# A SHORT HISTORY AND OUTLOOK FOR THE PATENT DATA AND ANALYTICS INDUSTRY

**PATENT KNOWLEDGE WEEK**
**NOVEMBER 2021**

**Associate Professor Frank Tietze**
**Head, Innovation and IP Management (IIPM) Laboratory**

frank.tietze@eng.cam.ac.uk

@franktietze

# DISCLAIMER

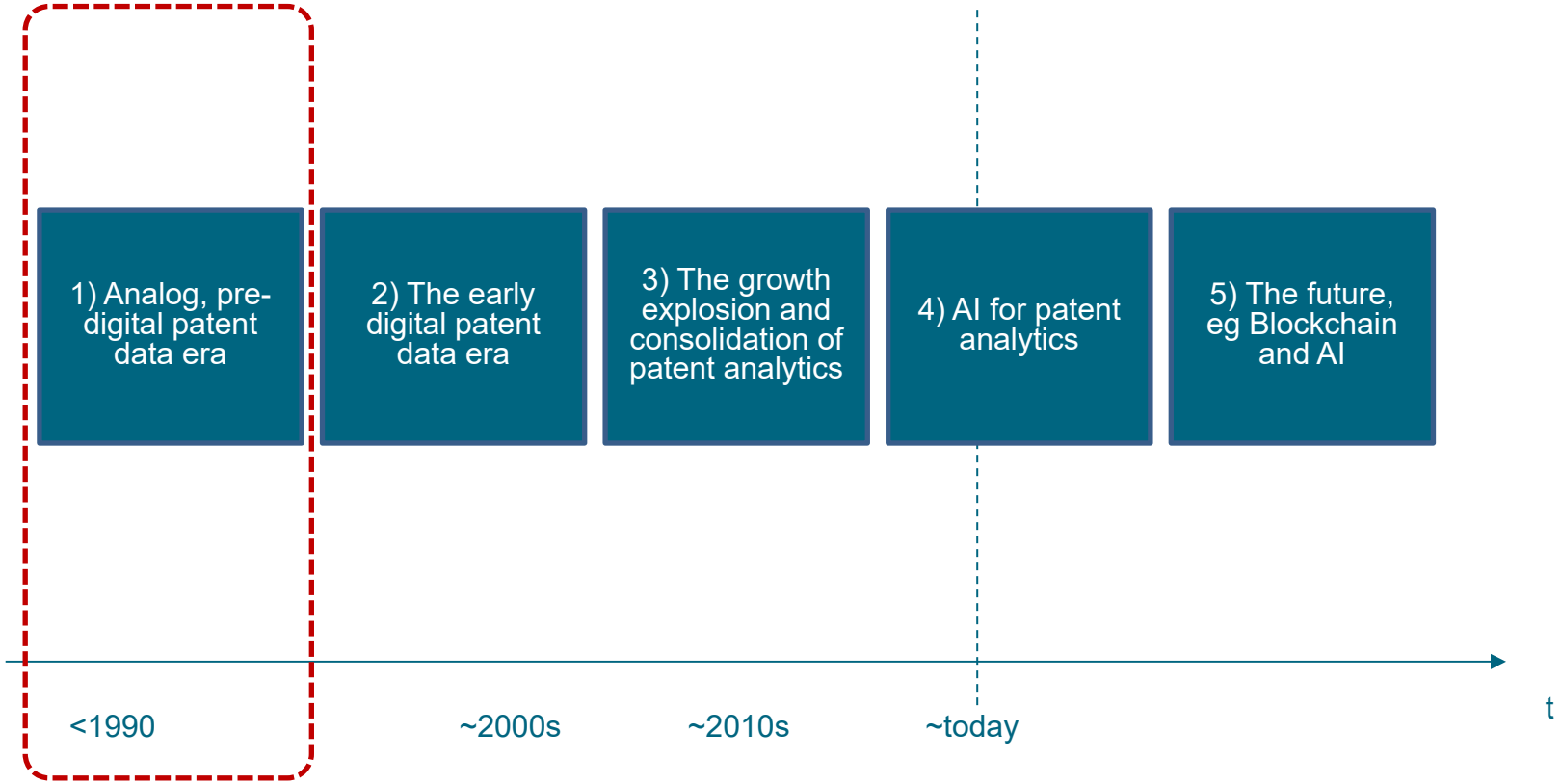
This talk is no attempt to accurately capture the history of the patent data industry

This talk is based on personal observations and interpretations, rather than a systematic collection of evidence

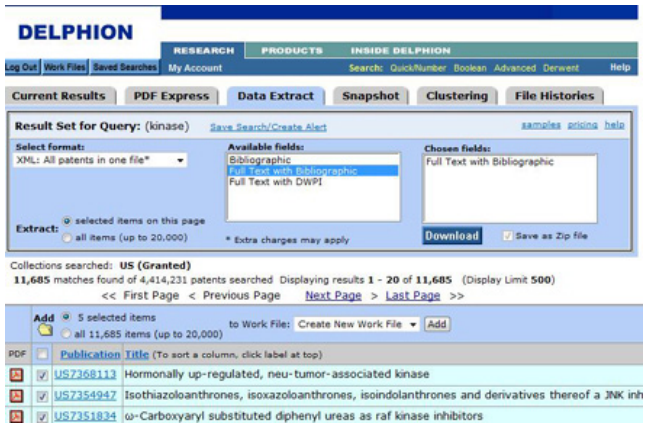
Any mentioning of companies in this presentation is not to be understood as an endorsement

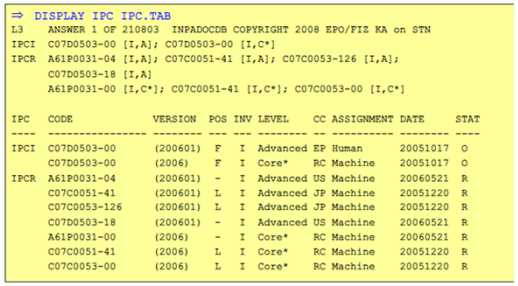
Accordingly, failing to mention any company is not meant to be a dis-endorsement

# A CHRONOLOGY OF THE PATENT DATA/ANALYTICS INDUSTRY



# EARLY DAYS: BIRTH OF FUTURE INCUMBENTS




- Delphion started life as an offshoot of IBM, the Intellectual Property Network in 1997.
- In 2002, it was acquired by Thomson

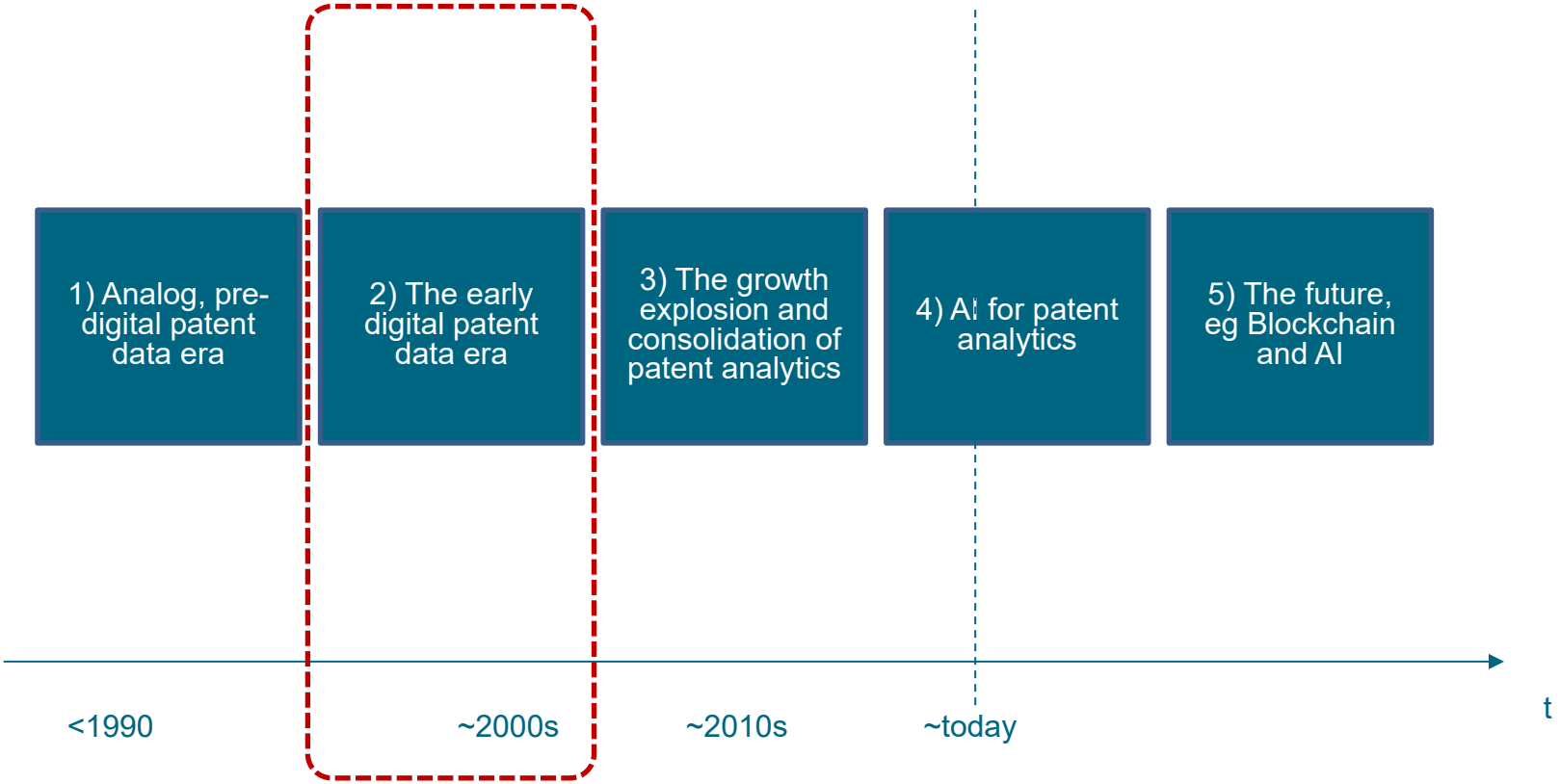
THOMSON

STN®
THE CHOICE OF PATENT EXPERTS™

# KEY LEARNINGS

- High search costs associated with using patent information
- Patent information profession rather focused on identifying and analysing individual patents
- Patent data as tool for large firms
- Large value in making the raw data available digitally, also at improved data quality (e.g. DWPI)

# A CHRONOLOGY OF THE PATENT DATA/ANALYTICS INDUSTRY



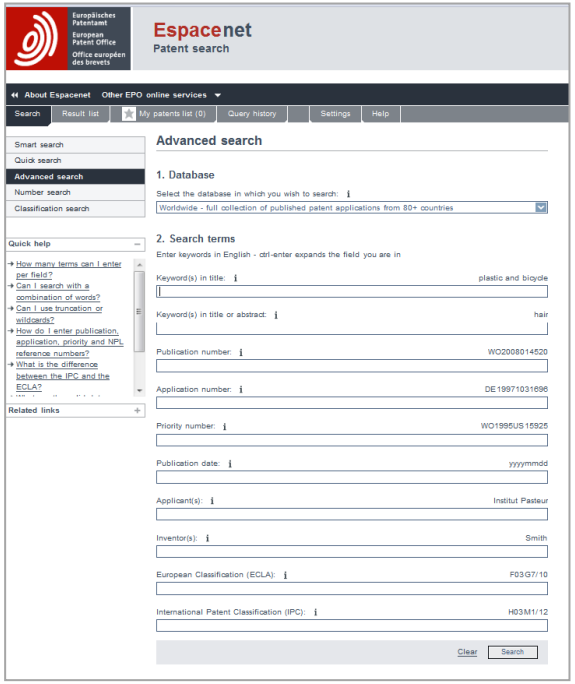
# COSTS TO ACCESS PATENT DATA STARTED TO DROP WITH THE DIGITIZATION OF PATENT DOCUMENTS

- 1984: BACON project (BAckfile CONversion) by the Trilateral offices
- 1992: TechSource project by Canadian IP office
- WIPOscan project to support less developed countries with digitization software

# DEMOCRATISING ACCESS TO PATENT DATA

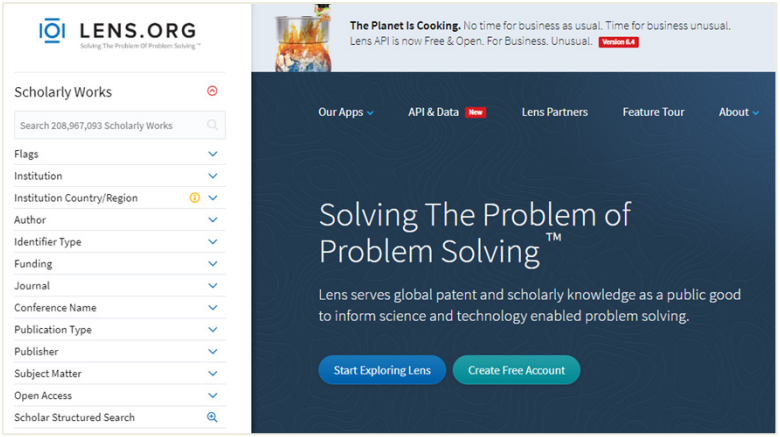
## Espacenet

Launched in 1998



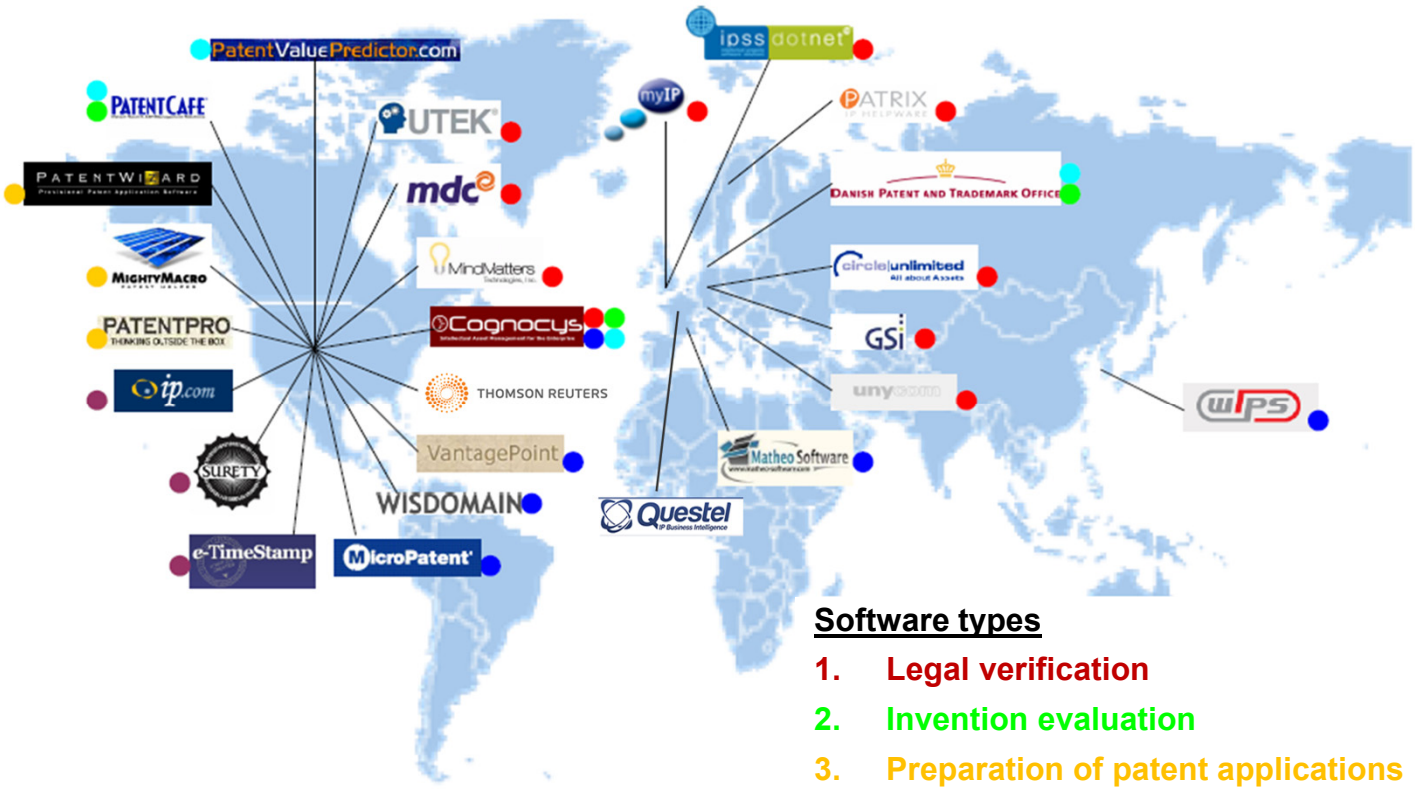

## The Lens.org

Launched in 2000



# SINCE THE DIGITIZATION OF PATENT DATA FIRMS INCREASINGLY DEVELOP TOOLS TO SUPPORT EFFICIENT PATENT INFORMATICS

"The number, depth and breadth of software tools for conducting patinformatics exercises have grown extensively over the past 5–10 years." *Trippe, A. (2003: 221)*




**Software types**

1. **Legal verification**
2. **Invention evaluation**
3. **Preparation of patent applications**
4. **Administration of IP rights**
5. **Patent valuation**
6. **Analysis of patent portfolios**

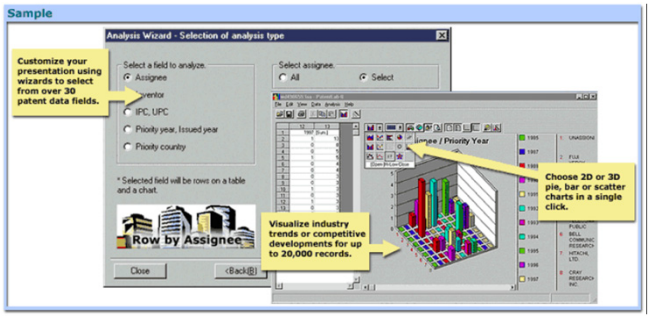
Source: Andrée, Tietze (2007)

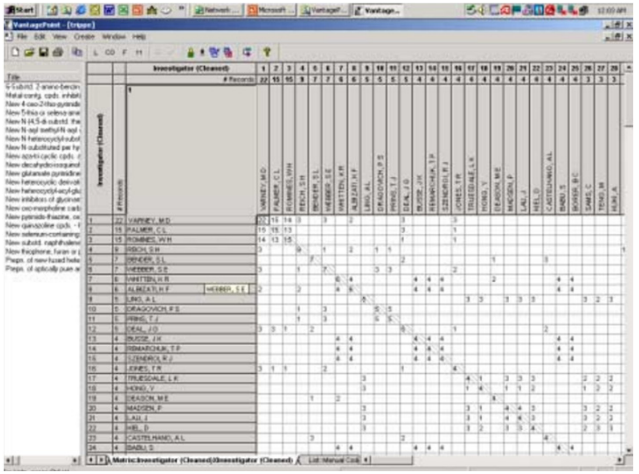
# DEFINING PATENT INFORMATICS (2002)

- Introduced by Trippe:

**What is Patinformatics?**

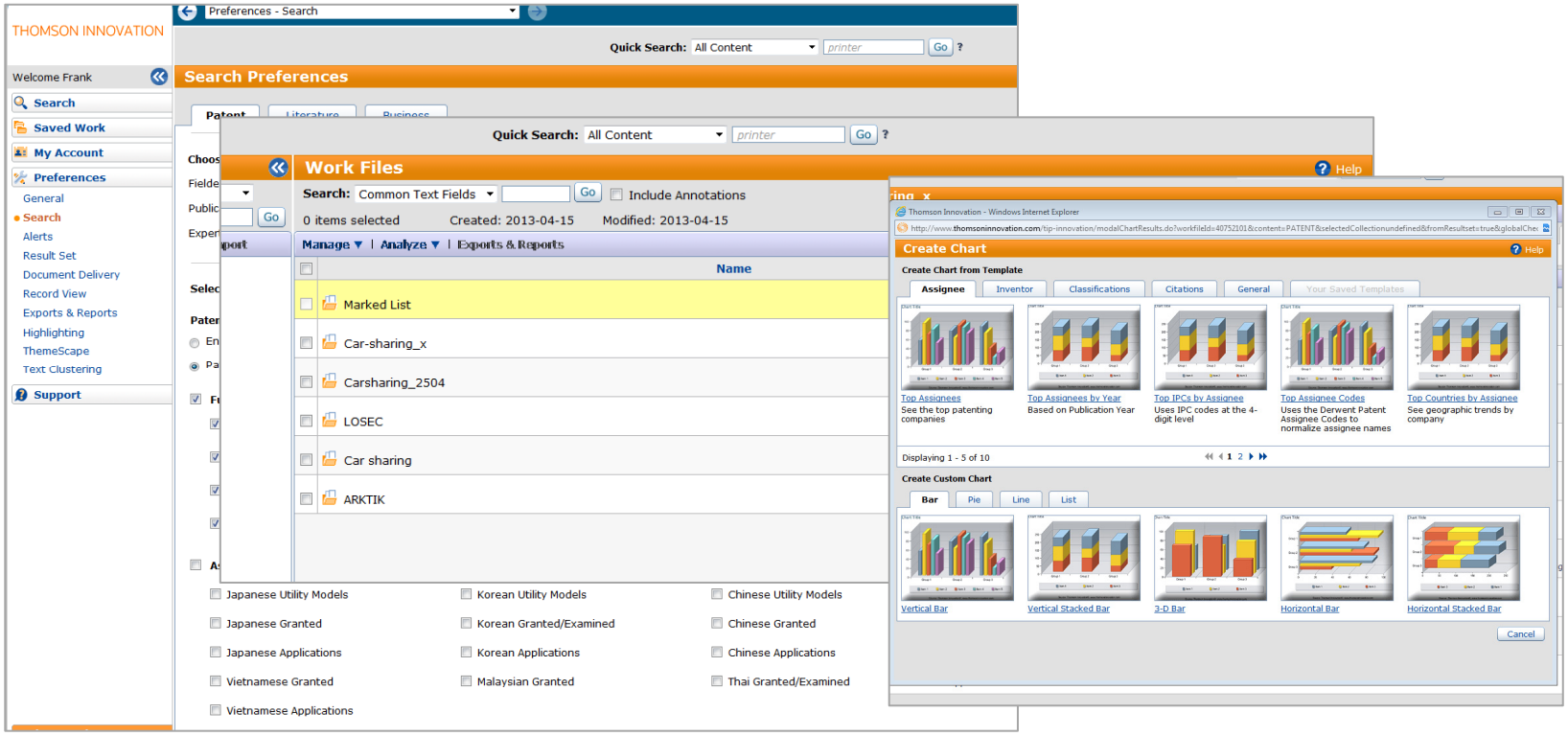
- Patinformatics
  - the science of analyzing patent information to discover relationships and trends, which would be difficult to see when working with patent documents on a one-on-one basis. The term encompasses all forms of analyzing patent information including:

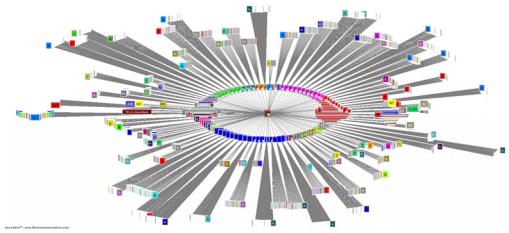


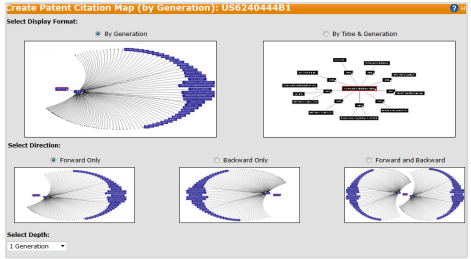



- In contrast to „classical“ in depth, one-by-one analysis of individual patent documents carried out by patent lawyers, patent informatics employ quantitative approaches to search and analyze patent data using dedicated software solutions.

# INDUSTRY INCUMBENTS FOLLOW STARTING TO OFFER SIMPLE, MOSTLY DESCRIPTIVE ANALYSES



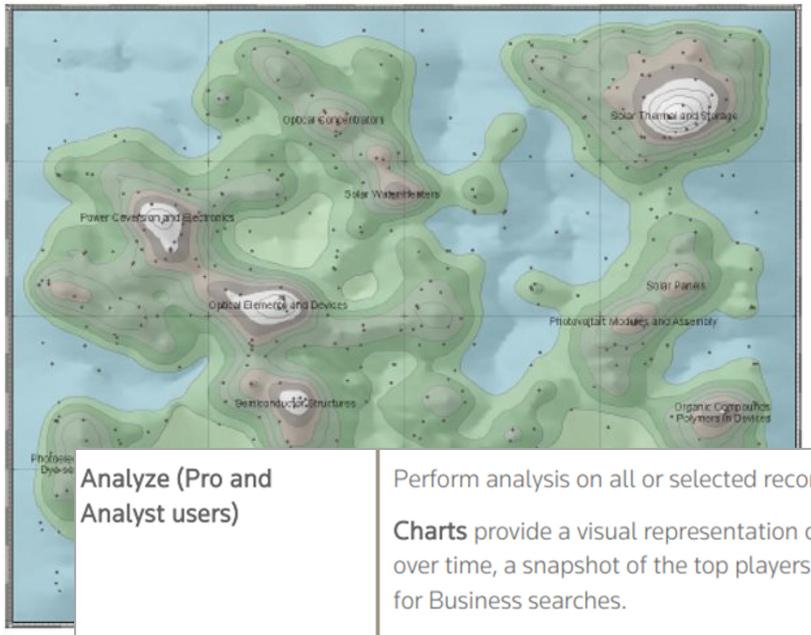





# FIRST VISUAL TOOLS FOR CREATING PATENT LANDSCAPE MAPS (~2002)

## Example of a Themescape Map

Themescapes are a powerful way to illustrate patent activity in any given technology area




| | | |
|---|---|---|
| | Search Histories | The current search history or a saved search history (identity icon) |
| | Watched Records | Watched record (identity icon) |
| | Work Files | Work file (primary or combined, see details panel for content type) (identity icon) |
| | Personal Folders<br>Public Folders | Personal or public folder (identity icon) |
| | Citation Maps | Citation map (identity icon) |
| | Charts | This fully-colored icon represents a saved chart (identity icon) |
| | Charts | This ghosted/outline icon represents a chart template (identity icon) |
| | Export Templates | Export template (exports and reports themselves are downloads and available via the Order Status screen in My account, not in saved work) (identity icon) |
| | ThemeScape Maps | Un-published ThemeScape map (identity/state icon) |
| | ThemeScape Maps | Published ThemeScape map (identity/state icon) |

| | |
|---|---|
| Analyze (Pro and Analyst users) | Perform analysis on all or selected records.<br><br>**Charts** provide a visual representation of your data, helping you see things like how a patenting landscape changes over time, a snapshot of the top players in a field, or how the top names change over the years. Charts are not available for Business searches.<br><br>**ThemeScape** creates content maps, displaying common conceptual terms in a two-dimensional map, with peaks representing a concentration of documents and showing the relative relationship of one record to another. ThemeScape is only available to Analyst level subscribers.<br><br>**Text Clustering** is a powerful analysis tool that automatically categorizes records through the linguistic analysis of text found in user-selected fields. Text Clustering is only available to Analyst level subscribers. |

# FIRST PATENT DATA INDICATORS PROPOSED BY THE ACADEMIC COMMUNITY IN THE EARLY 2000S

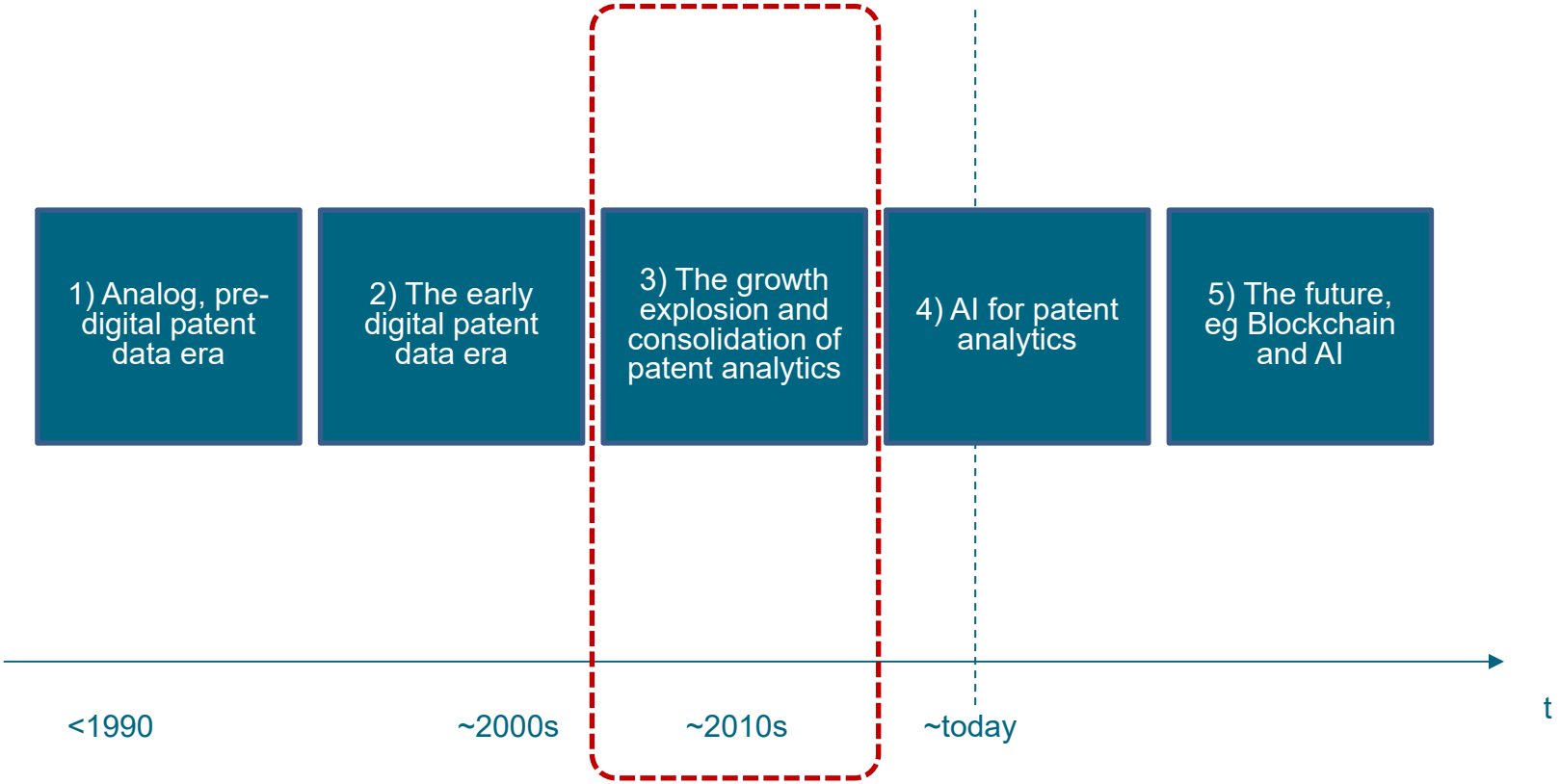
| Patent indicator | Definition | Meaning |
|---|---|---|
| Patent activity ($PA_{iF}$) | Patent applications (PA) of firm $i$ in technological field (TF) $F$ | Extent of R&D expenditures of firm $i$ in TF $F$ (interest of firm $i$ in TF $F$) |
| Technology share (based on patent applications) | $PA_{iF}$/PA of all competitors in TF $F$ | Competitive technological position of firm $i$ in TF $F$ (quantitative) |
| R&D emphasis | $PA_{iF}$/Number of firm's ($i$) total patent applications | Importance of technological field $F$ for firm $i$ (R&D emphasis) |
| Co-operation intensity | Number of joint patent applications with partners in TF $F$/$PA_{iF}$ | Access of firm $i$ to external knowledge (and identification of partners) |
| Share of granted patents ($Q_1$) | Granted patents of firm $i$ in TF $F$/$PA_{iF}$ | Technological quality of firm $i$'s patent applications |
| Technological scope ($Q_2$) | Diversity and number of IPC classes in firm $i$'s patent applications ($PA_{iF}$) | Technological quality of firm $i$'s patent applications |
| International scope ($Q_3$) | Size of patent family and share of triad (US, JP and EPO) patents of $PA_{iF}$ | Economic quality of firm $i$'s patent applications |
| Citation frequency ($Q_4$) | Average citation frequency of $PA_{iF}$ | Economic quality of firm $i$'s patent applications |
| Average patent quality ($PQ_{iF}$) | Sum of all indicators of patent quality ($Q_1$–$Q_4$) | Average total quality of all patent applications of firm $i$ in TF $F$ |
| Patent strength ($PS_{iF}$) | Product of average patent quality ($PQ_{iF}$) and patent activity ($PA_{iF}$) | Technological strength of firm $i$ in TF $F$ |
| Technology share (based on patent strength) | $PS_{iF}$/PS of all competitors in TF $F$ | Competitive technological position of firm $i$ in TF $F$ (qualitative) |
| Relative technology share | $PS_{iF}$/Max. patent strength of a firm in TF $F$ | Distance of firm $i$ to the technological leader in TF $F$ |

Source: Ernst, H. (2003)

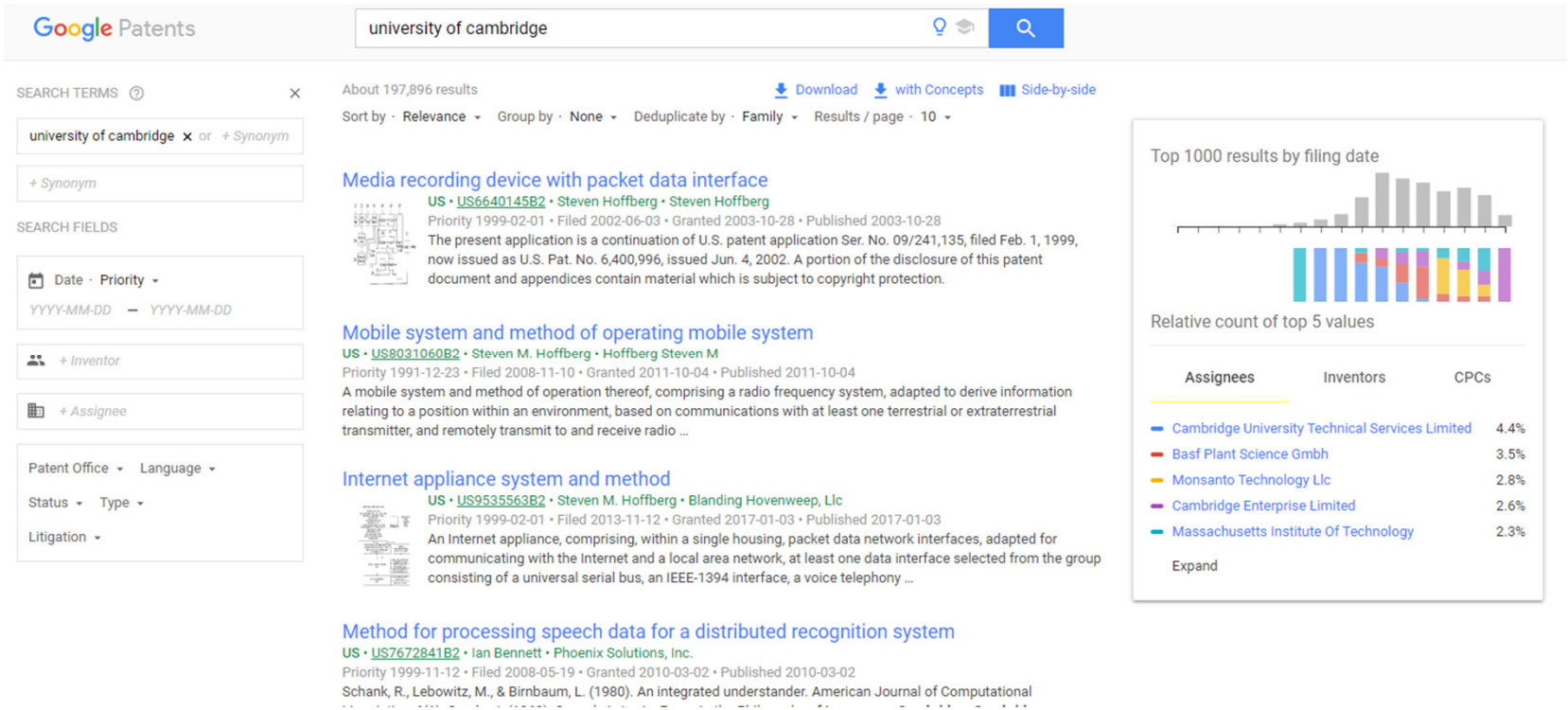
# KEY LEARNINGS

- Patent documents increasingly available digitally as PDF, but still not available in the cloud
- The value of data as basis for business models started to diminish
- Raw patent data quality remained relatively weak
- Industry still dominated by a few established patent data software providers
- The then pioneering firms build a business by providing digital patent data of higher quality, eg Derwent index
- Very limited (if at all only descriptive) capabilities for analysing patent data at scale / on portfolio level

# A CHRONOLOGY OF THE PATENT DATA/ANALYTICS INDUSTRY

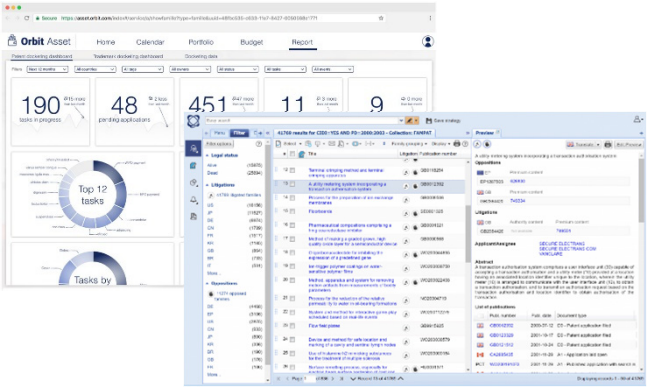


# LAUNCH OF GOOGLE PATENT (DEC. 2006)

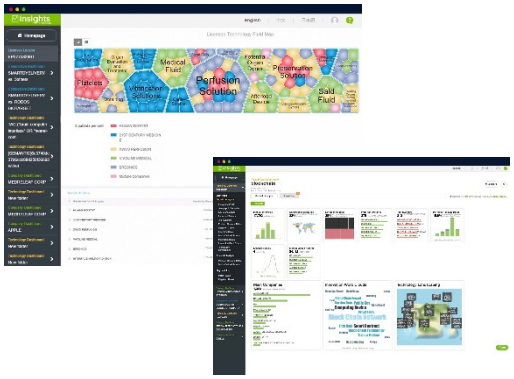


# NEW AGGRESSIVE-GROWTH PLAYERS ENTERING THE INDUSTRY

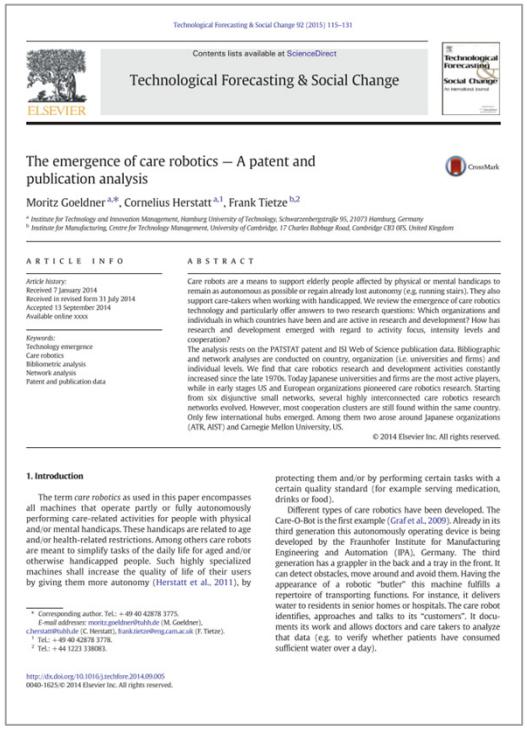
- Questel created in 1978 as subsidiary of Orange S.A., formerly France Télécom
- In 2001, it spun off from France Telecom and focused on IP
- With premises in Europe, US, and Asia, it has become a worldwide leader through an active acquisition strategy and constant innovations becoming standard in the industry.




- PatSnap founded in 2007
- Based out of the UK and Singapore, with locations in China and the U.S..
- The company started out as essentially a directory for IP, helping companies — and particularly enterprises — pull in data for R&D and product development purposes



# AN EARLY COMBINED ANALYSIS OF PATENT AND PUBLICATION DATA



Technological Forecasting & Social Change 92 (2015) 115–131

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

The emergence of care robotics — A patent and publication analysis

Moritz Goeldner [a,*], Cornelius Herstatt [a,1], Frank Tietze [b,2]

[a] Institute for Technology and Innovation Management, Hamburg University of Technology, Schwarzenbergstraße 95, 21073 Hamburg, Germany
[b] Institute for Manufacturing, Centre for Technology Management, University of Cambridge, 17 Charles Babbage Road, Cambridge CB3 0FS, United Kingdom

ARTICLE INFO

*Article history:*
Received 7 January 2014
Received in revised form 31 July 2014
Accepted 13 September 2014
Available online xxxx

*Keywords:*
Technology emergence
Care robotics
Bibliometric analysis
Network analysis
Patent and publication data

ABSTRACT

Care robots are a means to support elderly people affected by physical or mental handicaps to remain as autonomous as possible or regain already lost autonomy (e.g. running stairs). They also support care-takers when working with handicapped. We review the emergence of care robotics technology and particularly offer answers to two research questions: Which organizations and individuals in which countries have been and are active in research and development? How has research and development emerged with regard to activity focus, intensity levels and cooperation?
The analysis rests on the PATSTAT patent and ISI Web of Science publication data. Bibliographic and network analyses are conducted on country, organization (i.e. universities and firms) and individual levels. We find that care robotics research and development activities constantly increased since the late 1970s. Today Japanese universities and firms are the most active players, while in early stages US and European organizations pioneered care robotics research. Starting from six disjunctive small networks, several highly interconnected care robotics research networks evolved. However, most cooperation clusters are still found within the same country. Only few international hubs emerged. Among them two arose around Japanese organizations (ATR, AIST) and Carnegie Mellon University, US.



**1. Introduction**

The term *care robotics* as used in this paper encompasses all machines that operate partly or fully autonomously performing care-related activities for people with physical and/or mental handicaps. These handicaps are related to age and/or health-related restrictions. Among others care robots are meant to simplify tasks of the daily life for aged and/or otherwise handicapped people. Such highly specialized machines shall increase the quality of life of their users by giving them more autonomy (Herstatt et al., 2011), by protecting them and/or by performing certain tasks with a certain quality standard (for example serving medication, drinks or food).

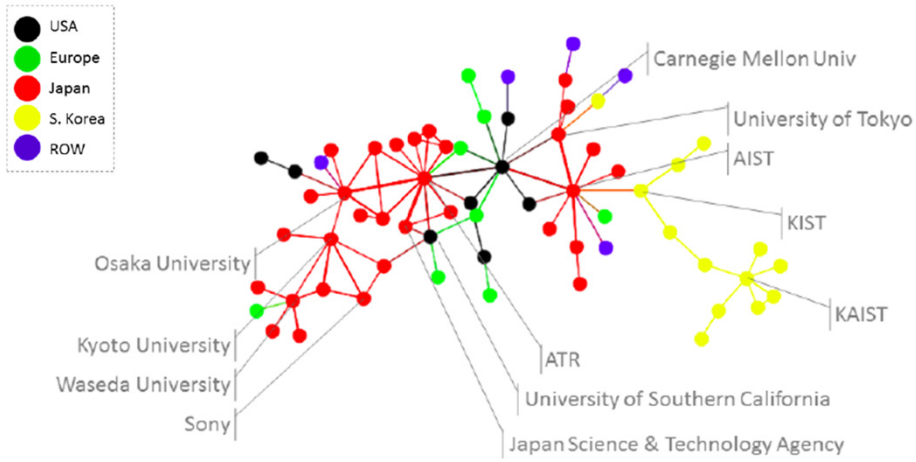
Different types of care robotics have been developed. The Care-O-Bot is the first example (Graefet al., 2009). Already in its third generation this autonomously operating device is being developed by the Fraunhofer Institute for Manufacturing Engineering and Automation (IPA), Germany. The third generation has a grappler in the back and a tray in the front. It can detect obstacles, move around and avoid them. Having the appearance of a robotic "butler" this machine fulfills a repertoire of transporting functions. For instance, it delivers water to residents in senior homes or hospitals. The care robot identifies, approaches and talks to its "customers". It documents its work and allows doctors and care takers to analyze that data (e.g. to verify whether patients have consumed sufficient water over a day).

* Corresponding author. Tel.: +49 40 42878 3775.
*E-mail addresses:* moritz.goeldner@tuhh.de (M. Goeldner), cherstatt@tuhh.de (C. Herstatt), frank.tietze@eng.cam.ac.uk (F. Tietze).
[1] Tel.: +49 40 42878 3778.
[2] Tel.: +44 1223 338083.

http://dx.doi.org/10.1016/j.techfore.2014.09.005
0040-1625/

**Table 6**
Top 10 ranked individuals for overall patent (left) and publishing activities (right).

| Rank | Individual | Organization country | – Pat. appl. | Rank | Individual | Organization country | – Sci. publ. |
|---|---|---|---|---|---|---|---|
| 1 | Ryota, H. | Mitsubishi (JP) | 9 | 1 | Breazeal, C. | MIT (US) | 19 |
| 2 | Ishiguro, H. | Osaka Univ. (JP) | 8 | 1 | Takanishi, A. | Waseda Univ. (JP) | 19 |
| 3 | Qiang, H. | Beijing Inst Tech (CN) | 7 | 2 | Kajita, S. | AIST (JP) | 17 |
| 3 | Akishi, K. | Kubota Corp (JP) | 7 | 3 | Ishiguro, H. | Osaka Univ. (JP) | 16 |
| 3 | Onishi, K. | Mitsubishi (JP) | 7 | 4 | Inaba, M. | Univ. Tokyo (JP) | 15 |
| 3 | Yamaguchi, J. | Sony (JP) | 7 | 5 | Cheng, G. | ATR (JP) | 14 |
| 4 | Gomi, H. | Honda Motor Co (JP) | 6 | 5 | Inoue, H. | Univ. Tokyo (JP) | 14 |
| 4 | Hiroaki, K. | Tech Co (JP) | 6 | 6 | Kagami, S. | AIST (JP) | 13 |
| 4 | Miyashita, K. | ATR (JP) | 6 | 6 | Kanda, T. | ATR (JP) | 13 |
| 4 | Furuta, T. | Tech Co (JP) | 6 | 6 | Okuno, H.G. | Kyoto Univ. (JP) | 13 |
| Other | | | 495 | Other | | | 1372 |
| Total | | | 564 | Total | | | 1525 |



# PROBLEMS PERSISTED WITH THE RAW PATENT DATA (~2015)

For instance,

- Corporate tree problem
- Incomplete inventor/application details


| Issue | Example |
|---|---|
| Spelling variations | "IBM" and "I.B.M." |
| Typographical errors | "INTERNATIONAL BUSINESS MACHINES" and "INTERATIONAL BUSINESS MACHINES |
| Addition of legal form | "IBM", "IBM CORP.", "IBM CORPORATION" and "IBM COPRORATION" |
| Errors | "INTERNATIONAL BUSINESS MACHINES" and "INTELLIGENT BUSINESS MACHINES" |
| Addition of establishment, business unit, department, subsidiary name or geographic identifier | "IBM" and "IBM JAPAN" |
| Acronyms | "IBM" and "INTERNATIONAL BUSINESS MACHINES" |

## Why the need to harmonize?

- Applicant and inventor names in patent databases: idiosyncratic inputs
- No standardized format
- Use of different name variants within and across databases
- Spelling variations, typos, legal form addition, abbreviations, etc.
- E.g. 658 name variants (~ 1.068 PERSON_IDs) of "I.B.M"; 488 name variants (~ 1.491 PERSON_IDs) of "PANASONIC CORPORATION"

Name harmonisation for better search results, Julie Callaert – KU Leuven ECOOM, EPOPIC 2015

# VARIOUS OPEN SOURCE TOOLS FOR SEARCHING AND ANALYSING PATENTS

World Patent Information 54 (2018) S4–S10

Contents lists available at ScienceDirect

World Patent Information

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

ELSEVIER

## The basics of patent searching

Nigel S. Clarke

*European Patent Office, Postfach 90, 1031 Vienna, Austria*

ARTICLE INFO

*Article history:*
Received 26 September 2016
Received in revised form
30 January 2017
Accepted 27 February 2017

*Keywords:*
Patent
Searching
Basics
Mind-set
Concepts

ABSTRACT

Some aspects of patent searching are similar to searching other kinds of technical literature, such as articles in popular science magazines or papers in professional and academic journals. However some aspects of searching patents are sufficiently different from searching conventional literature, that a dedicated mindset, strategy and tactics need to be applied to the task. The difference between searching conventional publications and patents arise, amongst other things, from the nature of the patent documents themselves, the interrelations between them and the unique way in which patent databases are constructed. This article describe an approach, and initial techniques for searching patents, which will enable the absolute beginner to make a start, and gain confidence is searching patent literature.



**1. Introduction**

This article is based on a presentation delivered by the author at the Search Matters Seminar held at the European Patent Office (EPO) in The Hague in June 2016. The presentation was commissioned by the Search Matters organisers expressly to cater for those participants who were complete newcomers to searching patents. This article covers some useful basic concepts in patent searching and some "tips and tricks" which hopefully the absolute beginner might find helpful. The article also incorporates ideas presented by the author over many years, in the "European Patent Information Beginners' Seminar" series and in introductory webinars [1,2].

The article is also about the psychology of patent searching, the mind-set, concepts, strategies and tactics, and is deliberately intended to be a general introduction only. It is not intended to promote any particular search tools or databases. However, by way of example only, search results and techniques in this article are illustrated by use of EPO tools and databases.

**2. The 64,000 dollar question: why search patents?**

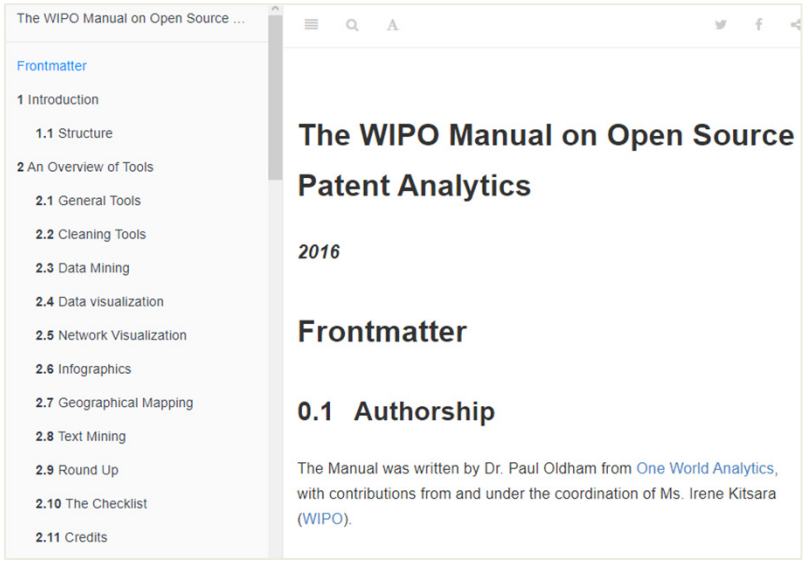
The trivial answer to this question is "because they are there" [3]. However there are many genuine and justifiable reasons for searching patents [4]; to find out about the most recent inventions, to study the development of a particular technology, or to find the patents by an eminent academic, for example. The publication of patent documents, which necessarily contain a full technical description of the invention, is a mandatory step in the patent process. Patent offices around the world are required by convention to publish the patents registered with them. *A patent cannot be a patent unless it is published* [5].

Most patent offices publish patent *applications* before any other publications. The existing volume of patent documents is huge. Publicly available free patent databases, for example Espacenet, PatentScope and DepatisNet [6] give access to tens of millions of patent documents. The cumulative volume of patent documents is growing rapidly too; 2.9 million patent applications are reported as having been filed in 2015 worldwide [7]. The patent process requires that all of these be published. Despite the volume and rapid growth in publications, searching patents is assisted by indexing or classification schemes which allow you to focus your search direction early on. Patents not only contain technical information about inventions. Patents are legal instruments and confer rights and privileges on their owners. So it may be that you would want to identify and characterise the legal status of particular patents. The patent documents also contain contact information about the applicants, the people or organisations who filed the patent application in the first place. Patents also list inventors, the clever individuals who had the ideas, and did the work in laboratories or factories or workshops. Inventors and applicants can be identified by searching patents, and this information can be commercially important.

*E-mail address:* nclarke@epo.org.

http://dx.doi.org/10.1016/j.wpi.2017.02.006
0172-2190/

https://www.sciencedirect.com/science/article/pii/S0172219016301030X



The WIPO Manual on Open Source ...



## The WIPO Manual on Open Source Patent Analytics

*2016*

## Frontmatter

### 0.1 Authorship

The Manual was written by Dr. Paul Oldham from One World Analytics, with contributions from and under the coordination of Ms. Irene Kitsara (WIPO).

https://wipo-analytics.github.io/index.html

# CHANGING PROFESSION: FROM PATENT INFORMATION RETRIEVAL EXPERT TO PATENT AND IP ANALYST

## Become a Qualified Patent Information Professional (QPIP®)

# KEY LEARNINGS

- Data available in the cloud

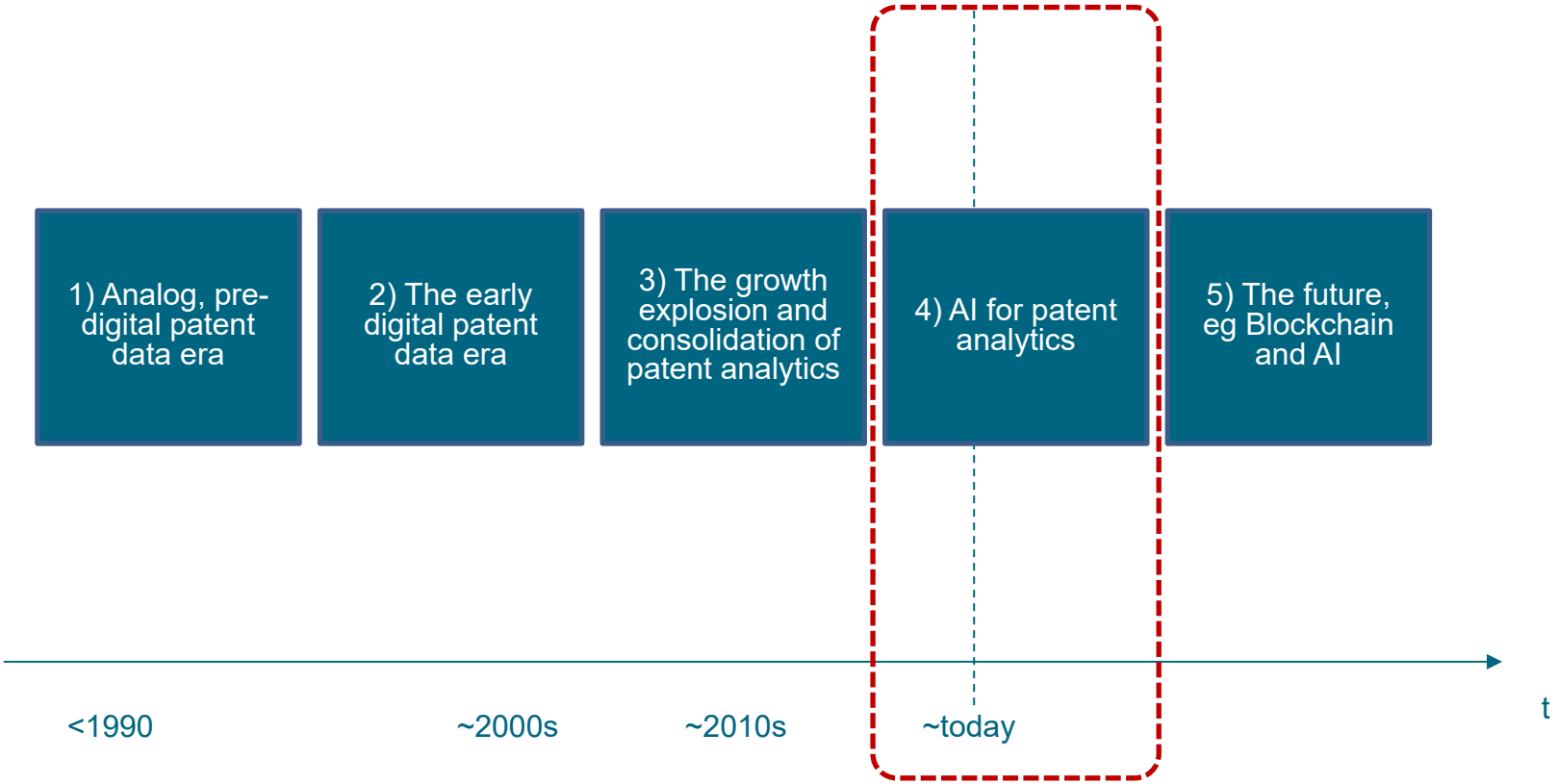
- High number of innovative new entrants
- Few companies with aggressive growth strategies

- Lower search costs for accessing patent data e.g. through online databases

- Democratized patent data – low value and competitive advantage from data alone

- Value proposition changes towards value in analytic capabilities

# A CHRONOLOGY OF THE PATENT DATA/ANALYTICS INDUSTRY



# A LARGE RANGE OF COMMERCIAL AND OPEN SOFTWARE TOOLS FOR ANALYSING PATENT DATA



## Patent Analysis, Mapping, and Visualization Tools

Created by Vinod Kumar Singh, last modified by Eric Podlogar on Mar 04, 2019

This page was created to encourage PIUG wiki participation in sharing knowledge about various tools and techniques relating to patent analysis, r brief style of other pages under the Patent Resources pages. Please leave details and marketing phrases to the websites to which the entries are li

**CandorMap -** Using Big-data, text mining technology CandorMap created a global semantic map of patents and scientific articles according to their con patent on its location), helping to understand its technological, IP and business significance. Users can navigate the map and retrieve all relevant informat actions can directly be drawn. www.candormap.com

**Clearstone FTO -** Clearstone FTO is an end-to-end freedom-to-operate and patent clearance management platform. The web-based application i reporting, and workflow management. Integrated patent analysis tools for efficient claim-by-claim assessment and multi-dimensional analytics bri pioneered by patent attorneys; based in California. www.clearstoneip.com

**Derwent Data Analyzer** - A desktop software solution for managing and extracting business-critical insights from patent and scientific data withi competitors, avoid or uncover patent and copyright infringement, and identify strategic development opportunities. It provides simple tools to im analyze data with a variety of tools and quickly create various custom reports.

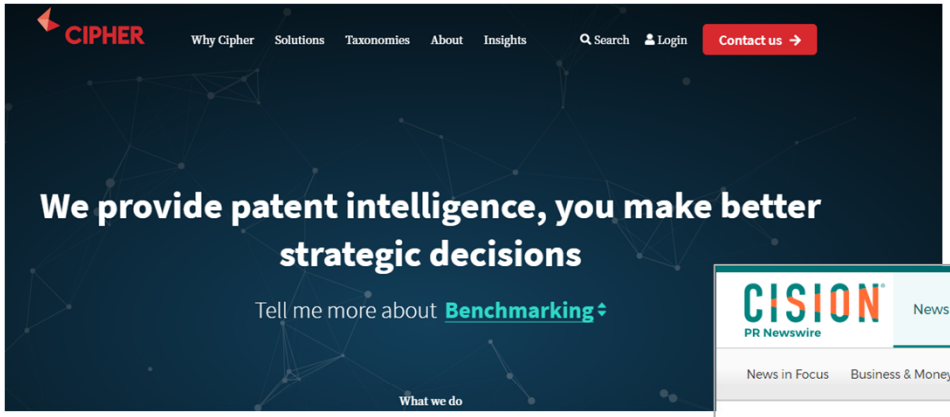
**Derwent Innovation** - A single, integrated solution that combines intellectual property, scientific literature, business data and news with analytic, enables searching of full text patent data from major patent authorities including English-language Chinese, Japanese and Korean information tog capabilities include Themescape mapping to provide interactive technology landscape maps and clustering, charting & citation mapping tools for

**IPriori IP Analytics Dashboard** (IPriori) - Current statistics and graphs on US patent litigation, reexamination, Top Patents In Suit by Class, Recent auctions. Statistics is based on IPriori's ePriori IP Knowledgbase.

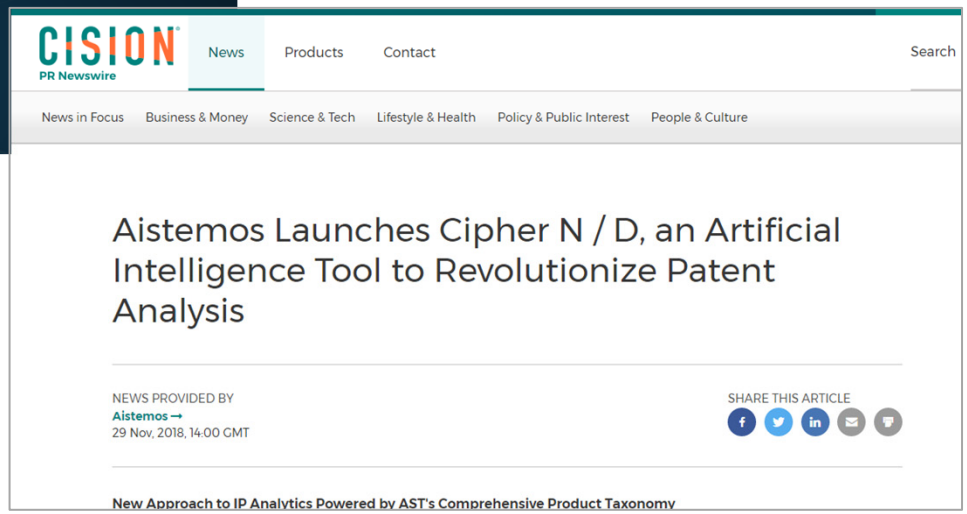
**Innography** - Innography provides better answers to questions about intellectual property to help organizations improve their business results. Ir technologies to enable users to quickly gain valuable insights for managing, extending and exploiting their patent portfolios. Founded in 2007 anc including Best Legal Information Service. Learn more and see demos at www.Innography.com.

**INTELLIXIR System** - A web application that provides patent and non-patent literature analysis. Used by large companies in various sci-tech dom highlight collaboration networks, identify experts in technological fields and a lot more. Everything is illustrated by numerous interactive graphical

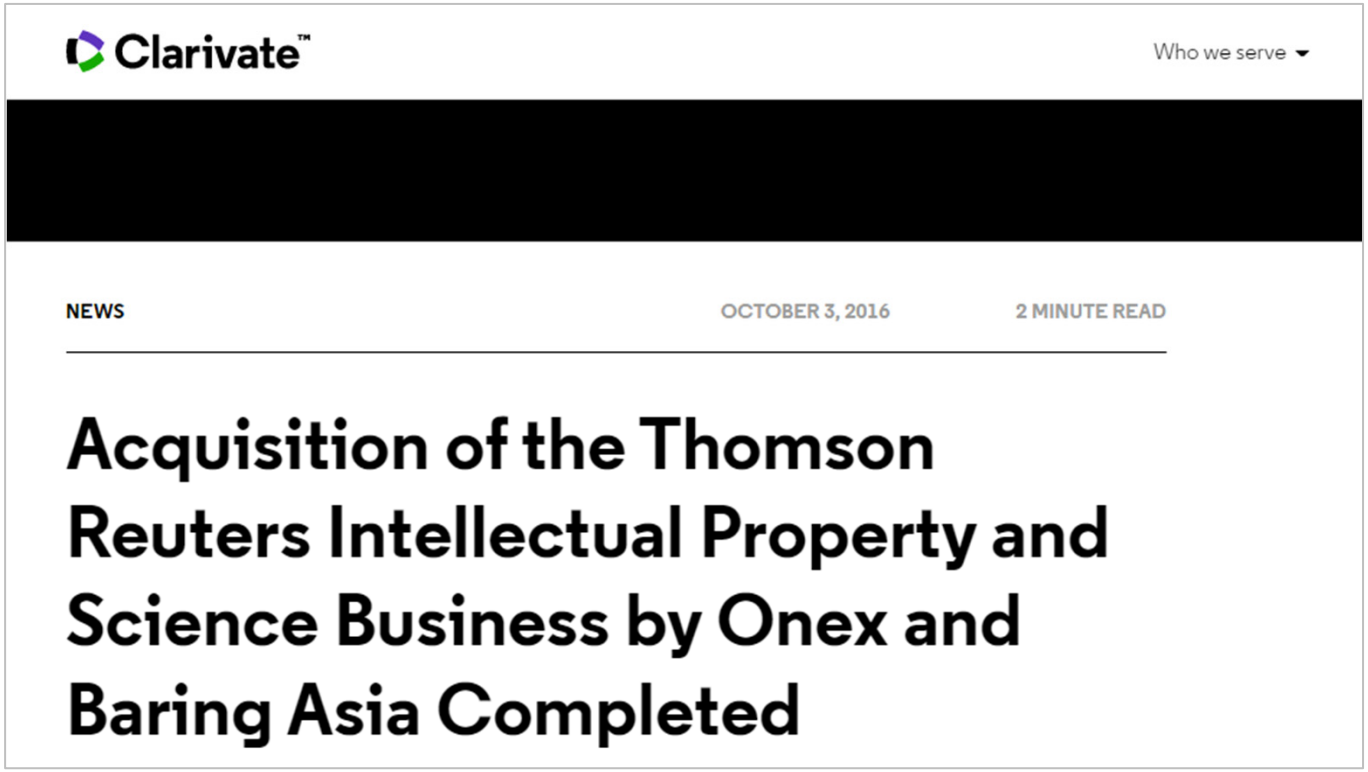
# SOME PIONEERING START-UPS WITH DEDICATED TECHNOLOGY ARCHITECTURES OPTIMISED FOR AI TO RUN ON PATENT DATA




Founded in 2013 and launched in 2014



# THE LARGE PLAYERS CONTINUE TO PLAY A ROLE, BUT CAN THEIR LEGACY ARCHITECTURES ADJUST TO THE USE OF AI?



Clarivate™

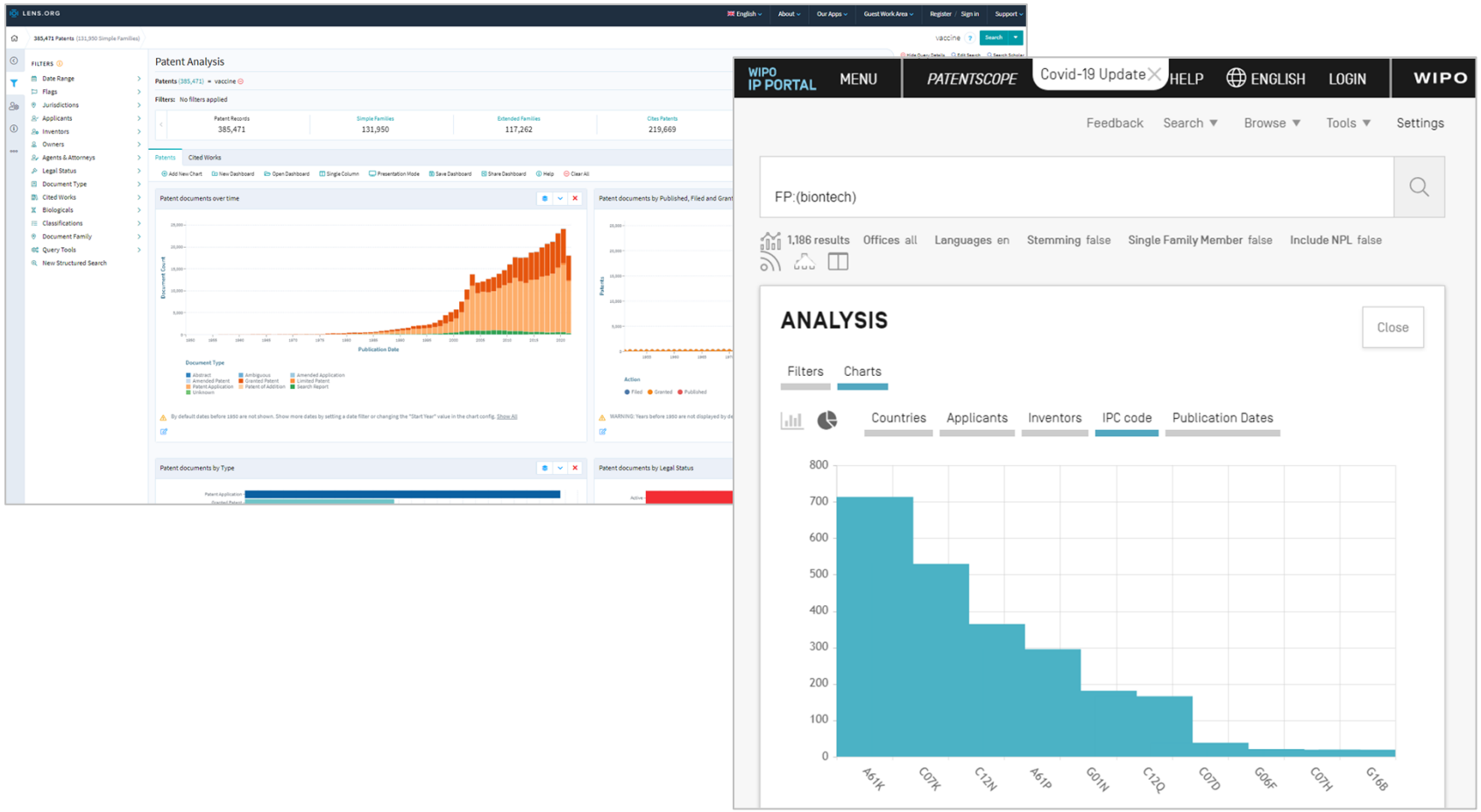
Who we serve

NEWS OCTOBER 3, 2016 2 MINUTE READ

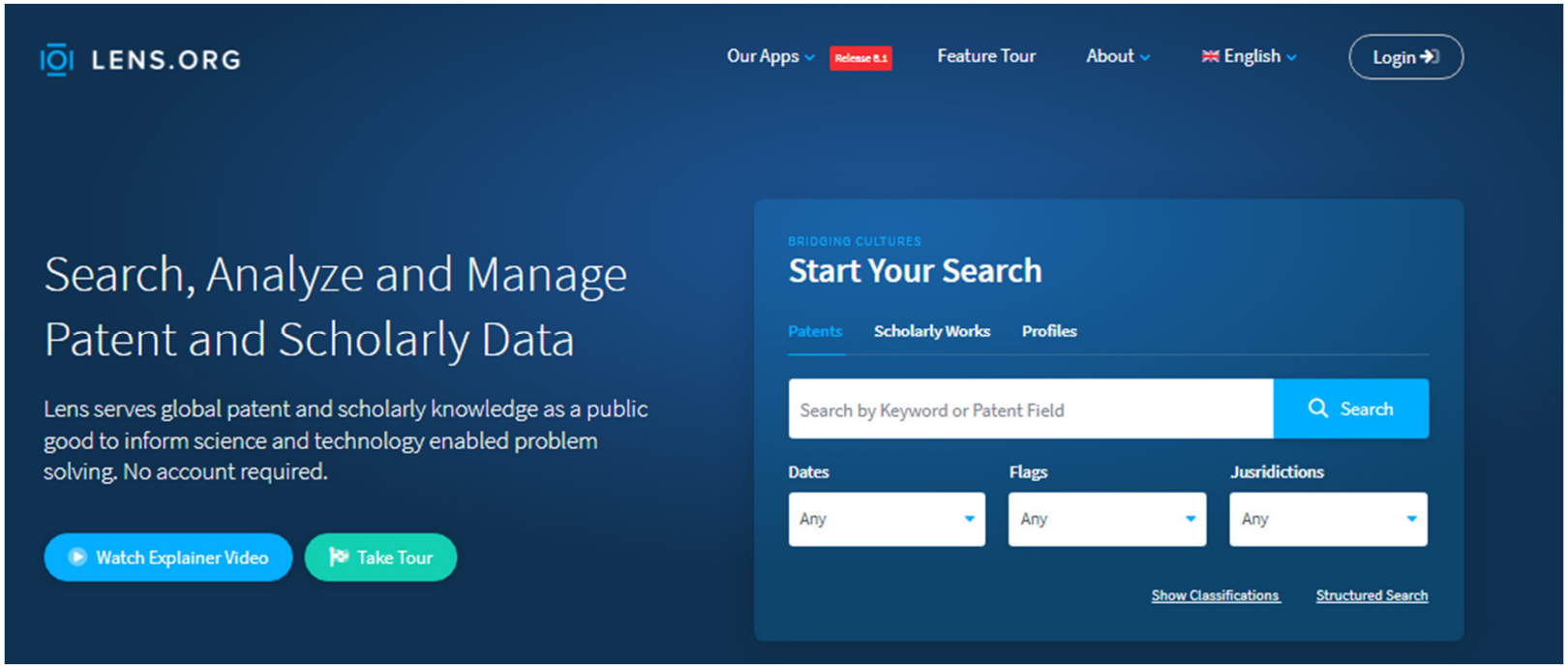
## Acquisition of the Thomson Reuters Intellectual Property and Science Business by Onex and Baring Asia Completed

https://clarivate.com/news/acquisition-thomson-reuters-intellectual-property-science-business-onex-baring-asia-completed/

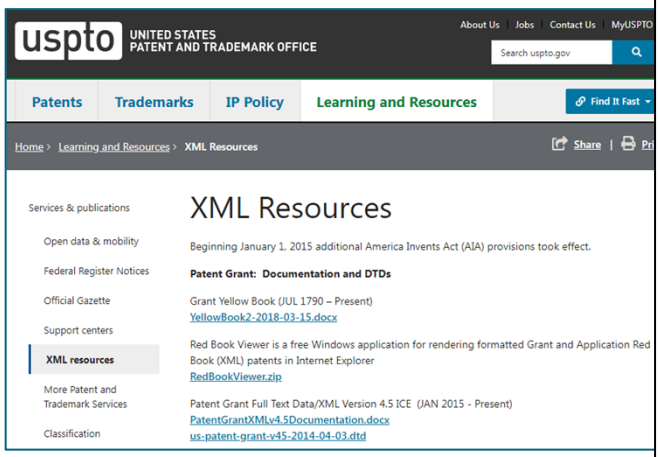
# DEMOCRATIZING PATENT ANALYSIS

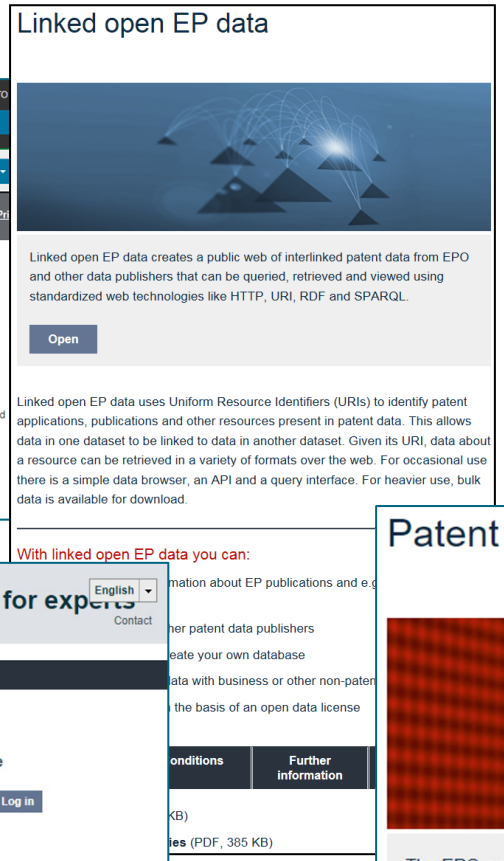


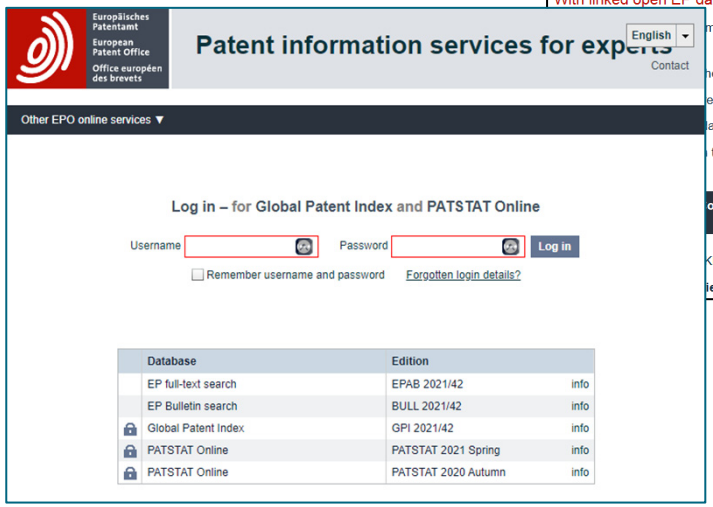
# PUBLICLY AVAILABLE INTEGRATION OF PATENT AND PUBLICATION (SCHOLARLY WORKS) DATA



# PATENT DATA IS BEING UPGRADED





Google Big Query

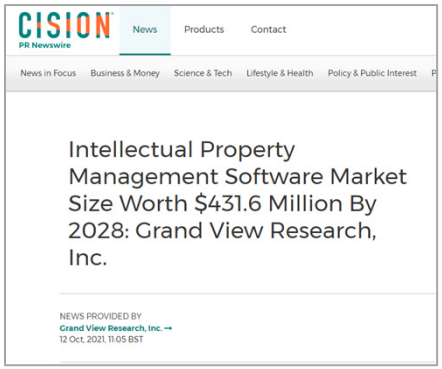
## Google Big Query

- Fast SQL queries/ Sampling large data
- Standardised means of access

# KEY LEARNINGS

- Fully democratized patent data
- Focus is on analysis large amounts of patent data, e.g. patent portfolios, emerging technologies
- Substantially lower search costs for accessing patent data e.g. through various free/open source tools
- Patent data analytics increasingly become a tool also for non-IP experts
- Patent information profession is adapting
- Data upgrades → new, more efficient ways of “crunching” large amounts of patent data
- New entrants build dedicated and AI compatible architectures for analysing patent data

# A CHRONOLOGY OF THE PATENT DATA/ANALYTICS INDUSTRY




Capturing value from offering digital patent data

Capturing value from offering patent analytics

1) Analog, pre-digital patent data era

2) The early digital patent data era

3) The growth explosion and consolidation of patent analytics

4) AI for patent analytics

5) The future, eg Blockchain and AI

Business models commercialise the data quality / making it available digitally

Business models that create value from offering insight tools/ analysis

???

<1990

~2000s

~2010s

~today

t

# WILL THE FUTURE VALUE PROPOSITION LIE IN THE ALGORITHMS?

"Defining IP intelligence (IPI) as the data science of analysing large amount of IP information, specifically patent data, with artificial intelligence (AI) methodologies to discover relationships and trends in the data for decision making." *

World Patent Information 55 (2018) xxx–xxx

Contents lists available at ScienceDirect

World Patent Information

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

The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data

Leonidas Aristodemou*, Frank Tietze

Intellectual Property and Innovation Management (IPM) research group, Centre for Technology Management (CTM), Institute for Manufacturing, Department of Engineering, University of Cambridge, United Kingdom

ARTICLE INFO

*Keywords:*
Intellectual property analytics
Patent analytics
Machine learning
Deep learning
Artificial intelligence

ABSTRACT

Big data is increasingly available in all areas of manufacturing and operations, which presents an opportunity for better decision making and discovery of the next generation of innovative technologies. Recently, there have been substantial developments in the field of patent analytics, which describes the science of analysing large amounts of patent information to discover trends. We define Intellectual Property Analytics (IPA) as the data science of analysing large amount of IP information, to discover relationships, trends and patterns for decision making. In this paper, we contribute to the ongoing discussion on the use of intellectual property analytics methods, i.e artificial intelligence methods, machine learning and deep learning approaches, to analyse intellectual property data. This literature review [illegible] where we present the state-of-the-art in intellectual property analytics [illegible] information of the articles are analysed, followed by a discussion [illegible] knowledge management, technology management, economic value, and [illegible] information. We hope research scholars and industrial users, may find this review [illegible] latest research efforts pertaining to intellectual property analytics.

1. Research background

Big data is increasingly available in all areas of manufacturing and operations [1]. Data as such presents value for enabling a competitive data-driven economy, which is at the heart of the Internet of things and Industry 4.0 [2,3]. Increased data availability presents an opportunity for better decision making and strategy development [4], to introduce the next generation of innovative and disruptive technologies [5].

Over the last two decades, there have been substantial developments in the field of patent analytics. Patent analytics describes the science of analysing large amounts of intellectual property information, in relation to other data sources, to discover relationships and trends [6–9]. With the digitization of patent data, the world's largest repository of technical information has become accessible for rapidly decreasing costs. While patent data has long been considered the world's largest repository of technological information, and only with its digitization since the BACON project in 1984 [10] and numerous gradual and cumulative improvements of the data quality and analytical techniques over the last decades, patent data has become increasingly accessible to and useful for a non-specialist audience [11]. With the rise of artificial intelligence (AI), and the increase in the usage of methods such as machine learning (ML) and deep learning (DL), a number of these have been applied to analyse IP data [6,12–14].

In a recent study, we have used the technology roadmapping approach [15] to explore the future of patent analytics [16,17]. We identify 11 priority technologies, such as artificial intelligence and artificial neural networks, that industry experts believe to be important to be adopted at a higher rate in the patent analytics domain [12]. While other domains have adopted such technologies already widely, the patent analytics domain seems to be catching up. We identify the need of adoption of these computer science techniques, to complement decision processes and provide decision support [11,12,18]. This is very much in line with the propositions by Refs. [19–22].

In this paper, we contribute to the ongoing discussion on the use of machine learning and deep learning approaches to analyse intellectual property data [12], by presenting the outcomes of a literature review on the state-of-the-art on intellectual property analytics. In particular, we focus the literature review on the use of artificial intelligence

* Corresponding author.
*E-mail addresses:* la324@cam.ac.uk (L. Aristodemou), frank.tietze@eng.cam.ac.uk (F. Tietze).

https://doi.org/10.1016/j.wpi.2018.07.002
Received 1 February 2018; Received in revised form 16 April 2018; Accepted 7 July 2018
0172-2190/ © 2018 Published by Elsevier Ltd.

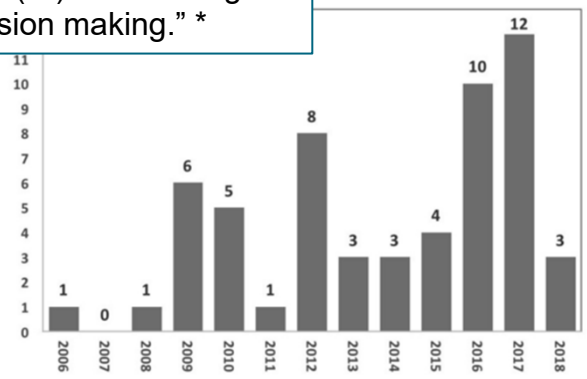



**Fig. 2.** Number of articles per year ($n_1 = 57$) since 2006 (< 2006 = 0 articles).

**Table 1**
Top 10 affiliations ($n_1 = 57$ articles, $n_2 = 128$ observations).

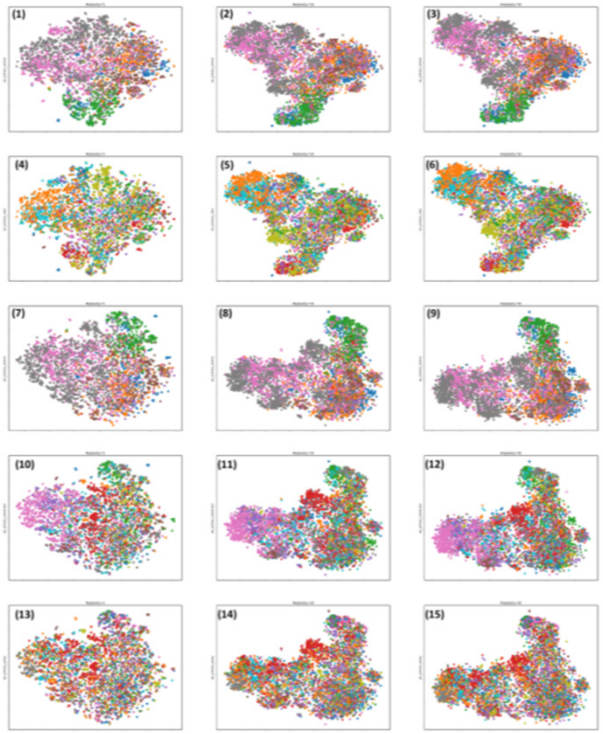
| Affiliation | No. of observations | Share of total (%) |
|---|---|---|
| National Tsing Hua University | 7 | 5% |
| National Chiao Tung University Taiwan | 6 | 5% |
| Korea University | 5 | 4% |
| Cheongju University | 5 | 4% |
| National Yunlin University of Science and Technology | 5 | 4% |
| University of Niš | 4 | 3% |
| Korea Institute of Science and Technology Information | 3 | 2% |
| Gainia Intellectual Asset Services, Inc. | 2 | 2% |
| Chung Hua University | 2 | 2% |
| Beijing Institute of Technology | 2 | 2% |
| Total | 41 | 33% |

Note: Articles with one or more affiliations are multi-counted.



* Aristodemou, L. (2020). Identifying Valuable Patents: A Deep Learning Approach. Department of Engineering, Institute for Manufacturing, Innovation and IP Management (IIPM) Laboratory. Cambridge, University fo Cambridge. **PhD.**

# INCREASING FOCUS ON UNSTRUCTURED TEXT DATA




[a]Categoric input feature determinants appear in order: IPC primary section, IPC primary class, CPC primary section, CPC primary subsection, CPC primary group (vertical axis), with perplexity variation (pervar) of [5, 30, 50] (horizontal axis).

[b]The associated numbers represent the following combinations: (1) IPC primary section vs. pervar(5), (2) IPC primary section vs. pervar(30), (3) IPC primary section vs. pervar(50), (4) IPC primary class vs. pervar(5), (5) IPC primary class vs. pervar(30), (6) IPC primary class vs. pervar(50), (7) CPC primary section vs. pervar(5), (8) CPC primary section vs. pervar(30), (9) CPC primary section vs. pervar(50), (10) CPC primary subsection vs. pervar(5), (11) CPC primary subsection vs. pervar(30), (12) CPC primary subsection vs. pervar(50), (13) CPC primary group vs. pervar(5), (14) CPC primary group vs. pervar(30), (15) CPC primary group vs. pervar(50).

Fig. 3.7 t-SNE visualisation for abstract vs. categoric input feature determinants[a,b]

Aristodemou, L. (2020). Identifying Valuable Patents: A Deep Learning Approach. Department of Engineering, Institute for Manufacturing, Innovation and IP Management (IIPM) Laboratory. Cambridge, University fo Cambridge. **PhD.**

## Leveraging the BERT algorithm for Patents with TensorFlow and BigQuery

November 2020, 2020
Rob Srebrovic[1], Jay Yonamine[2]

**BERT for Patents**



```
In [ ]: import collections
        import math
        import random
        import sys
        import time
        from typing import Dict, List, Tuple

        # Use Tensorflow 2.0
        import tensorflow as tf
        import numpy as np
```

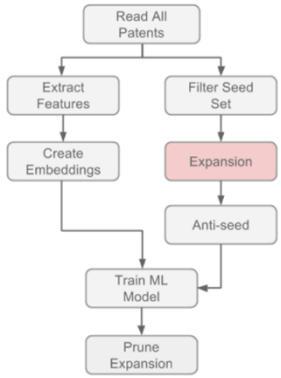
```
In [ ]: # Set BigQuery application credentials
        from google.cloud import bigquery
        import os
        os.environ["GOOGLE_APPLICATION_CREDENTIALS"] = "path/to/file.json"

        project_id = "your_bq_project_id"
        bq_client = bigquery.Client(project=project_id)
```

## Automated Patent Landscaping

Aaron Abood
Google, Inc.
aabood@google.com

Dave Feltenberger
Google, Inc.
feltenberger@google.com

### 3. AUTOMATED LANDSCAPING



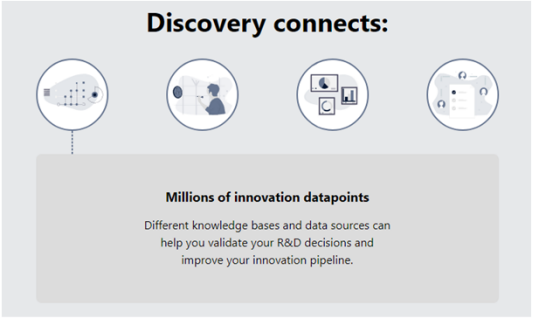
# DERIVING VALUE FROM COMBINING AN INCREASING NUMBER OF DATA SOURCES?


**Problem theme B: Database interconnectedness**

| WHEN | Short term (+5 years) | Medium term (+10 years) | Long term (+20 years) |
|---|---|---|---|
| Problem definition | • Link patents to owners<br>• Link patents to entities' data<br>• Entity disambiguation<br>• Apply classifications to non-patent literature | • Entities, acquisitions, licenses – link competitors to patents<br>• Link to non-patent literature<br>• Link patents to strategic aims and what is filed to support that | • Vision: link products to patents<br>• Cross-referencing of classifications across all spaces |
| Required research/ technology development | • Legislation – requirements to declare ownership<br>• Standardisation of patent numbers and ownership details<br>• Further citations and similarity indicators – requirements on applicants<br>• T23 – Artificial neural network analysis<br>• T17 – Citation analysis<br>• T2 – Semantic-analysis-based approaches<br>• T1 (NLP-based approaches) + T2 (semantic-analysis-based approaches) + T5 (neural-network-based approaches) combined to obtain more meaningful insights<br>• T2.4 – Domain ontologies and more comprehensive technology taxonomy | • Link owner ID to other numbers, e.g. company numbers<br>• T17.11 – NLP citations – encourage – standardise<br>M&A data, corporate trees<br>• Standardise owner names via similarity searching algorithms | |
| Milestones | • Related NPL fields in patent searches<br>• Standardisation actually happening (between intellectual property offices) | • Shift from an assignee to a legal entity<br>• Link to recognised legal entity from patent databases | • Products – list of parts – patents<br>• Give a technical specification of a product – get patents from it |
| Enablers | • Public funding<br>• Legislation/regulation | | |
| Risks | | | |
| High | • Lack of funding | | |
| Medium | • Lack of coordination | | |
| Low | • Slow return of investment | | |

*Figure 5 DQ.B Mini-technology roadmap*

- Technology intelligence
- Business intelligence
- Market intelligence



# EMERGING NEW USE CASES FOR PATENT DATA


LexisNexis

LOGIN LET'S TALK

PatentSight Press Center

Merger deals increasingly scrutinized using patent analysis

by LexisNexis PatentSight team, on Feb 19, 2018 11:41:00 AM

**Bonn, Germany, February 19th, 2018.** Patent metrics have increasingly been used to assess the innovative competitiveness of technology-oriented firms in recent years. The European Commission (EC) was no different when evaluating the Dow/DuPont merger. Interestingly

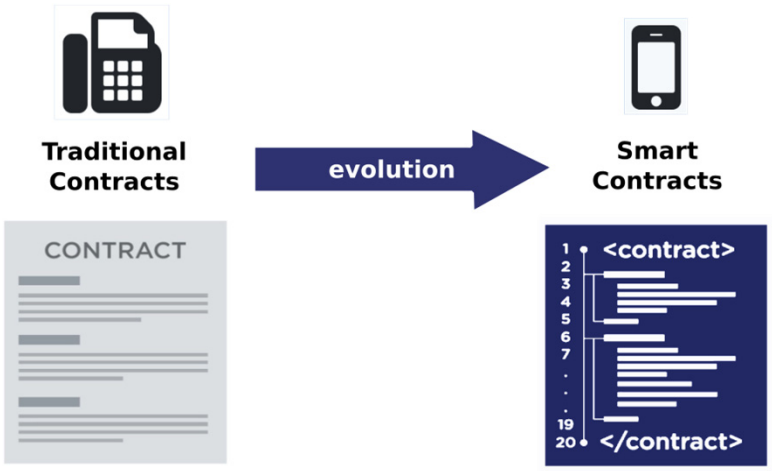

Iprova

**Invent disruptively.**
**Invent first.**

Iprova has developed ML and NLP based technologies which augment and enhance the human ability to invent. This enables the generation of commercially relevant, breakthrough inventions in real-time in direct response to day-day changes around the world.

# FORWARD INTEGRATION FROM PURE DATA ANALYTICS INTO THE ACTIVE PATENT LIFE CYCLE, E.G. LICENSING


Smart contracts for automatize licensing payments



# WHAT ARE THE IMPLICATIONS FROM AN EVENTUALLY DECENTRALIZED PATENT SYSTEM?

IBM

IBM News Room

IPwe and IBM Seek to Transform Corporate Patents With Next Generation NFTs Using IBM Blockchain

An ecosystem to tokenize patents can allow intellectual property to be treated as business assets, making patents easier to license, sell and commercialize

PARIS and ARMONK, N.Y., April 20, 2021 /PRNewswire/ --

Government Information Quarterly 38 (2021) 101559

Contents lists available at ScienceDirect

Government Information Quarterly

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

## Decentralising the patent system

Gaétan de Rassenfosse [a], Kyle Higham [b,*,1]

[a] École polytechnique fédérale de Lausanne, Switzerland
[b] Institute of Innovation Research, Hitotsubashi University, Japan

ABSTRACT

Modern patent systems are slow, inefficient, expensive, and may result in outcomes that actively harm technological progress. This paper proposes a substantive re-think of these systems and lays a foundation upon which practical solutions can be built. Many solutions proposed in the past, such as prior-art bounties, outsourced examination, and dynamic fee setting, have gone unheeded due to the cost of administering them and the rigidity of the patent system. We explore how distributed ledger technologies (DLTs) enable these major changes by altering the way stakeholders are able to interact with the patent records system. We find that transitioning to a DLT-based patent records system can enable many previously suggested improvements to current patent systems in a flexible, scalable, and transparent manner. The case for such a transition is strengthened when jointly considering the complex but common roots of problems facing modern patent systems, rather than a balkanized set of technical solutions to address each issue independently. Noting that a DLT-based system is not a panacea, we also provide comment on the political, legal, and organisational challenges that must be overcome for such changes to be implemented at scale.

### 1. Introduction

The core rationale behind the patent system has not changed since 15th century when the first well-documented patent system was concocted in Venice (Long, 1991; Prager, 1944; Sichelman & O'Connor, 2012). By granting inventors the right to exclude others from a newly discovered slice of technological space, the patent system provides an incentive to pour more resources into the inventive process. Furthermore, the disclosure of the technical details of the invention in the patent document is supposed to facilitate knowledge diffusion. However, it has been argued that current patent systems are not a particularly efficient means of achieving these goals (Bessen & Meurer, 2008; Jaffe & Lerner, 2004; Stiglitz, 2007; Thurow, 1997), and additionally come with such high costs to society that many have wondered why governments should grant these rights at all (Boldrin & Levine, 2008, 2013; Mirowski, 2011).

The shortcomings of modern patent systems are well-documented. We propose a reconfiguration of patent record systems that facilitates solutions to these problems through decentralisation of patent-related processes and information, in conjunction with increased transparency, flexibility, and efficiency. In particular, we ask whether distributed ledger technologies (DLTs) constitute an appropriate technical foundation for such a change, and consider how they may be practically implemented. The potential for DLTs to play this role is illustrated through a series of proposals. These proposals, many of which are already discussed in the literature, are aimed at alleviating well-known problems associated with patent systems, including long grant lags (Ackerman, 2011; Gans, Hsu, & Stern, 2008), high levels of patent invalidity (Frakes & Wasserman, 2017; Lemley & Shapiro, 2005), anti-competitive patenting strategies (Blind, Cremers, & Mueller, 2009; Macdonald, 2004), the high risk and cost of litigation (Bessen, Ford, & Meurer, 2011; Bessen & Meurer, 2013; Lemley, Richardson, & Oliver, 2018), and many other problems that have been covered in recent years (Eckert & Langinier, 2014; Hall & Harhoff, 2012; Sampat, 2018). We explore many of these issues throughout this work, and explain how a modernized patent system can address them.

It has become clear in recent years that it is not merely inefficiencies within patent offices that lie at the root of these problems, but the political and legal systems to which theses offices are duty-bound. It is easy to forget that patent offices are resource-constrained actors with little control over the legal and industrial environments in which they operate. Financial constraints and extended patent pendency times, the decisions around patent-eligible subject matter, or even differences in novelty and non-obviousness thresholds between the patent office and the courts, are crucial issues with current patent systems that are often entirely out of the control of patent offices (Jaffe & Lerner, 2004; Rai, 2008; Sohn, 2011).[2,3]

* Corresponding author.
*E-mail addresses:* gaetan.derassenfosse@epfl.ch (G. de Rassenfosse), higham@ir.hit-u.ac.jp (K. Higham).
[1] Most of this research was conducted while this author was working at École polytechnique fédérale de Lausanne.
[2] Of course, many of the benefits and drawbacks of current patent systems lie outside of these bureaucracies, and any proposed changes to patent systems must also consider the flow-on effects that these policies may have.
[3] In the United States, however, there have been recent developments towards increasing patent office autonomy (Wasserman, 2012).

https://doi.org/10.1016/j.giq.2020.101559
Received 6 July 2020; Received in revised form 18 December 2020; Accepted 20 December 2020
Available online 21 January 2021


Nokia Launches Blockchain-Powered Marketplace for Data Trading

The marketplace is aimed at providing enterprises and communication service providers access to trusted datasets.


Spotify acquires blockchain startup Mediachain to solve music's attribution problem

Sarah Perez @sarahintampa / 4:25 PM GMT+1 • April 26, 2017

WIPO

WIPO Work on Blockchain

Edward Elliott
Standards Section

Potential use cases in IP

- Patent Management
- Certification Trademarks
- Supply Chain Tracking
- Trade Secret Management
- Evidence of Creation
- Provenance Authentication
- Smart IP Rights and Registries
- IP Enforcement
- Smart Contracts
- Evidence of Use
- Anti-counterfeiting
- Digital Rights Management
- Traceability

BERNSTEIN

Own what you make

Secure your intellectual property in one click

EY Building a better working world

Search for insights, services and

Press release

21 Jun 2018

EY and Microsoft launch blockchain solution for content rights and royalties management for media and entertainment industry

# WHERE ARE YOU HEADING?

Thank you for your attention!

To discuss further please contact me at:

frank.tietze@eng.cam.ac.uk

Follow the IIPM research group at:

@IIPMcambridge