# Towards Open Research Data in the Economics Discipline

Piotr Kasprzak[1*], Magdalena Szuflita-Żurawska[2]

[1] Faculty of Management and Economics, Department of Finance, Gdańsk University of Technology (79 Traugutta Street, Gdańsk, Poland)
[2] Gdańsk University of Technology Library, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: pkasprzak@zie.pg.gda.pl; ORCID: 0000-0002-5570-4000

## Abstract

Nowadays, Open Research Data, as one of the three pillars of Open Science (along with Open Access and Open Scholarly Communication), is gaining enormous attention from different academic and commercial environments. A wide range of scientific disciplines represent and produce different types of data and at the same time, gather different issues and problems in terms of sharing and dissemination research output. This chapter aims to briefly describe the current state of the scientific data for the business and economics disciplines. It presents the opportunities and concerns in those particular scientific fields where very often conducting research is supported not only by public funds but also by business and industry providers. Specific data types (such as financial data) need a particular approach to maintain them, especially in terms of the licensing, preservation and sharing openly. Their proper management can have a significant impact on the knowledge-based economy.

**Keywords:** Open Research Data, economics discipline, tax system, taxation, open data, tax information source, tax reporting

https://doi.org/10.34808/x55q-sz53_dyr_roz4

## Introduction

Open Science has an undoubted impact on conducting research. It has changed the scientific methods and practices and scholarly communication and dissemination of research findings at national and international levels. According to the Organization for Economic Co-operation and Development (OECD), Open Science includes free access

to scientific articles, access to data from publicly funded research, and collaboration in research activities enabled by ICT tools (OECD, 2015). Nowadays, Open Access to scholarly papers is relatively well established in developed countries; however, making research data publicly available is during in transition period. Generally, Open Research Data (ORD) refers to all scientific data that is freely accessible and can be reused and redistributed without restrictions.

Standard practices include a Data Management Plan (DMP) to grant application process (e.g. for grants from Polish National Science Centre) to maximize knowledge about scientific projects from the initial phase or during the research project (such as those financed by Horizon 2020 and Horizon Europe, assume the sending of revised versions of plans during the project). The creation of DMP as a "living document" provides a description and evidence of produced data results in making research as time-efficient and safe as possible in terms of losing it due to, e.g. technical mishaps. Moreover, making accessible the research data is progressively common within Publishing Houses as part of their Open Science policies. To improve the quality of shared data and optimize their reusability, the community associated with research data management activities has developed the best practices called FAIR principles (Findable, Accessible, Interoperable, Re-usable). Those principles have human researchers in mind, but they focus on enhancing the availability of computers and machines to retrieve and use data automatically.

All scientific disciplines and sub-disciplines are characterised by their scholarly communication practices and methods, and their traditions in sharing knowledge and academic culture. Sharing research data varies between different fields and is determined by several factors such as geographical, technical and methodological features (Akers and Doty, 2013). Economics is considered to be a rather homogeneous field represented together by empirical and theoretical areas of conducting research. For instance, according to Dawson and Rascoff (2006), applied economics strongly depends on finding data (datasets) from commercial providers, governmental agencies, and academia, and collecting it through fieldwork. Additionally, supporting initiatives such as the Open Economics project (https://openeconomics.net/projects/) – a collaboration of Open Knowledge Foundation and Cambridge University – affect the popularisation of good practices of ORD available for preservation, transparency, analysis and replication.

It has always been known that information is a valuable asset. We strive to know more and more, we often incur high expenses to obtain information. More often, however, it is possible to obtain free access to knowledge, e.g. under the so-called Open Research Data. Although access to open data is rather identified mainly with science, the publication of research results and dissemination of knowledge, i.e. popularising activity, it is worth noting that access to knowledge is also becoming a significant competitive advantage, especially for enterprises. Companies that want to achieve more than their competitors are looking for data and information about their clients, contractors, changes in the market, and other business trends. Access to information becomes a valuable unit of assets (Schwab et al., 2011), although, for example, in accounting, this concept has long been known as an element of so-called intangible assets.

Currently, there is a growing trend in promoting universal and free access to up-to-date, reliable knowledge, including in the field of economy, finance and management (Sobczyk, 2016). The concept of open access to data is of general public interest, although one can find a number of examples of stakeholder groups who, due to the need to incur significant expenditures, especially financial ones, are reluctant to share their knowledge and experience (Skrzypek, 2018).

Economics is the science of making choices, allocating resources in the process of managing, producing and producing. Economics is also a science that collects and organises reliable knowledge about management. It is difficult to imagine making rational allocation decisions optimising a specific goal without having an appropriate amount of information (Czarny, 2011).

Enterprises are more often starting to use information and knowledge as factors of business success. The development of enterprises, and even their existence on the market, increasingly depends on the information and knowledge at their disposal (Olak, 2011). A new type of capital appears in enterprises, which consists mainly of knowledge, but also "raw" data and information. The contemporary success of a company is built on the basis of good knowledge of the market and the possibility of predicting what its environment will be like. The task of the information system is to recognise weak signals from the environment, analyse them and assess their suitability for building a competitive advantage for the enterprise (Górecka, 2018).

The aforementioned fields of science and researchers activity is based on so-called Research Data. More often, this data is a common good that is available in digital form for every Internet user (Strzelczyk, 2017).

According to a number of definitions in the literature, research data includes:

- Text documents,
- Numerical data,
- Surveys and survey research results,
- Video and sound recordings,
- Algorithms and schemes,
- Test and simulation results,
- Samples.

It should be noted that in the case of enterprises, the most important data are that about the market, customers, and their resources and preferences (Sobczyk, 2016). A problem for many companies, however, is the multitude of data (their quantity), and therefore skilful filtering of only those that are necessary for a given company is needed. Currently, we have access to many databases, information websites, financial statements and economic reports. In addition, companies obtain information on a daily basis by conducting their own research (Sosnowski, 2019). According to the literature, the methods of obtaining information are divided into basic research methods and detailed methods.

The first involves collecting information available to everyone from various types of publications, reports, etc. These are simpler and more informative methods. The detailed methods, on the other hand, require more effort and work to be performed. However, the

information is deeper in details and provides greater benefit to the company, and it must be conducted in order to find the solution for a specific problem. Usually one leading technique and two or more complementary ones are chosen and therefore used. It all depends on the needs of the enterprise and the budget (Rzemieniak, 2012).

The main test methods are:

- **Observation** – an imperceptible, intentional and systematic way of perceiving the tested objects in their natural conditions.
- **Participant observation** – in which the researcher tries to enter the environment she/he is interested in and collect information from the inside.
- **Open observation** – an observation in which research objects are aware that they are the subject of a study, which may lead to falsification of information.
- **Tests** – which belong to a wider group of projection techniques aimed at assessing the personality, attitudes and motives of research objects.
- **Survey research** – a set of questions prepared for a group of respondents, taking place without the participation of the researcher. It is strictly defined by her/him in the form of a questionnaire, and the manner of its delivery and receipt is also developed.
- **Panel method** – i.e. researching the same group of people several times in order to collect information on changes in the formation of various market or industry events at a specific time.
- **Interview** – a conversation between the researcher and the object of interest, the purpose of which is to obtain the necessary information.
- **Experiment** – examines the phenomena caused by the researcher in the environment and the conditions specified by her/him.

However, it should be noted that companies most often conduct research for their own needs and do not share the information obtained with others. Below (Tab. 4.1), the main conclusions in the form of opportunities and advantages, as well as threats and disadvantages, resulting from the widespread sharing of obtained and processed data, both by enterprises and specialised units (government agencies, economic intelligence agencies, research institutions) are presented.

In addition to in-house research, including marketing, enterprises obtain information from many other sources. In order to obtain information, entities often use the services of specialised agencies, consulting companies or public opinion polling centres (Sobczyk, 2016).

It is also common for companies to use databases which, due to their specific characteristics, can be divided into the following repositories:

- institutional/national: e.g. Central Statistical Office, European Statistical Office (EUROSTAT), Central Banks, Data Archiving and Networked Services (multi-domain science portal of the Netherlands created by a coalition of scientific governmental organisations of this country),
- thematic: e.g. UK Data Archive (data repository from social sciences and humanities), SeaDataNet (international oceanographic project), ADS Social Data Archive,

- general: e.g. FigShare, World Data System (open data from various fields, created in 2010 by the International Scientific Organization International Council for Science), OpenDOAR, RepOD,
- paid, with the possibility of ordering personalised data, e.g. Amadeus – European Business Information, EMIS Database.

**Tab. 4.1**

Open data in economics – opportunities and advantages as well as threats and disadvantages

| Opportunities and Advantages | Threats and Disadvantages |
|---|---|
| – possibility of establishing cooperation, especially at an international level and between various industries,<br>– access to market data from around the world,<br>– inventing new and improving existing inventions,<br>– reducing the cost of obtaining data that is particularly difficult to access,<br>– disseminating good practices in business,<br>– non-compliance with certain processes by several companies at the same time (e.g. analysis of the same data by several companies at the same time),<br>– possibility of specialisation and establishing cooperation on the business-science line,<br>– ORD increase the efficiency of economic research and its diffusion<br>– ORD increase cross-sector collaboration<br>– cost savings (reducing direct costs of scientific findings for R&B sector),<br>– provides for long-term safe storage of data,<br>– increases citation rates and enhancing scientific reputation,<br>– sharing data creates a public good,<br>– enhancing and popularisation of using research methods,<br>– maximising accountability and knowledge generation. | – possibility of "usurpation" of data, their inappropriate citation or failure to provide information about the author at all,<br>– concerns of some stakeholder groups about losing the status quo or fear of losing market share,<br>– reluctance to share data for free, especially when obtaining the information was associated with incurring significant expenditure,<br>– reducing the importance of companies specialising in the commercial collection and analysis of information,<br>– fear of access to enterprise data by government agencies and the tax administration,<br>– problems related to patenting inventions and scientific discoveries – making information available earlier may harm the patenting process,<br>– contractual obligations may prohibit data sharing,<br>– very complex data value cycle in economics<br>– data may be used for the wrong or inappropriate purposes,<br>– difficulty generating qualitative data and concerns about sharing data before the researcher(s) have had a chance to publish due to this. |

*Source***:** own elaboration

In addition to the "information" (considered to be true), there are also terms that have a completely different meaning, i.e. a negative meaning for both the entrepreneur, and contractors and customers (Łęgowik-Małolepsza et al., 2017). These are:

- pseudo-information – information that comes from different sources, but relates to the same fact or event,
- metainformation – information about information,
- misinformation – the concept of false information based on lies, manipulation and willful misrepresentation,
- parainformation – subjective information, based on an interpretation made in an individual way, which may lead to different conclusions or opinions.

Without the doubt, companies may resort to manipulating and misinforming about a given product or service, but in the long run, this is not profitable (after all, a company that does not apply an exemplary policy itself may one day fall victim to another entity that will deliberately want to introduce them. in error).

An example of actions aimed at curbing manipulation or misleading practices may be the amendment to the Act on Trading in Financial Instruments introduced in 2017 (Kwieciński, 2019). Pursuant to Polish law, the legislator introduced eight types of manipulative behaviour. The manipulative behaviour listed in the Act includes placing orders or entering into transactions that are misleading or may mislead as to the actual demand, supply or price of a financial instrument; placing orders or entering into transactions which result in an unnatural or artificial pegging of financial instruments; placing orders or concluding transactions with the intention of causing other legal effects than those for which the legal act is actually performed; placing orders or concluding transactions while misleading market participants or taking advantage of their error as to the price of financial instruments. It should be noted that transparency and universal access to information in the case of trading in securities is particularly important because the capital market is, next to the banking sector, one of the most important elements of the financial system and is subject to special protection. An efficiently operating, stable and balanced financial system plays a key role in the economy of any country, enables its development and, consequently, improves the wealth of the society. Its destabilisation may disturb the balance of this economy. Ensuring the stability of the financial system belongs to the state authorities which have been entrusted with the supervisory functions.

Moreover, due to the universal access to the financial data of listed companies, these entities are required to report data on a quarterly basis and publish them on the Internet, and, like other commercial companies, they are required to approve and publish the annual financial statements to the competent registry court. Consequently, the possibility of manipulation and deliberate misleading is negligible, as investors have virtually unlimited access to data, which allows them to quickly verify individual messages. This, in turn, proves that the exchange of information and access to data is in the interest of various groups of capital market stakeholders.

Promoting open, free and universal access databases is in everyone's interest. In the era of globalisation and more and more universal access to the Internet, companies are

not making any sense in terms of access to information and sharing it with other entities. Creating open databases to which various stakeholders have access is in the interest of everyone, as evidenced by numerous clusters, e.g. the automotive, craft, and agri-food clusters. A free database of an economic and financial nature is another such potential cluster.

An example of database including datasets meeting the criteria of openness described above (including publishing according to CC-BY policy) is the platform introduced by the Gdańsk University of Technology where both academic staff as well as external users may publish their data. The platform is available at mostwiedzy.pl, where datasets are published in English and can be used for free. One of the co-authors of the following paper regularly publishes the datasets obtained during the academic and scientific research[1].

As Rosa (2008) notes, a feature of knowledge-based clusters is a high concentration on a limited area of entrepreneurs, investors and scientists, and the existence of frequent formal and informal contacts between them. Industry or regional associations, often supported by public funds, are also a platform for these contacts. Therefore, the basis for the exchange of ideas and experiences in the case of economic, financial, reporting and statistical data can be common, open, data repositories. It should be clearly emphasised that nowadays, in the era of widespread use of the Internet, a cluster may also become a jointly, though independently, created open database, to which everyone has access. Everyone can also be the author of some part of the information contained within it, and thus participate in the general development of the given industry.

What is more, access to broad data also means building a knowledge-based economy. This is a concept that emerged in the 1990s. An economy in which knowledge is created, absorbed, transferred and used more effectively by enterprises, organisations, individuals and communities, fosters it quickly and has a positive impact on societies. The development of such an economy is based on the significant use of knowledge and experience; such a country can develop faster than others, because it generates faster technical and organisational progress, has better educated people and uses human and productive capital more effectively (Skrzypek, 2011).

## Conclusions

"Open Science" is an "umbrella" term where its different aspects such as ORD have to be singled out and investigated separately to better understand their opportunities and tangible problems. The above examples demonstrate the great challenge in opening research in the economics discipline, which has substantial implications for the business and industry sectors. In return, Open Science might greatly influence international visibility and recognition, and contribute to the knowledge diffusion and credibility that is one of the most critical issues for research worldwide. The ongoing debate on sharing data is far from settled, and scholars represent different opinions and practice it differently. Many of the concerns relate to the legal constraints and human participation; however, proper data management and sharing are at the core of valuable research. There is

[1] https://mostwiedzy.pl/pl/search/openResearchData?s=piotr%20kasprzak.

a severe concern that mostly commercially used open data leaves no trace and it is rather hard to assess its coverage and the actual state of affairs. However, the obvious benefits such as labour cost savings, productivity improvements, and reducing research duplications might decrease tensions related to commercialisation and openness. The right approach that will focus on developing policy positions on Open Science and commercialisation, promoting and supporting open collaboration between different stakeholders, and targeting business and industry in research findings can facilitate a transition and maximise the value of ORD in the economics discipline.

## References

Akers, K.G., and Doty, J. (2013) 'Disciplinary differences in faculty research data management practices and perspectives', *International Journal of Digital Curation*, 8(2), pp. 5–26. DOI: 10.2218/ijdc.v8i2.263 (Accessed: 6th January 2021).

Dawson, M., and Rascoff, M. (2006) '*Scholarly communications in the Economics discipline. A report commissioned by JSTOR*'. DOI: 10.18665/sr.22340.

Czarny, B. (2011) '*Podstawy ekonomii*'. PWE, Warszawa.

Górecka, K. (2018) 'Źródła informacji wykorzystywane w systemie wczesnego reagowania w przedsiębiorstwach', *Studia i Prace Kolegium Zarządzania i Finansów*, 163, pp. 143–164, DOI: 10.33119/SIP.2018.163.8.

Kwieciński, J. (2019) 'Przestępstwo manipulacji na rynku kapitałowym w świetle ustawy o obrocie instrumentami finansowymi i prawa wspólnotowego', *Studia i Prace Kolegium Zarządzania i Finansów*, 174, pp. 9–31, DOI: 10.33119/SIP.2019.174.1.

OECD (2015) 'Making Open Science a Reality', *OECD Science, Technology and Industry Policy Papers*, 25, OECD Publishing, Paris, DOI: 10.1787/5jrs2f963zs1-en.

Olak, A. (2011) 'Kapitał intelektualny i wiedza jako czynniki nowoczesnego zarządzania finansami – studium przypadku', *Zeszyty Naukowe SGGW - Ekonomika i Organizacja Gospodarki Żywnościowej*, 88, pp. 151–164, Available at: http://sj.wne.sggw.pl/article-EIOGZ_2011_n88_s151/ (Accessed: 25th May 2022).

Rosa, K. (2008) 'Klaster jako forma powiązań przedsiębiorstw w gospodarce opartej na wiedzy', *Zeszyty Naukowe Wyższej Szkoły Humanitas. Zarządzanie*, 1, pp. 69–78, Available at: https://www.sbc.org.pl/publication/11535#description (Accessed: 25th May 2022).

Rzemieniak, M. (2012) '*Badania marketingowe w podejmowaniu decyzji menedżerskich*', Lublin: Wydawnictwo Politechniki Lubelskiej, Available at: http://bc.pollub.pl/Content/1376/PDF/badania.pdf (Accessed: 25th May 2022).

Schwab, K., et al. (2011) 'Personal Data: The Emergence of a New Asset Class,' World Economic Forum Report. Available at: https://www.weforum.org/reports/personal-data-emergence-new-asset-class/ (Accessed: 25th May 2022).

Skrzypek, E. (2011) 'Gospodarka oparta na wiedzy i jej wyznaczniki', Nierówności społeczne a wzrost gospodarczy, 23, pp. 270–285, Available at: https://www.ur.edu.pl/storage/file/core_files/2012/7/2/8756a245950c4fa64366f06281bd5bcc/022.pdf (Accessed: 25th May 2022).

Skrzypek, E. (2018) 'Bariery dzielenia się wiedzą w organizacji w warunkach społeczeństwa informacyjnego', Nierówności Społeczne a Wzrost Gospodarczy, 53, pp. 34–48, DOI: 10.15584/nsawg.2018.1.2.

Sobczyk, G. (2016) 'Rola i źródła informacji rynkowych w działalności przedsiębiorstw – wyniki badania', Annales Universitatis Mariae Curie-Skłodowska. Sectio H. Oeconomia, 50(3), pp. 173–183, DOI: 10.17951/h.2016.50.3.173.

Sosnowski, P. (2019) 'Systematyzacja pojęć związanych z metodami i źródłami pozyskiwania informacji w kontekście infobrokeringu', Studia Politologiczne, 54, pp. 45–67, DOI: 10.33896/SPolit.2019.54.2.

Strzelczyk, E. (2017) 'Otwarte dane badawcze – kolejny krok do otwierania nauki, Bibliograficzne bazy danych: perspektywy i problemy rozwoju', III Konferencja Naukowa Konsorcjum BazTech, Kraków. Available at: http://ebibojs.pl/index.php/Mat_konf/article/view/616 (Accessed: 25th May 2022).