DELIVERABLE

# Interim consultation with expert

## Summary

This deliverable reports on the consultation held with experts to direct the efforts of the project. The session reported on three pilot cases namely 1) academy-industry collaboration, digitalization score and industry classification based on web scraped data. The report contains access to the webinar recording, webinar presentation and as an annex, the technical note shared with participants.

## Deliverable Information

| | |
|---|---|
| **Deliverable number and name:** | **Interim consultation with expert** |
| **Due date:** | 30 Sep 2021 |
| **Deliverable:** | D34 |
| **Work Package:** | WP5 |
| **Lead Partner for the Deliverable:** | UNU-MERIT |
| **Author:** | Arho Suominen, Arash Hajikhani, Sajad Ashouri, Scott Cunningham |
| **Reviewers:** | Hugo Hollanders |
| **Approved by:** | Arho Suominen |
| **Dissemination level:** | Confidential |
| **Version** | 7th April 2022 |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

## Disclaimer

This document contains a description of the **BIGPROD** project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium coordinator for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of BIGPROD consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (http://europa.eu.int/)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

# Interim consultation with experts

The purpose of this report is to summarize the interim consultation conducted with experts. The session focused on responding to the comments received in the two prior sessions with experts. The main comments received in the earlier sessions focused on the project creating better understanding and validation of the web scraped data. This led to the current session being planned to focus on explaining three cases in depth, with practical qualitative cases.

The list of invitees for the session included both scientific and practitioner stake-holders. The invitees included Caroline Paunov (OECD), Helene Dernis (OECD), Dan Andrews (OECD) Francesco Losma (OECD), Peter Gal (OECD), Lea Samek (OECD), Flavio Calvino (OECD), Peter Voigt (EC), Jan Einhoff (OECD), Katherine Quezada (EC), Mauro Vigani (EC), Julien Ravet (EC), Alessio Mitra (EC), Thomas Scherngell (Austrian Institute of Technology), Pierre Mohnen (Maastricht Universi-ty), Martina Neuländtner (Austrian Institute of Technology), Johannes Jasny (Fraunhofer ISI), Bart Verspagen (Maastricht University), Janna Axenbeck (Leibniz Centre for European Economic Research), Jan Kinne (Leibniz Centre for European Economic Research), Rene Belberos (Katholieke Universiteit Leuven), Gaétan de Rassenfosse (EPFL), Juan Mateos Garcia (NESTA), Luca Mora (Edinburgh Napier University), Carlo Bottai (UNIVERSITA' DEGLI STUDI DI MILANO - BICOCCA), G Zhang (University of Groningen), Juha-Jaakko Heiskari (VTT), Ville Valovirta (VTT) Juha Oksanen (VTT), Philip Shapira (Uni. Manchester), Catherine Beaudry (Polytechnique Montréal), David Howoldt (Fraunhofer ISI). The participants were also allowed to invite colleagues. The whole project team was also invited.

The session was held on Monday 4th April 2022 from 3:00 PM to 4:00 PM using Microsoft Teams. The session was joined by Arho Suominen (VTT), Ashouri Sajad (VTT), Hajikhani Arash (VTT), Quezada Katherine (EC), Ricardo Henrique da Silva (Polytechnique Montréal), Fabiana Visentin (MERIT), Bart Verspagen (Maastricht University), David Lenz (UNI-GIESSEN), Carlo Bottai (UNIVERSITA' DEGLI STUDI DI MILANO - BICOCCA), Genghao Zhang (University of Groningen), Angela Jäger (Fraunhofer ISI), Patrick Breithaupt (ZEW), Hugo Hollanders (MERIT), Ad Notten (MERIT), Lea Samek (OECD), Sarah-Jeanne Tourangeau (Polytechnique Montré-al), Matthias Deschryvere (VTT), René Belderbos (Katholieke Universiteit Leuven), Catherine Beaudry (Polytechnique Montréal), Juha Oksanen (VTT), Imene Roudjali (NEOMA Business School), Serdar Turkeli (MERIT), Philip Shapira (Uni Manches-ter), Lukas Pukelis (PPMI), Alvar Herrera (Polytechnique Montréal), Heiskari Juha-Jaakko (VTT) and Scott Cunningham (STRATH). In total the session had 26 partic-ipants.

The session was designed so that the participants were sent a technical note[1]. The session began with introduction of the project, a webinar presentation, and a Q&A session. The outcome from the session was a webinar presentation[2] and recording[3] shared publicly. The main outcomes from the consultation included highlighting the need to further validate the data. Overall positive comments were received from the audience on all three cases presented. The further validation should focus on better understanding of what data is shared on websites and how to further make sure that

[1] Technical note has been made publicly available and attached as an annex

[2] Presentation available in the Dataverse

[3] Webinar recording available in the Dataverse

text selected explains what is measured (innovation outcomes, organization processes etc.).

ANNEX

TECHNICAL NOTE / WORKING PAPER

# BIGPROD: Write-up of three pilot cases


## Summary

This technical note reports on a write-up of three pilot cases done during the BIGPROD project. Using a novel web scraped company website data of approximately 100 000 companies, the note reports on a comparison of NACE and Microsoft Academic Graph (MAG) based industry classifications, Field of Study (FOS) code based digitalization score and academy-Industry collaboration based on website data. The technical note is intended to open discussion of the potential of the web scraped data based indicators compared to existing innovation measures.


## Information

| | |
|---|---|
| **Deliverable number and name:** | **Technical Note** |
| **Date:** | 31th January 2022 |
| **Work Package:** | WP4 |
| **Lead Partner for the Deliverable:** | STRATH |
| **Author:** | Arho Suominen & Arash Hajikhani & Sajad Ashouri & Scott Cunningham |
| **Dissemination level:** | Public |


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

## Disclaimer

This document contains a description of the **BIGPROD** project findings, work and products. Certain parts of it might be under partner Intellectual Property Right (IPR) rules so, prior to using its content please contact the consortium coordinator for approval.

In case you believe that this document harms in any way IPR held by you as a person or as a representative of an entity, please do notify us immediately.

The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated in the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

This publication has been produced with the assistance of the European Union. The content of this publication is the sole responsibility of BIGPROD consortium and can in no way be taken to reflect the views of the European Union.

The European Union is established in accordance with the Treaty on European Union (Maastricht). There are currently 27 Member States of the Union. It is based on the European Communities and the member states cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice and the Court of Auditors (http://europa.eu.int/)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

# Introduction

This technical note is a write-up of work carried our during BIGPROD project[4] and in particular during its Work package 4 focusing on piloting the use of the novel web-scraped data. The technical note describes three case studies, which focus on explaining the potential of the webscraped data and verifying with a comparative approach to existing measures. The three cases presented in this note ar:

1. A comparison of NACE and Microsoft Academic Graph (MAG) based industry classifications
2. Field of Study (FOS) code based digitalization score
3. Academy-Industry collaboration based on website data

The first case shows the potential of creating novel industry classification based on the classification of scientific knowledge. Utilizing MAG allows the use of a hierarchical mapping of "the world's knowledge" to better understand the knowledge embedded in industry. The findings show that the webscraped data based analysis is able to produce a significantly more broad understanding of the knowledge embedded in industry, while allowing for a mapping to the existing NACE industry classification.

The second case focuses on creating a digitalization score based on the field of study codes mentioned on the company websites. The method proposed in this technical note allows for measuring companies digitalization score at two levels: products and capabilities. This offers a unique vantage point by looking at innovation outcomes but also the processes in the company. In addition, the method offers a practical way of analysing digital capabilities at scale overcoming disadvantage of questionnaires, such as CIS study, in terms of coverage.

The third case focuses on analysing collaborative patterns between companies and in particular companies and research and technology organization. The objective of the analysis is to extend the existing co-publishing and co-patenting based measures with a novel vantage point. The results show that the web scraped approach is able to capture a connections from companies which do not patent and publish. However, there is a significant relationship between companies patenting and publishing and the number of connections with research and technology organizations. The results also highlight case sample typologies of the types collaboration identified from websites.

The BIGPROD sample companies were generated from a sample of medium or high technology companies from the European Union and the United Kingdom. The details for the data are available from Ashouri, Suominen, Hajikhani, Pukelis, Schubert, Türkeli, van Beers, et al., (2021). The data contains roughly 24 700 companies based in Germany, 18 200 from the United Kingdom, 16 200 from Italy followed by France, Spain and Poland and smaller European companies. Previous work by the project can be seen on the project website[5], data repository[6] and ResearchGate page[7].

This technical note is work in progress and shared as a working paper toward three different publications.

[4] Refer to the project website https://www.bigprod.eu/
[5] https://www.bigprod.eu/output/
[6] https://dataverse.nl/dataverse/BIGPROD_Data_Sample
[7] https://www.researchgate.net/project/BIGPROD-Addressing-the-Productivity-Paradox-with-Big-Data

# MAG industry classification

The central element of technology and innovation management is continuous technological change. Foundations laid by Kontrantieff, Schumpeter and later Adner and Levinthal (2001) on the impact of technological change in innovation highlight the dynamism of our technological surroundings. The analysis of technological change within industry has relied on using classification like Standard Industrial Classification (SIC) or the European Classification of Economic Activities (NACE). The classification has been used to understand technological regimes (Leiponen and Drejer, 2007), technological change (Archibugi and Planta, 1996), regional impacts of technological change (Gitto, 2017), job creation (Gagliardi, Marin and Miriello, 2016) and policy impacts (Calel and Dechezleprêtre, 2012; Crespi, 2013) among a plethora of other technological change related questions.

Notwithstanding the importance of the economic activity classifications, the NACE-type of measures struggle with the dynamism of technological change (Kee, 2019; Zhou, Mu and Lin, 2021). The main challenge of existing measures is their slow process in classifying new industries emerging from novel technology (Kee, 2019). The methods also lack the inability to extend analyzing clusters of industries beyond the business operation, based on, e.g. technology adoption or industry networks (Zhou, Mu and Lin, 2021). Insufficiencies regarding innovation indicators have been assumed to be a possible cause for the so-called productivity paradox and a simultaneous productivity slowdown in developed countries (OECD, 2015). Measuring the outcomes of innovation at correct levels of aggregation is also challenging (Neuhäusler, Frietsch and Kroll, 2019). In addition, the linear and slow renewal of the classification systems limit the ability to correctly model the rapidly changing technological and industrial structure (Zhou, Mu and Lin, 2021). These challenges have previously been tackled with concordance tables (Schmoch *et al.*, 2003; Frietsch *et al.*, 2017; Neuhäusler *et al.*, 2017), which in part alleviate some of the issues faced in applying industry classification to innovation management and technological change research. Such concordance tables utilized patents and publications affiliated to firms to approximate the technology domain of firms and therefore create associations to their assigned NACE or SIC code.

The increasing possibility of creating machine learning-based classification from different source data has opened novel possibilities to create new type of industry classifications which can potentially be more dynamic to industrial and technological change. Already Bernestein et al. (2003) showed that a relational vector space model was competent in creating meaningful industry classifications. Since studies have extensively reported on different natural language based and machine learning employing methods to improve industry classification and research applying them. For example, Pant & Sheng (2015) created a model for competitor prediction superior to SIC classification based approaches, Goindani et al. (2017) and Chern et al. (2018) used machine learning to create an accurate industry classification method from job postings and Zhang et al. (2012) used semi-supervised text mining to identify industrial networks.

In developing a more dynamic industry classification, rather than focusing on an isolated innovation management question, Lamby & Isemann (2018) used word embeddings applied to newspaper articles to create a model of industry classifications. Their findings highlighted significant promise in creating text-based industry classification, but to be robust, more development is needed. Lyocsa & Vyrost (2011) also highlight

the potential of a more dynamic approach to creating industry classification while emphasizing that the methods should be transparent and easily replicable. While the measurement of innovative activity is critical for researchers and practitions, there is a call to find an approach or approaches that would create the needed dynamism while being transparent and replicable (Adams, Bessant and Phelps, 2006).

## Approach and findings

We propose an approach of using company website data at scale to develop an industry classification and relying on the well-established and transparent Microsoft Academic Graph (MAG) Field of Study (FOS) model to create an industry classification. Research has shown the potential of website data in creating valuable information on innovative activity at a firm-level. For example, website data can be used to better understand innovation outcomes, strategies, and relationships (Gök, Waterworth and Shapira, 2015). Running the analysis at scale also allows drawing industry-wide structures from the data. The BIGPROD project, reported in Ashouri et al. (2021), created a dataset from website data scraped from 96,921 medium-high and high-technology companies. Using the data, we infer a classification to company website data using natural language processing and the hierarchical topic model-based MAG FOS categories. Our study leverages the framework presented in Ashouri et al. (2021) to create a transparent approach to creating a classification. We compare our classification to the NACE company level classification. Our results show an average 20 percent expansion to the existing NACE classification. This expansions allows us to create a more multilayered and dynamic view of industry and the technologies used.

Figure 1 illustrates the methodological process for transforming companies' website scrapped content to their equivalent scientific literature and finally to NACE classifications. The process includes several separate processes. Identifying the sample and its associated metadata from a data-based, website scraping and mapping website data to a knowledge mapping, namely MAG FOS ids. Subsequently, the industry classification in company metadata (NACE) is modelled in relation to the knowledge mapping created with the MAG.

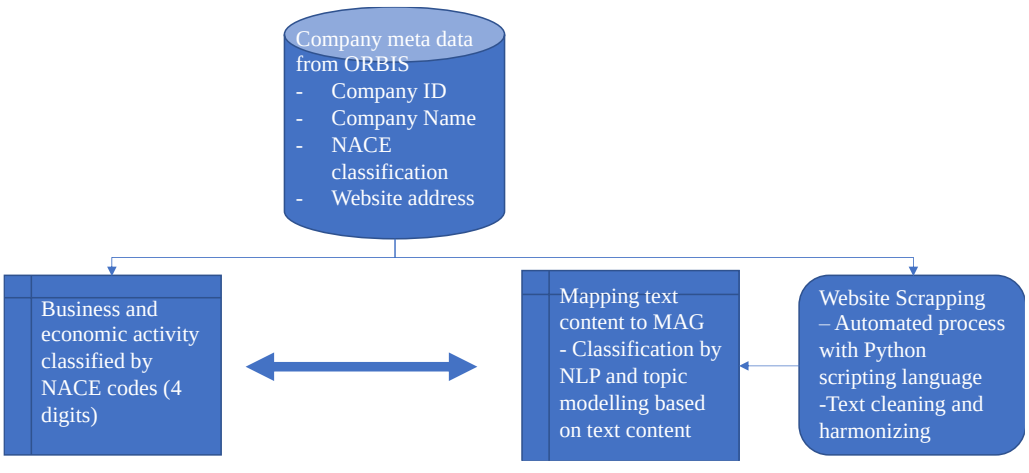


*Figure 1. Reallocation of NACE classification code to MAG hierarchical topic model code*

Involving the company's NACE codes resulted in a data structure where each NACE code is now allocated to a FOS code. Calculating a cosine similarity score for the FOS codes (continuous score between 0 to 1) the data was filtered to only consider 0.5 or above similarity scores for a NACE-FOS relationships. This resulted to 11,391 weighted conncetions between 65 NACE codes and 4,818 FOS ids. Figure 2 shows the count distribution of FOS codes to each NACE code. This is a skewed distribution where NACE 2899 is associated with 874 FOS codes and NACE 2894 with only 2. This difference is partly due to the unequal distribution of companies in the sample, further described in Ashouri et al. (2021).

The resulting network between NACE codes and FOS codes can be seen as a bipartite network, where nodes are represented by both NACE and FOS codes. The network does not have links between within either groups but only between the groups. The network representation allows for sense-making of the interconnections of the codes. A section of the network structure can be seen in Figure 3, where the network has been filtered based on indegree, which is the number of head ends adjacent to a node. Filtering the network to indegree to five reveals that there are over 100 FOS codes connected to 60 NACE codes or expanding the definition of the NACE codes. This illustration is presented in Figure 3 looking at individual cases shows the expansion of the NACE codes via a more detailed representation of business activities captured from the websites via FOS codes. For instance, in Table 1 one cluster of not very related NACE codes are affiliated to similar cluster of FOS codes relating to filtering technologies.

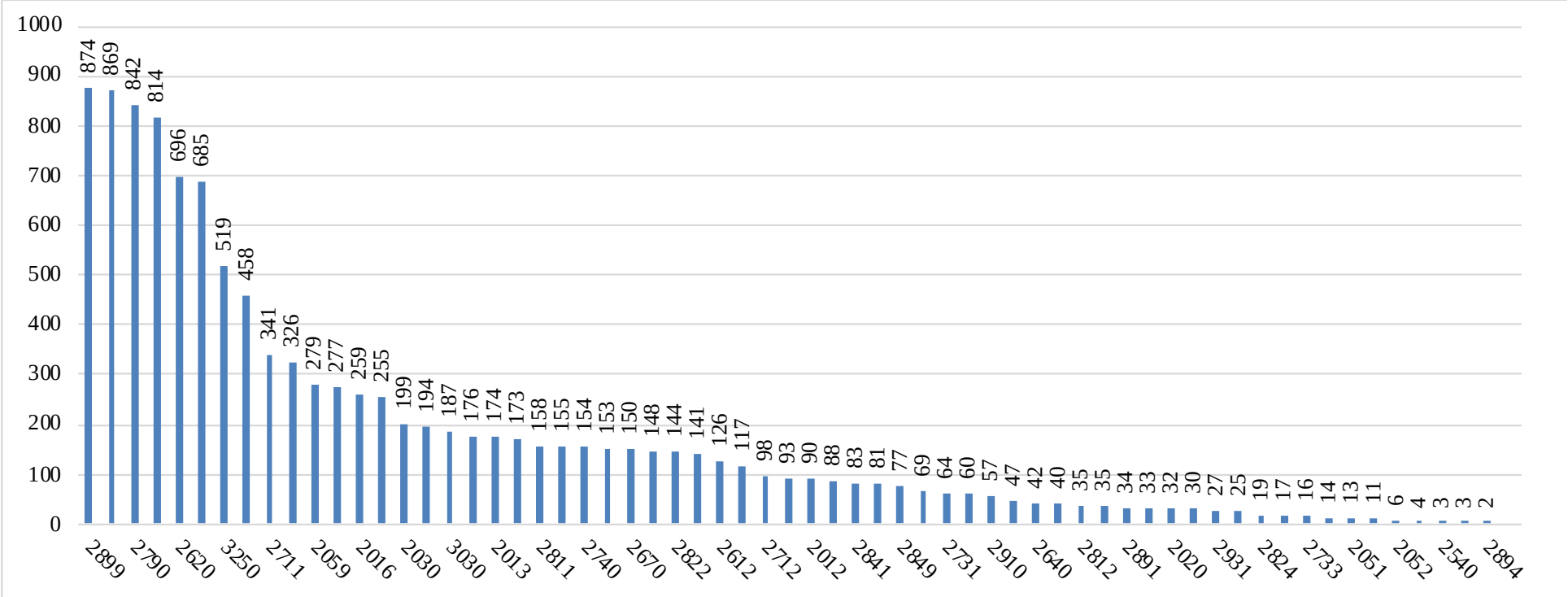


*Figure 2 NACE codes allocation to FOS codes quantity*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

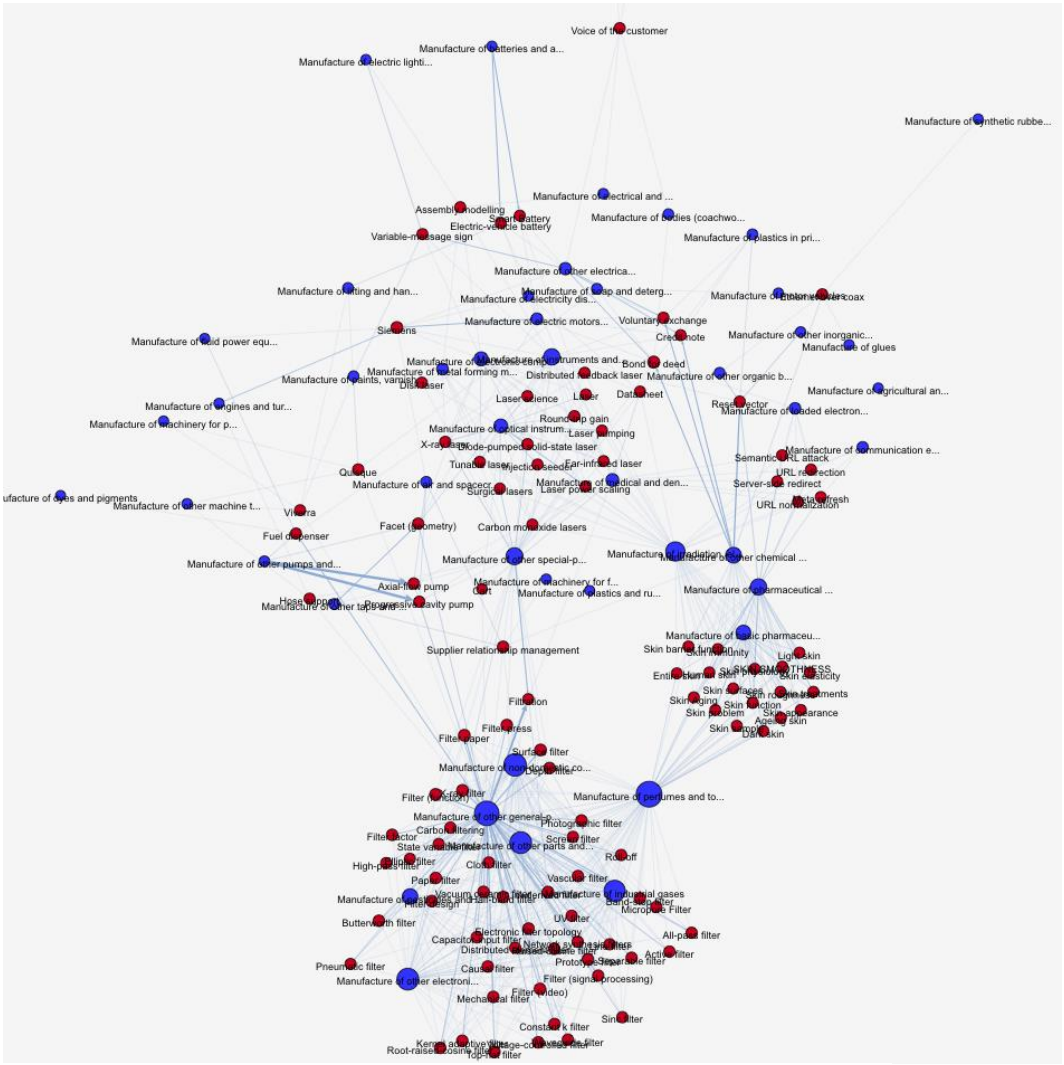


*Figure 3 Filter for nodes with minimum 5 indegrees*

*Table 1 NACE to FOS concordance table*

| NACE (4 digit code) and description | | Related FOS codes tag names | | |
|---|---|---|---|---|
| 2825 | Manufacture of non-domestic cooling and ventilation equipment | Filter paper | Butterworth filter | Vascular filter |
| 2829 | Manufacture of other general-purpose machinery n.e.c. | m-derived filter | UV filter | filter design |
| 2732 | Manufacture of other electronic and electric wires and cables | X-ray filter | Sinc filter | capacitor-input filter |
| 2932 | Manufacture of other parts and accessories for motor vehicles | Surface filter | All-pass filter | Separable filter |
| 2042 | Manufacture of perfumes and toilet preparations | Filter (function) | Scree filter | Prototype filter |
| 2011 | Manufacture of industrial gases | Vacuum ceramic filter | Filter press | Electronic filter topology |
| 2899 | Manufacture of other special-purpose machinery n.e.c. | High-pass Filter | Carbon filtering | Network synthesis filter |
| 2020 | Manufacture of pesticides and other agrochemical products | Pneumatic filter | Cloth filter | |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

We developed a system of mapping firms website content in the technological space that enabled us to develop a novel approach to capture firms' technological capabilities. Leveraging company website content at scale and using an established and transparent graph of knowledge our methodology offers a dynamic, transparent and replicable classification of industrial activity. The results were insightful as the new classification showed in average 20% expansion on the existing NACE classification category. The overlapping effect among NACE classification was spotted, which suggests additional breadth to the current NACE classification definitions.

The approach has a number of advantages. First, they leverage the most accurate source of companies' activities (their websites) as the raw data. Prior research has shown that company website data offers a rich information source on companies' technological activities (Gök et al., 2015). Second, the utilization of company website content eliminates lag in terms of business reporting of their activities. The data platform developed for the project (Ashouri et al., 2021) can operate in continuous mode offering a way to respond but also to track technological change at firm-level in a near-continuous way. Third, we map the website content to a curated database of scientific publications (MAG) which highlights the important activity of companies. MAG is freely available and thus offers the methods the transparency and replicability needed to serve as a practical classification tool for industrial activity.

We aim to expand this analysis by quantifying the similarity and difference of the FOS codes to better capture the various NACE codes distance to each other. This approach eventually enables us to have an accurate image of businesses and their activities by being able to benchmark companies' breadth and depth of activities compared to their peers. In an example Table 1 we demonstrated the correspondence of firm's NACE codes to our compiled FOS codes based on firm's website content. As can be seen, there are numerous adjacent identified FOS tags for any given NACE classification. Figure 4 encapsulates our approach visually.

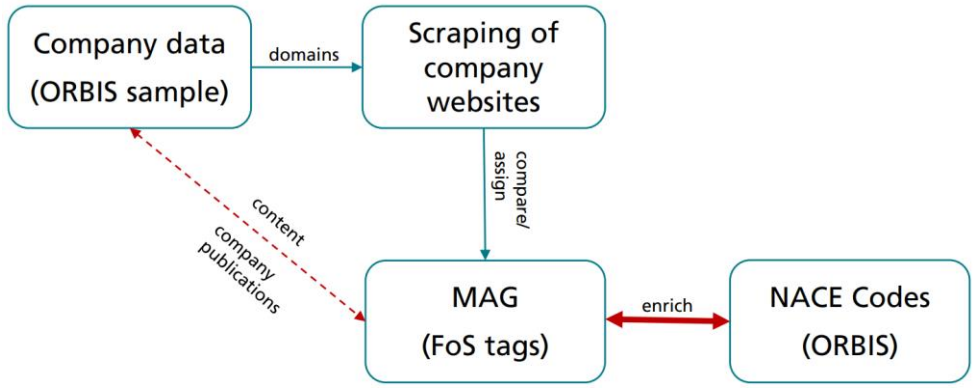


*Figure 4. Sensemaking of NACE code expansion*

Changing trends and events that happen with so much rapidity implies the need for tools and methods to capture changes in time, accurately and cost-effectively. The historical and longitudinal experimentation regarding businesses classification offers tremendous value. However, our contribution is to enhance the legacy classification of business activities and assist their granularity and relevance.

Our methodology offers immediate implications for practitioners, business managers, investors and policy makers. One benefit from the approach in to identify areas of growth and decline in terms of active investments or collaboration/competitor detection. For policy makers, this can help them make more astute decisions about supportive programs in form of grants that assist the development of incubators and research parks and foster research consortia that aligns with market and business dynamics.

# FOS digitalization score

The expansion of digital technologies and integration with numerous fields of science and technologies in recent years, has led firms toward digital transformation and digitalization. Digitalization is the use of digital technologies to innovate business routines toward more efficient and flexible performance, providing new revenue streams through defining new business models, and promoting competitive advantages by exploiting value-producing opportunities (Gobble, 2018). Successful adoption of digital transformation requires firms to identify and obtain the essential capabilities at various operational and organizational levels (Annarelli *et al.*, 2021).

Based on dynamic capabilities theory, digital capabilities of firms can be investigated under three different contexts of dynamic capability in terms of Sensing opportunities by managing digital ecosystem partnerships, Seizing firms' digital capabilities, Reconfiguring firms' digital resources and routines (Annarelli *et al.*, 2021). Pertaining the sensing opportunities and threats, Selander et al. (2013) state achieving benefits from digital ecosystems expect strong firms' network capabilities as well as absorptive capacities. Absorbing new digital business resources to promote novel innovations, products, or services, demands the strengthening of firms' capabilities in technical and social contexts (Sambamurthy, Bharadwaj and Grover, 2003).

Employing heterogeneous resources, deployment and of IT, managerial cognition in initiating changes, and organizing IT capabilities are the main challenges in seizing firms' digital capabilities (Annarelli *et al.*, 2021). Heterogenous resources can facilitate exchanging and processing information toward automated tasks (Mishra, Konana and Barua, 2007). Deployment of IT supports digital process and information flow, and consequently, advances innovative ideas in firms (Drnevich and Croson, 2013; Selander, Henfridsson and Svahn, 2013). Managerial cognitive capabilities also can direct searching processes in a new learning environment and supervise the evolution of firms' capabilities. Such cognitive capabilities can ultimately lead to successful digital transformation (Tripsas and Gavetti, 2000).

Considering the restructuring of firms' digital resources and practices, Nylén & Holmström (2015) note that embedding digital capabilities in suitable situations permit managing of process type of innovation in the firms. Similarly, in addition to the internal firms' capabilities, Wheeler (2002) acknowledges that customer value co-creation as firm capabilities in the development of business innovation using digital networks. Figure 5 shows the dimensions of digitalizations capabilities at different level.

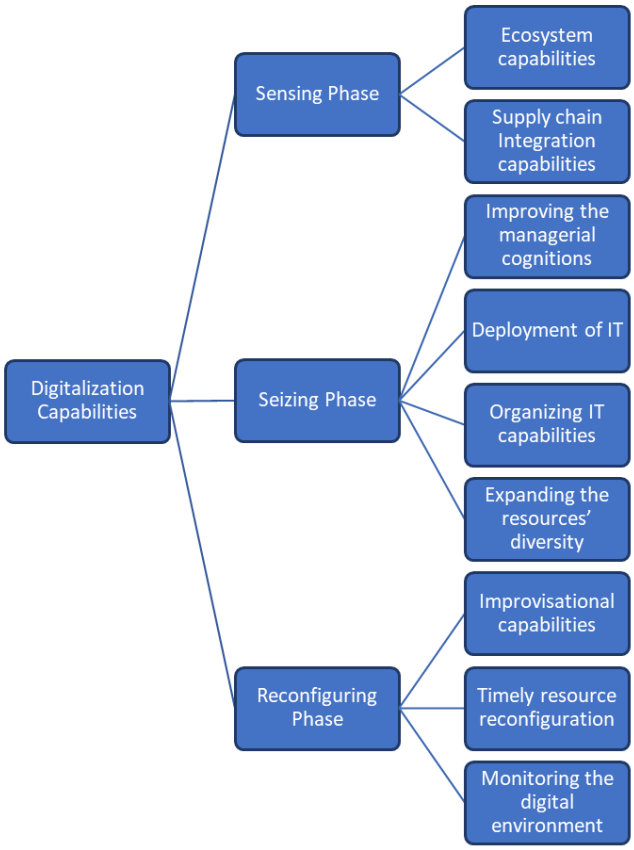


*Figure 5. Dimensions of digitalization capabilities based on the dynamic capability theory* (Annarelli et al., 2021a, adopted)

The complex structure of digitalization capabilities implies that the assessment of all various aspect as a single measures requires a source of information incorporating all the relevant data in one framework. Websites offer a rich source of information on company behavior (Gök, Waterworth and Shapira, 2015; Kinne and Axenbeck, 2020; Axenbeck and Breithaupt, 2021). The express purpose of websites is not exclusively to communicate technical expertise or innovative capability, nonetheless, previous work demonstrates the strong correlations of web texts with a variety of innovative measures including R&D expenditures, R&D employment, firms' capabilities, alliance network, patenting activity, and so on (Gök, Waterworth and Shapira, 2015). Communicating the firm capabilities throughout the website enables the use of webpage in the development of extensive internal capability measures at a large scale; as conventional data sources or surveys are hampered by their coverage. Moreover, utilizing webpages as data source facilitate more frequent and updated data in comparison with the conventional data source (Arora *et al.*, 2020).

The content analysis of 38 companies' websites, including large and SME firms in both B2B and B2C; showed that the website information can be categorized into 7 categories. The qualitative investigation revealed that not all the messages and signals cover the information over all categories, as the websites structures are strongly correlated with the firms' size. In contrary, firms can address several purposes using a single message. These categories are explained as follow:

1. C1. The website attempts to signal the competitive advantages of the firm's products and services. The competitive advantages can be related to quality

and technology level, multi-aspect oriented, affordability, and product/ service standards.

2. C2. The website communicates the firm's competencies and capabilities. This message also can be highlighted using the firm's knowledge/capabilities in offering diverse solutions as well as the firm's leadership and dominance in the market. The website also may remark the firms' competencies through their relations with other the firms/ brands and addressing the firms' position in the value chain.

3. C3. The website communicates corporate social responsibilities in terms of how social responsibility concerns are engaged in the company's business activities and policies. Corporate social responsibility may include sustainability issues, philanthropic activities, as well as inclusion and diversity.

4. C4. The website may communicate the firm's ethics and compliances, which can be explained through codes of conduct and ethical frameworks. The audience of such messages can be social, personnel, suppliers and logistics, peers, and also governments.

5. C5. The website describes the organizational structure, investors relations, and corporate governance. This message also may cover the firm's mission and vision, long-term strategies, and growth framework.

6. C6. The website presents the financial documents and earnings of the company, to show the profitability of companies. All relevant financial reports might be accessible through the website.

7. C7. The website targets the current and future suppliers and logistics partners to communicate the firm's strategies in alliances in growing the business.

The aforementioned categories elaborate how the website content manifest companies capabilities and competences. Digital capabilities also as a representative of dynamic capabilities are signalled through such inexpensive dissemination channels, in order to sustain the firms' competitive advantage among the competitors.

Besides the overall organizational level digitalization capabilities, it worths mentioning that in manufacturing industries as the competitive advantages are shaped based on the technologies and products, the assessment of digital capabilities requires considering the firms' attention on development of digital products (Björkdahl, 2020). Deployment of digital technologies in products can enhance the efficiency in product performance and design. Integrating product development with Artificial Intelligence also can offer new functions and engage customer value creations, which brings up the opportunity for new business models. Moreover, such high-tech industries maintain significant overlap between product development and R&D inputs, where R&D strategies and adoption of new R&D capabilities and competencies affect the development of products and technologies (Hagedoorn and Cloodt, 2003). Therefore, the assessment of digitalization in manufacturing industries requires to have particular attentions toward digital product development in order to not only does evaluate the impact of such digital transformation on firm performance and productivity, but also the companies determination toward digital transformation in R&D capabilities. (Björkdahl, 2020; Kohtamäki *et al.*, 2020)

## Digitalization measures

In the literature digitalization capabilities of firms have typically been measured based on survey data (Björkdahl, 2020; Kohtamäki *et al.*, 2020). Instead, this analysis measures digitalization using a classification of web-scraped text data based on the entire scientific body of knowledge.

According to the high dimensionality and complexity of unstructured text data, this analysis follows a novel approach by employing a large global publication database that can serve to measure the similarity between the structured data sources and the web-scraped text data. Microsoft Corporation has developed an open bibliometric database – that is similar to Google Scholar – named Microsoft Academic[8]. Microsoft Academic Graph (MAG) is a large heterogeneous graph comprised of more than 200 million publications and the related authors, venues, organizations, and fields of study. As of today, the MAG is the largest publicly available dataset of scholarly publications and the largest dataset of open citation data. Fields of Study (FOS) are the results of a hierarchical topic model run on the entire MAG data corpus. FOS IDs are introduced at five levels of detail, resulting in over 700,000 total topics and classifications. Certain data elements like FOS fields are calculated with the data provided from MS Academic Graph. FOS data and the underlying keyword distributions for each FOS are referenced from a dump of MAG. This entails the complete download of the entire MAG dataset to a single user's storage account. The analysis is based on the 2019-10-10 version of MAG dataset.

Based on website texts the new methodology introduced translates digitalization into single and easily read scores. The process begins with specific high-level FOS identified in the MAG associated with computer science. Both the parent, as well as all children of these FOS codes, are associated with digitalization. Identification of FOS identifiers is conducted in line with (Ashouri, Suominen, Hajikhani, Pukelis, Schubert, Türkeli, van Beers, *et al.*, 2021; Hajikhani *et al.*, 2021)

We propose two novel digitalization measures capture the firms' digitalization at two different levels. The first measure reflects the digital capability in product development, which is substantial for high-tech manufacturing industries, and the second measure reflects the organizational level of firms' digital capabilities.

**Product digitalization score:** To construct the product digitalization capabilities, the identified products on the company website are recorded and associated with these FOS IDs based on a high level of shared or overlapped text. Computer science FOS are scored as one, non-computer science FOS are scored as zero. The aggregation and average of all products available on the website results in a ratio-scored variable ranging from zero to one. The new measure of product digitalization score, therefore, uses actual product description rather than conventional wisdom to determine whether a product is digital or not. An added value is the resultant aggregation of the entirety of the listed products of a company.

**Capability score:** The firm's digital capability indicator on the other side measures the firms' communications related to the digital science and technologies, that embrace digital capabilities with sensing, seizing, and reconfiguring characteristics. Consequently, the firm's digitization capability score measures the weight of FOS ids associated with computer science in the company website FOS id subset, and it can vary from 0 to 1. This indicator is designed based on all website text of the firm. The relative importance (or weight) of all FOS codes identified in the Microsoft Academic Graph

[8] Source: https://academic.microsoft.com/home

associated with computer science is calculated (as a share of number of specific computer science FOS codes to total number of FOS codes) and summated. The indicator values sit between 0 and 1.

The proposed methodology of big data-based digitalization measures targeted two aspects of digitalization in companies. This methodology facilitates the assessment of digital capabilities at a very large scale and overcomes the disadvantage of questionnaires in terms of coverage. Studying of CIS2018 (Finland) surveys revealed that less than 10% of firms that participated in the survey, addressed the questions related to digitalization.

According to the significance of digital products in manufacturing industries, the first measure examines the digital deployment in products. The second measure summarizing all digitalization aspects in a single number also simplifies the investigation of general digitalization capabilities from the sensing, seizing, and reconfiguration phases in dynamic capability theory. However, this measure is restricted in distinguishing different digitalization capabilities. To measure any specific digitalization capabilities such as human resources, networks, supply chain and so on, different proxies are required to capture the competencies.

# Collaboration

The data has been downloaded in two data frames from two separate tables from the SQL table described in Ashouri, Suominen, Hajikhani, Pukelis, Schubert, Türkeli, van Beers, et al., (2021), namely “Companies Table” and “Collaborations Table”. The company data frame, with the sample companies, include in total 96,921 companies with 10 variables included. The variables included for the companies include a unique id, company name, website, country of company headquarters, NACE code, ISO standards, countries mentioned in the webpage and information on the thematic content of the webpage. The Collaborations Table created a data frame with 222 756 instances of collaboration.

For collaborations, we identified the name of the collaborator, category, country code and a unique identifier that links the collaborator to a sample company. In total, there were 57 899 unique collaborators in the dataset that were connected to 18 697 companies from the total sample, which means 19.4 % of the companies have mentioned their collaborative activities on their website. In total, the collaboration information forms a bipartite network G where nodes are defined as collaborators or sample companies and the edges linking the nodes. In total G is created by 76 596 nodes connected by 221 565 edges. The average degree of the network is 5,785. Using the variable category, built on identifying the type of organization being either a research and technology organization (RTO) or other, we can focus only on RTOs. This limits the network to 9 133 companies that are connected to 5 546 organizations. In total for 9,5 % of the companies the web scraping process was able to retrieve an RTO collaboration.

The nodes in network G were geocoded using Google Maps API and OpenStreetMap API. Taking the organizations name and country information, Google API was used to retrieve the latitude and longitude for each organization, would these be from the sample or collaborator. For the records not geocoded using Google Maps API OpenStreetMap API was used as a secondary source. From the collaborators 52 029 or-

ganizations were geocoded, which is 91,2 percent of all organizations. From the sample companies, 87 354 were geocoded, which represents in total 90,9 percent of the sample.

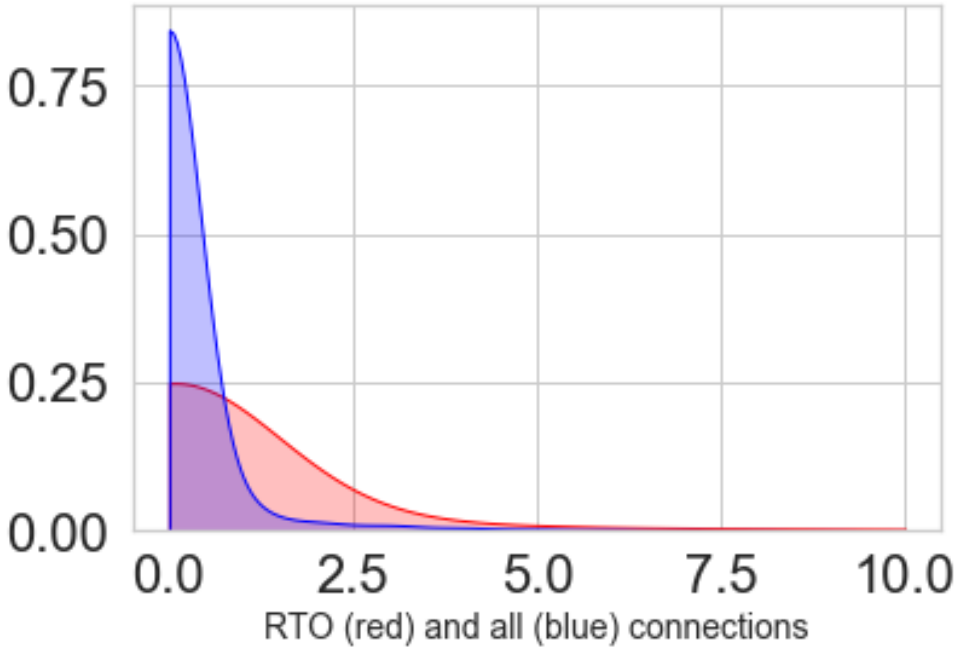


*Figure 6 Distribution of the number of connections in sample companies. (N=96107)*

The nodes in network G were geocoded using Google Maps API and OpenStreetMap API. Taking the organizations name and country information, Google API was used to retrieve the latitude and longitude for each organization, would these be from the sample or collaborator. For the records not geocoded using Google Maps API OpenStreetMap API was used as a secondary source. From the collaborators 52 029 organizations were geocoded, which is 91,2 percent of all organizations. From the sample companies, 87 354 were geocoded, which represents in total 90,9 percent of the sample. The collaboration sample does not have an equally large sample of companies in different industries. Seen in the Table 2 at NACE two-level, there is an over representation of NACE 28, while NACE 25 and 30 have a significantly low number of companies. This said, NACE 21 has a significant number of collaborations, having the highest mean and median values in both collaboration types.

*Table 2 Collaborative patterns between NACE codes.*

| | | All | All | RTO | RTO |
|---|---|---|---|---|---|
| NACE 2-level | count | mean | std | mean | std |
| **20** | 6634 | 0.67 | 3.24 | 0.19 | 1.26 |
| **21** | 1758 | 1.76 | 6.99 | 0.68 | 3.13 |
| **25** | 137 | 0.83 | 2.55 | 0.18 | 0.75 |
| **26** | 7149 | 1.02 | 4.36 | 0.35 | 2.27 |
| **27** | 6374 | 0.67 | 3.07 | 0.18 | 1.19 |
| **28** | 17854 | 0.49 | 2.54 | 0.13 | 1.27 |
| **29** | 3281 | 0.84 | 3.22 | 0.21 | 1.10 |
| **30** | 411 | 2.12 | 5.97 | 0.73 | 2.76 |
| **32** | 2459 | 0.77 | 4.43 | 0.33 | 2.29 |

Current literature focuses on the use of patent or publication collaboration as an indication of a academia-industry linkages. Using the BIGPROD data, we evaluate if there is a relationship with a company having publicatioin or patents and the companies collaborative behavior (specifically with RTO). Figure 7 and Figure 8 show the distribution of connections overlayed with the company having scientific publications or patents. Mann-Whitney U test shows that there is a statistically significant difference between companies with publications ($U= 15311413.5$, $p<0.05$) and and patents ($U= 24791264.5$, $p<0.05$) in the number of connections to RTO.

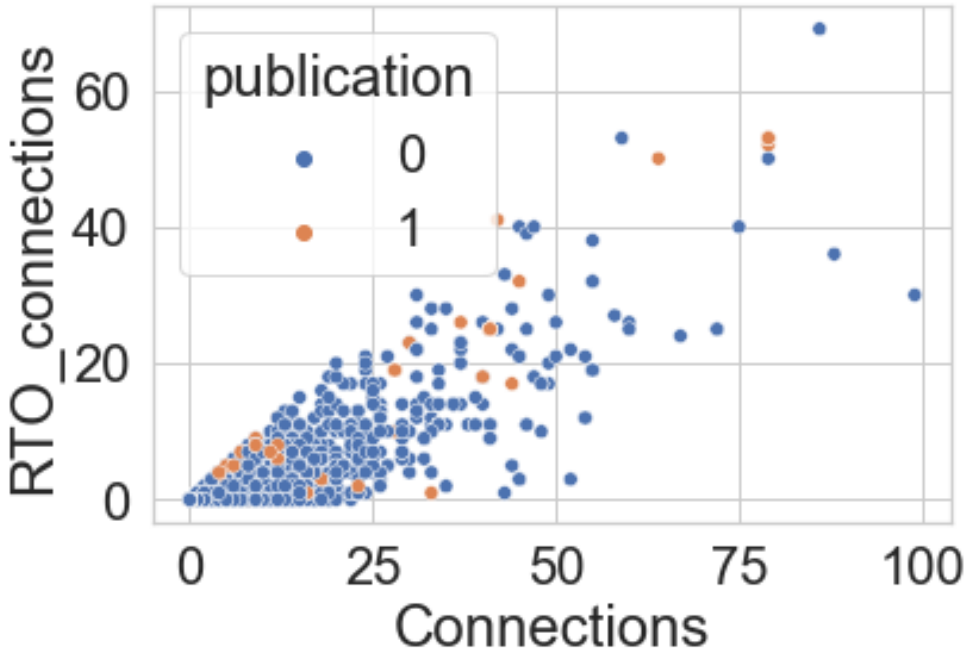


*Figure 7 Count of connections overall and to RTOs overlayed with company having publications.*

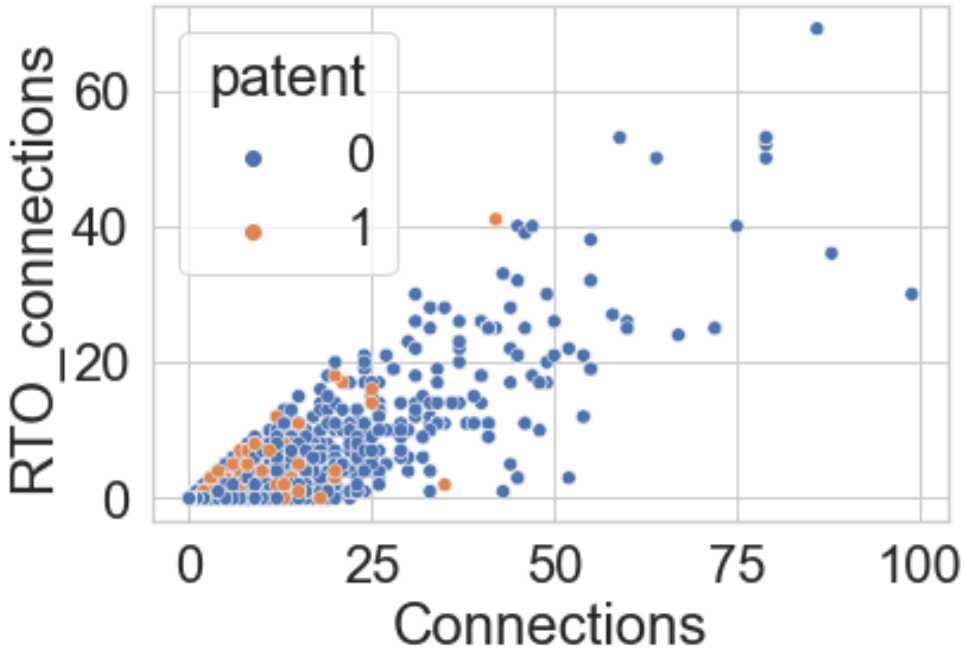


*Figure 8 Count of connections overall and to RTOs overlayed with company having patents.*

To validate our results, collaborations were also measured between the sample companies and the collaborators using co-publications. For an in-depth review process, we considered all the entity cases that have identified collaboration with research organizations on their websites as well as their scientific publications since the year 2020. The corresponded academic publications for each entity were obtained using Microsoft Academic Graph data. Among companies reporting collaboration with research organizations through their webpage, only 4.3% have co-publication activities during the mentioned period. Therefore, a relatively limited number of firms are interested in academic publishing, irrespective of their partnership with research organizations.

To reflect on the overlaps and differences in coverage of co-publications between the companies vs collaborations mentioned on their websites, we mapped the collaboration network using co-publishing data collaborators for the 81 entities with publication activities. The approach generated four additional variables for defining the overlap and difference between co-publishing data and web scraped data. These four variables are constructed upon three major groups 1) collaborations in web scraped data 2) collaborations in co-publishing bibliometric data, 3) “collaborations both in 1 and 2”. The results are seen in Table 3.

*Table 3 Verification table against web scraped data and academy-industry co-publishing.*

| | WEB SCRAPED COVERAGE COMPARED TO CO-PUBLICATION | WEB SCRAPED COVERAGE MAGNITUDE COMPARED TO ALL | WEB SCRAPED AND CO-PUBLICATION OVERALL | WEB SCRAPED ADDITIONALITY |
|---|---|---|---|---|
| AVG | 2,3620 | 0,3082 | 0,0418 | 0,2664 |
| MIN | 0,0087 | 0,0087 | 0,0010 | 0,0077 |
| MAX | 31 | 1 | 0,142857143 | 0,967741935 |
| MED | 0,1546 | 0,1370 | 0,0323 | 0,1137 |
| STDV | 7,0344 | 0,3052 | 0,0339 | 0,2934 |

Finally a case study verification of the sample was done. In the process, cases were randomly selected from the data. Identified collaborators was first search via Google (search query included the names as in data). If a link to the sample companies page was found it was used to retrieve data. Secondly, collaboration was search from the sample companies webpage. If a page to the sample companies website was found it was used to retrieve data. Data retrieve included “text” where collaboration was mentioned and the url. The type of collaboration was classified.

*Table 4 Case examples of collaborations.*

| Company | Collaborator | Description | Link | Type |
|---|---|---|---|---|
| ARM HOLDINGS PLC | ALTERA | Altera and ARM Announce Industry's First FPGA-Adaptive Embedded Software Toolkit | Altera and ARM Announce Industry's First FPGA-Adaptive Embedded Software Toolkit – Arm® | Product collaboration |
| CN Bio Innovations | Alnylam | CN Bio Innovations announces research collaboration with Alnylam | CN Bio Innovations announces research collaboration with Alnylam (cn-bio.com) | Research collaboration |
| Johnson Matthey | University of Oulu | Palladium Impurity Removal from Active Pharmaceutical Ingredient Process Streams | Palladium Impurity Removal from API Process \| Scale-up (matthey.com) | Human resources |
| THE SHADOW ROBOT COMPANY LIMITED | Lufthansa | ANA Avatar Unites Tech Leaders to Debut the World's First Touch-transmitting Telerobotic Hand at Amazon re:MARS Tech Showcase | ANA Avatar Unites Tech Leaders to Debut the World's First Touch-transmitting Telerobotic Hand at Amazon re:MARS Tech Showcase – Shadow Robot Company | Trial |
| Agilent | Eli Lily and company | Agilent is Partnering with Pharma to Help Your Immune System Fight Cancer | Agilent Technologies Blog Agilent is Partnering with Pharma to Help Your Immune System Fight Cancer | Research collaboration |
| AREVA H2GEN GMBH | European Marine Energy Centre | The Integrated Tidal Energy into the European Grid (ITEG) project … Led by the EMEC: European Marine Energy Centre, | Elogen (elogenh2.com) | Research collaboration |
| ENGCON POLAND SP. Z O.O. | JN Bentley (misidentified as Bentley) | JN Bentley increases excavator efficiency with Engcon's latest tiltrotator technology | JN Bentley increases excavator efficiency with engcon's latest tiltrotator technology \| engcon | Use case |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

*Table 5 Case examples of collaborations (cont.)..*

| Company | Collaborator | Description | Link | Type |
|---|---|---|---|---|
| FUJITSU (IRELAND) LIMITED | FRANKFURT STOCK EXCHANGE | Lists stock on Frankfurt Stock Exchange | Company milestones (Chronological table) : Fujitsu Global | Funding |
| NOTE TORSBY AB | NIMBUS Boats | Kommuniké från NOTEs årsstämma den 25 april 2007 | https://www.note-ems.com/press-releases/kommunike-fran-notes-arsstamma-den-25-april-2007/ | Human resource |
| JABIL POLAND SP. Z O.O. | Carl Zeiss | Everything we do at Jabil Optics is founded in our legacy of delivering optical solution excellence. Dating back to 2006 and our initial partnership with Carl Zeiss, strategic investments have made Jabil Optics a world-class optics design and manufacturing partner. | Jabil Optics \| Jabil | Strategic collaboration |

# References


Adams, R., Bessant, J. and Phelps, R. (2006) "Innovation management measurement: A review," *International Journal of Management Reviews*, 8(1), pp. 21–47. doi:10.1111/j.1468-2370.2006.00119.x.

Adner, R. and Levinthal, D. (2001) "Demand heterogeneity and technology evolution: Implications for product and process innovation," *Management Science*, 47(5), pp. 611–628. doi:10.1287/mnsc.47.5.611.10482.

Annarelli, A. *et al.* (2021) "Literature review on digitalization capabilities: Co-citation analysis of antecedents, conceptualization and consequences," *Technological Forecasting and Social Change*, 166, p. 120635. doi:10.1016/J.TECHFORE.2021.120635.

Archibugi, D. and Planta, M. (1996) "Measuring technological change through patents and innovation surveys," *Technovation*, 16(9), pp. 451–468.

Arora, S.K. *et al.* (2020) "Measuring dynamic capabilities in new ventures: exploring strategic change in US green goods manufacturing using website data," *Journal of Technology Transfer*, 45(5), pp. 1451–1480. doi:10.1007/S10961-019-09751-Y.

Ashouri, S., Suominen, A., Hajikhani, A., Pukelis, L., Schubert, T., Türkeli, S., Van Beers, C., *et al.* (2021) "Indicators on firm level innovation activities from web scraped data," *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.3938767.

Ashouri, S., Suominen, A., Hajikhani, A., Pukelis, L., Schubert, T., Türkeli, S., van Beers, C., *et al.* (2021) "Indicators on firm level innovation activities from web scraped data," *SSRN Electronic Journal* [Preprint]. doi:10.2139/SSRN.3938767.

Axenbeck, J. and Breithaupt, P. (2021) "Innovation indicators based on firm websites — Which website characteristics predict firm-level innovation activity?," *PLoS ONE*, 16(4 April). doi:10.1371/JOURNAL.PONE.0249583.

Bernstein, A., Provost, F. and Clearwater, S. (2003) "The Relational Vector-space Model and Industry Classification.," *International Joint Conference on Artificial Intelligence Workshop on Statistical Learning from Relational Models* [Preprint].

Björkdahl, J. (2020) "Strategies for Digitalization in Manufacturing Firms:," *https://doi.org/10.1177/0008125620920349*, 62(4), pp. 17–36. doi:10.1177/0008125620920349.

Calel, R. and Dechezleprêtre, A. (2012) "Environmental Policy and Directed Technological Change: Evidence from the European Carbon Market," *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.2024870.

Chern, A. *et al.* (2018) "Automatically Detecting Errors in Employer Industry Classification Using Job Postings," *Data Science and Engineering*, 3(3), pp. 221–231. doi:10.1007/s41019-018-0071-7.

Crespi, F. (2013) "Environmental policy and induced technological change in European industries," *The Dynamics of Environmental and Economic Systems: Innovation, Environmental Policy and Competitiveness*, 9789400750, pp. 143–157. doi:10.1007/978-94-007-5089-0_8.

Drnevich, P.L. and Croson, D.C. (2013) "Information technology and business-level strategy: Toward an integrated theoretical perspective," *MIS quarterly*, pp. 483–509.

Frietsch, R. *et al.* (2017) *Final Report on the Collection of Patents and Business Indicators by Economic Sector: Societal Grand Challenges and Key Enabling Technologies Collection and studies*.

Gagliardi, L., Marin, G. and Miriello, C. (2016) "The greener the better? Job creation effects of environmentally-friendly technological change," *Industrial and Corporate Change*, 25(5), pp. 779–807. doi:10.1093/icc/dtv054.

Gitto, S. (2017) "Efficiency change, technological change and capital accumulation in Italian regions: a sectoral study," *International Review of Applied Economics*, 31(2), pp. 191–207. doi:10.1080/02692171.2016.1240152.

Gobble, M.M. (2018) "Digitalization, Digitization, and Innovation," *Research-Technology Management*, 61(4), pp. 56–59. doi:10.1080/08956308.2018.1471280.

Goindani, M. *et al.* (2017) "Employer Industry Classification Using Job Postings," *IEEE International Conference on Data Mining Workshops, ICDMW*, 2017-Novem, pp. 183–188. doi:10.1109/ICDMW.2017.30.

Gök, A., Waterworth, A. and Shapira, P. (2015) "Use of web mining in studying innovation," *Scientometrics*,


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

102(1), pp. 653–671. doi:10.1007/S11192-014-1434-0/TABLES/5.
Hagedoorn, J. and Cloodt, M. (2003) “Measuring innovative performance: is there an advantage in using multiple indicators?,” *Research Policy*, 32(8), pp. 1365–1379. doi:10.1016/S0048-7333(02)00137-3.
Hajikhani, A. *et al.* (2021) *Connecting firm’s web scrapped textual content to body of science: Utilizing Microsoft Academic Graph hierarchical topic modeling*.
Kee, T. (2019) “Peer Firm Identification Using Word Embeddings,” *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, pp. 5536–5543. doi:10.1109/BigData47090.2019.9006438.
Kinne, J. and Axenbeck, J. (2020) “Web mining for innovation ecosystem mapping: a framework and a large-scale pilot study,” *Scientometrics*, 125(3), pp. 2011–2041. doi:10.1007/S11192-020-03726-9.
Kohtamäki, M. *et al.* (2020) “The relationship between digitalization and servitization: The role of servitization in capturing the financial potential of digitalization,” *Technological Forecasting and Social Change*, 151, p. 119804. doi:10.1016/J.TECHFORE.2019.119804.
Lamby, M. and Isemann, D. (2018) “Classifying companies by industry using word embeddings,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10859 LNCS, pp. 377–388. doi:10.1007/978-3-319-91947-8_39.
Leiponen, A. and Drejer, I. (2007) “What exactly are technological regimes?. Intra-industry heterogeneity in the organization of innovation activities,” *Research Policy*, 36(8), pp. 1221–1238. doi:10.1016/j.respol.2007.04.008.
Lyocsa, S. and Vyrost, T. (2011) “Industry Classification: Review, Hurdles and Methodologies,” *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.1480563.
Mishra, A.N., Konana, P. and Barua, A. (2007) “Antecedents and consequences of internet use in procurement: an empirical investigation of US manufacturing firms,” *Information systems research*, 18(1), pp. 103–120.
Neuhäusler, P. *et al.* (2017) *Identifying the Technology Profiles of R&D Performing Firms - A Matching of R&D and Patent Data*, *International Journal of Innovation and Technology Management*. doi:10.1142/S021987701740003X.
Neuhäusler, P., Frietsch, R. and Kroll, H. (2019) “Probabilistic concordance schemes for the re-assignment of patents to economic sectors and scientific publications to technology fields,” *Fraunhofer ISI Discussion Papers - Innovation Systems and Policy Analysis*, 60, p. 38.
Nylén, D. and Holmström, J. (2015) “Digital innovation strategy: A framework for diagnosing and improving digital product and service innovation,” *Business Horizons*, 58(1), pp. 57–67.
OECD (2015) *The Future of Productivity*, *The Future of Productivity*. OECD. doi:10.1787/9789264248533-en.
Pant, G. and Sheng, O.R.L. (2015) “Web footprints of firms: Using online isomorphism for competitor identification,” *Information Systems Research*, 26(1), pp. 188–209. doi:10.1287/isre.2014.0563.
Sambamurthy, V., Bharadwaj, A. and Grover, V. (2003) “Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms,” *MIS quarterly*, pp. 237–263.
Schmoch, U. *et al.* (2003) *Linking Technology Areas to Industrial Sectors*, *Final Report to the European Commision, DG Research*.
Selander, L., Henfridsson, O. and Svahn, F. (2013) “Capability search and redeem across digital ecosystems,” *Journal of information technology*, 28(3), pp. 183–197.
Tripsas, M. and Gavetti, G. (2000) “Capabilities, cognition, and inertia: Evidence from digital imaging,” *Strategic management journal*, 21(10-11), pp. 1147–1161.
Wheeler, B.C. (2002) “NEBIC: A dynamic capabilities theory for assessing net-enablement,” *Information systems research*, 13(2), pp. 125–146.
Zhang, W. *et al.* (2012) “Semi-supervised text mining for dynamic business network discovery,” *Proceedings - Pacific Asia Conference on Information Systems, PACIS 2012* [Preprint].
Zhou, Z., Mu, X. and Lin, X. (2021) “Constructing economic taxonomy reflecting firm relationships based on news reports,” *Data Technologies and Applications* [Preprint]. doi:10.1108/DTA-11-2020-0287.

# For more information, please contact

Dr. Arho Suominen (Consortium leader)
Tel. +358 50 5050 354
arho.suominen@vtt.fi

# About BIGPROD

BIFPROD is a research project focusing on Big Data based analysis of productivity using webscraped data. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

The project partners in the project are Quantitative Science and Technology Studies team, Foresight-driven Business Strategies, 1) VTT Technical Research Centre of Finland, Competence Center Innovation and Knowledge Economy (Coordinator), 2) Fraunhofer ISI, Economics of Knowledge and Innovation team, 3) UNU-MERIT, Maastricht University, 4) Public Policy and Management Institute, 5) Economics of Technology and Innovations, Faculty of Technology, Policy and Management, 6) Delft University of Technology, Economics of Technology and Innovations, 7) Faculty of Technology, Policy and Management, Delft University of Technology

www.bigprod.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822







This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822







This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822

# For more information, please contact

Dr. Arho Suominen (Consortium leader)
Tel. +358 50 5050 354
arho.suominen@vtt.fi

# About BIGPROD

BIFPROD is a research project focusing on Big Data based analysis of productivity using webscraped data. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870822.

The project partners in the project are Quantitative Science and Technology Studies team, Foresight-driven Business Strategies, 1) VTT Technical Research Centre of Finland, Competence Center Innovation and Knowledge Economy (Coordinator), 2) Fraunhofer ISI, Economics of Knowledge and Innovation team, 3) UNU-MERIT, Maastricht University, 4) Public Policy and Management Institute, 5) Economics of Technology and Innovations, Faculty of Technology, Policy and Management, 6) Delft University of Technology, Economics of Technology and Innovations, 7) Faculty of Technology, Policy and Management, Delft University of Technology

www.bigprod.eu