



European
Commission

CONCORDi 2021

Industrial innovation for competitive sustainability

22 - 25 November 2021

Virtual conference



Co-organised with



UNITED NATIONS
INDUSTRIAL DEVELOPMENT ORGANIZATION



Collaborative ecosystems between research organizations and companies from web scraped data: a vantage point to proximity of collaboration

Arho Suominen
VTT Technical Research
Centre of Finland
Tampere University, Tampere, Finland
arho.suominen@vtt.fi

Arash Hajikhani
VTT Technical Research
Centre of Finland
arash.hajikhani@vtt.fi

Lukas Pukelis
PPMI
lukas.pukelis@ppmi.lt

1 Project details and disclaimer

- This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 870822. <https://cordis.europa.eu/project/id/870822>



- A version of the data platform description has been published separately as a project deliverable (Pukelis 2020)
- A previous version of the paper has been presented as an oral presentation at the Global Tech Mining conference. (Suominen 2021)
- The work presented is in progress, with new data being available as we continue the work.



Collaboration

- Literature has extensively looked at the importance of collaboration and in particular the collaboration between industrial actors and research organization (Cohen 2002; Suominen 2018)
- Particularly ecosystems have seen a lot of research
- Methods to operationalize collaboration include, but are surely not limited to
 - Joint patenting (Petruzzelli 2009)
 - Research grants from industry to research organizations (D'Este et al. 2013)
 - Survey data (e.g. Community Innovation Survey) (Kobarg et al 2013)
 - Co-publishing between academia and industry (Abramo et al. 2009)

*“The relevance of industrial innovation to contribute to the multiple transitions mentioned above and sustainability objectives are reflected in recent key EU policy documents, such as the September 2020 renewed European Research Area (ERA) Communication, the May 2021 updated Industrial Strategy for Europe, and the 2030 Digital Compass. The updated Industrial Strategy raises EU industrial and technological vulnerabilities, and implements transition pathways for some of the identified industrial ecosystems. The renewed **European Research Area (ERA) strategy implements common industrial technology roadmaps to transfer R&I results into the economy through investment agendas developed among the stakeholders, supports industrial alliances to mobilise industrial capacities in key areas, as well as Horizon Europe partnerships with industry.**” (Diodato et al. 2021, emphasis added)*





Extending measures

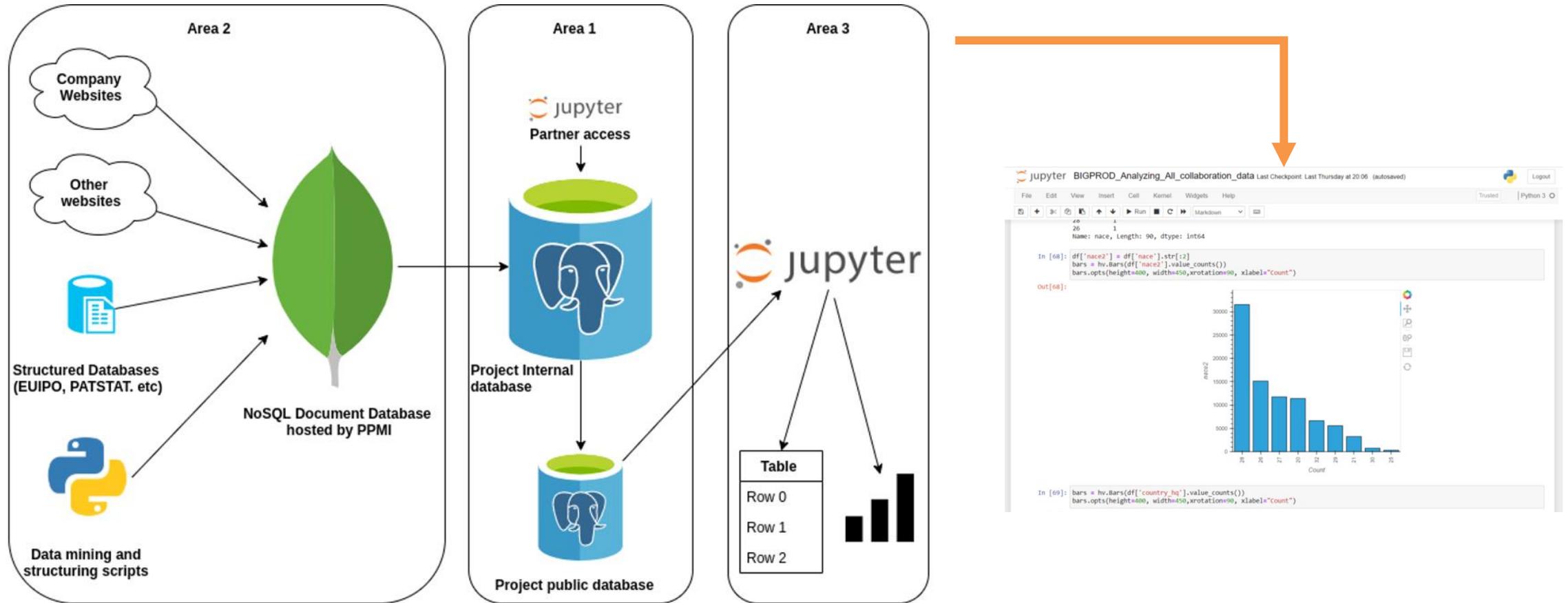
- What we are proposing here is an additional vantage point to measuring collaboration between industrial actors and between industrial actors and research organization.
- The measure is based on a web scraping a sample of medium-high and high-technology companies from EU and UK.
- The main motivation for the work is to offer additional measures to the partial views offered by existing measures.
- Making no claims of superiority we see that the webscraped data offers insights to:
 - Collaborative differences between different industries or by the size of companies
 - Deep analysis on a geospatial level, analyzing regional ecosystem and the importance of distance at scale-
 - Analyzing collaboration by thematic factors
 - Analyzing collaborative differences between industry-industry and industry-research organization

Sample

- Our analysis is based on a sample of 183,161 medium-high and high-tech companies in EU27+UK
 - Manufacture of basic pharmaceutical products and pharmaceutical preparations (21)
 - Manufacture of computer, electronic and optical products (26)
 - Manufacture of air and spacecraft and related machinery (30.3)
 - Manufacture of chemicals and chemical products (20)
 - Manufacture of weapons and ammunition (25.4)
 - Manufacture of electrical equipment (27)
 - Manufacture of machinery and equipment n.e.c. (28)
 - Manufacture of motor vehicles, trailers and semi-trailers (29)
 - Manufacture of other transport equipment (30) excluding Building of ships and boats (30.1) and excluding Manufacture of air and spacecraft and related machinery (30.3)
 - Manufacture of medical and dental instruments and supplies (32.5)



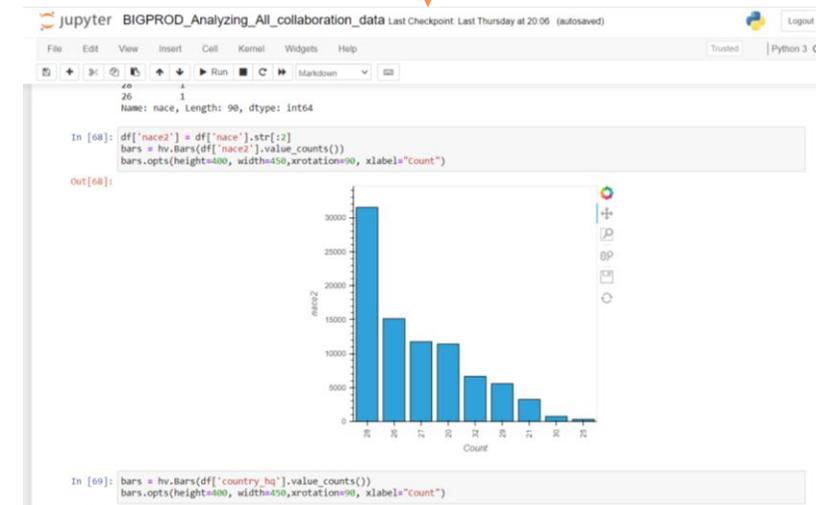
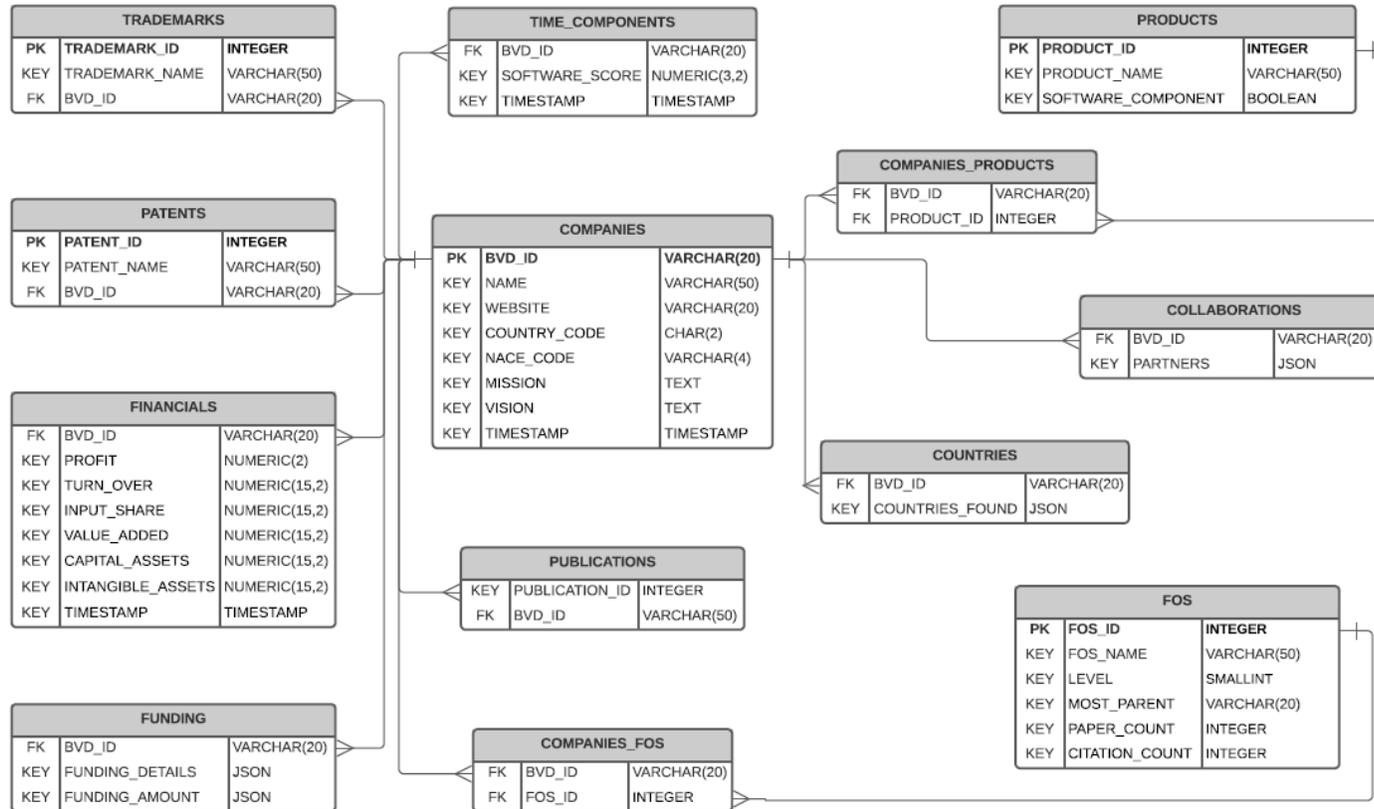
Data platform



Source: Pukelis 2020

Data platform

linguistic dependency parsing, entity recognition mechanisms with machine-learning to identify collaborations



Variable creation in Jupyter Notebook. Geospatial information created using Google Maps API

We define bipartite network $G = (U, V, E)$, where U are companies, V research organizations and E connections between the organizations

Results

Company	Collaborator	Text	URL	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 (Trial with ANA that Lufthansa is in venture partnership)



Results

- Analyzing 25 232 medium-high and high-technology companies
 - 2 455 research organizations and 1 100 unique companies in the collaboration graph (4.4 % have connections with research organizations.)
 - The average degree of the network is 1,704, which means that most organizations have a relatively low number of connections.
 - The network diameter is 12 reflecting that there is a path within the network, rather than being created by isolate links.
 - Filtering out all weakly connected nodes, a sub graph with 3138 and nodes and 5808 edges emerges.
 - This sub graph has an average degree of 3,702 and 26 communities.

Results

- There are particularly strong links between European companies and US research organizations.
- Industries have some differences in the amount of connection between research organizations. From the industry sectors,
 - NACE 25 (contains Manufacture of weapons and ammunition (NACE 25.4)) has the lowest number of connections and
 - NACE 32 (contains Manufacture of medical and dental instruments and supplies (32.5)) has the highest number of connections.
- In the proximity of collaboration, we note that linkages between research organizations and companies seems relatively high.

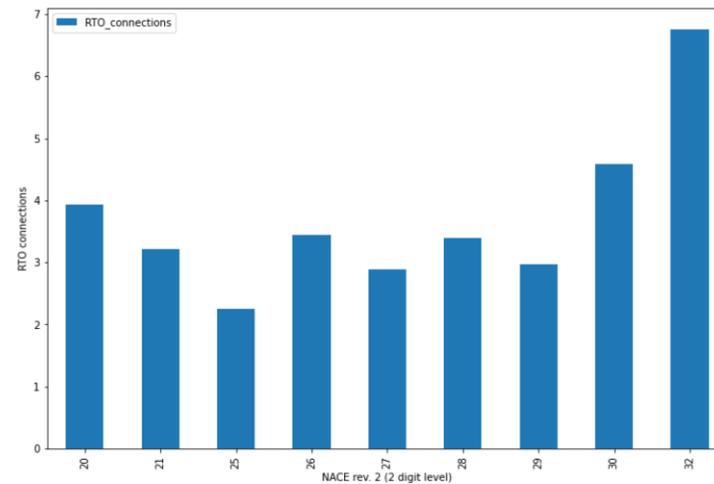
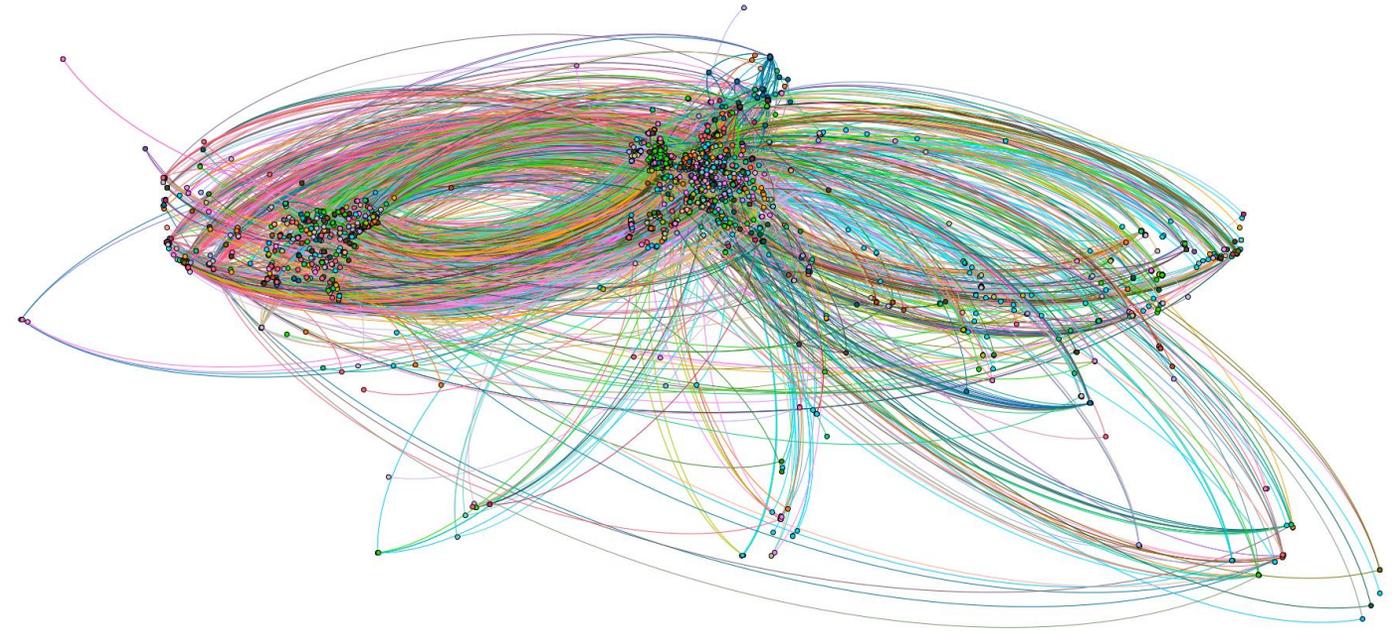


Figure 7. Average number of RTO connections in the sample.

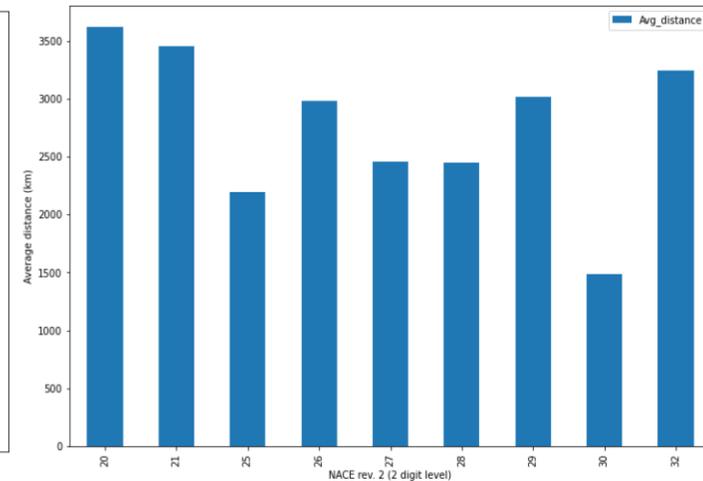


Figure 8. Average distance (km) of connections in the sample.



Conclusions

- **Method:** The work demonstrated a methodological procedure for utilizing a novel data source in the form of textual content from companies website to approximate collaboration.
- **Extension:** The study found that roughly five percent of our sample had mentions of research organizations on their webpage. From the ones mentioning research organizations, most of the companies had multiple research organization partners.
- **On proximity,** the data showed that in this sample, the average distance between the collaborators is relatively high. This questions the notion of proximity, which is often cited as being an important factor driving collaborations.
- **Limitations:** The work involves web scraping unstructured content from websites, which has the potential to omit data that would be relevant but also include false positives.

Keep in touch



BIGPROD website: <https://www.bigprod.eu/>



@ArhoSuominen

ResearchGate

BIGPROD project



<https://www.linkedin.com/in/arhosuominen/>

Pre-print available Ashouri, Sajad and Suominen, Arho and Hajikhani, Arash and Pukelis, Lukas and Schubert, Torben and Türkeli, Serdar and Van Beers, Cees and Cunningham, Scott, Indicators on firm level innovation activities from web scraped data (October 8, 2021). Available at SSRN: <https://ssrn.com/abstract=3938767> or <http://dx.doi.org/10.2139/ssrn.3938767>

Thank you!



- © European Union 2020
- Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.
- Slides 4-7 have illustrative pictures from Pixabay

References

- Abramo, G., D'Angelo, C. A., Di Costa, F., & Solazzi, M. (2009). University–industry collaboration in Italy: A bibliometric examination. *Technovation*, 29(6-7), 498-507.
- D'Este, P., Guy, F., & Iammarino, S. (2013). Shaping the formation of university–industry research collaborations: what type of proximity does really matter?. *Journal of economic geography*, 13(4), 537-558.
- Kobarg, S., Stumpf-Wollersheim, J., & Welpel, I. M. (2018). University-industry collaborations and product innovation performance: The moderating effects of absorptive capacity and innovation competencies. *The Journal of Technology Transfer*, 43(6), 1696-1724.
- Lundberg, J., Tomson, G., Lundkvist, I., Skerfving, J., & Brommels, M. (2006). Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics*, 69(3), 575-589.
- Pellens, M., Krieger, B., Blind, K., & Schubert, T. (2020). Are Firms Withdrawing From Basic Research? An Analysis of Firm-Level Publication Behaviour in Germany. *An Analysis of Firm-Level Publication Behaviour in Germany*, 20-083.
- Petruzzelli, A. M. (2011). The impact of technological relatedness, prior ties, and geographical distance on university–industry collaborations: A joint-patent analysis. *Technovation*, 31(7), 309-319.
- Pukelis, L. (2020) "D8 interim platform operation report,"BIGPROD deliverable
- Suominen, A., Hajikhani, A., & Pukelis, L (2021). Collaborative ecosystems between research organizations and companies from web scraped data: a vantage point to proximity of collaboration. *Concordi Conference 2021*.
- Suominen, A., Seppänen, M., & Dedehayir, O. (2018). A bibliometric review on innovation systems and ecosystems: a research agenda. *European Journal of Innovation Management*.
- Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). Links and impacts: the influence of public research on industrial R&D. *Management science*, 48(1), 1-23.