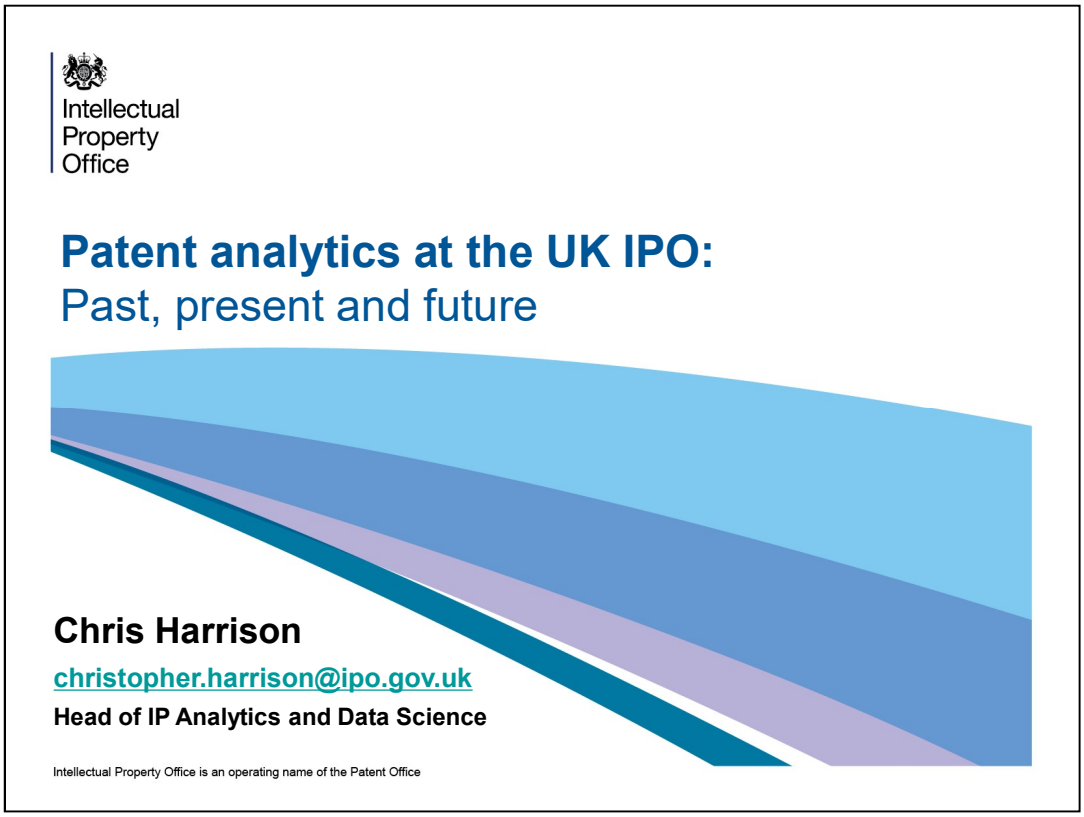


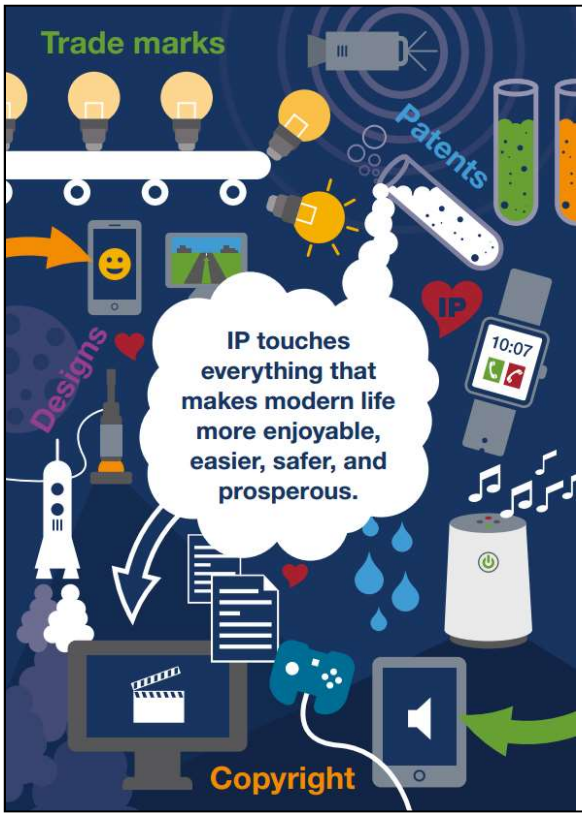
Head of IP Analytics and Data Science in the Informatics team at the IPO.

Useful to share how government uses patent information

Intellectual Property Office

Moved to Newport in South Wales in 1991 (small branch in London maintained for the benefit of the large professional IP community based there and for comms with central government)

In 2007 we changed our name to the IPO to more accurate reflect the work that we do

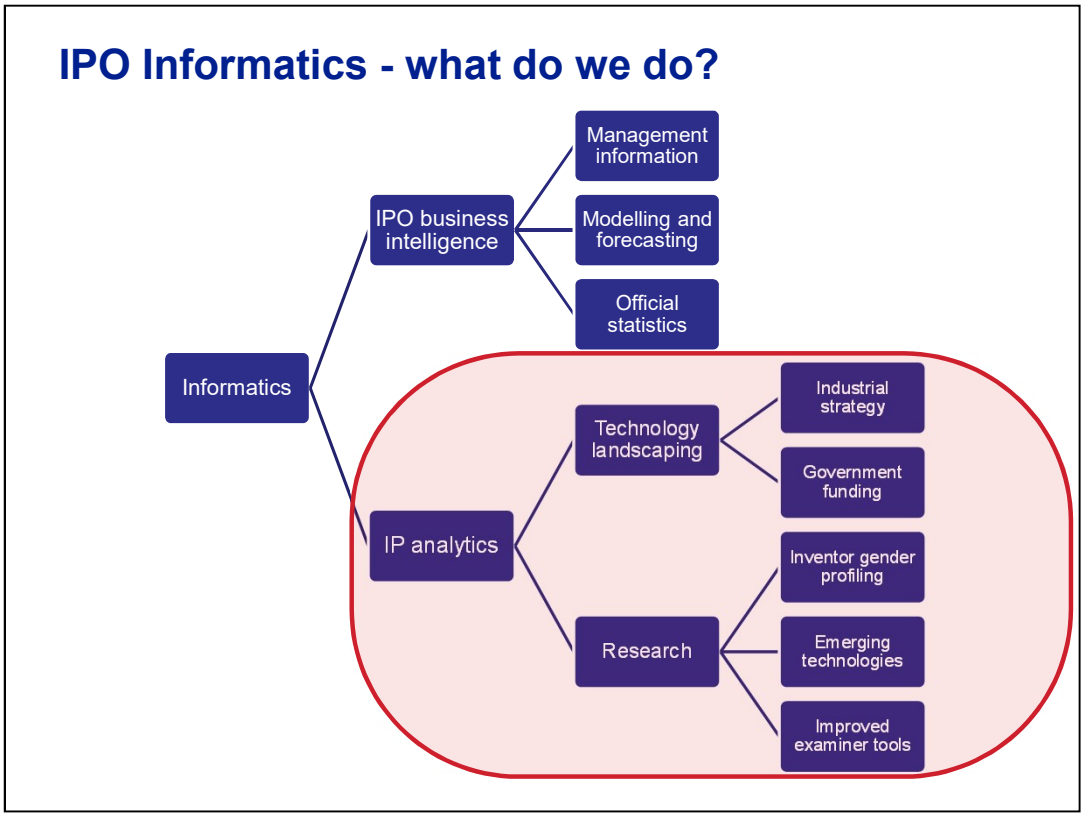


OUR ROLE

MAKING LIFE BETTER THROUGH IP

OUR AMBITION

TO BE THE BEST IP OFFICE



Informatics team at the IPO has two primary functions

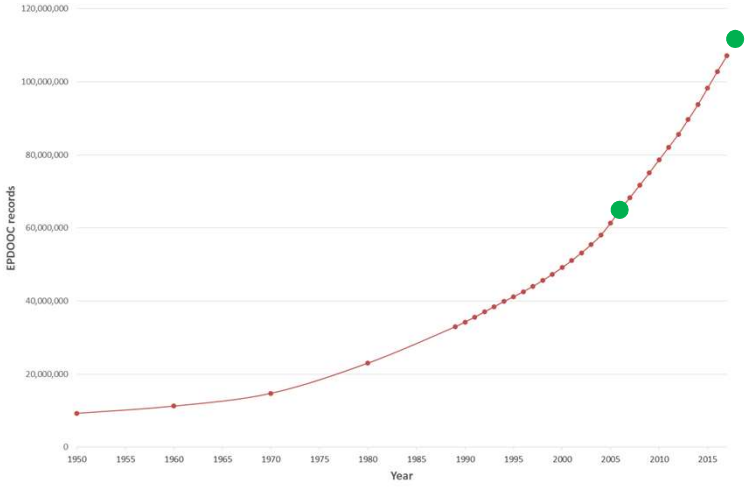
1) Business intelligence across all IP rights (patents, trade marks, and registered designs)

2) IP analytics

[click] This talk will focus on past, present and future developments in the patent analytics and research side of things.

Patent analytics

2006 Sainsbury Review of science and innovation policies gave IPO mandate for offering patent informatics “to inform planning and policy”, “to enable full access to patent resources” and to “aid Government strategy support”



Intellectual Property Office

Department for Business, Energy & Industrial Strategy

Government Office for Science

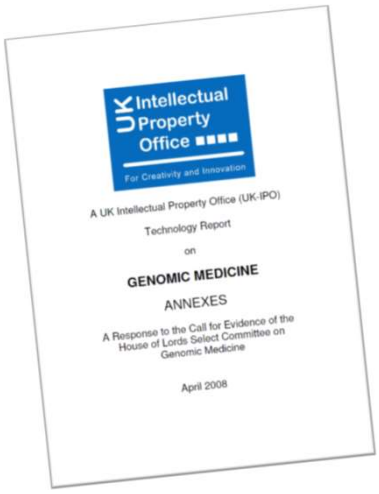
UK Research and Innovation

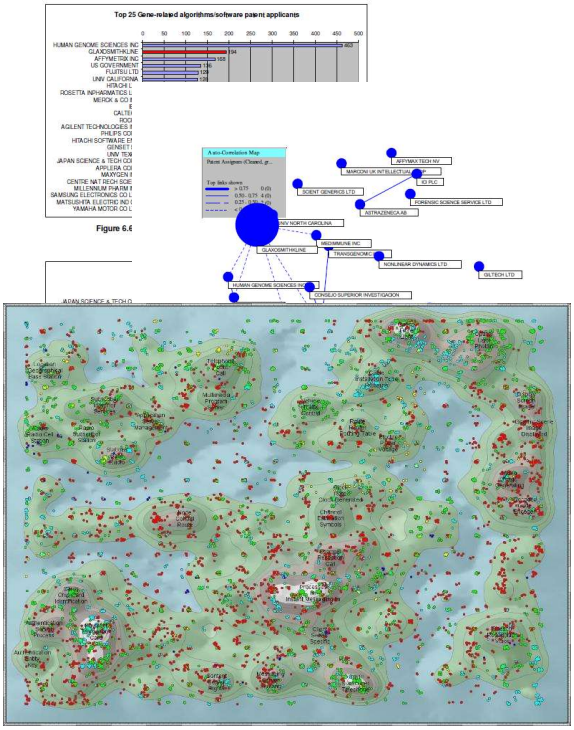
There are over 100,000,000 published patents worldwide. There is no doubt that there is a wealth of technical information contained within these documents.

Patent data is well suited for big data tools and techniques because of the size, speed and range of changes to the patent databases.

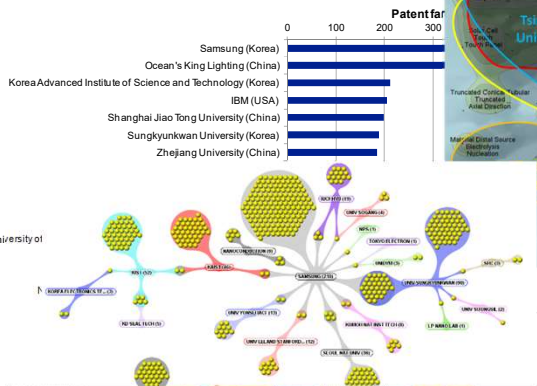
We have a range of stakeholders across Government, working closely with our parent department BEIS, as well as the Research Councils and Innovate UK who have an interest in the patent landscape when choosing which areas and companies to give research funding and grants to.

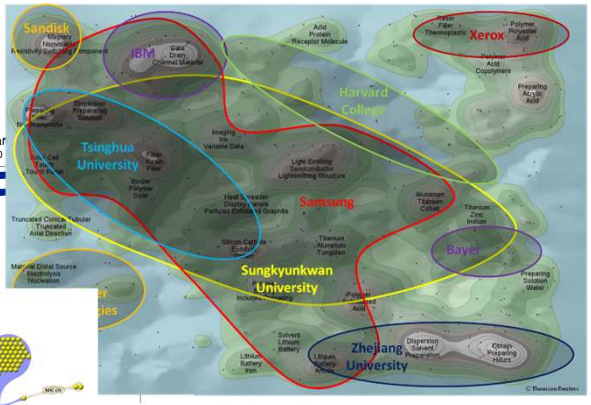
2008:
The early days…

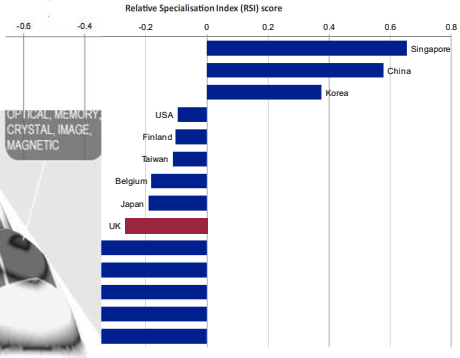




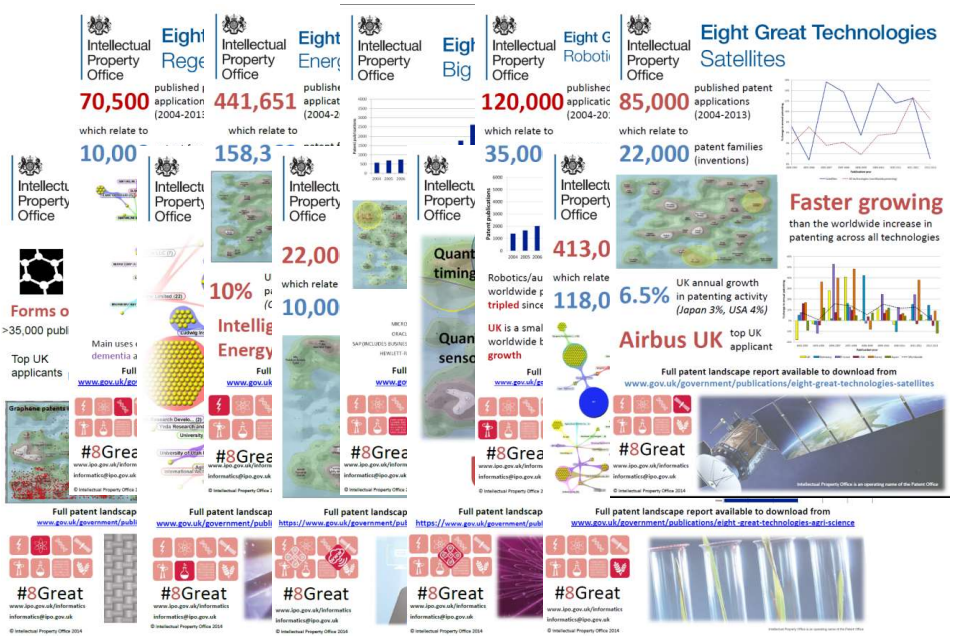
2012:
The graphene hype…



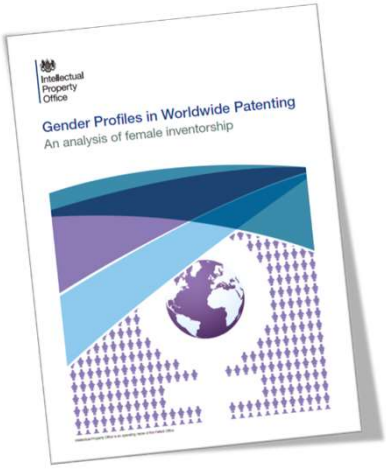


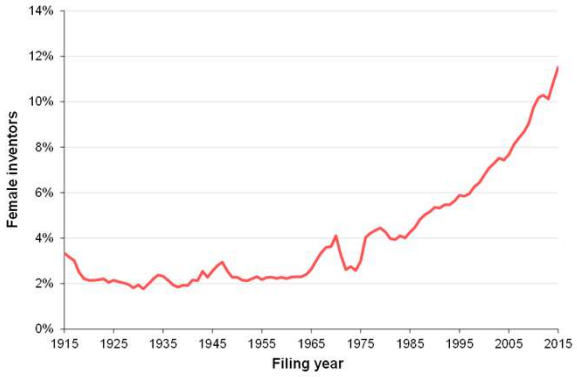


2014: Eight Great Technologies…



2016:
Wider research…





WIPO technology field	Male inventors	Female inventors
Chemistry: Biotechnology	80.2%	19.8%
Instruments: Analysis of biological materials	81.8%	18.2%
Chemistry: Pharmaceuticals	83.6%	16.4%
Chemistry: Food chemistry	83.7%	16.3%
Chemistry: Organic fine chemistry	87.1%	12.9%
…		
Mechanical engineering: Machine tools	96.7%	3.3%
Mechanical engineering: Transport	97.2%	2.8%
Electrical engineering: Basic communication processes	97.2%	2.8%
Mechanical engineering: Mechanical elements	97.3%	2.7%
Mechanical engineering: Engines, pumps, turbines	97.5%	2.5%

Historical profiling

60% increase in female inventors since 2000 (7.1% in 2001 to 11.5% in 2015)

2018: The Industrial Strategy…

Intellectual Property Office

Patent overview
Cyber security

Patent data can give a valuable insight into innovative activity, to the extent that it has been codified in patent applications. The dataset used for this top-level analysis was extracted from worldwide patent databases following detailed discussion and consultation with patent examiners from the Intellectual Property Office who are experts in the field and who, on a day-to-day basis, search, examine and grant patent applications relating to cyber security technologies. The dataset used for analysis includes a wide range of different cyber security technologies, including systems and methods protecting against viruses, malware, hacking, phishing and distributed denial of service (DDoS) attacks, amongst others.

This report is based on analysis of published patent application data and not granted patent data. Data for published patent applications gives more information about technological activity than the figures for granted patents because a number of factors determine whether an application ever proceeds to grant. These include the inherent lag in patent processing at national IP offices worldwide and the patenting strategies of applicants who may file more applications than they ever intend to pursue.

Figure 1: Text cluster analysis of cyber security patents from UK-based applicants (2008-2017)

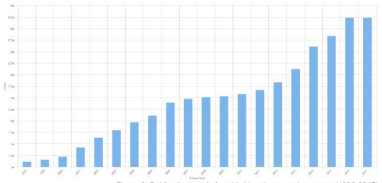
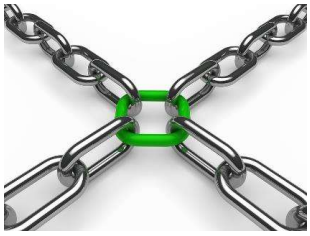


Figure 2: Publication trend of worldwide cyber security patents (1998-2017)

Figure 3: Top 50 applicants worldwide for cyber security patents (2008-2017)

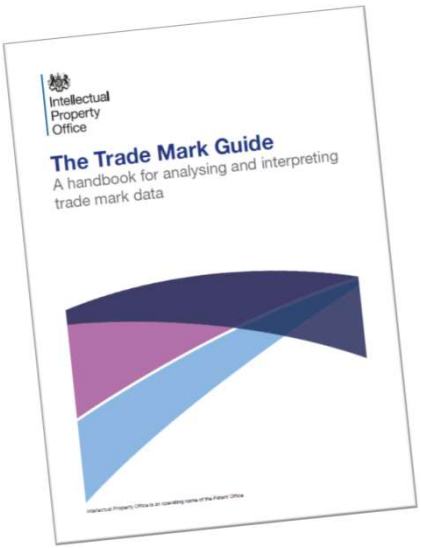

PATSTAT
EPO's Worldwide Patent Statistical Database
Clarivate
Analytics
vantage point
Search Technology
LexisNexis PatentStrategies™
PATENT
INSPIRATION
Powered by ΔULIVE
Patent iNSIGHT Pro
Transform Patents to intelligence
INNOGRAPHY®
LexisNexis™
TotalPatent
Questel
IP Business Intelligence
patsnap
PatBase

12 years and counting…



- Visualisations have improved, BUT… Power of simple charts
- Data cleaning
- Data linking
- Education

February 2015



November 2018

- Patent data
- Patent database interconnectedness
- Patent data analysis
- Patent information visualisation
- Patent quality

Patent data – data mgmt., data prep, data cleaning, data quality. No common standards for patent numbering; unharmonised def'n of patent families. Data quality is important – data is valuable beyond its orig purpose; urgent need for more structured and cleaner standardised data.

Patent database interconnectedness – plenty of issues to tackle worldwide when linking IP data to financial data, litigation data, market data etc.

Patent data analysis – understanding and deciding what type of analysis is more suitable for a certain dataset and why, and how to measure their effectiveness. Build a corporate memory of past analysis for future users to utilise, saving time and resources.

Patent information visualisation – what's available, how they need to be improved, their effectiveness for different decisions

Patent quality – how to determine patent quality and identify invalid patents by preventing granting de facto invalid patents.

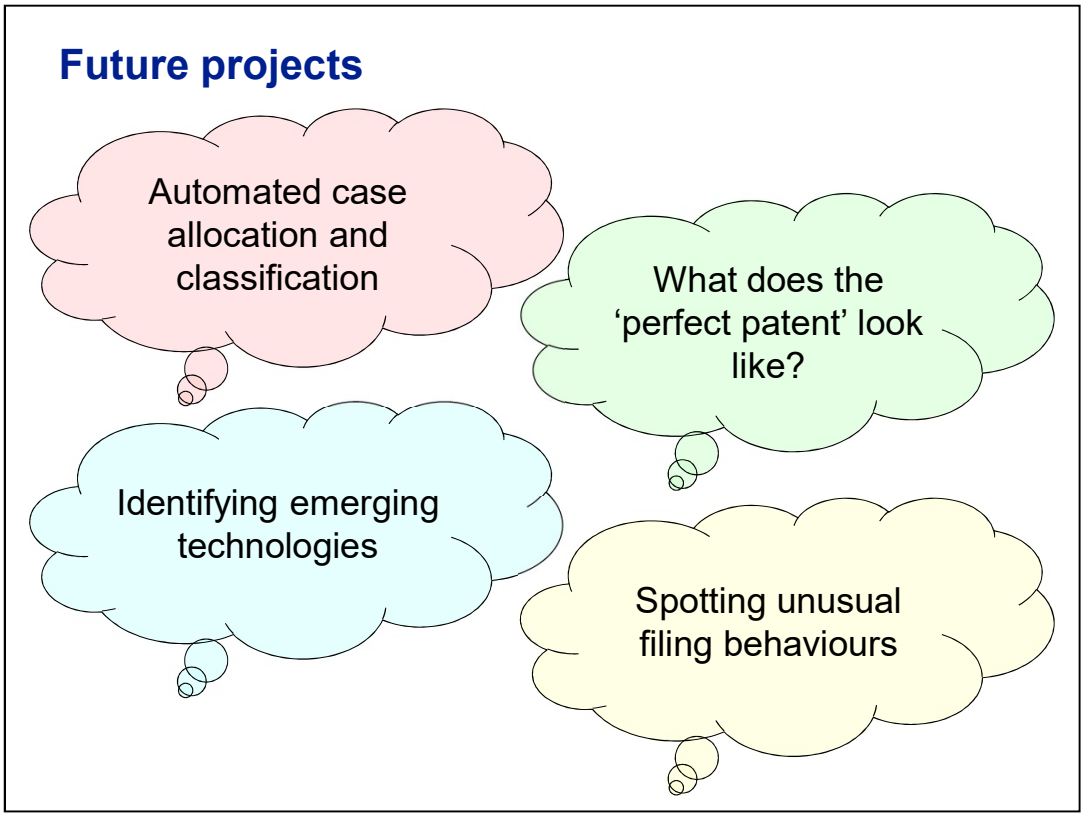
Modern patent analytics is evolving.

Patents have a highly formal syntax that is not well designed to optimally share knowledge with the public or with other scientists. This very structure, however, does mean that patents are uniquely positioned to benefit from machine learning and AI techniques to extract and present meaning in useful ways. Deep learning techniques on the contents of a document can be applied as a first pass to read the legalese of a patent disclosure and assign the technology areas mentioned in the invention. Then train machines to better decipher legalese laden documents and read patent disclosures.

At the IPO we are developing our data science capabilities, made a little easier by having some of the world experts in ONS' data science campus next door, to allow us to move with the times and we hope to remain one of the leading national offices for patent analytics.

This means going wider and deeper than just front page bibliographic data of patents.

Full-text analysis = unstructured data but potential for deeper understanding of the underlying technology.



Just a few other types of work we would like to undertake in the coming year

As we continue to increase our data science capability.

Use a variety of data science techniques.

Although the techniques to do these are quite complimentary.

Emerging technologies

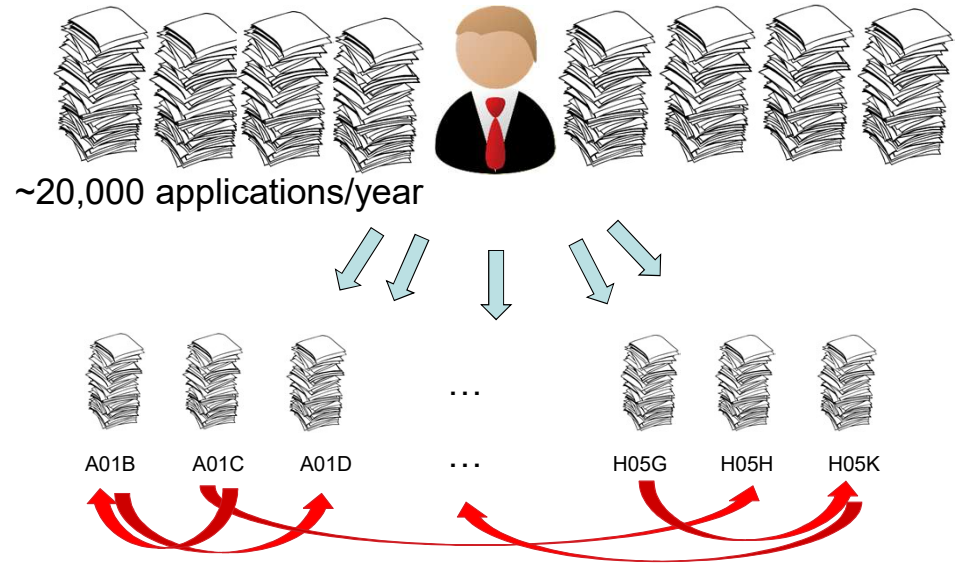
Can the patenting behaviour surrounding historically ground breaking products be used to identify ground breaking, disruptive or emerging technologies?

Can we use data science techniques to analyse patent text and identify technology groups?

The Informatics team supported government in their emerging technology work.

Important area of analysis to government to determine the place to focus initiatives and investment.

Auto-allocation: the business problem



- Need to allocate to a subclass – technical areas
- ~400 examiners, organised by CPC subclass
- Classification scheme has more detail, but not relevant here
- Examiners may reallocate if needed
- Estimate 2800 hours/year allocator time (blue arrows), and 1100 hours/year examiner time (red arrows)
- Goal is not necessarily to eliminate allocators! Saving examiner time is also a benefit.
- Supervised machine learning
- Multi-class, multi-label problem