18. New approaches for data in researching Science, Technology and Innovation (STI)

Tzameret H. Rubin,
Francisco Trincado-Munoz and Tim Vorley

18.1 INTRODUCTION

Although data are an essential component of innovation studies, and in particular, innovation policy, the importance of data is usually underestimated. However, in our experience as researchers, finding the fit-for-purpose data, granting access and manipulating data are essential steps to produce accurate policy advice with significant impact. Indeed, recent developments in data analytics, open data and data-driven technologies are opening up new opportunities. Increasing the availability of digitalisation and datasets can contribute to uncovering more diverse areas of policy activity, increase potential impact and become a foundation for more creative outputs and strategic design. Still, deciding what data to use, where to harvest these from and understanding its opportunities and limitations remain significant challenges for innovation studies scholars.

Throughout the years, the data available to use and capture innovation has transformed. The historical prevalence of manufacturing businesses led innovation studies researchers to use data inputs, such as the number of Science and Technology (S&T) engineers and R&D expenditure as main data sources. However, over time, consumers have become more demanding by changing their preferences towards more bespoke products and services, and businesses have adapted their strategies by putting a higher emphasis on the service rather than the products (Frank et al., 2019). Such transformation has manifested in the firms' business strategies; specifically, by pushing for more innovative ways to create value and alternative business models. In this context, the rapid advance of technologies (see van Meeteren et al., 2022) has urged strategy advisors in the public and private sectors, and those who study innovation, to reconsider how innovation emerges and what are the most effective ways to capture innovation.

Due to the rapid changes in industries, the boundaries between the manufacturing and the service sector have become blurred, and technological change can take directions that are based on digital transformation, which can generate changes in the technological trajectory. This evolution of industries challenges the academic community to find the right proxies to map innovation. More so, the sudden development of technologies, and in particular, the role of Artificial Intelligence (AI) in the economy, has expanded the limits of the data that scientists and researchers can access. Now, the options available for innovation and innovation research are broad in scope and nature, presenting significant opportunities but also limitations. Researchers can expand on traditional alternatives such as patent data to more revolutionary options such as big data, generated by the digital footprint of services, and user-digital footprint datasets.

In this chapter, we reconsider new types of data for Science, Technology and Innovation (STI) research and discuss six different data alternatives, reviewing their opportunities and limitations to capture innovation. The main aim of this chapter is to provide innovation studies scholars with an account of the data options and discuss their opportunities and limitations, from traditional data sources to more novel and challenging alternatives that have emerged throughout the digital revolution. Before doing this, however, we will briefly discuss the new context of innovation, including the new service economy and rapid technological change. This is followed by a discussion of the main data alternatives for innovation research, and finally, we conclude with a discussion on future trends and offer some recommendations on how researchers can reshape the way innovation is captured.

18.2 NEW SERVICE ECONOMY AND TECHNOLOGICAL CHANGE

As the service economy expands, information has acquired more relevance than tangible products. The focus on services, together with amplified digitalisation, has made advancements in information and communication technology (ICT) the cornerstone of innovation. Firms are presented with increasing opportunities for process and product/service innovation. For these firms, digitalisation means not only new forms of work, but it also influences how they create and capture value, which is mainly created from the intellectual work of employees, often acting as frontline workers in cooperation with clients.

Indeed, the digital transformation has shifted industries into service-oriented business models, becoming the new standard for economic growth. Yet, such transformation has not only impacted the service industries, but society as a whole. Technologies, like AI, Machine Learning (ML), Natural Language Processing (NLP), Robotics, Big Data, have become a cornerstone to accelerate and systematise innovation in firms and a driver to satisfy increasing and more complex needs from stakeholders.

The new panorama in the economy, dominated by new services and exponential technological change, in particular digitalisation, created challenges that are important to consider when researchers attempt to develop insights and capture innovation for the following reasons:

1. Exponentially rapid technological changes that transform industries, either through training and skills; business model innovation derived by colliding non-specific industries, and changes from developing products to services, or a mix of the two.
2. Predefined categorisations (for example patent classifications or static samples) present limitations in considering the rapid changes in 'real world' technologies.
3. Traditional statistical agencies struggle with the increasing cost of data collection as response rates are decreasing and costs are rising, urging researchers to reconsider data alternatives beyond the traditional sources.
4. New technologies have encouraged the emergence of new data companies that use web scraping, leveraging Internet of Things (IoT) data, among a range of sources. These new alternatives require careful consideration in order to allow the research community to understand their potential and inherent risks.

The following section will describe different data alternatives to study and capture STI, ranging from scientometrics (a more traditional source) to Big Data (a newer alternative). Each data type will be presented with its opportunities and limitations. Table 18.1 summarises the opportunities and limitations.

18.3 CURRENT DEVELOPMENTS IN DATA ALTERNATIVES IN THE STUDY OF SCIENCE, TECHNOLOGY AND INNOVATION

18.3.1 Patent Data

Patents are recognised as a very rich source for studying innovation as they contain a wealth of information. A patent is a document that grants legal ownership over an invention to a subject and is issued by an authorised governmental agency, permitting the right to exclude anyone else from the production or use of the invention for a stated number of years (Griliches, 1991). Patents have

Table 18.1 Opportunities and limitations of different data types in innovation studies

Data type	Opportunities	Limitations
Patents	• Allow for tracking changes. • Present a high variety of information. • Easy to access at an almost null cost. • Broadly used in innovation studies.	• Do not track practical utility. • Based on static codes that are manually changed. • Not all ideas/innovations can be patentable. • Not all organisations have incentives to patent their innovations.
Bibliometrics	• Cover broader fields. • Offer rich data. • Easy to access at an almost null cost. • Permit cross-disciplinary connections. • Can be analysed at different micro/macro levels. • Allow for tracking research trends.	• Bounded by the publication outlets. • Researchers need to adapt or operationalise the data by proxying constructs. • Citations can be biased and may lead to overestimating influential factors.
Commercialisation and government R&D funding	• May reveal patterns and information beyond the output of innovation. • May be connected to economic impact. • Value-added granularity. • Possible to capture innovation at different stages (for example, evolution).	• Difficult to access (bureaucratic processes). • Privacy concerns, bridging anonymity. • May require significant efforts to operationalise. • Potential sample bias towards large businesses. • Represent a small portion of the knowledge economy as they don't capture Business Expenditure on R&D (BERD).

Data type	Opportunities	Limitations
Labour and skills data	• Great granularity to focus on skills and knowledge. • Allow for tracking changes in the labour market associated with innovation. • May contain near-real time data. • Include multiple fields. • May be complemented with other labour and economic statistics.	• May be digitally biased. • May lack representativeness for certain occupations. • The direct connection with innovation can be more complex.
Real-time businesses' data	• Present data from multiple actors. • Near-real-time, dynamic data. • Great granularity. • Capture innovation in a broad range of organisations. • International and cross-industry data in one source.	• Lack of transparency in the data harvesting processes. • Expensive to access. • Geographical constraints depending on data suppliers. • Lack of flexibility in the data structures. • Potential selection biases.
User-generated and big data	• Large amounts of data. • Important range and variability of information. • Relatively easy to access with minimum costs. • May cover different geographies and sectors.	• Require significant knowledge and resources for their analysis and processing. • Data can be overwhelming, including unnecessary information. • Privacy concerns for the information shared. • May rapidly change, creating variability between extraction points.

been used to measure innovation because they contain a 'documented trail' based on standard classifications or classes (IPC edition 2023.01 consists of eight sections, 131 classes, 651 subclasses, 7,590 main groups and 78,378 subgroups) that permit changes to be tracked. Information such as patent numbers, document types, titles, inventors, International Patent Classifications (IPC), application dates, technical specifications, references to previous patents ('patent citation') and patent lifecycle details can be useful to capture and analyse innovation. Patents are also relatively easy to access, at an almost null cost.

Throughout the years, patents have been used in innovation research, and several examples can illustrate this. Trajtenberg (1990) was the first to develop a way to evaluate potential knowledge diffusion and the level of significance of an invention by counting the number of patent citations for Computed Tomography technology. Sakakibara and Branstetter (2003) looked at the total number of claims in the patent documentation. They concluded that the higher the number of claims per patent, the more novel it is, since it relies on a broader scope of other patents. More recently, the use of new methodologies borrowed from other research fields, such as Network Analysis or Natural Language Processing (NLP), has allowed the conversion of patent data into new types of input by extracting specific 'key-words', 'tags' or 'labels' to identify the inventions. These new methods allow for a 'breakthrough' from the traditional structural citation methods and turn the original patent data into other significant information that nurtures innovation research.

Patents, however, also present some limitations to capturing innovation. Although patents can provide extensive information regarding the inventions, their documentation fails to track their practical utility. In practice, this means that the 'real' economic value of the innovation is hard to determine, and therefore, the real impact of technological progress will not be able to be proxied by the patents alone (they will require connections to economic activities such as exports, productivity and labour). Second, patent data and industry classification are based on static codes (for example, IPC). This means that some significant aspects to characterise innovation can be omitted because of the static classifications. For example, the IPC for the biotech industry was last updated in its 8th Edition (OECD, 2009), which means that the new disruptive developments of AI in biotech are less likely to be accurately reflected by these classifications.

Furthermore, the new dominance of services and the constant pace of technological progress mean that not all ideas are patentable, including the failure to capture R&D activities or the adoption of new technologies that lead to innovative products or processes. This makes patents an indicator that does not necessarily capture all methods of creating knowledge. In general, patents present a lag in tracking innovation because the granting process in most countries takes around 18 months. This means that, in fast-moving innovative sectors, patents are not entirely useful to provide up-to-date evidence of technological change. Finally, technological companies lack urgency and incentives to apply for patents. This means that although, in some sectors, there is a common agreement that 'first to invent' gets priority over 'first to file'. This can be exemplified by software companies, where patents are not necessarily the common way to create a competitive advantage (Levin, 2004). In today's business environment, securing a growing initial customer base, or fast-revenue generation model, is more likely to give a company a competitive advantage rather than a patent. Technologies (such as software) are intangible assets, which can easily be copied, making infringement more challenging or not worthy of the resources to reinforce it. Some of the most popular tech companies have prospered by promoting open codes and developments, which can be used by others to innovate.

18.3.2 Bibliometrics Data

Similar to patent data, bibliometrics have a pre-structured taxonomy. However, they use and analyse scientific publications and draw upon two elements of information: the publications recorded in bibliographic repositories and the citations garnered. Bibliometrics use quantitative techniques, such as co-citations, bibliographic coupling, co-word analysis, and co-authorship to capture research impact and influence in specific topics, technologies or methods. Publications are, in general, highly accessible and contain rich data, which can then be processed through inexpensive and fast procedures. Yet, analysing such data might require extensive preparation and standardised methods. Bibliometrics have been central to the investigation of the role of technologies such as AI/ML, and design-rich STI frameworks that help to crystallise current thinking in different fields.

In recent years, bibliometrics have grown in business innovation research. This can be explained by the cross-disciplinary influence of bibliometric methodologies and the combination with newly available text-mining techniques. Scientific data have also become more accessible through platforms such as Scopus and Web of Science, which also permit the extraction of vast amounts of scientific publication data. The availability of data combined with scalable methodologies makes this data easy to analyse at different levels (for example individual, institutional, national or international level).

Although bibliometrics benefit from the wealth of data and the availability of software tools, there are some limitations. Publications can proxy only part of the knowledge produced, which is tailored to specific guidelines (presented by publishers) and differs in their intended purpose. The repositories do not necessarily include the original data, and what is available is at the discretion of the researchers. If researchers intend to use citations, these do not necessarily provide enough information as existing data are constrained by the structure of papers and citation conventions.

Another consideration relates to the availability of publications and the focus on most-cited works. While highly co-cited works can be regarded as more influential and can provide evidence of trends and research patterns' change, it is important to consider that a citation does not always arise for the right reasons, that is, an article can be cited a lot but for negative reasons. Dismissing information beyond the citation can lead to overestimating influential factors in capturing a phenomenon or the advance in a particular field. Similarly, only focusing on cited or published work ignores important data that might be more difficult to access because they were not cited, found in academic repositories that are less popular or perceived as not being of sufficient quality.

18.3.3 Commercialisation, and Government R&D Funding Data

Another significant way to capture innovation is through knowledge commercialisation of Research and Technology Organisations (RTO), including universities and government R&D funding data, both of which relate to public-funded programmes or publicly funded institutions that foster innovation. These sources of data, in contrast to patents, are more related to the exploitation or further advancement of innovation. Commercialisation data are based on the transformation of theoretical knowledge derived from research pursuits into tangible marketable activities, while government R&D funding data are mainly captured through the process of obtaining resources from government agencies. Both types of data can reveal patterns in the process of innovation that firms pursue beyond the outputs of innovation.

A significant source of innovation through knowledge commercialisation data can be found in knowledge exchange initiatives. Knowledge exchange is mutual fertilisation between RTO and the external stakeholders, in particular the business sector. It includes the exchange of resources and intellectual property, expertise and fundamental resources that propel innovation. Through this collaboration, RTOs can go beyond their primary focus on the exploration of new ideas and fundamental knowledge and tie these to businesses' practical considerations and the application of this knowledge, ultimately gaining profits from it. The whole knowledge transfer process and its economic impact (for example, revenues) can be captured, constituting an important source of data to understand innovation.

Government R&D funding data also relate to public expenditure on innovation. As funding is a necessary activity to facilitate innovation, firms engage in different processes to obtain this, including the application for external funds. For example, government agencies regularly subsidise research and development in new ventures (for example The Small Business Innovation Research

– SBIR – and Small Business Technology Transfer – STTR in the US, and UKRI in the UK). In such applications, firms may be required to disclose important aspects of their innovation process to compete for funding. Such data present good opportunities for researchers as they provide them with significant granularity and the possibility to link outcomes to processes (Pirog, 2014). For example, a researcher may use funding data to connect information between applicants' characteristics and the funding beneficiaries' organisational traits. Furthermore, as the application process requires applicant firms to provide a lot of firm-level details, it creates a source of rich and varied data, such as nuances in the firm's R&D stages.

Both commercialisation data and government R&D funding data have some limitations. The first and most important is the access to these data sources. Because of data privacy and bureaucratic processes, obtaining this data may require significant time and effort. Second, both sources are strongly biased towards large firms that are more likely to engage with public funding; hence SMEs that constitute the lion's share of the economy are less likely to be represented. Third, government R&D funding data covers only a small fraction of the Gross Domestic Expenditure on R&D, as it is very much private sector-led. Therefore, this source is more suitable for capturing fringe R&D fields which would not have obtained resources from the private sector and would be dependent on government investment. Hence, it is more likely to cover the 'R' in the R&D expenditure. Fourth, aspects such as the quality of the relationships in knowledge transfer partnerships, or the exact details to understand the process of research in a firm before a grant application, are not always straightforward to capture (hard to operationalise). In some cases, to measure significant variables, researchers might need to use proxies or rely on incomplete information. Finally, the processes of knowledge transfer and funding application might also involve interdependencies with other actors and asymmetries between applicants/funders or providers/recipients of knowledge, such that the relationship is unlikely to be linear and might happen through stages instead (Wehn and Montalvo, 2018). Therefore, measuring some effects might become a methodological challenge.

18.3.4 Labour and Skills Data

Another alternative to comprehend how knowledge is created and adopted, and in turn, how innovation emerges, is to look at the labour market. Labour data is key in technological change because it influences digital skills and allows a widespread adoption of new technologies. Looking at the labour market requires going beyond its relationship with economic growth and productivity, and further towards considering the digital footprint of job demand-related factors.

In this sense, the traditional sources for studying labour economics, and its link to knowledge generation and increase in productivity, are the Labour Force Surveys. However, as job advertisements are predominantly published online nowadays, it is possible to track their digital footprint to generate key labour data. For instance, Burning Glass Technologies (BGT) collects job advertisement data using web crawling techniques that browse online job boards and other websites. Hershbein and Kahn (2018) used this type of data to look at the transformation of the labour market with respect to technological skills. Using BGT allowed them to systematically track changes from unstructured data into structured data, which can then be categorised into specific predefined variables.

This type of data also has the advantage of producing near real-time information about job advertisements, such as trending skills and their geographic distribution. Furthermore, it can harvest multiple fields, for example location, industry, occupation, skill requirements, firm identifier and salaries, years of experience requested, education level required or preferred by the employer and so on. It also provides deeper granularity that allows for the tracking of changes over time (for example, by checking how the skill sets in job advertisements change).

Some limitations of this type of labour and skills data relate to their ability to represent the real-world demand and supply for labour in the economy. In principle, the data are digitally biased, that is, they only contain positions advertised online. This can affect representativeness in certain occupations. Furthermore, digital biases can lead to disproportionate results, especially when aggregating skills and industries or focusing on certain occupations. Finally, the connection with innovation or new ideas can be complex to achieve just by using this data source. This means that this data source could benefit from integration with other census sources, as argued by Hershbein et al. (2018).

18.3.5 Real-Time Business Data

Similar to BGT data, near real-time business data is a new alternative that allows capturing innovation. Businesses' big-data are generated (and later verified and validated) by harvesting public information, using web crawlers, AI and text-mining techniques, connecting through third parties' data companies' API, and recently even partnering with government ecosystem platforms. This process generates data on multiple actors (for example start-ups, scale-ups, investors, corporations, and their characteristics). The attractiveness of these data has facilitated the rapid emergence of business-to-business (B2B) data broker companies (for example Dealroom.co, Crunchbase, PitchBook, Beauhurst), which connect through APIs with specialised data companies (for example Owler, IPqwery, SimilarWeb, BuiltWith and so on), which create these large, dynamic data sets that offer better granularity and intelligence for the business analytics market.

In this respect, the volume of academic research using this type of data has also grown, with scholars focusing on insights regarding knowledge-intensive companies and their innovation financing. The main advantage of this type of data is that it has the possibility to capture innovation beyond what is patented, due to its wider coverage of SMEs that do not formally register innovation, and yet are in the constant process of adopting technology and improving their business models (Ménière et al., 2017). The rapid development of technology and increasing digitisation have also eased the capture of international and cross-industry data, offering more insights about service-based industries, which are usually more difficult to capture through other alternative data (van Meeteren et al., 2022).

In spite of the advantages of this type of data, one of the main limitations is the lack of transparency in the processes of data scraping and generation. As data companies profit from selling their data, the engines and algorithms used to obtain the data are curated internally, and validation methods are not disclosed. Also, some researchers have highlighted that some of the data companies have geographical constraints for Africa, Asia, and South America, which tend to have less presence. For example, Dalle et al. (2017) conducted a comparative study between the OECD Entrepreneurship Financing Database and the near real-time data platform Crunchbase, revealing favourable coverage of new ventures in the United States and Europe. Another disadvantage is the lack of flexibility in the data structures, as this is decided by the data companies.

Relying directly on web scrapers and third-party data that use their proprietary web scraper algorithms could also pose concerns regarding the variance of the results and potential selection bias. For example, Kinne and Axenbeck (2018) found that more innovative firms with registered patents tend to be overrepresented in web mining studies. Finally, data scraping is an expensive exercise in terms of resources, increasing the cost to access this type of data.

18.3.6 Digital Footprint of User-Generated Big Data

The widespread use of social media and IoT has increased the availability of structured data in large volumes. Digital markets (social networks, marketplaces, or any digital platform) allow users to share their interests and concerns about products and services, as well as engage with companies and communicate with them. Furthermore, using IoT sensors allows users' behaviours and preferences within physical environments to be registered. The data that emerge from these interactions are known as user-generated data.

Organisations and researchers can identify patterns using these large databases generated by the user digital footprint. For instance, Saura et al. (2021) explored Open Innovation by using user-generatedcontent on social media. They found eight Open Innovation patterns by combining users' opinions and sentiments through analytics. The novelty of the user digital footprint has so far shown the potential to lead to more innovative approaches by altering processes, developing new products and services, and increasing user engagement through constant interaction and even more data generation.

The promise of users' digital footprint big data has attracted great interest in recent years. This is partly because the development of ICT and web-based technologies facilitated the processes of knowledge creation, sharing and integration, which can ultimately lead to innovation. Yet, this (big) data is excessively complex, characterised not only by its large size but also the wide heterogeneity and high variability over time (Chen and Zhang, 2014; Lycett, 2013). Hence, its exploitation requires management and analysis capabilities beyond traditional databases and software tools. Technologies such as AI, ML and Deep Learning can be used for this purpose. However, they usually require sophisticated knowledge, skills and resources to be implemented, which can lead to excessive costs before exploiting the knowledge. Further concerns entail users' privacy, as some companies that collect these data prioritise economic objectives above ethical design and users' welfare (Saura et al., 2021). In many application sectors, users are unaware of not only issues related to data privacy, but more importantly, the full economic value of their data (Rubin et al., 2022).

18.4 CONCLUSIONS AND FUTURE RESEARCH IN INNOVATION

In this chapter, we have delved into multiple data sources and discussed their opportunities and limitations, from patents and bibliometrics, a well-known and standardised option to capture innovation, to the digital footprint of big data sources, representing a more novel approach led by digital transformation and servitisation across the economy. Researchers can choose among these options, looking at aspects such as granularity, real-time tracking and the flexibility to adapt to the objectives of their research. This is especially significant if the data will later inform policy makers and other scholars, so they can track the current state of innovation development. Although we have not talked about the connection between innovation research and policy making, it is undeniable that more suitable data sources can better inform public policy – hence, more accurate policy can be tailored. Indeed, the power of combining public intent and private intent data can offer more accurate insights that are relevant not only to some groups but to society as a whole.

The future of data for innovation research might develop in at least three alternative paths. First, researchers can take more holistic approaches to capture innovation by combining data sources (for example using pre-configured data, such as patents with either near-real-time data, or with advanced analytics that generate new structural data). As was demonstrated by Kinne and Lenz (2021) that developed new data sources, they not only trained an ML model using pre-configured patent data in order to identify innovating firms but also benchmarked their results with patent data and regional innovation indicators. The use of new technologies, such as text mining or machine learning, can create new variations of data, which can then be analysed using statistical analysis techniques. This point leads to the second path, which relates to the use of Generative AI. Large Language Models (LLM) such as ChatGPT, developed by OpenAI, and Gemini, developed by Google, are expected to transform the way we work, enhance processes and develop new applications that we can only imagine now. Generative AI has the potential to replace the generation of existing digital footprints and create new data points that go beyond current innovation indicators. The novelty of this technology also comes with increasing concerns and uncertainties that are yet to be explored. Finally, digital regulatory mechanisms, for example Open Banking, Open Finance and Smart Data, and the growing digital connectivity between different public services, could generate a new set of data that could be utilised by the scholar community.

In order to incorporate new sources of data relevant for STI studies, researchers should look at the opportunities and limitations that each option offers. It is also important to consider regulations, standards and other norms, especially when evaluating the more novel sources (for example, Generative AI). Researchers should always consider data privacy, control and governance, as well as keep in mind the alternative to democratise access to data. This also means striving to create the maximum value from data, informing policy makers, and promoting equality and trust among all stakeholders in the innovation ecosystem.

REFERENCES

Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, *275*, 314–347.

Dalle, J. M., den Besten, M., & Menon, C. (2017). Using Crunchbase for economic and managerial research. OECD Science, *Technology and Industry Working Papers 8*.

Frank, A. G., Mendes, G. H. S., Ayala, N. F., Ghezzi, A. (2019) Servitization and Industry 4.0 convergence in the digital transformation of product firms: A business model innovation perspective. *Technological Forecasting and Social Change*, *141*, 341–351 https://doi.org/10.1016/j.techfore.2019.01.014

Griliches, Z. (1991). Patent statistics as economic indicators. NBER Working Paper 3301. NBER Working Paper Series. Cambridge MA: National Bureau of Economic Research. 3301.

Hershbein, B., & Kahn, L. B. (2018). Do recessions accelerate routine biased technological change? Evidence from vacancy postings. *American Economic Review*, *108*(7), 1737–1772.

Hershbein, B., Macaluso, C., & Yeh, C. (2018). Concentration in US local labor markets: Evidence from vacancy and employment data. *Documento de Trabajo* (1336).

Kinne, J., & Axenbeck, J. (2018). Web mining of firm websites: A framework for web scraping and a pilot study for Germany. *ZEW-Centre for European Economic Research Discussion Paper*, 18–033.

Kinne, J., & Lenz, D. (2021), Predicting innovative firms using web mining and deep learning. *PLoS One*, 1–18.

Levin, R. (2004). A patent system for the 21st century. *Issues in Science and Technology*, *20*(4), 49–54.

Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386.

Ménière, Y., Rudyk, I., & Valdez, J. (2017). Patents and the Fourth Industrial Revolution. European Patent Office, Munich.

OECD (2009). 'Biotechnology patents', in OECD Science, Technology and Industry Scoreboard 2009, OECD Publishing, Paris.

Pirog, M. A. (2014). Data will drive innovation in public policy and management research in the next decade. *Journal of Policy Analysis and Management*, 537–543.

Rubin, T. H., Helge, A. T., & William, J. (2022). Big data and data ownership rights: The case of car insurance. *Journal of Information Technology Teaching Cases*, *13*(1) https://journals.sagepub.com/doi/full/10.1177/20438869221096859

Sakakibara, M., & Branstetter, L. (2003). Measuring the impact of US research consortia. *Managerial and Decision Economics*, *24*(2–3), 51–69.

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*, *60*, 102331.

Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations, RAND *Journal of Economics*, *21*(1), 172–187.

van Meeteren, M., Trincado-Munoz, F., Rubin, T. H., & Vorley, T. (2022). Rethinking the digital transformation in knowledge-intensive services: A technology space analysis. *Technological Forecasting and Social Change*, *179*, 121631.

Wehn, U., & Montalvo, C. (2018). Knowledge transfer dynamics and innovation: Behaviour, interactions and aggregated outcomes. *Journal of Cleaner Production*, *171*, S56-S68.