Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2022) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

Transport Research Arena (TRA) Conference

# The role of government and research organisations in the development of logistics networks as an integral area of physical internet

Skrickij Viktor[a,*], Šakalys Raimondas[a], Bazaras Darius[a], Šakalys Algirdas[a]

[a]*Faculty of Transport Engineering, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania*

**Abstract**

Digital technologies are fundamentally changing the development of economies and societies, affecting transport and logistics. Data and data-driven innovations are at the centre of this transformation. At the same time, the role of government and research organisations in the development of the physical internet is not sufficiently clear. This investigation has been performed to research the current limitations of cloud-based big data repositories, business attitudes toward sharing the data and identify priority areas for governmental and public research institutions. The study revealed the interest and desire of companies to cooperate while identifying the essential conditions for cooperation - data security, maintaining competitiveness, the mechanism of data preparation and provision, and possible pricing.

*Keywords:* Physical internet; data sharing; research institutions; governmental organisations; survey

## 1. Introduction

In recent years, digital technologies are fundamentally changing the development of economies and societies, affecting different sectors of activity, including transport and logistics. Data and data-driven innovations are at the centre of this transformation, and their importance will only increase in the future. Even more, the data-driven economy also changes the roles of market players. In terms of the logistics sector, there is a significant increase role of governmental and research organisations, especially in the development of the physical internet (PI) and databases containing supply chain data.

* Corresponding author. Tel.: +370 688 64981.
*E-mail address:*viktor.skrickij@vilniustech.lt


Peer-review under responsibility of the scientific committee of the Transport Research Arena (TRA) Conference

The PI as a new supply chain paradigm is probably the most ambitious concept for efficient and sustainable transportation of goods. It covers different technologies and areas where the main physical elements are load units (such as containers, pallets, and boxes), nodes and movers (vehicles and carriers) (Montreuil et al., 2010). The envisions of PI require a transition from logistics nodes and networks to PI nodes and PI networks, thus developing a network ecosystem with the possibility to access it for all supply chain stakeholders. For PI operation, governance rules and processes are needed that offer stakeholders transparency, security and control over their confidential data. In the development of logistics networks where governance processes, bodies and rules need to be defined, cooperation with governmental organisations (GO), authorities, and research institutions is recommended ( Ballot et al., 2020). At the same time, the role of government and research organisations in the development of logistics networks as an integral area of the PI is not defined from the perspective of digitalisation. This investigation has been performed to research current limitations of cloud-based big data repositories, business attitudes to share the data and identify priority areas for GO and public research institutions.

## 2. Literature review

The digitalisation of the processes is needed to implement the PI approach. Usually, digitalisation is understood as the process within the company which boosts its performance. It should be aligned with a clear innovation strategy that maximises effect (Fernández-Portillo et al., 2022). Digitalisation brings many challenges, such as the need for new IT architecture, data formats and protocols, data sharing policies, and data interchange tools. As a result, the weak implementation of digital processes and tools remains a challenge (Truant et al., 2021).

Combining supply chain big data with modern data processing techniques focusing on artificial intelligence (AI) may also provide significant opportunities for stakeholders. However, it is a very complex task as digital solutions must function between the companies, non-governmental organisations (NGOs), GO and authorities. NGOs and GO can use the supply chain data for statistics generation. Commonly cargo is transported worldwide, so GO plays an important role in data sharing/receiving with different regions, customs/border-crossing services, and taxation. Last but not least is the task of green supply chain management. Environmental issues commonly are self-regulated by companies. As a powerful stakeholder in the green supply chain, the GO can enforce the companies by influencing their internal and external resources (Nezakati et al., 2016). The GO, as a regulator, can provide incentives for green performing companies by subsidising or stimulating customers to buy green products, increasing the market share of these companies. At the same time, governments may impose taxes on companies that are not practising sustainability in their supply chain management (Clemens and Douglas, 2006; Nezakati et al., 2016).

The collaboration between stakeholders and research institutions can be used to create and implement digital innovations. Zinn and Goldsby (2017) analysed the role of researchers in supply chain practice. Previously there were many investigations in the field of optimisation and risk management tasks. However, there is a growing concern that today's business research, including supply chain research, strayed from the actual practice of management (Hoffman, 2015; Zinn and Goldsby, 2017). Currently, big data analytics and AI are hot topics in academia, and theoretically, new algorithms being under development could be used for supply chain practice. The actual situation is somewhat different. Scientists need the datasets to work with, and there are interested in developing new methods rather than solving applied tasks. There are a lot of valuable datasets in the field of medicine, automated driving and even car-sharing but not the supply chain (Ma et al., 2020; Song et al., 2022; Šabanovič et al., 2020).

Commonly datasets are available in registries, databases and repositories. Registries aim to collect comprehensive, uniform data on every case but are limited in depth and complexity. Databases are collections of local data that stakeholders might assemble for their own use. Repositories are multicentre, multimodal, well-designed and well-described databases. They collect consistent data but without the universal inclusion of registries. This often allows to collect much more detailed data but at the risk of bias (Wang and Williams 2022). Datasets are not uniform; they differ in terms of who holds the data and its nature. Commonly they are classified (Xafis and Labude 2019): *i)* Institutional; *ii)* Governmental; *iii)* Discipline-specific; *iv)* Generalist; *v)* Project/Program-specific; *vi)* Business.

There are no repositories with the supply chain datasets in open access; most sources are registries and databases. In Table 1, sources, where the datasets can be found are provided. Sources 1-6 are supply chain datasets. The majority of datasets can be downloaded for free; in some cases, registration is required. The main disadvantage is that most

data is older than two years old. The usefulness of data depends much on the task it is used for. It doesn't matter if it is used to develop methods, algorithms and tools. The lack of the newest data is a problem if the data must be used in an operational process (or to make a short-term prognosis). Source 7-8 provides economic data, which can be used with transportation data. Source 9 and 10 are examples of single datasets. The last one was created during ICONET project.

Existing datasets are available in different formats; however, there are no problems opening and processing the data in most cases.

Table 1. Sources of Datasets

| | Name | Source |
|---|---|---|
| 1 | Open data Europe | https://data.europa.eu/ |
| 2 | The Cloud-Native Data Catalog | https://data.world/datasets/transportation |
| 3 | US Department of transportation | https://data.transportation.gov/ |
| 4 | Eurostat | https://ec.europa.eu/eurostat/data/database |
| 5 | Data Hub | https://datahub.io/collections/logistics-data |
| 6 | Datarade | https://datarade.ai/data-categories/transport-logistics-data |
| 7 | Statista | https://www.statista.com/ |
| 8 | Trading Economics | https://tradingeconomics.com/ |
| 9 | Supply Chain Logistics Problem Dataset | https://brunel.figshare.com/articles/dataset/Supply_Chain_Logistics_Problem_Dataset/7558679 |
| 10 | ICONET datasets | https://doi.org/10.5281/zenodo.4564474 |

A review of data sources (Table 1) that provide datasets in the supply chain showed that there are no recent data. Often the data is fragmented; it is unclear if it will be updated and when. In such a case, the number of tasks that can be solved using such sources is limited. From publications, it can be seen that the academic community focus mainly on tasks that allow them to test their scientific ideas rather than tasks relevant to the business. It is essential to understand actual business needs to strengthen the cooperation between business representatives, GO, and research institutions.

In Section 3, we provide survey results, where we asked business representatives working in the field of supply chain about their attitudes to sharing the data with GO and research institutions, indicating the conditions and purposes for which the data may be used.

## 3. Priority areas for new sources of structured data for GO and public research institutions

The survey was selected as the main investigation tool, and the business representatives were defined as a target group. The survey aimed to understand business attitudes to sharing data with research institutions and GO and define the main tasks where researchers should focus. Also, identify possible trends and select appropriate recommendations on how stakeholders should respond to data availability policies with business companies. Survey questions are presented in Table 2.

Respondents who participated in the study are experts in the field of supply chain management and are familiar with digitalisation issues. Two groups of respondents were identified as experts. The first group was Horizon2020 project ePicenter partners and companies they work with. The project explores the potential of artificial intelligence, digitisation, automation, and innovation in freight transportation and handling technologies to create powerful solutions that enable the development of resilient, more efficient, and greener supply chains. The second group was experts who work for the largest transportation companies that are not project partners, but their role in the global supply chain is important. Surveys were sent to 40 experts from both groups. Totally 28 surveys were collected. The seven submitted questionnaires were rejected; as no business representatives filled part of them; another part of the questionnaires contained inconsistencies between the answers. As a result, 21 surveys were analysed, 13 from the first group and 8 from the second. Survey results are presented below in a graphical form.

Table 2. Survey questions

| Nr | Question |
|---|---|
| 1 | Please specify the transport node/company you represent. |
| 2 | Please specify your role in the company. |
| 3 | List the transport modes your company usually work with. Please specify the frequency. |
| 4 | How is BIG DATA used in your company? |
| 5 | Are you sharing BIG DATA with GO, authorities, and research institutions? |
| 6 | If you are not sharing the data, what are the main reasons? Please specify the importance. |
| 7 | If you are sharing (or could share) data with the GO and authorities, please select areas where this data is used (or could be used) in your opinion and score. |
| 8 | If you are sharing (or could share) data with the research institutions, please provide the main tasks of how this data is used (or could be used) in your opinion and score. |
| 9 | Provide your expectations regarding the tasks on which researchers should focus. Please specify the importance. |
| 10 | Provide the terms under which you would share the data with research organisations. |
| 11 | Define appropriate pricing mechanisms for data sharing with research institutions. Please specify possibilities. |
| 12 | If you want to receive survey results, please provide your email. |

### *3.1. Survey results and analysis*

The experts involved in the study represent the full range of supply chain actors; the distribution is presented in Figure 1. Respondents who selected option other represent software company, transport management software company; system integrator; visibility/planning platform, producer & shipper/final buyer of the freight; port authority, consulting company; logistics and marine transport representatives.

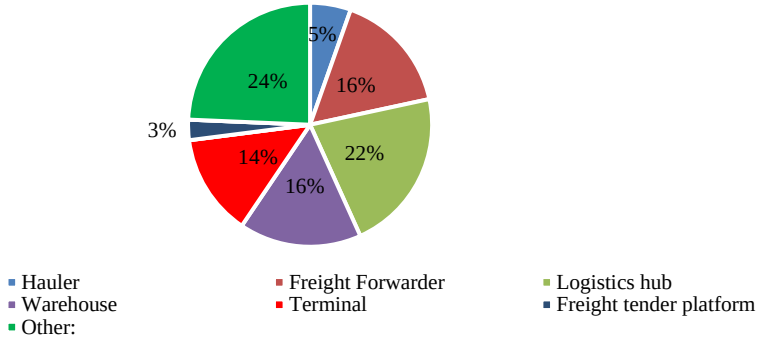


Fig. 1. The transport node/company respondent represent

The majority of the respondents were managers/data users (57%); data collectors/ providers were 14%; data analysts (scientist, engineer, architect) were 14%; database administrators, 5%. Respondents who selected option other (10%) defined their roles as data sharing provider, manager & data collector/provider.

Respondents were asked to provide the transport modes their company usually works with and specify the frequency. Four transport modes were provided as an option; each respondent could choose several modes. Results are presented in Table 3. Results demonstrate that the modal split reflects the existing situation in a market dominated by road transport. Only 5% of respondents answered that big data is not used in their companies. The rest choose the main task on how data is used: 41% of respondents use data for operational tasks, 36% for strategic, and 18% for tactical tasks. It was an option to select several answers.

The next question was about data sharing with GO and research institutions. Respondents were able to choose several options and answer this question. As a result, 13 respondents noted that they share data with GO (44% of all

answers); 6 respondents share data with research institutions (20% of all answers); 4 respondents noted that they are receiving data (13%), and 7 respondents (23% of all answers) do not share the data. There was an option to select "other", but no one chose such an option.

Table 3. Transport modes your company usually work with

| | Road transport | Railway transport | Air transport | Maritime transport |
|---|---|---|---|---|
| Average value | 8.4 | 6.5 | 3.6 | 7.9 |
| Rank | 1 | 3 | 4 | 2 |

The average value, standard deviation, and rank are used as metrics for the next questions. Rank is higher when the average value is higher. If the average values are the same for several answers, the answer with a lower standard deviation is most important. The higher value of standard deviations shows that this question is not important for some respondents and very important for others. If the standard deviation is rather small, it means that all the respondents are unanimous on this question.

Seven respondents mentioned that they were not sharing the data. From Table 4, it may be seen that the main reason why it happens is the lack of trust that data will be used properly. Also, a very important aspect is that companies do not generate/operate with data. The high costs of data preparation/ administration/ maintenance are in third place. Fourth place, cannot see benefits from data sharing, and the least important is technical challenges. It can be seen that the difference between all the proposed reasons is rather small, as average values vary in a narrow range. There was an option to provide additional reasons, and two respondents from seven used this option. The first mention that the data-sharing service is in a preparation (pilot) phase; the second notes that data is confidential and is used only internally for decision making.

In Table 5, the respondents who are sharing the data with GO ranked areas where data is or could be used. Rank 1 goes for economic statistical data generation. Respondents noted that the second most important is data sharing to develop digital tools. Rank 3 for end-to-end supply chain connectivity. Rank 4 goes for customs/border crossing services. It is important to point out that all the answer options are quite equivalent. The highest average value is 7.4; the lowest is 5.9.

Table 4. What are the main reasons not to share the data

| | Do not generate/operate BIG DATA | Technical challenges | High costs of data preparation/ administration/ maintenance | Lack of trust that data will be used in a proper manner | 'Can't see benefits from data sharing |
|---|---|---|---|---|---|
| Average value | 5.6 | 5.1 | 5.4 | 6.1 | 5.3 |
| Standard deviation | 3.2 | 2.7 | 3.1 | 3.3 | 3.8 |
| Rank | 2 | 5 | 3 | 1 | 4 |

Table 5. Areas where shared data can be used by governmental organisations and authorities

| | Data sharing/receiving with different regions all over the world | For economic statistics data generation | For digital tools development | For customs/border crossing services | For end-to-end supply chain connectivity | For taxation | For green supply chain management |
|---|---|---|---|---|---|---|---|
| Average value | 5.9 | 7.4 | 7.4 | 7.0 | 7.3 | 6.6 | 6.6 |
| Standard deviation | 2.8 | 2.7 | 3.0 | 3.3 | 2.7 | 3.3 | 2.9 |
| Rank | 7 | 1 | 2 | 4 | 3 | 6 | 5 |

In Table 6, the respondents who are sharing the data with research institutions ranked areas where data is or could be used. Most respondents selected it as the most crucial option for student teaching (Rank 1, average value 7.5).

Second place was the option for common research (Rank 2, average value 7.0). Rank three goes for company staff training.

Regarding data analysis provided in Table 7, the most important tasks are real-time (near real-time) data analytics, data exchange with the supply chain actors, and carbon footprint monitoring. The least important for business were two tasks: investigating big data's economic and societal impact and PI's economic and societal impact. However, the last one was mentioned as one of the most important for researchers in the Roadmap to Physical Internet prepared by the Alliance for Logistics Innovation through Collaboration in Europe, and it may be very important for making political decisions. Two respondents selected option other as well, one without providing any further details, second noted intermodality (mode switch in real-time).

Table 6. Areas where the research institutions can use shared data

| | For student teaching | For your company staff training | For scientific research without involving your company | For common research involving your company |
|---|---|---|---|---|
| Average value | 7.5 | 6.8 | 5.4 | 7.0 |
| Standard deviation | 3.0 | 2.9 | 3.3 | 2.7 |
| Rank | 1 | 3 | 4 | 2 |

Table 7. The tasks on which researchers should focus

| Task | Average value | Standard deviation | Rank |
|---|---|---|---|
| Traffic, timetable, mapping, routing | 8.0 | 2.1 | 6 |
| Tracking, monitoring, determining the availability of transport/warehouse | 8.0 | 1.4 | 5 |
| Carrier assignment, shipping order, warehouse booking | 7.1 | 2.2 | 11 |
| Customs procedures, CMR/e-CMR, other documentation | 8.0 | 2.3 | 7 |
| Coordination loading/unloading and changing transportation node | 8.1 | 2.4 | 4 |
| Management of assets, goods, clients | 7.7 | 1.9 | 9 |
| Optimised billing process | 7.6 | 2.6 | 10 |
| Data exchange with the supply chain actors | 8.9 | 1.4 | 2 |
| Carbon footprint monitoring, environmental challenges | 8.6 | 1.9 | 3 |
| Real-time (near real-time) data analytics | 8.9 | 1.1 | 1 |
| Price and demand forecasting | 7.9 | 2.1 | 8 |
| Investigate the economical and societal impact of Big Data | 6.6 | 2.6 | 12 |
| Investigate the economic and societal impact of Physical Internet | 6.6 | 2.3 | 13 |

The terms under which the company would share the data with research organisations. All the results are presented in Figure 2. Four respondents marked that data is available in open access, and ten noted that datasets are available or could be generated on request. Four respondents noted that in their companies, datasets are not generated, but they support the idea of data sharing. It can be seen that only three of twenty-one respondents noted that research institutions could not get access to the data. It can be seen that confidentiality is important for companies; twelve of twenty-one marked this answer. Regarding strategic, tactical and operational tasks, the highest rank goes to strategic tasks; second place for the operational and tactical tasks is less important and was selected only by three respondents.

As it can be seen from previous answers, respondents accept the possibility of sharing the data with research institutions. The most popular answer to this question was "data should be shared with minimal fees, including data preparation/administration/maintenance and other necessary costs". Rank 2 goes to answer that data between companies and research institutions could be shared for free. This data is presented in Table 8.

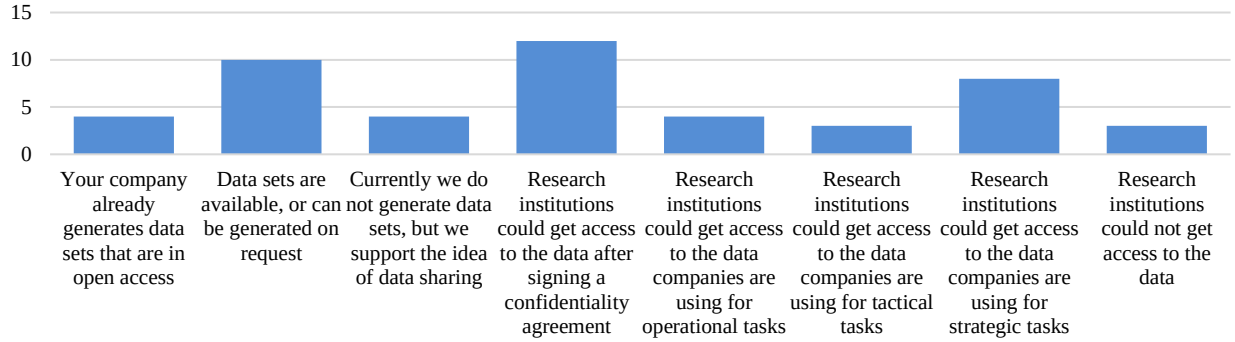


Fig. 2. Terms under which stakeholders would share the data with research organisations

Table 8. Pricing mechanisms for data sharing with research institutions

| | Data between your company and research institutions could be shared for free | Data could be shared with minimal fees that include only data preparation/ administration/ maintenance and other necessary costs | A one-time fee could be adopted for permanent access | Fixed fee for could be adopted annual access | Different fee planes could be adopted based on data type usage and the number of data needed | A post-paid monthly plan could be adopted, based on the number of issued documents per period |
|---|---|---|---|---|---|---|
| Average value | 6.8 | 6.9 | 3.5 | 3.7 | 3.6 | 3.3 |
| Standard deviation | 2.9 | 3.1 | 3.1 | 2.9 | 2.9 | 2.7 |
| Rank | 2 | 1 | 5 | 3 | 4 | 6 |

## 4. Discussion and conclusions

During the research, the obstacles, expectations and tendencies of the data exchange process between supply chain companies on the one side and GO's and research organisations on the otherwhere assessed, and the current situation was analysed from the perspective of business representatives.

Several key conclusions can be drawn from analysing the distribution of expert responses. Most experts confirmed that big data is an important topic in their work and are directly exposed to it. According to the preliminary provisions, the fact that Big Data is used for strategic purposes has been confirmed - it is likely to create business development plans, strategies and other long-term documents. However, experts pointed out that Big Data is used for operational purposes in the first place.

The issue of trust in data sharing is very important for businesses and can be explained in several circumstances. Transport and logistics companies operate in a highly competitive environment where information management, availability, and exchange are perceived as a potential loss of competitive advantage. In this context, a separate area of activity could be formed - a data governance model, including agreements and rules.

Understandably, businesses analyse tasks in a practical and utilitarian way. Therefore, it is logical that businesses see the role of GO and institutions in generating economic statistics and developing digital tools. Economic statistics would allow companies better to understand market trends, developments, and business prospects.

In collaboration with researchers, businesses pursue specific and practical goals, one of which is future staff training. The priority of teaching students can be understood in several aspects - first of all; it is a purposeful opportunity to influence the study process by providing relevant information to the study process, which creates favourable conditions for the purposeful training of specialists. Secondly, contact with students is an opportunity to update the activities of companies and form a circle of future clients or partners for the future. The second choice of experts in the ranking is the joint research in which the company is involved. It is also a matter-of-fact approach to collaboration, which would benefit both researchers and the business sector and allow it to address relevant operational

challenges or develop new technological or strategic solutions.

There is a need to highlight and further analyse the priority tasks that researchers could perform, such as real-time data analysis, data exchange with supply chain actors, and monitoring of carbon footprint and the use of environmental resources. The first two priorities are directly related to logistics supply chain management challenges; practice and theoretical research show that it remains a target. A functioning logistics supply chain comprising suppliers, manufacturers, distributors and traders generate a huge variety of real-time data. The successful operation of the logistics supply chain in delivering materials and products by synchronising time and space factors depends on the correct and efficient processing of real-time data. One of the biggest challenges in system management is forecasting and optimal use of resources, which is why real-time data analytics is relevant. In this survey, businesses pointed to the task of data exchange between the participants in the logistics supply chain, which has long been emphasised at the theoretical level.

Companies can generate datasets; however, the majority of them can only be generated on-demand, so there is no guarantee that the datasets will be systematised appropriately. There are no clear leaders in the field of compilation of datasets, except for the transport companies under state influence (e.g. railways and some airlines), and the confidentiality of the data will remain a factor for a long time. The approach to data sensitivity is also reflected in the fact that businesses prioritise data related to companies' strategic objectives, which could be presented for research. As a rule, strategic objectives underestimate the real-time data factor, with data being aggregated, depersonalised and reflecting trends and outcomes rather than processes.

The pricing mechanism proposed by the experts reaffirms the view that companies are willing to cooperate and do not see any fundamental problem with the need to charge researchers extra for the data submission process. The desire to have fixed prices for data preparation is quite understandable. However, on the other hand, it shows that companies still need additional administrative and time resources to prepare the datasets.

## Acknowledgements

This research was funded from the European Union Horizon 2020 Framework Program, under grant agreement no. 861584.

## References

Montreuil, B., Meller, R.D., Ballot, E., 2010. Towards a Physical Internet: the Impact on Logistics Facilities and Material Handling Systems Design and Innovation.

Ballot, E., Barbarino, S., van Bree, B., Liesa, F., Franklin, J. R., Hooft, D., Nettsträter, A., Paganelli, P., Tavasszy, L.A. Roadmap to the Physical Internet ALICE-ETP, 2020, p.31.

Fernández-Portillo, A., Almodóvar-González, M., Sánchez-Escobedo, M. C., Coca-Pérez, J. L., 2022. The Role of Innovation in the Relationship Between Digitalisation and Economic and Financial Performance. European Research on Management and Business Economics 28.3, 100190.

Truant, E., Broccardo, L., Dana, L.P., 2021. Digitalisation Boosts Company Performance: An Overview of Italian Listed Companies. Technological Forecasting and Social Change 173, 2021, 121173.

Nezakati, H., Fereidouni, M.A., Rahman, A.A., 2016. An Evaluation of Government Role in Green Supply Chain Management through Theories. International Journal of Economics and Financial Issues 6.6S, 76–79.

Clemens, B., Douglas, T., 2006. Does Coercion Drive Firms to Adopt 'Voluntary' Green Initiatives? Relationships Among Coercion, Superior Firm Resources, and Voluntary Green Initiatives. Journal of Business Research, 59.4, 483–491.

Zinn, W., Goldsby, T.J., 2017. The Role of Academic Research in Supply Chain Practice: How Much are We Contributing? Journal of Business Logistics 38.4, 236–237.

Hoffman, A.J., 2015. Isolated Scholars: Making Bricks, Not Shaping Policy. Chronicle of Higher Education 13. A48

Chaochao Ma, C., Wang, X., Wu, J., Cheng, X., Xia, L., Xue, F., Qiu, L., 2020. Real-World Big-Data Studies in Laboratory Medicine: Current Status, Application, and Future Considerations. Clinical Biochemistry 84, 21–30.

Song, H., Jo, K., Cho, J., Son, Y., Kim, C., Han, K., 2022. A Training Dataset for Semantic Segmentation of Urban Point Cloud Map for Intelligent Vehicles. ISPRS Journal of Photogrammetry and Remote Sensing 187, 159–170.

Šabanovič, E., Žuraulis, V., Prentkovskis, O., Skrickij, V., 2020. Identification of Road-Surface Type Using Deep Neural Networks for Friction Coefficient Estimation. *Sensors* 20.3, 612.

Wang, J.W., Williams, M., 2022. Registries, Databases and Repositories for Developing Artificial Intelligence in Cancer Care. Clinical Oncology 34.2, 97–103.

Xafis, V., Labude, M., 2019. Openness in Big Data and Data Repositories. Asian Bioethics Review 11, 255–273.