OECD
BETTER POLICIES FOR BETTER LIVES

CSTP-TIP Workshop:

Semantic analysis for innovation policy

Paris, 12-13 March 2018

Summary

Introduction to the workshop and semantic analysis

The digital transformation is generating new tools for policy making, including for analyses in support of innovation policy to address many open questions. With software developments and increased computing capacities, the opportunity to systematically exploit textual information has grown enormously, as illustrated by the semantic analysis of the reports of the OECD TIP Working Party that was conducted in the context of the TIP@50 event. The potential of semantic analyses is also boosted by large amounts of data available for analysis, including patent applications, research publications, official policy papers and evaluations, administrative data held in government databases, company websites, social media and online platforms.

The objective of this workshop was to **shed further light on both current and future applications of semantic analysis for innovation policy analysis**, highlighting core methodological considerations and next steps required for semantic analysis to better inform innovation policy.

What is semantic analysis?

Semantic analysis is a **branch of Artificial Intelligence** which refers to the activity of **extracting meaning from large sets of text**. Whereas traditional techniques use mainly metadata (e.g. the country of residence of a scientist) or quantitative data (e.g. the number of authors of a scientific paper), semantic analysis processes the text itself - of a scientific or policy paper, of a patent, etc. On that basis, it analyses the topics and opinions the document carries and it can compare individual documents from the perspective of their content, identify trends in the issues addressed etc. Semantic analysis can be implemented by various techniques and the field is in fast development due to expanding use and very active research.

The following characteristics make semantic analysis a particularly interesting tool to support innovation policy making:

- **Systematic** – Semantic analysis allows applying a model systematically on a large volume of textual data.
- **Fast** – Semantic analysis allows processing large volumes of textual data at high speed.
- **Granular** – Semantic analysis allows performing fine-grained analyses considering multi-dimensional parameters (e.g. networks formed of individual researchers for research areas).
- **Up-to-date –** Semantic analysis allows exploring near current data about a specific issue of policy interest (e.g. from social media, company and government websites). This contrasts with traditional indicators based for example on annual surveys that require time to validate, process and make them publicly accessible.
- **Relational** – Semantic analysis allows identifying relations that go beyond traditional indicators (e.g. research paper co-authorship), for example recognising researchers that share similar research trajectories or the increase in collaborations among researchers in specific areas.
- **Transparent** – Semantic analysis, if well implemented, examine text data following defined parameters that allow third parties to reproduce the analysis, question the validity of the methods and suggest changes to improve the accuracy of results.

7 takeaways from the workshop

1. **Semantic analysis, a branch of Artificial Intelligence that focuses on extracting meaning from large sets of text, is a fast-growing field that offers huge potential for innovation policy**. It allows applying analytical models systematically on a large volume of textual data at a high speed, considering multi-dimensional parameters at the granular level. It enables the exploration of up-to-date data (e.g. from social media) about specific issues of policy interest, and can help shed light on interactions that go beyond traditional research collaboration indicators.

2. **Semantic analysis techniques allow leveraging new information from large sets of text to inform science, technology and innovation policy**. Data sources include traditional ones, such as policy documents, public R&D funding databases, and patent and publications data; as well as new sources, such as company and university websites with information on ongoing research activities, as well as social media or newspapers articles that capture the public or stakeholders' opinion on STI-related issues.

3. **Semantic analysis provides new insights to characterise innovation ecosystems and detecting technology developments and innovation trends** by helping to address questions such as: What are past and emerging research and technology fields? Who is doing what in innovation ecosystems? What are the topics and patterns of collaboration in a specific innovation ecosystem?

4. **Semantic analysis also supports answering questions regarding innovation policy itself.** This includes questions such as: How are innovation policy debates evolving over time? How do policy debates relate to academic debates? What innovation policies are in place and what are they supporting? What are the views of the public on specific innovation policy issues?

5. **The right implementation of semantic analysis is needed for the tool to inform policy, requiring the development of guidelines and best practice codebooks.** This applies to all stages of analysis, from data selection to data processing and presentation of results. Transparency and replicability of the models used is important. Guidelines and best practice codebooks can help support good quality semantic analyses.

6. **Search and visualisation tools allow leveraging the potential of semantic analysis for innovation policy**, by converting results into valuable information that can be easily used by policy makers. Future efforts are likely to focus on developing more intelligent, user-assisted navigation tools (e.g. diagnostic toolkits).

7. **The adoption of semantic analysis as a policy support tool requires intensive exchange of experiences.** Exchanges on good practices, challenges and past failures among policy makers experimenting with their use can help accelerate the successful adoption of semantic analysis in policy. Sharing where possible efforts, for instance, when it comes to building taxonomies in the field of science, technology and innovation can help reduce the costs and time investments.

Structure of the document

This workshop was organised jointly by the OECD Committee for Scientific and Technological Policy (Dominique Guellec, Michael Keenan, Andrés Barreneche) and the OECD Working Party for Innovation and Technology Policy (Caroline Paunov, Martin Borowiecki, Teru Koide, Diogo Machado, Sandra Planes-Satorra, Blandine Serve, Clement Sternberger).

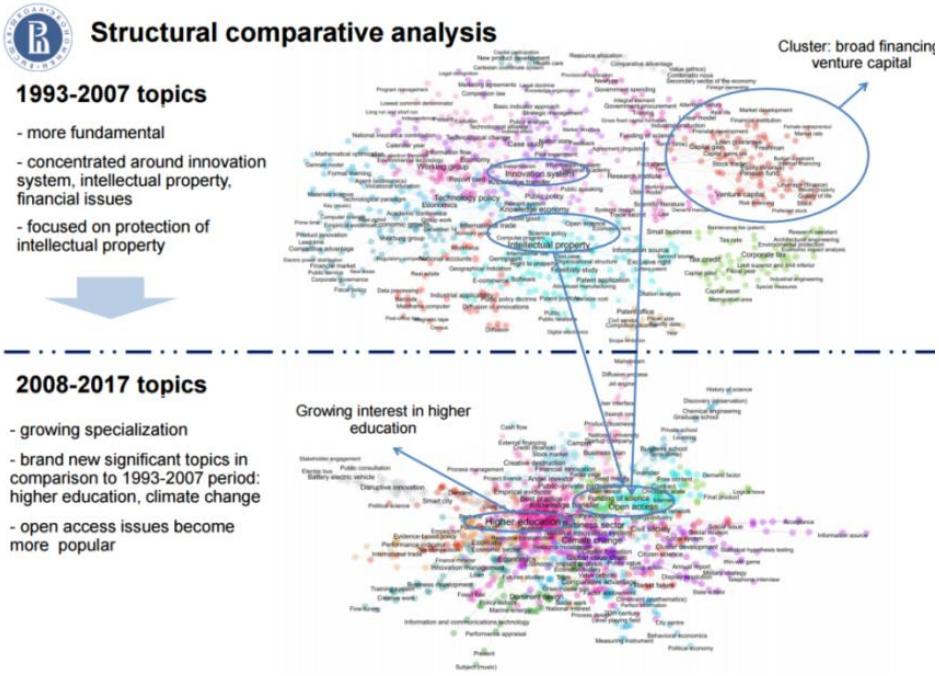
A. Application of semantic analysis for innovation policy

Semantic analysis enables the systematic exploitation of **textual information** relevant for science, technology and innovation (STI) policy. By doing so, semantic analysis can help provide answers to a range of **key questions for policy** that were nearly impossible to address in the past. Six of these key questions discussed during the workshop are presented below.

1. How are innovation policy debates evolving over time?

Semantic analysis can be used to **trace the evolution of policy debates over time**, as shown by a number of exercises conducted in the context of the 50th meeting of the OECD Working Party on Innovation and Technology Policy (TIP) – the TIP@50 Conference. Such exercises exploited past TIP documents and allowed addressing relevant questions, such as: *How has the focus of policy debates changed over time?* (Figure 1); *How do policy debates relate to academic debates?* Box 1 presents insights from those exercises.

Figure 1. Comparative analysis of TIP policy topics over time



Source: Presentation of Dirk Meissner and Ilya Kuzminov, available here.

Semantic analysis could also allow analysing the **time-lag between policy discussions in international fora** (e.g. the OECD) **and national policy debates** and new policy implementation, as suggested by **Byeongwon Park**, from the Korean Science and Technology Policy Institute (STEPI). While in the case of Korea time lags seem to be narrowing, he argued that these may depend on the complexity of the policy issue at hand and the specific country context.

Semantic analysis would also allow exploring what is **the impact of publicly funded research on future policy orientations**. For instance, De Mazière and Van Hulle[1] use semantic analysis to study the impact of EU funded research in Social Sciences and Humanities (SSH) on EU policies. By analysing the semantic similarity of research papers and policy documents, further insights on how public R&D funding impacts research outputs and how research outputs influence future policy orientations can be uncovered.

[1] Mazière, Patrick A. De and Marc M. Van Hulle. "A clustering study of a 7000 EU document inventory using MDS and SOM." *Expert Syst. Appl.* 38 (2011): 8835-8849

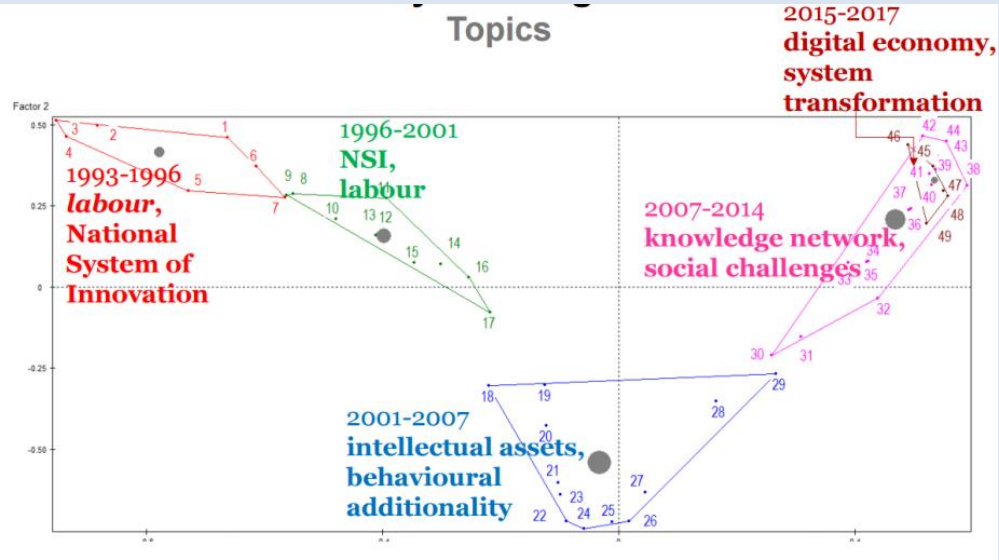
Box 1. TIP@50 event: Policy questions addressed using semantic analysis

In the context of the 50th meeting of the OECD Working Party on Innovation and Technology Policy (TIP) – the TIP@50 Conference – a number of semantic analysis exercises were conducted in order to exploit information from past TIP documents. These exercises address relevant innovation policy questions:

How have policy topics changed over time?

Based on the semantic analysis of 274 TIP reports, meeting agendas and minutes of plenary sessions and workshops between 1993 and 2017, **Margherita Russo** and **Pasquale Pavone**, from University of Modena and Reggio Emilia (Italy), found that since 2007 the TIP has focused on more specific areas, compared to the past. Main topics of work have also changed over the past 25 years, from a focus on national innovation systems and labour issues (1993-2001), to behavioural additionality and intellectual assets (2001-2007), knowledge networks and social challenges (2007-2014), and the digital and system transformation (since 2015) (Figure 2) (see their presentation here).

Figure 2. Main topics of plenary meetings, 1993-2017

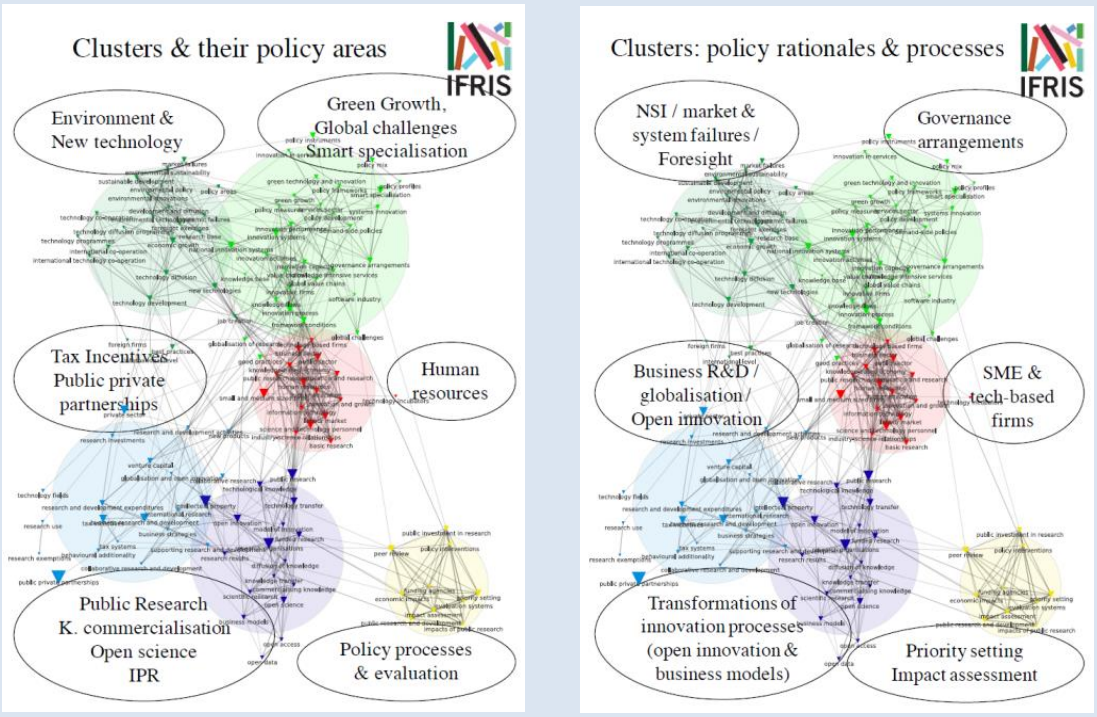


Source: Presentation of Margherita Russo and Pasquale Pavone, available here.

The semantic analysis of documents from past TIP conferences and workshops conducted by **Dirk Meissner** and **Ilya Kuzminov,** from the Higher School of Economics (Russian Federation), identifies common topics over time and finds increasing focus in recent years on cluster, spin-off and knowledge transfer policies, as well as venture capital and IT-related issues (see Figure 1 above).

Based on the semantic analysis of 330 TIP documents, which made use of the digital platform CORTEXT, **Philippe Larédo** (Université Paris-Est and Manchester University) and **Antoine Schoen** (ESIEE, Paris) identified six policy clusters and policy rationales that have varied in relevance over time (see their presentation here).

Figure 3. Semantic analysis of TIP documents: Policy clusters and policy rationales



Note: Based on the semantic analysis of 330 TIP documents using the digital platform CORTEXT
Source: Presentation of Philippe Larédo and Antoine Schoen, available here

How do TIP policy debates relate to academic debates?

Michael Keenan, Senior Policy Analyst at the OECD, presented the results of the comparison between the main topics of focus of the CSTP (based on 782 CSTP documents over 1994-2016) and the topics of research articles published in Research Policy, the leading academic journal on innovation policy studies (based on 2527 titles and abstracts over 1988-2017) (see their presentation here). Emerging topics over the past decade in CSTP include policy mix, global and societal challenges, tax incentives, impact evaluation, innovation strategy, and open access. Figure 4 presents the most frequent topics over the period.

Figure 4. 50 most frequent topics in CSTP reports, 1993-2017

Note: Based on the analysis of 782 full documents and the use of the IPP vocabulary of 1200 terms.
Source: Presentation of Michael Keenan, available here.

2. What are past and emerging research and technology trends?

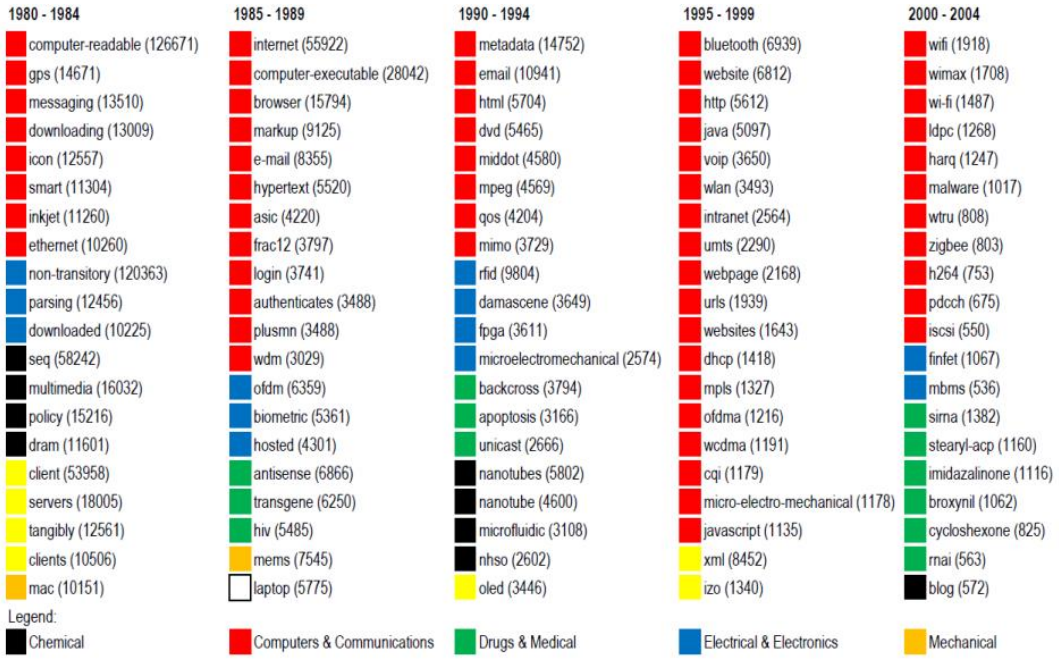
Semantic analysis can help identify emerging research and technology trends, allowing policy makers to adopt a forward-looking approach when designing new policies. Approaches to identify such trends using semantic analysis include the following:

- **Monitoring new collaborations among researchers in different fields** (i.e. interdisciplinary research). As explained by **Francesco Osborne**, from the UK Open University, an increase in the number of collaborations among researchers in specific fields can be a relevant indication of an emerging field of research or technology development.
- **Identifying new research-related terms in patents and academic publications. Sam Arts**, from KU Leuven, presented a method that uses semantic analysis to measure similarity among patents. This method requires selecting keywords in the documents. Measuring the number of occurrences of such words over time allows identifying new trends in scientific research. Figure 5 presents keywords at a given time and by technological category based on patent data. A similar analysis could be conducted using publications data.
- **Detecting the emergence of new clusters of terms in publications**. The Augur project, presented by **Francesco Osborne** from the UK Open University, follows three steps. First, using the corpus (scientific publications), terms that are identified as being commonly associated over time are clustered together. Second, the project monitors the occurrence of clusters of words in new publications. Finally, an algorithm detects the clusters that experienced the highest rise in utilisation in recent publications, which is used as a proxy to detect emerging technology trends.

Francesco Osborne

Sam Arts

Figure 5. Emergence and diffusion of new technologies



Note: Most reused novel words by five-year periods and coloured by NBER technology category, 1980.
Source: Presentation of Sam Arts, available here.

3. *Who is doing what in innovation ecosystems?*

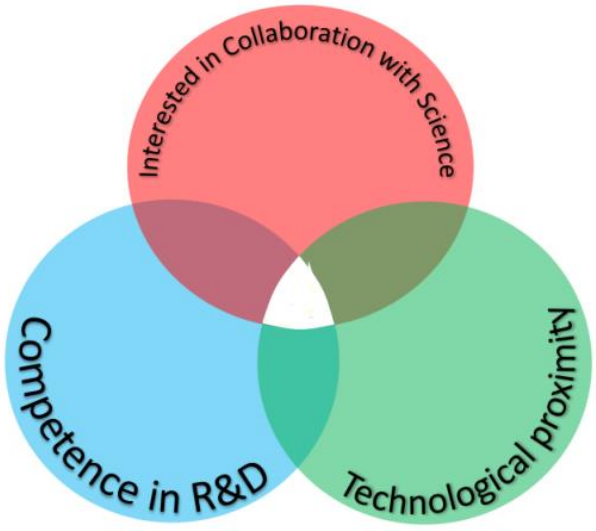

Antoine Schoen

Semantic analysis can be used to **produce indicators on a range of innovation-related issues that are relevant for policy**. This is the main objective of the KNOWMAC project, a web-based tool that provides indicators and interactive visualisations on knowledge co-creation. As explained by **Antoine Schoen**, professor at ESIEE, such indicators will allow answering questions such as: Which are the top European research actors on a certain research topic? Where are they located? In which topics is a given European region specialised?


Georg Licht

Universities searching for commercial applications of their technology developments can also benefit from semantic analysis. **Georg Licht** and his team at the Centre for European Economic Research (ZEW) are using patent data to explore the technological proximity between university patents and those of several firms, in order to **identify potential partners** capable of taking the new technologies to the market. Such analysis is complemented by firm-level data (e.g. from innovation surveys), data on publicly funded R&D projects, and geographical data in order to estimate the probability of those firms to collaborate with science (Figure 6). A similar approach could be used by universities to **identify possible academic research collaborations**, and by governments to identify what types of partnerships are worth supporting the most.

Figure 6. Three key factors defining the optimal partner



Note: Add the note here. If you do not need a note, please delete this line.
Source: Presentation of Georg Licht, available here

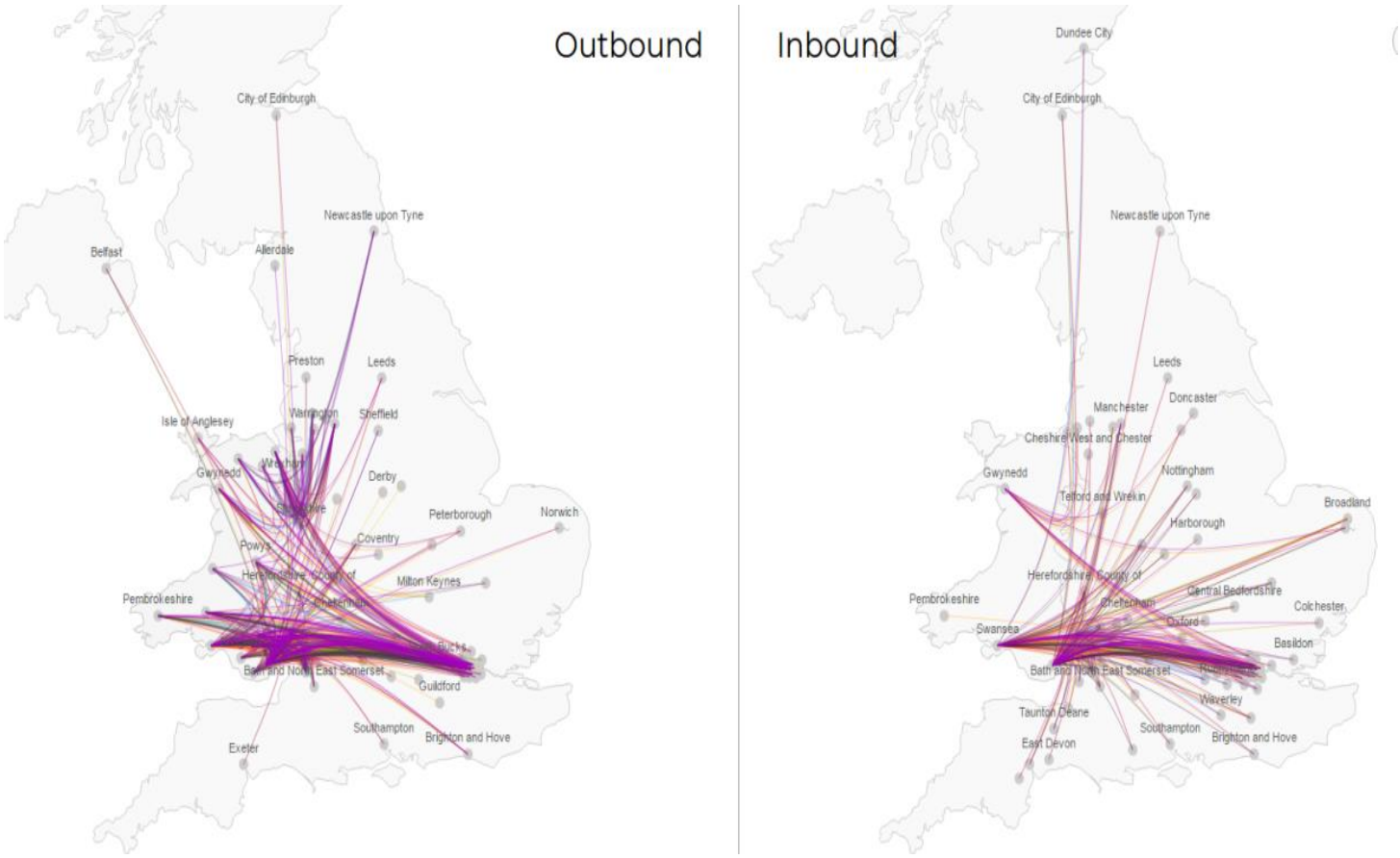
4. What are the networks in innovation ecosystems?

Going a step further from the identification of who is doing what in innovation ecosystems, semantic analysis also allows exploring the connections among such actors. **Juan Mateos-Garcia** and his team at NESTA have conducted a project (Arloesiadur) aimed at mapping innovation networks and research collaborations in Wales (see his presentation here). To identify innovation networks, they collected data through web-scrapping of company websites and a social media platform (Meetup) and geocoded such data.

Juan Mateos-Garcia

Figure 7 presents a mapping of those connections. To identify research collaborations, they have used data from Gateway to Research, an open dataset about UK publicly funded research and innovation projects. Research collaboration networks for different research fields can be explored on an interactive and visual platform (see here).

Figure 7. What are the connections between tech communities in Wales and other parts of the UK?



Note: This visualization shows the local authorities that tech Meetup users visited to attend events. The map on the left presents the probability that a user based in Wales visits another area, in or out of Wales, to attend a Meetup event. The map on the right shows the probability that users based elsewhere in the UK will attend a tech meetup event in Wales.
Source: Arloesiadur, available here , based on data from Meetup.com

5. What are innovation policies in place and what are they supporting?

Semantic analysis can also be used to explore **what are the R&D topics being funded by innovation programmes**. For instance, the Netherlands Enterprise Agency is applying a semantic analysis technique (topic modelling) to explore data on 1 122 projects funded by different EU innovation programmes (the European Fund for Regional Development, the SME Innovation programme, and the Public-private partnership programme) in order to better understand what R&D topics are these innovation programmes funding and whether they complement each other. This technique could also be used as a policy evaluation tool to better understand the impact of existing policy instruments, as it allows exploring whether financed projects are in line with initial policy goals (see **Maarten van Leeuwen**'s presentation here).


Marnix Surgeon


Frédérique Sachwald

Another application, as pointed out by **Marnix Surgeon**, from the European Commission, could be to use semantic analysis to assess **whether public R&D funds are allocated to support ground-breaking research**, by comparing information from documents on funding allocation with information from state-of-the-art science literature. Semantic analysis could also provide support for **assessing the impact of policy strategies**, as pointed out by **Frédérique Sachwald**, from HCERES in France.

6. What are the views of the public on innovation policy?

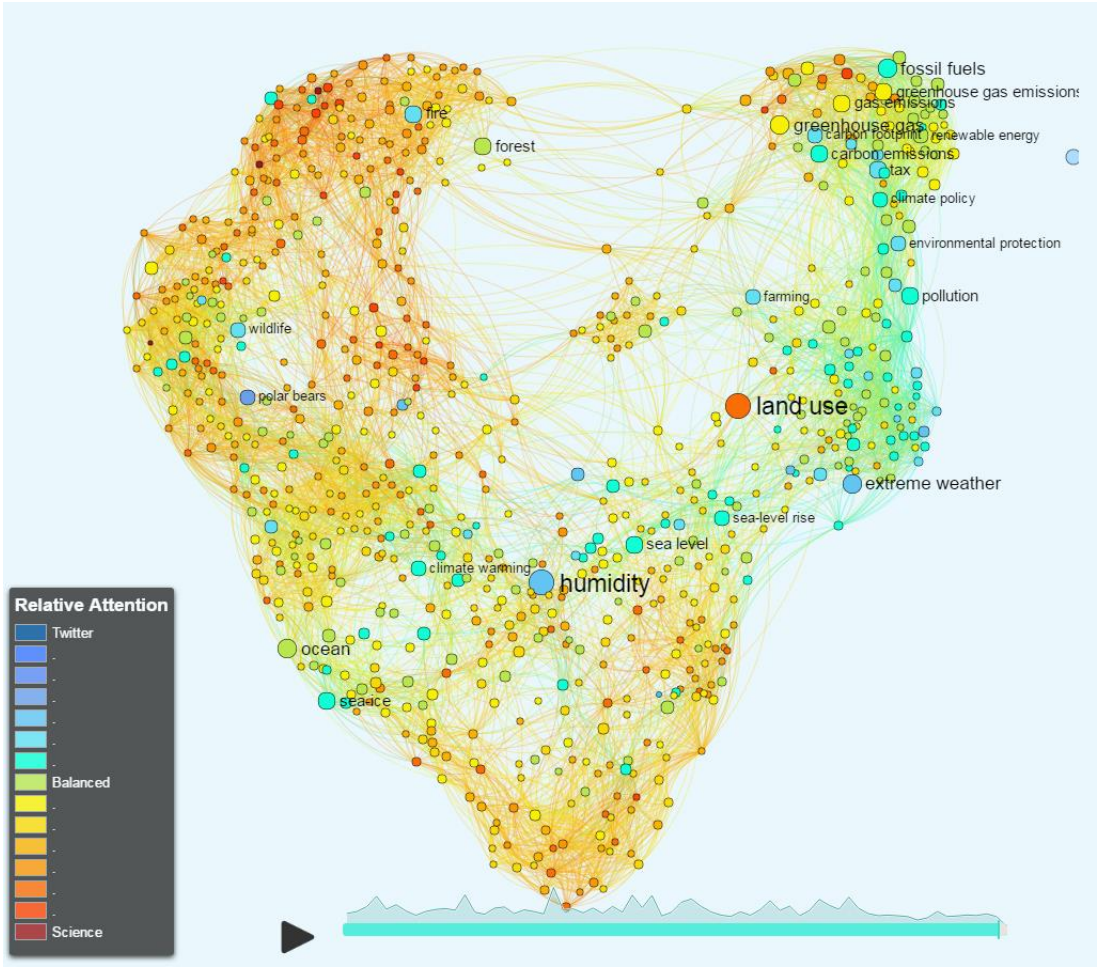
Semantic analysis opens opportunities for **conducting large-scale open public consultations** during processes of reform of regulatory or policy frameworks. As explained by **Marnix Surgeon**, such opportunities have already been exploited by the European Commission, with the inclusion of some open questions in their public consultation questionnaires. With several thousand responses of approximately 20 pages each, such information could not be systematically analysed by humans in a short period of time.

Semantic analysis also brings new opportunities for the **exploitation of data from social media** regarding the public opinions on specific science, technology and innovation topics. **David Chavalarias** and his team at the CNRS' Complex Systems Institute of Paris Île-de-France, have already developed several applications of these techniques in other areas. These are two examples:

- During the months preceding the 2017 French Presidential elections, the team used 'the Politoscope', an open platform, to analyse 60 million tweeter exchanges between 2.4 million users. During the campaign, they sorted tweets by political community of the author and topics of the tweet. Semantic analysis techniques were used to explore the evolution of the vocabulary used around various subjects during the campaign.
- The Climate Tweetoscope project collects tweets and scientific publications regarding specific climate change issues (e.g. coral reefs, biodiversity, greenhouse gas emissions, fire). Semantic analysis techniques allow measuring which topics are associated to each other, and identifying the topics that raise more interest in the scientific community and those that are more discussed by the public and media (Figure 8).


David Chavalarias

Figure 8. Comparing public interest and science focus on climate change issues



Source: Tweetoscope, available here

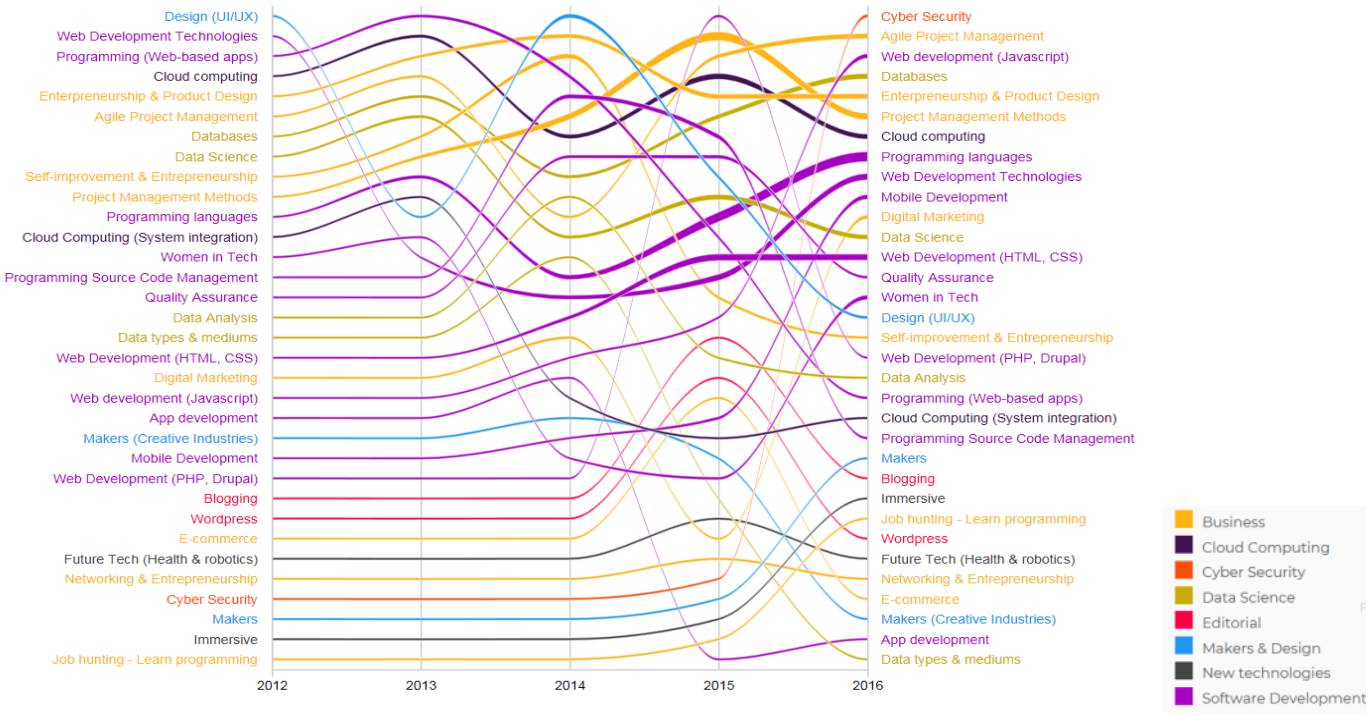
B. Leveraging new tools for innovation policy

A range of new tools allow leveraging the potential of semantic analysis for innovation policy, by converting the results of the analysis into valuable information that can be easily used by policy makers. Such tools are critical to fully leverage the potential of semantic analysis to provide more granulate detailed information on technologies, inventions and innovation policies. Rather than aggregating this information to synthesise the information as is usually done in policy reports, ways to allow navigating down to the granular level and up to the aggregate level can inform policy in important ways.

Search and visualisation tools present data in a visual and often interactive ways, such tools equip policy makers with rich inputs from semantic analyses exercises. Relevant examples of interactive search and visualisation tools include:

- ARLOESIADUR – A collaboration between Nesta and the Welsh government, this project has developed a platform that presents interactive data visualisations and open databases about industry, research and tech networks in Wales with the goal of informing innovation policies. As explained by **Juan Mateos**, visualisations answer a number of big questions, such as: What are research trends in Wales? Where do different industries concentrate in Wales? What are the connections between tech communities in Wales and other parts of the UK? What are tech networking trends in Wales? (Figure 9).

Figure 9. What are tech networking trends in Wales?

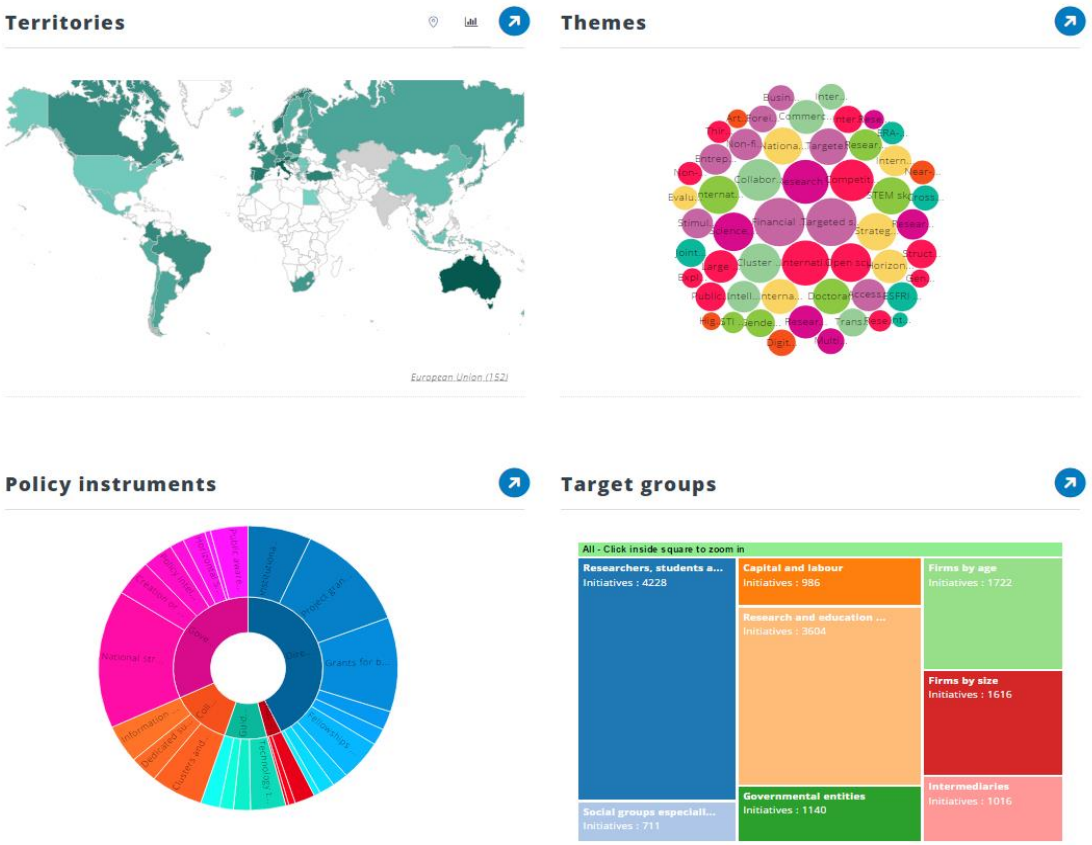


Source: Arloesiadur, available here , based on data from Meetup.com.
Note: The figure shows the levels of networking activity in different tech topics for all of Wales as well as its local economies. Topics are ranked with an index based on the number of attendants to events in that topic normalised by the UK average. If the index is big, this means that Wales has relatively higher levels of networking activity in that topic compared to the UK. The thicknesses of the lines represent the absolute number of attendants per event.

- STIP COMPASS – this initiative from the European Commission and the OECD collects quantitative and qualitative data regarding national innovation policies, gathered through the EC-OECD International Survey on Science, Technology and Innovation. As explained by **Andrés Barreneche** of the OECD, the STIP Compass incorporates more than 500 interactive dashboards and a sophisticated search tool with smart filtering (allowing to search by country, policy theme, policy instrument and target group) that facilitates policy discovery. These interfaces allow users to seamlessly query the database to identify country policies on a wide range of STI policy issues (Figure 10).

Andrés Barreneche

Figure 10. STI Policy Compass: content menu



Source: STIP Compass, available here.

- Climate Tweetoscope – this platform, presented above, provides visualisations of the evolution of main questions and debates on climate change issues. By identifying topics that receive more attention in social media (tweeter) and those that receive more in science, the platform allows to visually identify areas where there is higher public interest relative to the current research focus (e.g. extreme weather), and the reverse (e.g. land use receives high science focus and relatively low public interest) (see Figure 8 above).

Efforts are also currently devoted to developing more **intelligent, user-assisted navigation tools**. This could take the form of **diagnostic toolkits**, that would allow policy makers to assess, for instance, the innovation capacities and existing networks in a specific region, measure the impact of previous innovation policies and access information regarding policy tools implemented in other countries to address similar challenges.

C. Methodology

As is the case for any method, the right implementation of the tool is essential for it to build new evidence and inform policy. This applies to all stages of data selection, analysis and the use of the outcomes of the analysis.

Data preparation

An important step is identifying and preparing the data for policy analyses. Different data may be needed depending on the policy questions to be addressed. This concerns information about a) research and innovation itself - e.g. in patent and publications data- , about b) actors involved in innovation - e.g. on company and university websites (but also internal databases) -, about c) policies - e.g. in policy documents but also grant information/databases etc. as well as about d) public option on of STI among consumers e.g. in the press and on social media.

Data sources for semantic analysis include more **traditional sources** such as national policy reports to learn about policy trends and also patent application data and research articles that describe technology and research, supporting a better understanding of the progress of research and inventions. There are also **new sources of text information** on the internet, including Twitter to gather information on public opinions regarding technology or LinkedIn data to learn about the mobility of innovators. Company web sites may also prove useful to learn about their innovation activities. The use of such data is currently explored by Georg Licht and colleagues to complement information from the German company innovation survey.

As with traditional quantitative data, data quality needs to be assured to take into account the following:

Cinzia Daraio

- **The representativeness of the data.** This may be more challenging for data used from the web compared to using firm census information as not all firms may be active on the web or only some may be vocal on Twitter etc. Analyses may misrepresent the reality of firms by only capturing a small share of firms with distinct characteristics (those being active). As highlighted by **Cinzia Daraio**, from University of Rome La Sapienza, data are not always objective (they have been produced for a specific goal that needs to be taken into consideration e.g. company websites serve marketing purposes and may not describe the actual innovation capacity (but the image the company would like to give itself).
- **Potential biases in the text information**. Data may have been written with a specific purpose in mind (e.g. influencing public opinion) and consequently may not represent the reality that is being captured. Some topics may even be omitted altogether and could mislead analyses.
- **Text information is complex and context-dependent:** An analysis of text information needs to take this into account the complexity and often "dirtiness" of text data. The same word may have different meanings in different contexts (polysemy) while different words may have the same meaning (synonymy). Exploiting text information also requires substantial filtering of irrelevant information to specific analyses, e.g. spelling mistakes, punctuation, "stopwords" ("and" etc.), words which are generic and carry little specific meaning out of context ("system"). The meaning contained in text can also be entirely different depending on how words are combined: Current semantic techniques do not exploit syntax, they just analyse the list of words ("bag of words"), although there is active research for improving that. There is active research also for identifying semantic units beyond individual words ("n-grams").

Analysis and use

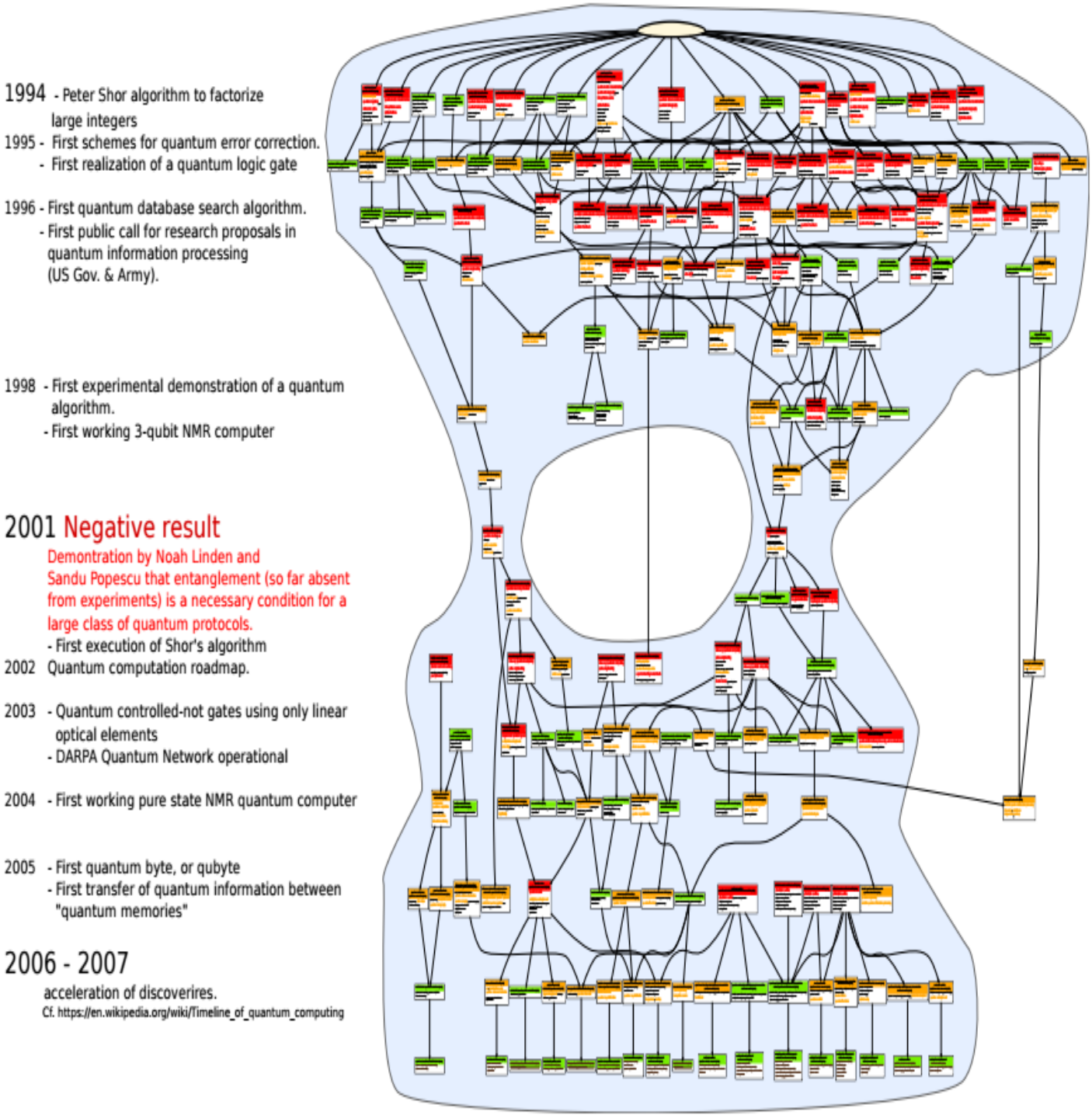
Various techniques are used for the semantic analysis of texts, the most common being Latent Semantic Analysis (LSA) and Topic Modelling. Both of them start with representing documents as vectors of words, which put together form a matrix. LSA is a heuristic approach which implements standard data analysis and linear algebra techniques in order to reduce the dimensionality of the matrix. That allows to compress the information contained in each document without changing the meaning, to calculate distances between documents, to identify clusters of documents, do recommendations ("you liked documents X and Y, then you might like document Z"), detecting outliers etc.

Topic modelling is a Bayesian technique which aims at identifying the "topics" that constitute each document in a corpus. A topic is defined by a set of words (with weights) which tend to co-occur in documents and are presumably semantically related (a same word can belong to several topics). The technique consists in 1) inferring the topics from the words; 2) identifying the topic structure of document. Like LSA, topic modelling can compress information, compare documents and identify clusters, but in addition it provides a direct interpretation of the results (in terms of topics). The technique allows tracing for a number of interesting analyses, including tracing the evolution of research fields such as quantum computing (Figure 11). This technique however relies much on human intervention and does not always have high robustness.

Importantly, semantic analysis has many complementary contributions to other tools and should not be seen as silver bullet that will respond to all questions. Rather, semantic analysis is one new tool for the toolbox available to inform innovation policy. Much benefit can be derived from combining semantics with other tools. Multi-dimensional analysis provides for exploratory techniques that support the identification of emerging topics or temporal categorization, independently from the experts' intervention and complementary to it. For instance, combining semantic analysis on progress in research across specific fields with econometric analysis can allow evaluating how policy affected such progress rather than simply tracing the evolution of progress.

Qualitative methods and expert consultations have an important role to play in informing about developments (aside from providing important validation of semantic analysis and other studies)..

Figure 11. Evolution of quantum computing



Source: Presentation of Chavalarias, available here.

D. Next milestones for semantic analysis for innovation policy

All workshop participants agreed to the importance of moving forward for semantic analysis to contribute to answering a number of important questions to contribute to innovation policy. While examples illustrate the huge potential, semantic analysis is not yet fully deployed for a number of challenges related to conducting better analyses and using new approaches more in policy.

Improvements in semantic analysis and its tools to better inform innovation policy are important along the following dimensions:

- **Address policy questions.** The evaluation will depend on how well semantic analysis delivers on what are questions policy needs an answer for. A focus of new tools on the core questions policy has will consequently be important. Close interactions with policy will help in that regard.
- **Build best practice codebooks/guidelines and promote transparency.** There are many ways of conducting analyses and choices made affect results. To allow for building the right standards to produce best results writing best practice codebooks are essential. It is also important to be being transparent about what analysis was conducted for the community of experts to comment and validate or improve on such practices, as emphasised by **Cinzia Daraio**, from University of Rome - La Sapienza.
- **Set up the infrastructure in collaboration.** The sharing of research and innovation data, and efforts in building taxonomies in the field of science, technology and innovation policy reduces the costs and time investments of conducting semantic analysis. As the quality of results depends on the quality of the infrastructure, large upfront investments cannot be cut short but should be combined as best possible. Creating shared depositories openly and building them collectively can help in a similar way that online libraries help users of open source software. There are several areas where efforts can be shared (e.g. disambiguation of names of institutions, organisations, programmes, technologies, etc.) to avoid unnecessary duplication of efforts, as pointed out by **Frédérique Sachwald**, from HCERES in France, and **Margherita Russo**, from the University of Modena and Reggio Emilia in Italy.
- **Build a community**: Community building that gathers domain specialists in the field of STI, policy makers and technical specialists (including computer scientists, statisticians, linguists) experienced in text analysis will help ensure policy questions are addressed in optimal ways by semantic analysis tools. Feedback processes are also essential to develop tools use by policy experts.
- **Data access to build interconnected systems**: There is most to gain from combining different data sources from different sources, providing much benefit efforts in improving interconnectivity. An example is combining the information gathered from firm surveys with information from those firms' websites is an example in point.

There is also an additional step regarding use of semantics in STI policy processes:

- **Establish trust.** The integrity of results yielded by semantic analysis is a pre-requisite, requiring common standards, more transparency in methodologies and source data quality assurance, as stressed by **Armin Mahr**, from Austria's Federal Ministry of Science, Research and Economy. Quality assurance mechanisms will help build trust among policy makers in results using those tools in different policy processes.

Margherita Russo

- **Incentivise the use of semantic analysis to inform STI policy.** Changing habits can be challenging and promoting best practice examples of applications. As pointed out by **Margherita Russo** flagship initiatives can help drawing more attention to the improvements semantic analysis can bring to policy analysis. She emphasised that the change in mindsets needed requires active change management and would not happen otherwise, even if tools are much improved.
- **Exchange actively experiences.** There is much experimentation under way at the moment aimed at applying semantic analysis to address policy questions. Exchanging on good experiences and potential failures including in the development of ontologies and tools can help avoid doing the same mistakes others did and be a means to accelerate progress.
- **Develop smart use cases of semantic tools**. As is the case for all tools, there are good and bad ways of using results from high quality semantic analysis and tools. This requires training.
- **Be realistic**. Exercises today are often large investments and very time-consuming, involving difficult validation and expert involvement is often critical. Realistic expectations of what can be delivered are important to avoid disappointment that may lead to not using those tools.

List of speakers

Sam Arts, Assistant Professor, KU Leuven, Belgium

Andrés Barreneche, Policy Analyst, OECD

David Chavalarias, Research Director at the National Centre for Scientific Research (CNRS) and Director of the Complex Systems Institute of Paris Ile-de-France (ISC-PIF), France

Cinzia Daraio, Associate Professor, University of Rome La Sapienza, Italy

Mary-Ann Grosset, Digital Practice Team Manager at OECD

Dominique Guellec, Head of Division, Directorate for Science, Technology and Innovation, OECD

András Hlács, Counsellor, Permanent Delegation of Hungary to the OECD

Michael Keenan, Senior Policy Analyst, Directorate for Science, Technology and Innovation, OECD

Joel Klinger, Data Developer, Innovation Mapping, NESTA

Philippe Larédo, Research Director at Université Paris-Est and Professor at Manchester University

Georg Licht, Head of Department of Economics of Innovation and Industrial Dynamics, Centre for European Economic Research (ZEW), Germany

Maarten van Leeuwen, Lecturer, Leiden University, the Netherlands

Armin Mahr, Head of STI locations and Regional Policies, Federal Ministry of Science, Research and Economy (BMWFW), Austria

Göran Marklund, Deputy Director General for External Matters, VINNOVA, Sweden, and Chair of the TIP Working Party

Juan Mateos, Head of Innovation Mapping, NESTA, United Kingdom

Dirk Meissner, Professor, National Research University - Higher School of Economics, Russian Federation

Francesco Osborne, Research Fellow, Knowledge Media Institute, The Open University, UK

Byeongwon Park, Research Fellow, Science and Technology Policy Institute (STEPI), Korea

Caroline Paunov, Senior Economist, Directorate for Science, Technology and Innovation, OECD

Pasquale Pavone, Researcher, Research Centre for the Analysis of Public Policies (CAPP), University of Modena and Reggio Emilia, Italy

Margherita Russo, Professor at University of Modena and Reggio Emilia, Italy

Frédérique Sachwald, Director, Science and Technology Observatory (OST), HCERES, France

Antoine Schoen, Professor, ESIEE Paris, France

Jan-Anno Schuur, Information Systems Specialist, OECD

Marnix Surgeon, Deputy Head of Unit, European Commission

Chantale Tippett, Principal Researcher, Innovation Mapping, NESTA

Workshop agenda

Monday, 12 March 2018

Semantic analysis of TIP documents & best practice in semantic analysis

Welcoming and introduction to the workshop

- **Dominique Guellec**, Head of Division, Directorate for Science, Technology and Innovation, OECD
- **Michael Keenan** (Senior Policy Analyst, OECD) and **Caroline Paunov** (Senior Economist, OECD)

Session 1: Potential, best practice and caveats in using semantic analysis

The session introduced the basics of semantic analysis (methods, software tools, principles and challenges) to establish a common ground for discussion between experts and non-experts alike. It focused notably on the following issues:

- What can be done using semantics analysis in the field of innovation policy? What are some illustrative examples?
- What are the strengths and caveats involved in such an analysis compared to more traditional types of analyses?
- What are the steps that need to be undertaken to undertake such an analysis?
- What applications of semantic analysis relevant to innovation policy can be implemented as of today? What applications need further development?

Chair: **Dominique Guellec**, Head of Division, Directorate for Science, Technology and Innovation, OECD

Roundtable discussion involving **Mary-Ann Grosset**, Digital Practice Team Manager, OECD; **Philippe Laredo**, Research Director at Université Paris-Est and Professor at Manchester University; **Juan Mateos**, Head of Innovation Mapping, NESTA, United Kingdom; and **Margherita Russo**, Professor, University of Modena and Reggio Emilia, Italy.

Session 2: In-depth exploration of the semantic analysis of TIP documents

The session allowed for an in-depth discussion of the semantic analysis that was conducted in support of the TIP@50 event and focused in particular on the following questions:

- How has innovation policy thinking changed in what is reflected in 25 years of TIP, CSTP and national policy discussions?
- What trends can be identified over the 25-year period for the TIP, CSTP and beyond?
- What lessons can be taken from the work for the future of the TIP, the CSTP and innovation policy analysis more generally?

Chair: **Caroline Paunov**, Senior Economist, Directorate for Science, Technology and Innovation, OECD

Speakers:

- **Michael Keenan**, Senior Policy Analyst, Directorate for Science, Technology and Innovation, OECD
- **Margherita Russo**, Professor, University of Modena and Reggio Emilia, Italy; and **Pasquale Pavone**, Researcher, Research Centre for the Analysis of Public Policies (CAPP), University of Modena and Reggio Emilia, Italy
- **Dirk Meissner**, Professor, National Research University - Higher School of Economics, Russian Federation
- **Philippe Laredo**, Research Director at Université Paris-Est and Professor at Manchester University; and **Antoine Schoen**, Professor, ESIEE Paris, France
- **Byeongwon Park**, Research Fellow, Science and Technology Policy Institute (STEPI), Korea

Session 3: The making of the TIP@50 analyses

The session focused on what lay behind producing the TIP@50 analyses and offered interested participants an opportunity to do a simple analysis:

- What material was used and how did choices on how the material was organised and analysed affect the analysis?
- What software decisions were made? What are the costs involved in using such software? What is freely available and what is proprietary?
- How can these analyses be applied to other types of analysis of innovation policy?

Chair: **Michael Keenan**, Senior Policy Analyst, Directorate for Science, Technology and Innovation, OECD

Speakers:

- **Jan-Anno Schuur,** Information Systems Specialist, OECD; **Andrés Barreneche**, Policy Analyst, OECD
- **Margherita Russo**, Professor, University of Modena and Reggio Emilia, Italy; and **Pasquale Pavone**, Researcher, Research Centre for the Analysis of Public Policies (CAPP), University of Modena and Reggio Emilia, Italy
- **Dirk Meissner**, Professor, National Research University - Higher School of Economics, Russian Federation
- **Philippe Laredo**, Research Director at Université Paris-Est and Professor at Manchester University; and **Antoine Schoen**, Professor, ESIEE Paris, France

Session 4: Perspectives on semantic analysis

- **David Chavalarias**, Research Director at the National Centre for Scientific Research (CNRS) and Director of the Complex Systems Institute of Paris Ile-de-France (ISC-PIF), France: Reconstruction and monitoring of the scientific debates from digital traces: the cases of science and politics

Session 5: Hands-on exercise

The session allowed participants to produce a simple semantic analysis with the support of those involved in the TIP@50 analysis.

- **Antoine Schoen** : Replicating results from the TIP@50 analysis using Cortext

Tuesday, 13 March 2018

Semantic analysis to investigate knowledge transfer and innovation policy

Brief summary of main conclusions of 12 March regarding semantic analysis

Session 6a: Examples of semantic analysis conducted to study knowledge transfer

This session presented different national initiatives that use semantic analysis to shed light on innovation policy and in particular on knowledge transfer between industry and science.

- What new questions can be addressed using semantic analysis?
- What information are they semantically analysing and how is this accessed?
- What infrastructure requirements were needed to set them up?
- What new answers have these studies identified, with a special focus on knowledge transfer?

Chair: **Frédérique Sachwald**, Director, Science and Technology Observatory (OST), HCERES, France

Speakers:

- **Antoine Schoen**, Professor, ESIEE Paris, France : KNOWMAK
- **Sam Arts**, Assistant Professor, KU Leuven, Belgium
- **Cinzia Daraio,** Associate Professor, University of Rome La Sapienza, Italy
- **Francesco Osborne**, Research Fellow, Knowledge Media Institute, The Open University, UK

Session 6b: Examples of semantic analysis conducted to study knowledge transfer

Chair: **András Hlács**, Counsellor, Permanent Delegation of Hungary to the OECD

Speakers:

- **Andrés Barreneche**, Policy Analyst, Directorate for Science, Technology and Innovation, OECD
- **Juan Mateos**, Head of Innovation Mapping, NESTA, United Kingdom Mapping and strengthening research and innovation networks with open data in Arloesiadur
- **Maarten van Leeuwen**, Lecturer, Leiden University, the Netherlands [by WebEx]
- **Georg Licht**, Head of Department of Economics of Innovation and Industrial Dynamics, Centre for European Economic Research (ZEW), Germany

Session 7: Hands-on experience (parallel sessions)

Session A: Using semantic visualisation tools to analyse data from the 2017 EC/OECD STI Policy Survey

Presenter: **Andrés Barreneche,** Policy Analyst, OECD

Description: The OECD Secretariat has recently revised the data collection methodology of the STI Policy Survey, run jointly with the European Commission. Data provided by countries is now more firmly structured on taxonomies-ontologies, which not only improves the comparability of responses but also facilitates analysis. Visualisation tools, accessed through a web interface, take advantage of semantic structures to aggregate data and present insights on over 6000 policies initiatives submitted by over 50 countries. This hands-on exercise provided participants the opportunity to get familiarised with these tools and learn how they can readily use the database for their own purposes

Session B: From text to impact in ninety minutes

Presenters: **Juan Mateos**, Head of Innovation Mapping, NESTA; **Chantale Tippett**, Principal Researcher, Innovation Mapping, NESTA; **Joel Klinger**, Data Developer, Innovation Mapping, NESTA

Description: This session gave participants an opportunity to turn their text into insight by walking them through key phases of the project pipeline. Starting with an overview of how semantic analyses can be used to explore questions of interest in the domain of knowledge transfer, the session was then broken down into a series of practical exercises covering data collection, analysis and outputs. These consisted of a combination of presentations drawn from Nesta's vast experience and small group exercises. The practical component concluded with a short session on the practicalities of semantic analyses, allowing participants to reflect on how they might implement projects in their own institutional context. The session concluded with a Q&A on the semantic analysis pipeline and a recap of key points

Session C: Supporting research policy makers with semantic technologies

Presenter: **Francesco Osborne**, Research Fellow, Knowledge Media Institute, The Open University, UK

Description: The number of papers and the available scientific knowledge is growing rapidly, making it harder to keep track of all the relevant knowledge that could inform research policy makers. In this workshop, we discussed how semantic technologies, that are being increasingly used to represent and analyse research data, can help to tackle this issue. We addressed some novel solutions for predicting trends, collecting research materials, analysing the research landscape, and tracking the evolution of technologies. The workshop also included a hand-on session in which the attendees were given access to the demos of some of these prototypical systems. More information about the technologies that were showcased at the workshop are available at http://skm.kmi.open.ac.uk/#projects

Session 8: Concluding panel and next steps

- What do you think is the potential and feasibility for semantic analysis in your own context?
- What could be useful next steps to facilitate exchange on best practice?
- How can CSTP and TIP provide support to help exploit the potential of these tools?
- What other tools should be part of the basket to consider?

Chair: **Göran Marklund**, Deputy Director General for External Matters, VINNOVA, Sweden, and Chair of the TIP Working Party

Speakers:

- **Marnix Surgeon**, Deputy Head of Unit, European Commission
- **Frédérique Sachwald**, Director, Science and Technology Observatory (OST), HCERES, France
- **Armin Mahr**, Head of STI locations and Regional Policies, Federal Ministry of Science, Research and Economy (BMWFW), Austria
- **Cinzia Daraio**, Associate Professor, University of Rome La Sapienza, Italy

Wrap up

Michael Keenan and **Caroline Paunov**, OECD

Workshop Website:
www.innovationpolicyplatform.org/semantics

OECD Working Party on Innovation and Technology Policy:
www.innovationpolicyplatform.org/cstp/tip

EC-OECD STIP Monitoring and Analysis:
www.innovationpolicyplatform.org/reiter

OECD Digital Science and Innovation Policy and Governance:
oe.cd/DSIP