EXPLORING **THE FUTURE OF PATENT ANALYTICS**

UNIVERSITY OF CAMBRIDGE
Department of Engineering

IfM MANAGEMENT TECHNOLOGY POLICY

Acknowledgements

This research was funded by the United Kingdom Engineering Physical Science Research Council (EPSRC), through the Cambridge Big Data initiative as part of an EPSRC Institutional Sponsorship Grant 2016 – Small Partnership Awards, supported by Aistemos Ltd as the industrial partner. The project ran from late 2016 until spring 2017 with the aim of exploring the future of patent analytics using a technology roadmapping approach. We would like to thank Nigel Swycher (Aistemos) and Steve Harris (Aistemos) for supporting this research. Furthermore, we thank Susan Bates (Shell), Nigel Clarke (EPO), Richard Corken (UKIPO), Steve Harris (Aistemos) and Irene Kitsara (WIPO) for reviewing earlier versions of this report. In addition, we would also like to thank Nicky Athanassopoulou (IfM ECS) and Michèle Routley (IfM ECS) for running the roadmapping workshop. Finally, we thank all the experts and participants who have contributed to the different stages of this research. As with all research, this report is also subject to limitations given the resource constraints. We nevertheless hope the report will contribute to coordinating the research efforts of the patent analytics community.

Exploring the Future of Patent Analytics

By Leonidas Aristodemou and Dr Frank Tietze
Centre for Technology Management, Institute for Manufacturing

First published in Great Britain 2017 by the University of Cambridge Institute for Manufacturing,
The Alan Reece Building, 17 Charles Babbage Road, Cambridge CB3 0FS, UK

ISBN: 978-1-902546-84-1

FOREWORD

This report is the culmination of many months of work involving surveys, interviews, an international workshop and hours of roadmapping. I took part in the survey and the workshop and I was delighted to be invited to write this foreword from my own personal perspective. I thank the authors and the researchers for these opportunities.

Some years ago I was involved in a scenario-building process, which was intended to scope out plausible relevant and challenging options for possible future IP worlds. The time horizon stretched to 2025 and the four main drivers for IP regimes were identified at the time as:

- the demands of business and commerce;
- the conscience and changing mores of society;
- the ever-increasing complexity of technology; and
- the impact of geopolitics.

At the time 2025 seemed far enough away, but in 2017, barely a decade after these scenarios were created, they have come together before their time, in a perfect storm, to dispel any complacency (if it ever existed) about the IP constellation today.

Indeed, the future is already here; it is just not evenly distributed.

Today we see increases in patent applications across the board in terms of numbers and diversity, as businesses seek to protect their inventions. There is increasing concern about the impact of technology on the environment, and on society. The Internet of Things and Industry 4.0 are transforming the nature of, and drastically increasing the complexity and sophistication of the subject matter forming the basis of, patent applications. Emerging economies are changing the distribution of inventive activity, increasing the contribution of third countries compared with the established industrialised countries. National economic growth and national innovation support policies skew the distribution even more acutely. The volume, nature and characteristics of patent data are changing in response.

On the other hand, modern technologies such as artificial intelligence, machine learning, deep learning, cloud computing, big data/linked data, text mining/data mining, image searching and visualisation, the very core technologies of I4.0, need to be harnessed to make sense of the patent world of today. That is to say, patent search, analysis, informatics and visualisation have to embrace these technologies in order to establish, retain and deepen relevance and legitimacy.

The present work initially distilled the major challenges facing the patent analytics community today, and into the future as (I paraphrase):

- source data;
- linked data;
- data analysis;
- data visualisation; and
- patent quality.

These challenges were presented online for comment by survey respondents prior to a workshop held in Cambridge in March 2017. To set the scene for the discussion, each challenge was briefly presented at the beginning of the workshop by its own champion. The participants then developed and elaborated the discussion, and break-out syndicates addressed possible impacts and solutions. After the workshop finished and the day's flipchart and post-it work were over, the hardest task began, namely, to consolidate the input from the survey interviews and workshop, and to define the key actions/next steps. A draft report was subsequently issued and sent to the workshop participants for review. Their feedback was collected and analysed, and at that stage trends could at last be distilled from the roadmap.

This necessary and timely study has identified a sort of "needs pyramid", with the above 5 challenges at the apex, requiring primary and complementary measures in the intermediate strata, but notably and commendably emphasising 21 supporting enablers at the base. These enablers comprise four themes: technology developments, legislation and standardisation, cooperation between industry and academia, and continued professional development.

I am particularly pleased at the inclusion of the latter, since the inevitable disruption brought by I4.0 on patentable subject matter, and the technologies to be brought to bear on patent search and analysis, will have a significant human impact. Tasks, jobs, occupations and professions will be affected, and this is not to taken lightly.

There is clearly untapped potential in the application of patent analytics. And there is certainly much more to be done in order to make the results and conclusions of patent analysis intelligible and actionable to the expert and non-expert alike.

This study has provided a snapshot of the current status, but it is clear that further research is necessary. Such research will need to focus on information retrieval: AI, machine learning; on new data structures: big data, linked data; on communication and storage: the cloud and nG; and on reporting protocols and visualisation. This is likely to be an enduring, but absolutely necessary, activity. It is not inconceivable that the establishment of a dedicated "observatory" to monitor progress and predict future directions will be required.

The present report will do much to focus the community's attention on the current status and future directions of patent analytics and its applications. It is to be hoped that this report is the first of many.

Dr Nigel S. Clarke
Head of Patent Information Research, European Patent Office, Vienna.
Editorial Board Member "World Patent Information".

Vienna, August 2017

This Foreword represents his personal opinions and not necessarily those of the EPO.

EXECUTIVE SUMMARY

In a connected world, where successful technological development increasingly depends on collaboration between different partners, effectively utilising patent data analytics has significant, yet untapped, potential. Given the right analytics solutions, this high-quality data can be used for decision-making at a strategic level in a variety of organisation types.

This project and report contribute to expanding the field of patent analytics for more effective exploitation of the largest worldwide repository of technological information. This has been done by developing a domain-level technology roadmap following a three-stage technology roadmapping and problem-solving approach. First, from desk research and expert discussions, we identified five main problem themes in the patent analytics field (patent data, database interconnectedness, data analysis, data visualisation and patent quality). Second, we verified and expanded these problem themes through an online survey of 70 respondents. Third, we explored the future direction of the field through a workshop, with input from the preparatory stages above, with 28 leading experts.

The technology roadmapping approach has served to develop a technology roadmap to facilitate collaboration and coordinated action within the patent analytics community. We have identified 11 priority technologies, such as artificial intelligence and neural networks, 5 additional technologies, such as technologies for linking databases, and 15 complementary technologies, such as block chain, to be adopted in the field and which are important in terms of overcoming the problems. We have also identified 21 enablers for potential breakthrough progress in the field that cluster around 4 themes: technology development cycles and methodologies; legislation and standardisation for patent data quality; continuous professional development; and cooperation between industry and academia. Key next actions include the generation of use cases for different users, the standardisation and harmonisation of patent ontologies and the implementation of reporting standards.

TABLE OF CONTENTS

INTRODUCTION

BACKGROUND

In a connected world, where successful technological development increasingly depends on the collaboration of different partners (Tietze & Lauritzen 2016; West & Bogers 2014; Randhawa et al. 2016), effectively utilising patent data has significant, yet only partially exploited potential (Lee et al. 2011; OECD 2016). Given the right analytics solutions, patent data, in particular, can be used for regular decision-making at a strategic level in all kinds of organisation, small or large, private or public, young or old. Patent data has long been considered the world's largest repository of technological information (WIPO 2016; WIPO 2015). Only with the digitisation of patent data since the BACON project in 1984 (Dintzner & Van Thieleny 1991) and gradual improvements in analytics over the last decades has patent data become increasingly accessible to a non-specialist audience. While the quality of patent data has increased enormously over the last two decades or so, and increasingly better software tools for analysing the data are continuously being developed, still today a significant potential for utilising patent data remains untapped (Lupu et al. 2011; Tietze & Probert 2015; Lupu 2017).

Patent analytics describes the science of analysing large amounts of patent information to discover relationships and trends (Trippe 2003; Trippe 2015). Abbas et al. (2014) provide an overview of a set of tools and approaches, with their key features and weaknesses, for the analysis of patent documents for the purpose of forecasting future technological trends, conducting strategic technology planning and identifying technological hotspots and patent vacuums. Moehrle et al. (2010) apply a business process model to the patent analytics process, which maps the main tasks in patent analytics to the available tools and techniques. In this report, we adopt the definition of patent analytics as the science of analysing large amounts of patent information to derive meaningful insights to support strategic decision-making, which constitutes the deployment of different technologies, techniques and approaches.

The recent advancements of data technologies (OECD 2016), such as machine learning, deep learning and artificial intelligence, potentially seem to deliver breakthrough progress to enable completely new use cases for patent data with substantial economic benefits (Lupu 2017). While these technologies already impact several areas, their impact on patent analytics is yet to be understood. These technologies, which are either well established in other fields or emerging, have been used in a limited way to explore and exploit the patent data repository. At the same time, in patent analytics, there exists a large number of problems that remain unsolved today (Lupu et al. 2011; Raturi et al. 2010; Trippe 2003).

Involving numerous key stakeholders, such as technical experts, lead-users of patent analytics solutions, patent specialists and decision-makers, the report presents the results from a foresight study aimed at developing a technology roadmap (Phaal 2015) for the future of patent analytics (similar to Ferrari et al. 2014). The technology roadmap contributes to identifying these breakthroughs and enhancing further academic and industrial development of the patent analytics field for more effective exploitation of the largest worldwide repository of technological information. We hope that the roadmap will contribute to coordinating further activities in the field of patent analytics and helping the industry to explore, understand and overcome a number of unresolved problems with the potential to expand the boundaries of the field.

The report presents the outcomes of the project, with the following objectives:

- Help improve patent analytics for more effective exploitation of the largest worldwide repository of technological information in order to enable new use cases supporting better strategic decision-making and partnerships of R&D pursuing organisations.
- Develop a public roadmap to facilitate collaboration and coordinated action of actors in the patent analytics community to further develop the capabilities for analysing patent data.
- Bring together the relevant stakeholders in a research setting to enable new collaborations, thereby enhancing the progress of patent analytics.

The report is designed to help the following audiences: the patent analytics community, patent information specialists, patent offices, governments, industrial users and industrial patent database providers. It is structured in the following order: firstly we present the problem themes that arose from the technology roadmapping and problem-solving approach we followed (see Appendix I for Methodology). This is followed by the findings from the exercise and finally the conclusion.

PATENT ANALYTICS PROBLEM THEMES AND DISCUSSION QUESTION (DQ) FORMULATION

Over the last decade, there has been a large push to improve areas of the patent analytic field and expand its capabilities (Baudour & van de Kuilen 2015; Bonino et al. 2010). However, even today there is a very large number of problematic areas (Lupu et al. 2011). Overcoming these issues should enable us to improve and expand the boundaries of the patent field. The problem themes were identified through desk research and expert discussions in stage 1, verified through the survey in stage 2, and formulated into the discussion questions used in stage 3 (Appendix I).

Problem theme A – Patent data

This is concentrated around the patent data itself. It tackles issues arising during the pre-processing stage of patent analytics (Bonino et al. 2010; Moehrle et al. 2010) in relation to data management, data preparation, data cleaning and data quality. First, a sub-theme emerged with the existence of several unharmonised patent family definitions (Martinez 2010; Martínez 2011; Oldham & Kitsara 2016). Second, there are no common standards for data preparation or a current best approach. In addition, the data is often inconsistent and inaccurate (Baudour & van de Kuilen 2015), and there is no global standard for patent numbering across different patent offices. Furthermore, patent taxonomies are in need of improvement and ontologies are largely absent.

Discussion question A is therefore formulated as:
How can the quality of patent data be substantially improved?

Problem theme B – Patent database interconnectedness

This theme focuses on database interconnectedness, and tackles the issue whereby different types of data, such as intellectual property data, financial data, litigation data and market data, can be combined for more comprehensive analysis. Currently, patent data is linked primarily to legal data, mainly litigation.

Discussion question B is therefore formulated as:
How can the interconnectedness of patent databases with other data sources be enabled?

Problem theme C – Patent data analysis

This theme concentrates on data analysis effectiveness (Brügmann et al. 2015; Gassmann et al. 2012; Lupu et al. 2011), and tackles the problem of understanding and deciding what type of analysis is more suitable for a certain data set, and why. Several sub-themes emerged in relation to this problem, such as the type of analytic techniques available (Abbas et al. 2014; Raturi et al. 2010), how to deploy them, how to measure their effectiveness (OECD 2009; Squicciarini et al. 2013), and which are more suitable for which decisions. In addition, sub-themes include the building of a "corporate memory" of past analysis for future users to begin utilising deep-learning and machine-learning capabilities, saving time and resources, and changing the analytic perspective to a prospective/adaptive framework, to enable a future-oriented approach of patent analytics.

Discussion question C is therefore formulated as:
How can better use be made of the valuable information contained in the patent data?

Problem theme D – Patent information visualisation

This theme focuses on the problem of information visualisation and its effectiveness (Masiakowski & Wang 2013), and tackles issues whereby one needs to decide and understand visualisations arising from patent analysis. Sub-themes concentrate on the types of visualisation available, how these can be improved and their effectiveness for different decisions.

Discussion question D is therefore formulated as:
How can the results from patent analysis be visualised more effectively for better decision-making?

Problem theme E – Patent quality

This theme concentrates on the problem of determining patent quality (Squicciarini et al. 2013; Trappey et al. 2012) and identifying those "invalid" patents that have been granted but should not have been. Sub-themes include the definition of patent quality, how it is measured, how we can make judgements about it and how can we identify existing invalid patents and prevent granting de facto invalid patents.

Discussion question E is therefore formulated as:
How are patent quality and patent invalidity determined?

RESULTS

The technology roadmap that was developed provides a glimpse into the future of patent analytics, identifying key milestones/breakthroughs and enabling factors for fundamental problems in the field. The technology roadmap is designed to contribute to coordinating further activities in the field of patent analytics by helping research and industry to explore potential breakthroughs and by increasing collaboration. The results section is structured as follows: first, we present the results and give an overview of the survey; second, we present the five mini-technology problem roadmaps (DQ A–E); third, we present the technology ranking and classification matrix; and, finally, the overall technology roadmap is presented, with key actions and output arising from the discussions that took place during the workshop. The patent analytics technologies, techniques and tools tree map to which the technology numbers in the roadmaps and matrices refer can be found in Figure 9.

SURVEY OUTCOMES

In the second stage, the verification stage, we used an online survey to verify and prioritise the technologies and problems identified in stage 1 (Appendix I). Seventy respondents participated in the survey, and participants came from a range of sectors within the intellectual property domain (Figure 1). Under the section "other", there are sectors such as patent information specialists, patent analytics consultants, intellectual property offices, governmental organisations, research and development organisations, and more. Moreover, respondents appear to have substantial experience in the field (Figure 2), with 51 per cent of the survey participants having 10+ years of experience in the patent analytics field, and a total of 68 per cent of participants having 5+ years of experience.

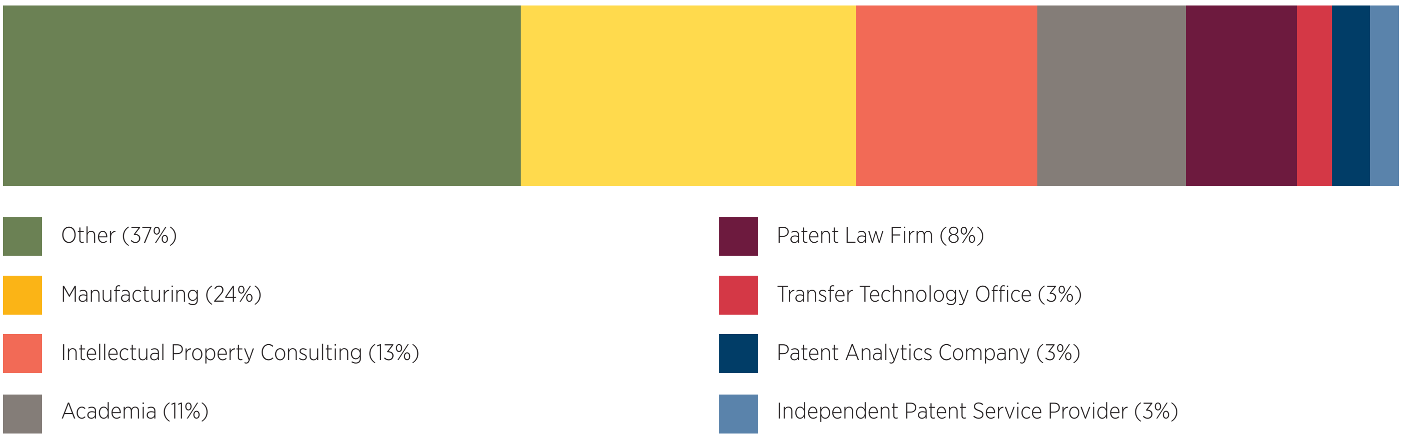


Figure 1 *Sector distribution, percentage of survey participants (n = 70)*

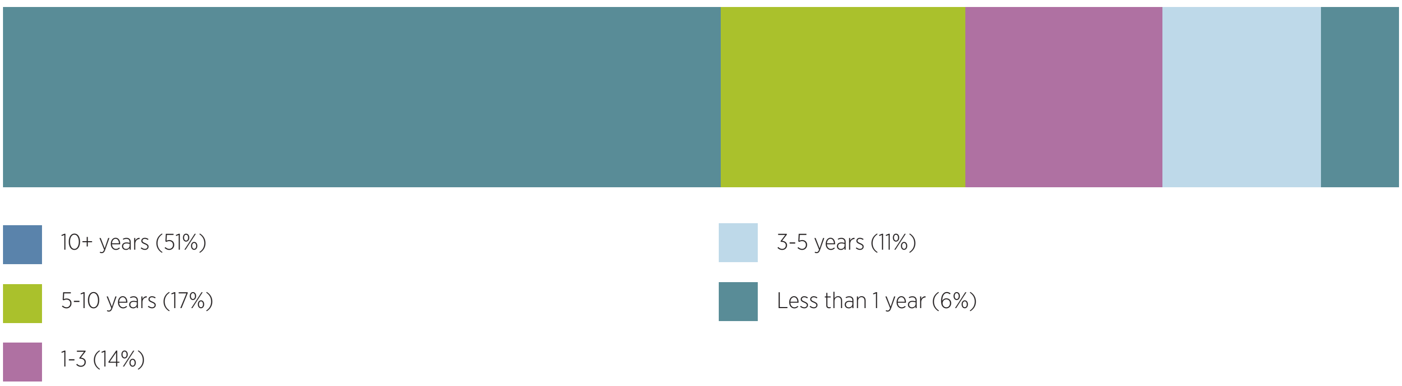


Figure 2 *Experience in the patent analytics field, percentage of participants (n = 70)*

Figure 3 shows the priority ranking for the problems, as selected by the survey participants. The survey responses verified the validity of the identified problem themes (B–E), indicating that the four problems appear to be relatively equally distributed by importance. The results, however, show that the two most important problem themes to be tackled first appear to be: problem theme C, data analytic techniques, and problem theme D, visualisation techniques. In addition, problem theme A, patent data, emerged from the available text box for participants to enter other important problems in the domain, and was subsequently turned into a discussion question.

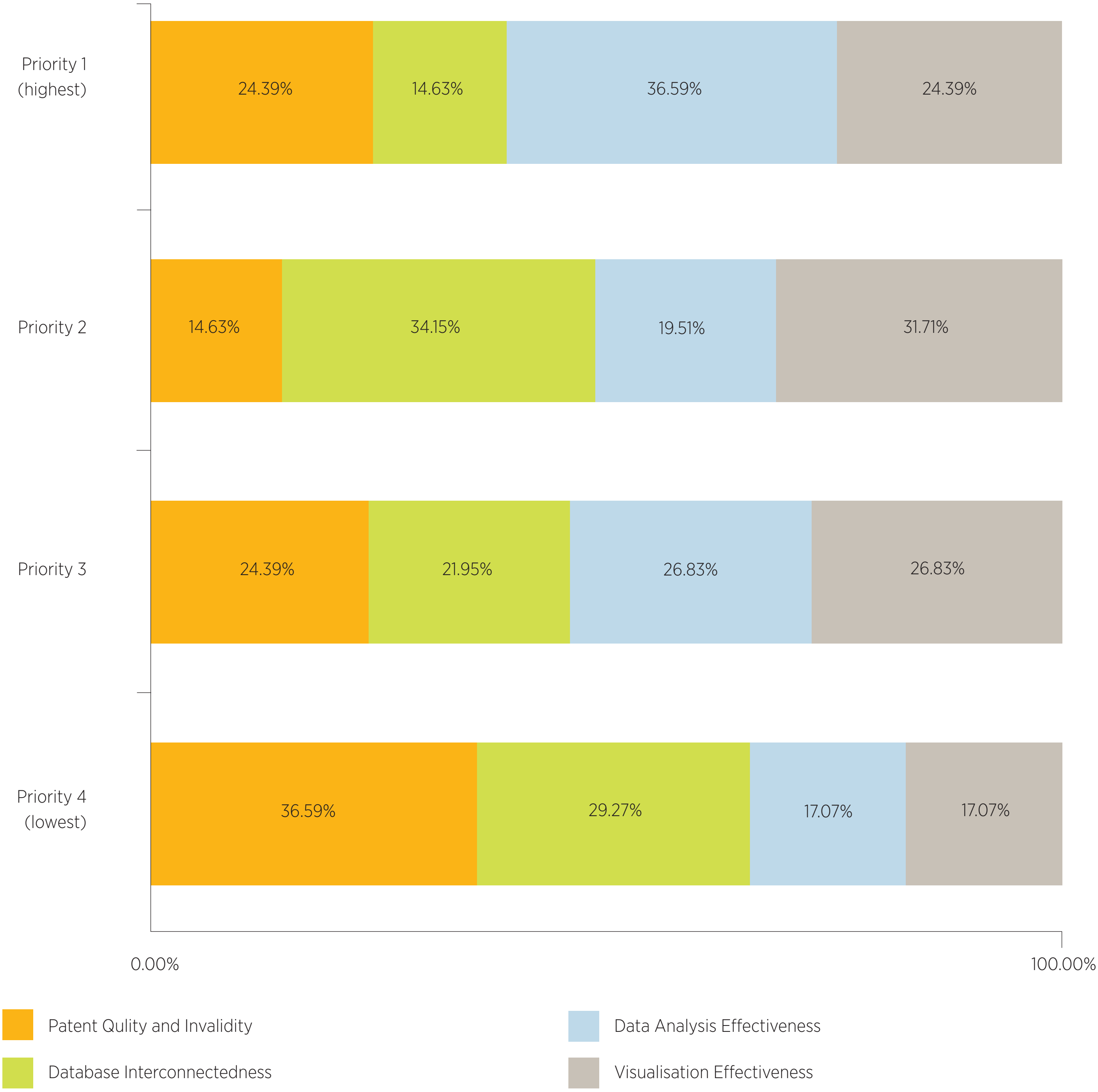


Figure 3 *Problem priority ranking by the survey participants (n = 70)*

MINI-TECHNOLOGY ROADMAPS (DQ A–E PROBLEM-FOCUSED TECHNOLOGY ROADMAPS)

In this section, the five mini-technology roadmaps constructed in phase 1 of stage 3, the explorative workshop stage, are presented. Each discussion question was used to initiate the technology roadmap for that particular patent domain problem theme, from which the discussion question arose. We deployed the technology roadmapping approach (Gerdsri 2013; Jeong et al. 2015; Phaal 2015; Phaal 2004; Phaal et al. 2001; Probert et al. 2003) to identify enabling factors and technologies for the patent analytics domain, ultimately producing a technology roadmap for the future of patent analytics. Technology roadmapping is a flexible technique that is widely used within industry to support strategic and long-range planning. The approach provides graphical and structured means for exploring and communicating the relationships between vision, problems, evolving and developing technologies and enabling factors, over time (Phaal 2004).

Each technology roadmap has three main layers: the problem milestones layer, the technology development layer and the enablers layer. Along the technology roadmap, the time axis is split into short term, medium term and long term, which are defined by the groups. Furthermore, every mini-technology roadmap has a vision. The patent analytics technologies, techniques and tools tree map to which the technology numbers in the roadmaps refer can be found in Figure 9.

Figure 4 shows the technology roadmap for discussion question A and problem theme A, patent data. Figure 5 shows the technology roadmap for discussion question B and problem theme B, database interconnectedness. Figure 6 shows the technology roadmap for discussion question C and problem theme C, data analysis effectiveness. Figure 7 shows the technology roadmap for discussion question D and problem theme D, visualisation effectiveness. Figure 8 shows the technology roadmap for question E and problem theme E, patent quality and invalidity.

DQ A: How can the quality of patent data be substantially improved?

The mini-technology roadmap around discussion question A and the problem theme of patent data portrays a vision of a completely harmonised open source patent data set, which includes patent ownership, litigation data, image search and image analysis. Figure 4 shows the DQ A mini-technology roadmap.

To address this discussion question, a quality parameter can be introduced in the data, which, in the short term, can include classification, and, in the medium term, concept identification. The relevant stakeholders should agree on the proposed quality parameter in terms of definition and value. In the long term, an overall agreement on the harmonisation of names, data ownership and litigation data would help to achieve improved data quality.

In terms of technology developments, in the short term, meta-database harmonisation is essential to make progress towards better patent data quality. This database should be linked to developments in citation and classification analysis to enable more comprehensive analysis of the meta-data, leading to more insightful results. In the medium term, full text analysis should be enabled by natural language processing and latent semantics. The latter also includes artificial intelligence, machine learning, neural networks and any other technology that assists in extracting semantics from data. In the long term, implementing further developed image searching and image analytics techniques will contribute substantially to progress in the patent analytics domain.

The main enablers identified for solving this problem include a substantially larger number of non-expert users and increasing demand for this type of analysis in industry, academia and technology-transfer offices. Availability of funding also appears to be an important enabler. Linking funding schemes with technology-transfer offices, industry, technology adopters and developers (e.g. academics, contract research or commercial vendors) is critical for accelerating the solution of this problem.

Problem theme A: Patent Data			
WHEN	**Short term (+3 years)**	**Medium term (5–10 years)**	**Long term (+10 years)**
Problem definition	• Concordance patent class vs industry class • Prescreening "bad" patents • Pre-classification text mining • WIPO ST legal status codes	• Harmonisation of XML standards • WIPO ST legal status codes • Litigation information • Patent ownership information	• Vision: harmonised complete data, including patent ownership, litigation data, image-searching analysis
Required research/ technology development	• T1 – NLP-based approaches • T1.1 – Keyword-based analysis • T1.3.3 – Similarity detection • T1.3.5 – Bibliographic coupling • T1.3.6 – World phase and action-object categorisation • T.10 – Patent quality • T11 – Geographical analysis • T13.3 – IPC classification • T13.1.1 – Claim analysis • T13.6 – Technology strength • T14 – Classification algorithms • T17.6 – Self citation analysis • T17.12 – Originality	• T.8 – Legal analysis • T.8.1 – Patent opposition • T.8.2 – Patent infringement • T.8.3 – Patent validity	• T18 – Open source
Milestones	• Classification 80%	• Major patent offices agree on a common XML standard • Concept identification – in agreement with human performance	
Enablers	• Public research funding bodies • Technology-transfer offices • Demands of corporate multinationals	• Open data/open datasets	
Risks			
High			
Medium		• Key commercial players fail	
Low			

Figure 4 *DQ A Mini-technology roadmap*

DQ B: How can the interconnectedness of patent databases with other data sources be enabled?

The mini-technology roadmap around discussion question B and the problem theme of database interconnectedness visualises databases that are connected to products and cross-referencing across all data streams. Figure 5 shows the mini-technology roadmap, which concentrates on how we can combine patent data with other data sources. Data interconnectedness is a combination of three essential sub-themes: standardisation, entity disambiguation and technology classification. Standardisation covers patent publication and application numbers.

In the short term, this can be achieved by standardising how patent numbers are generated. For example, legislation could facilitate reaching an agreement on the patent family definitions, so that the same patents are associated with the family throughout the lifecycle. In the medium term, an entity disambiguation, and a systematic way for technology classification with documented and widely accepted keys to link technologies together, are essential to enable the interconnectedness of patent databases with other data sources.

In terms of technology developments, there are existing technologies such as fuzzy logic algorithms, which can be used to progress the issues in the domain. In the long term, to resolve the problem of entity disambiguation and technology classification, more advanced systems and technologies are needed, which can be aided by significant progress in natural language processing, artificial intelligence and the development of ontologies.

Finally, legislation/regulation appears to be the biggest enabler for this discussion question, where strong cooperation between patent offices is essential to establish and implement common standards. For instance, better concordances can enable standardisation, and a coordinated effort by industry and academic experts can help to facilitate that.

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

Figure 5 *DQ B Mini-technology roadmap*

DQ C: How can better use be made of the valuable information contained in the patent data?

The mini-technology roadmap around discussion question C and the problem theme of exploring the effectiveness of data analysis envisions a fully automated, highly intelligent, highly adaptive analysis artificial intelligence system. Figure 6 shows the mini-technology roadmap.

In the short term, an expanded understanding of the existing analytic tools and techniques can be helpful for both industry and academia. Across all the time axes, a better understanding of the analytical requirements of users also appears essential, by understanding user analytic needs and the decisions that need to be made (Oldham & Kitsara 2016; Trippe 2015). Moreover, training, awareness-raising and certification are seen to be important. Although the above are not technical developments, they are necessary to establish transparency in terms of how the analytic tools are used, and can be used, to avoid a "black box" solution. This is an ongoing process with many strong enablers. There are governance bodies working on certification, such as the Qualified Patent Information Professionals (QPIP), the Patent Documentation Group (PDG), IP offices, the World Intellectual Property Organisation (WIPO), the European Patent Office (EPO), the United States Patent and Trademark Office (USPTO), the Organisation for Economic Co-operation and Development (OECD) and National IP offices.

In the medium term, the enablers for the development of the tools for effective data analysis are academic and research communities, and the private sector. In the long term, in approximately 15 years or more, reliable and efficient data-analysis methods will be essential for an increasingly large amount of data. Technologies such as full analysis automation through deep learning, artificial intelligence (AI), machine learning and predictive analytics could revolutionise this domain. Additional initiatives such as an open source community, open data, patent information community and working groups could facilitate these developments.

Problem theme C: Data-analysis effectiveness			
WHEN	**Short term (+1 year)**	**Medium term (+3 years)**	**Long term (+10 years)**
Problem definition	• Variety of analysis, but no one really knows what to use them for – should be appropriate to the user • What is the "heart" of the invention? (CPC/codes not enough) • Appropriateness of analysis types • Purpose-driven analysis (user-centric) • Define purpose of analysis (tailor to decision) • Transparency of tool – pathways of decisions	• Validations: making tools/data open source for analysis • Use orthogonal methods • Availability of complete data sets • Widespread adoption/ standardisation, • Complete assignee database • Openness of data	• Vision: full automations, highly intelligent artificial intelligence systems • Complete open, accurate data • Artificial intelligence
Required research/ technology development	• Data cleaning to generate reliable data sets to analyse • Analytical tools are black boxes, e.g. cluster analysis • Characterisation of technology – able to evaluate strengths/ weaknesses of a technology • Tools needed for efficient categorisation (CPC/IPC classifications are not good enough) • Customisation needed, automated and reliable analysis • T17 (citation analysis) and T13 (technology analysis) • Decisions need to be mapped to analytics tools	• Take advantage of big data • Generate meaningful answers • Maintain transparency of methods • Allow flexibility of approaches • Adapted to different users • Evolution of patent scientist/ analyst profile and skills (data, analytics) • All countries' data open and free • T27 (adaptive analysis) + T32 (predictive analytics) + T25 (In memory analytics) • Modelling/customisation of tools/preparation phase to deep learning/automation (decision support analysis)	• Crowd intelligence • Quality validation despite AI (extreme trust of users) • T24 (artificial intelligence) • T26 (deep learning analytics) • T30 (machine learning)
Milestones	• Professional certification of patent searchers • Standards in analysis • Purpose-driven approach to analysing data and transparency of process (often get "black box") • Matching user needs (decisions, criteria) and technology analysis tools	• Meaningful measurement of value/ worth and quality • Good estimates of data quality • Meaningful communication among various user certification groups	• Full artificial intelligence
Enablers	• Open source tools • Open patent data	• Patent info and working communities developing open source tools • Awareness-raising and training on tools and techniques • T24 – AI development and understanding of the capabilities of the technology, what happens within the "black box"	• Deep learning as an enabler to AI • Commercial vendors adopt proven open source tools
Risks			
High		• Separation of legal analytics from data analytics, leading to poor understanding	
Medium	• Tools, trends, lack of appropriateness • Misuse of tools (context)		• Users over-trusting AI
Low			

Figure 6 *DQ C Mini-technology roadmap*

DQ D: How can the results from patent analysis be visualised more effectively for better decision-making?

The mini-technology roadmap around discussion question D and the problem theme of effective visualisations of data analysis envisions an adaptive, interactive, intelligent, personalised search analysis with visualisation and interpretation. Figure 7 demonstrates the mini-technology roadmap.

In the short term, defining more precise user requirements and identifying use cases are essential, which can be achieved by establishing a patent analytics community, enabling the appropriate development of data visualisations. In addition, the production of guidelines and standards for the different interpretations of the visualisations can be helpful to minimise error and ambiguity.

Regarding the relevant technologies, machine learning has significant potential to be very useful in the short term, as long as it can be reconfigured for patent analytics. In terms of feasibility, collaboration between IP departments and machine-learning technology experts is important.

In the long term, efficient cartography methods, different types of visualisation for different usage, as well as adaptive/interactive visualisations, can have a substantial impact. In the long term, the ideal process can be to move directly from analysis to text and conclusions, eliminating the need for graphs, and thus eliminating the need for interpretation. This kind of technique already exists in other domains, which can be reapplied.

Problem theme D: Visualisation effectiveness			
WHEN	**Short term (+3 years)**	**Medium term (+5 years)**	**Long term (+10 years)**
Problem definition	• Unrealistic user expectations • Interpretation of visualisations • Need for multiple perspectives/ visualisations based on user need • Inconsistent visualisations when integrating with other data visualisations	• Automated generation of images from text • Comparison with images/drawings • Cross-visualisation of data; • truly big data visualisation	• Vision: adaptive, interactive, intelligent, personalised, search, analysis, visualisation and interpretation • Summarisation of results directly to text
Required research/ technology development	• Interactive visualisations are not "there yet" • Context-sensitive help • Use case analysis • T29 – cloud computing • T33 – data lakes • Measurement/KPIs for visualisation evaluation • More collaboration outside IP industry	• Tailoring visualisation depending on user groups • T28 – virtual reality + user interface • New visualisation techniques	• T30 – machine learning • T23 – artificial neural network analysis • T24 – artificial intelligence • T1 – NLP-based approaches • T2 – Semantic-analysis-based approaches • Natural language generation/ summarisation, automatic interpretation, predictive analysis
Milestones	• Established active/committed user community specific to visualisation • Certification for landscape analysis (best practice)	• AI is mature enough to help with visualisation • Dependent on developments in DQ A, DQ E, DQ B, DQ C in order of importance	
Enablers	• Data scientists • Demonstration of value of a good visualisation • Research into content, how a user would visualise data, best way to visualise data • People that specialise in technologies outside the IP industry	• Relevant computer power "on desk" of users • Research in user interaction specific to patents • Open cleaned data	
Risks			
High			
Medium	• Organise IT policies, security, etc.	• Continuity of service (someone pulls the plug)	• Validation/measurement of visualisation • Patent system replaced/ disappears/superseded
Low			

Figure 7 *DQ D Mini-technology roadmap*

DQ E: How are patent quality and patent invalidity identified?

The mini-technology roadmap around discussion question E and the problem theme of patent quality and invalidity has a vision of transparency and inter-linkage of data, where there is the ability to match patents with products in a level playing field. Figure 8 depicts the mini-technology roadmap.

An essential activity appears to be connecting different data sets and having multiple indicators that can be used in combination to determine patent quality. However, one should note that patent quality has different meanings for different stakeholders such as inventors, applicants, patent attorneys, opponents, information specialists, and so on. There are two main milestones: the integration of data from different databases; and the use of these integrated data sets. It is also likely that the system would benefit from the integration of additional information such as patent ownership, legal status data, economic data, product linkage data, licensing, transactions and standard essential patents. Ultimately, the integration of different data sources could lead to the availability of more data to determine patent quality.

In terms of technology development, improvements and refinements of the existing analytic metrics indicators and the development of identifiers are key to enabling the matching of different databases. Improvements in models using natural language processing, neural networks and deep-learning approaches can better address the inclusion of both structured and unstructured data into the databases. A key enabler of the above approach is the establishment of international standards in this area, supported by WIPO or national IP offices. Changing patent legislation is also likely to create the desired benefits, but can be very time-consuming. Changing accounting rules on how to incorporate and value intangible assets can also be a strong driver of change in this domain.

In future, having one organisation or a small number of organisations that coordinate data gathering is a valid consideration, which could be assisted by either private or public intermediaries to enhance data transparency. For database integration, the appropriate secure infrastructure is essential.

Problem E: Patent quality and invalidity			
WHEN	**Short term (+1–5 years)**	**Medium term (+5–15 years)**	**Long term (+15–20 years)**
Problem definition	• Young versus mature patents • 1) investments 2) assertion 3) citations 4) patent family 5) SEP as value parameter and licensing (in and out) • Multiple stakeholders have different visions/needs • Quality = value, legal assertion survived	• Lack of information for patents not taken to court/new patent applications • Identification of patent clusters • Description supports claims • Accurate/full disclosure • Worth evolves over time • Transparency of licensing and transactions ownership	• Vision: ability to match patents with products • Transparency and inter-linkage • Level playing field • Status quo favours larger corporations
Required research/ technology development	• Value citations • T1 – NLP-based approaches • Quality/value/worth – understand what is disclosed • Categorisation of patents • Link existing data better	• T30 – Automated (advanced computing tools) • Value of quality/value/worth against metrics for various stakeholders • T5 – neural network licence information (neural network for vectorisation)	• Patent data quality control from patent offices
Milestones	• Identification of similar patents to those litigated • T8 – legal status partly available	• Data availability • Link to other databases, e.g. with product/licensing information	
Enablers	• Open source • Open data • IP5 China, JPO, US, EPO, Korea • Legal changes to improve patent raw data • WIPO standards (WTO/WIPO standards)	• Faster examination by patent offices • Organisation/intermediary for linking data • Publish info when patent has expired	
Risks			
High			
Medium	• Inertia/industry resistance • Misalignment of incentives (competition vs transparency)	• Certain countries need to change further	
Low			

Figure 8 *DQ E Mini-technology roadmap*

TECHNOLOGY AS A KEY ENABLING FACTOR

Technology is regarded as a key enabling factor to help resolve many of the problems and the discussion questions presented in the earlier sections. The technology layer from each mini-technology roadmap has been carefully analysed to extract and identify developments and trends. In this section, we present the key technology impact analysis, ranking and classification analysis. We then summarise the main organisations contributing to these technologies and other related technologies that are currently in the technology-development process. The patent analytics technologies, techniques and tools tree map to which the technology numbers in the matrices refer can be found in Figure 9.

Priority technologies for the patent analytics domain

Priority technologies for the patent analytics domain (Figure 10) from the technology roadmapping have emerged as priorities across different problem groups from a scoring expert exercise that took place during the workshop, in priority order. The matrix is ranked according to the highest ranked technology. The matrix can also be read from the problem perspective, and the collective and individual level of impact for each technology on the specific problem theme. Also, Figure 10 shows the technology classification according to its current maturity status and use in the patent analytics domain, resulting from another scoring exercise that took place during phase 2 of the workshop (stage 3). Most priority technologies are in the "growth" phase, with some potentially high impact ones, for example, T8 – legal analysis, including legal status data worldwide and oppositions contested, and T10 – patent quality emerging from basic research. There is a good balance of technologies across the different process levels, with no evident gaps.

Some "additional" technologies (shown in bold), not included in Figure 9 and which it is essential to develop and use in this domain, are identified as important. These technologies, such as the ones for linking databases, were shown to have the highest impact in the domain, together with artificial intelligence technology, incorporating artificial neural network analysis, deep-learning analytics and machine learning. These are followed closely by classification algorithms and concordance. New technologies that allow databases to be linked and combined can potentially have a substantial impact on progressing the field. From the priority and new technologies identified, the majority complement DQ A, followed by E, D and C. It is also clear from the matrix that there is a gap in the technology for database interconnectedness, and therefore a need for it.

Complementary technologies

During the workshop, 15 complementary technologies were identified that may potentially play an important enabling role in accelerating the adoption and/or integration of the priority technologies into the patent analytics domain. These have been split into three categories and can be seen in Table 1.

Table 1 Complementary technologies (accelerating the adoption of priority technologies)

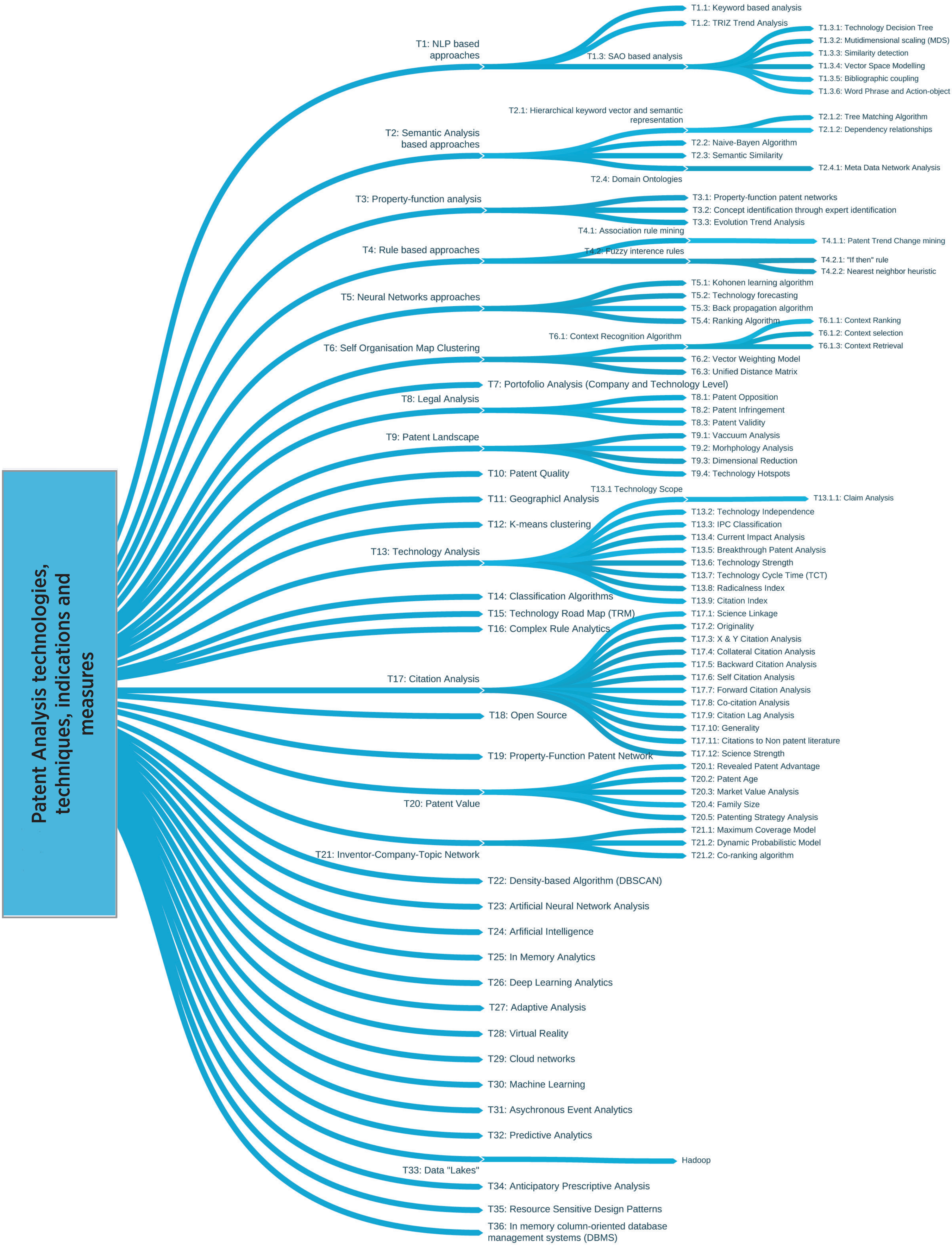
Technology Categories	Complementary Technologies
Tools and Methods	• Block chain • Automated effectiveness evaluation • Automated patent document translation • Automated drafting of patent applications, taking into account analytics while drafting • Quantum computing • Tools to facilitate NPL search • Technology forecasting • Computer-aided design
Databases	• Building concordance between existing taxonomies • OECD database of standardised names • OROPO ownership database • Better open source database software • Technologies for loading databases
Integration of existing technologies	• Integration of machine learning with other techniques • Inexpensive cloud computing and enabling platforms to harness cloud analysis

Main organisations working on priority technologies

During the workshop, a broad and non-comprehensive list was created of the academic and commercial organisations that are working on the development of some of the priority technologies. These include, but are not limited to, the ones shown in Table 2.

Table 2 List of academic and commercial organisations working on priority technologies

Priority Technologies	Organisations
T1 – NLP-based approaches, T14 – classification algorithms	Ambercite, Apache, Clarivate Analytics (formerly Thomson Reuters), Google, IBM, KU Leuven, Linguamatics, Lucene, Microsoft, Minesoft, NII Torio, Questel, Sheffield U Gate, Sole, TU Wien
T5 – neural-network-based approaches, T17 – citation analysis, T18 – open source, T24 – artificial intelligence	Aistemos, Alt legal, Anaqua, Article One Partners, Aulive, CB Insight, Clarivate Analytics, CPA Global, Darts – IP, Lexis Nexus, MaxVal, Patseer, Patient Vecto, Patentcore
T28 – virtual reality	Microsoft
T30 – machine learning/AI	Aistemos, Elsevier, Facebook, Google, IBM, Palbase, Quebec



Patent Analysis technologies, techniques, indications and measures
T1: NLP based approaches
T1.1: Keyword based analysis
T1.2: TRIZ Trend Analysis
T1.3: SAO based analysis
T1.3.1: Technology Decision Tree
T1.3.2: Mutidimensional scaling (MDS)
T1.3.3: Similarity detection
T1.3.4: Vector Space Modelling
T1.3.5: Bibliographic coupling
T1.3.6: Word Phrase and Action-object
T2: Semantic Analysis based approaches
T2.1: Hierarchical keyword vector and semantic representation
T2.1.2: Tree Matching Algorithm
T2.1.2: Dependency relationships
T2.2: Naive-Bayen Algorithm
T2.3: Semantic Similarity
T2.4.1: Meta Data Network Analysis
T2.4: Domain Ontologies
T3: Property-function analysis
T3.1: Property-function patent networks
T3.2: Concept identification through expert identification
T3.3: Evolution Trend Analysis
T4.1: Association rule mining
T4.1.1: Patent Trend Change mining
T4: Rule based approaches
T4.2: Fuzzy interence rules
T4.2.1: "If then" rule
T4.2.2: Nearest neighbor heuristic
T5.1: Kohonen learning algorithm
T5: Neural Networks approaches
T5.2: Technology forecasting
T5.3: Back propagation algorithm
T5.4: Ranking Algorithm
T6.1.1: Context Ranking
T6.1.2: Context selection
T6.1.3: Context Retrieval
T6.1: Context Recognition Algorithm
T6: Self Organisation Map Clustering
T6.2: Vector Weighting Model
T6.3: Unified Distance Matrix
T7: Portofolio Analysis (Company and Technology Level)
T8: Legal Analysis
T8.1: Patent Opposition
T8.2: Patent Infringement
T8.3: Patent Validity
T9: Patent Landscape
T9.1: Vaccuum Analysis
T9.2: Morhphology Analysis
T9.3: Dimensional Reduction
T9.4: Technology Hotspots
T10: Patent Quality
T13.1 Technology Scope
T13.1.1: Claim Analysis
T11: Geographicl Analysis
T13.2: Technology Independence
T12: K-means clustering
T13.3: IPC Classification
T13.4: Current Impact Analysis
T13: Technology Analysis
T13.5: Breakthrough Patent Analysis
T13.6: Technology Strength
T13.7: Technology Cycle Time (TCT)
T13.8: Radicalness Index
T13.9: Citation Index
T14: Classification Algorithms
T15: Technology Road Map (TRM)
T17.1: Science Linkage
T16: Complex Rule Analytics
T17.2: Originality
T17.3: X & Y Citation Analysis
T17.4: Collateral Citation Analysis
T17.5: Backward Citation Analysis
T17.6: Self Citation Analysis
T17: Citation Analysis
T17.7: Forward Citation Analysis
T17.8: Co-citation Analysis
T18: Open Source
T17.9: Citation Lag Analysis
T17.10: Generality
T17.11: Citations to Non patent literature
T19: Property-Function Patent Network
T17.12: Science Strength
T20.1: Revealed Patent Advantage
T20.2: Patent Age
T20.3: Market Value Analysis
T20: Patent Value
T20.4: Family Size
T20.5: Patenting Strategy Analysis
T21.1: Maximum Coverage Model
T21.2: Dynamic Probabilistic Model
T21: Inventor-Company-Topic Network
T21.2: Co-ranking algorithm
T22: Density-based Algorithm (DBSCAN)
T23: Artificial Neural Network Analysis
T24: Arfificial Intelligence
T25: In Memory Analytics
T26: Deep Learning Analytics
T27: Adaptive Analysis
T28: Virtual Reality
T29: Cloud networks
T30: Machine Learning
T31: Asychronous Event Analytics
T32: Predictive Analytics
Hadoop
T33: Data "Lakes"
T34: Anticipatory Prescriptive Analysis
T35: Resource Sensitive Design Patterns
T36: In memory column-oriented database management systems (DBMS)

Technologies	Discussion Question Impact Level					Maturity Level			Process Level		
	A	B	C	D	E	Embryonic	Growth	Mature	Pre-processing	Core-processing	Post-processing
Technology for linking databases; combination of patent data with economic and product life data	High	High	High	High	High			•	•		
T24 – artificial intelligence, incorporating T26 – deep-learning analytics, T30 – machine learning, T5 – neural network approaches, and T23 – artificial neural network analysis	High	Medium	High	High	High		•		•		
T1 – natural language processing (NLP) approaches	High	Medium	High	High	Medium		•	•	•	•	
T14 – classification algorithms and concordance with data system (e.g. NACE)	High	Medium	High	High	High		•		•		
T18 – open source	High	Low	High	Medium	High		•		•	•	•
T10 – patent quality (need to define "quality")	High	Low	High		High	•		•		•	•
T13 – technology analysis, including T13.1.1 – claim analysis and white space technology scouting	High	Medium		Medium	High		•			•	•
T17 – citation analysis, including T17.11 – citation to non-patent literature, and T17.1 – science linkage, as well as network analysis and applicant litigations	High	Medium	Medium	Low	Medium			•		•	
New visualisation techniques	High		High	Medium	Medium		•				•
T2.4 – domain ontologies	High	Low	Low	Medium	Medium		•		•		
T8 – legal analysis, including legal status data worldwide and oppositions contested	High	High	Low		Medium	•	•		•		
Automated document translation technology to ensure access to all international patents	Low	High	Low	Medium	Medium		•		•	•	
T2 – semantic analysis approaches and latent semantics	High	Low		Medium	Medium			•	•		
Empirical – conceptual/theoretical; use case analysis			High	High			•				•
Automatic interpretation – natural language generation (NLG)	Low	High		High		•					•
T28 – virtual reality and user interface (UI)			Low	Low			•				•

Notes: Dark colour indicates high impact, whereas blank indicates no impact. Technologies in bold are new technologies that have been identified, whereas all the others are priority technologies. Technology numbers (T1, etc.) refer to the technology numbering shown in Figure 9.

High (+3)	Medium (+2)	Low (+1)	No impact (0)

Impact colour key

Figure 10 *Impact scores for priority and new technologies for the patent analytics domain (consolidated) and priority technologies assessment in terms of maturity and process-level utilisation (consolidated)*

PATENT ANALYTICS DOMAIN TECHNOLOGY ROADMAP

In stage 3 of the research approach (Appendix I), the technology roadmap for the patent analytics domain was constructed, after integration of the five mini-technology roadmaps. The patent analytics domain technology roadmap has a vision of a fully adaptive, interactive, intelligent, personalised system with searching, analysis, visualisations and interpretation. The time frame that is envisioned is approximately fifteen years or more. The integrated technology roadmap is an integration of the mini-technology roadmaps and helps to provide enablers in overcoming with a joint consolidated approach the five problem themes under discussion.

Figure 11 shows the overall technology roadmap, with three clearly articulated layers – the problem-solving milestones layer, the technology developments layer and the key enablers layer – which are required over time to progress the field. Different pathways are highlighted for resolving the most pressing problems in the domain, for instance, either through the further development of AI technologies and their integration with neural networks and related citation protocols of technologies, or by facilitating implementation of the key enablers (Table 3) necessary for the resolution of the issues in this area.

The four main problem-solving milestones for the patent analytics domain are: first, automating patent classification; second, transparent and consistent clarification and clustering of information; third, having cleaner, standardised and interlinked patent data with other data; and, fourth, the creation of appropriate use cases for user groups, for understanding decision needs. The required technology developments are focused around the integration and validation of artificial intelligence, neural networks and citation protocols. This is complemented by the alignment of different databases to enable the compatibility of data and visualisations.

The field can benefit from more emphasis being placed in key enablers, especially on the cooperation of different organisations, such as WIPO, EPO and OECD. Furthermore, incentives to applicants to write clearer abstracts that enable easier classification of patent applications can act as an enabler, followed by a standardised (harmonised) legislation. In terms of the technology cycle enablers, these can be identified as funding resources, open source community development and infrastructure development for security to protect the patent data with the interconnected databases. In terms of legislation, enablers such as legislation for cooperation between intellectual property offices and internal standards are important for harmonising and converging the patent data. The main issues in this field are the lack of appropriate data tagging, ontologies or taxonomies, and the data not being well organised.

Three key insights from the process of synthesising the patent domain technology roadmap were derived. First, in order to progress the field of patent analytics, a number of use cases can be generated, which can help to link user group needs to technology developments and decision-making. Second, the technologies that are most required are already known, some of which are currently in use by the patent analyst experts. One requirement to aid and guide the technology adoption is to create more specialised training for both developers and end-users in key technologies, adopting a data science profile. The aim of this is that these analytic technologies stop being regarded as “black box” solutions and can be customised for specific needs. Finally, the greatest impact in the domain can be achieved only with the cooperation of different organisations and standardised legislation.

Table 3 *Patent analytics domain technology roadmap enablers*

Theme	Enablers
Technology development cycle/ methodologies	• Market (users) demand – industry, academic, technology transfer office, policy and decision-makers • Funders – resources – staff, premises • Technology transfer – academia and commercialisation • Producers of the technology – academics, contract research, commercial vendors • Open source development • Clarify choice and definition of families • Open source tools and community. Open data patent information communities working groups • Cooperation between academics and the private sector • Infrastructure to protect the linked data security standards
Legislation	• Legislation cooperation between IPOs • Standardised legislation • International standards (e.g. WIPO) IP5 and legal changes for patents
Training/continuous professional development	• Changes in evolution of patent scientist/analyst profile • Training, awareness certification • Transparency (no black box tools) • Training of developers and end-users in patent analytics and visualisation • Training for QPIP/ISBQPIP PDG
Cooperation	• “5–10” collaborations between IP tool suppliers and external visualisation experts and data sciences • Increased cooperation between WIPO, EPO, USPTO and OECD • Incentives to write informative abstracts, requiring applicants to classify the application • Organisation(s) to run the integrated data, for example, patent offices, private intermediary firms • Concordance, collaboration with industry experts

Roadmap for Patent Analytics March 2017		Short term 1–3 years
Problem solving milestones	Technical analysis	Automation of patent clarification – 80%
	Data quality	Cleaner data standardisation
	Other	Understand user groups/use cases (visualisation)
Technologies	Miscellaneous technologies	Methods to define application of technologies – tools
	AI, neural networks citations NLP, and related methods	AI, neural networks and related citation methods established preliminary protocols
	Gaps	What technologies have not been developed yet? What existing technologies are not used in this area?
Enablers	Technology development cycle/ Methodologies	Market (users) demand – industry, academic, TTO, policy and decision makers Funders –resources – staff, premises, money Technology transfer – academia and commercialiser Producers of the technology – academics, contract research, commercial vendors Open source development Clarify choice and definition of families
	Legislation	Legislation cooperation between IPOs
	Cooperation	"5-10" collaborations between IP tool suppliers and external visualization experts and data sciences Increased cooperation between WIPO, EPO, USPTO, OECD Incentives to write informative abstracts, require applicants to classify the application
	Training/Continuous professional development	Changes evolution of patent scientist/analyst profile Training, awareness certification. Transparency (no black box tools) Training of developers and end users in patent analytics and visualisation Training for QPIP/ISBQPIP PDG
	Gaps	Lack of appropriate tagging of data Data mess

Figure 11 *Patent analysis roadmap*

Medium term 3–10 years	Long term 10–20 years
Automated document clustering based on content	Cross referencing – linking patent and non patent classification
Legal information ownership. Transparent and consistent	
	Alignment/compatibility of databases – data – analytics – visualisations Search and analytics of images, designs drawings
AI, neural networks and related citation protocols are optimised, validated and incorporated into commercial offerings	AI, neural networks and related citation protocols are fully accepted validated alongside other orthogonal tools
	What technologies have not been developed yet? What existing technologies are not used in this area?
Open source tools and community. Open data pat Information communities working Cooperation between Acadgermouicpss/R TD and the private sector Infrastructure to protect the linked data security standards	
Standardised legislation International Standards (e.g. WIPO) IP5 and legal changes for patents	
Organisation(s) to run the integrated data e.g. patent offices, private intermediary firms	Concordance, collaboration with industry experts

Vision 15+ years

Adaptive, interactive, intelligent, personalised, search, analysis, visualisation and interpretation

CONCLUSION

This project contributes to expanding the field of patent analytics for more effective exploitation of the largest worldwide repository of technological information to enable new use cases, supporting better decision-making and partnerships of R&D pursuing organisations. This was achieved by developing a public technology roadmap to facilitate the collaboration and coordinated action of actors in the patent analytics community to further develop the capabilities for analysing patent data. Using a technology roadmapping problem-solving approach, the research design involved more than a hundred experts from academia and industry in the patent analytics domain, to develop a patent analytics domain roadmap (Figure 11). We have identified 16 technologies that can contribute to solving the 5 problem themes and 15 complementary technologies. In addition, of these 16 technology families, we have identified 5 additional technologies, which can complement and aid the process. Finally, 21 enablers have been identified, which play an important and equal role in resolving the 5 problem themes in the domain, and are classified under the themes of technology development cycle/methodologies, legislation, training/continuous professional development and cooperation. This section summarises the findings in terms of technology development identification, enabler identification and key messages and next steps.

PATENT ANALYTICS DOMAIN ENABLERS

Twenty-one enablers have been identified, which were found to play an equal, if not more important, role than the technologies in enabling solutions for the five problem themes in the domain. The enablers are classified under the themes of technology development cycle/methodologies, legislation, training/continuous professional development and cooperation.

TECHNOLOGY DEVELOPMENT IDENTIFICATION

The top five priority technologies identified are:

- Technologies for linking databases; combination of patent data with economic and product life data;
- T24 – artificial intelligence, incorporating T26 – deep-learning analytics, T30 – machine learning, T5 – neural network approaches, and T23 – artificial neural network analysis;
- T1 – natural language processing (NLP) approaches;
- T14 – classification algorithms and concordance with data system (e.g. NACE); and
- T18 – open source.

KEY MESSAGES AND NEXT STEPS

The key messages derived from this work are as follows:

- First, better data quality is important. Data is valuable beyond its original purpose and there is an urgent need for more structured and cleaner standardised data. In addition, open data can contribute to data quality and data repair.
- Second, the identification of key stakeholders and different user group needs is important in extracting the information needs and use cases. By understanding users, appropriate use cases can be defined.
- Third, there is a need for training in using different technologies. This can be aided by increasing the transparency and traceability of using different analytic technologies, techniques and tools, and by the provision of guidelines and online material.
- Fourth, legislation and standardisation can aide transparency and the adoption of technologies in the patent domain.

The next key actions from the technology roadmap should be to generate use cases for different users and/or user groups (these could possibly be created by technology vendors) and the standardisation and harmonisation of patent ontologies by WIPO and member states. The final action would be to implement standards of reporting that are disclosed.

APPENDIX

APPENDIX I – METHODOLOGY

This study deployed a technology roadmapping approach (Gerdsri 2013; Jeong et al. 2015; Phaal 2015; Phaal 2004; Phaal et al. 2001; Probert et al. 2003) consisting of three stages, with the first two preparatory stages providing input for the third stage, a workshop that was run in March 2017 as a core element of this approach for developing a patent analytics domain roadmap. The research was guided by principles that are commonly used to establish the quality of research: validity and reliability (Bryman 2012; Creswell 2013; Flick 2009), increasing quality and robustness of the research design. Figure 1 illustrates the research process.

First, in the identification stage, we conducted desk and literature reviews (Creswell 2013; Cronin et al. 2008), as well as expert consultations, to identify problem themes and technologies that could have a substantial impact on the patent analytics domain.

Second, in the verification stage we reached out to the relevant stakeholder communities using an online survey (Bryman 2012; Flick 2009). Seventy respondents provided input to further identify, prioritise and eliminate technologies and problem themes arising from stage 1.

In the third exploration stage we ran a workshop with 28 carefully selected experts covering a variety of stakeholder perspectives from both academia and industry. The workshop had three main phases: in the first phase participants followed a problem-solving approach to develop five mini-technology roadmaps in groups. Second, they extracted information on technologies from the technology layer of the mini-technology roadmap, which can enable the field. In the third phase, the technology roadmap was synthesised by combining the key elements from the initial mini-technology roadmaps (phases 1 and 2) created for each of the five patent analytics domain problems, the information from stages 1 and 2 of the research design and the examination of the three layers (problem milestone, technology and enablers).

Stage 1 Identification

In the first stage, we conducted desk and literature reviews (Creswell 2013; Cronin et al. 2008) and expert consultations to identify problems and technologies that could have a substantial impact on the patent analytics domain. Our aim was twofold:

- First, we aimed to identify technologies with the potential to deliver breakthroughs to enable previously untapped use cases. These can be either emerging or relatively mature in other fields and not yet deployed in the field of patent analytics. We developed a technology literature map (Creswell 2013) for the different patent analytic technologies, techniques and tools.
- Second, we aimed to identify current problems within the field that cannot be solved with the currently available technologies, and grouped them into themes. From the data analysis, we derived four themes, which led into the four discussion questions elaborated above. In addition, we identified experts through snowballing, networking, interviews and desk research.

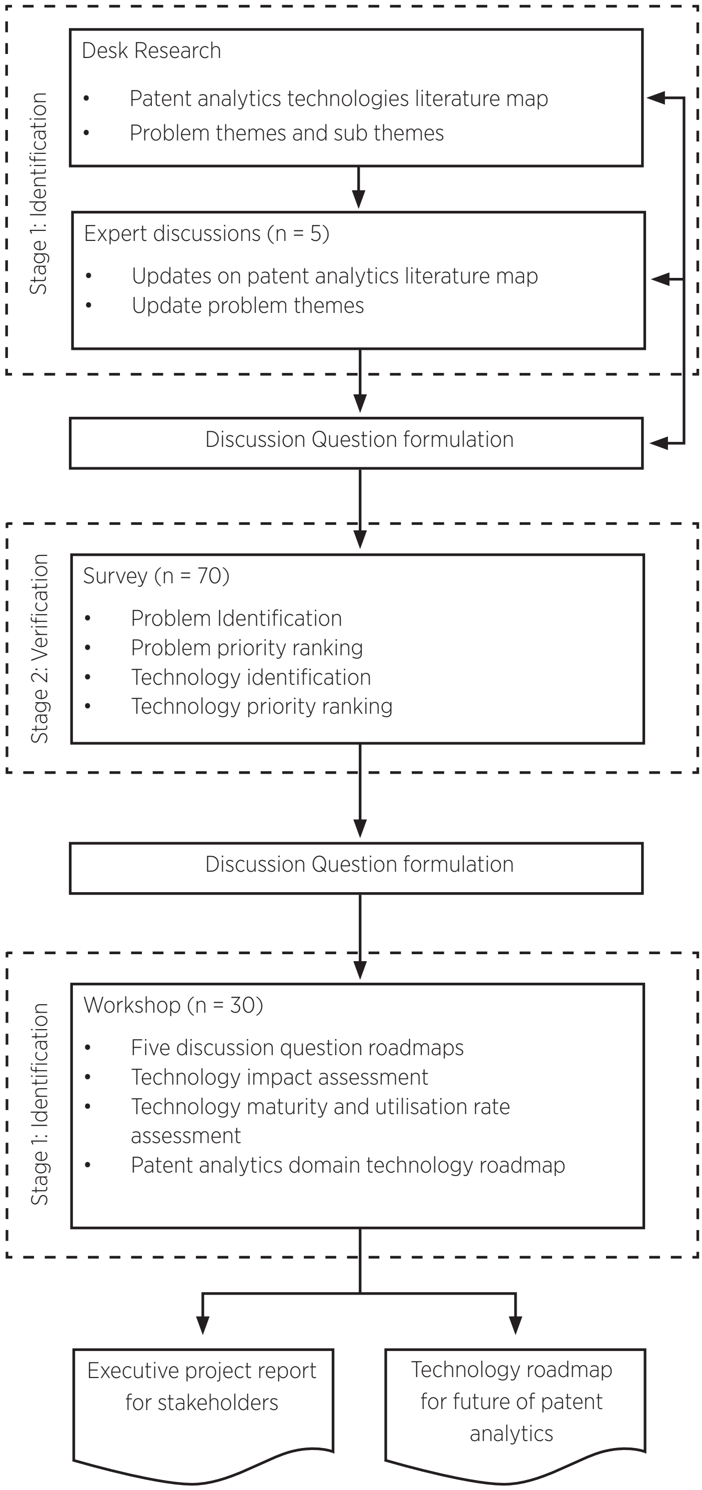


Figure 13 *Research design*

Stage 2 Verification

In the second stage, the verification stage, we reached out to different stakeholder communities using an online survey (Bryman 2012) to verify and prioritise the technologies and problems identified in the first stage. At the same time, we looked to identify more problem themes and sub-themes, technologies and experts. The survey had three parts:

- First, participants ranked and prioritised the technologies and problems identified in stage 1. There were five technologies and four problem themes for prioritisation and ranking.
- Second, participants identified other technologies and problem themes related to the patent domain. From the data analysis, one more problem theme, A, arose as being important in the patent analytics domain.
- Third, the participants were asked to identify experts that we can contact for further information regarding technologies and problem themes in the patent domain.

The online survey was aimed at the wide variety of stakeholders (Robson 2011), which falls into five main groups. These groups are shown in Table 5, and they are the same ones that were used to identify participants for stage 3 of the project. The survey had 70 participants from a variety of sectors and experience within the patent analytics domain. The data collected from the survey was then analysed to ensure consistency of the problem themes identified in stage 1. From the results, another problem theme arose, problem A, with several sub-themes, which was then rephrased in a discussion question, DQ A. The five verified discussion questions were then used for stage 3.

Table 4 *Survey stakeholder groups*

Group	Stakeholders
1	Patent information specialists, patent analysts, technology specialists
2	Industry lead-users of patent analytics solutions, chief technology officers, decision-makers
3	Patent analytics providers, patent offices
4	Lead academics in the patent domain
5	Patent attorneys, technology-transfer offices, intellectual property consultants

Stage 3 Exploration

In the third stage, the exploration stage, we ran a workshop with 28 carefully selected experts from the patent analytics field. The experts came from a variety of countries (such as the USA, the UK, France, Germany, Switzerland, Austria, Turkey) and were selected to cover a variety of stakeholder groups in the patent analytics domain, from academia, governmental organisations (EPO, UKIPO, WIPO) and small, medium and large companies. These stakeholder groups are shown in Table 4, which are also the ones at which the survey was targeted in stage 2. As part of the technology roadmapping approach used throughout the project (Gerdsri 2013; Jeong et al. 2015; Phaal 2015; Phaal 2004; Phaal et al. 2001; Probert et al. 2003) to identify enabling factors and technologies for the patent analytics domain, stages 1 and 2 were the preparatory stages leading to stage 3 and provided the input for it.

Technology roadmapping is a flexible technique that is widely used within industry to support strategic and long-range planning. The approach provides graphical and structured means for exploring and communicating the relationships between vision, problems, evolving and developing technologies and enabling factors, over time (Phaal 2004). The roadmap allows the integration and alignment of a number of different perspectives across a broad time range. In this way, the development of currently developing, or short-term, underpinning science and technology to support the long-term vision can be explored.

The workshop was divided into three main phases:

- **Phase 1: Exploration of the five discussion questions in groups**
  In this phase, the participants explored the five discussion questions in groups of six and the identification of common important technology. Each group had one of each of the stakeholders from Table 5, leading to the development of five mini-technology roadmaps. In this phase, the participants first articulated and specified the discussion question and the desired vision. Then, they summarised the research and technology developments required to address this problem, followed by specifying the milestones required to deliver a solution to this problem. Second, the participants specified the key organisations that lead technology developments globally, other complementary technologies and the ones accelerating the adoption of main technologies. Third, the participants identified resource requirements, enablers, key risks, gaps, barriers and weaknesses to enable the technology in order to overcome the problem, leading to the vision.
- **Phase 2: Technology development identification**
  In addition, in this phase each group was asked to summarise the five most important technologies from the five mini-technology roadmaps developed in phase 1 to enable their discussion question. Then, they first assessed the technologies (Figures 10, 11 and 14) in terms of their potential impact (Featherston & O'Sullivan 2014). "Impact" is defined as "How this technology will impact the solution of current problems in the patent analytics domain". The impact scores ranged from (-3) to (+3). Then, they assessed the technology's maturity level (Ilevbare 2013; Kohli et al. 2010) and process level (Moehrle et al. 2010). The criteria for the above assessments can be found in Tables 6–8.
- **Phase 3: Integration and development of the technology roadmap**
  In this phase, each group constructed a summary narrative of their mini-technology roadmap, populating the overall technology roadmap from the synthesis of the five mini-technology roadmaps. The participants reviewed, in a plenary session, each layer for pathways in time, grouping similar ideas and synthesising pathways within the layers. Summarised key messages and key pathways were created across the technology roadmap. Through constructive discussion, the key risks and potential gaps were also identified, creating a required action plan to eliminate these. The final technology roadmap illustrates the future of patent analytics, identifying the key milestones/breakthroughs and pointing out how and when fundamental problems can be overcome.

Table 5 *Impact assessment criteria*

Impact Rating	Description
+3	Creates significant change that accelerates the solution in patent analytics
0	Creates no change or effect in the solution of the problem in patent analytics
-3	Strongly delays or stops altogether the solution in the patent analytics

Table 6 *Technology maturity assessment criteria*

Technology Maturity	Description
Embryonic	Fundamental research, where technological possibilities are conceived or discovered, or both
Growth	Technology, which can be configured possibly with other technologies, to form a proprietary technology in a "proof of concept"
Mature	Established technologies and processes that are routinely used in patent analytics.

Table 7 *Patent analytics process-level assessment criteria*

Patent analytics process level	Description
Pre-processing technology	Technologies that support the pre-processing stage of patent analysis, for data management, vectorisation and preparation
Core-processing technology	Technologies that support the core-processing of patent analysis and can be used to analyse and identify relationships
Knowledge/ post-processing technology	Technologies that support the visualisation of patent analytic information and provide and meaningful insights arising from the information

APPENDIX II – PRIORITY TECHNOLOGIES AND SCORES FOR PATENT ANALYTICS PROBLEMS

	Patent analytics domain problem					
	A	B	C	D	E	Total
T1 – NLP-based approaches	3	2	2	2	3	12
T10 – patent quality (need to define "quality")	3	1	3	0	3	10
T13 – technology analysis, including T13.1.1 – claim analysis and white space technology scouting	3	2	0	2	3	10
T14 – classification algorithms and concordance with data system (e.g. NACE)	3	2	3	3	2	13
T17 – citation analysis that also includes applicant litigations	3	2	2	1	2	10
T1 – NLP-based approaches	3	2	2	2	3	12
T2 – semantic-analysis-based approaches and latent semantics	3	1	-1	2	2	7
T2.4 – domain ontologies	3	1	1	2	2	9
T24 – artificial intelligence, incorporating T26 – deep-learning analytics, T30 – machine learning, T23 – artificial neural network analysis, and T5 – neural-network-based approaches	3	2	3	3	3	14
T17 – citation analysis, including T17.11 – citation to non-patent literature, and T17.1 – science linkage	3	2	2	1	2	10
T17 – citation analysis, especially network analysis	3	2	2	1		8
T24 – artificial intelligence, including T26 – deep-learning analytics	3	2	3	3	2	13
T1 – NLP-based approaches	3	2	2	2	3	12
T5 – neural-network-based approaches	3	2	3	2	2	12
T18 – open source	3	1	3	2	3	12
Empirical – conceptual/theoretical Use case analysis	0	0	3	3		6
T30 – machine learning, including T24 – artificial Intelligence, and T1 – NLP-based approaches	3	2	3	3	3	14
New visualisation techniques	3	0	3	2	2	10
T28 – virtual reality + user interface (UI)	0	0	1	1		2
Automatic interpretation (natural language generation (NLG))	1	2	0	3		6
T1 – NLP Natural language processing for analysing patent contract data	3	2	2	2	2	11
Technology for linking databases; combination of patent data with economic and product life data	3	3	3	3	3	15
Automated document translation technology to ensure access to all international patents	1	2	1	2	2	8
T8 – legal analysis, including legal status data worldwide and oppositions contested	3	2	1	0	2	8
T30 – machine learning for 1) state of the art, 2) incomplete data, and 3) value versus objectives	3	2	3	3	2	13
Total	65	41	50	50	51	

Figure 14 *Priority technologies and scores for each patent analytics problem*

REFERENCES

Abbas, A., Zhang, L. & Khan, S.U., 2014. A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37, pp.3–13. Available at: http://dx.doi.org/10.1016/j.wpi.2013.12.006.

Baudour, F. & van de Kuilen, A., 2015. Evolution of the Patent Information World - Challenges of yesterday, today and tomorrow. *World Patent Information*, 40, pp.4–9. Available at: http://dx.doi.org/10.1016/j.wpi.2014.10.001.

Bonino, D., Ciaramella, A. & Corno, F., 2010. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, 32(1), pp.30–38. Available at: http://dx.doi.org/10.1016/j.wpi.2009.05.008.

Brügmann, S. et al., 2015. Towards content-oriented patent document processing: Intelligent patent analysis and summarization. *World Patent Information*, 40, pp.30–42. Available at: https://doi.org/10.1016/j.wpi.2014.10.003

Bryman, A., 2012. *Social Research Methods*, Oxford University Press

Creswell, J.W., 2013. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* 4th Edition

Cronin, P., Ryan, F. & Coughlan, M., 2008. Undertaking a literature review : a step-by-step approach., *British J ournal of Nursing*, 17(1), pp.38–43. Available at: https://doi.org/10.12968/bjon.2008.17.1.28059

Dintzner, J.P. & Van Thieleny, J., 1991. Image handling at the European Patent Office: BACON and first page. *World Patent Information*, 13(3), pp.152–154. Available at: https://doi.org/10.1016/0172-2190(91)90070-L

Featherston, C. & O'Sullivan, E., 2014. The different roles of technology in technological emergence, *Briefing note, Centre for Science Technology and Innovation Policy (CSTI), Institute for Manufacturing, University of Cambridge*

Ferrari, A.C. et al., 2014. Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems. *Nanoscale*, 7(11), pp.4598–4810. Available at: http://pubs.rsc.org/en/Content/ArticleLanding/2014/NR/C4NR01600A

Flick, U., 2009. An introduction to qualitative research. Sage, 4th Edition

Gassmann, O. et al., 2012. The role of IT for managing intellectual property - An empirical analysis. *World Patent Information*, 34(3), pp.216–223. Available at: http://dx.doi.org/10.1016/j.wpi.2012.03.005.

Gerdsri N. (2013) Implementing Technology Roadmapping in an Organization. In: Moehrle M., Isenmann R., Phaal R. (eds) Technology Roadmapping for Strategy and Innovation. Springer, Berlin, Heidelberg, Available at: https://doi.org/10.1007/978-3-642-33923-3_12

Ilevbare, I. M. (2013). An investigation into the treatment of uncertainty and risk in roadmapping: a framework and a practical process (doctoral thesis). https://doi.org/10.17863/CAM.14064

Jeong, Y. et al., 2015. Development of a patent roadmap through the Generative Topographic Mapping and Bass diffusion model. Journal of Engineering and Technology Management, 38, pp.53–70. Available at: http://dx.doi.org/10.1016/j.jengtecman.2015.08.006

Kohli, R. et al., 2010. Decision Gate Process for Assessment of a Technology Development Portfolio. The Aeropsace Corporation, Civil and Commercial Operations Houston Programs

Lee, C., Jeon, J. & Park, Y., 2011. Monitoring trends of technological changes based on the dynamic patent lattice: A modified formal concept analysis approach. *Technological Forecasting and Social Change*, 78(4), pp.690–702. Available at: http://dx.doi.org/10.1016/j.techfore.2010.11.010.

Lupu, M., Tait, J., & Trippe, A. J. (2011). Current challenges in patent information retrieval (Vol. 29). K. Mayer (Ed.). Berlin: Springer

Lupu, M., 2017. Information retrieval, machine learning, and Natural Language Processing for intellectual property information. *World Patent Information*, 49, pp.A1–A3. Available at: https://doi.org/10.1016/j.wpi.2017.06.002

Martinez, C. (2010), Insight into Different Types of Patent Families, OECD Science, Technology and Industry Working Papers, No. 2010/02, OECD Publishing, Paris. Available at: http://dx.doi.org/10.1787/5kml97dr6ptl-en

Martínez, C., 2011. Patent families: When do different definitions really matter? *Scientometrics*, 86(1), pp.39–63. Available at: http://doi.org/ 10.1007/s11192-010-0251-3

Masiakowski, P. & Wang, S., 2013. Integration of software tools in patent analysis. *World Patent Information*, 35(2), pp.97–104. Available at: http://dx.doi.org/10.1016/j.wpi.2012.12.010.

Moehrle, M.G. et al., 2010. Patinformatics as a business process: A guideline through patent research tasks and tools. *World Patent Information*, 32(4), pp.291–299. Available at: http://dx.doi.org/10.1016/j.wpi.2009.11.003.

OECD, 2016. Enabling the Next Production Revolution: the Future of Manufacturing and Services - Interim Report, In: OECD (2017), The Next Production Revolution: Implications for Governments and Business, OECD Publishing, Paris, Available at: http://dx.doi.org/10.1787/9789264271036-en

OECD, 2009. OECD Patent Statistics Manual, OECD, ISBN 978-92-64-05412-7, Available at: http://www.oecd-ilibrary.org/science-and-technology/oecd-patent-statistics-manual_9789264056442-en.

Oldham, P. & Kitsara, I., 2016. The WIPO Manual on Open Source Patent Analytics. *World Intellectual Property Organization*. Available at: https://wipo-analytics.github.io [Accessed August 8, 2017].

Phaal, R., 2015. Roadmapping for strategy and innovation., Briefing note, March, Centre for Technology Management, Institute for Manufacturing, Department of Engineering, University of Cambridge

Phaal, R., 2004. Technology roadmapping - A planning framework for evolution and revolution. Technological Forecasting and Social Change, 71(1–2), pp.5–26. Available at: http://doi.org/10.1016/S0040-1625(03)00072-6

Phaal, R., Farrukh, C.J.P. & Probert, D.R., 2001. Technology Roadmapping: linking technology resources to business objectives. *International Journal of Technology Management*, 26(1), p.2. Available at: http://www.brm-toolkit.com/index_bestanden/technology_roadmapping.pdf.

Probert, D.R., Farrukh, C.J.P. & Phaal, R., 2003. Technology roadmapping Developing a practical approach for linking resources to strategic goals , Proceeding of the Institution of Mechanical Engineeris, Part B: Journal of Enginering Manufacture, Vol. 217, Issue 9, pp.1183–1195. Available at: https://doi.org/10.1243/095440503322420115

Randhawa, K., Wilden, R. & Hohberger, J., 2016. A Bibliometric Review of Open Innovation: Setting a Research Agenda. Journal of Product Innovation Management, 33(6), pp.750–772. Available at: http://doi.org/10.1111/jpim.12312

Raturi, Mamta Kumari and Sahoo, Prabhat Kumar and Mukherjee, Susmita and Tiwari, Amit Kumar, 2010. Patinformatics – An Emerging Scientific Discipline (March 6, 2010). Available at: http://dx.doi.org/10.2139/ssrn.1566067

Robson, C., 2011. Real world research. Edition. Blackwell Publishing. Malden.

Squicciarini, M., Dernis, H. & Criscuolo, C., 2013. Measuring Patent Quality: Indicators of Technological and Economic Value. *OECD Science, Technology and Industry Working Papers*, (3), p.70. Available at: http://www.oecd-ilibrary.org/science-and-technology/measuring-patent-quality_5k4522wkw1r8-en.

Tietze, F. & Lauritzen, G.D., 2016. IP challenges in multi-partner collaboration. *Institute for Collaborative working: The Partner*, (May), pp.90–91.

Tietze, F. & Probert, D., 2015. Patent Informatics for collaboration management. *The Partner*, May, pp.109–111.

Trappey, A.J.C. et al., 2012. A patent quality analysis for innovative technology and product development. *Advanced Engineering Informatics*, 26(1), pp.26–34. Available at: http://www.sciencedirect.com/science/article/pii/S1474034611000486.

Trippe, A., 2015. *Guidelines for Preparing Patent Landscape Reports*, Available at: http://www.wipo.int/edocs/pubdocs/en/wipo_pub_946.pdf.

Trippe, A.J., 2003. Patinformatics: Tasks to tools. World Patent Information, 25(3), pp.211–221. Available at: http://doi.org/10.1016/S0172-2190(03)00079-6

West, J & Bogers, M 2014, ‘Leveraging External Sources of Innovation: A Review of Research on Open Innovation’ Journal of Product Innovation Management, vol 31, no. 4, pp. 814–831. DOI: 10.1111/jpim.12125

WIPO, 2016. *WIPO IP Facts and Figures 2016*, Available at: http://www.wipo.int/publications/en/details.jsp?id=4157

WIPO, 2015. World Intellectual Property Report: Breakthrough innovation and economic growth, Available at: http://www.wipo.int/publications/en/details.jsp?id=3995

BIOGRAPHIES

Mr. Leonidas Aristodemou

Leonidas is a doctoral researcher in Technology, Innovation and Intellectual Property Management, at the Centre for Technology Management, Institute for Manufacturing, University of Cambridge. His current research revolves on the usage of Big Data Intellectual Property Analytics to make informed decisions along technology development innovation models. He is an executive committee member of the Cambridge University Engineering Association and a member of St. Edmund's College. His research interests include big data, strategic decision making, artificial intelligence and machine learning (LinkedIn: https://uk.linkedin.com/in/leonidasaristodemou).

la324@cam.ac.uk

Dr. Frank Tietze

Frank Tietze is Lecturer in Technology and Innovation Management at the University of Cambridge, Department of Engineering. At the Centre for Technology Management he leads the Innovation and Intellectual Property Management research group. Frank serves on the Exploitation Committee of Chartered Institute for Patent Attorneys (UK) and is a member of the UKIPO Research Expert Advisory Group for Registered Rights. He is Affiliate at the Cambridge Centre for Intellectual Property and Information Law and member of the Innovation and IP research group at Chalmers University of Technology (Sweden).

frank.tietze@eng.cam.ac.uk

Institute for Manufacturing (IfM)

The IfM is part of the University of Cambridge. With a focus on manufacturing industries it creates, develops and deploys new insights into management, technology and policy.

Its mission is scholarship, world-leading research and education that:

- Creates knowledge, insights and technologies of value to established and emerging manufacturing industries and the associated policy community.
- Develops the skills and capabilities of researchers, students and people working in industry and government.
- Has an impact on industry and government through knowledge transfer and application, as well as by influencing the direction of future research.

It aims to be the partner of choice for businesses and policymakers, as they enhance manufacturing processes, systems and supply chains to deliver sustainable economic growth through productivity and innovation.

The Centre for Technology Management (CTM)

The CTM is one of several research centres within the IfM. CTM focuses on helping managers to make the most appropriate use of current and future technological resources. It aims to provide comprehensive support to managers, based on an integrated understanding of science, engineering and business management. A CTM research group focuses on strategic considerations of intellectual property, its management and IP analytics within technology development projects and processes.

Centre for Technology Management, Institute for Manufacturing,
Department of Engineering, University of Cambridge,
17 Charles Babbage Road, Cambridge CB3 0FS, United Kingdom

+44 (0)1223 766141 | ifm-enquiries@eng.ca.ac.uk
www.ifm.eng.cam.ac.uk | Twitter@IfMCambridge

Design: www.cambridgedesignstudio.org

EPSRC
Engineering and Physical Sciences Research Council

CAMBRIDGE BIG DATA