliation, keyword, funding agency, reference, or article DOI or jump directly to a particular section of a journal by entering the desired volume, issue, and page number. **Results sorting** can be done by relevance (match percentage), date, or publication title.
200. The CrossRef/Google Search interface permits searching across all content from Wiley and other scholarly publishers participating in the CrossRef program.⁴⁰ This interface permits the use of the same operators as Google itself, i.e. Boolean operators but not wildcard operators. Results are returned from Google using the Google search and ranking algorithms, and using the article's DOI whenever possible to link from the search results to the published article.
201. For the basic and advanced search interfaces, subscribers can track and save their past searches through their **session history**. **Citation export** functionality will allow any user to download the citation and abstract for journal articles and book summaries. **Citation tracking** allows any user to track where an article has been cited in other online content via links within the article content. Wiley InterScience also provides links on some pages of relevant journals that allow direct access to articles for authenticated users of services such as Highwire, articles archived on hosting services such as JSTOR, or correspondence websites and blogs. An **Acronym Finder** is made available to help find acronym definitions. Acronyms can be searched for exact matches, matches at the beginning of terms, or wildcard matches, and results can be limited according to specific definition keywords or particular subject areas.
- (h) Access to Research for Development and Innovation (aRD*i*) and related services**
202. The Access to Research for Development and Innovation (aRD*i*) program was launched by the World Intellectual Property Organization together with its partners in the publishing industry with the aim to increase the availability of scientific and technical information in developing countries.⁴¹ By improving access to scholarly literature from diverse fields of science and technology, the aRD*i* program seeks to:

³⁹ Wiley InterScience, "Home", <http://www3.interscience.wiley.com/cgi-bin/home> (Retrieved 3 November 2009)

⁴⁰ See: Crossref.org, "Publishers and Societies", <http://www.crossref.org/01company/06publishers.html>

⁴¹ For more information about the aRD*i* program, please refer to <http://www.wipo.int/ardi>

- reinforce the capacity of developing countries to participate in the global knowledge economy; and
 - support researchers in developing countries in creating and developing new solutions to technical challenges faced on a local and global level.
203. Within the framework of the aRD_i program, academic and research institutions and intellectual property offices in least-developed countries as defined by the United Nations may apply for free access to the journals made available through the aRD_i program. Intellectual property offices in other developing countries may obtain access to these journals for a nominal fee. At the time of writing, the aRD_i program provides access to 50 journals from the PCT NPL Minimum Documentation for 107 developing countries in cooperation with 12 major publishers including the American Institute of Physics, Elsevier, John Wiley & Sons, National Academy of Sciences, Oxford University Press, Royal Society of Chemistry, Sage Publications, Springer Science+Business Media, and Taylor & Francis.
204. The aRD_i program has been developed in close cooperation with similar programs already established by the World Health Organization (WHO), the Food and Agriculture Organization of the United Nations (FAO), and the United Nations Environment Programme (UNEP).
205. The HINARI program, established by the World Health Organization, was developed in the framework of the Health InterNetwork, introduced by the United Nations' Secretary General Kofi Annan at the UN Millennium Summit in the year 2000. The HINARI Programme enables health institutions in 107 developing countries to gain access to over 6400 journal titles in biomedical and health fields.
206. The Access to Global Online Research (AGORA) program, set up by the Food and Agriculture Organization of the United Nations in 2003, enables not-for-profit national academic, research or government institutions in agriculture and related biological, environmental and social sciences located in developing countries to gain access to a collection over 1270 journals in the fields of food, agriculture, environmental science and related social sciences.
207. The Online Access to Research in the Environment (OARE) program, launched in 2006 by the United Nations Environment Programme (UNEP), enables public institutions and all local non-governmental organizations in 100 developing countries to gain access to over 2,990 peer reviewed journals.

4. ANALYSIS

(a) Introduction

208. As mentioned in the Introduction (see above section 1. (a)), the scope of the current Paper is to review the major existing searchable patent and non-patent databases and provide an indication to Member States as to which databases may be useful for them. A decision as to the usefulness of a database can be made from a country specific needs analysis which reflects the search requirements, particularly as regards the kind of information and fields of technology sectors users are searching, and the depth of such searches. The selection of databases can be decided also from the viewpoint of whether such needs could be best met from the information on data coverage, features and search functions of a database.

(b) Needs analysis

209. As the Terms of Reference suggest, the countries' needs are assessed on the basis of statistics representing national patenting activity (patenting propensity), e.g., number of first patent filings per million population and other data available to WIPO⁴², in particular those indicating the fields of technology in which developing countries have most patenting activity, as well as the questionnaire sent with Circular C.N 3024 to Member States' intellectual property offices for a self-assessment of their needs.

210. Annex II reflects the compilation of both sets of data in a tabular form, while conclusions from both sets of data, as well as their graphical representation, are given below.

(c) User profiles and needs

(i) Definition and methodology

211. According to Point 2.2 ("Objectives") of the Project Document of the Development Agenda Recommendation No. 8, target beneficiaries are among others IPOs, individual inventors, small and medium enterprises (SMEs), industry, researchers in technology centers and universities, academia, IP professionals and government policy-makers. This framework is useful for the definition of the users needs.

212. The user needs analysis focuses on two groups of users: IPOs and other user groups in general. Moreover, the reasons for carrying out a patent search by the various users will also give an indication for using certain specialized patent databases.

(ii) Conclusions from statistics on patent activity

213. The WIPO Indicators and the World Patent Report, as well as the National Collections Statistics provide among others data on patent applications by field of technology. These statistics reflect the national and international patent activity of developing countries which are the subject of the current Study Paper. The fields of technology in which patent applications are filed indicate trends and are useful to identify specialized patent databases which would be of interest for IPOs and other user groups of developing countries and LDCs. As can be seen from the above mentioned statistic, the main technical fields of interest in developing countries in general are in the Chemistry, Medicine/medical technology, Pharmaceuticals and Biotechnology fields.

(iii) Conclusions from responses to questionnaire by IPOs of WIPO Member States

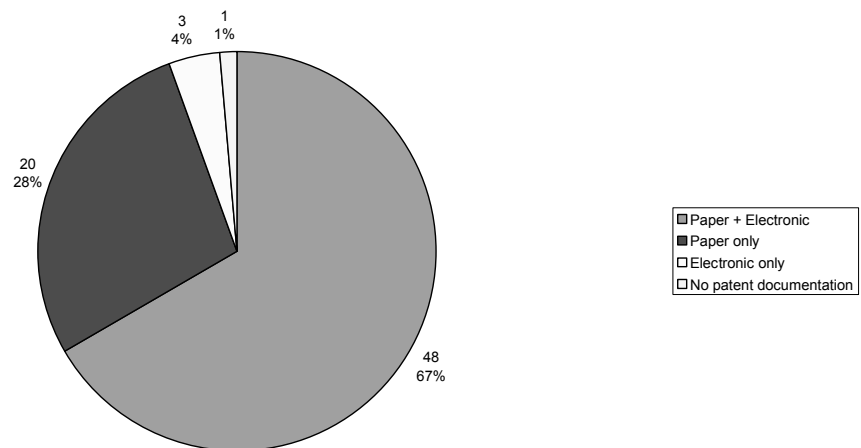
214. In order to assess the present situation and future needs of IPOs and their stakeholders with respect to technology databases, a questionnaire was sent with Circular C.N 3024 to IPOs within the framework of Recommendation 8 of the Development Agenda. The responses from 72 offices (70 national offices and two regional offices) were received

⁴² <http://www.wipo.int/ipstats/en/statistics/patents/index.html>

and are summarized below. The numbers in brackets represent the numbers of responses belonging to each item, as well as the percentage with regard to the overall total number of replies received. When reviewing the overall results, it should also be taken into account, that out of the 70 WIPO member states from whose national offices responses were received, 53 are eligible for PCT fee reductions, while the two responding regional offices largely cover countries eligible for fee reductions as well. Thus, they could be considered to broadly represent the situation faced by developing countries. It may be added, in any case, that even the replies from the 17 other national offices nonetheless provide useful input as regards eventual needs and solutions for developing countries and their content leads to the same conclusions.

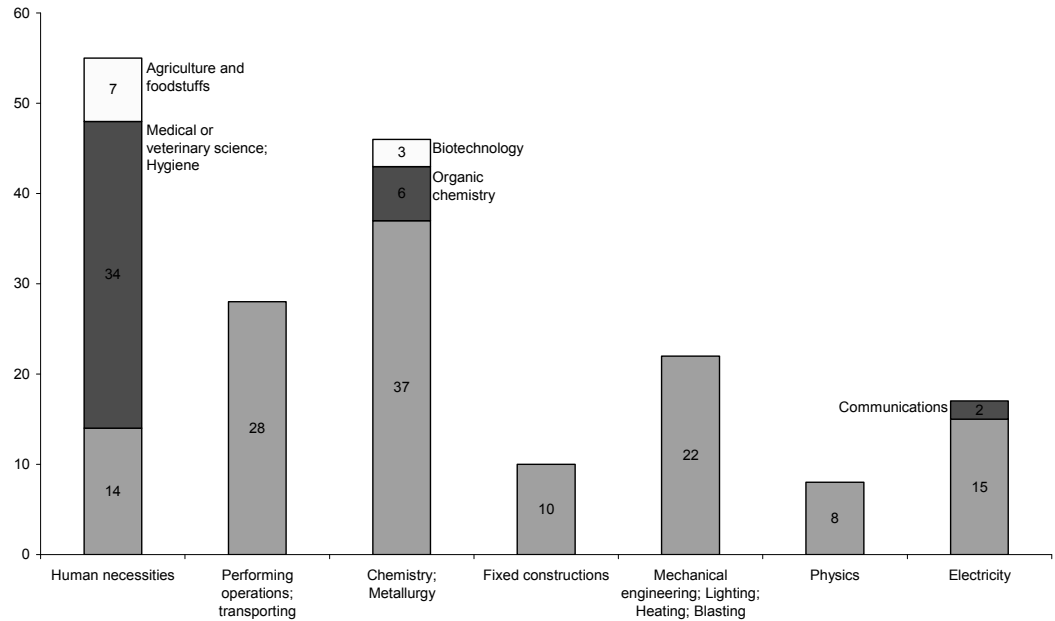
- 215. Most of the responding offices carry out prior art searches (58, 80.56%) and perform substantive examination (52, 72.22%). At offices that perform prior art searches, approximately 30 percent of staff is employed as examiners who carry out such searches.
- 216. Most responding offices maintain their patent documentation in electronic form (51, 70.83%). However, a substantial number (20, 27.78%) only have access to their documentation in paper form. Certain offices maintain their patent documentation fully in electronic format.

Figure 2. In what form does your Office maintain its patent documentation?



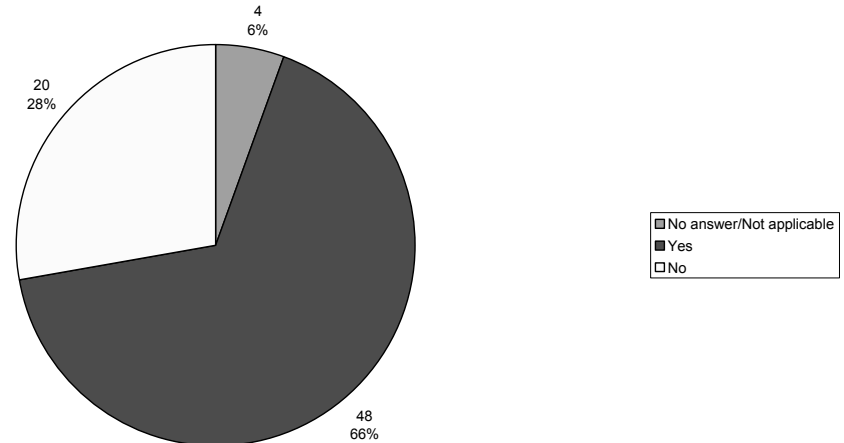
- 217. The main fields of technology in which patent applications are filed at responding offices are the following: human necessities (51, 70.83%), in particular medical or veterinary science and hygiene; chemistry and metallurgy (45, 62.50%), in particular organic chemistry; performing operations (28, 38.89%); mechanical engineering (22, 30.56%); and electricity (17, 23.61%), in particular communications.

Figure 3. What are the top three technical areas of patent applications filed at your Office?



218. The majority of the offices that carry out prior art searches use a patent search system (48, 66.67%), of which EPOQUE/EPOQUENet is the most commonly used (25, 34.72%), followed closely by national database systems (21, 29.17%). Several offices also use EAPATIS (Eurasian Patent Information System) and commercial search systems. Only 20 of the offices, or 27.8 percent of responding offices, do not use any particular search system at all.

Figure 4. Does your office use a patent search system for prior art searches?

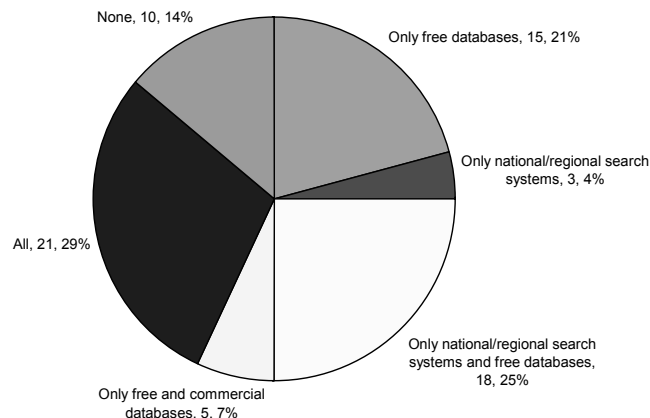


219. The most commonly cited free-of-charge patent databases used by the offices are esp@cenet (56, 77.78%), PATENTSCOPE® search service (46, 63.89%), USPTO (36, 50.00%), JPO/NCIPI (24, 33.33%), German PTO (14, 19.44%) and Korean IPO/KIPRIS (13, 18.06%) databases. A significant number of offices also use Google Patents (16, 22.22%) to carry out searches.

220. Most responding offices do not use commercial patent databases available on the internet. It may be partly because offices do not have sufficient reasons to justify the payment of subscription fees of commercial databases due to the fact that increasingly the public sector databases offer many functionalities which available in commercial databases. As the survey results indicate (see Figure 6 below), intellectual proper offices as one of the main users of patent databases, provide essentially basic searches. Basic searches alone do not require the more sophisticated search and analysis tools available in commercial databases.

221. The two most widely used commercial patent databases are STN (10, 13.89%) and Delphion (8, 11.11%).
222. Overall, 13.9 percent of responding offices neither use a national or regional patent search system nor any kind of free or commercial patent database available online, while 20.8 percent of responding offices use only free databases. 25.0 percent of responding offices use only national or regional search systems and free databases, while 29.2 percent use all national or regional search systems and all types of available online patent databases.

Figure 5. Does your office use national/regional patent search systems, free patent databases, or commercial patent databases?

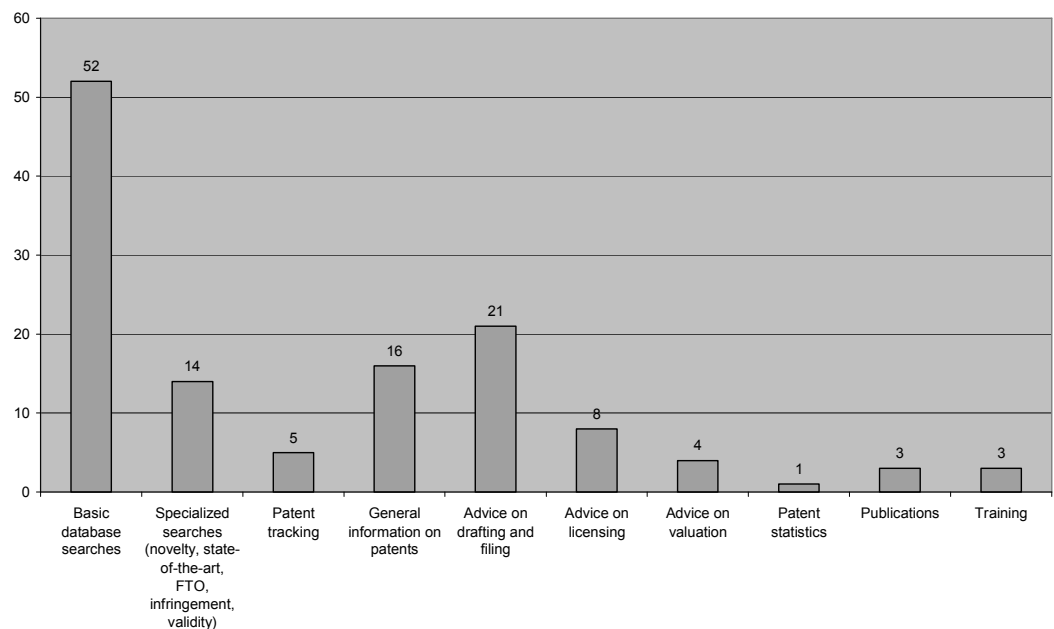


223. The three most commonly cited free-of-charge non-patent databases and search services used by the offices include the NIH National Library of Medicine databases and services (e.g., PubMed and PubChem) (25, 34.72%), Google Scholar (14, 19.44%), Scirus (8, 11.11%), and ScienceDirect (free search) (8, 11.11%). Notable content databases include the Directory of Open Access Journals (5, 6.94%) and SciELO (4, 5.56%). A large number of other databases, in particular non-English language databases, are additionally used by individual offices.
224. Few responding offices use commercial non-patent databases, which may also be linked to overlaps with public sector databases. IEEE Xplore (13, 18.06%) and ScienceDirect (11, 15.28%), and STN (7, 9.72%) are the three most widely used commercial non-patent databases.
225. The additional commercial patent databases considered most useful include Derwent (Innovations Index and World Patents Index) (23, 31.94%), STN (8, 11.11%), and Delphion (4, 5.56%).
226. The additional commercial non-patent databases considered most useful include STN (9, 12.50%), ScienceDirect (8, 11.11%), Inspec, Medline, and Wiley InterScience (each: 4, 5.56%).
227. The most widely cited reason for obtaining additional patent and non-patent databases were their additional content (27, 37.50%), greater efficiency in searching (5, 6.94%), and ease of use (4, 5.56%).
228. Nearly all of the responding offices have websites (65, 90.28%), the majority of which contain information on national patent documentation (54, 75%). The average number of hits received by these websites was over 1.5 million. However, this average was skewed by an extraordinary number of hits received by a few offices, namely the Russian Patent Office (30 million), USPTO (20 million), and JPO (8 million). Excluding these offices' figures, the average was closer to 200 000. An average of around 19 500 inquiries on patent information were made per year at responding offices, a figure again somewhat skewed by the large number of inquiries made at certain offices, in particular the Korean

Intellectual Property Office (756 000) and the German Patent and Trade Mark Office (175 000). Excluding these two offices, the figure was closer to 4500 inquiries.

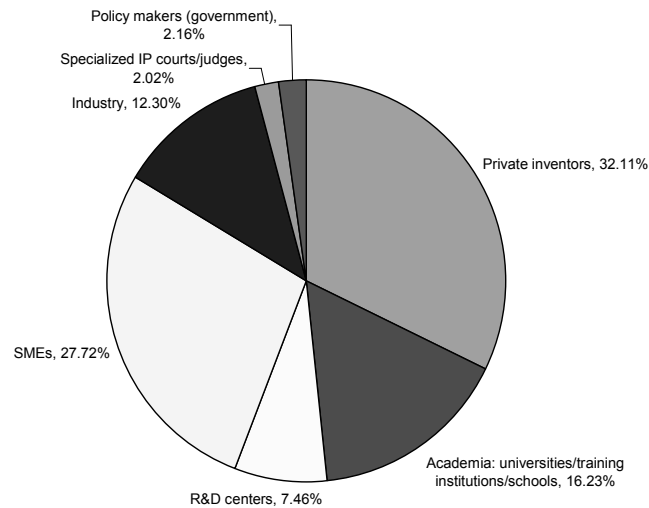
229. Major sources of patent information in various countries and regions are the national offices (69, 95.83%), patent information centers (35, 47.22%), and regional patent offices (21, 29.17%). Two offices (Afghanistan and Bahrain) do not make available patent information, while one office (United Republic of Tanzania) only makes its patent information available through patent information centers. Other non-commercial sources of patent information include chambers of commerce, libraries, and technology transfer centers.
230. Most offices provide basic database searches (52, 72.22%). Other services provided include: advice on drafting and filing (21, 29.17%), on licensing (8, 11.11%), and on valuation (4, 5.56%); specialized searches (e.g., novelty, state-of-the-art, freedom-to-operate, infringement, and validity searches) (14, 19.44%), patent tracking services (5, 6.94%), and training (3, 4.17%).

Figure 6. What are the services offered, e.g. basic databases searches information on other aspects of innovation support (licensing), etc.?



231. On average, over 1100 national patent agents/attorneys were registered per country in various countries and regions, though this average is strongly skewed by the number of patent attorneys registered in the United States (38'729) and China (over 8000). Excluding these offices, the average was closer to 170.
232. As estimated by offices, the most sizeable groups of potential users of patent information are private inventors and SMEs, which were considered to represent an average of around 32.1 percent and 28.6 percent of the user base, respectively. Academia and industry are estimated to represent 16.2 and 12.7 percent of the user base, R&D centers 7.7 percent, and specialized courts and policymakers around 2 percent each.

Figure 7. Please estimate the approximate number of potential users of patent and non-patent information services.



233. Only a small number of responding offices (10, 13.89%) belong to consortia providing access to e-resources. These consortia are often organized among agencies of the national government.
234. According to responding offices' assessment, potential users of patent information in most countries (17, 23.61%) would likely require access only to free databases, though many responding offices considered access to free and commercial (12, 16.67%) or only commercial (12, 16.67%) to be important to meeting users' requirements.
- (iv) User needs**
235. Beneficiaries from the access to patent and non-patent information through databases are typically patent examiners of intellectual property offices, thus people directly involved in the patenting process in patent offices, or those interested in or involved with patent application procedures such as patent applicants and their agents. Potential patent applicants are research organizations, academia, small and medium enterprises (SMEs), individual inventors etc. They would benefit even more from patent databases if intellectual property offices provided them with search assistance.
236. Patent searches can be carried out to retrieve various types of information relevant to users' needs. They may be used to determine the state of the art in a certain field of technology (state-of-the-art searches). Patent offices, applicants and their agents may also seek to determine whether or not an invention is novel (novelty/patentability search). A higher quality of patent search results, crucial for patenting and examining procedures, can be achieved when access to the broadest possible coverage of patent information is ensured.
237. Increasingly, information contained in patent documents is also supplemented by non-patent literature, i.e. scientific and technical information published in other forms, such as journals or research databases. In some countries and under international or regional patent systems such as PCT and the European Patent Convention, a patent application is published with a search report that contains the result of the prior art search performed by a searching authority with citations which may be patent documents and/or articles in scientific and technology journals.

(d) Comparison between free-of-charge and commercial databases**(i) General**

238. Only the most general of comparisons of the differences between free-of-charge and commercial databases can be made with particular reference to the coverage, tools and general use of these two “types” of databases, namely free-of-charge and commercial.
239. In general, free-of-charge databases, most of which are public sector databases, continue to evolve from simple electronic publications of patent gazettes containing basic and “raw” data (data extracted from an application as filed), to the full data containing machine searchable text of an application electronically filed. These databases were mainly meant to meet a statutory requirement of the public disclosure of a patent application, and not specially designed to meet the needs for searching known prior art. On the other hand, commercial patent databases are expected to meet needs of industry and potential applicants as regards a prior art search and for ensuring the freedom to operate. Thus, they offer more value-added data, which has been reviewed and if necessary corrected, to enhance the quality of raw data as well as more sophisticated search and analytical tools to improve the accuracy (less noise) and “recall” (higher return of relevant documents). Clearly, commercial databases compete and must offer more value-added content and functionalities than their free-of-charge “rivals” in order to remain viable and competitive.

(ii) Coverage

240. As regards content of patent information, there are no major differences in the coverage offered by free-of-charge patent databases and the commercial ones. However, as just mentioned, the quality of that data, whether it is bibliographic data, abstract quality, etc. is reviewed by commercial database providers which is clearly of added value over the free-of-charge databases as regards the likely quality of a prior art search. Commercial databases do not cover all the public sector databases, although their geographical coverage has been recently enlarged. This may be due to the fact that a number of countries (in particular developing countries and LDCs) have not fully digitized the data of their national patent applications so as to make them available in electronic form, and that most applications filed with such countries are second filing applications which have been included in databases of the first filing countries.
241. The situation regarding non-patent literature is quite different. Open source journal databases do exist but are not as widespread as commercial databases and even though the latter allow searching free-of-charge, access is only granted to the title and abstract for free, while access to the complete article content is only granted on payment.

(iii) Tools

242. The availability of data from various collections is just the first step for an accurate patent search. Patent searchers are assisted in retrieving relevant data by using search engines with sophisticated search and analysis tools. For instance, no public sector database service provides search facilities to allow for an in-depth search of a new chemical entity by chemical formula or for searching similar amino acid sequence listings by statistical analysis. Specialized commercial databases exist for such purposes. Obviously the availability and the efficiency of such tools is greater in commercial database search services than in free-of-charge ones.
243. Moreover, a distinction should be made between search and analysis tools, as the majority of the sophisticated tools are analysis-related, which is not always useful or necessary for IPO users. Therefore, IPOs which do not carry out the full substantive examination, may not necessarily be interested in the most sophisticated analysis tools and may be satisfied using those increasingly available in free-of-charge database search services.

(e) Issues and Recommendations**(i) Database selection**

244. The Study Paper provides an overview of the most representative databases including both public sector patent databases and commercial patent and non-patent services available on the internet so as to satisfy one of the primary aims of the Paper, which was to give member State IPOs and users in general an initial indication of which databases may be useful for them to carry out specific technology searches, whether they are state of the art searches (to give a broad overview of a specific technological field), patentability searches (related to the possible grant of a patent application), infringements searches (to verify the right to use or freedom to operate with respect to possible patents still in force) or validity searches (in response to the grant of a specific patent with the possible aim of opposing it).
245. Moreover, the analysis of developing countries' needs in this Paper also provides IPOs with a first indication in identifying databases which could satisfy their specific search needs, e.g. in specific technical fields.
246. Given the vast choice of databases, it is imperative that IPOs and users identify which databases are relevant and of interest to them, just as much as deciding eventually on search strategies using keywords and search tools, such as the patent classification, Boolean operators, etc.
247. No country specific recommendations can obviously be given in such a general review. No "one-size-fits-all" recommendation can be given to countries either. However, each member State IPO should be further encouraged to review its users' needs more specifically and apply the results of this Paper more closely to its needs. As just mentioned, a first step in this direction was taken by means of the questionnaire in Circular C.N 3024. The result of the present survey indicates the current and general trends with regard to the choice of patent databases. It should facilitate the decision by each country with regard to the choice of patent databases which meet the needs of its various users.
248. It is further intended to define country groups so as to identify countries according to their needs and means and thereby target assistance in accessing technical databases through established programs such as WIPO's program "Access to Research for Development and Innovation" (aRD_i) and through possible new programs for access to specialized commercial patent databases, such that countries in a particular country group may (though this is a broad generalization) more likely to be interested in similar technology-based databases.
249. WIPO's aRD_i program already lists two groups of countries which reflect both overall country needs. Group 1 where not-for profit research and academic institutions and patent offices in Least Developed Countries (LDCs) are permitted free access, and publishers' commercial realities, as in the case of aRD_i Group 2 where patent offices in certain publisher authorized developing countries have very low-cost access to scientific and technical journals.
250. The Terms of Reference suggested a cost-benefit analysis. However, the costs of subscribing to a commercial database are very much customized to: the specific type of user, (e.g. individual, commercial or institutional); the size of an institution; the region or country where the user resides (clearly users in countries with transitional economies are likely to pay more than in developing countries); different pricing structures by different vendors (e.g. a fixed annual subscription may also have additional costs regarding the number of searches performed or the number of downloads or simply the time connected to the service). Moreover, this information is not given on the internet and must be requested by each new user or institution directly from the respective database provider. Because of the large number of variables needed to be taken into consideration to provide

a simple comparable cost-benefit analysis, even if this information had been available online, it is unlikely that the resultant analysis would be intelligible or in the end useful. Moreover, the different needs of different intellectual property offices and users are most likely to lead to different conclusions as to whether it is worth paying a certain level of fee in the light of the degree to which such databases meet specific needs and how much value they offer.

(ii) Access to commercial patent databases

251. It is planned to establish a corresponding program for specialized (commercial) patent databases maintaining the country groups already defined in aRDi. However, given that a number of patent offices in countries outside these groups are still struggling to establish adequate access to critical patent resources, a third group of countries is envisaged in which access to patent databases would be provided to patent offices at a reduced flat rate.
252. Negotiations are still ongoing with commercial patent database vendors, but are sufficiently well advanced to envisage launch of such a program soon. Two meetings (the word “forum” was used in the terms of reference) of participating database vendors took place in February and June 2010 so as to discuss a common approach within the new program.

(iii) Enhancing PATENTSCOPE[®]

253. PATENTSCOPE[®] coverage has been expanded to include regional and national collections. This is based on WIPO’s strategy to meet Strategic Goals IV and V to create global databases and use it as a WIPO global reference for enhancing public goods. WIPO work programs will gear towards these goals. For instance, through the digitization, automation and modernization programs in LDCs, developing and transition countries and through arrangements with other countries and stakeholders (such as publishers and other commercial providers), the content and coverage of WIPO’s global databases including PATENTSCOPE[®] will be progressively extended. PATENTSCOPE[®] analytical and language tools have been greatly enhanced with new graphical analysis and the introduction of the Cross Language Information Retrieval (CLIR) tool (for further details see Section 2. (e) (i)). Contacts are also ongoing with popular internet portals, such as Google, and patent database service providers so as to seek possible cooperation in enhancing search technologies.

(iv) WIPO Patent Information Services for Developing Countries (WPIS)

254. WIPO Patent Information Services (WPIS) were specially developed for and used by developing countries for the last three decades to meet their patent search and examination needs. These services have been offered for many years free of charge as a result of contributions made by industrial property offices (“donor IPOs”⁴³) throughout the world and by the International Bureau of WIPO.
255. The following services are currently provided within the framework of WPIS:
- (a) State-of-the-art search reports for individuals and institutions carried out by donor IP Offices (on average, approximately 1,000 reports are produced every year);
 - (b) Based on agreements within the International Cooperation for the Search and Examination of Inventions (ICSEI), donor IPOs establish search and examination reports on patent applications filed with IP Offices (requesting IPOs), which have no

⁴³ IPOs of Australia, Austria, Brazil, Bulgaria, Canada, Cuba, Finland, France, Germany, Japan, Mexico, Norway, Republic of Korea, Russian Federation, Switzerland, the United Kingdom and Venezuela.

priority or for which no search reports have been established
(on average, approximately 300 reports are established every year);

(c) Information on equivalent patent documents published in other countries (family patents); and

(d) Copies of individual patent documents.

256. Given that a number of free of charge patent databases available on the Internet allow novice users to search and print out copies of individual patent documents, services of WPIS and ICSEI should be shifted from basic (such as item (d) above) to more value-added services. It is also important to redesign WIPO services in such a way that synergies between WIPO, IPOs and Technology and Innovation Support Centers (TISCs) will be created to cater for diversified needs of users of the patent system and local players of innovation, and to build national capacity using patent information.

257. WPIS in the future could provide the following two kinds of services regrouped according to their main focus and target audiences, i.e., (1) targeting users of patent information for promoting innovation with focus on patent information search, and (2) assisting IPOs in their work of patent search and examination, respectively:

(a) Patent Information Analysis service ("PIA") consisting of:

- the State-of-the-Art search report service (continuation of service referred to in (a) above)
- patent family search service (continuation of service referred to in (c) above with the possibility of extension to broader search of similar patents)
- patent legal status information service using currently available databases (a new database of patent legal status may be created as an extension of Patentscope®, subject to the approval of member States of a separate study to be presented to CDIP in connection with "Intellectual Property and the Public Domain")
- Patent Landscape reports service (a thematic project was approved by the CDIP for "Developing Tools for Access to Patent Information")

(b) International Cooperation for the Examination ("ICE") as a refocus of ICSEI service referred to in (b) above). WIPO will continue to seek more contribution from donor IPOs to increase the number of examination reports that they could prepare on demand from IPOs in developing countries. Such reports would be not only delivered but also be used as training material in training services for patent examiners of beneficiary IPOs in developing countries with a view to building their capacity of patent examination.

(v) Analysis of Training Gaps

258. The project implementing Recommendation 8 provides for the establishment of Technology and Innovation Support Centers (TISCs) in developing countries. Within this capacity building element of the project, TISC personnel are trained to assist their institutional and local users in accessing and searching technology databases. Training, therefore, is a very important element of the project and needs to be effectively provided in cooperation with national IPOs and local stakeholders. Similarly, dissemination and awareness raising activities need to be effectively coordinated at a national level.

259. As discussed in connection with ICE, an increasing number of intellectual property offices wish to strengthen their capacity of patent information search for substantive examination of patent applications. The WIPO Secretariat is prepared to further study and identify

specific needs for training of patent examiners. The scope and method of such training may go beyond the issue of access to patent databases. The current Study Paper indicates that there may be a need for another study concerning training of patent examiners particularly in developing countries which member States may wish to use for further discussions and possible reorientation of relevant programs of WIPO.

[Annex I follows]

WIPO Patent Information Needs Analysis Questionnaire

To be completed and returned by August 31, 2009 by email to andrew.czajkowski@wipo.int or by facsimile to (+ 41 22) 338 8495

Please use a separate sheet for your answers if necessary

A. Intellectual Property Office Needs	Answer (tick or fill in)	
1. Does your Office carry out prior art search?	Yes	No
2. Does your Office carry out substantive examination?	Yes	No
3. What are the top three technical areas of patent applications filed at your Office?		
4. How many patent examiners carry out prior art searches in your Office? 4.1 What is the total number of staff in your Office?		
5. In what form does your Office maintain its patent documentation: <ul style="list-style-type: none"> • paper • electronically • other (please give details) 	Yes Yes Yes	No No No
6. Does your Office use a patent search system for prior art searches? 6.1 If yes, which system does it use (e.g. Derwent, EPOQUE)? 6.2 Which databases are available on this system?		
7. Which free-of-charge patent databases (e.g. PATENTSCOPE [®] , espa@cenet, etc.), available on the internet, do your examiners currently use for prior art searches? 7.1 Which commercial (subscription/payment) patent databases on the internet do your examiners currently use for prior art searches?		
8. Which free-of-charge non-patent (scientific and technical journals) databases, available on the internet, are currently used? 8.1 Which commercial non-patent (scientific and technical journals) databases, available on the internet, are currently used?		

9. Which additional commercial patent databases do you believe would be of benefit for your examiners to access?		
9.1 Which additional commercial non-patent (scientific/technical journal) databases do you believe would be of benefit for your examiners to access?		
9.2 Why do you think these databases are of benefit?		
B. National User Needs		
10. Does your Office have a website?	Yes	No
10.1 Does this website include information on national patent documentation?	Yes	No
10.2 How many website users/hits does your Office have on average per month?		
11. How many external inquiries on patent information (physical visits or via e-mail) are made to your Office?		
12. Is patent information available today from the: <ul style="list-style-type: none"> • National Patent Office • Regional branch of the Patent Office • Patent information centers • Other 	Yes Yes Yes Yes	No No No No
12.1 If yes, what are the services offered, e.g. basic databases searches, information on other aspects of innovation support (licensing), etc?		
13. How many national patent agents/attorneys are there in your country?		
14. Please estimate the approximate number of potential users of patent and non-patent information services: <ul style="list-style-type: none"> • Private inventors • Academia: universities / training institutions / schools • R&D centers • SMEs • Industry • Specialized IP courts/judges • Policy makers (government) 		
15. Is your Office part of a library consortium providing access to e-resources, i.e. access to scientific and technical journals?		
15.1 If yes, to which journals and via which consortium?		
16. To which patent and non-patent databases mentioned in question sections 6 to 8 above are the above users likely to need access?		

Thank you for your cooperation!

[Annex II follows]