



European IPR Helpdesk

Fact Sheet

Automatic Patent Analysis

July 2015¹

Introduction.....	2
1. What does Automatic Patent Analysis consist in?	3
1.1 One field analysis	3
1.2 Multiple fields analysis: the matrices.....	4
1.3 The networks.....	5
1.4 An advanced practice: the use of strategic groups	6
2. Construction of a set of patent bibliographic references.....	6
2.1 Access to patent data	6
2.2 Download of patent bibliographic references	8
2.3 The example of Espacenet.....	9
3. APA practical application	10
3.1. The software solutions	10
3.2. Illustrated example (using Matheo Patent)	11

¹ This fact sheet was initially published in December 2013 and updated in July 2015.

Introduction²

Often patents are used as a tool to protect intellectual assets of companies and inventors. In 2012, 257,744 European patent filings were made at the European Patent Office (EPO)³. This number does not account for the filings in other offices, including European and non-European. Just in Espacenet⁴, a free of charge patent database, you have access to nearly 80 million patent documents, the majority consisting of patent applications and only a part of granted patents currently in force.

These documents provide technical information and many of the state of the art technologies can be found there. Companies can and should benefit from analysing this unique source of information. It is indeed considered by experts that much of the technical information contained in patents is not available anywhere else, and that it contains around 80% of the worldwide technical knowledge.

However, patent information analysis is not easy: patent documents are abundant, lengthy and are written in very technical language. Thus, reading and analysing patent documents can be complex and time consuming. This is where the use of Automatic Patent Analysis (APA) can help.

The strength of APA indeed resides in speeding up the analysis of patent documents and in its ability to develop new information, generally called “elaborated information”, to be used by companies in their internal business decision making process. Other advantages of APA are:

- Allowing for a rapid overview of technology sector patenting trends;
- Screening competitors’ patenting activities;
- Elaborating indicators that are not available anywhere else, such as the R&D and commercial strategies of companies, their collaborations and also identification of potential business partners.

It is therefore recommended that companies, including SMEs, perform such an analysis once or twice a year in order to understand the evolution of their technical and competitive environment. Patents also allow making the link between science and technology. Researchers should therefore be able, through APA, to integrate

² This fact sheet was first published in November 2013 and revised in December 2013

³ Further information on patent statistics at the EPO can be seen in its website: <http://www.epo.org/about-us/annual-reports-statistics/annual-report/2012/statistics-trends/patent-filings.html>.

⁴ <http://www.epo.org/searching/free/espacenet.html>.

patent information more easily into their activities, especially as patent information is easily available through many free of charge patent databases⁵.

1. What does Automatic Patent Analysis consist in?

Patent searches can lead to different results according to the type of search undertaken.

Indeed, a first way to search for patents is to try to get the most precise answer possible to an initial question, which leads to a search results page presenting the minimum patents possible and the most relevant answers (for example when searching for infringing patents, or performing a search related to an invention). In such conditions it is not possible to analyse the full environment of the terms used in the search query, that is to say to have a global idea of what is going on in the field (trends, technologies, applicants, etc.).

This introduces the second way to perform a patent search: to have the best overview of the subject analysed (for example when exploring a particular technological field, or for a state of the art). To this end, the search query must be large enough to embrace the various developments of the subject. In such a case, the number of answers to the query will be large. Consequently, it will not be possible by a simple reading of the results to get a complete analysis including possible correlations determined by combination of the various bibliographic fields present in the patent bibliographic references, such as: applicants or inventors names, dates, priority information, patent countries and International Patent Classification (IPC) codes. To get these answers it is necessary to use the APA.

1.1 One field analysis

The APA analysis is performed most of the time on sets of downloaded patent bibliographic references retrieved following the query of a patent database. Different types of analysis of these results can be performed. A simple analysis is the one field analysis based on lists.

In fact, all the data available in the patent bibliographic references (and specifically in the fields composing the reference) can be listed. Resulting lists can show very useful information on the set of patents/technology studied, such as:

- The trend in patent applications over years (importance of the subject analysed),
- The number of applicants (for example, high number of applicants at low frequency indicates a subject relatively open),

⁵ For more on patent information search, please see the European IPR Helpdesk fact sheet "How to search for patent information", available in the [library](#).

- The number of IPC codes with their frequencies (this indicates if the subject is scattered among different orientations),
- The number of inventors and frequencies (this indicates a possible threshold of investment in research),
- The number of patent applications filed at European or Worldwide levels (most valuable patents, the ratio with the other patents is a kind of competition index; this is also linked to the “triadic” patents),
- The listing of the vocabulary used in titles or abstracts (this allows spotting the best related patents to a sub-subject), etc.

Examples of information provided according to the kind of informational field analysed:

Field	What information can be retrieved
IPC Code	<ul style="list-style-type: none"> • Main and secondary technologies • Diversity of technology • Macro or micro vision depending on the truncation of the code (e.g. IPC code analysed at 4-digit level: Macro vision)
Patent Assignee	<ul style="list-style-type: none"> • Number of patents for a company • Leader, followers in technological sectors
Priority Date	<ul style="list-style-type: none"> • Number of patents per year for a company or a technological sector • Evolution of patent policy
Inventors	<ul style="list-style-type: none"> • Researchers for the sector
Citations	<ul style="list-style-type: none"> • Technological importance of inventions • Indications on sales and profits (for pharmaceutical industry) • Speed of innovation of companies

1.2 Multiple fields analysis: the matrices

Even if the lists provide interesting information, they do not give a deep understanding of the subject analysed. Combining different types of information (i.e. developing matrices) is a way to have a deeper analysis.

For example, a matrix combining the names of applicants with IPC codes (at 4 to 8 digits according the precision required) provides a direct benchmarking of companies; a matrix mixing, for a specific domain, the applicants’ names with the patent publication year provides elements on the involvement of the companies in this domain (if a company appears several time in different years this means a

strong involvement in the subject). The table provided under paragraph 1.3 shows several examples of cross-referred information.

1.3 The networks

Networks provide another representation of the correlations between two different bibliographic fields of the patent bibliographic notices, for instance:

- Link between co-applicants (network applicants/applicants),
- Links between inventors (network inventors/inventors),
- Inventors of applicants (network inventors/applicants),
- Technology development trends in the subject (network IPC/application date), etc.
- ...

The following table provides you with the type of information provided when cross-referring data from informational fields:

	Inventor	Priority Year	Priority Country	IPC Code	Patent Assignee	...
Inventor	Inventors network					
Priority Year	Inventors' activities evolution					
Priority Country	Inventors' country of origin	Evolution of the activity per country				
IPC Code	Experts per domain	Evolution of technologies	Activity fields per country	Links between technologies		
Patent Assignee	Link inventor – patent assignee	Evolution of firms' activities	Filing countries for firms (market)	Activity fields of companies	Collaborations between firms	
Patent Family		Evolution of the interests for national markets	Filing strategies at national levels	Activity fields exploited per country	Filing strategies	
Patent Family size		Evolution of patent extensions	Filing strategies at national levels		Filing strategies	

1.4 An advanced practice: the use of strategic groups

When you have to deal with hundreds or thousands of patents, the lists, matrices or networks constructed on the totality of the downloaded data are almost impossible to interpret. This is the reason why in such a case an approach using strategic grouping has been developed. The principle is to use the information contained in a list of relevant patents constructed according to, for example, countries, dates, IPC codes, words (vocabulary) from title or abstract, series of applicants or inventors... These patents constitute strategic groups integrated in the analytical process. On each of the groups, the same analysis can therefore be performed (lists, matrices, networks). This provides a deeper insight in the group and gives detailed and interpretable information.

For example, the network of the priority patents group elements analysed using the IPC codes (at 4 or 8 digits level) will help to discriminate the different approaches of R&D on the subject according to the different countries, etc.

2. Construction of a set of patent bibliographic references

Whatever the kind of analysis to perform, the first task to undertake is to gather a set of patent bibliographic references. The first step is to access the patent data and therefore to construct the set of patent references to be analysed.

2.1 Access to patent data

Patents are available in several ways; large organisations such as the EPO or WIPO as well as some National Patent Offices have opened the way to a free access of patent information on their respective websites. Some private initiatives have also emerged, such as for example Google that offers a free access to full text US patents.

Patent information is generally available either in the form of a full text document or in the form of a bibliographic reference of the full patent.

The material to be used in order to perform APA must be constituted of bibliographic references available from the patent databases and formatted with the same pattern for all the data.

To query patent databases, it is recommended to use a "boolean" combination of:

- keywords (foreseen to be present in the title or the abstracts),
- applicants and inventors names,
- International Patent Classification (IPC) codes

and if appropriate:

- patent number,
- priority number,
- patent dates,
- country codes (using the two first digits representing the countries: FR=France, US=USA, CN=China...-) either in the patent number code or the priority number code.

It is the combination of such elements that will help the users get the right information.

According to the database consulted, searches could be performed in English or another language.

For more information on patent information searching, we recommend you to refer to our fact sheet on "*How to search for patent information*"⁶.

Examples of free of charge patent databases⁷

- Google patent http://www.google.com/advanced_patent_search
- Espacenet <http://worldwide.espacenet.com/>
- USPTO <http://portal.uspto.gov/>
- Patentscope (WIPO) <http://patentscope.wipo.int/>
- Eurasian Patent Organization <http://www.eapo.org/en/>

Commercial patent databases

These are available on commercial database servers, such as:

- Proquest Dialog <http://www.proquest.com/products-services/ProQuest-Dialog-Patents-Collection.html>
- STN <http://www.cas.org/products/stn/dbss>
- Questel-Orbit <http://www.questel.com>

or as dedicated databases, such as:

- PatBase from MineSoft and RWS Group, which offers searching patents worldwide as well monitoring, analysis and other functions <http://www.patbase.com>
- Thomson Innovation (from Thomson Reuters) claims world's most complete source of patent data to deliver a complete patent solution (search, full text, analysis, history) <http://info.thomsoninnovation.com/>

⁶ The fact sheet "How to search for patent information" is available in the European IPR Helpdesk [library](#).

⁷ The examples of databases and tools in this fact sheet are not intended to be exhaustive.

- TotalPatent from LexisNexis, which offers searching on patents from 102 patent authorities and provides the way to analyse them among other characteristics <http://www.lexisnexis.com/>
- Matheo Software patent database contains more than 80 million patents worldwide based on the European Patent Office data <http://www.matheo-software.com/>

Types of databases to access patent information

Type of databases	Advantages	Disadvantages
Free of charge databases	<ul style="list-style-type: none"> • Free of charge 	<ul style="list-style-type: none"> • Geographic coverage • Patents not grouped by families • Very often query limitations • No mass download of patent bibliographic references
Commercial patent databases on multidisciplinary servers	<ul style="list-style-type: none"> • No limitation concerning the queries • Possibilities to save searches • Availability of APA tools • Sectoral databases are available • Some databases group patent families • Easy download of patent notices 	<ul style="list-style-type: none"> • Cost • When available, APA tool is not specific to patents
Commercial patent databases on specific servers	<ul style="list-style-type: none"> • No limitation concerning the queries • Possibilities to save searches • Availability of APA tools specifically developed • Some databases group patent families • Easy download of patent bibliographic references 	<ul style="list-style-type: none"> • Cost • When available, APA tool cannot be fully configured

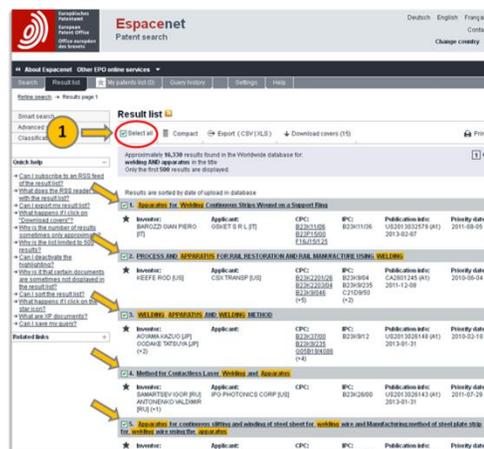
2.2 Download of patent bibliographic references

When using commercial databases, sets of bibliographic references are constructed directly as a result of the query used. Specific commands allow downloading sets of references or storing them in a dedicated space of the server to be used for further analysis. When using free of charge databases, the problem of data downloading is different. Free of charge solutions do not allow the advanced options as are available in fee-based solutions. For example, it is not possible to easily download references from Google patent or Patentscope.

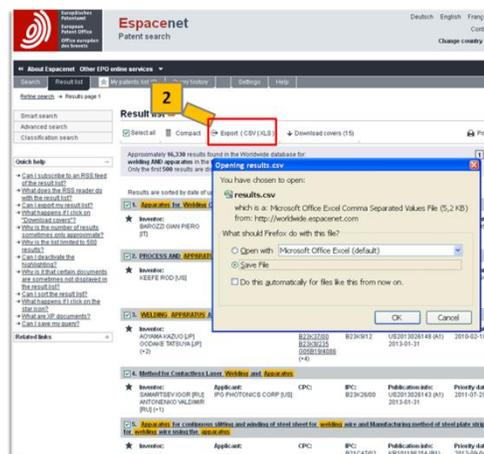
2.3 The example of Espacenet

A database that offers the opportunity to download patent bibliographic references is Espacenet. The procedure is as follows:

1. Click on the "Select all" box on your results pages to automatically select all the references of the result page you are consulting



2. Click on the "Export" option to export your data either in CSV (Comma Separated Value) or XLS (Excel) format



Processing in this way will generate several files that you will have to merge in order to get a full set of data available for further analysis.

TIP: When using Espacenet, whatever the format you choose to download your data, compact the results using the "Compact" option made available in the interface. It allows getting 30 results on one page, so to download 30 patent notices. To download further notices, you will have to connect to each result page and to re-iterate the downloading procedure

3. APA practical application

3.1. The software solutions

3.1.1. Online solutions

The most important commercial database servers have developed an offer integrating automatic analysis solutions directly into their interface. So, after having performed a patent search that provides hundreds of patent references as a result, it is possible to directly analyse the set of results using the online solutions.

Concerning patents, some of the main solutions made available are from:

- Questel-Orbit (<http://www.questel.com>)
- ThomsonReuters (<http://thomsonreuters.com>)
- STN International (<http://www.stn-international.de>)

3.1.2. Desktop software

Alternative solutions exist under the form of software to be installed on one's own computer. They either need the download of a set of patent bibliographic references from commercial servers or free of charge databases, or they can work directly with their own patent database.

Examples of commonly used solutions are:

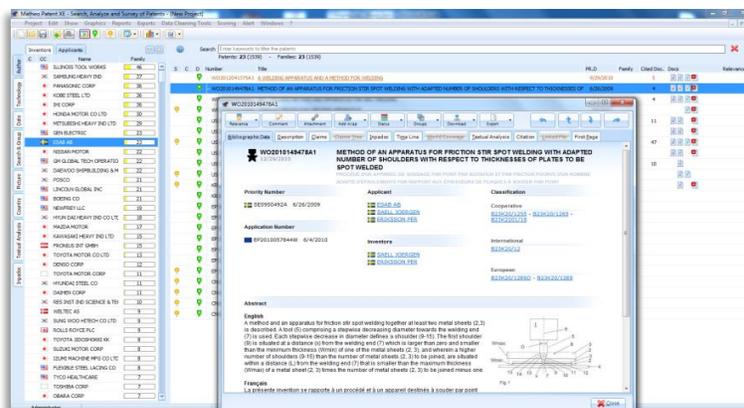
- Matheo Patent (<http://www.matheo-software.com>) - patent research and analysis platform that uses the Espacenet database and US Patents databases;
- Matheo Analyzer (<http://www.matheo-software.com>) - decision making software able to analyse any type of downloaded bibliographic information (patent or scientific for example);
- Patent Insight Pro (<http://www.patentinsightpro.com>) - patent research and analysis platform for any type of downloaded bibliographic information;
- Vantage point (<http://www.thevantagepoint.com>) - analysis solution for patent and scientific information.

Patent Databases vs Patent Software		
	Advantages	Disadvantages
Online solutions	<ul style="list-style-type: none"> • No need to install additional software • No need for data downloading (depending on the solution used, just connection costs) 	<ul style="list-style-type: none"> • Not always easily configurable (limitations in term of possible analyses) • Could be expensive • Need online connection
Desktop solutions	<ul style="list-style-type: none"> • Fully configurable (theoretically, any types of analysis can be performed) • No fees for data downloading (depending on the solution used) • Allows analysis to be performed offline • Full confidentiality 	<ul style="list-style-type: none"> • Need to install the software • Download of data may imply high costs depending on the database used

3.2. Illustrated example (using Matheo Patent)

A patent search about “welding apparatus” has been performed in the Espacenet Worldwide database. The query used was “welding AND apparatus” limited to the patent’s titles and also to the period 2000-2012. It provided as a result a total of 1,581 patent applications.

The set of patents has thus been downloaded into the local databases of the software for analysis using its manual, semi-automated and automated functionalities.



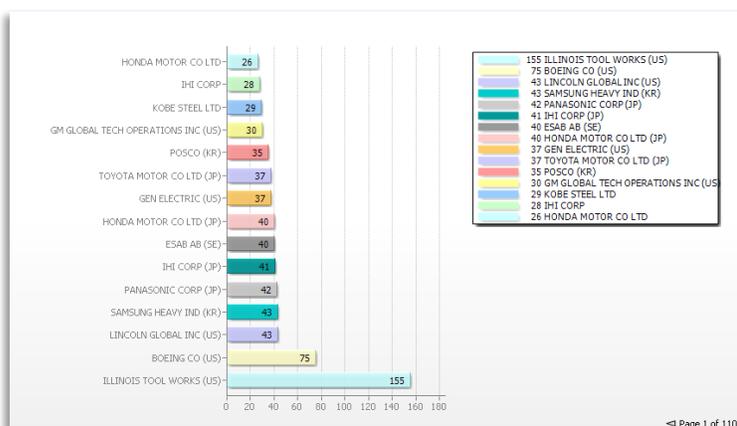
Matheo Patent main interface

Using the analysis facility of the Matheo software, different groups of welding technologies have been identified (using the titles, abstracts, claims, descriptions and IPC codes of the patents):

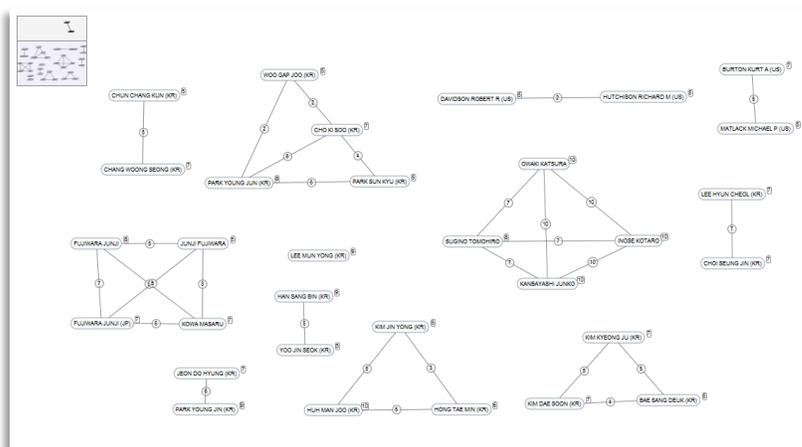
Groups	Number of families	Number of patents
Laser	153	408
Friction	63	187
Ultrasound	26	95
Vibration	17	48
Plasma	10	32

Thereafter, different analyses have been performed in order to better understand the environment of the different technological fields:

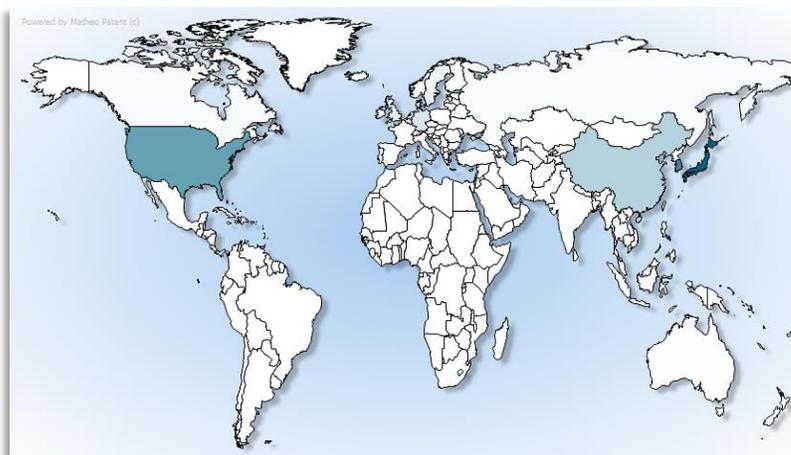
- Main applicants in the vibration group (top of companies filing patents in the area), by listing patent assignee's names



- The inventor network in the friction group that shows the relationships/collaborations between experts in a company or between different companies.



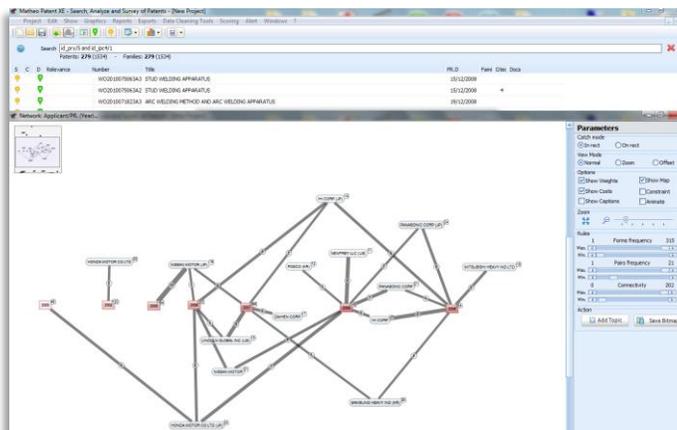
- Another interesting analysis is provided by the location of inventors. It shows countries of excellence in the area.



- Cross-referencing information on IPC and publication dates illustrates the trends of technologies. It highlights the technologies emerging, those declining as well as the relative evolution of core technologies. When based on a single company patent portfolio, such an analysis provides a complete mapping of the company's technological policy.

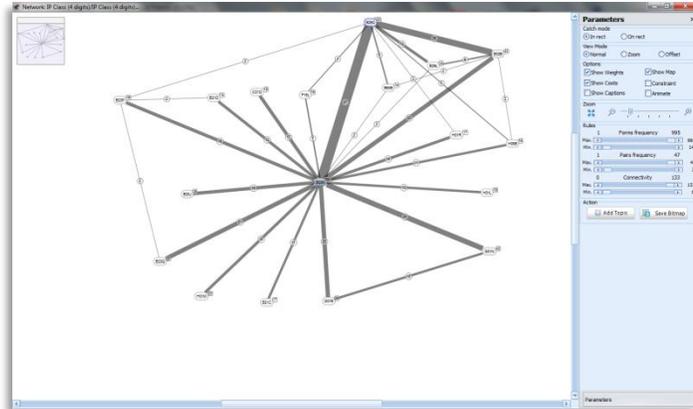
	B23K	B29C	B32B	H05B	B65B	H01R	H01M	H01L	F27B	B23P	B65G	H01S	H02M	G05D	F16L	G02B	B29L	B25J	B23Q	A61B	F27D	F42B	C21D	B05B	B65H
2012	12					1					1														
2011	90	6	3	1	2	1	3					1													
2010	175	14	3	4			2	2		1	2	1		1	2			1	2				1		
2009	220	26	3	6	1	3	2	7		3	4	1			4	4			2	2			1	1	2
2008	278	26	3	3	1	4	2	1	2		1	1	5	5	1	1	2	5	2					1	
2007	234	21	2	4		1	5	1	3		3	5	3	7		2		1		3					
2006	160	29	3	2	7	3	1	1		6	3	3	1		2	1	4	1		6			3	3	1
2005	143	28	3		2	1	2	3		1	2	3			1	3	1	1					4	4	4
2004	97	18	2	1		3		1	3		1	1			1	4									1
2003	42	5	1	4					4				1								4	4			
2002	35	1		1		3			6												6	6			
2001	11	1								1													1		
2000	7						1	1											1	1					
1999	15	1			1	1							2												
1998	10	2			2														1						1
1997	4																		1						
1996	2																								
1995	1																								

- The evolution of players in a technological field: newcomers, core companies and companies giving up could be provided when analysing information related to applicants over the years. The main interest is to obtain in a single map the global understanding and trends of a competitive technological sector.



- In this analysis, the network illustrates the different technologies and the intensity of links between the technologies (always, often, sometimes or never together). This analysis is used to cluster the different technologies and consequently to create technological groups among a patent set.

The network visualisation is based on two visual effects: the size of the box containing the information (number of patents having a technology) and the intensity of the line linking two boxes (number of patents having two technologies simultaneously).



GET IN TOUCH

For comments, suggestions or further information, please contact

European IPR Helpdesk
c/o infeuurope S.A.
62, rue Charles Martel
L-2134, Luxembourg

Email: service@iprhelpdesk.eu

Phone: +352 25 22 33 - 333

Fax: +352 25 22 33 - 334



©istockphoto.com/Dave White

ABOUT THE EUROPEAN IPR HELPDESK

The European IPR Helpdesk aims at raising awareness of Intellectual Property (IP) and Intellectual Property Rights (IPR) by providing information, direct advice and training on IP and IPR matters to current and potential participants of EU funded projects. In addition, the European IPR Helpdesk provides IP support to EU SMEs negotiating or concluding transnational partnership agreements, especially through the Enterprise Europe Network. All services provided are free of charge.

Helpline: The Helpline service answers your IP queries within three working days. Please contact us via registration on our website – www.iprhelpdesk.eu – phone or fax.

Website: On our website you can find extensive information and helpful documents on different aspects of IPR and IP management, especially with regard to specific IP questions in the context of EU funded programmes.

Newsletter and Bulletin: Keep track of the latest news on IP and read expert articles and case studies by subscribing to our email newsletter and Bulletin.

Training: We have designed a training catalogue consisting of nine different modules. If you are interested in planning a session with us, simply send us an email at training@iprhelpdesk.eu.

DISCLAIMER

This Fact Sheet has been initially developed under a previous edition of the European IPR Helpdesk (2011-2014). At that time the European IPR Helpdesk operated under a service contract with the European Commission.

From 2015 the European IPR Helpdesk operates as a project receiving funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 641474. It is managed by the European Commission's Executive Agency for Small and Medium-sized Enterprises (EASME), with policy guidance provided by the European Commission's Internal Market, Industry, Entrepreneurship and SMEs Directorate-General.

Even though this Fact Sheet has been developed with the financial support of the EU, the positions expressed are those of the authors and do not necessarily reflect the official opinion of EASME or the European Commission. Neither EASME nor the European Commission nor any person acting on behalf of the EASME or the European Commission is responsible for the use which might be made of this information.

Although the European IPR Helpdesk endeavours to deliver a high level service, no guarantee can be given on the correctness or completeness of the content of this Fact Sheet and neither the European Commission nor the European IPR Helpdesk consortium members are responsible or may be held accountable for any loss suffered as a result of reliance upon the content of this Fact Sheet.

Our complete disclaimer is available at www.iprhelpdesk.eu.

© European Union (2015)