



Public Intelligence

The Use of Administrative
Data for Monitoring and Evaluation
of Public Policies

Public Intelligence

The Use of Administrative
Data for Monitoring and Evaluation
of Public Policies

Public Intelligence.

The Use of Administrative Data for Monitoring and Evaluation of Public Policies

Editors:

Agnieszka Chłoń-Domińczak, Andrzej Jędrzejowski, Jolanta Perek-Białas, Jacek Pokorski

Scientific reviewers: Józef Oleński, Mirosław Szreder

Authors:

A. Chłoń-Domińczak; M. Jasiński; J. Perek-Białas; M. Taracha – Chapter I

I. Vollbracht; G. Mazzarella; P. Paruolo – Chapter II

A. Chłoń-Domińczak; A. Ptak-Chmielewska – Chapter III

A. Ptak-Chmielewska – Chapter IV

R. Wiśła; M. Niklewicz-Pijaczyńska; M. Gołacki – Chapter V

J. Kotowski; D. Nowak; P. Skórska – Chapter VI

M. Bożykowski; A. Chłoń-Domińczak; M. Jasiński – Chapter VII

J. Adamski – Chapter VIII

A. Kowalewska – Chapter IX

A. Jędrzejowski; J. Pokorski – Chapter X

The views of the authors of the chapters contained in the publication express their own opinions and do not reflect the official position of PARP.

Citation: Chłoń-Domińczak A., Pokorski J. *et al.* (eds.). (2022). *Public Intelligence. The Use of Administrative Data for Monitoring and Evaluation of Public Policies*. Warsaw: Polish Agency for Enterprise Development (PARP).

Technical editing: Michał Taracha

Cover design: Pracownia C&C Sp. z o.o.

Translation (Chapters I; III-X): GTC AMG Sp. z o.o.

Publisher: Polish Agency for Enterprise Development, ul. Pańska 81/83, 00-834 Warsaw

© Polish Agency for Enterprise Development, Warsaw 2022

ISBN: 978-83-7633-485-1

Preparation for printing and printing: Pracownia C&C Sp. z o.o.

Delivery of original works, technical editing, layout:

SGH Warsaw School of Economics and Polish Agency for Enterprise Development (PARP)

The purchase of select original works and translation was co-financed by the European Regional Development Fund under the Smart Growth Operational Programme, 2014–2020



Scientific review

Professor Mirosław Szreder, the University of Gdańsk

Addressing the issues of administrative data being used in various types of research, analyses, evaluations, and current monitoring of specific activities, especially in the public sphere, should be considered as desirable and, from the scientific point of view, interesting and with growth potential. (...) I look positively on the issues of using administrative data for monitoring and evaluation of public policies discussed in the compilation. The level of knowledge of the topic is relatively low in Poland, therefore, it deserves a thorough discussion by specialists, scientists, and practitioners alike.

Professor Józef Oleński, the Centre of Cross Border Areas Research and Statistics for Euroregions, Statistical Office in Rzeszów

I am convinced that the compilation under review, thanks to its comprehensive approach to the problem, in terms of methodology and IT, and particularly thanks to the presentation of practical applications, will not only provide researchers and practitioners with extensive, up-to-date knowledge of the use of administrative data in statistical surveys, but also help scientists and practitioners in public sector institutions, bodies of public statistics, and scientific institutions to consider administrative data as a source equal to individual statistical data when generating microaggregates and aggregates of statistical indicators. Hopefully, the publication will encourage potential users to turn to administrative data as basis for research, analyses, and evaluations as well as decision-making for other public sector policies.

Table of contents

Scientific review	3
Foreword	9
Introduction	11
1. From data collected traditionally to administrative data	
– dilemmas and perspectives	15
Introduction	15
1.1. Use of public register data in evaluation.	17
Analytical challenges	18
Legal challenges	19
Institutional challenges	22
1.2. Current state of affairs and perspectives for using public register data for the purpose of evaluation	24
References.	27
2. Learning to learn: lessons from five years of implementing the “what works” approach to evaluation.	28
Introduction	28
2.1. The need to learn	29
Why evaluate?	29
Learning ingredients.	32
2.2. Good practice: challenges and drivers	32
Challenges	32
Drivers of best practice.	34
2.3. Examples	38
Evaluation of Job-plus in Ireland	38
Evaluation of the youth employment initiative in Portugal.	41
2.4. Conclusions	43
References.	43
3. Use of data collected in Social Insurance Institution registers and tax registers to monitor the situation of the small and medium-sized enterprise sector at the local level.	45
Introduction	45
3.1. Data on the SME sector collected by public registers	49

3.2. Analytical potential of data from administrative records for local SMEs' diagnosis . . .	54
3.3. Conclusions and recommendations.	62
References.	63
4. Use of register data for development of a predictive model to enable identification of companies with high probability of success of funded projects.	65
Introduction	65
4.1. Enterprise data in public registers	66
REGON	66
VAT Register.	68
Central Register of Entities – National Register of Taxpayers	68
National Court Register	70
Central Register of Social Contribution Payers	72
Business intelligence services	72
4.2. Examples of models supporting the assessment of potential success of enterprises: models for assessing SMEs' resilience to crisis	74
Enterprise bankruptcy forecasting model.	76
Crisis resilience model	78
Summary.	81
References.	82
5. Patent Information Metadata in Evaluating the Effects of Innovation Support for the Business Sector.	85
Introduction	85
5.1. Development of statistics on science, technology, and innovation	87
5.2. Patent information in research on the effects of innovation support for the business sector	90
A. Patent activity of Polish residents at home and abroad:.	96
B. Non-residents' interest in the Polish market:.	98
C. Legal monopoly evaluation, patent value:	99
D. Evaluation of the diffusion effect of scientific and technical knowledge:	100
E. Other obtainable information supporting the evaluation of innovative support for the business sector:	100
5.3. Obstacles in the process of obtaining and using patent data	101
Conclusion.	104
References.	105

6. Past experience and future prospects of Statistics Poland in the use of administrative data in enterprise surveys.	107
Introduction	107
6.1. Methodology for estimating the effects of support for enterprises vs. requirements for existing data.	111
6.2. Specific nature of administrative data and data collected by official statistics	113
6.3. Use of administrative data in the evaluation of support for enterprises – opportunities and challenges	116
6.4. Opportunities of supplementing result indicators of enterprises' activity in the area of competitiveness, innovation, and employment	121
Summary and recommendations	125
References.	127
7. Examples of supporting the implementation and monitoring of public policies based on administrative data	130
Introduction	130
7.1. The study process in selected areas of public policies in the ZPA project	131
Study preparation	134
In-depth diagnosis of resources.	134
Development of a set of metrics and their interpretation	137
Development of intervention logic	139
Analytical solution prototype.	142
Study implementation phase.	143
7.2. Legislative background and organisational arrangements for the functioning of the Integrated Analytical Platform	144
Summary.	147
8. Use of Administrative Data to Create and Evaluate Health-related Policies.	148
Introduction	148
8.1. Health data – an international perspective.	149
8.2. The role of the health needs map in the creation and evaluation of health policy.	150
Historical background.	150
Role in the health care system.	152
Current scope of data and analysis	154
Health needs assessment.	155
8.3. Principal health data sets in Poland.	157
HEALTHY DATA (Healthy Data, 2021).	158

ProfiBaza and EpiBaza	159
Polish National Cancer Registry	159
System and Implementation Analysis Database (Disability-adjusted life years, DALYs, 2021)	160
8.4. Health data and evaluation of public policies	161
Health care	161
Health in all policies	162
The future of health data and its use	163
References	164
9. Possibilities of using administrative data of the Public Procurement Office for the evaluation of public programmes	168
Introduction	168
Significance of the public procurement market for the economy	169
Principle of openness in public procurement	172
9.1. Information needs related to the public procurement market	174
Enterprises' participation in the public procurement market	174
Impact of training and advisory programmes/projects on the activity of enterprises on the public procurement market	177
9.2. The scope and potential of data collected in the public procurement system ...	179
Data published in the BZP	180
9.3. Ways of accessing data collected in the public procurement system	190
The "old" BZP	190
The "new" BZP	191
Information on requests to participate and tenders submitted under the PZP of 2019	192
Summary	193
References	194
10. Using company financial statements in evaluating the impact of economic development programmes	198
Introduction	198
10.1. Data... the fuel of evaluation?	200
10.2. Description of the EKRS Financial Document Browser	206
Entity scope	206
Content of financial statements	207
Time scope	209

Access to resources	210
10.3. Characteristics of the particular elements of financial statements	210
Profit and loss account	212
Balance sheet	216
Statement of changes in equity	220
Cash flow statement	222
10.4. Use of KRS data in impact evaluation practice	225
Assumptions for the analysis	225
Main outcomes of the analysis	229
Conclusions	231
References	233
List of abbreviations	235

Foreword

Dear Readers,

I am pleased to present the new publication by the Polish Agency for Enterprise Development (PARP), ***Public Intelligence. The Use of Administrative Data for Monitoring and Evaluation of Public Policies***. The compilation is the result of cooperation of a team of outstanding researchers, experts, and practitioners from the area of science and administration. It is a response to trends, already well-established, and to the latest revolutionary phenomena in the light of which analytics of rich data resources from administrative systems is gaining importance in formulating public policies, according to the needs, challenges, and current technological possibilities.

The dynamic trend of using administrative data (as well as any new big data from private and public resources) can significantly support the process of monitoring and evaluation during EU's new programming period 2021–2027, as well as analysis of the SME sector's condition. Today, at the time of industrial revolution 4.0, data is the new fuel for business: it drives the activity of enterprises and enables the development of economy. Also, in order to properly perform its tasks, the administration should be close to the realities in which the economy and society operate; it should act on the basis of objective register data. New methods of data collection, going beyond the declarations of respondents and limitations of the representative method, will allow analysts and evaluators to find previously unrecorded trends and identify unknown mechanisms of programme operation, which will serve to design new measures in a more informed manner.

The concept of this publication, prepared in PARP's Evaluation Unit, in the Analysis and Strategy Department, is part of PARP's notable achievements in the field of research and evaluation, which help us to design and improve our services. Moreover, the undertaking is in line with the objectives of a broader cooperation agreement signed by PARP and Warsaw School of Economics: the authors and co-authors of this compilation are researchers from Warsaw School of Economics and other experts invited by our scientific partner (including the University of Warsaw and Jagiellonian University), as well as PARP experts.

I believe that the compilation will encourage and guide other units, national and regional public intelligence services in Poland, to make wider use of administrative data in their research and analytical challenges. For researchers and scientists, I hope the publication will be an inspiration to develop, theoretically and empirically, the concepts and examples which it presents. It is worth emphasising that the publication has received very positive scientific reviews from renowned experts with recognised achievements in social research and public statistics, who list a wide range of its target readers.

Dariusz Budrowski

President of Polish Agency for Enterprise Development

Introduction

Nowadays, the use of administrative data in designing, implementing, and evaluating public activities (e.g., in the economic and educational policy, active labour market policies, and the health policy) is part of the strategic direction of the state, called data policy. On the one hand, the impulse for development in this direction was provided by the COVID-19 pandemic. The pandemic has shown that without precise data, analysed in an agile way in real time, it is impossible to respond adequately and effectively, let alone forecast or counteract. Conversely, the limitations caused by the pandemic forced a change in the approach to social research (including programme evaluations). When field surveys relying on direct contact with the respondent became impossible, other approaches to research gained in importance (including remote techniques, but especially secondary use of administrative data). On the other hand, experience and knowledge with respect to the use of register data, which have been accumulated for many years by the Polish administration and academia, are also not without significance. Analysts and researchers are developing the readiness of the data ecosystem and best practices in this area. This, in turn, leads to gradual reduction of the siloed approach and development of cooperation between data managers and data users. In Poland, good examples in this area include systemic solutions in the field of education and higher education (the system of monitoring the economic lives of graduates), projects implemented for the purposes of many policies (the Integrated Analytical Platform), as well as projects, in particular PARP's projects, concerning counterfactual impact evaluation of the programme (referred to in this publication).

The aim of the publication is to present European and Polish experience in using data collected in public registers for monitoring and evaluation studies, and show the associated opportunities and limitations. Primarily, the publication aims to offer practical solutions that are used in selected projects (analyses) relying on administrative data. Furthermore, it illustrates the potential of analyses based on such data for the accuracy and usefulness of conclusions and recommendations, and points to new possibilities which open up before the evaluation of public policies, both in Poland and at international level. The summary of the experiences and reflections on the solutions adopted in them, and the analysis of the potential of the methods and tools used so far, will serve to formulate conclusions for further initiatives of this type.

The publication starts with the chapter, *From Traditional Data Collection to Administrative Data: Dilemmas and Perspectives*, which is a systematic introduction to the problems discussed, taking into account the existing theoretical concepts and the achievements of Polish and foreign researchers. In addition, the chapter is the first in this publication to discuss the analytical, legal, and institutional challenges of working with data collected in public registers. Apart from that, the authors provide a diagnosis of the current state and the prospects for the use of this type of data in monitoring and evaluation processes.

The introduction to the main issues covered in the publication is complemented by the second chapter which discusses the experience of promoting the idea in question and the use of administrative data in evaluations conducted by the Competence Centre for Microeconomic Evaluation affiliated with the European Commission Joint Research Centre.

The chapter presents:

- Academic and policy foundations of counterfactual impact evaluation from the perspective of European Commission's experience;
- Progress observed in applied microeconometrics, known as the causality revolution, in relation to evaluation;
- Examples of good evaluation practices in analyses based on data from public registers;
- Conclusions on the differences between scientists and politicians' approach in their perception of projects and scientific work (e.g., research based on randomised control trials).

After the two introductory chapters, the publication presents the potential and best practices of using administrative data in enterprise research.

The third chapter discusses the achievements in enterprise analysis from the regional and local perspective, which come from the Regiogmina project (Gospostrateg). The project used data from administrative registers to evaluate the situation of the SME sector in municipalities of the kujawsko-pomorskie voivodship.

The next (fourth) chapter assesses the possibility of using large administrative data sets to define a predictive model for selecting enterprises with high chances of project success (crisis-proof enterprises). As proposed, the model can support the process of selecting companies in grant schemes.

The fifth chapter presents the potential of using patent data to evaluate the effects of the policy of pro-innovation support for the enterprise sector. Among other things, it presents a list of indicators related to the innovative activity of enterprises, which can be analysed using data collected by patent offices.

The sixth chapter explores the possibilities of using administrative data held by Statistics Poland (e.g., REGON, KEP, ZUS, PIT, CIT and VAT) for the purposes of monitoring and evaluation of enterprise support. Moreover, it outlines the perspectives and directions of Statistics Poland's activities with regard to integration of different data sources and practical use of data in enterprise research.

The seventh chapter highlights the experience of research and monitoring of the effects of public policies under the Integrated Analytical Platform project. In particular, it introduces the approach to developing analytical solutions as well as the legal solutions adopted that enable this research.

The eighth chapter shows opportunities and challenges for the use of public data collected in the health care information system. It investigates the past experience in collecting, compiling, and using public data, referring to the example of health need maps, which were introduced into the Polish health care system so that the priorities in health care can be objective, and to confirm the rationale of financing activities from EU funds. The chapter explains the methodology of estimating health needs in Poland and the practice of planning and evaluating health policy measures.

The ninth chapter looks into the scope of information collected in the Public Procurement Bulletin (in Polish, Biuletyn Zamówień Publicznych, or BZP) and indicates the directions of its potential use in evaluations, especially of programmes aimed at strengthening the activity of small and medium-sized enterprises in the public procurement market. The wide range of data available in the BZP makes it possible to use it for monitoring and evaluation and for other purposes underlying the new regulation of the state's purchasing policy.

The last, tenth chapter delves into the use of financial reports of companies for the evaluation of enterprise support programmes. Implementation of a counterfactual impact evaluation, which aims to examine the impact of an intervention on enterprises, involves obtaining data on the economic situation of beneficiaries and relevant control cases. In

Poland, this data is contained in documents like companies' financial reports submitted annually to the National Court Register and published on the website of the Ministry of Justice. Appropriate use of this information makes it possible to significantly improve the quality of data and reduce the cost and time of obtaining data for evaluation.

We hope that you will find the publication interesting and that it will encourage you to develop your analytical and research projects using administrative data. We also believe that the concepts and examples we have provided will encourage data managers to be more ready to develop systemic solutions and procedures to open public data for secondary use in various scientific and research contexts.

Editorial Team

1. From data collected traditionally to administrative data – dilemmas and perspectives

Agnieszka Chłoń-Domińczak

*Collegium of Economic Analysis
SGH Warsaw School of Economics*

Mikołaj Jasiński

*Faculty of Sociology
University of Warsaw*

Jolanta Perek-Biała

*Center for Evaluation and Analysis of Public Policies
Jagiellonian University
Collegium of Economic Analysis
SGH Warsaw School of Economics*

Michał Taracha

*Collegium of Economic Analysis
SGH Warsaw School of Economics*

Introduction

Global progress in evaluation research is increasingly based on the use of data from administrative registers. This shift may be explained by several reasons. Firstly, public registers contain detailed information on the developments in various areas of social and economic life, including information on businesses, taxpayers and taxes, debtors, employment and unemployment rates, social transfers, health services or, more broadly, public services, to name but a few. Secondly (as opposed to quantitative research), secondary use of such data does not involve high costs of collecting information. Thirdly, the collected data cover the entire population, which reduces the occurrence of errors in the estimation of evaluation indicators, typical for sample studies.

In Poland, data collected by public institutions are relatively rarely used for policy monitoring and evaluation, especially if policies are unrelated to the field for which they were originally established. There are over 600 different types of domestic public registers¹ but so far, their potential has not been fully used.

At present, public register data are mostly used in official statistics. Annually, Statistics Poland collects 438 data sets from official registers and public administration IT systems. These data sets have been used in 123 studies under the 2021 official statistics research programme (57% of studies conducted by the President of GUS, independently or jointly with other authorized bodies) within 31 areas of official statistics². Such data use has some advantages, in particular:

- reduced administrative burden;
- more flexibility and faster response to users' new information needs, and provision of information adequate to such needs;
- lower costs of statistics production;
- better accuracy and timeliness of output data and estimates;
- shorter data processing times.

Moreover, the use of registers and IT systems in public statistics has a positive impact on:

- better integration of the official statistics system with administrative systems;
- possible combination of administrative and statistical data sets;
- possibility of providing new output data and services.³

Challenges and dilemmas around the usability, quality, and availability of data in large administrative sets become evident in the daily work of evaluators, who assess the effectiveness of public policies in various areas of social life.

This chapter aims to provide the reader with an overview of European and national experiences in the use of public register data in evaluation research, and show the relevant opportunities and limitations. Therefore, on the one hand, we discuss the practical solutions that have been used in Poland in selected projects (analyses) based on administrative data,

¹ Over 600 IT systems are included in the GUS Repository of Information Standards.

² <https://stat.gov.pl/badania-statystyczne/przekazywanie-danych-z-systemow/wykorzystanie-rejestrow-urzedowych-i-systemow-informacyjnych-administracji-publicznej-w-statystyce-publicznej/>

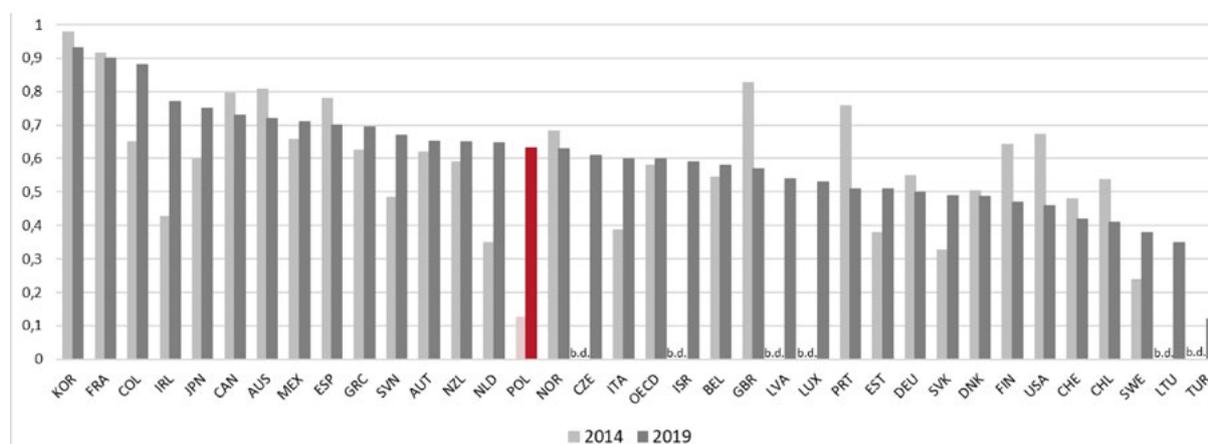
³ Ibid.

and on the other hand, we address the potential of analyses produced with the use of administrative data in respect of accuracy and usefulness of the ensuing conclusions and recommendations, as well as opportunities which open up for the evaluation industry. We sum up the experiences acquired so far, reflect on the adopted solutions, and examine the potential of the existing methods and tools to facilitate the development of methods and research, including evaluation, based on the use of administrative data in Poland.

1.1. Use of public register data in evaluation

The state has huge quantities of data which can potentially be reused to support the implementation of public policies (see Bożykowski, Chłoń-Domińczak, Jasiński, Zajac, 2019). Public administration bodies collect various types of data, but due to legal barriers or insufficient skills, these resources are now rarely used. The 2014 OECD Open Government Data Survey clearly showed that the situation in Poland was far from satisfactory. The availability and access to government data in Poland scored lowest among all the OECD countries covered by the study (Figure 1.1). The most recent OECD analyses (2019 report) have revealed that the situation in many countries has improved over the recent years, and Poland is now above the OECD average.

Figure 1.1. OURdata Index: Open, useful and reusable administrative data – 2014 and 2019⁴



Source: OECD Survey on Open Government Data

⁴ For the U.S. and Turkey, data unavailable for 2019 were replaced with 2017 statistics.

Below, we discuss the key challenges of using public data for evaluation: analytical, legal, and institutional.

Analytical challenges

The use of public register data requires a slightly different approach, both in terms of data content, information which they can provide, and the size of such collections, as illustrated in Figure 1.1. It should be noted that at the stage of study design only information already collected in specific registers can be taken into account and that no information excluded from collection can be added or included. As a result, a study based on register data cannot be “tailored” to specific research or evaluation questions. Consequently, these are analyses that are adapted to the contents of information available in the particular registers rather than the other way round. If data are of lower quality, or if errors have been identified, data can be corrected or supplemented only if the register holder responds in an adequate manner, as they may not necessarily be interested in making such adjustments.

Mortelmans & Pasteels (2014) emphasize that the use of such data has its limitations. Firstly, the time between the subsequent rounds of data collection will affect their immediate usefulness, in particular, when information is updated for instance only once a year. Therefore, any ad hoc research is excluded. Secondly, due to the complex data structure, advanced skills are required to analyse and use the data for analytical purposes. Thirdly, the nature and volume of data need to be carefully controlled and prepared for further analysis. Finally, analyses and the presentation of their findings are often challenging as well; the main challenge for analytical teams is to know how information resources, which are often quite extensive, should be presented in an adequate, user-friendly, and clear manner.

On the other hand, some advantages should be emphasized as well. Register data are free from errors caused, *inter alia*, by interviewers’ influence, or other factors (e.g., missing responses of study participants, or non-response to the study). Besides, such data are often gathered over a considerable period of time, as their collection is subject to legal requirements that establish the obligation to keep registers. Consequently, analyses can be performed at a specific point of time or in a given year, as well as over a longer period. Therefore, it is possible to monitor changes that occur in the analysed populations over time (longitudinal data). Moreover, administrative data help to analyse small populations, which are difficult to assess based on data from surveys.

At the same time, data mismatches can be limited by combining information from different registers. Such an approach significantly broadens analytical potential of register data.

Exhibit 1.1. Partial study, full study, and study based on register data: basic similarities and differences

Partial (representative) study	Full study (census)	Study based on register data
Independent study	Included in statistical registers – may become the basis for various register-based studies	
Use of register data to define populations and as a data source		
Sampling, estimates, uncertainty measures	Systemic thinking and coordination with other register-based research is essential (combination of data from registers)	
Independent collection of information – development of own questionnaire	Use of information collected in registers	
Data correction – option to contact respondents		Data correction – intermediated by the register holder
Refusal to participate (non-response), tight deadline of the fieldwork		Data mismatch
Data quality: sampling and measurement errors	Data quality: measurement errors	Mismatch: missing data, under-coverage
Small sub-samples – no estimates for small groups	Presentation – large tables with several variables	

Source: Wallgren & Wahlgren (2007)

Legal challenges

Currently, there are no horizontal regulations in Poland that govern the possible use of administrative data, in particular for the purposes of evaluation in the broad sense.⁵

Principles underlying the establishment and purpose of use of all public registers have been defined in relevant statutory provisions. Data in public registers are to be used solely for the purposes determined in specific legal acts. Specific public register data can only be shared and used for research and evaluation (i.e., purposes other than the intended purpose of the registers) when the data holder allows it in the case of publicly available data, and when legal provisions expressly define such modalities.

⁵ Understood as comprehensive regulations specifying, in particular, the purposes and principles of processing and analysing data collected in administrative registers and the means of publishing the results obtained.

One of the few examples of extensive use of public register data is their employment in official statistics by Statistics Poland (including as part of censuses), which is regulated by the Official Statistics Act. Another example of data use for public policies deliverables monitoring, based on relevant registers, are the analyses produced by the Ministry of Finance. The tax-contribution model developed at the Ministry of Finance relies on joint database of the Ministry and Social Insurance Institution (ZUS). The work starts with the data held by the Ministry of Finance – collections of tax returns filed by natural persons, gathered under the tax law, while Social Insurance Institution uses data on the declared social and health insurance contributions (Chrostek et al., 2019).

Pursuant to Article 50 paragraph 3 (a) and (c) of the Social Insurance System Act, data from ZUS are also used to evaluate programme outcomes. Based on the ZUS data, estimates are made of indicators for the implementation of projects financed under the European Social Fund (ESF) (Ministry of Investments and Development, 2016). These analyses are produced as part of current activities of public administration, and their outcomes are published in the evaluations (made by the ESF Management Department of the administrative office of the Minister responsible for regional development) regarding the estimated long-term result indicators of ESF, mainly those related to employability of people who receive support under the Knowledge, Education, Development Operational Programme and Regional Operational Programs (Pokorski 2020).

Another example is also worth mentioning: the system which has been operating since 2016, to track graduates' economic path (ELA). The solution is based on the provisions of Article 352 of the Act on Higher Education and Science. The Act defines the sources of data used in the ELA system, their scope, and the information processing procedure. In essence, the system uses two sources: POL-on System registers, and the data of Social Insurance Institution (registers of the insured and insurance contribution payers). In particular, the Act identifies all variables in data sets covered by the ELA system. The procedures under the Act ensure full anonymity of the surveyed subjects (Jasiński et al., 2017).

Solutions implemented in the ELA system have inspired the development of a career tracking system for graduates of public and non-public secondary schools, as defined in Article 26b of the Education Law Act. The provisions entered into force in April 2021. The graduates' career tracking system is managed by the Ministry of Education. The solution relies on:

- data from the Education Information System, referred to in the Education Information System Act of 15 April 2011;
- data from the POL-on Integrated Information System on Higher Education and Science, referred to in Article 342 paragraph 1 of the Act on Higher Education and Science of 20 July 2018;
- data collected by district examination boards;
- data collected by Social Insurance Institution on the accounts of the insured or contribution payers.

The latest legal solution established to regulate the use of administrative data is the Act of 11 August 2021 on open data and the reuse of public sector information, which entered into force on 12 December 2021 and integrated the provisions on the Open Data Program, initially included in the Act on Access to Public Information. The goal of the programme is to increase the quantity and improve the quality of public sector information, private data, and metadata used to describe their structure, all made available on the open data portal. The Seven-Year Programme will be adopted by the Council of Ministers, at the request of the minister responsible for computerization, by way of a resolution. The Programme will be implemented by members of the Council of Ministers and the President of GUS, but will also stay open to other entities. Coordination tasks will be performed by the minister responsible for computerization. The same Act provides for another new solution: establishment of a network of open data officers. The officers will perform tasks intended to increase the quantity and improve the quality of public sector information made available on the open data portal (<https://dane.gov.pl/pl>) for their further reuse. The portal is a pool of data that are updated on an on-going basis and made available free of charge for reuse. Data of more than 160 providers, i.e., public administration bodies and private entities are gathered in one place. The portal is targeted at:

- citizens interested in the activities of the state;
- companies which develop innovative data-based products and services;
- non-governmental organizations which use data in their daily work;
- scientists involved in research projects;
- officials who produce reports and analyses;

Registration is not required to use the data available on the portal. However, to access extra functionalities, such as development of own data compilations and convenient tracking

of data updates, the user needs to create an account. The data portal is managed by the minister responsible for computerization.

The said Act introduced modifications to the Act of 17 February 2005 on computerization of activities of entities performing public tasks; the amendment included a new chapter 3b entitled *Integrated Analytical Platform (ZPA)*. This change has come as a result of a project implemented under the same name and discussed in greater detail in further parts of the paper. ZPA is to be used for analyses that support the development of key public policies, based on data provided by entities responsible for public tasks and collected in public registers and ICT systems. The Platform is to become the central analytical tool to gather data held by public administration bodies, as identified by the analysis initiator. However, it should be noted that any new initiatives will require a review in order to check whether the existing legal framework allows the collection of relevant data and their analysis. Such assessment will be conducted based on agreements signed by the minister responsible for digitisation with public bodies involved in ZPA analyses. Once introduced, such legal changes will significantly expand the possibilities of using public data for evaluation purposes.

Institutional challenges

As proven by international experiences (including those of Sweden, the UK, and Australia), effective use of administrative data resources requires clear and adequate coordination between entities involved in the monitoring and evaluation process and systems integrating such data and setting up an organizational framework for their use. This is particularly true when new analytical and evaluation projects are initiated, required information resources are identified and diagnosed, rules for setting up analytical teams (involving members from public administration, private sector, and the research community) are defined, and mechanisms to ensure anonymity and security of analyses are determined. An adequate regulatory framework in the form of horizontal legal provisions is needed to regulate these issues, identify goals, and define rules applicable for the integration of information collected in registers kept by various holders for the purpose of their public tasks. In particular, the framework should define the rules for the use of such resources as part of public policy analyses, including the scope and method of data integration, information resource management, and tools to support appropriate data integration and use (Bożykowski, Chłoń-Domińczak, Jasiński, & Zajac, 2019).

In Sweden, public register data are made available by two systems: MONA⁶ and LISA⁷, both managed by Statistics Sweden. The MONA System: Microdata Online Access provides secure Internet access to microdata in the Swedish official statistics system and offers processing and analytical tools in the form of a ready-made application. The LISA System (*Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier*) is another solution, which enables the use of administrative data for research and evaluation. This is an integrated base of longitudinal data on health, labour market, and education.

In the United Kingdom, institutional framework for the use of administrative data has been defined by Administrative Data Research Partnership (ADRP).⁸ ADRP works with the Office for National Statistics (ONS) in order to strategically capture, manage, and share data in a secure manner, for the purpose of high-quality research and public policy evaluation. The governmental initiative labelled Data Integration Partnership for Australia (DIPA) is another example of a systemic institutional solution developed to foster the use of administrative data, with the Multi-Agency Data Integration Project (MADIP) as one of its core components.⁹ The MADIP is an agreement between six Australian government agencies to facilitate the integration of information from public registers. MADIP covers a wide range of public policies, including healthcare, education, social benefits, income and taxation, social mobility, and family policy.

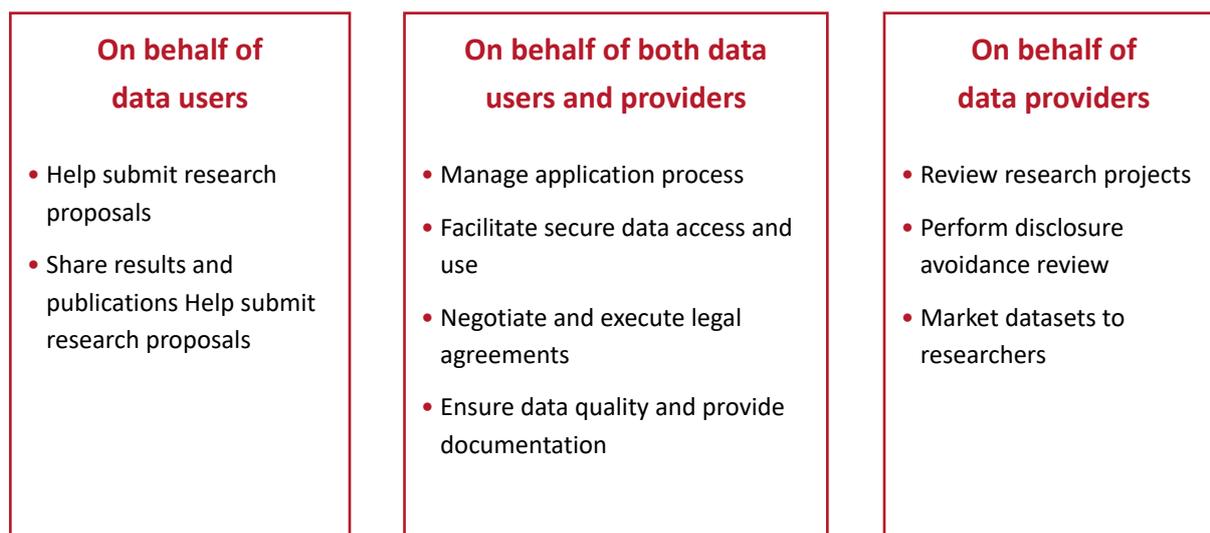
The international good practices outlined above are based on the functions of institutions which mediate access to data on the one hand, and ensure security of administrative data provided by their holders on the other. These functions of data intermediaries are presented in Figure 1.2 below.

⁶ <https://www.scb.se/en/services/ordering-data-and-statistics/ordering-microdata/mona--statistics-swedens-platform-for-access-to-microdata/>.

⁷ <https://www.scb.se/LISA>.

⁸ <https://www.adruk.org/>.

⁹ <https://www.abs.gov.au/about/data-services/data-integration/integrated-data/multi-agency-data-integration-project-madip>.

Figure 1.2. Data Intermediary Functions: What do they do and on whose behalf?

Note: Figure translated and adapted for the Polish version based on the following license: Creative Commons Attribution-Non-commercial-NoDerivs 4.0 International (CC BY-NC-ND 4.0). Translation and adaptation in (Bożykowski, Chłoń-Domińczak, Jasiński, & Zajęc, 2019).

Source: King (2018) in: (Bożykowski, Chłoń-Domińczak, Jasiński, & Zajęc, 2019).

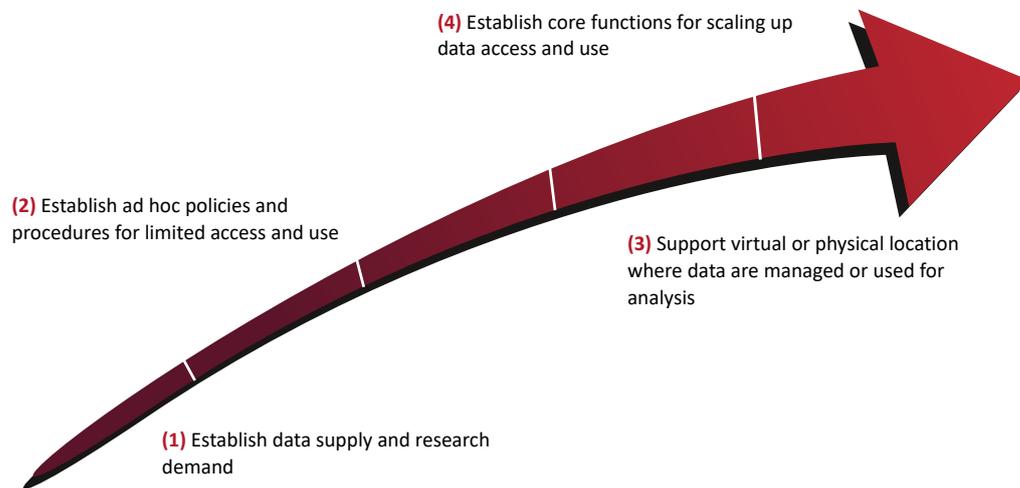
1.2. Current state of affairs and perspectives for using public register data for the purpose of evaluation

Specific examples and experiences of the use of administrative data for analyses and evaluations in Poland have been compared to solutions available in Europe and discussed in the particular chapters of this document. We focus on prospective and practical solutions of administrative data use for the analysis of the operation of companies, as well as Polish best practices developed in projects based on large datasets, including infrastructure solutions, e.g., the Integrated Analytical Platform, and sector practices, for instance in the health care system.

It should be noted that the way towards a fully-fledged public data management and sharing system (data intermediary) is a process which usually takes time and involves several development stages. Based on a series of interviews with practitioners from 17

organizations in the United States and Canada (including universities, public institutions, and non-profit organizations), King (2008) proposed a model of a journey towards a mature data management model, which is presented in Figure 1.3 below.

Figure 1.3. Model of the way towards integration of the state’s information resources



Note: Figure translated and adapted for the Polish version based on the following license: Creative Commons Attribution-NonCommercial- NoDerivats 4.0 International (CC BY-NC-ND 4.0). Translation and adaptation in (Bożykowski, Chłoń-Domińczak, Jasiński, & Zając, 2019).

Source: King (2018) in: (Bożykowski, Chłoń-Domińczak, Jasiński, & Zając, 2019).

In the first step, institutions providing data for analytical purposes match data supply and demand: on the one hand, they receive data requests from researchers wishing to do their analyses, on the other hand, they are informed by providers whether such data can be shared. The Statistical Information Centre of Statistics Poland, the Department of Statistics and Actuarial Forecasts of Social Insurance Institution, and the IT Centre for Education are examples of such intermediaries who will run consultations if analysts flag data demand.

The next step involves working with both parties (the institution which requests data and the data holder) in order to establish ad hoc rules and procedures regulating limited access to data and their use. For instance, data may only be used for a specific research project by a specific research team. Such solutions include e.g., provision of anonymized data from the unemployment register for research projects on the functioning of the labour market and its selected aspects.

In the third step, the data broker identifies or develops a technology solution to enable data management and use in research. The first three steps are a kind of a pilot procedure to build trust and demonstrate potential success of the analytical project. For example, Statistics Canada provides access to an array of individual and free administrative data through its network of over 30 selected data centres. Successful pilot projects could be used as an argument for expanding access to administrative data. An example of such pilot initiative in Poland is the Integrated Analytical Platform – project under which administrative data from several holders are to be made available for the implementation of six specific research projects described in this material.

After the completion of the first three steps, the data broker can determine their core functions to ensure better data access and usage. This requires development and implementation of verified and stable rules and procedures, and definition of a legal framework to enable wider use of administrative data (administrative and operational capacity of the data broker is enhanced accordingly). Technology solutions are also introduced to handle larger volumes of transactions and data analyses.

Polish practice and experience of public policy evaluation and monitoring based on administrative data show that Poland is currently at the third stage of the process of integration of the state's information resources. Practices of using administrative data for specific purposes are in place and pilot projects are now underway. However, systemic solutions similar to those applied in Sweden are still missing. Therefore, further actions and initiatives are needed to develop an institutional solution which will facilitate scaling and sharing of data for the purposes of their analysis and evaluation, and will use the experience gathered in Poland and good international practices. This should be done by introduction of new legal and institutional solutions (Statistics Poland could be the institution with such a role), and development of systemic procedures that could enable the use of administrative data.

References

- Bożykowski, M., Chłoń-Domińczak, A., Jasiński, M., & Zając, T. (2019). Dane publiczne – nowy impuls dla rozwoju Polski (8/2019; Policy Paper. Polish Economic Institute).
- Chrostek, P., Klejdysz, J., Korniluk, D., & Skawiński, M. (2019). Wybrane aspekty systemu podatkowo-składkowego na podstawie danych PIT i ZUS 2016. *MF Opracowania i Analizy*, 1, 1–91.
- Jasiński, M., Bożykowski, M., Chłoń-Domińczak, A., Zając, T., & Żółtak, M. (2017). Who gets a job after graduation? Factors affecting the early career employment chances of higher education graduates in Poland. *Edukacja*, 4(13), 17–30. <https://doi.org/10.24131/3724.170402>.
- King, M. (2018). Data Sharing Governance and Management. ADRF Network Working Group Reports, <https://texaserc.utexas.edu/wp-content/uploads/2018/11/Data-Sharing-Governance-and-Management.pdf>
- Ministry of Investments and Development. (2016). Appendix 6. Method of measuring ESF long-term result indicators based on the data source from the evaluation study, (in:) Guidelines for monitoring substantive progress in the implementation of 2014-2020 operational programs, Warsaw https://www.funduszeuropejskie.gov.pl/media/2279/wytyczne_nr_13_220415_zal_6.pdf
- Mortelmans, D., & Pasteels, I. (2014). Using register data in the social sciences. In: SAGE Research Methods Cases. <https://www.doi.org/10.4135/978144627305013497458>
- Pokorski J. (2020). Ewaluacja wsparcia kompetencji osób na rynku pracy i promocji zatrudnienia. Pełny obraz efektów z mikro- i makroperspektywy. In: B. Ciężka, A. Rybińska (2020). Ewaluacja w działaniach edukacyjnych, FRSE, Warsaw, pp. 88-107.
- Wallgren, A. & Wallgren, B. (2007). Register-based Statistics. Administrative Data for Statistical Purposes, John Wiley & Sons, Chichester.

2. Learning to learn: lessons from five years of implementing the “what works” approach to evaluation

Ian Vollbracht

European Commission Joint Research Centre

Gianluca Mazzarella

European Commission Joint Research Centre

Paolo Paruolo

European Commission Joint Research Centre

Introduction

This chapter discusses challenges and opportunities associated with *ex-post* evaluation as a learning opportunity about “what works” and for whom. It draws lessons from the experience of the Competence Centre on Microeconomic Evaluation (CC-ME) active at the Joint Research Centre (JRC) of the European Commission (EC) since 2016¹⁰.

These reflections come at the time of the 2021 award of the Nobel Prize in Economics to the initiators of the so-called “credibility revolution” in economics, which is the basis of the “what works” approach¹¹. There is also renewed emphasis on learning from evaluation in

¹⁰ The Competence Centre for Micro-economic Evaluation (CC-ME) was set up in 2016 in line with the launch of Better Regulation. The CC-ME is based at the Joint Research Centre of the European Commission, the Science-for-policy Department of the Commission. Its mission is to enhance the effectiveness of EU policies through data-driven micro-econometric analyses and provide causal evidence on what works. It does this by focusing on quantitative evaluations of EU policies contributing to the Better Regulation Agenda, the European Semester, and EU Spending Programmes, in collaboration with all EU Institutions and Member States.

¹¹ Cf. The Royal Swedish Academy of Sciences. (2021). *The Prize in Economic Sciences 2021*. <https://www.nobelprize.org/prizes/economic-sciences/2021/press-release/>, accessed October 25, 2021; see also Angrist & Pischke (2009).

the Better Regulation (BR) Communication of the European Commission (EC) called “Joining forces to make better laws” (see EC (2021) and the associated Guidelines).

The chapter is organized as follows: Section 2.1 describes how the search for causal effects of policies is fully embedded in BR and describes the learning ingredients necessary for an effective overall approach, which are grounded in data and the use of causal methods. Section 2.2 lists the current challenges for good practice, and drivers of good performance. Section 2.3. contains examples drawn from the work of CC-ME in JRC, while Section 2.4 provides a conclusion.

2.1. The need to learn

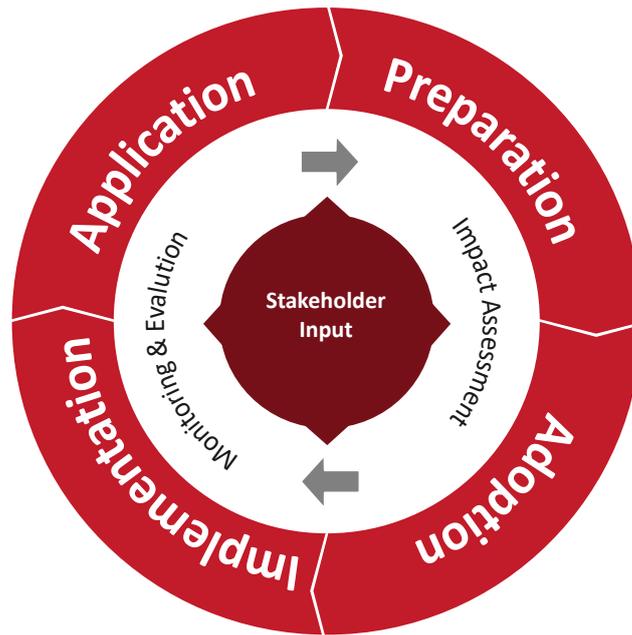
This section discusses why learning from evaluations is important in institutions like the EC and what are the necessary basic elements of such learning.

Why evaluate?

This subsection observes how learning about causal effects of policies is the focus of BR and briefly discusses the potential associated benefits.

BR is the EU’s agenda for improving the quality of legislation throughout the policy cycle, straddling ex-ante, implementation, and ex-post phases. The EC has been developing and seeking to improve its legislative model for many years. Communication EC (2002) was an important staging post; see Renda (2006, pp. 43 – 56) for a useful summary. Since the Communication “Better regulation for better results – An EU agenda” in 2015, see EC (2015), the EC has adopted a comprehensive policy cycle approach, composed of the ex-ante, implementation, and ex-post phases. The most recent Communication on BR entitled “Better regulation: Joining forces to make better laws” was adopted on 29 April 2021. Figure 2.1, taken from the latest Communication, depicts the EU policy cycle.¹²

¹² EC documents on BR guidelines and the BR toolbox can be found at: https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how_en, accessed October 25, 2021.

Figure 2.1. EU policy cycle

Source: https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how_en

The 2021 Better Regulation Guidelines (EC, 2021, p. 6) set out the principle of “learning from experience” in the following broad manner:

“laws and regulations can always be improved, partly because the world is in constant flux, but also due to the EU’s inherent diversity (which EU policies have to capture and to which they have to adapt). Policymakers need to learn from the experience of implementing and applying EU rules on the ground. This is why the ‘evaluate first’ principle should apply whenever legislation is revised”.

The principle of “evaluate first” is the idea that, in order to amend legislation, one needs to learn what worked and what didn’t work in the implementation of previous policies. More specifically (EC, 2021, p. 9):

“Evaluations gather evidence to assess how a specific intervention has performed (or is working), taking account of earlier expectations in the context of an impact assessment and/or ensuing from the adopted legislation and whether there were

unintended or unexpected effects that were not anticipated and taken into account in the impact assessment or the adopted act. They also draw conclusions as to whether the EU intervention:

- *remains fit for its purpose;*
- *should be adjusted for greater effectiveness, relevance and coherence, and/or to eliminate unnecessary burdens or inconsistencies; or*
- *should simply be repealed.”*

In terms of overall objectives (EC, 2021, p. 23): *Evaluation is an evidence-based assessment of the extent to which an intervention:*

- *is effective in fulfilling expectations and meeting its objectives;*
- *is efficient in terms of cost-effectiveness and proportionality of actual costs to benefits;*
- *is relevant to current and emerging needs;*
- *is coherent (internally and externally with other EU interventions or international agreements); and*
- *has EU added value — i.e., produces results beyond what would have been achieved by Member States acting alone.*

The above text shows what evaluations are meant to discover, namely, the causal effects of policies.

Potential benefits of learning what works in a policy context similar to the one of the EC is comprehensively discussed in the chapter “The potential of administrative microdata for better policy-making in Europe”, Langedijk et al. (2019), in the book “Data-driven Policy Impact Evaluation”, see Crato and Paruolo (2019). They can be summarized as a more efficient use of resources and increased institutional transparency.

The following subsection discusses in greater detail the necessary ingredients of learning what works.

Learning ingredients

In order to learn what works, field data and causal methods need to be combined wisely. In order to apply appropriate causal techniques developed for quasi- or natural experiments¹³, granular data¹⁴ are usually needed. Finally, results from the analysis need to meet the interest of policy makers in order for the analysis to be useful in the future ex-ante policy design phase. As discussed in the rest of the chapter, lack of any of these elements impairs learning what works.

2.2. Good practice: challenges and drivers

This section discusses challenges and success drivers in learning what works.

Challenges

The three main potential challenges are as follows:

- lack of appropriate and sufficiently reliable and granular data;
- data privacy and/or data security constraints that prevent the data from being shared with those conducting the evaluation project;
- an overarching policy environment whereby rigorous research findings may be deemed too sensitive to be placed in the public domain or used in the subsequent policy design.

One illustration of data challenges and opportunities is provided by Rettore and Trivellato (2019) who discuss both the potential and limitations of administrative data for evaluating public interventions with reference to the Italian “Liste di Mobilità” – a programme to handle collective redundancies. The paper compares two “generations” of studies, the first of which relied on poor information, while the second covered studies that over the years were based on access to much richer data, which in turn provided deeper insights into policy, unknown to first-generation studies.

¹³ Natural experiments refer to experiments run by Nature, i.e., not designed and controlled by scientists. The use of this term in Economics goes back at least to Trygve Haavelmo, see Haavelmo (1944).

¹⁴ Detailed data at the microeconomic level (free from averaging characteristic for aggregated macroeconomic data).

The inability to access administrative microdata¹⁵ for rigorous policy evaluation may result from “data privacy and data security concerns”, see Langedijk et al. (2019). In particular, meeting all the legal requirements for the transmission of (or access to) microdata sets between the authority that holds the data and the research entity is a necessary, but not sufficient, condition for the data transfer to actually take place. Transmission of data may require secure and strongly encrypted information and communication infrastructure. In some cases, data handling requires even a dedicated secure room with appropriate access registers and security protocols. Naturally, purchase and installation of such facilities often involve a financial cost.

In practice, although it would be convenient to assume that microdata is always and everywhere both collected and available to researchers, it is in fact far from being the case. One type of data that poses a particular challenge in terms of access is sensitive, personal data such as health records, regarding which national data protection authorities tend to have a very cautious approach to allowing data sharing, even when intended exclusively for research purposes.

Another challenging aspect concerns the scope of integrating data, for example from a range of different countries or different ministries within one country, if no overarching national law is in place to coordinate the sharing and linking of information at micro level.

Unlike for macro-economic data on GDP, unemployment etc., which are coordinated by EUROSTAT, there is no centralized and coordinated approach at EU level for administrative data. The collection of taxation, health, social security, and other individual-level data is rarely performed primarily for research purposes. It is also the case that different EU Member States have very different approaches, histories, and traditions regarding the collection, use, and potential for sharing personal data.

At the institutional level, the EC as a whole is cognizant of this situation, its challenges, and its potential benefits; the latest version of the Better Regulation Guidelines (EC, 2021, p. 34) places strong emphasis on data collection:

¹⁵ The term “microdata” is used as a catch-all term for data-sets that include large amounts of granular data at individual or company level. It therefore encompasses “administrative registry data” held by governments (for example tax and social security records) as well as commercial data sets on various aspects of company performance.

“The impact assessment should outline monitoring and evaluation arrangements (including the definition of a set of indicators and any relevant data collection aspects) to track progress in implementation of the legislation and assess its ultimate impacts on the ground.”

As this text clarifies, already at the impact assessment stage (i.e. during the policy preparation stage prior to adoption), the EC department (or Directorate General, DG) responsible for the policy must ensure that it has a plan in place to track the progress of implementation and its impact. This data plan should be explicitly linked to the future evaluation of impacts.

Therefore, there are no reasons for future optimism that the EC and EU will, more broadly, become “smarter” as regards the data aspects of policy design, implementation, and subsequent evaluation.

Drivers of best practice

This subsection discusses a set of drivers that can help to overcome the aforementioned hurdles in practice. Three potential drivers of successful learning are:

- developing a common language and learning culture in the science-for-policy community;
- developing a community of practice that can match demand and supply of experts in causal methods;
- fostering / facilitating access to microdata that has already been collected for administrative purposes to the community of experts.

These drivers are analysed here, drawing from practical experience at CC-ME.

Developing a common language

JRC in general and CC-ME in particular operate in the area between science and policy-making. Finding a common language between policy officials and scientists is essential to communicate effectively in the fields where communication is key to create and use rigorous evaluation results. Moreover, the purpose of evaluation, which can sometimes be perceived as judgement of the personal success or failure of certain decision-makers, should be viewed

rather as a common learning exercise, at the service of improving the effectiveness of future policies. However, reaching this shared understanding of the purpose of the exercise can take time: the core element is building trust, which is rarely an overnight process.

Scientists and policy officials also operate in different environments. For instance, they have quite different boundary conditions: scientists may have a (subconscious) tendency to look for confirmation of the results published in scientific literature. Policy-makers tend to face other pressures regarding acceptability of outcomes and results in areas for which they feel directly responsible in some way. Also, policy officials typically work to shorter time horizons and deadlines imposed by legislative requirements than scientists, who can typically devote more time to a given research project, often over a longer time horizon.

In order to develop a common language, the concept most commonly used within the EC is the “co-design” of a policy intervention and evaluation plan from the early stages of the process. This means that the EC department(s) directly responsible for the policy due to be evaluated need to consult widely, both within the EC, and with relevant stakeholders, in terms of project design, evaluation methodology, and data plan.

This is coherent with the overall spirit of the Better Regulation Guidelines (EC, 2021, p. 5):

“‘Better regulation’ is not about regulating or deregulating. It is a way of working that allows political decisions to be prepared in an open and transparent manner, informed by the best available evidence, including via the comprehensive involvement of stakeholders.

‘Evidence’ refers to multiple sources of data, information, and knowledge, including quantitative data such as statistics and measurements, qualitative data such as opinions, stakeholder input, conclusions of evaluations, as well as scientific and expert advice.”

A community of practice

This Community of Practice (CoP) approach is a facilitating factor to enable good practice to spread and the quantity and quality of evaluations to grow and improve. A community of practice consists of a range of actors (a community) including scientists, policy officials, and,

possibly, officers of data-holding institutions to discuss what works in terms of project design and methodology, and compare problems and solutions across teams.

What worked in one Member State may or may not work in another (this issue is called a problem of external validity). However, if one team listens to the experience of another team from another Member State, this may inspire the design of a new policy implementation or of evaluation of similar policies. This is the peer effect in learning to learn.

This requires a community of practice that includes both experts in causal methods, i.e., the methodology to infer cause-and-effect relations from appropriate data, policy officials, and perhaps also data holders. Experts in this community usually come from the academia, international organizations, consultancies, or research departments of private enterprises and public institutions.

Examples of activities in this area include the CoP for evaluations of European Social Fund (ESF) interventions, which is coordinated by DG EMPL and supported by CC-ME. It involves policy experts from Managing Authorities in Member States as well as consultants; the CoP meets regularly to discuss problems and solutions for the evaluations of ESF interventions.

Another activity in this area is the COMPIE conference series¹⁶ (COunterfactual Methods for Policy Impact Evaluation) organised by CC-ME, which brings together practitioners and academics to discuss developments in the theory and applications of counterfactual methods for policy impact evaluation. Applications can cover a broad range of interventions, from labour and social policy to regulation and funding programs, ideally focussing on EU policies.

Facilitating access to microdata

Finally, all evaluation work is based on the availability of reliable micro-level data that can be securely accessed, analysed, and integrated into usable data-sets, whilst fully respecting the relevant legislation on data protection and privacy. The efforts by data holders to share data in secure conditions should be incentivised by all levels of government, that in return would

¹⁶ Cf. European Commission (2021). *Competence Centre on Microeconomic Evaluation*. https://knowledge4policy.ec.europa.eu/microeconomic-evaluation/compie-conference-series_en, accessed October 25, 2021.

receive a wealth of analyses to see which policies worked best, as well as ideas on how to improve them in the future.

How these three drivers can be combined is shown by evaluation reports conducted by CC-ME in the area of the ESF evaluation. A list of these studies is presented in Table 2.1 below. All of the projects were based on partnerships with relevant Commission departments (DG EMPL) and Managing Authorities in EU Member States. The studies involved safe access to microdata and attempted to maximise communication and the scope for peer learning and information sharing within the CoP.

Table 2.1. List of studies using administrative data

Country	Year	Title	CC-ME staff	Managing Authority	Link
Umbria (Italy)	2017	Counterfactual Impact Evaluation of Work Experience for Graduates (WELL)	Corinna Ghirelli, Enkelejda Havari, Giulia Santangelo	Marta Scettri	JRC report 106430
Latvia	2018	Vocational training and labour market outcomes: Evidence from Youth Guarantee in Latvia	Massimiliano Bratti, Corinna Ghirelli, Enkelejda Havari, Giulia Santangelo	Janis Leikucs, Normunds Strautmanis	JRC report 110247
Belgium	2020	Active Labour Market Policies in Flanders. Evaluation of the ESF “Work Experience for Young Persons” program	Giulia Canzian, Elena Claudia Meroni, Giulia Santangelo	Josephine Foubert	JRC report 113899
Ireland	2019	JobsPlus Evaluation	Antonella Ferrara, Andrea Geraci, Gianluca Mazzarella, Giulia Santangelo	Hugh Cronin, Saidhbhín Hardiman, Ciaran Judge	JRC report 119523
Portugal	2020	Evaluation of the Youth Employment Initiative in Portugal using Counterfactual Impact Evaluation methods	Andrea Geraci, Silvia Granato, Gianluca Mazzarella	Nuno Duarte, Maria Joao Mortagua	JRC report 120942

Source: Elaborated by author

2.3. Examples

This section reports sample results from two evaluation studies performed in CC-ME on social impact of policies in the Member States. They provide insights into how the hurdles

listed in Section 2 can be overcome. They are selected from the support work of CC-ME in collaboration with DG EMPL¹⁷.

These studies are examples of the dual approach CC-ME applies for its wide range of policy evaluations with a number of departments within the Commission (DG AGRI, BUDG, CNECT, COMP, EAC, EMPL, RTD). Wherever possible, short versions of the reports are later published in peer-reviewed scientific journals, sometimes with a more technical focus. This twin-track approach is intended to engage with the policy community on the one side and the research community on the other.¹⁸

Evaluation of Job-plus in Ireland

In 2017, the Irish Department of Employment Affairs and Social Protection (DEASP) approached CC-ME regarding an evaluation of the effects of a hiring incentive scheme named “JobsPlus” launched in Ireland in 2012 and partly financed by ESF. This initiative aims to support long-term unemployed individuals in finding new jobs. It has the structure of a financial incentive for firms which hire long-term jobseekers. The amount of the subsidy is either € 7,500 or € 10,000 over two years, and depends on the length of the qualifying period of each jobseeker.¹⁹

Data received for this evaluation contain records of all individuals that from 2007 to 2018 received benefit payments from DEASP. However, not all benefit payments recorded in the database were useful for the purposes of the evaluation. DEASP administers a series of benefits that are not related to unemployment (such as maternity leave, trainings, disability benefits, etc.). Therefore, the first step of the evaluation was to recode the full set of

¹⁷ More details of this specific initiative can be found at: European Commission. *Competence Centre on Microeconomic Evaluation*. https://knowledge4policy.ec.europa.eu/microeconomic-evaluation/quality-assurance-support-qas-cie_en, accessed October 25, 2021.

¹⁸ The list of publications of people in CC-ME reflects this twin track, see the CC-ME yearbooks available at: European Commission. (2021). *Competence Centre on Microeconomic Evaluation*. https://knowledge4policy.ec.europa.eu/microeconomic-evaluation/cc-me-yearbooks_en, accessed October 25, 2021.

¹⁹ The €7,500 subsidy is payable to firms hiring individuals who have been at least 12 months (312 days) unemployed in the previous 18 months. €10,000 is provided to firms hiring jobseekers over 25 years of age who have been at least 24 months (624 days) unemployed in the previous 30 months.

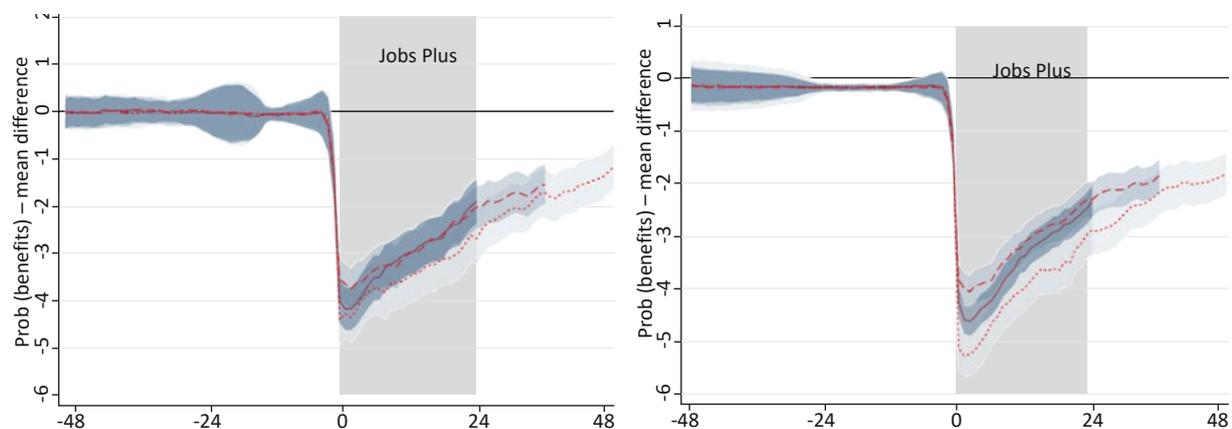
benefits in order to include only the relevant ones. This produced a database recording each individual who received unemployment-related benefits in each month from 2007 to 2018.

In principle, it is quite easy to reply to questions like: “How many participants continue to be employed 24, 36, or 48 months after JobPlus?”: it simply requires data about the working careers of the participants. However, in order to conduct proper evaluation, it is necessary to compare this (factual) outcome with an estimate of what would have happened to these individuals if they had never participated in the intervention (counterfactual outcomes).

In general, algorithms match participants with the most similar non-participants, according to a series of characteristics. In order to be reliable, the algorithms require a large number of control (non-participant) units: a higher number of control units means that the algorithm has a higher probability of finding an individual with (almost) exactly the same characteristics as a given participant in the intervention.

Thanks to the abundance of individuals in this information set, the evaluation was conducted by first selecting individuals who became eligible for the JobsPlus scheme in exactly the same month as certain “non-JobsPlus-recipient” individuals. In the second step, a selection was made among the non-recipients to identify a “matching” case with the most similar work-history (in the 48 months before the start of JobsPlus) with respect to each “treated” individual receiving JobsPlus benefits.

Figure 2.2. Effects of participation in the “Jobplus” scheme in respect of probability of receiving unemployment benefits relative to the first month of intervention



Left: € 7 500 subsidy. Right: € 10 000 subsidy.

Source: Cronin et al. (2021)

The results are reported separately for participants hired with the € 7 500 subsidy (left panel in Figure 2.2) and those hired with the € 10 000 subsidy (right panel in Figure 2.2). The horizontal axis represents the relative time with respect to the beginning of the intervention, starting from 48 months before to 48 months after the intervention. The vertical axis is the difference between the participants and the matched not-participants in their probability of receiving unemployment benefits. Moreover, each graph represents separate lines for individuals observable for 24, 36, and 48 months after the beginning of JobsPlus.

Figure 2.2 shows the main results of the evaluation. The first part of each chart shows the number of months until the relevant individuals in the study received (or did not receive) the JobsPlus support. As one would expect, there is virtually no difference between the two groups in the pre-treatment period: the treated group that received the JobsPlus benefits and the selected (matched) control group of individuals who were deliberately selected to have identical work experience histories prior to the start of the policy.

Such precise control group could be obtained from the data set that was available. Data could be accessed for the entire population of benefit recipients in Ireland. This made it possible to select “matched” individuals with identical, pairwise, work histories as their JobsPlus counterparts covered.

The positive numbers on the x axis in both charts in Figure 2 show the 48-month period after the policy intervention (from 0 to 48 months). As such, this shows the effect of the intervention. JobsPlus participants and non-participants demonstrate sizable differences in their probability of receiving unemployment benefits, even 48 months after the initial receipt of the benefit, with JobsPlus recipients significantly more likely to have gained formal employment during that period.²⁰

More details on the study can be found in Cronin et al. (2021).

²⁰ The results of this evaluation were presented by CC-ME to DEASP in Dublin in October 2019. Currently, CC-ME staff involved in the evaluation are working on publishing the paper in a peer-reviewed academic journal, giving more focus on the methodological aspects used in the study. The paper has been presented at the AIEL (Italian Association of Labour Economists) conference in 2019 and the COMPIE conference in 2021.

Evaluation of the youth employment initiative in Portugal

Another example is provided by an evaluation conducted by CC-ME, of the Youth Employment Initiative (YEI) in Portugal. The YEI targets young individuals who often have no (or very little) work experience. Consequently, the study does not involve data on the previous work experience of its participants.

In this context, CC-ME relied on administrative data obtained from the Portuguese public employment service (IEFP), as well as the Social Security (SS) register of Portuguese social security contributions. The data cover all the Portuguese individuals below 30 years of age as of September 1st, 2013 registered in the Public Employment Service (1,131,804 individuals in total).

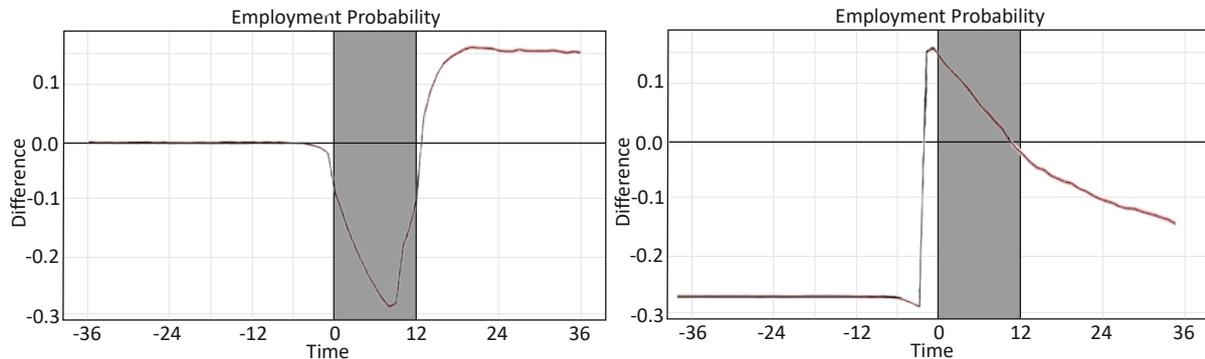
The data contain very detailed information on individual characteristics, including gender, exact date of birth, place of residence, nationality and education. In addition, they include monthly information about the employment status of each individual, which can be classified as:

- **unemployed:** the applicant is registered in an Employment Centre but is not employed;
- **employed:** the individual has a full-time or part-time job, or submits a monthly income statement;
- **occupied:** the individual is attending vocational training, taking part in active employment measures, or is involved in socially necessary work;
- **inactive:** the applicant is registered in an Employment Centre but does not meet immediate conditions for work.

Thanks to the great amount of data, the treated individuals benefiting from the initiative and the control individuals not covered by the scheme were first matched based on gender, year of birth, education, and district of residence. Next, the matching procedure selected counterpart “control” individuals who in the 36 months prior to the start of the YEI had the most similar work history according to the data.

The analysis was performed separately for internships and hiring support schemes. This was possible thanks to the large data set, covering the entire population of YEI participants.

Figure 2.3. Left: internships of 12 months. Right: hiring incentives. Effects on the employment probability relative to the start of the program.



Source: Duarte et al. (2021)

Figure 2.3 shows the results for young people who participated in a 12-month internship programs (left chart) and for those who were receiving hiring incentives for 12 months (right chart). The vertical axis represents the difference between the participants and the matched non-participants in their likelihood of being employed.

The left part of each graph (between -40 and 0) confirms the validity of the design, as it shows no difference between the treated and control units before the intervention. During the intervention period (between 0 and 12), a negative difference is visible between the treated and control units as internships of individuals participating in internship programs were not considered as a working period.

The situation was different for individuals receiving hiring incentives, properly employed during the period of the intervention (between 0 and 12 months). In both cases, the results show positive effects even 36 months after the start of the intervention (i.e., 24 months after its end).

More details on the study can be found in Duarte et al. (2021).

2.4. Conclusions

Micro data collected for administrative purposes can be combined in secure environments for causal inference on what works and for whom. These analyses underpin the quantification of the effectiveness, efficiency, and EU value added of policies.

Several measures can be implemented to foster best practices in evaluation, including facilitating data access, participating in a community of experts, policy officers, and data holders and creating the dialogue necessary to take stock of evaluation results for the revision of future legislation and funding. A sample of these measures from the JRC of the EC are the activities of the CC-ME discussed in this chapter.

Similar measures can be implemented by institutions in the Member States, for instance participation in communities of practice on evaluation, and providing support for legislation that would enable fast and secure access to administrative microdata.

References

- Agosti F., Dusatko, J., Meroni, E., Vrobel, J. (2021). Evaluation of the support to schools project in Czech Republic. *JRC Report 125324*. <https://op.europa.eu/en/publication-detail/-/publication/1086e5c6-3219-11ec-bd8e-01aa75ed71a1/language-en>.
- Angrist, J. & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton: Princeton University Press.
- Cronin H., Ferrara, A., Geraci, A., Mazzarella, G. & Santangelo G. (2021). Matching using sequences: the evaluation of the Irish JobsPlus wage subsidy scheme. *Mimeo*.
- Crato, N. & Paruolo, P. (2019). *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design*. Cham: Springer. <https://link.springer.com/book/10.1007%2F978-3-319-78461-8>, accessed October 25, 2021.
- Duarte, N., Geraci, A., Granato, S., Mazzarella, G. & Mortagua, M. J. (2021). The evaluation of the Youth Employment Initiative in Portugal using Counterfactual Impact Evaluation methods. *Mimeo*.
- European Commission. (2002). Communication on impact assessment. *COM, 276 final*. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2002:0276:FIN:EN:PDF>, accessed October 25, 2021.
- European Commission (2015). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. “Better regulation for better results – An EU agenda”. *COM 215 final*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0215&from=MT>, accessed October 25, 2021.

- European Commission (2021). Commission staff working document “Better Regulation Guidelines” https://ec.europa.eu/info/sites/default/files/swd2021_305_en.pdf, accessed April 25, 2022.
- Haavelmo, T. (1944) The probability approach in econometrics, *Econometrica* 12, iii-vi+1-115, <https://doi.org/10.2307/1906935>, accessed October 25, 2021.
- Langedijk, S., Vollbracht, I. & Paruolo, P. (2019). The Potential of Administrative Microdata for Better Policy-Making in Europe. In Crato N. and Paruolo, P. (Ed.). *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design* (333-346). Cham: Springer.
- Renda, A. (2006). Impact Assessment in the EU. *CEPS*. <https://www.ceps.eu/ceps-publications/impact-assessment-eu-state-art-and-art-state/>, accessed October 25, 2021.
- Rettore, E. & Trivellato, U. (2019). The Use of Administrative Data to Evaluate the Impact of Active Labor Market Policies: The Case of the Italian Liste di Mobilità. In Crato N. and Paruolo, P. (Ed.). *Data-Driven Policy Impact Evaluation: How Access to Microdata is Transforming Policy Design* (165-182). Cham: Springer.

3. Use of data collected in Social Insurance Institution registers and tax registers to monitor the situation of the small and medium-sized enterprise sector at the local level

Agnieszka Chłoń-Domińczak

Collegium of Economic Analysis

Warsaw School of Economics

Aneta Ptak-Chmielewska

Collegium of Economic Analysis

Warsaw School of Economics

Introduction

Small and medium-sized enterprises (SMEs) play an important role in countries' economies, including Polish – the SME sector represents around 99% of enterprises in Europe. This sector significantly contributes to economic development, both regionally and locally, through jobs' creation, formation of economic capital, and investments. Therefore, the condition of the SME sector is an important element in the development potential of a particular area. At the same time, the importance and spatial distribution of SMEs across countries is uneven – both regionally and locally. Among other things, this is due to differences in local development, entrepreneurs' access to infrastructure or other resources, and the market on which entrepreneurs operate. From the perspective of economic policy, especially related to entrepreneurship development, access to information about this sector and its basic characteristics on each of the mentioned levels is therefore crucial.

It is worth noting that local governments in Poland do not have mechanisms in place for monitoring entrepreneurship development at the local level. A 2017 audit by the Supreme Audit Office (NIK, 2017) proved that in most of the audited municipalities where the municipal council had adopted long-term strategic documents covering tasks related to entrepreneurship development, no systemic monitoring of strategy implementation was carried out, or it was not carried out in accordance with the adopted rules. Furthermore, if strategy implementation monitoring was carried out, it is not possible to compare the condition of the sector across municipalities due to the variety of approaches adopted.

When developing solutions for monitoring of the SME sector, it is worth taking advantage of models developed in other countries. For example, European Commission's communication (COM(2008)394) *Small Business Act* (EC, 2008) sets out priorities for promoting entrepreneurship in the EU and facilitating development of small businesses. The SBA identifies a number of priority areas from the perspective of the SME sector: promoting entrepreneurship, improving access to finance, reducing the regulatory burden, improving access to markets, and enterprises' activities internationalisation. These priorities are set out in 10 principles that aim to guide development policymaking and implementation at EU and national levels (Box 3.1).

Box 3.1. Ten principles for promoting entrepreneurship in the EU, as set out in the SBA

1. Create an environment in which entrepreneurs and family businesses can thrive and entrepreneurship is rewarded.
2. Ensure that honest entrepreneurs who have faced bankruptcy quickly get a second chance.
3. Design rules according to the „Think Small First” principle.
4. Make public administrations responsive to SMEs' needs.
5. Adapt public policy tools to SMEs' needs: facilitate SMEs' participation in public procurement, and better use State Aid possibilities for SMEs.
6. Facilitate SMEs' access to finance and develop a legal and business environment supportive of timely payments in commercial transactions.
7. Help SMEs to benefit more from the opportunities offered by the Single Market.
8. Promote the upgrading of skills in SMEs and all forms of innovation.
9. Enable SMEs to turn environmental challenges into opportunities.
10. Encourage and support SMEs to benefit from the growth of markets.

These principles are monitored by the SBA Factsheet²¹ using 9 groups of statistical indicators (indicators for principles 3 and 4 are presented jointly). The monitoring, as assessed by (Stano and Ghisett, 2016), provides a good basis for assessing implementation of business development priorities.

The main objective of the SBA is to monitor the development of the SME sector at the level of EU member states. However, as pointed out among others by the OECD (2017), it is also necessary to develop monitoring and evaluation of the SME sector at regional and local level. Good evaluation at local level can help policy makers to design local policies that support enterprise development. The OECD points out that countries will need more effective, efficient regional and urban policies in the coming years. The policies must be more effective because latest trends are indicative of increasing spatial socio-economic disparities across countries, leading to significant disparities between different units of local government. Therefore, local policies, including those related to enterprise development, should become increasingly effective, on the one hand, considering the expectations regarding policies' effects, and on the other hand, considering the existing budgetary constraints and the need to improve the effectiveness of support instruments implemented at local level. Consequently, one of OECD's recommendations is to choose an appropriate spatial level for designing, implementing, and evaluating policies. In conducting evaluations, it is also important to take into account the links and externalities related to the unit's environment and the fact that the choice of a spatial unit to be analysed may not be neutral for the results of empirical analyses and their interpretation. This means that the measured impact of a policy at one spatial scale may be significantly different at another spatial scale (e.g., municipalities versus regions).

The experience of the REGIOGMINA project has shown that the SME sector is regionally diversified, which requires appropriate adaptation of the policy design process to support enterprise development at the local level, based on relevant local data and information about the sector. Data from registers can be used for this purpose. For instance, administrative data can be used in estimations of indicators to assess the implementation of the first SBA principle, related to promoting enterprise development, i.e., "create conditions in which entrepreneurs and family businesses can thrive and entrepreneurship is rewarded" (Chłoń-Domińczak and Ptak-Chmielewska, 2020). These data can also be used to estimate, at

²¹ https://ec.europa.eu/growth/smes/sme-strategy/sme-performance-review_en#sba-fact-sheets

local level, some of the indicators included in the *SME Performance Review* conducted by the European Commission, such as:

- 1.1. Total early-stage entrepreneurial activity (% adults who started a business or are taking steps to start one).
- 1.2. Total early-stage entrepreneurial activity of the female working age population (% women who started a business or are taking steps to start one).
- 1.3. Established Business Ownership, measuring the percentage of persons aged 18-64 who have been self-employed for at least 42 months.
- 1.4. Share of high growth enterprises (percentage of enterprises with annual increase in employment of over 20% over a three-year period).

Data collected by Social Insurance Institution (ZUS) allow, in particular, to identify persons who register as ZUS contributions' payers when establishing business activity (including by gender in the case of the self-employed), monitor the length of the periods of newly-registered payers' contributions' payment (enterprises' stability), and monitor employment growth by assessing the number of insured persons for whom social insurance contributions are paid.

This section aims to present the experience with obtaining and using information collected by public registers to diagnose the SME sector at the local (municipal) level. This experience is related to the implementation of the project "Locating support instruments for small and medium enterprises (SMEs) operating on the basis of multi-level region management model at the level of local governments (REGIOGMINA)", carried out by a consortium led by the local government of kujawsko-pomorskie voivodship, with the participation of Warsaw School of Economics and Nicolaus Copernicus University in Toruń, under the initiative of the National Centre for Research and Development entitled "GOSPOSTRATEG" and financed from national funds.

The analyses aimed to develop a methodology for obtaining information on SMEs from databases compiled by public administrators, to ensure their geospatial identification and take into account the structures of local administration. In this chapter, we present the conclusions from this project – referring to the possibility of using the developed solutions for monitoring and evaluation of the development of the SME sector in Poland.

As the project was of exploratory nature, the information obtained mainly aimed to verify the potential and feasibility of obtaining information from administrative records. At this stage of analysis, the data do not allow an estimation of the abovementioned indicators. The data obtained were used to estimate indicators at municipal level: the number of enterprises by size, the number of employees (total, and average per enterprise by size), contribution bases (by enterprise size, Polish Classification of Activity (PKD) section, and gender of the insured), revenue and income of enterprises by size and type of tax, as well as spatial analyses and grouping of municipalities according to SME sector characteristics and other economic and social characteristics.

3.1. Data on the SME sector collected by public registers

Data on the population of businesses, particularly from the SME sector, are collected by a number of public registers (to the extent necessary for the performance of duties and obligations of register administrators). The REGIOGMINA project used registers kept by three entities:

- **Statistics Poland (GUS)** – enterprises database, obtained from the National Official Register of Business Entities (REGON),
- **Social Insurance Institution (ZUS)** – data aggregated at municipality level, including information about social security contributions' payers and insured persons,
- **Tax Chamber in the kujawsko-pomorskie voivodship** – data aggregated at municipality level from tax returns.

Data from Statistics Poland and the Tax Chamber were obtained at the request of the Voivodship Marshal, while data from Social Insurance Institution were provided at the request filed with the President of ZUS by the Rector of Warsaw School of Economics.

Current legislation does not allow to obtain micro-data (even if anonymised) on taxpayers and social security contributions' payers and insured persons. However, it was possible to obtain aggregated indicators at the level of local government units. Therefore, data from ZUS and Tax Chamber registers were obtained in aggregated form (at the level of municipalities). The REGON database for the voivodship was submitted to the Marshal's Office (in accordance with the regulations in force) as of the end of 2018. Subsequently,

geolocalisation of enterprises was carried out on the basis of anonymised data. During the research phase of the project, information was obtained from ZUS and the Tax Chamber, for 2017 and 2018. During the implementation phase conducted in 2021, data for 2019 and 2020 were obtained. The project developed a model for local data acquisition that can potentially be used by other researchers in the future.

National Official Register of Business Entities (REGON) is the main register which collects information on enterprises operating in Poland. It is an up-to-date collection of information on national business entities, kept in an IT system as part of public statistics (the register is managed by the President of Statistics Poland). The REGON register keeps data on:

- legal persons,
- organisational units without legal personality,
- natural persons conducting business activity, including those running individual agricultural holdings.

The REGON database for kujawsko-pomorskie voivodship contained **over 193 thousand records** on entities conducting business activity. It contained the following information on business entities:

- full name
- REGON and NIP (taxpayer identification) numbers
- registered address (if entered in the register)
- predominant activity according to PKD 2007 (section and division, if recorded), as well as other activities
- primary legal form
- specific legal form
- form of ownership (if recorded)
- state of legal and economic activity
- dates of: entity establishment, commencement of business activity, suspension and resumption of business activity, entry in the register, completion of bankruptcy proceedings, business activity termination (removal from the register),
- number of persons employed.

Other registers used in the project include the **register of contributors and register of insured persons of ZUS**. The former collects information on contributors, i.e., entities paying social insurance contributions for employed persons (these may be entrepreneurs paying

contributions for themselves, or employers paying contributions for their employees). The latter contains information on persons for whom social insurance contributions are paid.

Register of *contributors* contains the following information:

- registration of contributors – from the moment when the obligation to pay contributions arose (registration in the Central Registration and Information on Business (CEIDG), or the National Court Register (KRS) in the case of companies);
- contributions paid for insured persons – monthly reports submitted by contributors employing insured persons;
- deregistration of contributors – the moment when the obligation to pay contributions ceases (similarly to registration).

The following information is collected in the register of *insured persons*:

- registration of a person for insurance purposes, including inter alia: personal data (name, surname, identification data, year of birth, gender) of a person covered by social insurance
- monthly information on contributions paid, including: insurance entitlement (employment contract, contract of mandate, other entitlements), contribution assessment basis (gross remuneration being the basis for calculating social insurance contributions);
- deregistration of a person from insurance as a result of termination of a contract of employment or another contract resulting in the obligation of social insurance coverage.

Therefore, the registers of contributors and insured persons contain detailed information on the number of contributors – entrepreneurs, by number of employees (which, in a simplified way, allows to determine the size of an enterprise), section of PKD, and enterprises' field of business.

The third register from which data were obtained was **the tax register**. It provides information on taxpayers paying the following taxes:

- CIT 8
- PIT 28A
- PIT 28B
- PIT 36
- PIT 36 L
- VAT 7-K

The information obtained included the amount of taxes paid, and revenues by size of the taxpayer, taking into account revenues from imports and exports as well as from intra-Community supplies and acquisitions (for VAT returns)²².

A summary of the number of entities contained in the databases from the abovementioned registers is presented in Table 3.1. The number of enterprises contained in the ZUS and Tax Chamber registers of the analysed voivodship is significantly smaller than the number of enterprises in the REGON database, covering 176,874 business entities. The discrepancies may be due to the fact that REGON continues to keep data of business entities that suspended or terminated their business activity without submitting a formal application to update their status. There are also differences in the number of entities paying contributions and taxes. This discrepancy is highest for self-employed persons. Such taxpayers do not have to pay contributions if they are covered by other insurance (e.g., are also employed elsewhere). The numbers of micro-, small- and medium-sized enterprises in the voivodship's ZUS and tax registers were similar (45,920 and 45,055 respectively).

On the other hand, only some enterprises are VAT payers. There are also differences between groups of enterprises, which may suggest differences in the criteria for enterprise size assessment.

²² In the REGIOGMINA project, data from the tax register were used in segmentation models (e.g., average revenue per company) and innovation scoring model (other variables from the register).

Table 3.1. Number of entities included in the 2018 database²³

	self-employed	micro	small	medium	Total
ZUS	47.759	35.705	7.572	1.778	92.814
Taxes, including:	81.312	40.642	4.442	836	127.232
PIT 36	44.288	17.733	433	13	62.467
PIT 36L	14.273	10.158	2.011	119	26.561
PIT 28A	22.458	5.297	51	1	27.807
PIT 28B	293	46	10	–	349
CIT 8	–	7.408	1.937	703	10.048
VAT	38.378	33.770	4.977	1.020	78.145

Source: REGIOGMINA project

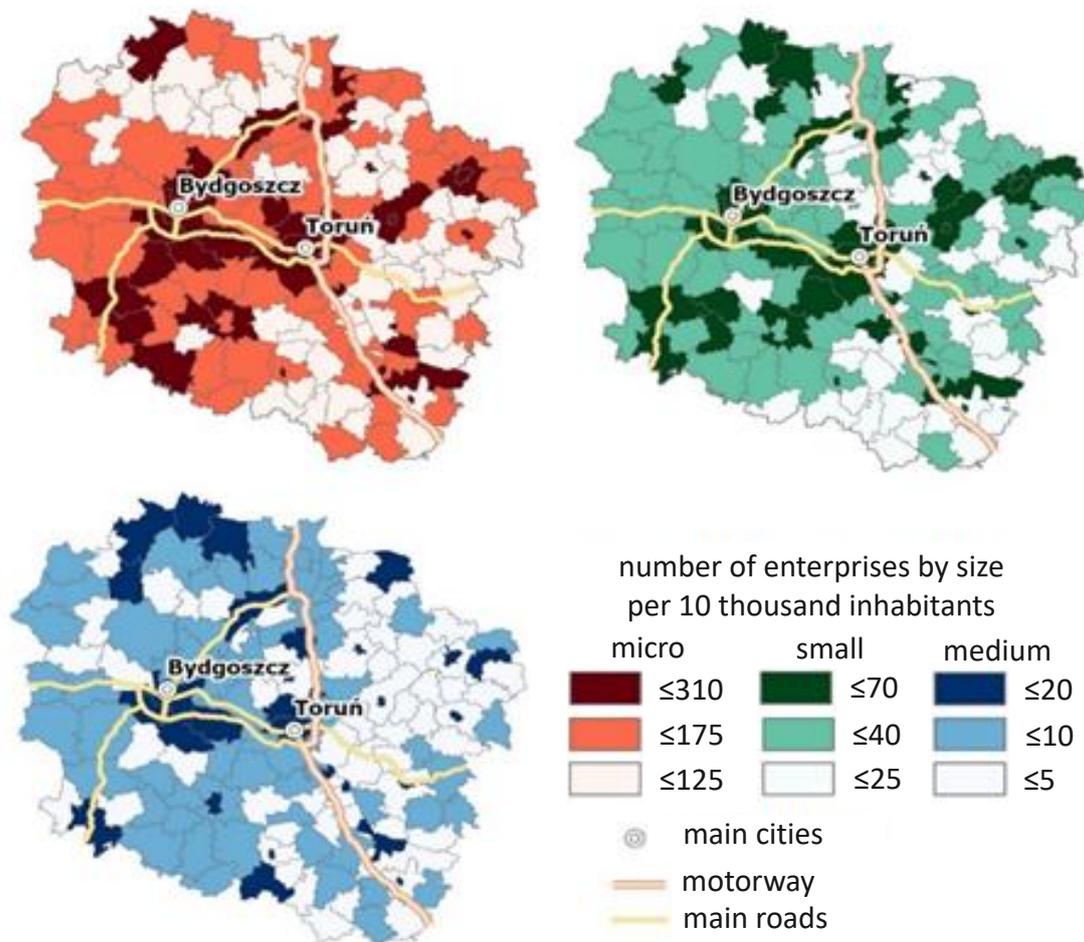
It should be emphasized that due to their scope, data collected in the particular registers are complementary. In particular, data from the REGON register provide basic statistical information for assessing enterprise population in municipalities. The data also allow to assess spatial distribution of enterprises in a given area (see Figure 3.1). Data from Social Insurance Institution (ZUS) provide information about employment numbers and their changes over time. They are also a source of information about employees' earnings by enterprise size, location, and PKD classification section. Data from tax registers contain information on the revenue and income of enterprises, thus enabling assessment of the sector's financial condition.

²³ Data from the tax and social insurance registers processed in the project were aggregates of enterprise data at municipality level (i.e., information about the number of entities in a given municipality by enterprise size, information about selected measures of central tendency of variables at aggregate level, etc.). In the case of low numbers of entities in an aggregate, data were not shared (e.g., average revenue level in the group of medium-sized companies in a given municipality).

3.2. Analytical potential of data from administrative records for local SMEs' diagnosis

Analytical potential of the collected data for monitoring and evaluation of SME sector's condition was assessed in an analysis carried out under the REGIOGMINA project in kujawsko-pomorskie voivodship. Information obtained from the registers includes data that are not available in the Local Data Bank of Statistics Poland, e.g., information on the number of contributors (by enterprise size, PKD section, number of insured persons by PKD section, age, and gender, and contribution assessment bases according to the same characteristics), as well as the number of taxpayers by the type of tax return filed, and revenues (including revenues from exports). Most importantly, the obtained administrative data provide **a general overview of the SME sector** – both in the voivodship, and in its particular municipalities. General data on the number of enterprises by size are presented in Table 3.1. The data also show uneven distribution of the number of enterprises in the particular municipalities. In the five largest cities of the voivodship (Bydgoszcz, Toruń, Włocławek, Grudziądz, Inowrocław), the self-employed account for 50.0%, micro-enterprises – 47.0%, small enterprises – 44.7%, and medium-sized enterprises 50.2% (of all contributors in the voivodship). The diversity of the SME sector's enterprises' distribution in the voivodship confirms that municipalities differ significantly in terms of size and structure of both SMEs and the self-employed. Figure 3.1. illustrates spatial distribution of enterprises (measured by the number of contributors) in the SME sector in relation to the number of inhabitants. It shows that enterprises are concentrated around Toruń and Bydgoszcz, and in municipalities located along major roads.

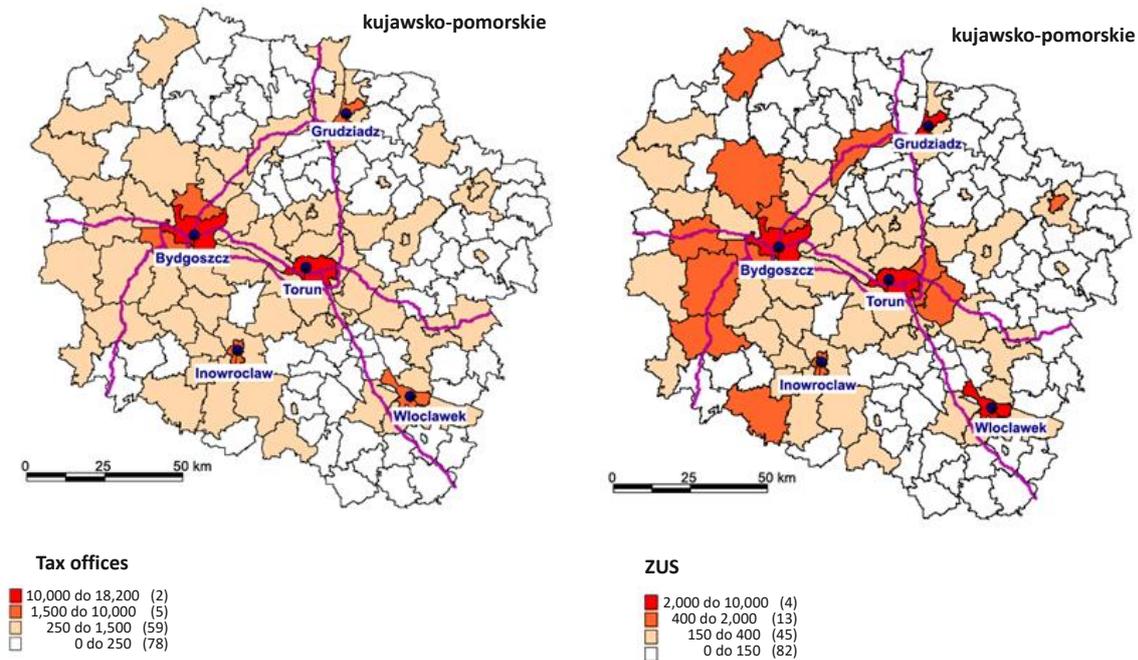
Figure 3.1. Number of micro-, small- and medium-sized enterprises per 10,000 inhabitants in municipalities of kujawsko-pomorskie voivodship in 2018.



Source: (Chłoń-Domińczak et al. 2020)

Figure 3.2. also shows a discrepancy in the number of the self-employed, both according to ZUS registers and tax offices – as indicated earlier, these numbers vary across the voivodship, which also translates into differences at local level.

Figure 3.2. Number of self-employed persons per 100,000 inhabitants by municipality in 2018.



Source: REGIOGMINA project.

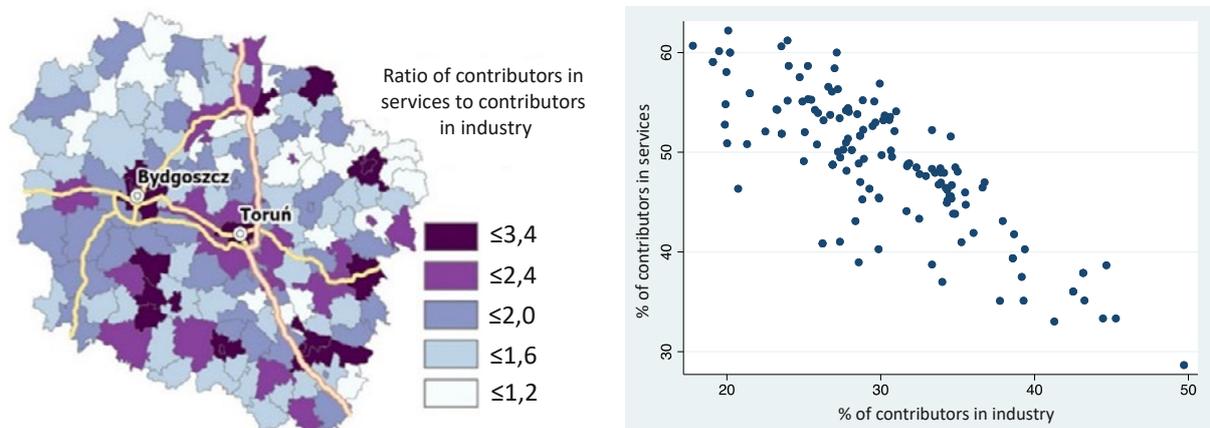
Uneven distribution of contributors is also related to uneven distribution of the number of insured persons (for whom contributors pay contributions). In December 2018, micro-enterprises paid contributions for 115 thousand persons (insured employees), small enterprises paid contributions for 166 thousand persons, and medium-sized enterprises paid contributions for 184 thousand persons. Like contributors, persons insured by contributors are typically concentrated in the largest cities of the voivodship – regardless of enterprise size. About a half of the insured are employed in the 12 largest cities and towns of the voivodship, and about 3/4 work in 1/3 of municipalities largest in terms of employment (including cities with powiat rights).

The acquired administrative data also make it possible to determine the structure of enterprises by PKD sections. The largest numbers of SMEs in the voivodship are from the PKD section G (trade) – 22%, section F (construction) – 13%, and section C (processing industry) – 10%.

Distribution of insured persons in the region is different and depends on the size of companies. Among micro-companies, less than 30% of insured persons (employees) work in trade companies (i.e., the share is higher than in this section), for construction businesses, the share is 15%, while slightly more than 10% of insured persons employed in micro-companies work in the processing industry. In small enterprises, the proportions are different, the share of insured persons working in the processing industry is higher, while the share of those working in trade and construction is lower, although trade still remains the sector with the highest employment share. Slightly over 10% of insured workers of medium-sized companies are employed in trade, while 14.4% work in construction. In this group of enterprises, employment share is largest in the processing industry, with the share of insured persons working in this sector being highly diversified depending on the municipality.

The ratio of contributors (enterprises) in industry and in services in the particular municipalities (Figure 3.3.) should be noted. The data clearly show the dominance of service enterprises in municipalities with the highest number of enterprises (where the ratio of contributors in services to contributors in industry is highest). At the same time, the relationship between the share of contributors in industry and in services in the particular municipalities is almost linear (municipalities with a high share of contributors in services have a low share of contributors in industry and vice versa).

Figure 3.3. Share of contributors in the industry and service sectors in the total number of contributors in municipalities and the ratio of contributors in services to contributors in industry by municipality in 2018.



Source: (Chłoń-Domińczak et al. 2020)

Diversity of the SME sector is also reflected in salaries, which approximate the contribution bases for pension insurance. The size of contribution assessment bases increases with the size of enterprises. In the three largest PKD sections, highest contribution assessment bases are observed in construction, and lowest – in trade. Gender pay gap (where men are paid more) increases with the size of enterprise. In the case of micro contributors, it is less than 4%, while in the case of the group of contributors employing from 10 to 49 persons it is 15.5%, and in the case of medium contributors – as much as 28.3%.²⁴

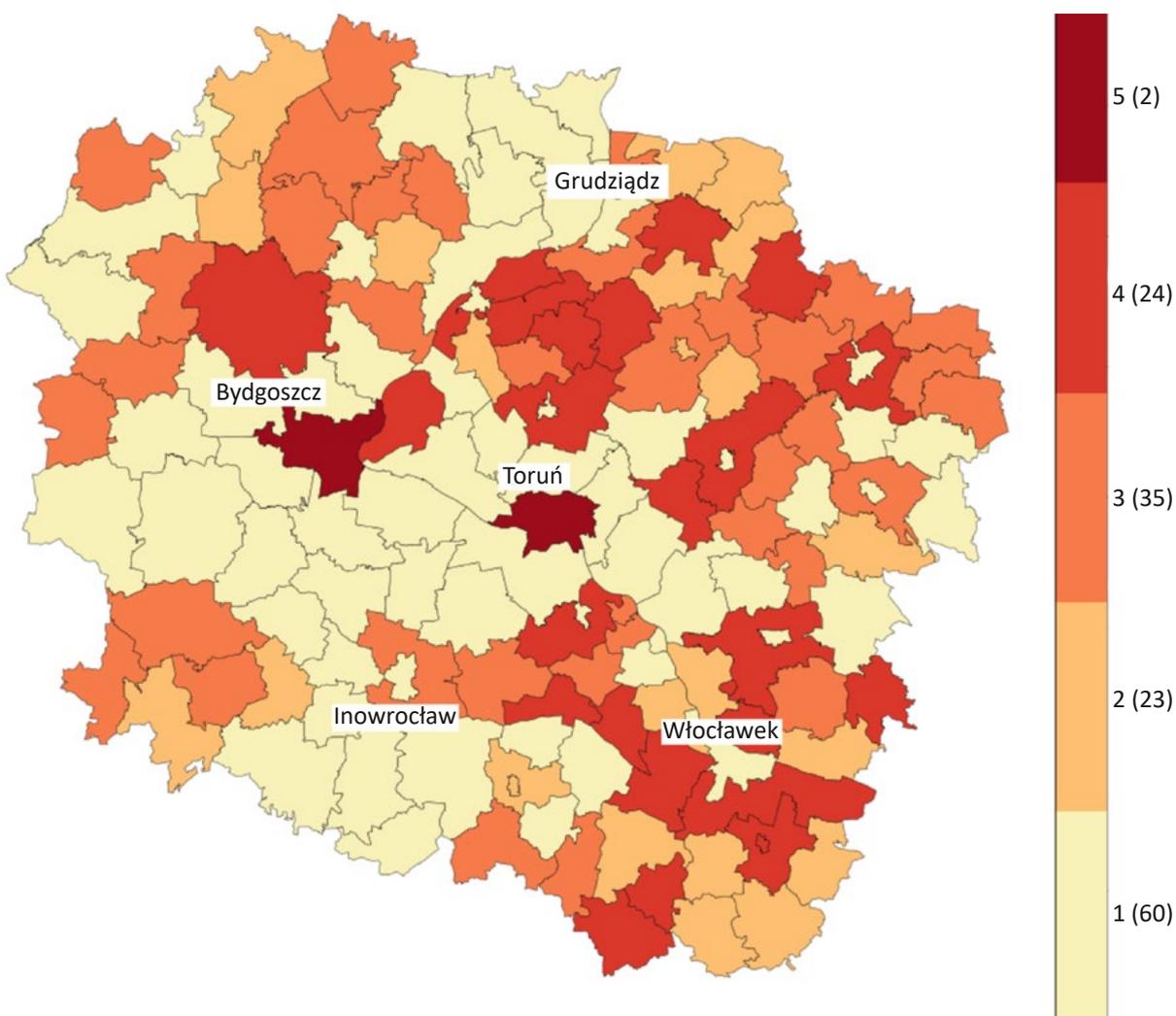
Therefore, data on contributors aggregated at municipality level show that the SME sector is highly diversified spatially. The data sample from kujawsko-pomorskie voivodship confirms the concentration of both contributors (enterprises) and the insured (employees) in the largest towns. Differentiation in size and the role of the SME sector at the local level depends on the sector as well.

The obtained administrative data in combination with data from the Local Data Bank of Statistics Poland also allow an analysis of grouping of individual municipalities. An example of such analyses can be found in the publication (Ptak-Chmielewska and Chłóń-Domińczak, 2021) which groups municipalities by their characteristics using the k-means clustering method and Kohonen's SOM. The results of the analysis using the latter method are presented in Figure 3.4. They show that statistics on the SME sector and its environment make it possible to distinguish municipalities characterised by different features. As a result of the above-mentioned method's application, 5 groups of municipalities were identified. The first consists of municipalities where micro and small enterprises are predominant, while investment expenditures are low. Moreover, in this group we can observe a slightly higher average salary in SMEs (60 municipalities in total). Another group consists of 23 municipalities with slightly more medium-sized enterprises and slightly higher average salaries. In these municipalities, low own revenues are observed, combined with high expenditures of local governments. The third and fourth groups consist of 35 and 24 municipalities respectively, with fewest SMEs, a low share of enterprises in services, but a higher share in industry, and relatively low per capita income of local governments. The fifth group includes two largest cities, Toruń and Bydgoszcz, with the largest number of enterprises, but also the best socio-economic situation in the region.

²⁴ The project was of exploratory nature, but the developed analysis methodology can also be used for in-depth research, e.g., on how gender pay gap depends on the section of economy in which a given enterprise operates.

Results of the segmentation are also indicative of significant variations in the condition of the SME sector in the particular groups of municipalities. This means that instruments used to support entrepreneurship development should be differentiated in each group. At the same time, it is worthwhile for municipalities with similar characteristics to cooperate in order to develop and implement good practices and projects in the field of entrepreneurship development.

Figure 3.4. Structure of municipalities in kujawsko-pomorskie voivodship by groups of municipalities according to Kohonen's SOM grouping method (data from 2018)

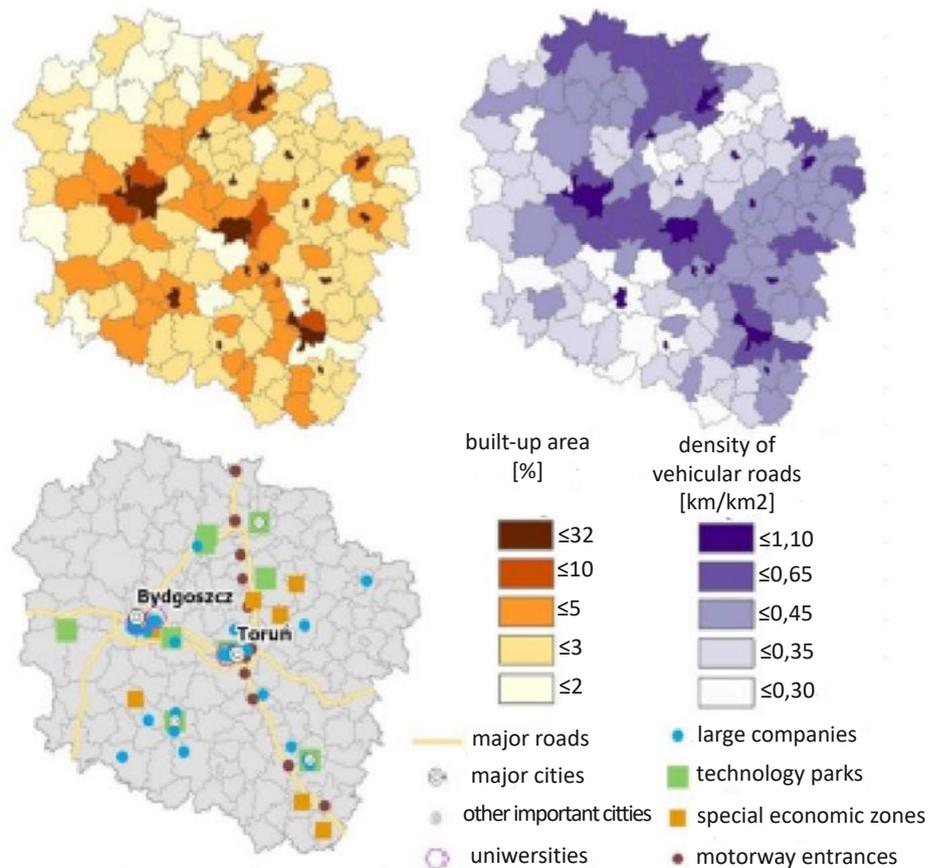


Source: (Ptak-Chmielewska and Chłóń-Domińczak, 2021)

Administrative data at the level of municipalities also allow **more in-depth spatial analyses**. Such analyses were carried out e.g., in (Chłoń-Domińczak et al. 2020). Spatial regression models were used in the analyses, taking into account characteristics like the share of built-up areas, agricultural areas and orchards, density of the road network and railway network, distance from major cities, industrial parks, special economic zones, academic centres and largest companies in the region (Figure 3.5.). In addition, economic and social characteristics of the municipalities were taken into account, such as population, its age structure, and unemployment. Another group of variables was the assessment of pro-entrepreneurship policies, as well as the number and value of European Union funded projects implemented in the municipalities.

The estimated analytical models showed that the share of built-up areas, road density, and rail network density were statistically significantly correlated with the dependent variable, which was the number of enterprises per 10 thousand inhabitants – the higher the value of these features, the higher the number of enterprises in relation to population. On the other hand, the relationship was negative for the share of agricultural areas, distance from Bydgoszcz and Toruń (for micro- and small enterprises), distance from large enterprises, technology parks, and the nearest university. The number of enterprises in relation to the number of inhabitants also depends on socio-economic characteristics. In municipalities with more inhabitants and higher own per capita income, there are also relatively more enterprises. On the other hand, higher unemployment is connected with lower density of enterprises. In the municipalities of kujawsko-pomorskie voivodship which pursue a pro-entrepreneurship policy and implement more EU-funded projects aimed at enterprise development, there are also relatively more micro- and medium-sized enterprises.

Figure 3.5. Main spatial characteristics of municipalities in kujawsko-pomorskie voivodship in 2018.



Source: (Chłoń-Domińczak et al. 2020)

Another application of the acquired administrative data was to **combine information from administrative records and data obtained from a sample survey** of enterprises, on innovation, competitiveness and productivity, conducted among enterprises in the voivodship. The survey covered 1,109 enterprises with 0-9 employees, 1,017 enterprises with 10-49 employees, and 269 enterprises with 50-249 employees (Zastempowski et al., 2020). Having both types of data made it possible to carry out an original experiment of combining them and building a statistical model that allows to forecast municipality's innovation index and identify key indicators related to the level of innovation (Przanowski, 2020). The results showed, among other things, that innovation is lower in municipalities where enterprises earn their revenues mainly from domestic sources, and higher in municipalities where enterprises pay higher VAT for selling goods and services to EU countries. Innovation is higher

also in municipalities where there are more contributors connected to the education sector, which may be indicative of the role of human capital in creating innovation at local level.

Apart from analyses conducted at the level of the voivodship and its municipalities, data obtained under the REGIOGMINA project were also used to **prepare a system of automatic reports** presenting basic information on the SME sector at municipality level, including comparisons with municipalities with similar characteristics, as well as with the voivodship's average. Municipal reports contain a summary of the main characteristics of a municipality: municipality type, population, income per capita, as well as municipality characteristics based on grouping conducted on the basis of data obtained from registers and official statistics. The report also presents synthetic information on the number of enterprises in particular (covered by the project) registers, number of enterprises / contributors by PKD section, persons insured according to the size of the contributor and its PKD section, as well as by age group and gender. Information on average contribution assessment bases by size of enterprise, gender, and PKD section is also presented. The reports also contain information about payers of income taxes, as well as their income. As part of the research phase of the REGIOGMINA project, reports were prepared for the years 2017 and 2018, and in the implementation phase, they were extended to cover the subsequent two years – 2019 and 2020. Ultimately, it is assumed that the reports will be extended to cover the subsequent years, by the Regional Economic Observatory established in the voivodship, whose purpose will be, inter alia, to collect and make available information on the SME sector at the local level, using the experience of the implemented project.

3.3. Conclusions and recommendations

The experience of the REGIOGMINA project and the obtained results of the analyses confirm, firstly, that it is possible to use register data to assess the state of the SME sector at the regional and local level. Data aggregated at municipality level allow the assessment of regional differentiation of the SME sector's potential and the influence of selected spatial and socio-economic characteristics on this potential.

Secondly, administrative data provide information supporting the diagnosis of the level of SME sector's development at municipality level, which would be impossible to carry out on the basis of data from public statistics (e.g., data published in the Local Data Bank of

Statistics Poland). The analyses prepared for kujawsko-pomorskie voivodship can be easily duplicated in other voivodeships, and obtaining information is not expensive. In particular, the experience of the REGIOGMINA project shows that appropriate use of register data meets the recommendations of the OECD (OECD, 2017) to increase knowledge about SME sector's development at local level.

Experience of the REGIOGMINA project allows us to recommend the use of information aggregated at municipal level for various spatial analyses. These analyses can help to identify factors related to the level of development of the SME sector at local level. The solutions can be used not only in kujawsko-pomorskie voivodship, but also in other regions, and enable local and interregional comparisons. Moreover, such analyses conducted in a long-term series may allow to monitor changes in the SME sector in the particular municipalities – not only in kujawsko-pomorskie voivodship, but also the whole country. Extension of the observation period to include the subsequent years significantly increases analytical possibilities (e.g., by enabling panel analyses), which may allow to identify the occurring changes and potential factors that influence them. As a result, it will be possible to formulate recommendations concerning policies and programmes supporting SME sector's development on local, regional, and national level.

References

- Chłoń-Domińczak, A., Ptak-Chmielewska, A. (2020). The use of register data to monitor micro, small and medium-sized enterprises' economic situation at the local level on the basis of the example of kujawsko-pomorskie voivodship. *Wiadomości Statystyczne. The Polish Statistician*, 65(6), 52–68. <https://doi.org/10.5604/01.3001.0014.2345>
- Chłoń-Domińczak, A., Fiedukowicz, A., Olszewski, R. (2020). Geographical and Economic Factors Affecting the Spatial Distribution of Micro, Small, and Medium Enterprises: An Empirical Study of the kujawsko-pomorskie region in Poland. *ISPRS International Journal of Geo-Information*, 9(7). <https://doi.org/10.3390/ijgi9070426>
- European Commission. (2008). Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions "Think Small First" A "Small Business Act" for Europe. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0394:FIN:en:PDF>
- NIK. (2017). Wspieranie przedsiębiorczości przez gminy. Informacja o wynikach kontroli, 1– 119.
- OECD. (2017). Making policy evaluation work. The Case of Regional Development Policy (OECD Science, Technology and Innovation Policy Papers No. 38).

- Przanowski, K. (2020). Model prognozy wskaźnika innowacyjności gminy.
- Ptak-Chmielewska, A., Chłoń-Domińczak, A. (2021). Spatial conditions supporting sustainable development of enterprises on local level. *Sustainability (Switzerland)*, 13(4), 1–15. <https://doi.org/10.3390/su13042292>
- Stano, P. M., & Ghisett, C. (2016). Monitoring Europe SMEs ' performance in Europe. Methodological assessment of the SME Scoreboard 2016. JRC Technical Reports – EUR 28244 EN. <https://doi.org/10.2788/940531>
- Zastempowski, M., A. Glińska-Neweś, E. Wędrowska, W. Glabiszewski, D Grego-Planer, I. Escher, K. Liczmańska-Kopcewicz. (2020). Sektor mikro-, małych i średnich przedsiębiorstw w województwie kujawsko-pomorskim. Wydawnictwo naukowe Uniwersytetu Mikołaja Kopernika, Toruń <http://regiogmina.kujawsko-pomorskie.pl/wp-content/uploads/2020/10/Sektor-mikro-malych-i-srednich-przedsiębiorstw-w-województwie-kujawsko-pomorskim.pdf>

4. Use of register data for development of a predictive model to enable identification of companies with high probability of success of funded projects

Aneta Ptak-Chmielewska

Collegium of Economic Analysis

Warsaw School of Economics

Introduction

A number of public policy instruments at the central, regional, and local level aim to support enterprise development. The instruments include guarantees, loans, and subsidies e.g., for research and development. One of the significant challenges for such support is appropriate selection of project beneficiaries to ensure high chances of their success. The success, depending on the support instruments, may be understood in many ways, for instance as:

- project implementation without agreement termination (meeting minimum implementation conditions);
- achievement of all the objectives/indicators of the project within the planned timeframe;
- significant improvement of selected indicators of enterprise's operation (enterprise development understood as growing revenue, profit, employment, etc.);
- keeping the newly established enterprise on the market for a defined period of time;
- avoiding the risk of enterprise bankruptcy.

However we define project success, tools (e.g., analytical models) that can support the decision-making process regarding co-financing of the particular projects and entities (potential beneficiaries) are useful. Such tools should be based on the widest possible

knowledge of enterprise operation, based on data on the enterprise sector from various sources (including administrative).

Application of such tools requires access to information from reliable and complete sources, whose acquisition should be cost-effective. An important source that meets these criteria is administrative data.

There is a growing number of sources and large data sets on the operation of companies, in particular in administrative registers, whose accumulated information resources are partly available to the public. This chapter presents a brief overview of the kind of information that is currently available in registers covering the population of enterprises (all, or its significant part) and discusses which of this information can potentially be used to assess the condition of enterprises and the chances of their projects succeeding. The chapter also offers an assessment of the opportunities and limitations associated with the use of data that are collected but not yet used for analytical purposes in Poland. The particular sections of the chapter present examples of models that support the assessment of the potential success of enterprises (models for assessing SMEs' resilience to crisis) and propose measures for extending their use to include new analytical purposes.

4.1. Enterprise data in public registers

In Poland, data on enterprises are collected in a number of administrative registers that serve different public policy purposes concerning enterprises.

REGON

The main register covering the population of enterprises is the **National Official Business Register (REGON)** established pursuant to the provisions of the Act on Public Statistics.

REGON Register:

- serves to achieve identification coherence of national economy entities which are recorded in other official registers and information systems of public administration,
- provides uniformity of descriptions used in conceptual and classification nomenclature,

- provides general characteristics of entities operating in the national economy, from the following categories: territorial, ownership, types of activity, legal forms, etc.,
- allows the list of addresses of active entities to be drawn up,
- is the basis for compilation of databases on national economy entities,
- is one of the important sources for supplying the database of units selected for statistical research (sampling).

The register is not a direct source of information about the current economic standing of enterprises, but it makes it possible to: generate random samples for enterprise surveys, describe enterprises in general according to the attributes present in REGON, and geolocate enterprises.

Access to REGON data can be obtained in at least three ways. The first one is to apply to Statistics Poland for access to data of a specified population of entities (e.g., all entities from a given voivodeship). The data may be made available to institutions defined in the Act for purposes of their statutory tasks. The institutions include: agencies maintaining other official registers and information systems of public administration, government administration authorities and local authorities (JST), other government institutions, the Supreme Audit Office, the National Bank of Poland, the President of the Personal Data Protection Office, the State Labour Inspectorate, and the State Sanitary Inspection²⁵.

The second way is to use REGON's Internet Database²⁶, which displays data for a selected, single entity. Obtaining data requires stating a specific NIP, REGON, or KRS number.

The third way is to obtain data via the REGON API (Web Application Programming Interface)²⁷.

²⁵ On the basis of this authorisation, it is e.g., possible for JST to obtain information on enterprises in order to carry out public tasks related to supporting enterprise development. For more information, refer to the chapter on using data collected in ZUS and tax registers to monitor the condition of the SME sector at the local level – the REGIOGMINA project.

²⁶ Access: <https://wyszukiwarkaregon.stat.gov.pl/>

²⁷ More on this in the chapter on examples of supporting the execution and monitoring of public policies based on administrative data – the Integrated Analytical Platform project.

VAT Register

Another register which contains information about enterprises is the **VAT Register**. This is an electronic, free of charge database of VAT taxpayers, containing data on taxpayers who: (i) have been registered as active or exempt VAT taxpayers, (ii) have not been registered or have been removed from the register, (iii) have been reinstated as VAT taxpayers. This database is maintained by the Head of the National Revenue Administration (KAS). It is used, among others, to provide the so-called whitelist of VAT taxpayers²⁸. The list is effective from 1 September 2019. It is a register which helps companies verify contractors quickly and effectively. The VAT Register replaced the previously functioning lists of VAT taxpayers: registered, unregistered, as well as removed and reinstated in the VAT Register. The whitelist makes it possible to:

- check whether a company is an active VAT taxpayer;
- find out why a company was refused registration, removed from the register, or reinstated as VAT taxpayer
- confirm the bank account number to which payments are to be made to a contractor.

Using the list requires searching for information by taxpayer identification number (NIP), REGON, bank account number, or company name. This means that it is not possible to obtain a database of the entire population of enterprises contained in this register without knowing their identification numbers, but it is possible to obtain information on specific, known entities (e.g., applying for support).

Central Register of Entities – National Register of Taxpayers

Central Register of Entities – National Register of Taxpayers (CRP-KEP) also belongs to the group of registers related to tax administration. CRP-KEP keeps records of taxpayers who pay taxes on goods and services, and income taxes on natural persons and legal entities. These records include information collected on the basis of the submitted tax declarations (PIT, CIT, and VAT), including information about revenues, tax deductible costs, taxpayers' income, in accordance with the scope of information specified in tax declarations. Particularly noteworthy is the information provided in the VAT Standard Audit File for Tax (JPK). As of 1

²⁸ Access: <https://www.podatki.gov.pl/wykaz-podatnikow-vat-wyszukiwarka>

July 2018, all taxpayers who keep tax books and prepare accounting evidence in electronic form submit JPK structures at the request of tax authorities. The on-demand JPK includes 7 structures:

- accounting books – JPK_KR,
- bank statements – JPK_WB,
- warehouse – JPK_MAG,
- VAT invoices – JPK_FA,
- VAT RR invoices (of a flat-rate farmer) – JPK_FA_RR,
- tax revenue and expense ledger – JPK_PKPIR,
- revenue register – JPK_EWP.

Data from PIT income taxes, including information about the income of a taxpayer applying lump-sum income tax on registered income from activity conducted in the form of a partnership(s) of natural persons, civil or general (PIT 28B), income of natural persons who conducted non-agricultural business activity subject to taxation on a general basis in accordance with the tax scale or special divisions of agricultural production subject to taxation on a general basis in accordance with the tax scale (PIT 36) and the amount of income earned (loss incurred) by a corporate income taxpayer (CIT-8), provide annual information concerning, i.a. corporate income depending on the type and scale of business and legal form. The rules of maintaining the register and making the data available are set out in the Act on the rules for recording and identification of taxpayers and payers. According to Article 15 of the Act, data collected in tax registers may be made available only to the following: courts, prosecutors, tax authorities, customs authorities, representatives of the Supreme Audit Office, the Police, the Internal Security Agency, the Central Anti-Corruption Bureau, the Military Counterintelligence Service, and the Military Police – in connection with ongoing proceedings; the Head of the National Centre for Criminal Information, for their statutory tasks; judicial enforcement officers and enforcement bodies specified in separate provisions – in connection with conducted enforcement or insolvency proceedings or execution of a decision on securing the estate or drawing up an inventory; the Internal Supervision Inspector – for their statutory tasks; supreme and central government administration authorities – in connection with proceedings for granting a business license or a permit to conduct business activity; a taxpayer – in the scope concerning their number; public statistical services – in the scope and for the needs resulting from the provisions on public statistics; Social Insurance Institution (ZUS) and the State Fund for Rehabilitation of Disabled Persons (PFRON) – for the execution of tasks and purposes specified in separate

acts; authorities maintaining official registers pursuant to separate regulations, upon request including data necessary to identify entities referred to in Article 2, within the scope of NIP – in connection with the execution of their statutory tasks; the clearing house within the meaning of Article 119zg item 3 of the Act of 29 August 1997 – Tax Ordinance. In most cases, data are made available for control purposes. Currently, use of information collected in tax registers for the evaluation of public policies is not planned.

National Court Register

Another register which collects information on the condition of enterprises in Poland is the **National Court Register (KRS)**. The National Court Register is a centralised IT database consisting of three registers: (1) the register of entrepreneurs, (2) the register of associations, other social and professional organisations, foundations and public health care institutions, and (3) the register of insolvent debtors. KRS was established pursuant to the Act of 20 August 1997 on the National Court Register.

Its purpose is to provide universal access to fast and reliable information on the legal status of a registered entity (KRS Central Information), most important elements of its financial standing, and the way it is represented. Additionally, KRS contains other data on the entrepreneur, important for economic transactions, including information on tax and customs duty arrears, arrears in ZUS contributions, creditors, and amounts of outstanding debts.

For each entity entered into KRS, courts keep registration files which include, in particular, documents confirming the information entered in the register, and specimens of signatures of persons authorised to represent the entity. Any interested party may, at the seat of the competent court, browse registration files of entities entered in the register, with no limitations. As of 2018, data on, i.a. financial statements are collected only in electronic form and are available through the EKRS Financial Document Browser²⁹. However, as in the case of VAT taxpayers whitelist, information on the particular entities is available only for specified identification numbers. For instance, without a list of identification numbers, it is impossible to obtain data on the population of enterprises from a selected area. Financial data of

²⁹ Access: https://ekrs.ms.gov.pl/rdf/pd/search_df

companies (including SMEs) are available in KRS on a relatively current basis (as a rule, just after the end of the financial year).

It should be noted that before 2018, KRS data were also available, but not in database form. They were processed by business intelligence services, e.g., Beck KRS which, as part of the Legalis system, provided electronic access to data collected from scans of paper documents available in KRS or Monitor Polski B.

KRS register provides full information on an entity's financial statements, including data from the balance sheet, profit and loss account, statement of changes in equity, and cash flow statement.³⁰ The available KRS data make it possible to obtain financial indicators useful from the point of view of the subject of this chapter.

The **National Register of Insolvent Debtors** (KRDN) is also an element of KRS³¹. KRDN is a state institution established within the National Court Register, based on the provisions of the Act of 20 August 1997 on the National Court Register (Polish Journal of Laws of 2001 no 17, item 209, as amended). Establishment of the register is the legislator's response to the need for instruments that will help make business transactions more reliable and protect businesses and citizens against dishonest and insolvent persons. Through the KRDN portal, one can verify specific companies or natural persons. For those interested in using the service on a wider scale, KRDN offers a paid service in the form of access packages and comprehensive service of preparing letters and documents necessary to recover debts via the KRDN service.

Another supplement to KRS is the **Court and Commercial Journal** (MSiG)³², which contains information on entities in respect of which a bankruptcy or restructuring motion has been filed. Information included in MSiG is published and available in the form of daily summaries in PDF format (with separate document being published for each day of the year). A commercial tool facilitating work with data collected in MSiG is the iMSiG (Court and Commercial Journal Online).

³⁰ More on the content and analytical value of company data collected in KRS is presented in the chapter on the use of company financial statements in programme evaluation.

³¹ Access: <https://krdn.pl/>

³² Access: <https://ems.ms.gov.pl/msig/przegladaniemonitorow>

Central Register of Social Contribution Payers

Another administrative register collecting data on enterprises is the **Central Register of Social Contribution Payers (Comprehensive Information System of the Social Insurance Institution)** established on the basis of the provisions of the Act on the Social Insurance System. The purpose of the register is to support statutory activities of ZUS, in particular those related to the recording of contributions on the accounts of insured persons, record keeping, entitlement determination, granting, calculation of payment amounts, and payment of social insurance benefits. Among others, the system makes it possible to carry out multidimensional, interactive analyses of data in the area of registering the insured and collecting receivables for social insurance contributions. Data from the register's sets are made available on the basis of an application from authorised institutions, either in electronic or paper form. These data are made available mainly for the operation of the social insurance system, and for control purposes. The register of contribution payers (entrepreneurs) collects territorial information about the seat of the company, form of ownership, type of business activity, data on insured persons, including names of insurance (e.g., employment contract, civil law contract), the amount of contribution assessment bases for particular insurances (retirement, disability, sickness, accident). Moreover, information is collected on the timeliness of social security contributions payments and arrears incurred. Analysis of longitudinal data on contributions' payers can also support knowledge on the development of the company (e.g., increase in employment, increase in salaries). Due to the lack of formal basis for using this information, as a rule, data collected in ZUS registers are not used for analytical purposes related to the assessment of enterprises' operation. The exceptions here are analyses conducted by ZUS and the Ministry of Finance.

Business intelligence services

Business intelligence services are also a source of data on the financial standing of enterprises, including financial statements. However, these are commercial sources (partially or fully paid) and do not include information on the entire population of enterprises. Nevertheless, the data, studies, and sector reports published and made available by these services can be a source of complementary information. Examples include:

- iMSiG³³ provides access to financial statements submitted to the Repository of Financial Documents in the National Court Register. Under the financial statements service, a daily updated set of data is made available on the profiles of particular entities registered in the National Court Register.
- EMIS³⁴ (Emerging Markets Information Service) is a service containing extensive economic, financial, and political information on Poland and other countries of Central and Eastern Europe. It provides full texts of press and agency articles, financial reports of public and non-public companies as well as various analyses and industry reports. Users can also take advantage of stock market quotations, statistics, and data on the particular sectors of the economy as well as macroeconomic data provided by the service.
- ORBIS³⁵ is an expert service that collects, standardizes, and provides information on over 400 million companies from all countries of the world – including detailed financial information on 41 million companies.

Despite the development of administrative data resources, both in the past and at present, access to complete databases at unit level is restricted. This is mainly due to legal and organisational constraints. If data are made available, using them on a wider scale frequently requires considerable effort. For instance, access to financial statements of the EKRS involves downloading unit files (manually), or using a special script. Such solutions are applied, but these are only a few examples. Apart from them, studies conducted by the Polish Economic Institute (PIE) can be mentioned, which, during the economic crisis related to the coronavirus pandemic, analysed enterprise data in weekly cycles, using data from the Ministry of Finance³⁶. These are changes that have taken place in recent years that have altered the situation and use of this data. Nevertheless, using the full potential of these information resources continues to be impossible and remains a challenge. On the one hand, this is due to the analytical potential of administrations and data hosts. On the other hand – there are formal restrictions in data processing for purposes other than the original purpose of data

³³ Access: <https://www.imsig.pl/>

³⁴ Access: <https://www.emis.com/pl>

³⁵ Access: <https://www.bvdinfo.com/en-gb/our-products/data/international/orbis>

³⁶ See https://www.youtube.com/watch?v=ZgFdzm5dXBk&list=PLanTI4bY8k77nPjS_zPu-qz86FXXz85aU&index=1 (from 31 min.)

transfer by data subjects, limitations in transfer of data, or combining information from different databases.

4.2. Examples of models supporting the assessment of potential success of enterprises: models for assessing SMEs' resilience to crisis

This section presents a potential approach to the use of enterprise data for building models providing decision-making support for projects or instruments aiding enterprise development. The key in building a model is to define what we want to find out, e.g., by selecting an dependent variable that will be explained on the basis of the data we have (independent variables). When developing a model for predicting project success, it is necessary to determine success definition – e.g., meeting the requirements specified in the agreement and achieving all contractual indicators, positive return on investment, or meeting certain enterprise development indicators. The above has not yet been a subject of forecasting models known to the authors. However, the process of their development may be based on modification of forecasting models dedicated to similar issues:

- Models used to assess the risk of survival or bankruptcy of an enterprise, used as an early warning system in business.
- Models used to assess the chances of liabilities' repayment (or the risk of non-payment) in financial institutions (mainly banking, e.g., scoring models widely used and developed by banks to support credit decision-making) (Matuszyk, 2008; Thomas et al., 2002; Anderson, 2007).³⁷

The literature on bankruptcy predictions in Poland is rich and its beginnings date back to the 1990s (the transition period, when researchers began to take interest in the topic due to the observed economic phenomena. One of the first authors to apply the Altman model (Altman,

³⁷ Credit scoring methodologies are statistical verification procedures aiming to assess customers and classify them under one of the two categories: (1) bad customer, (2) good customer. Scoring models aim to predict whether a bank customer will repay their credit commitments in the future.

1968) to Polish enterprises was E. Mączyńska (1994). At that time, authors mainly applied financial indicators and focused on models dedicated to domestic entities (Wędzki, 2000; Stępień, Strąk, 2003; Prusak, 2005; Pogodzińska, Sojak, 1995; Gajdka, Stos, 1996; Hadasik, 1998; Wierzba, 2000). They mainly used discriminant analysis and based their models on data obtained from small samples. Later analyses were extended to include larger samples and more advanced statistical modelling techniques, such as logit models (Hołda, 2001; Sojak, Stawicki, 2001; Gruszczyński, 2003; Michaluk, 2003; Mączyńska, 2004; Appenzeller, Szarzec, 2004; Korol, 2004; Hamrol et al., 2004; Wędzki, 2004; Stępień, Strąk, 2004; Prusak, Więckowska, 2007; Jagiełło, 2013; Pocięcha et al., 2014; Karbownik, 2017). Logit models belong to the group of qualitative models, where the variable is explained by a qualitative variable (related to the occurrence of a specific phenomenon, e.g., risk of project failure).

Recent works also include data mining and survival models (Pocięcha et al., 2014; Ptak-Chmielewska, 2016; Korol, 2010b; Gąska, 2016; Zięba et al., 2016). Numerous models were estimated for different sectors of the economy and company sizes (Jagiełło, 2013; Brożyna et al., 2016; Balina, Bąk, 2016; Karbownik, 2017), based mainly on financial data from a sample of analysed companies. Works and models that took account of macroeconomic variables and non-financial variables were much less common (Korol, 2010a; Ptak-Chmielewska, Matuszyk, 2017). Few studies considered the economic cycle, suggesting the inclusion in models of measures reflecting current economic conditions (Pocięcha, Pawełek, 2011). One example of empirical application was the analysis of bankruptcy as one of the reasons for discontinuation of company operations (Pocięcha, Pawełek, 2011; Nehrebecka, Dzik, 2013).

However, models for assessing solvency of enterprises are mainly internal models used by banks and financial institutions and developed for internal capital and provision management. In research literature, numerous applications of this type of models with data of Polish enterprises can be found. Jagiełło, 2013, used discriminant analysis and logistic regression models for enterprises by industry. The works of Matuszyk and Ptak-Chmielewska (2015), were devoted to the use of traditional methods and machine learning methods. On the other hand, data mining methods were applied to assess the chances of failure (corporate insolvency) in the work of Ptak-Chmielewska (2019).

Two selected models developed on data for Polish companies are presented below. They provide examples of the use of companies' financial data in assessing companies' resilience and risk of bankruptcy (or the opposite – chances of continued operation). The first model

is based on a large set of data (from the business intelligence service EMIS) and predicts bankruptcy of companies. The second model is based on a small set of data (NBP internal data) and forecasts insolvency of enterprises. The selected approaches enable presentation of the existing practices of data sets' use in forecasting specific phenomena. As noted earlier, modification of the presented models can serve as a starting point for the development of models predicting success of projects (including those co-financed from European funds).

Enterprise bankruptcy forecasting model

Among studies conducted in Poland, the approach based on **large data sets** proposed by a team led by Józef Pocięcha (Pocięcha ed., 2013) is worth mentioning. The proposal included a model built on the basis of data from the EMIS service, covering company reports from the 2005-2009 period (and bankruptcies in the 2007-2010 period), i.e., including the 2007-2009 financial crisis. Ultimately, 35 financial indicators were designed and used as the model's input variables. The database contained information on 1,719 so-called "healthy" industrial processing enterprises and 133 so-called "bankrupts" (of which as many as 65 went bankrupt in 2009).

An attempt was made to forecast bankruptcies within one- or two years from the date of the financial statement³⁸. The collected data were divided into a section dedicated to model development and section dedicated to testing³⁹. These procedures can also be applied to registry sets, e.g., data available in the eKRS system. The following model estimation methods were used: linear discriminant function, logistic regression, classification trees, and neural networks.

Noteworthy is the variant of the bankruptcy forecast model taking the perspective of two years prior to the event, where the authors took into account reports from only one year,

³⁸ It is worth mentioning that analyses of smaller subsamples of data and the so-called matching of companies at risk of bankruptcy with "healthy" companies with similar characteristics (e.g., company size, sector, or age), but different from those used in the model, were also considered.

³⁹ The intention was to build a universal model, one that would operate accurately not only on the set of data for the adopted period. For this purpose, the set of available data is divided into two subsets, with the first one being used to develop the model. The second ("test") subset is used to verify the accuracy of the model's operation. This makes it possible to check whether the developed model has similar accuracy also in other data sets (not only the one used to develop the model).

i.e., 2007, and bankruptcy perspective in the years 2008-2009, i.e., during the financial crisis. The application of such a model provided an opportunity to determine enterprises' resilience to the global crisis. The approach applied in this analysis could also be used in analyses of project success from the perspective of **enterprises' resilience to unfavourable external environment** during project execution.

It is worth demonstrating that few variables are needed to develop such a model. As indicated by the description of the equation based on logistic regression (Pociecha ed., 2014, p. 125), the main categories from the financial statements were used to assess risk⁴⁰:

$$P(Y = 1) = \frac{1}{1 + e^{-(5,1560 - 8,8165 \cdot R_{10} - 1,1540 \cdot R_{14} - 0,1909 \cdot R_{18} + 7,2012 \cdot R_{28})}}$$

where:

$P(Y=1)$ – bankruptcy probability,

R_{10} – R10 – fixed assets/balance sheet total

R_{14} – R14 – (equity + long-term liabilities)/fixed assets

R_{18} – 100*net result/net sales revenues

R_{28} – stock/net sales revenues.

The model output determines the probability of bankruptcy within two years from the date of financial statement (based on 2007 financial statements only)⁴¹. The probability is in the [0,1] range. The higher the probability of bankruptcy, the lower the company's resilience to crisis.

The model, based on logistic regression, obtained correct classification in 81.82% of cases (for both bankrupts and non-bankrupts). For data from an independent "test" subset (not used to develop the model), model accuracy was only slightly lower.

⁴⁰ The coefficients in the regression model are the result of the model's estimations using empirical data.

⁴¹ Use of data from only one year already offers the possibilities of analyses that can be used for bankruptcy predictions. Availability of data covering several periods is not a prerequisite for modelling.

Crisis resilience model

Another approach, based on a set of data for 400 entities, is proposed by Robert Jagiełło (Jagiełło, 2013). It consists in the development of industry models based on financial indicators from the 2008 end-of-year statements, for entities for which the National Bank of Poland recorded irregular payments in 2009 (i.e., a negative event in the banking system). The year 2009 was chosen due to the model's presentation, which referred to the companies' resilience to the economic crisis.

The author confirmed high efficiency of this approach, also in the case of models developed based on small samples using the traditional approach based on discriminant function and logistic regression. However, the biggest advantage of these models was the division into sectors of activity. A separate model was estimated for each of the main sectors: industry, construction, trade, transportation, and services. The applied logistic regression models allowed estimation based on small subsamples of enterprises. The models used **16** uncorrelated **indicators** (presented in Table 4.1.).

Table 4.1. Financial indicators used in company insolvency prediction model

No.	Indicator name	Indicator design
1	Current liquidity	$\frac{\text{current assets}}{\text{short - term liabilities}}$
2	Fast liquidity	$\frac{\text{current assets} - \text{stock} - \text{RMK}^{42}}{\text{short - term liabilities}}$
3	Cash flow liquidity	$\frac{\text{cash}}{\text{short - term liabilities}}$
4	Share of operating capital in assets	$\frac{\text{current assets} - \text{short term liabilities}}{\text{total assets}}$
5	Gross margin	$\frac{\text{gross result on sales}}{\text{operating costs}}$
6	Operating profitability of sales	$\frac{\text{operating result}}{\text{total revenue}}$
7	Operating profitability of assets	$\frac{\text{operating result}}{\text{total assets}}$
8	Net profitability of equity	$\frac{\text{net result}}{\text{equity}}$
9	Asset turnover	$\frac{\text{total revenue}}{\text{total assets}}$

⁴² Short-term prepayments

No.	Indicator name	Indicator design
10	Current asset turnover	$\frac{\text{total revenue}}{\text{current assets}}$
11	Receivables turnover	$\frac{\text{total revenue}}{\text{liabilities}}$
12	Stock turnover	$\frac{\text{total revenue}}{\text{stock}}$
13	Share of equity in liabilities (capital ratio)	$\frac{\text{equity}}{\text{total liabilities}}$
14	Coverage of short-term liabilities with equity	$\frac{\text{equity}}{\text{short term liabilities}}$
15	Coverage of fixed assets with equity	$\frac{\text{equity}}{\text{fixed assets}}$
16	Share of net financial surplus in total liabilities	$\frac{\text{net result} + \text{depreciation} + \text{interest}}{\text{total liabilities}}$

Source: Jagiełło (2013), pp. 21-22.

Of the 16 indicators used in the models, 11 appeared to be relevant for the assessment of enterprises' solvency, while the remaining five appeared to be irrelevant and were excluded from further models. The indicators used in solvency probability models for the particular sectors are presented in Table 4.2.

Table 4.2. Indicators in solvency probability model – sectoral models

Coefficient	industry	construction	retail	transportation	services
Fast liquidity					++
Cash flow liquidity			++		
Share of operating capital in assets		++		++	
Gross margin	++		++		++
Net profitability of equity		++		++	
Asset turnover	+		+		
Current asset turnover		+		+	
Receivables turnover					+
Share of equity in liabilities (capital ratio)	++	++			
Coverage of short-term liabilities with equity			+		+
Share of net financial surplus in total liabilities				++	

Source: Based on Ptak-Chmielewska, Kuleta, 2018, p. 100.

It should be noted that the set of indicators applied varies from sector to sector. **Gross margin** and **capital structure** (capital ratio) are most important when assessing the resilience of industrial enterprises. On the other hand, when assessing the condition of enterprises from the construction sector, the **share of operating capital** and the **share of equity**, as well as the **return on equity** in general, are of greatest importance. At the same time, enterprises from the trade sector are most vulnerable to **liquidity** and the **level of cash**, and their leverage in assessing the condition is **gross margin**. In the assessment of transport enterprises, **return on equity** is most important, but also the **share of operating capital** and **financial surplus in relation to liabilities**. In the assessment of their vulnerability to crisis, companies from the service sector are characterised by **liquidity** and **gross margin** indicators. Effectiveness of the developed models can be assessed by the share of accurate forecasts for the group of solvent and insolvent companies, presented in Table 4.3. Effectiveness was particularly high in the case of industrial companies, but in each of the sectors forecasts' accuracy was equal to or higher than 85%.

Table 4.3. Effectiveness of solvency forecasting models (Jagiełło, 2013)

	Industry	Construction	Retail	Transportation	Services
Solvent	97.5%	85.0%	87.5%	87.5%	87.5%
Insolvent	90.0%	90.0%	85.0%	87.5%	92.5%

Source: Jagiełło, 2013, p. 95 et seq.

In summary, Professor Pocięcha's model makes it possible to estimate the probability of bankruptcy: *the higher the probability of bankruptcy, the lower the resilience to crisis and success in terms of project execution*. Professor Jagiełło's model makes it possible to estimate the probability of solvency: *the higher the probability of solvency, the higher the enterprise's resilience to financial crisis*.

The presented Pocięcha and Jagiełło models can be recommended as ready-made algorithms for identifying enterprises resilient to crisis. Therefore, the models may support the process of selecting projects for co-financing under programmes financed from the European funds, in particular, qualification of enterprises for development support of a specific type, controlling the probability of bankruptcy or loss of liquidity, depending on the purpose of support, allowing or limiting participation of such entities in the support scheme.

The above models do not directly involve evaluation of the implemented project's chances of success (as mentioned above, Professor Pocięcha's model makes it possible to estimate the probability of bankruptcy, while Professor Jagiełło's model makes it possible to estimate the probability of solvency). However, both can serve as a starting point for the development of new prediction models, predicting the prospects of project success (according to a specific definition of project success, depending on the type of support to be granted). While applying the discussed models in practice one should:

- Generate, for the selected support measures to be implemented, a database with information on completed projects, including at least the beneficiary's identification number (e.g., NIP or KRS) and information on project success or failure (in accordance with the previously adopted definition).
- The above-mentioned database should be supplemented with data to enable calculation of indicator values applied in the above-discussed models (including e.g., data from financial statements available in the eKRS system).
- Recalibrate the models discussed above, i.e., recalculate them, whereby the independent variables will be the indicators applied in the previous model (supplied with the new database), and the dependent variable – project success (in accordance with the previously adopted definition).
- If the accuracy of the new model is verified (on the basis of historical data), the model can be used to support the decision-making process in future calls for proposals or newly implemented support measures.

It should be noted that basing models on register data collected in a systematic way may be relatively simple and cheap. The system of continuous collection of financial data about companies and monitoring of the condition of the subsidized entities, including the outcomes of projects which they implemented, may also enable assessment of the extent to which the models' predictions are accurate and, consequently, improve them to increase their future predictive capacity.

Summary

The chapter presented the existing administrative sets of data on enterprises, their information resources, and the practices used to access these data. It also discussed the possibilities of using large data sets on enterprises' condition to build models supporting

decisions on the provision of support and assessment of risks associated with project execution. The presented approach seems to be particularly useful for analyses using administrative data. These are models for assessing bankruptcy probability (Pociecha, 2014) and crisis resilience (Jagiełło, 2013). The chapter proposed the use of the discussed models as tools for preliminary assessment of enterprise resilience and preselection in the system of selecting enterprises' projects for funding under specific support instruments.

Identification of crisis-resilient enterprises makes it possible to effectively limit the costs of failed investments and maximise the benefits from investments that have high chances of success. It should be emphasized that the discussed models can also be adjusted to assess the chances of success of individual enterprises depending on the adopted definition of success, specific to a given group of support instruments.

References

- Altman, E.I., 1968. Financial ratios, Discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance* 23(4): 589-609.
- Anderson, R. (2007), *The Credit Scoring Toolkit. Theory and Practice for Retail Credit Risk Management and Decision Automation*, Oxford.
- Appenzeller, D., Szarzec, K., 2004. Forecasting the bankruptcy risk of Polish public companies. *Rynek Terminowy* 1, pp. 120–28.
- Balina, R., Bąk, M.J., 2016. Discriminant Analysis as a Prediction Method for Corporate Bankruptcy with the Industrial Aspects. Waleńców: Wydawnictwo Naukowe Intellect.
- Brożyna, J., Mentel, G., Pisula, T., 2016. Statistical methods of the bankruptcy prediction in the logistics sector in Poland and Slovakia. *Transformations in Business & Economics* 15, pp. 80–96.
- Gajdka, J., Stos, D., 1996. Wykorzystanie analizy dyskryminacyjnej do badania podatności przedsiębiorstwa na bankructwo in: *Przedsiębiorstwo na rynku kapitałowym*, (ed.) Duraj, J. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Gąska, D., 2016. Predicting Bankruptcy of Enterprises with the use of Learning Methods. Ph.D. dissertation, Wrocław: Wrocław University of Economics.
- Frątczak, E. (ed.), 2009. *Wielowymiarowa analiza statystyczna. Teoria i przykłady zastosowań z systemem SAS*. Warsaw: Warsaw School of Economics.
- Gruszczyński, M., 2003. Models of microeconometrics in the analysis and forecasting of the financial risk of enterprises. Warsaw: *Zeszyty Polskiej Akademii Nauk* no. 23.
- Hadasik, D., 1998. Upadłość przedsiębiorstw w Polsce i metody jej prognozowania. Poznań: Wydawnictwo AE w Poznaniu, Zeszyt 153.

- Hamrol, M., Czajka, B., Piechocki, M., 2004. Enterprise bankruptcy—discriminant analysis model. *Przegląd Organizacji* 6, pp. 35–39.
- Hołda, A., 2001. Prognozowanie bankructwa jednostki w warunkach gospodarki polskiej z wykorzystaniem funkcji dyskryminacyjnej ZH, *Rachunkowość*, no. 5/2001.
- Jagiełło, R., 2013. Analiza dyskryminacyjna i regresja logistyczna w procesie oceny zdolności kredytowej przedsiębiorstw, *Warsaw: MATERIAŁY I STUDIA*, Zeszyt no. 286.
- Karbownik, L., 2017. *Methods for Assessing the Financial Risk of Enterprises in the TSI Sector in Poland*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Korol, T., 2004. *Assessment of the Accuracy of the Application of Discriminatory Methods and Artificial Neural Networks for the Identification of Enterprises Threatened with Bankruptcy*. Gdańsk: Doctoral dissertation.
- Korol T., 2010a. *Early Warning Systems of Enterprises to the Risk of Bankruptcy*. Warsaw: Wolters Kluwer.
- Korol, T., 2010b. Forecasting bankruptcies of companies using soft computing techniques. *Finansowy Kwartalnik Internetowy e-Finanse* 6, pp. 1–14.
- Matuszyk, A. (2008), *Credit Scoring*, CeDeWu, edition II.
- Matuszyk, A. Ptak-Chmielewska, A., Profile of the fraudulent customer; *Bezpieczny Bank*; no. 2(59)/2015, pp. 7-23.
- Mączyńska, E., 1994. Assessment of the condition of the enterprise. Simplified methods. *Życie Gospodarcze* 38: pp. 42-45
- Mączyńska, E., 2004. Early warning systems. *Nowe Życie Gospodarcze* 12, pp. 4–9.
- Michaluk, K., 2003. Effectiveness of corporate bankruptcy models in Polish economic conditions. In *Corporate Finance in the Face of Globalization Processes*. Edited by Leszek Pawłowicz and Ryszard Wierzb. Warsaw: Wydawnictwo Gdańskiej Akademii Bankowej.
- Nehrebecka, N., Dzik, A.M., 2013. Zdolność przetrwania przedsiębiorstw w Polsce. *Wiadomości Statystyczne*, no. 5.
- Pocięcha, J. (ed.), 2014. *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*. Kraków: Fundacja Uniwersytetu Ekonomicznego w Krakowie.
- Pocięcha, J., Pawełek, B., 2011. Prognozowanie bankructwa a koniunktura gospodarcza. *Metody analizy danych*. Kraków: Zeszyty Naukowe UE w Krakowie, no. 873.
- Pogodzińska, M., Sojak, S., 1995. Wykorzystanie analizy dyskryminacyjnej w przewidywaniu bankructwa przedsiębiorstw w AUNC. *Toruń: Ekonomia XXV*, Zeszyt 299.
- Prusak, B., 2005. *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*. Warsaw: Difin.
- Prusak, B., Więckowska, A., 2007. Multidimensional models of discriminant analysis in the study of the bankruptcy risk of Polish companies listed on the WSE. In *Economic and Legal Aspects of Corporate Bankruptcy*. Edited by Błażej Prusak. Warsaw: Difin.
- Ptak-Chmielewska, A., *Predicting Micro-Enterprise Failures Using Data Mining Techniques*. *Journal of Risk and Financial Management* 2019, no. 12, 0; pp. 1-17

- Ptak-Chmielewska, A., 2016. Determinanty przeżywalności mikro i małych przedsiębiorstw w Polsce. Warsaw: Oficyna Wydawnicza SGH.
- Ptak-Chmielewska, A., Kuleta, P., 2018. Statistical models in enterprises default risk assessment – an example of application. *Econometrics, Advances in Applied Data Analysis*, no. 1(59), pp. 94–106.
- Ptak-Chmielewska, A., Matuszyk, A., 2017. The importance of financial and non-financial ratios in SMEs bankruptcy prediction. *Bank i Kredyt* 49(1), pp. 45-62.
- Sojak, S., Stawicki, J., 2001. Wykorzystanie metod taksonomicznych do oceny kondycji ekonomicznej przedsiębiorstw, *Zeszyty Teoretyczne Rachunkowości*, vol. 3 (59), ed. Bednarski, L. Warsaw.
- Stępień, P., Strąg, T., 2003. Signs of the threat of bankruptcy of Polish enterprises—Empirical study. In *Time for Money*, vol. II. Edited by Dariusz Zarzecki. Szczecin: Wydawnictwo Uniwersytetu Szczecińskiego.
- Stępień, P., Strąg, T., 2004. Multidimensional logit models for assessing the risk of bankruptcy of Polish enterprises. In *Time for Money*, vol. I. Edited by Dariusz Zarzecki. Szczecin: Wydawnictwo Uniwersytetu Szczecińskiego.
- Thomas, L., Edelman, D., Crook, J., *Credit. (2002), Scoring and Its Applications, Mathematical Modeling and Computation*, SIAM.
- Wędzki, D., 2000. The problem of using the ratio analysis to predict the bankruptcy of Polish enterprises—Case study. *Bank i Kredyt* 5: 54-61
- Wierzba, D., 2000. Wczesne wykrywanie przedsiębiorstw zagrożonych upadłością na podstawie wskaźników finansowych – teoria i badania empiryczne, in: *Zeszyty Naukowe no. 9*. Warsaw: Wydawnictwo Wyższej Szkoły Ekonomiczno-Informacyjnej w Warszawie.
- Zięba, M., Tomczak, S.K., Tomczak, J.M., 2016. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems With Applications* 58, pp. 93-101.

5. Patent Information Metadata in Evaluating the Effects of Innovation Support for the Business Sector

Rafał Wiśła

*Institute of Economics, Finance and Management
Jagiellonian University in Krakow*

Małgorzata Niklewicz-Pijaczyńska

*Institute of Economic Sciences
University of Wrocław*

Michał Gołacki

*International Cooperation Department
Patent Office of the Republic of Poland*

Introduction

The development of official statistics in the area of science, technology, and innovation is a consequence of changes in the factors of economic growth and development. Innovations (product-related, in business processes, and social) have ultimately become important characteristics in explaining the reasons for different competitive positions and different levels of economic development. As a development factor, innovation has a complex genesis:

- it is the result of interaction between people with different qualifications, creativity levels, interests, and abilities,
- it requires initiation of interactive, dynamic cooperation structures in its formation and development,
- it is the result of crossing and merging of more knowledge areas than ever before, and of interaction of people with different needs,
- in economic terms, results of the innovation development process are a particularly sought-after good that are becoming the subject of market transactions,

- universities, laboratories, and companies are diversifying their strategies and action programmes towards innovation over time.

The shift from the economics of capital, labour, and allocation to the economics of innovation and intangible resources has necessitated the development of new methodological approaches to data collection and quantification of phenomena and processes related to innovative activity, as well as evaluation of its results.

In the context of the growing importance of multidisciplinary research, links between science and industry, commercialization of the results of product and business processes development, and the increasing scale of public intervention aiming to stimulate innovative attitudes and activities, traditional data collection and development of indicators do not come close to providing quick answers to questions about the impact of innovation processes on economic and social development at the macro-scale. They also do not help in in-depth evaluation of the effects of innovation support for the business sector – at the meso-scale. Meanwhile, the last 200 years have been a period of unprecedented economic growth dynamics in the conditions of rapid technological changes. The analysis of many innovations (including implemented inventions) shows that they have clearly increased microeconomic business efficiency. It seems reasonable that technical progress and technological change are strongly related to economic growth. But what exactly is the nature of this relationship at the meso- and macro-scale?

The economics of innovation is an area of research that remains open. It includes research into the causes of innovation, the mechanisms that drive it, and the factors that inhibit it. Although there is already a well-developed system for measuring the results of innovation processes, the question of how to use the data obtained remains a big challenge (Kotowicz-Jawor, 2016). While the methodological discussion on the scope and ways of using patent statistics in research is not extensive compared to methodological discussions in the areas of innovation implementation and bibliometrics, a catalogue of several fundamental rules or areas useful for designing evaluation of the effects of innovative support for the business sector emerges.

This chapter aims to identify the potential of using patent data for evaluating the effects of innovation support policy for the business sector. It discusses the genesis of the development of research in the area of science, technology, and innovation and presents selected

indicators for the measurement and diagnosis of the possibilities of and barriers to using patent statistics in studies on the effects of innovation support policy.

In this publication, the use of **patent data** is understood as the use of data collected in public registers for evaluation and monitoring of innovation policy. One should bear in mind the important differences between the concept of *invention* as the subject of patent statistics research and the concept of *innovation* defined in the methodology of the Oslo Manual (2018). In accordance with the Oslo Manual guidelines, innovation is the implementation of a new or significantly improved product, process or method (technical, organisational, or marketing-related), which is new or significantly improved at least from the point of view of the enterprise that implements it. The material scope thus outlined is therefore relatively broad. Patent statistics, on the other hand, focus on empirical data concerning only inventions, i.e., new solutions of technical nature which represent a specific inventive step and can be applied in any field of industry.

5.1. Development of statistics on science, technology, and innovation

Technological change in the context of innovation has been the subject of intensive empirical research and stimulus for data collection in this area since the mid-1950s (Gomułka, 1998). A wide range of conceptual works was undertaken in the 1980s and 1990s to develop models and analytical approaches to studying technological change. But back in the 1980s, Pavitt (1984) or Pavitt, Robson, Townsend (1989) relied predominantly on their own survey forms and expert opinion in their research procedures. The characteristic features of research approaches at that time were as follows:

- focusing on individual (single) cases,
- relatively short time frame of analysis,
- high variability of the sets of items covered by the survey,
- scope of research – usually covering individual sectors of the economy.

As a consequence, it was difficult to achieve the effect of credibility and comparability over time and across countries.

The need to measure research and technological activity was recognised by the OECD as early as the 1960s. The first methodological guide (dedicated to official statistics services) was the 1963 *Frascati Manual* on the measurement of research and experimental development. In 2015, its 7th edition was published under the full title *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development* (OECD, 2015; GUS, 2018). The manual “is the recognised global standard for collecting and publishing internationally comparable statistics on financial and human resources devoted to research and development activities” (OECD, 2015, p. 20).

Currently, public statistics conduct studies of business innovation activity based on a standard, international methodology described in the *Oslo Manual* developed by the OECD and Eurostat (2018). The results of these studies are used in innovation analyses conducted by academic units, governments, and local authorities.

Evolution of methodological approaches to the measurement of the effects in science, technology, and innovation ran in parallel with the emergence of new development paradigms and the ensuing changes in scientific, industrial, and innovation policies. Increasing involvement of the state in these areas forced the necessity of evaluating the efficiency and effectiveness of these processes.

Table 5.1 presents, in evolutionary terms, the development of methodological tools for measuring the effects of scientific, technological, and innovative activities (S+T+I).

Table 5.1. Development of S+T+I indicators

Years	Groups of indicators
Third decade of the 21st century	the existing S+T+I measurement tools have been enriched with: measurement of innovation development in the following sectors: government, non-profit institutions serving households, households; measurement of innovation taking into account internal and external factors influencing innovativeness of enterprises; integrated subject matter approach of innovation development studies (linking statistical studies with other sources, such as administrative data, and complementary methods for obtaining data on the key innovation of an enterprise);
First and second decades of the 21st century	the existing S+T+I measurement tools have been enriched with: tax incentives; biotechnology and nanotechnology statistics; research commercialisation statistics; internationalisation and globalisation statistics; knowledge economy statistics; intellectual capital indicators;
1990s	the existing S+T+I measurement tools have been enriched with: surveys of innovation activities in the manufacturing sector; review of manufacturing technologies; innovations described in the technical literature; budget support for innovation activities; investment in intangible assets; ICT indicators; productivity; venture capital;
1980s	the existing S+T+I measurement tools have been enriched with: high technology products; bibliometrics; human resources statistics; surveys of innovation activities;
1970s	the existing S+T+I measurement tools have been enriched with: patent statistics; balance of payments in the field of technology;
1950s -1960s	expenditure on R&D activities;

Sources: Archibugi D., Sirilli G. (2001), *The Direct Measurement of Technological Innovations in Business*, [in:] *Innovation and enterprise creation: Statistics and indicators*, European Commission, Luxembourg, 2001, quoted after: Michał Górczyński (2005), *Przegląd wskaźników monitorowania systemów wspierania innowacyjności w krajach UE i wybranych krajach pozaeuropejskich – wnioski i rekomendacje dla Polski*, Warsaw; Kozłowski J., (2009), *Statystyka nauki, techniki i innowacji w krajach UE i OECD. Status and problems of development*, Ministry of Science and Higher Education, Warsaw; OECD/Eurostat (2018), *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation*, 4th Edition.

In general, it should be emphasized that the changing ways of practicing science, and the nature of R&D and innovation activities do not facilitate full understanding of the nature of the innovation phenomenon and innovation process. The data and indicators used so far enable understanding of only certain elements of the innovation process, in particular those concerning the **input**. A serious limitation is the lack of information on the **output** and **impact**.

A method commonly used in the European Union to collect data on innovation and innovation activities is to survey companies on various aspects of the innovation process in industry and services, and the resources allocated to innovation activities. The applied

entity-based approach makes it possible to analyse innovation in connection with other economic variables (such as production volume, value added, employment), both at the level of enterprise and industry (this data collection perspective is recommended and described in the *Oslo Manual*)⁴³. Currently, the main innovation indicators are as follows:

- share of enterprises in the studied group (enterprises from the sectoral/regional perspective) which introduced a product innovation and/or business process innovation in a specific period of time,
- share of enterprises in the studied group (enterprises from the sectoral/regional perspective) which conduct R&D works in a systematic way,
- impact of product innovations on sales volume/dynamics/profitability,
- impact of process innovations on changes in costs of the manufacturing process and the size/structure of employment,
- impact of innovations (product-related, process-related) on business efficiency,
- intensity of cooperation in the field of R&D activity, and intensity of implementation of its results,
- share of individuals employed in R&D in the total number of economically active/employed population,
- expenditure on innovation activity in enterprises per person/in relation to GDP,
- degree of commercialisation (marketisation) of new solutions (implemented from 2013).

5.2. Patent information in research on the effects of innovation support for the business sector

Despite the variety of indicators used in this area, studies on innovation conducted in the current formula do not provide answers to the question how public innovation support for the business sector translates into its creativity, innovativeness, and business effectiveness.

⁴³ The examples of national data collection schemes within the European research programme entitled *Community Innovation Survey* are: Report on research and development activities (PNT-01); Report on research and development activities (R&D) and resources allocated to R&D in government units (PNT-01/a); Report on innovation in industry (PNT-02); Report on nanotechnology research and development activities in research units (PNT-05); Report on nanotechnology activities in enterprises (PNT-06).

This is mainly due to the fact that the most comprehensive and complete studies in the field of innovation are carried out primarily by Statistics Poland (GUS)⁴⁴ and, therefore, have certain limitations inherent in the methodology applied. These include, among others, the issue of the random sample of enterprises covered by innovation studies, and statistical secrecy of the results of these studies, which limits the possibility of an extended analysis of individual data, which would evaluate the results of the innovation development process.

This problem can be illustrated using the example of the study of changes in innovative activity of pharmaceutical enterprises. In 2019, 131 business entities (employing more than 9 people) conducted this kind of activity, with 39 largest producing over 90% of the value of the industry's sold production (GUS, 2020, p. 37). An even more difficult situation is encountered when trying to analyse the innovative activity of enterprises operating in medical device manufacturing (PKD codes: 26.60 and 32.50, i.e., third level of business activity classification). Due to the fact that "No data obtained in statistical studies may be published or made available (...) characterising the economic and financial results of national economy entities carrying out economic activity if a given aggregation consists of less than three entities or the share of one entity in a given aggregation is greater than three quarters of the total" (Article 38 of the Act of 29 June 1995 on Public Statistics), usefulness of this type of microdata is very limited. Moreover, in this situation, even spatial analysis of data – concerning innovative activity of enterprises – at the level of municipalities, counties, city counties, and particularly cities, is practically impossible.

In this context, patent information is of key importance for research in the field of innovation, as it offers considerable opportunities for exploration. It must be admitted that, compared to other methodological approaches, it reaches narrowly, but most deeply, a specific section of the innovation development process: **open and free access** (public registers), **high flexibility of aggregation and disaggregation of the studied processes** (high scalability), and **diversification of research areas** for which it is applied.

An important event in the context of discovering the potential of patent data was the Inaugural Conference⁴⁵ organised by the OECD in 1996, devoted to the creation of new indicators for innovation, science, and technology policies in the twenty-first century. In

⁴⁴ Cyclical studies "Innovative activities of enterprises in ..." (2001-2019, next in the series "Statistical analyses" of Statistics Poland).

⁴⁵ The OECD conference of November 1996, with Poland joining the conference for the first time.

the summary of the meeting, it was admitted that the indicators have a very wide range of applications. Patent data make it possible to study the mobility of human resources, track the flow of not only tangible knowledge but also the so-called tacit knowledge⁴⁶, assess the economic value of new solutions, analyse the network of scientific connections, measure the internationalisation of scientific and technical activity, and create guidelines for scientific and technical policy, e.g., in the area of research, on the impact of tax reliefs on the intensity of R&D work. Discussions on the creation of new and verification of the existing research methods used in these areas coincided with a change in the approach to the innovation process, noticeable in politics and economics. It consisted in gradual modification of the hitherto prevailing linear model, where innovations are mainly created through internally concentrated research and development works (referred to in the first *Frascati Manual*), towards a systemic theory of innovation, according to which a key role in its creation is played by diversified, internal and external knowledge flows (as illustrated in the first *Oslo Manual*). The findings of the conference confirmed the belief, repeatedly expressed earlier by, among others, Z. Griliches (1979), J. Schmookler (1966), that patent data are a valuable source of knowledge not only on the stimuli, course, and structure of innovation processes, but also their connection with economic growth and development.⁴⁷

Currently, information included in patent dossiers can be divided into five main pillars:

- information on the field of technology to which a given solution belongs, established on the basis of classification carried out with the use of various systematics;
- information enabling determination of the potential value of a new solution, established on the basis of, among others, the number of citations, the number of licences granted, the rhythm of change of the patent holder resulting from market transactions, the number of years patent monopoly has been in force, the geographical scope of protection;
- information enabling identification of the actors involved in the process of creating new technologies (by analysing, inter alia, the structure of the team of authors of an

⁴⁶ Much of human knowledge is tacit knowledge. This is clearly visible, for instance, in the case of operational skills or know-how acquired through experience. This type of knowledge is action-oriented. It has an individual dimension. It is difficult to formalize. Unlike explicit knowledge, tacit knowledge transfer requires a high level of interaction between actors in the sharing of knowledge.

⁴⁷ OECD (2017). *What Indicators for Science, Technology and Innovation Policies in the 21st Century?* (<https://www.oecd.org/science/inno/37082579.pdf>, accessed: 15 September 2021).

invention and their affiliations, the structure and type of applicants, the progress of the development of patent families, including the “triadic family”⁴⁸);

- information enabling reconstruction of the history of the application, using the dates available in the documentation, such as: application date, publication date, refusal or revocation date, patent grant date, monopoly termination date;
- information enabling determination of the invention’s legal status (patent in force, patent revoked).

The documentation which comprises patent literature requires a standardised bibliographic description of its contents (formal and pragmatic considerations). The data included in patent literature are collected according to the objectives and typology characteristic for a given classification. The standard recommended by the World Intellectual Property Organization for bibliographic description of a patent document is the **ST.9 Standard**⁴⁹. The main elements of the standard are: data identifying the patent document, data on the patent application, technical information on the invention, references to patent and scientific literature, identification of parties related to the patent. According to the WIPO ST. 9 Standard, the numerical identification codes that precede the information on inventions stand for:

- (21) – application number;
- (22) – application date;
- (23) – data relating to exhibition priority (date and reference of exhibition);
- (30) – data relating to convention priority (priority application number, date of priority application, country code);
- (41) – date of making available to the public;
- (51) – International Patent Classification symbol;
- (54) – title of invention;
- (61) – number of parent patent;

⁴⁸ “Triadic patent family” refers to an invention that has been applied for simultaneously in the modes: (1) the EPO (European Patent Office, an executive body of the European Patent Organisation tasked with granting European patents), (2) the JPO (Japan Patent Office, a Japanese governmental body responsible for granting and maintaining industrial property rights; one of the three largest patent offices in the world, (3) the USPTO (United States Patent and Trademark Office, a patent office under the US Department of Commerce, granting patents and registering trademarks), and obtained a patent in the group of the most economically developed countries in the world. The underlying assumption is that technical and economic importance of this industrial property right is above average.

⁴⁹ WIPO (2013). STANDARD ST.9. Recommendation concerning bibliographic data on and relating to patents and SPCP. (<http://www.wipo.int/export/sites/www/standards/en/pdf/03-09-01.pdf>).

- (62) – number of parent patent application (under which the given application was published);
- (68) – number of the basic patent for supplementary protection right;
- (71) – surname and forename, or name of the applicant; additionally, place of residence and country of the applicant (country code) are stated;
- (72) – surname and forename of the author, additionally, place of residence and country of the author (country code) are stated;
- (73) – surname and forename, or name of the right holder, additionally, place of residence or registered office and country of the right holder are stated (country code);
- (76) – surname and forename of the author who is also the right holder; additionally, place of residence and country of the right holder are stated (country code);
- (86) – date and number of international application;
- (87) – date and publication number of international application;
- (93) – number and date of the first authorisation to place the product covered by supplementary protection right on the market, and identification of the authority which granted the authorisation;
- (94) – date from which supplementary protection right arises, and the right's duration;
- (95) – designation of the product described in the authorisation for product covered by supplementary protection right;
- (97) – date and publication number of European patent;
- (98) – date of application for renewal of supplementary protection right.

The main objective of standardization of patent description attributes, based on WIPO ST.9 Standard, is to increase the efficiency of patent literature searches. Arranging data in a standardized way enables data aggregation at any level, making it possible to identify interdependencies, regularities, and trends, predict the direction of technological processes, development and diffusion of specialist engineering knowledge, and structural changes in the economy. At the same time, intensive technological development is taking place in the field of IT infrastructure of data repositories. An example of measures aiming to improve access to patent data is the new service: e-Search Engine (e-Wyszukiwarka) available on the website of the Patent Office of the Republic of Poland. The e-Search Engine service enables searches for public data about industrial property objects from a single search engine with an intuitive user interface. The modern search system meets the requirements set for such services and enables searching in the databases of the Polish Patent Office domain systems,

as well as in publications (*Bulletin of the Patent Office; Patent Office News*). The e-Search Engine also enables subscription for information concerning a defined search.

One of the essential elements of the patent description is information on the entities involved in the process of creating a new solution and obtaining patent protection. An example of a patent bibliographic description record is presented in Figure 5.1.

Figure 5.1. Standard bibliographic description record⁵⁰:

PATENTS	
RIGHTS GRANTED (from n. 238 811 to n. 238870)	
W) Part of the submission will be considered as a submission of a trade dress	
(B1) (11) 238819 (41) 2019 04 23 (51) A01G 25/06 (2006.01) A01G 25/06 (2006.01) (21) 423220 (22) 2017 10 20 (72) BURSZTA-ADAMIAK EWA, Strzelin (PL); LICZNAR PAWEŁ, Wrocław (PL) (73) MIEJSKIE WODOCIĄGI I KANALIZACJA W BYDGOSZCZY SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ, Bydgoszcz (PL) (54) Układ solarny wspomagający regenerację zbiorników infiltracyjnych	(B1) (11) 238817 (41) 2018 10 08 (51) A24C 5/34 (2006.01) A24C 5/345 (2006.01) B65B 19/30 (2006.01) (21) 421217 (22) 2017 04 07 (72) OWCZAREK RADOSŁAW, Radom (PL) (73) INTERNATIONAL TABACCO MACHINERY POLAND SPÓŁKA Z OGRANICZONĄ ODPOWIEDZIALNOŚCIĄ, Radom (PL) (54) Urządzenie przemysłu tytoniowego do odrzutu wadliwych artykułów prętopodobnych z przepływu masowego artykułów prętopodobnych przemysłu tytoniowego oraz system określania jakości w przemyśle tytoniowym
	(B1) (11) 238818 (41) 2018 12 03 (51) A24D 3/02 (2006.01) (21) 421723 (22) 2017 05 29

Source: Patent Office News, 28/2021, p. 4.

Attributes of the patent bibliographic description (e.g., 51, 72, 73) can be combined and enable multithreaded analysis in combination with a range of other data (PKD classification, ranking lists of projects qualified for financing by the National Centre for Research and Development – NCBiR, Polish Agency for Enterprise Development – PARP).

For example, data relating to the author of a technical solution (attribute 72) and other persons or entities designated in patent documentation (attributes 73 and/or 76) can be analysed along the following dimensions:

- gender and region of origin of the national natural persons stated in the patent documentation as the author or applicant,

⁵⁰ For example, see: Polish Patent Office (2021). Patent Office News, 28/2021, pp. 4-5. (https://uprp.gov.pl/sites/default/files/wup/2021/10_27-30/28//wup28_2021.pdf)

- legal status and other features that define the legal and economic situation⁵¹ of the entity and its role in the process of obtaining patent protection, e.g., natural person, legal entity, joint-stock company, limited liability company, enterprise, university, author, applicant,
- structure of the team of authors and patent applicants/rights holders in terms of patent applications and their success in obtaining patent protection; patent information makes it possible to analyse joint applications made by entities representing different sectors, e.g., scientific and economic sectors,
- structure of the team of authors and patent applicants/rights holders in terms of the intensity of quotability of the solution in other patent descriptions; patent information makes it possible to determine whether technological solutions created by research teams consisting of representatives of different fields of science will have wider application and be quoted more frequently,
- structure of the team of authors and patent applicants/rights holders in connection with the technological field where the new solution is located,
- structure and size of the team of authors and patent applicants/rights holders in relation to the geographical scope of protection; analysis of the characteristics of technological solutions developed by international teams or in cooperation between entities from different countries,
- structure of the team of authors and applicants/right holders from the perspective of economic use of the solution (e.g., by using information about granted licences for this solution); analysis of the impact of the size and internationalisation of research teams on the effectiveness of commercialisation of a given solution.

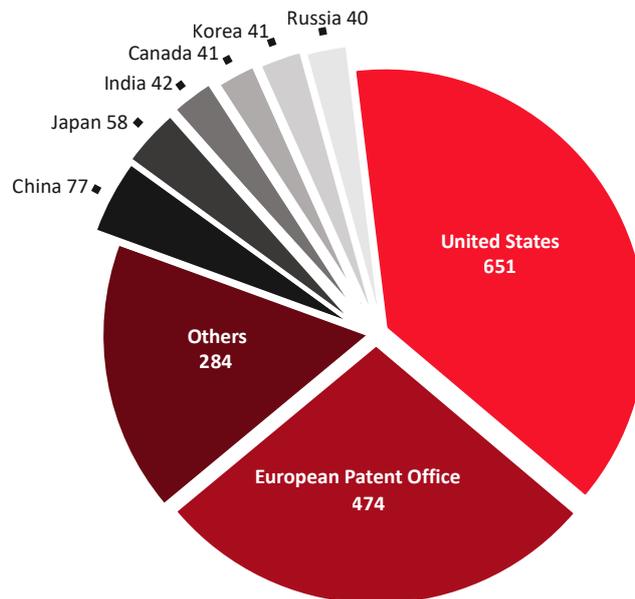
From the point of view of evaluating the effects of innovation support for the business sector, patent documentation makes it possible to analyse the following aspects:

A. Patent activity of Polish residents at home and abroad:

number of patent applications of Polish enterprises filed with the Polish Patent Office (PPO) in a specific technological field (covered by public support, e.g., energy technologies, material technologies, information technologies, telecommunications, mechatronics, etc.) vs. number of all applications from domestic entities filed with the PPO in a given period,

⁵¹ This information can be partially identified directly on the basis of patent documentation (e.g., based on entity's name), or using other available data sources, such as the REGON register.

Figure. 5.2. Number of invention patent applications filed in 2019 by Polish entities for protection in selected countries



Source: Own study based on data available from IP Statistics Data Centre <https://www3.wipo.int/ipstats/>

- 2) number of patent applications filed by Polish enterprises with the PPO in a given technological field (as mentioned above) taking into consideration: enterprise size, leading PKD code, regional location, and other connections⁵²,
- 3) number of patent applications by Polish enterprises filed abroad in a given technological field (covered by public support, e.g., energy technologies, material technologies, information technologies, telecommunications, mechatronics, etc.) vs. number of all Polish applications filed abroad in a given period.

Information expected when using the analytical approach, points: 1-3:

- aggregated, general link between multiannual public support programmes (in any technological field) and the performance of the business sector in protecting R&D results; in this respect, it is important to establish a time lag;
- aggregated evaluation of protection right's value; one of its indicators is the pursuit of protection not only domestically but also abroad (the number of countries in which the monopoly is to be valid matters) as proof of value (industrial applicability); international

⁵² Ibid.

patent protection from the point of view of Polish entities is still a considerable expense (and a challenge, in the broad sense);

- number of patent applications by Polish enterprises filed with the PPO, which obtained specific public support for innovation (from any governmental agency),
- number of patent applications by Polish enterprises filed abroad, which obtained specific public support for innovation (from any governmental agency).

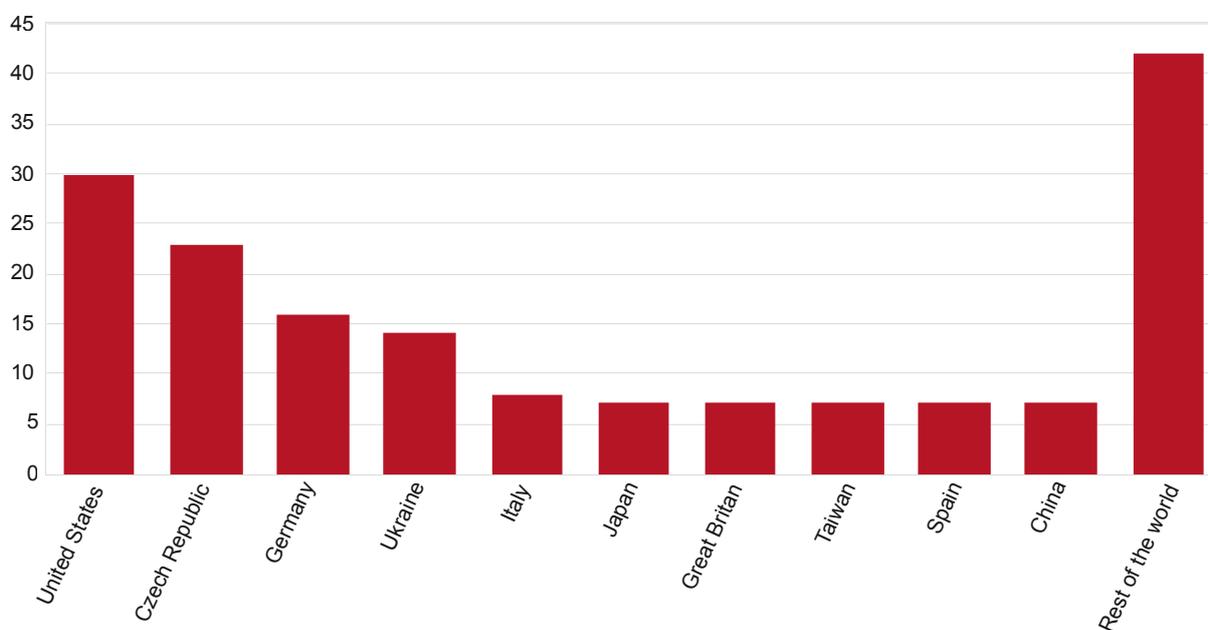
Information expected when using the analytical approach, points: 4-5:

- list of names of entities which receive financial support (e.g., from NCBiR, PARP) confronted with patent applications of these entities; both lists are public.

B. Non-residents' interest in the Polish market:

- 6) number of European patents validated in Poland in a specific technological field considered as a priority in Poland and covered by public support,
- 7) number of foreign applications filed domestically in a specific technological field considered as a priority in Poland and covered by public support,

Figure 5.3. Applications for inventions and utility models in Poland in 2019 by foreign entities according to their origin



Source: Annual Report 2019, Polish Patent Office, p. 11.

- 8) number of PCT⁵³ applications filed in Poland in a specific technological field considered to be a priority in Poland and covered by public support (e.g., technological field linked to a specific national or regional smart specialisation),
- 9) most active entities seeking protection in Poland in a specific technological field identified as a priority in Poland and covered by public support,

Information expected when using the analytical approach, points: 6-9:

- R&D intensity in the world in a given technological field and the importance of Poland as a location for protection,
- geographical and sectoral structure of entities seeking patent protection in Poland and the technological structure of the solutions they file.

C. Legal monopoly evaluation, patent value:

- 10) scope of the patent family (e.g., the number, importance of patent offices, geographical coverage of protection) in a specific technological field considered to be a priority in Poland and covered by public support,
- 11) number of citations in other patent descriptions (as manifestation of the patent's value, applicability of research results),
- 12) duration of patent protection (payment of fees) by sector and technological field as proof of invention's actual market value,
- 13) launching of an objection proceeding,
- 14) number and type of licences granted on the basis of the exclusive right in question (as manifestation of the deployment effect and technological accumulation which ultimately materialises itself in industrial terms, contributing to changes in technical work equipment and technological progress).

⁵³ Procedure for patent applications filed by non-residents at the patent office of a country under the Patent Cooperation Treaty, abbreviated as PCT. The Patent Cooperation Treaty, abbreviated as PCT, is an agreement signed on 19 June 1970 in Washington (effective as of 1 June 1978) establishing international patent applications with the same effects as national applications in all the states-parties to the treaty. International applications are directed to the World Intellectual Property Organization (WIPO) and then examined by the European Patent Office or the respective national patent office (acting as the so-called International Searching Authority, ISA). The intermediary in the filing of resident patent applications under the PCT procedure is the Patent Office of the Republic of Poland. The PCT application means that the applicant seeks protection of the invention in a number of countries (at least three, and maximum all the signatory states, of which there are 142 at present). Patents for inventions filed under the PCT are, however, still granted by the particular national patent offices. Poland joined the Patent Cooperation Treaty in December 1990 (after: Statistics Poland: Concepts used in public statistics).

Information expected when using the analytical approach, points: 10-14:

- aggregated picture of changes in the value of technical knowledge accumulated by Polish companies in a given technological field,
- aggregated picture of changes in the value of a given technological solution owned by a given company,
- aggregated picture of changes in the value of the accumulated technical knowledge owned by residents and non-residents,
- aggregated picture of changes in the value of accumulated technical knowledge belonging to the particular sectors of national economy.

D. Evaluation of the diffusion effect of scientific and technical knowledge:

- 15) networks of collaboration and diffusion of knowledge involving actors from different sectors (including authors and/or applicants/right holders; continuity vs. incidence of collaboration),
- 16) degree of internationalisation of research teams, including determination of spatial mobility of scientific and industrial authors,
- 17) technology diffusion, flow of technical knowledge between industries (using licensing information in databases, patent holder changes).

Information expected when using the analytical approach, points: 15-17:

- dynamics and scope of dissemination of technical knowledge in different spatial, temporal, and sectoral settings.

E. Other obtainable information supporting the evaluation of the effects of innovative support for the business sector:

- 18) total number of all applications in a specific technological field in the world vs. Polish applications; broken down by country of origin of the application; taking account of the structure of entities and broken down by sector (e.g., science sector, economy sector, etc.),
- 19) number of cited patent claims from a specific technological field made in Poland, broken down by sector,
- 20) cooperation in creating technical solutions protected by patents between different sectors/within a given sector (most active in cooperation),
- 21) capital of origin of the entities examined, if the register of the PPO is integrated with other registers, or based on information available e.g., on official websites of these entities,

- 22) most active applicants from the industry sector and entities with foreign patent families,
- 23) most active applicants from the science sector and science sector entities with foreign patent families,
- 24) most active inventors (industrial creators),
- 25) territoriality – in different variants (e.g., according to the origin of the applicant, the origin of the inventor, up to municipality level),
- 26) number of applications per technological field,
- 27) average annual growth rate of patent applications for each technological field (for residents and non-residents; most active applicants in each technological field),
- 28) RCA Index (*Revealed Comparative Advantage Index*)⁵⁴ in various configurations (e.g., technology field, country, region).

5.3. Obstacles in the process of obtaining and using patent data

Patent information is accessible in various modes. The main international patent databases maintained and published by the particular international organisations include:

- **Espacenet** – the international patent document database maintained by the European Patent Office (<https://worldwide.espacenet.com>),
- **PATENTSCOPE** – the international patent document database maintained by the World Intellectual Property Organisation (<https://patentscope.wipo.int>),
- **DEPATISnet** – database and information service of the German Patent Office (<https://www.dpma.de>),
- **USPTO** – full-text database of US patent applications and grants (<https://www.uspto.gov>).

The above patent databases are characterised by free access and the possibility of defining a set of metadata. Importantly, however, they are essentially designed for the processes of examining novelty, inventiveness level, etc., conducted by patent attorneys and other persons interested in a particular technical solution or particular field of technology.

⁵⁴ See: Nesta, L, & Patel, P., (2005). National Patterns of Technology Accumulation: Use of Patent Statistics, [in:] Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent, red. H. F. Moed, W. Glänzel, U. Schmoch, KLUWER Academic Publishers New York, Boston, Dordrecht, London, Moscow, pp. 536-539.

On the other hand, **Clarivate** (Thomson Innovation) – a commercial database enabling exploration of large, structured sets of patent applications and grants – has many additional, paid functionalities helpful in comprehensive, structured exploration of all records and attributes of a set.

Other, frequently subject-specific databases include: Cippix® (chemistry); GenomeQuest (biology); LexisNexis; Delphion (integrates the following databases: European Patent Office, United States Patent and Trademark Office, World Intellectual Property Organization); Japan Patent Data Service.

Despite the existence of such diversified sources, access to a complete set of metadata can turn out problematic in practice. Some of the barriers to accessing complete data sets identified by the authors are as follows:

- national patent offices provide open access functionalities and tools for extracting metadata automatically and in bulk, but to a very limited extent (e.g., only limited numbers of records can be exported from publicly available search engines for further analysis),
- public digital repositories of collected patent documentation continue to have rather simple architecture, and only limited functionalities are available to users, although the situation in this respect is systematically improving, as exemplified by the consistently updated and developed services, e.g., by the European Patent Office (e.g., Espacenet), WIPO (e.g., IP Statistics Data Centre), and the Office of the Patent and Trademark Office (e-Search Engine, Register Search Engine),
- potential results depend on the search method used (e.g., patent number, right holder, etc.), as a result, the obtained information may be misleading, e.g., by generating a message that a given document is unavailable in a situation where a search according to a changed key will display the searched technical documentation – this is the so-called artifacts problem,
- it happens that researchers identify patents in the 1:1 ratio (one invention assigned to one protection right) omitting the issue of the so-called patent portfolio, which results in misinterpretation of patent statistics (Niedbalska, 2009, p. 10),
- using imprecise descriptions of the ownership structure in the case of multinational enterprises significantly hinders the assessment of the level of globalisation and internationalisation of inventive activity (Niedbalska, 2009, p. 11),

- the statutory 18-month lag between the publication of patent applications in the Bulletin of the Patent Office of the Republic of Poland and the filing date makes it difficult to assess inventive activity of the analysed entities, in both qualitative and quantitative studies,
- it is problematic to determine the correct date for the prepared statistical data – which not only identifies different legal status of an invention, but also distorts the assessment of the current state of the art,
- patent reporting by national and regional patent organisations to statistical offices is generic and their continued visibility in official statistics does not allow in-depth research,
- despite the increasing harmonisation of regulations, the existing procedural discrepancies do not enable reliable comparative analyses (e.g., inventive activity of American and Japanese inventors); legal factors shaping the country-specific safeguard procedure result in a privileged situation of domestic enterprises, which are better-acquainted with the specificity of domestic procedural solutions than foreign entities – this phenomenon is referred to as the *home advantage* (Niedbalska, 2009, p. 11),
- in the case of inventions developed in international teams, there is a risk of multiple counting of the same solution – a way to avoid this problem is to use the *fractional count* method, but it is not always applied consistently (Niedbalska, 2009, p. 12).

The identified technical and organisational limitations are being addressed by commercial patent information distributors. Their patent information acquisition is functional, but also costly for the end user. Moreover, commercial patent data distributors often do not have patent information collections from countries characterised by low values of global or European innovation indices. An attractive alternative to these products is the PATSTAT database administered by the European Patent Office. The database provides access to over 100 million patent documents from all over the world. Its specificity lies in the fact that it provides the possibility of retrieving, using SQL queries, source data in the form of extracts containing up to 700 thousand records at a time. In addition, the source data are enriched with information such as the number of citations, size of the patent family, and reference of patent classification to economic activity codes (according to NACE Rev 2). The database administrators are also constantly working on the quality of the data by, among other things, harmonising the names, terminology, and concepts, enabling uniform validation of technical documentation and improving the efficiency of searches for information archived in the system.

At the same time, it should be emphasized that the above-mentioned imperfections of patent data use do not question their general usefulness at practically all stages of the innovation process. At the stage of initial research, they provide information on the current state of the art, profitability of efforts related to the development of a new solution, possibility of cooperation with external entities, and the current inventive activity of competitors, which makes it necessary to e.g., obtain a licence. On the other hand, at the final stage of the process, they make it possible to select the optimum strategy for securing exclusive rights and estimate the costs related to possible commercialisation of the registered solution. Moreover, the problems of data publication delays referred to above are relatively minor compared to delays occurring in other publicly available repositories, such as GUS and EUROSTAT databases, where data are published only in aggregated form and with an additional time lag related to their necessary processing.

Conclusion

A patent application is a significant economic event, one of the many steps in the process of innovation development. Frequently, it is the culmination of research and development work, most important economically, as it is linked to the creation of inventions, i.e., potential technical innovations. The obtained protection right is a potential resource for the commercial activity of the organisation, which can evolve into a production factor. That makes patent information a specific bridge between the results of the enterprise's R&D and implementation activity. The collection of items, their further embedding in a given context (i.e., giving them meaning and content) is a key stage in the cognitive process. In the scientific layer, the way in which a data collection is organised, processed, inferred, and presented determines the approach or distance from the truth. In the application layer, it allows to manage the area of uncertainty related to the decision-making process. It is intended to support resource allocation, monitor performance levels, identify and recommend directions for societal development, as well as forecasting. In this context, patent information metadata represent information potential which is still untapped. The main reason for this is the imperfect methods of patent statistics use, which are far from the holistic approach. In Poland, the present state of research works, especially methodological, on the evaluation of usefulness and usability of patent descriptions and statistics, can be described as initial, while globally it is much more advanced. However, despite the reservations signalled in the study, it should be emphasized that the strength of patent

data in their structured form enables comparative analyses of international nature. The use of a detailed classification enables, among other things, precise identification of the most sensitive areas of technology, observation of global trends of technological development, estimation of the value of patents, improvement of the effectiveness of the activities of engineers and researchers, and establishment of cooperation networks between the particular participants of the innovation process. Moreover, this type of data, after thorough substantive verification, is made available to all the interested parties free of charge, under equal conditions, which is a consequence of the obligation to make technical documentation on the requested protection generally available. It should be stressed, however, that broadly understood comprehensive analysis of the effects of innovation development processes, apart from the use of patent information, requires the inclusion of other compatible data repositories.

References:

- Archibugi, D., Sirilli, G. (2001), The Direct Measurement of Technological Innovations in Business, [in:] *Innovation and enterprise creation: Statistics and indicators*, Luxemburg: European Commission.
- Archambault, E. (2002), Methods for using patents in cross-country comparisons, *Scientometrics* Vol. 54, pp. 15-30.
- Godin, B. (2002), Measuring Output: When Economics Drives Science and Technology Measurements, Project on the History and Sociology of S&T Statistics Working Paper No. 14.
- Gomułka, P. (1998), *Teoria innowacji i wzrostu gospodarczego*, Warsaw: CASE.
- Griliches, Z. (1990), Patent Statistics as Economic Indicators: A Survey, *Journal of Economic Literature*, Vol. 28, pp. 1661-1707.
- Griliches Z. (1979), Issues in Assessing the Contribution of Research and Development to Productivity Growth, *Bell Journal of Economics*, The RAND Corporation, vol. 10 (1).
- GUS (2020), *Rocznik Statystyczny Przemysłu*. Statistics Poland, Warsaw.
- Kotowicz-Jawor J., (ed.). (2016), *Innowacyjność polskiej gospodarki w przejściowej fazie rozwoju*, Instytut Nauk Ekonomicznych Polskiej Akademii Nauk, Warsaw.
- Kozłowski J., (2009), *Statystyka nauki, techniki i innowacji w krajach UE i OECD. Stan i problemy rozwoju*, Warsaw: Ministry of Science and Higher Education.
- Kuznets P., Inventive Activity : Problems of Definition and Measurement Problems of Definition, [in:] *The Rate and Direction of Inventive Activity: Economic and Social Factors*, 1962, pp. 19-52.
- Larivière, V., Ni, Ch., Gingras, Y., Cronin, B., Sugimoto C.R. (2013), Bibliometrics: Global gender disparities in science, *Nature*, Vol. 504, pp. 211-213.

- Nesta, L. & Patel, P., (2005), National Patterns of Technology Accumulation: Use of Patent Statistics, [in:] *Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent*, red. H. F. Moed, W. Glänzel, U. Schmoch, KLUWER Academic Publishers New York, Boston, Dordrecht, London, Moscow, pp. 536-539.
- Niedbalska G., (2009), Podręcznik statystyki patentów. OECD Patent Statistics Manual, MSN Working Papers nr 5, Warsaw.
- Niklewicz-Pijaczyńska M., (2019), System patentowy w gospodarowaniu wiedzą. Ekonomia wiedzy technicznej skodyfikowanej, Prawnicza i Ekonomiczna Biblioteka Cyfrowa, Wrocław.
- OECD (2001), *Using patent counts for cross-country comparisons of technology output*.
- OECD (2009), *Patent Statistics Manual*.
- OECD (2015). Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development.
- OECD (2015). Podręcznik Frascati 2015: Zalecenia dotyczące pozyskiwania i prezentowania danych z zakresu działalności badawczej i rozwojowej, GUS 2018.
- OECD/Eurostat (2018), Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris/Eurostat, Luxembourg, <https://doi.org/10.1787/9789264304604-en>
- Pavitt, K. (1984), Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory, *Research Policy* No. 13 (6), pp. 343-373.
- Pavitt, K., Robson, M., Townsend, J. (1989), Technological Accumulation, Diversification and Organisation in UK Companies, 1945-1983, *Management Science*, Vol. 35, No.1, pp. 81-99.
- Schmookler J. (1966), *Invention and Economic Growth*, Harvard University Press, Cambridge.
- Stephan P. E., Gurmu P., Sumell A. J., Black G. (2005), Who's Patenting in the University?, Evidence from the Survey of Doctorate Recipients, Evidence from a survey of doctoral recipients Forthcoming *Economics of Innovation and New Technology*.
- Verbeek A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., Deleus, F. (2001), *Linking Science to Technology: Using Bibliographic References in Patents to Build Linkage Schemes*, Katholieke Universiteit Leuven.
- Wiśła, R., (2014), Regionalne wzorce akumulacji wiedzy technicznej w krajach Europy Środkowo-Wschodniej. Warsaw: Wydawnictwo Naukowe PWN.

6. Past experience and future prospects of Statistics Poland in the use of administrative data in enterprise surveys

Jacek Kotowski

Statistics Poland

Dominika Nowak

Statistics Poland,

Collegium of World Economy

SGH Warsaw School of Economics

Paulina Skórska

Idea Instytut Sp. z o. o.,

Centre for Evaluation and Analysis of Public Policies

Jagiellonian University

Introduction

The objective of the broadly understood development policy is to implement legislative amendments, programmes, and projects leading to planned changes in its target areas. Therefore, an indispensable element of the process of programming and implementing development policies should be to measure the size of change brought by the intervention undertaken and compare the results obtained against the adopted assumptions and defined objectives. The complexity and different generality levels of the particular policies' objectives are the reason why, in most cases, it is impossible to directly record the level of their accomplishment. Therefore, it is necessary to capture them indirectly, by measuring the features or variables that operationalise the objectives in the best possible and most accurate way. For this, indicators are useful, which, thanks to their measurability, allow institutions supervising the programmes to establish whether a given feature, or element, of the intervention's objective is present, and measure it (Górniak and Keler, 2007). However, indicators are frequently problematic, because we often forget that abstract, unobservable

features, such as competitiveness of the economy or innovativeness of enterprises, are not indicators. Appropriate matching of indicators with the measured socio-economic phenomena (including the level of achievement of intervention's objectives) is not always a simple task, particularly because most of the unobservable features, such as the above-mentioned competitiveness, can adopt many definitions (see e.g. Gorynia, 2000; Grzebyk and Kryński, 2011; Stankiewicz, 2005; Szymanik, 2016).

At the same time, the challenge involved in using indicators for policy objectives' operationalisation is supplying indicators with data. The data should meet a number of criteria, for instance in terms of data source stability, quality, comparability, time range, etc. Besides, their collection and processing should be cost-effective. In this context, the possibility of multiple use of public data i.e., data collected by official statistics and in administrative systems is becoming increasingly important in the process of programming and evaluating public policies. At the same time, institutions supervising the implementation of public interventions must assess the usefulness of the above-mentioned data in the processes of monitoring and evaluation of measures, with particular emphasis on differences in the methodologies and methods of data collection used by the particular units of public administration.

In the second half of 2021, under the guidance of the National Evaluation Unit (KJE) in the Ministry of Development Funds and Regional Policy, work was started on the guidelines for the evaluation of cohesion policy in the years 2021-2027. Although the national document is not ready yet⁵⁵, the guidelines of the European Commission make it possible to outline the expected directions and recommended approaches for the process of cohesion policy evaluation in Poland past 2020. In July 2021, the Commission published a working document setting out the approach to the implementation, monitoring, and evaluation of programmes financed by the European Regional Development Fund (ERDF), the Cohesion Fund (CF), and Just Transition Fund (JTF) for the years 2021-2027 (European Commission, 2021). In the context of lessons learned from the previous programming period (2014-2020), the role of using – in the current period – of the data of EUROSTAT, national statistical offices, and administrative systems is highlighted:

⁵⁵ As of 15 November 2021,

“EUROSTAT and national statistical indicators (...) could also provide valuable insights [in the monitoring and evaluation of the 2021-27 perspective] (...). Use administrative data wherever possible to assess the effects of interventions, so as to avoid the pitfalls of using survey or interview data from a sample of respondents that may not be representative or may be prone to recollection error. Particular attention should be given to public data registers storing time series data covering samples or total populations of entities” (European Commission, 2021, p. 4; 15)

In the 2021-2027 perspective, emphasis is to be placed on the analysis of result indicators, which are to approximate the change in the situation of public interventions’ beneficiaries, achieved as a result of programme implementation. They refer to the immediate effects of the measure, visible among direct recipients of support. This is a fundamental change in comparison to the previous perspective, which did not emphasise the use of this type of data in monitoring and evaluation. On the other hand, the new perspective will not include impact indicators, which were present in the 2014-2020 perspective and referred to measures’ impact on the economy and society, not just direct recipients of support. If good immediate effect indicators are available, a wide-ranging impact evaluation may not be necessary:

“Even if measures are strategic, likely to be replicated, etc., there may still be circumstances when an extensive impact evaluation may not necessarily take place. For example: the ‘outputs’ of the measure may be ‘administrative’ rather than substantive in which case monitoring using established direct result indicators should be sufficient.” (European Commission, 2021, p. 14).

With this approach, waiting for the applicable period of time from the end of intervention for the effects to be revealed becomes less important. Therefore, it is possible to more broadly include individual data collected by official statistics and administrative systems in the evaluation. In many cases, it will be possible to use the data in the subsequent reporting years, during and immediately after the end of intervention, which provides an opportunity to obtain a fuller picture of the changes in the indicators, related to the activities of the supported entities.

With regard to the objectives of interventions targeting the SME sector, also in the Innovation Survey, the Commission expects that at the level of national interventions

(grant and non-grant), selected result indicators in innovation and competitiveness will be compared with result indicators regarding job creation and private investment increase.

In this respect, statistical and administrative data will be useful for isolating the effects of interventions from the impact of other factors, which will make it possible to understand the functioning of the programme and carry out a reliable assessment of its effects. Therefore, they can be helpful in understanding how an intervention generates effects (explanation of the mechanism), to what extent these effects were intentional, and whether unintentional effects occurred (whether significant, positive, or negative). The latter may be of key importance in analysing whether support for enterprises disturbs market balance.

In the studies assessing the effects of the programmes, the Commission explicitly recommends that relying solely on survey data and interviews with intervention recipients be avoided. As one of the important criteria for the selection of indicators, the Commission mentions the possibility of reconstructing the way of measuring the indicator, and its replication in the future. Public statistics data and administrative data, subject to procedural rigours with regard to organising data collection and analysis, seem to be well-matched with the above assumptions. These data may also be relevant in the context of the growing importance of measuring added value, which requires determination of the effects of national impacts, as well as impacts at the European level (European Added Value).

All these recommendations by the Commission should be implemented in the form of an evaluation plan, i.e., tool used also in the previous perspective of EU's cohesion policy funds. As a rule, the plan should contain all evaluations to be carried out during the financial perspective. For the submission of the evaluation plan, the Monitoring Committee will have one year from the approval of the programme. Moreover, along with the objective, methodology, schedule, and budget of the evaluation, it should also include a detailed statement of data availability. The challenge in this regard, which the Commission highlights, is to build a cooperation model based on knowledge transfer among managing authorities, evaluators, and relevant data holders, especially in priority areas, among others, development of the SME sector.

“If some particular data are needed, the managing authority may need to negotiate with the relevant authority (e.g., national statistical office) sufficiently in advance to ensure the data are available when required. Equally, if the data required involve

confidentiality concerns, this needs to be identified and negotiated well before the data come to be used.” (European Commission, 2021, p. 15).

In short, programme managing authorities are tasked with securing data access in due time. As pointed out by the Commission, the experience of the previous perspective showed that shortcomings in providing access to existing data significantly increased the cost of studies (as data needed to be collected), or led to the situation where managing authorities and/or evaluators selected less rigorous research approaches. Therefore, synergy of activities with regard to ensuring an adequate range of data, their processing, and transmission of the resulting information is a critical element and should be of interest to entities responsible for the monitoring and evaluation of the 2021-2027 programmes.

6.1. Methodology for estimating the effects of support for enterprises vs. requirements for existing data

In view of the guidelines outlined by the European Commission, evaluators will be faced with the challenge of providing adequate data for cause-and-effect analyses, including the implementation of counterfactual models. The answer to the question about the impact of support targeting enterprises requires definition of the causal change observed after receiving support. Causal interpretation of the observed changes requires a closer description of the situation that would have taken place if the intervention had not happened (the so-called counterfactual situation). Next, counterfactual situation should be compared to what actually happened. In other words, it is important to imagine the situation of the entrepreneur that would have taken place if they had not received support. Originally, the counterfactual approach was developed on the grounds of statistics (Rosenbaum and Rubin, 1983) and econometrics (Heckman, 1979; 1989). The result of this work is an analytical model enabling evaluation of the effects of public interventions in situations where a classic experiment (Random Controlled Trial, RCT) would be difficult or, for various reasons, impossible to implement, or when only existing data is available. If, in theory, companies could be randomly assigned to a group receiving support (beneficiaries) and a control group (no support), the groups would be equivalent, i.e., differences between them,

in the size of a given effect indicator (e.g., revenue), could be interpreted as the result of the intervention exclusively. However, due to the fact that in reality enterprises differ from one another in terms of factors other than only receiving support (e.g., size, age, location, etc.), it is necessary to take into account the so-called selection bias (Heckman, 1979; Stolzenberg and Relles, 1997). Statistics offer a number of methods to reduce the bias, the most popular being *propensity score matching* (PSM) (Morgan and Winship, 2014). The same applies to selectable methods of estimating the so-called net effect⁵⁶, among which one of the most commonly used is the difference-in-differences (DID) method. The approach takes its name from the fact it accounts for two differences: difference between the supported and non-supported groups, and difference in time (before and after intervention). An abundance of methods, both of statistical matching and impact estimation, provides the opportunity to select the tools that will make it possible to estimate net effects of impacts on entrepreneurs accurately and precisely. However, if existing data are to be properly used in impact evaluations, they must meet a number of requirements that will be faced by the institutions involved in the evaluation of interventions. The most important may include:

- requirement for a sufficient number of samples and sub-samples (beneficiaries, and companies without support);
- measurement of effect indicators at two points in time minimum (before and after intervention);
- equivalence/invariability of indicator measurement at different points in time;
- schedules for the collection of public statistics data and administrative data, resulting in a time shift in their availability⁵⁷;
- sufficiently large pool of variables used in statistical matching (procedure for selecting a control group in the analysis of the net effect), necessary to reduce selection bias thus ensuring similarity of the compared groups. As characteristics of enterprises, the variables should be adequately reliable and cross-verified in various sources/databases;
- sufficiently big pool of control variables used in the regression model necessary to estimate the intervention's effect indicators;
- variables measured on an appropriate scale (preferably quotient scale);
- appropriately large variance of the variables used;

⁵⁶ Net effect is a change which can be attributed to the impact of intervention only, i.e., excluding change independent of the support provided and resulting from other factors (e.g., economic situation).

⁵⁷ In the case of administrative data, the process of corrections and validation may take as long as several years from the moment of the enterprise submitting a relevant report.

- variables remaining in specific correlational relationships with each other.⁵⁸

Above, we list just a few challenges connected with the stage of using existing data in the process of estimating the impact of support on the situation of enterprises. To be properly included in the evaluations prepared, these issues should be considered as early as possible at the stage of formulating evaluation plans for the 2021-2027 programmes.

6.2. Specific nature of administrative data and data collected by official statistics

In Poland, the experience related to the use of public data includes the use of data collected by official statistics, both for the purposes of general monitoring of development and evaluation of specific public interventions. Among others, they are reflected in the implementation of the STRATEG development monitoring system⁵⁹ and evaluation of EU support programmes for enterprises based on the counterfactual pattern⁶⁰. In the former case, we are dealing with a system of support for the process of monitoring the development and evaluation of the effects of activities aiming to strengthen social cohesion. In the latter case, the rich data resources collected by official statistics as part of regular

⁵⁸ The nature of relations among the variables is dependent on their use in a given analytical model. For example, some control variables should show a low level of mutual correlation, whereas they should be correlated with the *propensity score* coefficient in the model of matching a group with support and a control group in the evaluation based on the counterfactual pattern. In turn, estimation of the net effect requires control in the regression model of the variable affecting the effect and thus characterised by a high level of correlation.

⁵⁹ STRATEG Development Monitoring System: <https://strateg.stat.gov.pl/>

⁶⁰ See e.g. "Determining the values of selected economic indicators for beneficiaries of Regional Operational Programmes, Operational Programme Innovative Economy and for selected control groups"; "Determining the values of selected economic indicators for beneficiaries of the Operational Programme Human Capital and for selected control groups", <https://stat.gov.pl/statystyka-regionalna/statystyka-dla-polityki-spojnosci/>; "Return on investment of public support to SMEs and innovation in Poland, <https://www.worldbank.org/en/country/poland/publication/poland-return-on-investment>; Determining the values of selected economic indicators for recipients of support provided through PARP as part of the Operational Programme Smart Growth and Operational Programme Eastern Poland, and for selected control groups – stage 1. Feasibility study <https://stat.gov.pl/statystyka-regionalna/statystyka-dla-polityki-spojnosci/statystyka-dla-polityki-spojnosci-2016-2018/badania/badania-kontrfaktyczne/>

studies covered by the Statistical Survey Programme of Official Statistics (PBSSP)⁶¹ made it possible to operationalise a number of result indicators in order to assess the so-called net effects of support programmes for enterprises. Several years of experience in this area enabled development of effective models of cooperation among institutions implementing programmes, evaluators, and Polish official statistics, facilitating understanding of the requirements regarding statistical confidentiality as well as the potential and limitations related to the use of statistical data in evaluations.

Growing interest in the use of administrative data in the processes of formulating and evaluating public policies inspires analysis of the potential of wider use of the above-mentioned data in monitoring and evaluation, particularly in the context of Statistics Poland's previous experience in using data from statistical sources for these purposes.

Administrative data are information stored in administrative systems, usually kept by public entities (Elias, 2014), such as, e.g., tax administration, social security institutions, healthcare institutions, education institutions, etc. Administrative data are defined as existing data, collected for the purpose of public entities meeting their obligations and implementing their statutory tasks. In principle, it can be assumed that administrative registers cover large volumes of systematically collected and regularly updated data, characterized by a relatively high level of accuracy and reliability, covering the fundamental section of, and sometimes the whole population in question (Künn, 2015). The above features of administrative data are perceived as the source of their particular potential for being used in monitoring and evaluation of public policies. Stable and complete sources of quantitative data make it possible to conduct longitudinal studies and measurements at selected points in time, as well as significantly lowering the cost in relation to that of data collected specifically for the purpose of a given study (Card et al., 2010).

It should be emphasized that administrative data are collected for the implementation of public tasks specifically defined by the law for a given entity, that is, purposes fundamentally different than evaluation studies or public policies monitoring (Connelly et al., 2016). The tasks of a given data holder determine the information scope of the collected data, which is most often limited from the point of view of evaluation needs. This is particularly relevant

⁶¹ Statistical Survey Programme of Official Statistics is a document adopted by way of a Regulation of the Council of Ministers, specifying the detailed scope of studies conducted annually by public statistics: <https://bip.stat.gov.pl/dzialalnosc-statystyki-publicznej/program-badan-statystycznych/>

in the context of analyses of multidimensional phenomena and theoretical constructs (e.g., smart economy, socio-economic cohesion, inclusiveness), for which administrative data rarely provide information in a direct manner: they require advanced works at the level of methods, processing, combining of multiple sources and sets. Besides, these data are collected in formats most useful for the public institution that gathers them to achieve its statutory objectives (primary purpose of the data) rather than formats optimal for public policy researchers. Finally, administrative registers are not free from errors or data gaps, therefore, in order to obtain reliable information, it is necessary to implement validation and imputation methods (Hashimoto et al., 2014).

As in the case of administrative data holders, data collected by national statistical institutes are used to support the achievement of a number of public objectives by generating independent, reliable, high-quality statistical information on socio-economic phenomena and environment. This information is used to shape the development policies of the country and monitor and evaluate the changes that are taking place. The mission of official statistics gives it the mandate to collect rich resources of data on a number of aspects of social and economic life. It can be said that official statistics have the greatest abundance of variables, and thus the widest information scope of data of all public data holders. At the same time, standard studies conducted by official statistics are based on data (collected from households, economic entities and other legal persons, natural persons, etc.) which cover the whole population of entities analysed in the given cross-sections or study sample. Therefore, despite the rigour and high standards applied by official statistics when conducting studies⁶², problems related to data shortages, exclusion of entities from the study, and measurement errors may prove more challenging than in the case of administrative data.

As the above description of data sources shows, while administrative data are characterised by a much higher, frequently almost full level of coverage of a given population, the advantage of official statistics is their internal consistency and information scope of data, covering a large number of issues and thematic cross-sections. Both data sources may contain errors, thus requiring advanced, time-consuming work in the field of methodology and data engineering at the stage of sets' preparation for analysis. Similarly, they are characterised by strictly defined time limitations of data sharing (e.g., in the case of official

⁶² Cf. e.g., International recommendations on improving quality in official statistics: <https://bip.stat.gov.pl/dzialalnosc-statystyki-publicznej/jakosc-w-statystyce/rekomendacje-miedzynarodowe/>

statistics data, resulting from international and national obligations with regard to data publication).

In Poland, access to individual data collected in both administrative registers and official statistics systems is limited by the law. This stems mainly from the need to guarantee data security and statistical confidentiality. The Act on Official Statistics (Act, 1995)⁶³ highlights absolute protection of identifiable individual data⁶⁴ collected in statistical surveys. The same guidelines apply to data from administrative systems provided to Statistics Poland.

6.3. Use of administrative data in the evaluation of support for enterprises – opportunities and challenges

While the possibilities of using Polish official statistics data for the purpose of evaluating support programmes for enterprises have been verified in evaluation studies, broader use of administrative data for this purpose is only marginal. That may be due to the absence of coherent systemic solutions at the state level, regarding the provision of non-identifiable individual data⁶⁵, which play a special role in estimating result indicators. In practice, a large number of independent data holders means that provision of administrative data for purposes other than a given entity's public tasks takes place by way of bilateral arrangements with the institution that keeps a relevant register.

⁶³ Act on official statistics of 29 June 1995 (Journal of Laws No 88 item 439 as amended), cf., among others, Article 10, Article 38, Article 38(2).

⁶⁴ Pursuant to the Act on official statistics, identifiable individual data are “statistical data containing information regarding a specific entity of the national economy or a natural person, identifying directly this entity or person, by name, name and surname, address or publicly available identification number and allowing to indirectly identify this entity or person using other means than those enabling the direct identification, exclusive of means requiring excessive costs, amount of time or activities”.

⁶⁵ Non-identifiable individual data do not allow to identify, directly or indirectly, entities of the national economy or data subjects. The guarantee of non-identifiability is possible, among others, with the use of data anonymisation or pseudonymisation techniques. These are complex operations, in the case of which of key importance is the guarantee of privacy and, at the same time, the lowest possible loss of informative value of datasets.

Given the above conditions, currently, the opportunity of using administrative data in evaluation studies of enterprises is provided by the rich resources of registers provided to Statistics Poland by other public data holders⁶⁶. The experience of Polish official statistics in the implementation of tasks related to monitoring and evaluation is an excellent basis for extending the scope of cooperation with institutions supervising support programmes with information generated based on administrative data. It is also worth emphasizing that Polish official statistics are moving towards intensifying the use of administrative data in statistical research.⁶⁷

However, the use of administrative data in monitoring and evaluation of support for enterprises, in the context of the possibility, and, sometimes, necessity (e.g., for the purpose of validation) of combining them with statistical data, entails a number of challenges. When analysing administrative registers as a source of information about enterprises pursuing business activity in Poland, and comparing them with the statistical register, which is a sampling frame for surveys conducted by Polish official statistics, one can identify two key problems with the comparability of units that appear in them. First, the units of observation in the particular registers differ from each other. Secondly, the material scope of the variables and the way of their updating, and, consequently, the data defining basic information describing the units, are not identical. This particularly applies to the status of enterprises' activity (e.g., active, bankrupt, inactive), type of activity they pursue according to the Polish Classification of Activities (PKD), and basic data enabling the definition of the unit's size class (information on the average number of employees, income, ties, e.g., capital ties).

Definition differences appear at a level so basic as the subject of evaluation studies, i.e., the enterprise. The legal basis for the designation of observation units in statistical registers is provided by Article 10 of the Act on official statistics of 29 June 1995⁶⁸, which authorises the

⁶⁶ More information on datasets provided to Statistics Poland by administrative data controllers: GUS Public Information Bulletin / Official statistics activity / Statistical surveys / PBSSP for 2020 – interactive version / Part II. Information on provided data. List, and: Repository of Information Standards (knowledge base, including metadata on information systems of public administration): <https://stat.gov.pl/metainformacje/repozytorium-standardow-informacyjnych/>

⁶⁷ Cf. e.g., "Development of the methodology and estimation of the number of persons working in the national economy by place of residence and the major workplace at the NTS 4 level, registered unemployment rate at the NTS 5 level, and gross remuneration measures at the NTS 4 level", <https://stat.gov.pl/statystyki-eksperymentalne/>

⁶⁸ The Act on official statistics, op. cit.

President of Statistics Poland to use statistical data to create a sampling frame for statistical surveys. The personal and material content of the register is consistent with Regulation of the European Parliament and of the Council (EC) No 177/2008 of 20 February 2008, establishing a common framework for business registers for statistical purposes, and the observation applies to units referred to in Council Regulation (EEC) No 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community (OJ L 76, 30.3.1993, p.1). Therefore, the statistical register, along with units like natural persons pursuing business activity, legal persons, organisational units without legal personality, includes units characteristic for statistics only, such as the *enterprise* (the smallest combination of legal units, i.e., an organisational unit producing goods and services which benefits from a certain degree of autonomy in decision-making, in particular, in the allocation of current resources), which have no counterparts in administrative systems. The problem applies to the largest units and those having greatest impact on the Polish economy⁶⁹.

In administrative registers, the subject of registration under Article 4 of the Act of 6 March 2018 Entrepreneurs' Law (Act, 2018) has been defined as an *entrepreneur*. In the light of the above-mentioned Act, an entrepreneur is a natural person, a legal person, or an organisational unit without legal personality granted with legal capacity by a separate act, pursuing business activity, and partners of a civil law partnership in the scope of their business activity. Thus, official statistics use the term *enterprise*, while administrative registers use the term *entrepreneur*, and the terms' definitions are not consistent.

When comparing the registration system of official statistics with the systems of administrative registers, it should be stressed that in each administrative register the same unit subject to registration may be treated in a completely different way. The non-uniform approach to units pursuing business activity is due to the purpose for which the given register has been established. Among others, the following similarities and differences in the personal scope can be observed between the REGON register, essential for statistics, and other administrative registers:

⁶⁹ In statistics, one enterprise may be composed of several legal entities (e.g., limited liability companies). For example, in this case, these companies are subject to one management and a common policy while having no possibility of making independent decisions on, e.g., closing their activity or changing its type. Therefore, they have no freedom in decision-making. In statistical terms, a group of such units is one economic entity, such as the enterprise.

- natural persons pursuing business activity are registered in each of the systems in question. However, there are differences regarding the concept of “natural person pursuing business activity” between the REGON register and other registers, resulting from the fact that each register records the same natural person in terms of their different features or functions. In the National Register of Taxpayers (KEP), it is referred to as “natural person independently pursuing business activity”, while in the Central Register of ZUS Contribution Payers as “natural person”. In other words, a natural person registered with KEP or ZUS and treated there as a natural person pursuing business activity does not have to be included in the REGON register, due to the differences in the personal scopes of the particular registers.
- in the National Register of Taxpayers, there is often no legal person equivalent to an enterprise within the meaning of statistical nomenclature, obliged to submit consolidated financial results. Often, for a legal person, the results are submitted by an “establishment (branch) of the legal person” or “separate internal unit of the entity”. However, a separate internal unit is included only if its parent unit is a legal person or organisational unit without legal personality.

Another example to illustrate the problem of data compilation for entities present in various information systems of public administration may be the civil law partnership of natural persons. In public statistics, such companies are treated as separate entities pursuing business activity, regardless of partners. Both the company and the partners provide the required statistical data to the extent to which they pursue business activity (in the case of partners, they do not show the effects of business activity achieved in the company). Meanwhile, in the registers of the Ministry of Finance, a civil law partnership is treated in two ways. With respect to income tax, all partners are obliged to make individual settlements, while with respect to VAT, the company is treated as one economic entity, like in the statistical register.

Additional challenges for creating a coherent base of results stem from the specific nature of the Ministry of Finance’s resources, which are divided into a number of databases, covering only specified types of units, e.g., VAT payers, legal persons (payers using the CIT form in their settlements), natural persons (payers using the PIT form in their settlements). Each of these databases contains a different amount of information, which, during the process of base of results preparation, must be compiled in a way that will make it possible to maintain their consistency in line with the applied definitions.

It should also be noted that statutory obligations of administrative data holders and failure to take account of the needs of statistics result in gaps when data are generated. An example of this is the impossibility to obtain, directly from ZUS resources, full information on persons working in the main workplace, or information on remunerations paid to all employees of the enterprise. Only information about employees who the employer must report to the system is available⁷⁰.

Both statistical registers and administrative registers are characterised by nuances in the field of registration of entities, which makes it difficult to compare the units directly. However, use of appropriate techniques of data linking and validation and use of the degree of similarity between registration entities of these systems makes it possible to develop links between units to enable development of coherent information. The table below presents a list of identification indicators for the economic entities present in the resources of official statistics and in administrative registers, broken down into entities keeping accounting books and other types of records⁷¹. In addition to being useful at the level of merging sets, these features are also of key importance when conducting evaluation studies based on the counterfactual pattern.

Table 6.1. Identification indicators of entities present in the Polish official statistics system (STP) and administrative registers (ADM)

Indicator	Description of the indicator	Accounting books		Other accounting records	
		STP	ADM	STP	ADM
REGON	Statistical identification number	X	X	X	X
NIP	Tax identification number	X	X	X	X
Voivodeship	Symbol of the voivodeship where the enterprise is seated	X	X	X	X
Powiat	Symbol of the powiat where the enterprise is seated	X	X	X	X
Municipality	Symbol of the municipality where the enterprise is seated	X	X	X	X

⁷⁰ Frequent amendments to the tax law may complicate the issue and make certain procedures related to registration of employees outdated. For example, previously, a person working based on a contract of mandate, who received minimum remuneration from other sources thus paying social security and health insurance contributions, did not have to be registered with ZUS, while today they must be reported by their employer.

⁷¹ The accounting principles are included in the Act on accounting of 29 September 1994 (as amended) and in the Act of 26 July 1991 on personal income tax (as amended).

Indicator	Description of the indicator	Accounting books		Other accounting records	
		STP	ADM	STP	ADM
Polish Classification of Activities (PKD)	Symbol of the primary type of activity pursued by the enterprise	X	X	X	X
Date of starting the activity	Date of starting activity/issuing the first invoice	X	X	X	X
Legal form	Symbol of basic legal form (legal person/natural person)	X	X	X	X
Specific legal form	Symbol of specific legal form	X	X	X	X
Accounting books	Type of accounting records	X	X	X	X
Size of enterprise	Enterprise size class (specified pursuant to Annex I to Commission Regulation (EU) 651/2014)	X	X	X	X
Subsidy	Information on support granted for operating activity from public funds	X	X	X	X

Source: own study.

6.4. Opportunities of supplementing result indicators of enterprises' activity in the area of competitiveness, innovation, and employment

Official statistics data enable estimations of the effects of support programmes for enterprises, based on a number of competitiveness, innovation, and employment indicators. Data sets collected in the course of studies, such as, among others, Annual Enterprise Survey (SP), Report on Research and Development (R&D) Activities (PNT-01), or Industrial Innovation Report (PNT-02) make it possible to use most rigorous research designs, including counterfactual models.

At the same time, the use of administrative data for preparation of a basic data set regarding enterprises business activity requires consideration of a number of sets containing dispersed information, whose holders are the Minister of Finance and the President of ZUS. For instance, as regards data held by the Ministry of Finance, data from complex financial

statements should be used to develop result sets referring to revenues of units keeping full accounting records, i.e., units entered in the register of entrepreneurs of the National Court Register and natural persons keeping accounting books. For other units, i.e., those that do not keep full accounting records, data obtained on forms, e.g., VAT-7, PIT-36, PIT-36L, PIT-28, CIT should be taken into account. In addition, in order to take account of adjustments of income earned by business entities within civil law partnerships, information from the PIT-B annex and register of partners of civil law partnerships is additionally used. It should be emphasized that despite the complex, demanding procedure for generating information based on administrative registers, these sources can be a supplement to information obtained in representative statistical surveys, as they cover the full population of micro-enterprises.

Table 6.2. presents examples of result indicators used in the existing studies of the effects of support provided to enterprises, along with information on the possibility of supplementing them with statistical and/or administrative data. Having information from two sources provides additional opportunities for results validation.

Table 6.2. Review of sources for supplementing the indicators in the study of the effects of support for enterprises in the Polish official statistics system (STP) and administrative registers (ADM)

Indicator	Description of the indicator	Accounting books		Other accounting records	
		STP	ADM	STP	ADM
Competitiveness					
Net revenues from sale and equivalent revenues	Net revenues from sale; increase (+) or loss (-) (increasing or decreasing revenues from sales of products) in the value of stocks; treated on an equal footing with revenues, costs of self-production of benefits increasing assets	X	X	X	X
Net revenues from sale of goods and materials	net revenues from sale of goods and materials (excluding value added tax), i.e., amounts due for goods and materials sold	X	X		
Net revenues from sale of products	net revenues from domestic sale and export of products manufactured by the unit (finished products, intermediate products, and services), as well as external services, if invoiced to customers along with products	X	X		

Indicator	Description of the indicator	Accounting books		Other accounting records	
		STP	ADM	STP	ADM
Net revenues from sale of products for export	total value of revenues from sale of products, which relates to sale for export	X	X		
EBIT – earnings before deducting interest and taxes	operating profit/loss	X	X		
EAT – earnings after taxes;	net profit/loss	X	X		
ROE – return on equity	$(EAT * 100) / \text{equity (fund), as of the end of the year}$	X	X		
ROA – return on assets	$(EAT * 100) / \text{total assets, as of the end of the year}$	X	X		
ROS – return on sales	$(EAT * 100) / \text{net revenues from sale}$	X	X		
PP – labour productivity	net revenues from sale / average number of employed persons, in FTEs	X	X	X	X
Cost level indicator	total costs to total revenues ratio in %	X	X	X	X
Revenue dynamics indicator	ratio of revenues in the year of project completion to revenues in the year of joining the project	X	X	X	
Indicator of dynamics of the number of working persons	ratio of the number of working persons in the year of project completion to the number of working persons in the year of joining the project	X	X	X	X
Innovations					
Capital expenditure	financial or material expenditures aimed at creating new fixed assets or improving (alteration, extension, reconstruction or modernisation) of existing fixed assets as well as expenditures on the so-called first equipment of the project	X	X	X	X
Internal R&D expenditures	financial expenditures incurred for all R&D works carried out by the unit, i.e., both R&D work performed for own purposes and on request of other entities (provision of R&D services) in a given reporting year	X			
External R&D expenditures	amount spent on R&D activities performed outside the reporting unit	X			
R&D staff in persons	all persons directly involved in R&D work, and persons providing direct support (e.g., R&D work managers, administrative and office employees), whose full-time equivalent was at least 0.1 or whose contribution in the R&D work performed was very important	X			
Number of patent applications	Number of applications submitted by the unit with the Patent Office of the Republic of Poland in 2020	X			
Number of patents obtained	Number of patents obtained from the Patent Office of the Republic of Poland in the reporting year	X			

Indicator	Description of the indicator	Accounting books		Other accounting records	
		STP	ADM	STP	ADM
Employment					
Employment in FTEs	Average number of persons employed after conversion of part-time employees to full-time employees	X	X	X	X
Number of working persons	Number of persons working in the unit as of the reporting day	X	X	X	X
Gross remunerations	gross remunerations, i.e., including income tax as well as pension, disability pension, and sickness contributions (regardless of the sources of financing payments, i.e., both from own and reimbursed resources, e.g., from the State Fund for Rehabilitation of Disabled People) including profit sharing and balance surplus payments in cooperatives and payments from the company's bonus fund; without remuneration for external employees	X	X	X	X

Source: own study, as of 15.11.2021

It is also worth emphasizing that in addition to the existing ways of conceptualising competitiveness and innovation, used in the existing studies, including evaluation studies, the scope of important indicators for enterprises is directly specified at EU level. The Regulation of the European Parliament and of the Council (EU) 2021/1058 of 24 June 2021 on the European Regional Development Fund and the Cohesion Fund (Regulation 2021/1058, 2021) provides a list of common output indicators for instruments addressed to enterprises. This list is common for the EU countries and includes output and direct result indicators monitored at both project and programme level. Therefore, this property should translate into the possibility of aggregating the effects of the cohesion policy at EU level. In the case of instruments offered to entrepreneurs, the role of Policy Objective 1, i.e., “More competitive and smarter Europe by promoting innovative and smart economic transformation and regional ICT connectivity (PO1)” should be highlighted. Under the five specific objectives of CP1, several dozen of indicators measuring the properties and achievements of enterprises have been categorised. Some of these indicators may be clearly and directly identified in official statistics data and administrative data. However, some of them will be subject to the process of thorough conceptualisation and operationalisation to match both data resources. Therefore, the basic criterion determining the quality of measurements of the effects of impacts on the enterprise sector will remain the accuracy of decisions in the definition process (Górniak and Keler, 2007).

Summary and recommendations

Years of experience of European countries, including Poland, in implementing tasks related to the monitoring and evaluation of EU-funded programmes have resulted in a number of good practices and conclusions regarding the approach to estimating the effects of interventions. The European Commission's emphasis on the use of existing data, i.e., official statistics data and administrative data is undoubtedly in pursuit of improving the quality of studies (their relevance, accuracy, and reliability), including cause-and-effect analyses. Such an approach should also guarantee improved cost-effectiveness and reduced burden for respondents, when compared to research processes based on primary data, collected to monitor numerous support programmes and conduct specific evaluation studies. At the same time, the recommended complementary use of the above-mentioned data sources stems from the accurate prediction and awareness of different purposes of these data, and the related limitations of the particular sets, which can be reduced by combining the potential of various data sources.

This chapter attempts to outline the possibilities of a wide use of the existing public data held by Polish official statistics for the purposes of monitoring and evaluation of enterprises' support. The area of particular interest for the authors is the use of data from administrative systems for this purpose. The possibilities of their application have been presented in relation to the existing experience in implementing evaluations based on data collected by Statistics Poland (GUS). The role of official statistics and the conditions of their functioning in the state's information system, including, among others, access to data from multiple registers and data sets, and the possibility of their integration, are indicative of the potential of generating rich and complex statistical information. While some administrative registers offer public access or the possibility of obtaining data under the law by individual agreements with a given data holder, their individual usefulness for complex analyses covered by monitoring or evaluation is limited, often providing only a fragmented picture of the phenomena taking place in the analysed population.

Proper design and implementation of the process of using existing public data in monitoring and evaluation of support for enterprises is a complex issue that requires a number of procedures at the political level – legislative or institutional, and technical level – with respect to data methodology, quality, engineering, and analysis. In the absence of a uniform system for public data collection and sharing, providing a researcher with access to individual

data that will remain non-identifiable is a challenge. Dispersion of data across multiple systems and registers, specific methods of their collection, and lack of consistency in the definition of variables are just some of the difficulties whose solution is of great importance for studies assessing public policies and development strategies.

Both statistical and administrative data have great potential for use in the monitoring and evaluation of support for enterprises. By definition, administrative data sets are characterised by full coverage of the analysed population, but they are not free from errors and data gaps. Moreover, they contain a narrower scope of variables that can be used in estimating project result indicators. The opportunity of obtaining the widest possible scope of information, with the highest level of population coverage, is provided by combining statistical and administrative data sets. Administrative data are not collected in a standardised manner, but in the formats and shape necessary for a given data holder to implement specific public tasks. Therefore, it is necessary to standardise them properly and, later, process them into statistical data, applying the highest methodology and quality standards applicable to official statistics.

It should also be emphasized that it is the objectives and assumptions of monitoring and evaluation that determine the manner and possibilities of using data, by generating new challenges each time. Therefore, the process of developing consistent databases for the purposes of a specific evaluation is exceptionally time-consuming, requires the involvement of high-class experts with extensive substantive, methodological, data engineering, and analytical knowledge. Above all, it requires early involvement of data holders and evaluators in the process of study planning as well as integration and coordination of the work of communities pursuing public policies. This is also important in view of the fact that, as indicated above, the list of common output indicators for the EU countries regarding support instruments for enterprises includes certain measures which cannot be directly supplemented with statistical or administrative data, but require complex operations with regard to developing appropriate data sets and the accompanying processing procedures.

In the face of the above challenges, in Poland, it is extremely important to involve representatives of official statistics in the processes of programming interventions and planning monitoring and evaluation of support for enterprises. Early-stage comparison of the expectations of institutions supervising public interventions in the area of the desired methodology and scope of the information evaluation study, with genuine assessment of the

possibility of supplementing specific objectives of analyses with the required indicators and/or developing approximation measures (on the basis of both statistical and administrative data) constitutes the basis for implementing highest-quality research designs.

Experience and conclusions from the implementation of monitoring and evaluation of EU programmes may be successfully used and developed in the assessment of all development policies, also those financed from the national budget. Evidence-based formulation and pursuit of policies is, in fact, the basis for the purposefulness and relevance of activities, as well as for effective allocation of public funds. Given the difficulties currently faced by institutions responsible for assessing the effects of interventions, and the resulting limitations for research projects as such, it would be recommended to build an integrated system for collecting and processing public data at the state level. It is worth highlighting that these processes have already been partially initiated – official statistics are increasingly using administrative data, improvements are being introduced to the flow of information among public institutions (e.g., introduction of the so-called “one-stop-shop” rule when registering a company, implying cooperation and consistency of information among registers, treating the NIP number as a key identifier in business transactions, etc.). Managing the flow of information on the scale of all public administration units and implementing good practices that take account of information needs in the process of designing data systems at the interinstitutional level should be the next stage of development.

In view of the special role of official statistics in the state’s information system, as well as previous experience and effective models of cooperation between Statistics Poland and institutions implementing tasks related to assessing the effectiveness of support for enterprises, it appears that official statistics should play a key role in this process. Locating data from various data registers and systems in one place would be a milestone in further development of the processes of public policies’ monitoring and evaluation in Poland.

References

- Act on Official Statistics (1995) [Ustawa o statystyce publicznej z dnia 29 czerwca 1995 r.], (Polska). <http://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=wdu19950880439>
- Act Entrepreneurs’ Law (2018) [Ustawa Prawo przedsiębiorców z dnia 6 marca 2018 r.], (Polska). <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20180000646>

- Card, D.E., Chetty, R., Feldstein, M. S., & Saez, E. (2010). Expanding Access to Administrative Data for Research in the United States. Downloaded on 19 October 2021 from <https://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf>
- Connelly, R., Playford, C.J., & Gayle, V. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, 59, 1-12. <https://doi.org/10.1016/j.ssresearch.2016.04.015>
- Elias, P. (2014). Administrative data. W A. Duża, D. Nelle, G. Stock, & G. Wagner, G. (Red.), *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences* (pp. 47-48). SCIVERO.
- European Commission. (2021). *Performance, monitoring and evaluation of the European Regional Development Fund, the Cohesion Fund and the Just Transition Fund in 2021-2027*. COMMISSION STAFF WORKING DOCUMENT. https://ec.europa.eu/regional_policy/en/information/publications/evaluations-guidance-documents/2021/performance-monitoring-and-evaluation-of-the-european-regional-development-fund-the-cohesion-fund-and-the-just-transition-fund-in-2021-2027
- Gorynia M. (2000). Koncepcja i metodyka badania konkurencyjności przedsiębiorstwa. W J. Bossak (Red.), *Konkurencyjność gospodarki Polski w dobie integracji z Unią Europejską i globalizacji* (p. 89). Oficyna Wydawnicza SGH.
- Górniak J., & Keler K. (2007). Wskaźniki w ewaluacji ex post programów publicznych. W A. Haber (Red.), *Ewaluacja ex-post. Teoria i praktyka badawcza* (pp. 155-181). Polska Agencja Rozwoju Przedsiębiorczości.
- Grzebyk M., & Kryński Z. (2011). *Konkurencja i konkurencyjność przedsiębiorstw: ujęcie teoretyczne, Nierówności społeczne a wzrost gospodarczy. Zeszyty Naukowe Uniwersytetu Rzeszowskiego*, 20, 449-450.
- Hashimoto, R.E, Brodt, E.D., Skelly, A.C., & Dettori, J.R., (2014). Administrative database studies: Goldmine or goosechase? *Evidence-Based Spine-Care Journal*, 5(2), 74–76.
- Heckman, J. (1979). Selection bias as a specification error, *Econometrica*, 47(1), 153–161.
- Heckman, J. (1989). Causal inference and nonrandom samples, *Journal of Educational Statistics*, 14(2), 159–168.
- Künn, S. (2015). The challenges of linking survey and administrative data, *IZA World of Labor*, 214, 1-10.
- Morgan, S., & Winship, C. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)*. 2nd ed. Cambridge: Cambridge University Press.
- Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70(1), 41–55.
- Regulation 177/2008 of 20 February 2008 establishing a common framework for business registers for statistical purposes and repealing Council Regulation (EEC) No 2186/93, Regulation nr 177/2008 (2008) (UE). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008R0177&from=EN>
- Regulation 696/93 of 15 march 1993 on the statistical units for the observation and analysis of the production system in the Community. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31993R0696&from=EN>
- Regulation 2021/1058 of 24 june 2021 on the European Regional Development Fund and on the Cohesion Fund, Regulation nr 2021/1058 (2021) (UE). <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021R1058&from=PL>

- Stankiewicz, M.J. (2005). *Konkurencyjność przedsiębiorstwa. Budowanie konkurencyjności przedsiębiorstwa w warunkach globalizacji*, Towarzystwo Naukowe Organizacji i Kierownictwa SWU Dom Organizatora.
- Stolzenberg, R., & Relles, D. (1997). Tools for Intuition about Sample Selection Bias and Its Correction, *American Sociological Review*, 62(3), 494–507.
- System Monitorowania Rozwoju STRATEG. <https://strateg.stat.gov.pl/>
- Szymanik, E. (2016). Konkurencyjność przedsiębiorstwa – główne aspekty, *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, 5(953), 107-124.

7. Examples of supporting the implementation and monitoring of public policies based on administrative data

Marek Bożykowski

Faculty of Sociology

University of Warsaw

Agnieszka Chłoń-Domińczak

Collegium of Economic Analysis

SGH Warsaw School of Economics

Mikołaj Jasiński

Faculty of Sociology

University of Warsaw

Introduction

The aim of this chapter is to outline the lessons learnt from studies supporting the development of public policies and monitoring of their outcomes conducted by the University of Warsaw and SGH Warsaw School of Economics. These experiences are mainly related to the implementation of the Integrated Analytical Platform (*Zintegrowana Platforma Analityczna – ZPA*) project. The ZPA project was created to address the need for tools to support the state governance process. The main objective of the project is to increase the effectiveness of public administration activities in selected socio-economic areas by supporting decision-making processes with comprehensive information. The project consists of six profiled studies on important social and economic issues covering the areas of education, labour market, social security, health, and access to public services. The studies are conducted in close cooperation between analytical teams and study owners, which is in line with international best practices.

The chapter presents the experience of conducting a diagnosis of resources collected in public registers for the purpose of analyses, including the role of cooperation with hosts, and the necessary elements of analyses under the ZPA. In particular, we present the framework for the study process adopted under the project, covering a diagnosis of resources and development of intervention logic for public policies, including the design of indicators and analytical models based on data from public registers for the areas covered by the studies conducted under the ZPA. Next, we discuss the legal and organisational circumstances related to conducting studies using administrative data with the help of the ZPA platform. Finally, we demonstrate the potential of the ZPA as a tool to support the development of evaluation research.

7.1. The study process in selected areas of public policies in the ZPA project

Since the beginning of the ZPA project, a specific model of study execution has been adopted. Primarily, the development and application of this model aims to ensure a uniform methodological approach throughout the project. Secondly, the adopted model enables the involvement of respective study owners at the particular stages of the study, which ensures consistency between their needs and expectations and the analytical solutions developed.

In the ZPA project, the study owners were respectively:

Study 1: Analytics of public e-services and support of strategy governance for the development of services for citizens and entrepreneurs – Chancellery of the Prime Minister (formerly Ministry of Digitalisation);

Study 2: Optimizing the allocation of funds for health care and social policy: Improving the health system's efficiency and return to work – Ministry of Health;

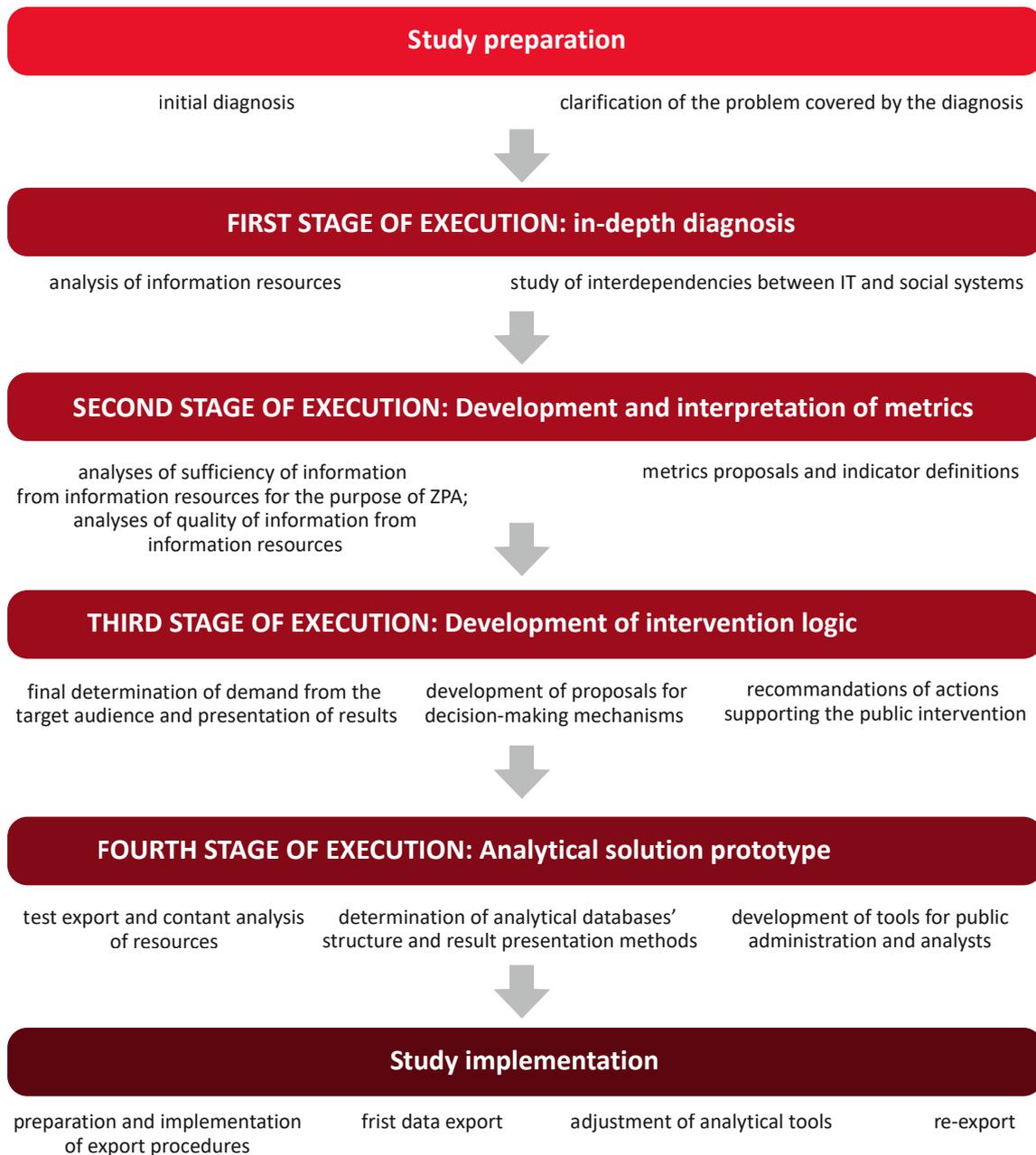
Study 3: Job activation of school graduates – Ministry of Education and Science;

Study 4: Analysis of the use of social security system benefits and net public transfers of individuals and households – Ministry of Family and Social Policy;

Study 5: Education pathways of children and youth – Ministry of Education and Science;**Study 6: Optimizing the allocation of funds for health care and social policy: Environmental and social determinants of health inequalities – Ministry of Health**

The ZPA project is implemented under the Digital Poland Operational Programme 2014-2020, priority axis 2. “E-administration and open government.” Measure 2.2 “Digitalisation of back-office processes in government administration.” Its outcome is the development of solutions enabling the initiation and execution of studies supporting the development of public analytics using administrative data within the ZPA platform. After the project’s completion, it is assumed that it will be possible to carry out new studies on the basis of ZPA regulations.⁷² Recommendations for the processes of initiating and executing later studies are being developed as part of the project. Figure 7.1. presents the different stages of studies under the ZPA.

⁷² Information on legal regulations is presented in the chapter entitled “From traditional data collection to administrative data – dilemmas and perspectives.”

Figure 7.1. Study stages under the Integrated Analytical Platform project

Source: The ZPA project

The next section details the particular stages of the study process under the ZPA.

Study preparation

Preliminary diagnosis is the starting point for study conceptualization. The purpose of this stage is to, first, clarify the topic to be covered by the analyses – in consultation with the relevant study owner. During this stage, public policy instruments to be covered by the analyses are assessed (e.g., use of specific health benefits, or use of specific social benefits), and the population to be covered by the analyses is defined.

For example, Study 4, in consultation with the study owner, identified three groups of benefit recipients covered by the analyses:

- families, in particular families with young children;
- unemployed people registered in employment offices;
- people with disabilities.

Subsequently, research questions (analytical issues) concerning the surveyed population are formulated, e.g., related to the use of specific benefits, characteristics of persons or households using selected public policy instruments, or economic activity of the selected population. Moreover, public registers are identified where information and data are collected that can be used in the process of seeking answers to the questions (analytical issues) identified in the study.

In-depth diagnosis of resources

In-depth resource diagnosis is the first of the four stages study execution. It aims to establish a reference point for further analysis. This stage includes a needs analysis of the target audience, based on consultations with the study owner as well as the available data and information on the selected population. Also identified are the potential benefits that the study can provide for the pursuit of public policies.

Besides, the in-depth diagnosis includes a detailed analysis of the registers covered by the study. The main objective of resource diagnosis is to identify the actual content of the surveyed registers, determine the quality and reliability of the data, as well as their usefulness in the context of a given study. Diagnosis of the particular registers also covers identification of the legal basis for their functioning, and current practices related to data

export. Diagnosis of resources facilitates demand assessment among the recipients of the analyses' findings. It enables the preparation of analysis proposals and assessment of the feasibility of analyses requested by recipients. Further, the results of the diagnosis form the basis for designing analytical variables and indicators. They also facilitate future arrangements regarding data export from register hosts. During this stage, resource diagnosis cards are developed to enable proper structuring of the diagnosis. Diagnosis cards are filled in by analysts conducting the particular studies, in cooperation with data hosts of the particular registers.

Thanks to this measure, each register included in the study⁷³ has an appropriately structured diagnosis, which constitutes an element of the Knowledge Base developed under the ZPA. The Knowledge Base can also be used in later initiatives related to the use of these registers for the evaluation of public policies exceeding the scope covered by the ZPA. The database will be a repository of reports, models, and analytical processes generated in connection with the execution of studies conducted in the ZPA environment. The Knowledge Base will be a resource available free of charge to all entities interested in initiating and executing studies under the ZPA as well as other recipients interested in using public registers to carry out analyses. The conclusions of the in-depth resource diagnosis enable the description of the legal infrastructure, resource structure, data quality register, data export practices, and database structure for each of the registers used in the particular studies.⁷⁴

Another element of the in-depth diagnosis is the analysis of the interdependence between social and IT systems. It shows the extent to which information from registers corresponds to the fragment of social reality that they cover. In particular, this part of the analysis covers the potential differences between information entered in the registers and their sources. In addition, it covers the possible origin of data quality issues and the solutions applied in order to ensure the quality of data collected in the registers and reduce the risk of errors made by those responsible for the input of data into the registers. Moreover, a diagnosis is carried out of the possibilities of using the data collected in the registers covered by the diagnosis for analyses, including evaluations and scientific research.

⁷³ Under the ZPA project, a total of 33 registers from 12 hosts are covered by in-depth diagnosis (cf. Table 7.1. List of hosts and registers included in the ZPA project).

⁷⁴ As a deliverable of the ZPA project, the Knowledge Base will be made publicly available at the end of the project, in 2023. It will also be further developed as part of ZPA operation.

The final element of the in-depth diagnosis is a cross-sectional analysis of the registers, including identification of the type of characteristics of the surveyed population present in the particular registers. In particular, it is important to determine whether the registers use a common identifier of persons (PESEL number) and entities, which enables the integration of information resources collected in different registers.

Table 7.1. List of hosts and registers included in the ZPA project

Host	Register
Agricultural Social Insurance Fund (KRUS)	Farmer
	nSIU
	KACHNA
The Chancellery of the Prime Minister (formerly Ministry of Digitalisation)	Civil Registry Office registers – BUSC (marriages, births, deaths)
	PESEL (personal identification number)
	RDO (identity cards)
Ministry of Education and Science/IT Centre for Education	Examination databases maintained by Regional Examination Boards (OKE)
	Educational Information System registers (SIO2)
	Registers of the IT System for Handling Exams Confirming Qualifications in the Occupation (SIOEPKZ)
Ministry of Education and Science/National Information Processing Institute (OPI PIB)	POL-on registers
Ministry of Finance	Register of Personal Income Tax (PIT)
Ministry of Family and Social Policy	Registers of social assistance beneficiaries: central (CAS) and registers in municipalities
	Registers concerning benefits for families: childcare benefit, family benefit with supplements (CAS, CBB)
	Electronic National System of Disability Assessment Monitoring (EKSMOoN)
	Registers of unemployment benefits: central register in the Ministry of Family, Labour and Social Policy and registers in district employment offices for the registered unemployed (Syriusz)
State Fund for Rehabilitation of Disabled Persons (PFRON)	System for Subsidies and Reimbursements (SODiR)
Ministry of Health – e-Health Centre (CeZ):	Pharmaceutical register (RL)
	Reimbursed prescription register
	Register of Entries to Poland (EWP)
	Register of people vaccinated against COVID-19

Host	Register
National Health Fund	Central List of Insured Persons
	Register of health care services financed from public funds
	Pharmaceutical Programs (Monitoring System for Therapeutic Programs)
Social Insurance Institution	Central Register of Insured Persons
	Central Register of Beneficiaries
	Central Register of Social Contribution Payers
	Central Register of Family Members
	Register of Social Insurance Institution (ZUS) medical certificates

Source: The ZPA project

It is also determined which register is the primary source of data such as sex, date of birth, date of death, information on marriage, parents' data (enabling determination of composition of families, in particular information on parents and their children), and identification of household members, addresses, levels of education and examinations passed, educational activity, disability, incapacity for work, occupational activity and labour market status, employer characteristics, sources of livelihood, social contributions paid, use of social benefits, information on periods of unemployment, medical history including diagnoses and treatments (as relevant for the particular field of study).

Development of a set of metrics and their interpretation

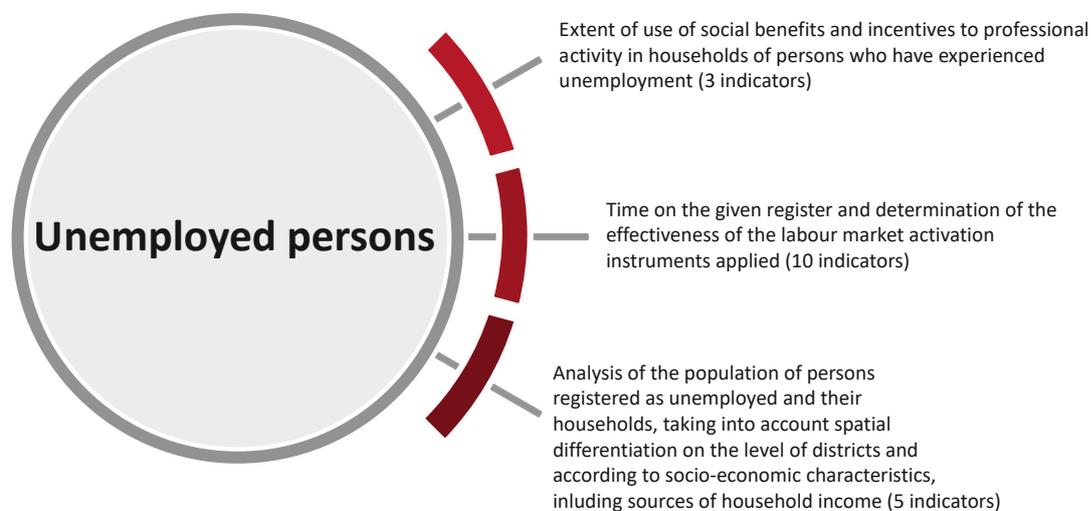
The second stage of study execution is the development of a set of metrics. During this stage, a set of indicators is developed assigned to the analysed populations. This includes the indicators' definitions, calculation algorithm, and a statement in which cross-sections they can be calculated.⁷⁵ Moreover, this stage includes identification of the registers used for the calculation of a given indicator as well as potential limitations concerning the current possibilities of using the resources collected in the listed registers (e.g., it is not possible to use the information from the database of social welfare beneficiaries due to the lack of beneficiaries' identification data; certain fields in the databases contain different information

⁷⁵ The decision whether to make the indicator base public is taken by study owners.

due to the lack of predefined codes i.e., standardised methods for filling in data). Indicators for each population are also grouped by analytical category.

Figure 7.2. presents an example of such a grouping for the unemployed population, developed as part of ZPA Study 4.

Figure 7.2. Indicators identified in Study 4 for the population of the unemployed: division into categories



Source: The ZPA project

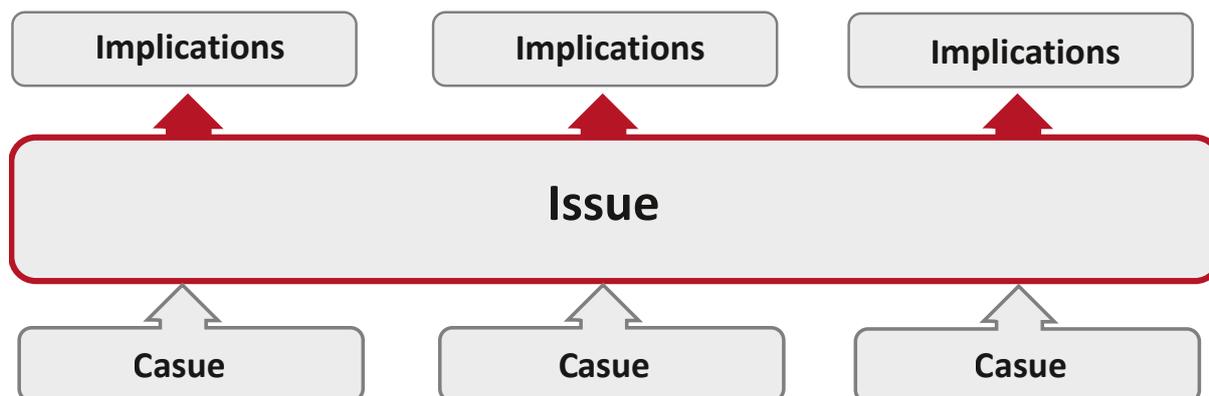
At this stage of the study, the set of indicators developed by the analytical team is also verified using a survey of the target audience including representatives of the study owner and other potential users of the analyses.

In the case of ZPA Study 4, a survey was conducted, as part of the target group analysis, of the representatives of the study owner – the Ministry of Family and Social Policy. Its aim was to collect information concerning the ranking of indicator proposals for the three surveyed populations, to identify the preferred public policy instruments for addressing potential issues and determine the preferred indicator visualisation methods. Results of the survey made it possible to identify a set of 30 indicators (10 for each of the surveyed populations), which were subjected to the subsequent stages of analytical work within the framework of the executed study.

Development of intervention logic

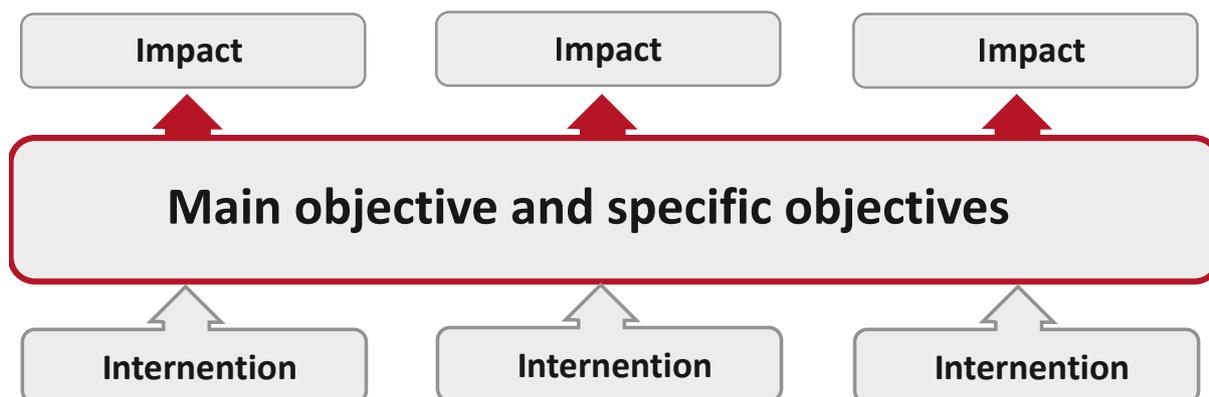
The third stage of study execution is the development of intervention logic. It aims to elaborate on and present how the indicators designed in the framework of the study can support decision-making in public policy interventions. Support of decision-making processes with information generated under the ZPA is related to the support in diagnosing emerging problems. On the one hand, indicators developed as part of the particular studies may serve to diagnose problems and, on the other hand – evaluate the results of public interventions, i.e., whether the applied intervention (change of public policy) led to reduction or elimination of the diagnosed problem.

Monitoring of indicators defined in the framework of the study makes it possible to determine whether a specific problem occurs, and what its causes and implications may be, as shown in the problem tree presented in Figure 7.3. For example, poverty can be caused by low socio-economic status, low level of education, or loss of employment, and lead to deteriorating health, addictions, and social exclusion. It is important to use existing register data to identify problems, as well as their causes and consequences. This kind of understanding of the occurring phenomena enables appropriate shaping of public policies that address both the causes and implications of the identified problems. The problem tree is a starting point for defining the objectives of public intervention, which can address the causes of problems (e.g., active labour market policies support activation of people who lost their jobs) as well as their implications (e.g., social assistance benefits help to protect households from the consequences of the lack of income).

Figure 7.3. Problem tree.

Source: the ZPA project elaboration based on: https://archiwum.ncbr.gov.pl/fileadmin/user_upload/import/tt_content/files/logika_intervencji.pdf (access: 17/02/2021)

On the other hand, an objective tree can be used to define objectives. It identifies the aspects of the target situation to be achieved at the end of the intervention, as shown in Figure 7.4.

Figure 7.4. Objective tree.

Source: the ZPA project elaboration based on: https://archiwum.ncbr.gov.pl/fileadmin/user_upload/import/tt_content/files/logika_intervencji.pdf (access: 17/02/2021)

The logical public intervention model adopted in the ZPA project includes:

- problem identification using diagnostic indicators developed in the framework of the particular studies⁷⁶;

⁷⁶ Within the framework of the studies carried out under the ZPA, it is also possible to identify/verify problems that have not yet been thoroughly investigated and which is allowed by the administrative data collected in public registers.

- identification of the method of public intervention, accounting for the legal and organisational instruments used to pursue public policy;
- identification of the expected short- and long-term change resulting from the intervention;
- monitoring of the change using evaluation indicators.

The following public policy instruments were taken into account as part of the intervention logic developed in Study 4 of the ZPA project (the list of instruments covers the areas of public policies analysed in the ZPA. It is not exhaustive and, depending on the context, may be modified to include other intervention instruments):

- Introduction of new regulations or amendment of existing regulations:
 - Amendment of a law;
 - Amendment of a regulation;
- Direct intervention at the central level:
 - Establishment of a central programme;
 - Changes to the scope of a central programme;
 - Establishment of a problem team;
 - Changes to the level of funding of particular public policies;
- Indirect intervention at regional/local level:
 - Minister's recommendations for local government policies
- In-depth evaluation and monitoring of implemented policies:
 - Launch of cyclical monitoring of a given public policy;
 - Launching of *ex post* evaluation;
- Changes to the register:
 - Register audit;
 - Modification of the collection/processing of data in the register;
- Education, information, and consulting:
 - Public and citizen consultation regarding public policies;
 - Public education activities/campaigning;
 - Workshops/training for individuals;
 - Identification and description of good practices.

It is worth noting that at this stage of study execution, it is necessary to consider the preferred methods of indicator visualisation. As public administration practice demonstrates, an appropriate graphic presentation of indicators, their changes over time, and their

variation with respect to the particular features of the monitored population, including geographical variation, supports identification of issues and monitoring of the extent of the objectives' achievement.

Analytical solution prototype

The last stage of study execution is the development of a prototype. During this stage, the outcomes of the activities conducted during the execution phase are summarised, which includes formulation of the purpose of the analytical solution developed, a summary of the analysis of needs and problems covered by a particular study, as well as analysis of the legal environment, inventory of hosts, analysis of sources, and analysis of stakeholders.

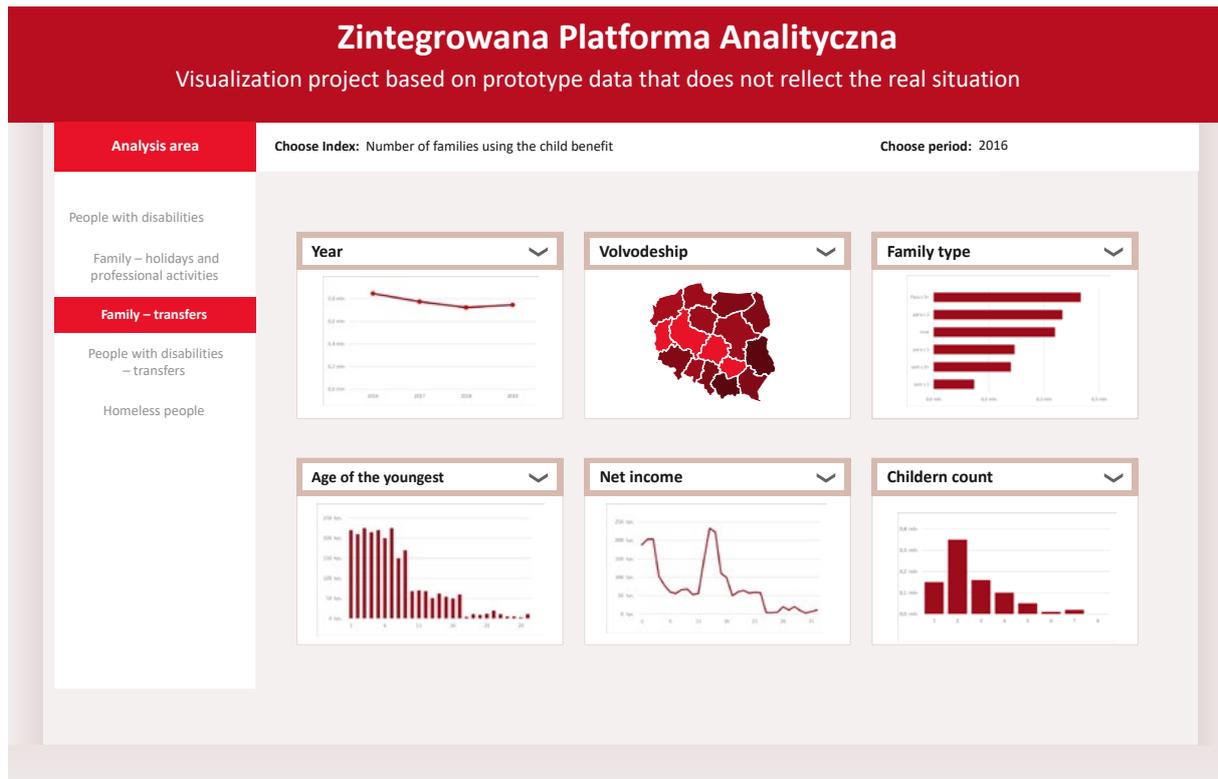
In ZPA Study 4, the stakeholder analysis covered the following groups:

- central administration and related entities, including public policy makers (e.g., keepers of registers covered by the project and their subordinate institutions);
- local authorities;
- labour market institutions, including public employment services, Voluntary Labour Corps, employment agencies, training institutions, and institutions of social dialogue and local partnership;
- broadly understood academia;
- citizens and NGOs.

The prototype is to include:

- detailed algorithms for calculating the indicators defined in the study;
- descriptions of the export structure for the particular registers included in a given study;
- attribute descriptions including a list of registers, list of attributes with their names and names of registers, sets and subsets where they are located, data type, database and validation limitations, and information about the glossaries used;
- draft proposals for visualisation of indicators, including dashboard proposals.

Figure 7.5. presents an example of indicators visualisation in ZPA Study 4.

Figure 7.5. Data visualisation project in ZPA Study 4.

Source: The ZPA project

Study implementation phase

The four-stage execution phase described above is followed by the implementation phase. The first stage of implementation covers the period from the moment the prototype is ready to the first data export. It includes preparation and implementation of export procedures at the hosts, and the first data export. The second stage of implementation comprises analytical work carried out in the ZPA analytical environment. The starting point is the data obtained during the first export from the hosts. They form the basis for preparation of the first analytical report and adjustment of the analytical tools prepared in the framework of a given study. The second stage of implementation also envisages re-execution of the validation and export stage. Upon completion of validation, final analyses and a summary of this phase are prepared, which are subject to approval by the study owner. This work is summarised in an implementation report and publication of results in the form approved by the study owner. In parallel with the preparation of the final export, user manuals are developed and training is provided for the target audience of the study.

7.2. Legislative background and organisational arrangements for the functioning of the Integrated Analytical Platform

Implementation of the ZPA project required the development and implementation of a specific legal framework for its operation. As described in the chapter “From traditional data collection to administrative data – dilemmas and perspectives,” as part of the amendment to the Act on computerisation of the bodies performing public tasks, Chapter 3b “Integrated Analytical Platform” was added to the Act.

These provisions provide the legal basis for conducting studies under the ZPA. In particular, they define the tasks of the minister responsible for computerisation, concerning:

- operation of the Integrated Analytical Platform understood as an organisational and technical solution for conducting analyses supporting key public policies;
- processing of data provided by public registers and the possibility of data provision by hosts, including the possibility of combining data, if necessary for the specific analyses defined in the ZPA;
- pseudonymisation and anonymisation of data submitted to the ZPA.

The adopted provisions also contain a mandate to develop regulations concerning:

- the scope of data and the list of public registers and ICT systems which provide the data needed for the analyses under the ZPA, and the list of entities that operate them obliged to provide data from these registers and systems,
- the manner of publishing these data, so as to ensure effective collection of data and compliance of data provision with personal data protection regulations.

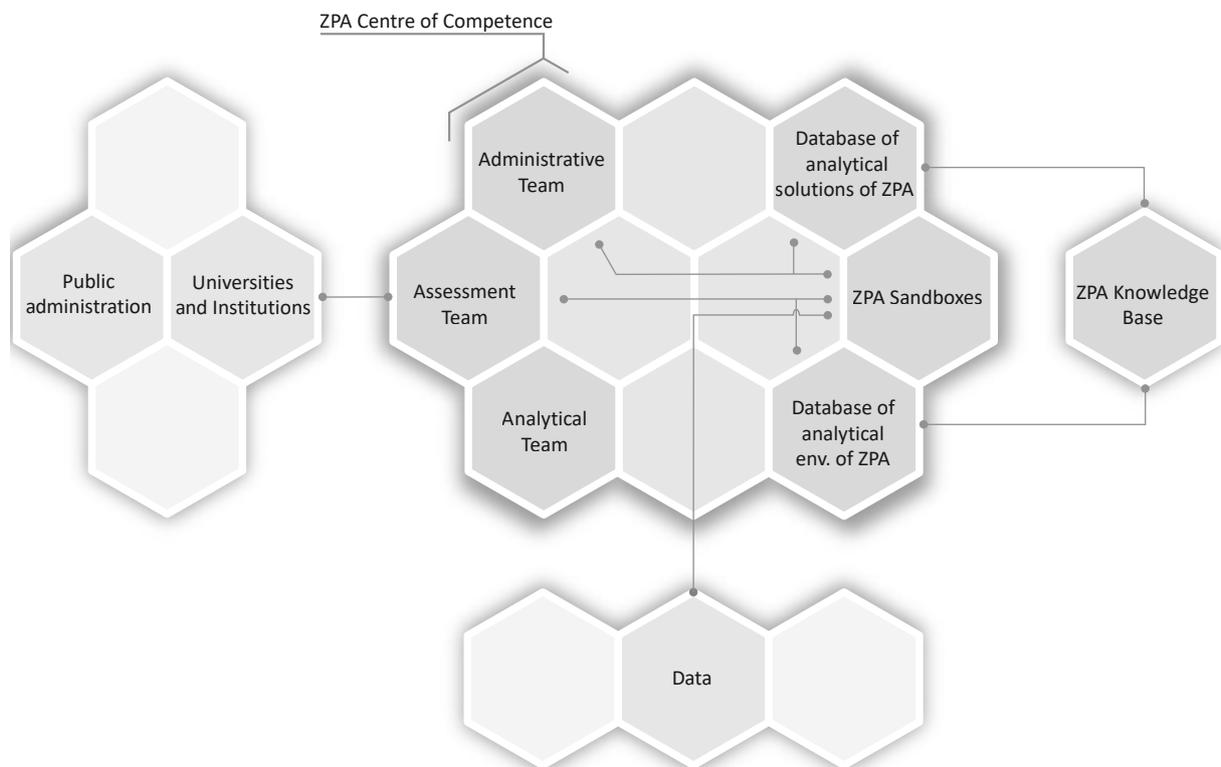
The detailed scope of data covered by the export will also be included in agreements concluded by the minister responsible for digitization with the particular data hosts. The agreements will provide for the purpose of analysis, scope of data, and public registers containing data needed for specific analyses.

The ZPA provisions also define the obligations of the minister responsible for computerisation with regard to ensuring security and integrity of data collected in the ZPA.

In order to ensure sustainability of the ZPA platform, organisational solutions need to be developed, including for identification of institutions involved in the development and operation of the ZPA. To ensure continued operation of the platform, a mechanism needs to be set up that will guarantee the development of analytical concepts for future studies carried out under the ZPA, development of a methodology for these studies, as well as ensuring stable cooperation between the academia and public administration.

The ZPA Centre of Competence (CK) will serve this purpose. According to the existing arrangements, the CK is to be a platform for substantive communication between the public administration and academia. In order to achieve this purpose, the CK should gather information on analytical centres ready to cooperate in projects carried out using the ZPA, as well as analytical concepts that can potentially be used in this type of analyses. Knowledge gathered in the CK can also support the process of expressing opinions on the feasibility, substantive value, and technical correctness of studies planned and submitted to the ZPA. Figure 7.6. presents the links between the components of ZPA and CK.

Figure 7.6. Integrated Analytical Platform and Centre of Competence – functional links.



From the perspective of evaluation studies to be conducted through the ZPA, a procedure for initiating and executing new studies is crucial. According to the current project assumptions (as of November 2021), such studies may be initiated by a government administration body (as study owner). Prior to a study's execution, the study initiation application should undergo a feasibility assessment, which may be carried out in the CK. The assessment should take account of the following:

- substantive feasibility of the study using the data stated by the applicant;
- financial and organisational feasibility of the study, based on the schedule and budget proposed by the applicant;
- validity of the study from the point of view of its potential impact on improving the effectiveness of public policies.

On the basis of the assessment, the ZPA decision-maker will decide about the validity of the study, taking into account the planned, ongoing, and completed scientific studies on the improvement of effectiveness of public policies. A solution like that is in line with the practices applied in other countries (e.g., Sweden).

It is also worth noting that the planned CK should disseminate information on the ZPA among the academia and public administration. This can be done through activities like:

- seminars and workshops involving representatives of public administration and academia, dedicated to the discussion of the available ZPA study results and development of new analytical concepts using administrative data;
- analysis of the potential and quality of data in the state information resources (including the coherence of administrative registers of different hosts);
- public debates on the possibilities of and limitations to using administrative data to conduct analyses and develop recommendations for state governance, as well as for scientific studies;
- technical assistance and training in the use of state information resources as a tool supporting the improvement of public policies (including their evaluation), for representatives of both public administration and academia;

Another important element in operation of the ZPA is technical infrastructure. The ZPA platform will be a closed environment where designated analysts will be able to conduct analytical work on anonymised data sets. Study results, in the form of aggregated indicators, can be accessed in various forms. One of the possible forms of publishing the information is

the use of web API. Like Eurostat or Statistics Poland (GUS) data, they will allow the entities registered as ZPA users to download processed, anonymised data or indicators (depending on the study) for further analytical work using statistical software. Such solutions may be particularly useful in the case of cyclical studies. They can improve productivity and convenience for regular users of ZPA data (study owners, analytical communities, and other stakeholders registered as ZPA users).

Summary

The ZPA project is a significant step towards building solutions supporting systemic analyses, monitoring, and evaluation of public policies using data collected in administrative registers. The approach presented in this chapter stems from the authors' experience related to conducting analyses using administrative data, e.g., as part of the preparation and implementation of the editions of the Polish Graduate Tracking System (ELA system). It should be emphasised that the studies carried out in the framework of the ZPA are to provide a basis (pilot) for elaboration of permanent mechanisms to enable the development of public data analytics, including institutionalisation of solutions related to providing access to and use of administrative data from the perspective of data hosts, institutions, and those interested in using them for analytical, monitoring, and evaluation works in the area of public policies.

8. Use of Administrative Data to Create and Evaluate Health-related Policies

Jakub Adamski

*Office of the Commissioner for Patients' Rights,
Faculty of Sociology
University of Warsaw*

Introduction

At the end of the 20th century, a trend extremely important for the future of medicine began – evidence based-medicine, EBM. Over time, the postulate of EBM i.e., the need for constant evaluation and improvement, has become central to the culture of health care. At the same time, its development was largely conditioned by the emergence of new tools and opportunities for data collection and analysis. It is natural that health data has become one of the most important resources in health care, while planning and evaluation of health-related policies are among its primary uses. It is estimated that up to 30% of all data collected worldwide is health data (DigitalEurope, 2021).

This chapter aims to introduce the topic of secondary data usage in health care. The particular sections address the topic of health data (their types and sources) in a broader perspective, present the beginnings of the use of health data in evaluations in Poland (on the example of health needs maps as tools, data, and analyses developed for the purpose of creating and evaluating health policy), as well as discussing the main, publicly available, sources of health data in Poland and the potential use of health data in public policies and their evaluation.

8.1. Health data – an international perspective

Several types of global health data sources are commonly referred to (National Library of Medicine, 2021). These are distinguished by their origin and collection method, as well as their availability and use. They include:

- Medical records, or electronic medical records, containing information on the patient's health conditions (diagnoses) and the services provided (examinations, treatments, etc.), serving to document the treatment process, but also to evaluate its effectiveness. This data is considered sensitive – as a rule, only patients, persons authorised by them, and persons providing them with medical services have access to such data,
- Claims data collected and processed for the purpose of accounting for services provided by a payer (e.g., the National Health Fund). This data includes detailed information on the services provided. It is often questioned, as its reliability is susceptible to fraudulent practices applied in order to obtain higher payments for the provided services,
- Vital records are administrative data collected by administrative authorities and contain basic information on citizens (e.g., births, deaths, marriages, etc.). They are a confirmation of the most important facts about a person,
- Public health surveillance data are data collected and analysed to observe and respond to changes in population's health (e.g., prevalence of specific infectious diseases). These data is present in medical registers, where detailed data on selected diseases is collected to determine their prevalence in the population, as well as studying the risk factors and treatment effectiveness,
- Other data, in particular results of clinical trials or representative surveys.

Health data analysis encounters a number of challenges related to data quality (e.g., in the case of claims data or research bias), data completeness (e.g., in the case of surveys or public health surveillance), data fragmentation (where the actors only have access to part of the data collected), and ensuring data security due to the sensitivity of the matter (Vayena, Dzenowagis, Brownstein i Sheikh, 2017).

In recent years, big data analysis has become extremely useful, and also fashionable, enabling formulation of far-reaching conclusions from data collected on large populations. So far, these analyses have made it possible to, among other things, identify patients with various diseases (in order to ensure appropriate prophylaxis), identify abuses, support the management of the health care system and medical entities (Durcevic, 2020). Big data

analysis is used in particular to gather scientific evidence for the safety and efficacy of treatments (*real world evidence, RWE*) (Maissenhaelter, Woolmore, Schlag, 2018).

Recently, analysis of small data, i.e., data collected directly at a given medical facility for the purpose of improving process efficiency or reducing operating costs, has also gained popularity (Miller, 2020).

The collection and further processing of health data has become one of the key areas of health system management. It is an area of intense activity for all the countries trying to improve the quality of health care and system efficiency (OECD, 2015).

Studies conducted by the OECD provide a good overview of the range of key data collected by countries (such as data on primary health care, hospital treatment, causes of death, but also data on prescriptions, cancer or cardiovascular disease registries), as well as the challenges they face. These challenges include ensuring a high level of coverage of the population with the available data, automation of data collection and processing, validity (which proved crucial, for example, during the COVID-19 pandemic), and the possibility of combining different data sets. Another challenge is the practical use of the data, both in the health care system and e.g., for scientific purposes (Oderkirk, 2021).

8.2. The role of the health needs map in the creation and evaluation of health policy

Historical background

Since the very beginning, the map of health needs, as a tool of health policy, has been directly related to the evaluation of EU funded projects. Its creation was driven by the negative evaluation of how EU funds were spent on health in the 2007-2013 financial perspective. At the same time, it was created in order to provide an appropriate tool for the evaluation of health-related projects. Its historical background points to the measures and policies for whose creation and evaluation the map of health needs was originally developed.

In the previous programming periods, the spending of EU funds involved numerous difficulties – from the perspective of each participant in the process. In the first programming period (2004-2006), the main problem was the lack of experience and difficulties in purposeful allocation of funds and proper prioritisation of their use to maximise the benefits for the public (Świstak, 2016, p. 53). In 2007-2013, the situation improved, but there were very frequent situations where demand for certain solutions was created only because European funds were available. There was a tendency to raise funds for ‘whatever works’ and provide the necessary impetus in the right direction only to a limited extent. Therefore, prevention of financing numerous isolated, often small-scale projects ended in failure. They received funds, but generated hardly any sustainable competitive capabilities (Świstak, 2016, p. 53).

The largest number of examples of inexpedient project selection and implementation are provided by the results of two audits carried out by the Supreme Audit Office (NIK): “Purchase and use of medical equipment co-financed from the funds of the Regional Operational Programmes”⁷⁷ and “Activity of local government hospitals transformed into capital companies”.⁷⁸ In the first of the audits, NIK revealed that in nearly one in three (30.3%) of the audited entities, one year after completion of project implementation, the planned outcome indicators i.e., the number of specialist medical examinations/treatments carried out with the use of the purchased equipment, and the number of new jobs created had not been achieved (NIK, 2012a, p. 29). In the opinion of NIK, this was the result of decisions on granting subsidies being made without analysing the health needs or assessing the availability of services provided with the use of this type of equipment. Another reason was failure to ensure appropriate coordination of activities related to the equipment’s purchase. The situation was also influenced by irregularities found in hospitals:

- failure to provide appropriate conditions for the provision of services – in particular, lack of appropriate specialists,
- delays in preparation of rooms in diagnostic laboratories,
- lack of previously held or approved contract with the National Health Fund (NFZ) for the relevant services, and limits on services provided for in the concluded contracts.

⁷⁷ Carried out in 2012 for the period from January 2009 to October 2012., KZD-4101-03/2012, no. 3/2013/P12123/KZD.

⁷⁸ Completed in March 2015 for the period 2011 to 2014, KZD-4101-004/2014, no. 196/2014/P/14/061/KZD.

As a result of the investments, some entities even experienced a deterioration in their financial performance (NIK, 2012a, p. 10).

NIK pointed out that the institutions managing the operational programmes under which these activities were financed did not have sufficient knowledge on the distribution of medical devices and equipment for entities carrying out medical activity, which resulted in failure to ensure equal access to specialist medical services to the residents of the Podlaskie voivodship under analysis. For example:

- by the end of 2011, MRI scanners were not installed in 13 out of 15 counties,
- out of seven mammographs used in the voivodship at the end of 2011, six were located in Białystok and one in Łomża; as a result, residents of the most remote locations had to travel more than 150 km to the nearest device,
- in the county of Suwałki (105 thousand residents), there was one echocardiography machine at the end of 2011, while in Białystok (295 thousand residents), 31 machines were available, and the voivodship's average was 4.9 machines per 100 thousand residents,
- in the county of Siemiatycze (46.9 thousand residents), there was no X-ray machine with visual track, while in the voivodship there were 46 such machines at the end of 2011 (NIK, 2012a, p. 19).

These phenomena led to making the spending of European funds in the next perspective (2014-2020) dependent on the introduction of a system of mapping the health needs in Poland. The system was to create national and regional coordination frameworks aiming to increase access to high-quality health services and stimulate efficiency in the healthcare sector, as well as monitoring and reviewing of health services (NIK, 2012b, p. 41). The European Commission introduced these requirements as part of the so-called ex ante conditionality (condition 9.3 of Annex 9 to the Regulation of the European Parliament and of the Council EU No 1303/2013 laying down common provisions for the European Funds).

Role in the health care system

In their original form, maps of health needs did not meet the expectations. In the results of the audit "The creation of maps of health needs", NIK stated that "maps of health needs have not yet become an important tool enabling the reconstruction of the health care system

and adjustment of the system to the health needs of the population”, referring to, inter alia, limited influence on the establishment of priorities for the regional health policy, contracting of services, the so-called network of hospitals, and evaluation of investment applications (NIK, 2012a, p. 10).

This situation underwent a significant formal change in July 2021, with the entry into force of the amendment to the Act on Health Care Services Financed from Public Funds (Act, 2021). It introduced changes to enable planning and implementation of actions under the so-called transformation plans. They provide main directions and objectives for actions in the health care system in Poland and on the territory of the particular voivodships, based on objectivised indications resulting from the map of health needs. This will serve to define the priorities, making it possible to select the tasks which will, first of all, bring greatest health benefit to the public. In particular, they are to focus on the areas of Poland where the currently achieved health results are furthest from those achievable and achieved in the other countries, including the European Union (Explanatory, 2021).

From 1 January 2022, the **National Transformation Plan** developed by the Minister of Health, and voivodship plans proposed by the voivodes and voivodship health needs councils, will be in force. These documents, prepared on the basis of the map of health needs, will reveal:

- health needs, and challenges to the organisation of the health care system that require coordinated action at supra-regional level,
- activities that require coordination at supra-regional/regional level,
- planned year(s) in which the actions will be implemented,
- bodies responsible for the implementation of the actions,
- estimated costs of the actions,
- expected results of the actions,
- indicators for the implementation of the particular actions, including those concerning the provision of health care services.

The map of health needs, together with transformation plans, will also be the basis for drawing up plans for the President of the National Health Fund’s purchase of services. In its current form, the map of health needs, along with transformation plans, can, in the hands of government administration, become an effective mechanism for planning and implementing changes in the health care system, including, among others, by ensuring that ideas for reforms are reflected in programme documents at the national and local level, which have a

direct bearing on the reality of financing health care services from public funds. At the same time, the adopted programme based on the consensus of the main stakeholders, including local governments, has a chance of ensuring sustainability of the designed solutions.

The above also shows for what type of measures the health needs map has been developed and evaluated. From the perspective of evaluation of other public policies, also worth considering are the indicators adopted in the transformation plans: as a source of contextual information about projects implemented or planned in this area. Due to the links between the documents, data necessary for the measurement of a given indicator will often be available in the map of health needs, or the related analysis database.

Current scope of data and analysis

Following the 2021 change in the regulations governing the map of health needs (Act, 2021), the newly published report based on the map (for 2022-2026) includes:

- demographic and epidemiological analyses, analyses of the state and use of resources, including medical personnel;
- challenges of the health care system and recommended courses of action, for the whole country and for the particular voivodships.

In comparison to the editions developed under the previous legislation, the current map of health needs significantly broadens the scope of analyses covered. It includes:

- Demography, including population status and structure, vital statistics, life expectancy and population projections,
- Epidemiology and epidemiological forecasts,
- Key risk factors, actions implemented to improve, promote and protect public health, as well as prophylaxis programmes and preventive medical examinations,
- Most important types of services: primary health care, outpatient specialist care (including out-of-pocket services), hospital treatment (including staff, hospital bed occupancy, and quality of care), psychiatric care and addiction treatment, medical rehabilitation, long-term care, palliative and hospice care, and medical emergency services,
- Personnel, including doctors, dentists, nurses, midwives, and physiotherapists
- Medical equipment, its age, use, and availability.

The map of health needs presents analyses which make it possible to look at health needs in a broader context. The document presents not only different perspectives of health needs assessment, but also – on the basis of surveys and public statistics – provides analyses of citizens' satisfaction with health care in Poland and analyses of the level of use of health care services not financed from public funds.

Another important novelty is direct determination of the challenges faced by the health care system in the particular areas, and formulation of recommended courses of action.

As indicated by the Ministry of Health, the document aims to help in planning investments and optimal spending of public funds in health care. The analyses and recommended courses of action are to be used by institutions to prepare strategic documents at the operational level, such as the National Recovery and Resilience Plan, national and voivodship transformation plans, and operational programmes under the EU cohesion policy. The map was also used in the development of a major directional strategy document for the health sector, "Healthy Future", which is due to be published in autumn 2021.⁷⁹

Health needs assessment

Assessing health needs and the level of their fulfilment is crucial for planning and evaluation of public policies. Properly assessed health needs can provide a measure of desirability of any intervention. At the same time, properly selected indicators enable the assessment of the current situation, and whether the desired change has been achieved.

It is generally agreed that one of three approaches can be adopted to assess health needs (Wright, Williams, Wilkinson, 1998):

- Supply, defined as health care services provided – the supply of services depends on numerous factors other than the objective health needs: availability of effective medical technologies, priorities for financing services, available financial resources, tools influencing the choice of services provided by medical entities (e.g., financing limits, incentive allowances),

⁷⁹ <http://mpz.mz.gov.pl/2021/09/01/mapa-potrzeb-zdrowotnych-w-nowej-formule-opublikowana/>, accessed October 14, 2021.

- Demand, defined as services requested by patients, regardless of advisability of their provision (i.e., possibility of their positive impact on health) and possibility of their provision (e.g., limited resources for financing health care services) – in this sense, demand depends primarily on patient awareness and preferences;
- Need, understood as a situation where the provision of a service can bring clinical benefit to the patient – the fulfilment of this need is possible only if appropriate (effective) medical technology exists.

Figure 8.1. Examples of health needs in terms of potential benefits, demand, and supply of services (Topór-Mądry, 2017)



Historically, health needs in Poland were assessed from the supply perspective – services financed from public funds, in particular from the NFZ budget. This approach was largely dictated by practical considerations, i.e., the availability of data. It also involved significant limitations, e.g., did not take into account whether a given type of service was really necessary, or even whether the service paid for was actually provided.

The current map of health needs also draws on international scientific output in the area of health needs assessment, presenting objectivised health needs, as diagnosed at the national and individual voivodship level. For this purpose, indicators developed by the Global Burden of Diseases (GBD) project currently run by the University of Washington in Seattle have been used.

The use of disability-adjusted life years (DALYs), years of life lost (YLLs), and years lived with disability (YLDs)⁸⁰ enables the best possible representation of the burden of disease resulting from particular health problems in a population⁸¹. Where 1 (one) DALY means one year of life lost in full health, regardless of whether due to death or disability. This approach makes it possible to compare, inter alia, diseases which result in premature death (e.g., cancer, myocardial infarction) and those which cause disability and may significantly reduce the quality of life (e.g., cataract, depression, alcohol abuse).

Full comparative data for the assessment of the situation in Poland in comparison to other countries and regions is available on the GBD platform⁸². In 2020, the Lancet journal published the latest updates to its systematic review of disease and injury outcomes and risk factors (Lancet, 2020a, 2020b).

8.3. Principal health data sets in Poland

Collection and advanced analysis of health data in Poland is currently developing extremely rapidly. In the first stage (in 2014-2016), it was necessary to review and assess the available data, and start analysing it. The main role was played by the aforementioned maps of health needs, implemented as part of two EU-funded projects.⁸³

Further development of these works is associated primarily with the increased digitalisation of the health care system and introduction of services using ICT systems (in particular, introduction of the e-prescription, e-referral, and e-sick leave). We are currently on the threshold of another revolution, related to the implementation of electronic medical records

⁸⁰ <https://www.aotm.gov.pl/projekty-aotmit/gbd/gbd-2019/>, accessed October 14, 2021.

⁸¹ <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/158>.

⁸² <https://vizhub.healthdata.org/gbd-compare/>.

⁸³ Projects:

- a) *Improvement of the quality of healthcare management through support for the development of regional health maps as a tool for improving the quality of healthcare management – training of estimation of health needs* – co-financed by the European Union from the funds of the European Social Fund under the Operational Programme Human Capital. The project was carried out in the 2014-2015 period.
- b) *Maps of Healthcare Needs – Database of Systemic and Implementation Analyses* – co-financed by the European Union from the European Social Fund under the Operational Programme Knowledge Education Development. The project has been running since 2015 and is currently scheduled to run until the end of 2022.

and collection of data from them on one central platform (P1 – Electronic Platform for Collection, Analysis, and Sharing of Digital Medical Records, 2017).

Most of the collected health data, including administrative data, is not publicly available. This section presents the most relevant public data sets, which can also be used in the evaluation of other policies. Data collected in the said sets covers the whole territory of Poland and is available for further use free of charge.

HEALTHY DATA (Healthy Data, 2021)

Since 2019, the National Health Fund has been running a website entitled **Healthy Data**, where it publishes various types of data, statements, and reports. It comprises the following sections:

- Facility comparison – section with indicators for the comparison of healthcare providers (e.g., in oncology, in terms of the number of treatments and average length of hospitalisation),
- Monitoring – section that provides reports on the various phenomena occurring in the health care system while performing the functions of a public payer (e.g., in areas like diagnostics, childhood diseases),
- Reports – section presenting a variety of topics: one-off (e.g., tobacco-related diseases), or cyclical (e.g., hospitalizations according to the Diagnosis-Related Patient Group, Primary Health Care activities),
- Summary data – section where selected data prepared at the request of other institutions is presented (National Health Fund, 2019a)

It should be noted that apart from the Healthy Data service, the NFZ also provides other important data on its website – e.g., about contracts concluded with healthcare service providers (National Health Fund, 2019b) and available treatment dates (National Health Fund, 2019c).

ProfiBaza⁸⁴ and EpiBaza⁸⁵

The ProfiBaza and EpiBaza are two platforms run by the National Institute of Public Health of the National Institute of Hygiene – National Research Institute.

The first platform presents selected aspects of disease prevention and health promotion, health programmes, and health policy of local government units and the National Health Programme, as well as the health status of the Polish population. It aims to assess what public health initiatives are being implemented, whether they meet the health needs of local communities, how much they cost, and what their effects are (Project of the National Institute of Public Health – National Institute of Hygiene, 2019).

The second platform presents data collected in the communicable diseases and food safety monitoring system, such as the number of cases and suspected cases of influenza and other communicable diseases (by voivodship, in the particular years, providing an overview of the effectiveness of the preventive actions taken).⁸⁶

Polish National Cancer Registry⁸⁷

The Polish National Cancer Registry (KRN) is a registry where data on cancer incidence and cancer-related deaths is collected and used to assess the epidemiological risk of cancer in Poland. The KRN presents trends in incidence, geographical differentiation, and frequency of particular cancers. The data is available by county and year. The target platform, which is currently being developed, is to collect data not only within the basic cancer registry, but also oncological organ registries, integrated with databases of health care service providers that treat oncological diseases (MedExpress, 2019).

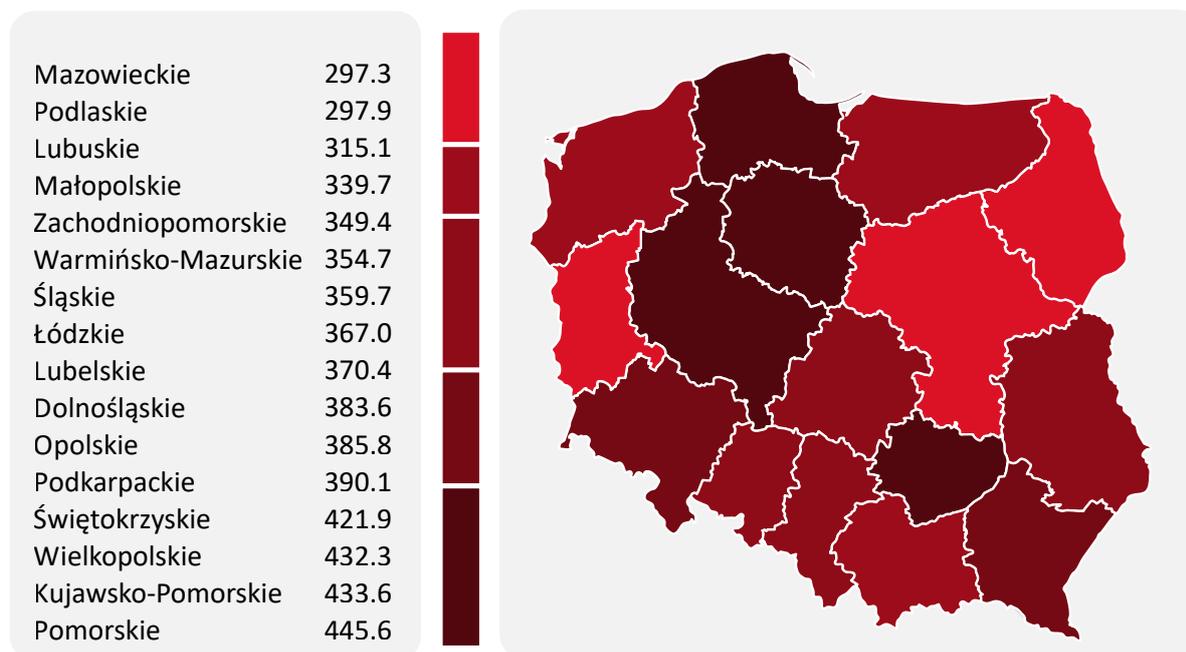
⁸⁴ <https://profibaza.pzh.gov.pl/>, accessed October 15, 2021.

⁸⁵ National Institute of Public Health. <https://epibaza.pzh.gov.pl/>, accessed October 15, 2021.

⁸⁶ <https://epibaza.pzh.gov.pl/story/sprawozdanie-z-konferencji-podsumowuj%C4%85cej-realizacj%C4%99-projektu-epibaza>, accessed October 15, 2021.

⁸⁷ Polish National Cancer Registry. <http://onkologia.org.pl/>, accessed October 15, 2021

Figure 8.2. KRN sample report – standardised⁸⁸ incidence of cancer in men in 2018 by voivodship⁸⁹



System and Implementation Analysis Database (Disability-adjusted life years, DALYs, 2021)

The System and Implementation Analysis Database is an original database developed by the Ministry of Health for the purpose of presenting the results of analyses carried out in the Ministry on the occasion of and for the purpose of creating the editions of maps of health needs.

In addition to the Healthy Data website, this database contains the widest range of analyses, including in particular data from the National Health Fund (NFZ), unique analyses of equipment resources and medical personnel, as well as an objectivised analysis of health needs and priorities at national and voivodship levels.

⁸⁸ Standardised rate defines e.g., morbidity in a given population, assuming that its age structure would correspond to the age structure of the population selected as the standard (e.g., a voivodship with the population age structure corresponding to the population of the whole Poland, Europe, or the world).

⁸⁹ Report generated at <http://onkologia.org.pl/raporty/>, accessed October 15, 2021.

This database originally reflected the data used in the first maps of health needs, including demographic and epidemiological analyses, analyses of the status and use of resources, and forecasts of health needs (in terms of demographic forecasts and the supply of publicly funded services, hospital beds or inpatient days). In principle, the scope of these analyses was defined by the Regulation of the Minister of Health of 26 March 2015 on the Scope of Content of the Health Needs Maps (Regulation, 2015). At the same time, in-depth analyses were published, developed within selected disease groups such as oncology, cardiology, and others. In June 2021, data for 2019 was published for the last time under the old methodology.⁹⁰

Currently, the platform also includes data and analyses collected for the latest edition of the health needs map of August 2021.⁹¹

Apart from the above analyses – structured and comparable for the 2014–2019 period – other thematic analyses are available as well, covering selected problems identified in the health care system (e.g., analysis of the functioning of the emergency medical service, and analyses aiming to identify systemic problems among selected conditions, such as diabetes, chronic obstructive pulmonary disease, and others). These analyses may not be developed or updated in the future, but they may significantly contribute to finding solutions to the challenges affecting the health care system.

8.4. Health data and evaluation of public policies

Health care

In recent years, thanks largely to the use of European funds, evaluation of public policies and projects aiming to implement changes in the health care system has come a long way. We have moved from total reliance on eminence-based medicine (BMJ, 1999), which, in evidence-based medicine, is considered to be of lowest quality (Burns, P. et al., 2011),

⁹⁰ <http://mpz.mz.gov.pl/2021/06/01/mapy-potrzeb-zdrowotnych-dane-za-2019-rok-dostepne-na-platformie-internetowej/>, accessed October 15, 2021.

⁹¹ Published on 27 August 2021 by an announcement of the Minister of Health, Official Journal of the Ministry of Health, item 69.

through the use of unscientific and unverifiable assessment criteria⁹², to the first examples of multi-criteria decision-making systems (Instrument for the Assessment of Investment Applications in Health, 2016) and big data analysis. Another step into the future will be modelling the effects of designed public policies – some countries have already taken this step (e.g., modelling the effects of making hearing aids available to people with hearing loss in the EVOTION project) (Brdarić et al., 2020).

The approach based on analyses of results of observational studies, in particular analyses of big data, despite the impossibility of determining the causality of phenomena, will become standard due to the fact that introduction of experimental (randomised) studies in public policies is not practically possible (Scott et al., 2020). Use of such data in the evaluation of public policies is presented in documents of different institutions and organisations (US Department of Health and Human Services, 2014), as well as becoming the subject of increasingly popular monographs and other publications (Crato and Paruolo, 2019).

Health in all policies

The slogan “health in all policies” has become popular in recent years (World Health Organization, 2013). Its main elements are as follows: health as a human right, searching for and counteracting the effects of social determinants of health inequalities (resulting from the distribution of money, power, resources, birth, as well as living and working conditions)⁹³, acting on the assumption that citizens’ health is not the sole responsibility of the Ministry of Health (it is estimated that air pollution may be a significant risk factor for up to one in eight deaths worldwide)⁹⁴, promoting intersectoral cooperation in policy-making and participation of other stakeholders in policy-making (Cairney et al., 2021). However, implementation of this slogan encounters serious difficulties, in spite of the impact of the COVID-19 pandemic, which brought hope for faster absorption of such ideas (e.g., given its particularly visible impact on the education system) (Bucciardini et al., 2020).

⁹² In 2015, in the Operational Programme Infrastructure and Environment, one of the criteria was the impact of the project on reducing the risk of hospital-acquired infections in wards or other organisational units of the hospital, where the maximum score was awarded if the applicant indicated that the project would reduce the risk of hospital-acquired infections by 80%.

⁹³ <https://www.who.int/teams/social-determinants-of-health>, accessed: October 15, 2021.

⁹⁴ https://www.who.int/social_determinants/publications/health-policies-manual/key-messages-en.pdf, accessed October 15, 2021.

Undoubtedly, an advantage of the emergence and attempts to implement the “health in all policies” postulate is the growing volume of publications and scientific evidence from the intersection of different sectors, in particular health, work, and education. These can serve as a model for selecting indicators and data for evaluation of the proposed and implemented public policies. For instance, international comparative studies have shown that in Poland differences in education have significantly greater impact on men’s health and life expectancy than in many other countries (Mackenbach et al., 2008). However, it is not only health data and the visible positive effects of interventions – in e.g., the labour market – on mental health⁹⁵ that should be the basis for interventions in other sectors. The effects visible in other sectors should be an important argument for specific interventions in the health care system as well. This is perfectly illustrated by studies linking children’s health to its impact on the labour market or future wealth (Smith, 2009).

To propose public policies based on data, three questions must be answered (Fay, 1996):

- What effect (in different areas, including health) will the actions have?
- Is this effect big enough to outweigh the associated costs (including social costs)?
- Is this the best way to spend the money available? What could be alternative projects/policies?

The future of health data and its use

For many years to come, the main barrier to using health data for creating and evaluating public policies will be the limited scope of the data available. However, there is no doubt that the trend of using health data to assess the impact of population’s health on the other sectors (in particular work and innovation) will continue, as actions characterized by the highest possible efficiency and bringing biggest benefits to the public are needed.⁹⁶ This will require collection of more and more administrative data and linking them together.

Some administrative data sets are still awaiting potential public release (ensuring no publication of personal data, or potential pseudonymisation, so as to prevent linking them together). Making the data sets available can significantly help in understanding the health care system, as well as identifying potential and actual impacts of actions and policies. Such

⁹⁵ <https://www.instituteofhealthequity.org/file-manager/economic-active-labour-market-full-report.pdf>.

⁹⁶ https://www.oecd.org/els/emp/Impact_evaluation_of_LMP.pdf.

sets include data collected in the course of provision of e-services in healthcare (such as electronic prescriptions), and data of Statistics Poland. The latter, although they do not relate directly to individuals, cannot be published by official statistics services in a way that would make it possible to link them to a specific entity.

At the same time, apart from administrative data sets already collected and available but not published, it is worth pointing to those that are in the process of being created, or are at an early stage of development. This applies in particular to clinical (aggregated) data collected in electronic medical records, medical registers, and the future register of adverse events (referred to in the draft Act on Quality in Healthcare and Patient Safety) (Draft Act on Quality in Healthcare and Patient Safety, 2021), as well as the results of clinical trials, expected to be conducted in Poland in increasing numbers as the activity of the Medical Research Agency grows.

In the future, in addition to being used for the improvement and development of the health care system, this data may contribute to increasing Poland's role in yet another area – science and innovation.

References

- Act. (2021) [Ustawa z dnia 20 maja 2021 r. o zmianie ustawy o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych oraz niektórych innych ustaw, Dz. U. poz. 1292.] <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20210001292>
- BMJ. (1999). Seven alternatives to evidence based medicine. 319(7225), 1618.
- Brdarić, D., Samardžić, S., Huskić, I., Dritsakis, G., Sessa, J., Śliwińska-Kowalska, M., Pawlaczyk-Łuszczczyńska, M., Basdekis, I. & Spanoudakis, G. (2020). A Data-informed Public Health Policy-Makers Platform, *Int. J. Environ. Res. Public Health*, 17(3271). <https://doi.org/10.3390/ijerph17093271>
- Bucciardini, R., Contoli, B., De Castro, P., Donfrancesco, C., Falzano, L., Ferrelli, R., Giammarioli, A., Mattioli, B., Medda, E., Minardi, V., Minelli, G., Palmieri, L., Pasetto, R., Pizzi, E., Rossi S. & A. Venerosi, A. (2020). The health equity in all policies (HEiAP) approach before and beyond the Covid-19 pandemic in the Italian context. *Int J Equity Health* 19, 92. <https://doi.org/10.1186/s12939-020-01209-0>
- Burns, P., Rohrich, R. & Chung, K. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and reconstructive surgery*. Vol. 128,1: 305-310. <https://doi.org/10.1097/PRS.0b013e318219c171>
- Cairney, P., St Denny, E. & Mitchell, H. (2021). The future of public health policymaking after COVID-19: a qualitative systematic review of lessons from Health in All Policies [version 2; peer review: 2 approved]. *Open Research Europe*. 1(23) <https://doi.org/10.12688/openreseurope.13178.2>

- Crato, N. & Paruolo, P. (2019). *Data-driven Policy Impact Evaluation*, Springer, ISBN 978-3-319-78460-1, JRC107550, DOI <https://doi.org/10.1007/978-3-319-78461-8>
- DigitalEurope. (2021). DIGITALEUROPE's recommendations for the European Health Data Space. https://www.digitaleurope.org/resources/digitaleuropes-recommendations-for-the-european-health-data-space/#_ftn2
- Disability-adjusted life years (DALYs). (2021). <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/158>, access: 15.10.2021
- Draft Act on Quality in Healthcare and Patient Safety [Projekt ustawy o jakości w opiece zdrowotnej i bezpieczeństwie pacjenta] (2021). <https://legislacja.rcl.gov.pl/projekt/12349305/katalog/12804503#12804503>
- Durcevic, S. (2020). 18 Examples Of Big Data Analytics In Healthcare That Can Save People. Datapine. <https://www.datapine.com/blog/big-data-examples-in-healthcare/> access: 14 października 2021.
- Explanatory. (2021) [Uzasadnienie do projektu ustawy o zmianie ustawy o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych oraz niektórych innych ustaw], <https://sejm.gov.pl/Sejm9.nsf/druk.xsp?nr=1013>, access: 14.10.2021.
- Fay, R. (1996), *Enhancing the Effectiveness of Active Labour Market Policies: Evidence from Programme Evaluations in OECD Countries*. OECD Labour Market and Social Policy Occasional Papers, No. 18, OECD Publishing. https://read.oecd-ilibrary.org/social-issues-migration-health/enhancing-the-effectiveness-of-active-labour-market-policies_560806166428#page1
- Healthy Data [Zdrowe Dane]. (2021). <https://ezdrowie.gov.pl/portal/home/zdrowe-dane>, access: 15.10.2021.
- Lancet (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. 396, 1223–1249.
- Lancet. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. 396, 1204–1222.
- Mackenbach, J., Stirbu, I., Roskam, A.-J., Schaap, M., Menvielle, G., Leinsalu, M. & Kunst, A. (2008). Socioeconomic inequalities in health in 22 European countries. *New England J Med.* 358(23): 246881.
- Maissenhaelter BE, Woolmore AL, Schlag PM. (2018). Real-world evidence research based on big data: Motivation-challenges-success factors. *Onkologie (Berl)*. 2018;24(Suppl 2), 91-98. <https://doi.org/10.1007/s00761-018-0358-3>. Epub 2018 Jun 7. PMID: 30464373; PMCID: PMC6224010.
- MedExpress. (2019). Powstaje e-KRN+, czyli innowacyjna platforma gromadząca dane o chorobach nowotworowych. <https://www.medexpress.pl/powstaje-e-krn-czyli-innowacyjna-platforma-gromadzaca-dane-o-chorobach-nowotworowych/74950>, access: 15.10.2021.
- Miller, L. (2020). Big Data in Healthcare Isn't Enough – Small Data Is The Answer. *VIE Healthcare*. <https://viehealthcare.com/big-data-isnt-enough-small-data-is-the-answer/>, dostęp: 14.10.2021.
- National Health Fund [Narodowy Fundusz Zdrowia] (2019a). <https://www.nfz.gov.pl/aktualnosci/aktualnosci-centrali/zdrowe-dane-nowy-portal-informacyjny-nfz,7393.html>, access: 15.10.2021.
- National Health Fund [Narodowy Fundusz Zdrowia] (2019b). <https://aplikacje.nfz.gov.pl/umowy/Provider/Search>, access: 15.10.2021.

- National Health Fund [Narodowy Fundusz Zdrowia] (2019c). <https://terminyleczenia.nfz.gov.pl/>, access: 15.10.2021.
- National Institute of Public Health [Narodowy Instytut Zdrowia Publicznego]. <https://epibaza.pzh.gov.pl/>, dostęp: 15.10.2021.
- National Library of Medicine. (2021). Finding and Using Health Statistics, https://www.nlm.nih.gov/nichsr/stats_tutorial/section3/index.html
- NIK (Supreme Audit Office) [Najwyższa Izba Kontroli], Health Department. (2012a). Informacja o wynikach kontroli KZD-4101-03/2012.
- NIK (Supreme Audit Office) [Najwyższa Izba Kontroli], Health Department. (2012b). Informacja o wynikach kontroli „Tworzenie map potrzeb zdrowotnych” z lutego 2018 r., KZD.430.008.2017 (nr ewid. 191/2017/P/17/059/KZD).
- Oderkirk, J. (2021). Survey results: National health data infrastructure and governance, OECD Health Working Papers, 127, <https://doi.org/10.1787/55d24b5d-en>, access: 14.10.2021.
- OECD. (2015). Health Data Governance: Privacy, Monitoring and Research – Policy Brief. <https://www.oecd.org/health/health-systems/Health-Data-Governance-Policy-Brief.pdf>, access: 14.10.2021.
- P1 – Electronic Platform for Collection, Analysis and Sharing of Digital Medical Records [P1 – Elektroniczna Platforma Gromadzenia, Analizy i Udostępniania Zasobów Cyfrowych o Zdarzeniach Medycznych]. (2017). <https://www.gov.pl/web/popcwsparcie/p1-elektroniczna-platforma-gromadzenia-analizy-i-udostepniania-zasobow-cyfrowych-o-zdarzeniach-medycznych>, access: 15.10.2021.
- Polish National Cancer Registry (KRN) [Krajowy Rejestr Nowotworów]. <http://onkologia.org.pl/raporty/>, access: 15.10.2021.
- Prezentacja pt. „Zaspokajanie potrzeb zdrowotnych w oparciu o dowody – idea i rzeczywistość” dr Roman Topór-Mądry, dostępna pod adresem: <https://bit.ly/3DFQycd>, access: 14.10.2021.
- Project of the National Institute of Public Health – National Institute of Hygiene [Projekt Narodowego Instytutu Zdrowia Publicznego – Państwowego Zakładu Higieny]. (2019). <https://www.pzh.gov.pl/wp-content/uploads/2019/05/notka-prasowa.pdf>, access: 15.10.2021.
- Regulation of the Minister of Health of 26 March 2015 on the Scope of Content of the Health Needs Maps [Rozporządzenie Ministra Zdrowia z 26 marca 2015 r. w sprawie zakresu treści map potrzeb zdrowotnych] (Dz. U. poz. 458). (2015) – now the act has expired. <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20150000458>
- Regulation No 1303/2013 of the European Parliament and of the Council of 17 December 2013 laying down common provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund, the European Agricultural Fund for Rural Development and the European Maritime and Fisheries Fund and laying down general provisions on the European Regional Development Fund, the European Social Fund, the Cohesion Fund and the European Maritime and Fisheries Fund and repealing Council Regulation (EC) No 1083/2006 , 320–469. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013R1303&from=pl>
- Scott, J., Schwartz, T. & Dimick, J. (2020) Practical Guide to Health Policy Evaluation Using Observational Data. *JAMA Surg.* 155(4), 353–354. <https://doi.org/10.1001/jamasurg.2019.4398>

- Smith, J. (2009). The impact of childhood health on adult labor market outcomes. *Rev Econ Stat.* 91(3): 478–89.
- World Health Organization. (2013). Helsinki Statement on Health in All Policies. 8th Global Conference on Health Promotion.
- Świstak, M. (2016). Fundusze Unii Europejskiej w Polsce: dekada doświadczeń. *Przegląd Europejski.* 3(41).
- US Department of Health and Human Services. (2014). Centers for Disease Control and Prevention. Using Evaluation to Inform CDC's Policy Process. Atlanta, GA: Centers for Disease Control and Prevention.
- Vayena E., Dzenowagis, J., Brownstein, J. & Sheikh, A. (2017). Policy implications of big data in the health sector, *Bulletin of the World Health Organization*; ID: BLT.17.197426.
- Wright, J., Williams, R. & Wilkinson, J. (1998). Development and importance of health needs assessment. *BMJ (Clinical research ed.)* vol. 316,7140. 1310-3. <https://doi.org/10.1136/bmj.316.7140.1310>

Internet sources:

- <http://mpz.mz.gov.pl/2021/06/01/mapy-potrzeb-zdrowotnych-dane-za-2019-rok-dostepne-na-platformie-internetowej/>, dostęp: 15.10.2021.
- <http://mpz.mz.gov.pl/2021/09/01/mapa-potrzeb-zdrowotnych-w-nowej-formule-opublikowana/>, dostęp: 14.10.2021.
- <https://epibaza.pzh.gov.pl/story/sprawozdanie-z-konferencji-podsumowuj%C4%85cej-realizacj%C4%99-projektu-epibaza>, dostęp: 15.10.2021.
- <https://profibaza.pzh.gov.pl/>, dostęp: 15.10.2021.
- <https://vizhub.healthdata.org/gbd-compare/>.
- <https://www.aotm.gov.pl/projekty-aotmit/gbd/gbd-2019/>, dostęp: 14.10.2021.
- <https://www.instituteofhealthequity.org/file-manager/economic-active-labour-market-full-report.pdf>.
- https://www.oecd.org/els/emp/Impact_evaluation_of_LMP.pdf.
- <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/158>
- https://www.who.int/social_determinants/publications/health-policies-manual/key-messages-en.pdf,
dostęp: 15.10.2021.
- <https://www.who.int/teams/social-determinants-of-health>, dostęp: 15.10.2021.

9. Possibilities of using administrative data of the Public Procurement Office for the evaluation of public programmes⁹⁷

Anna Kowalewska

Department of analysis and strategy

Polish Agency for Enterprise Development (PARP)

Introduction

The value of public procurement contracts awarded in 2020 in Poland under the procedures set out in the Act on Public Procurement (PZP) was nearly PLN 184 billion, accounting for 8% of gross domestic product. The position of the public sector on the domestic procurement market is unique, not only because of the value of the purchases made. Public procurement contracts have great potential for achieving goals other than those related to the day-to-day provision of the necessary goods and services to the state. The other goals include creating demand for innovation and environmentally friendly solutions, creating stable jobs, supporting social inclusion, and promoting digitisation. In this way, they become a tool for the country's development policy. Greater inclusion of public procurement in the achievement of strategically important development goals is supported by the new Act on Public Procurement (PZP), in force since 2021. One of the objectives of the new regulation is to facilitate access to public procurement contracts for micro-, small and medium-sized enterprises (SMEs). The Public Procurement Bulletin (BZP), an electronic publication of contract notices and procedure results published by the Public Procurement Office (UZP) provides a comprehensive database of public procurement contracts in Poland. This chapter discusses the scope of information collected in the BZP and points to the directions of its

⁹⁷ All conclusions in this chapter are the opinions of the author. They do not reflect the official position of the Public Procurement Office (UZP) and have not been verified by the Office.

potential use in evaluation, especially of programmes aiming to strengthen SME activity on the public procurement market. The wide range of data available in the BZP also enables its use for the purposes of monitoring and evaluation of other objectives of the state's procurement policy that underlie the new regulation.

The public procurement contract is the main method which public entities use to make purchases for the purpose of their tasks aiming to address social needs. Legal regulations on public procurement contracts in Poland have quite a long history, dating back to the inter-war period⁹⁸. Due to the historical context and the related social and economic circumstances (World War II and the centrally managed, almost entirely state-controlled economy in the post-war period), for a large part of that period, the role of public procurement contracts was only marginal. In 1992, the change that had occurred in the social and economic system at the turn of the 1980s and 1990s, as well as Poland's planned accession to the European Union, prompted works aiming to regulate procurement procedures under the changed socio-economic conditions, taking account of the implications of Poland's future EU membership (Dolecki, 2011, p. 26). Therefore, the Act on Public Procurement of 10 June 1994 (Act, 1994) was passed, and remained in force for 10 years.

As the Polish law had to be aligned with EU requirements, it was replaced by the Act on Public Procurement of 29 January 2004 (PZP of 2004) (Act, 2004). Numerous amendments to the PZP of 2004, the changing economic landscape, as well as the need to implement new EU directives necessitated the drafting of a new law, which was passed on 11 September 2019 (PZP of 2019) (Act, 2019), took effect on 1 January 2021, and is currently in force.

Significance of the public procurement market for the economy

The value of Poland's public procurement market has been growing steadily. Data released by the UZP in its annual *Reports on the functioning of the public procurement system* reveal that in the last 20 years the value of contracts awarded under the Act on Public Procurement has grown almost eight-fold, from PLN 23 billion in 2000 to PLN 183.5 billion in 2020⁹⁹. The share of the value of public procurement in the GDP has also increased, from 3.1% in 2000 to 7.9%

⁹⁸ The first piece of legislation regulating government procurement was the 1933 Act on supplies and works for the State Treasury, Local Government, and Public Law Institutions (Act, 1933).

⁹⁹ In 2018 and 2019, the value of the market was even higher, amounting to approximately PLN 200 billion.

in 2020 (author's calculations based on UZP reports and data from the GUS Macroeconomic Data Bank). In 2010, spending on public procurement under the Public Procurement Law accounted for nearly 12% of the GDP.

If public procurement contracts also included those awarded on the basis of exemptions from the obligation to apply the Act¹⁰⁰, total value of the market in the 2013¹⁰¹-2020 period would range from approximately PLN 200 to 300 billion (UZP, 2013-2020), accounting for approximately 12%-13% of the GDP (author's calculations based on UZP reports and data from Macroeconomic Data Bank).

These amounts illustrate the importance of public procurement for the national economy. The service sector, as well as industrial production related to construction and infrastructure, are largely dependent on the public procurement market (Ministry of Entrepreneurship and Technology [MPiT], UZP, 2018 p. 12).

However, the role of public procurement cannot be seen as merely economic i.e., consisting in rational spending of public funds and stimulating the economy. The objectives assigned to the public procurement system have evolved over time. Initially, the regulations were designed to ensure rational, corruption-free public spending and support the national economy. Over time, it has been recognised that the Act on Public Procurement can support additional economic, environmental, innovation-related, and social policy objectives, secondary to the primary purpose of public procurement i.e., the acquisition of goods and services by the state. This is because awarding public procurement contracts can support participation of entities from the SME sector, foster acquisition of innovative products, minimise the impact on the environment, and address the needs of socially and professionally disadvantaged individuals (Ołdak-Bułańska, 2014, p. 6). In Poland, the principles of the so-called new approach to public procurement were outlined in the government document *New approach to public procurement – public procurement and small and medium enterprises, innovation, and sustainable development*, drafted by the Ministry of Economy and the UZP and adopted by the Council of Ministers on 8 April 2008 (2008, after: Włodarczyk, 2010). The document identifies actions whose implementation will serve

¹⁰⁰ The basis for the exclusions were Articles 4 and 136-138 of the Act on Public Procurement of 2004, and, from 2016, also Articles 4b and 4d of the Act on Public Procurement of 2004.

¹⁰¹ The value of contracts awarded on the basis of exemptions from the obligation to apply the Act has been published by the Public Procurement Office since 2014 (for each preceding year).

to increase the participation of SMEs in the public procurement system and create demand for innovative, environmentally friendly products (goods and services) through public procurement.

The need to steer public procurement towards achieving additional economic policy goals (increased innovation, resource efficiency, support for SMEs) is clearly highlighted in the EU's *Europe 2020 strategy for smart, sustainable and inclusive growth* (European Commission, 2010). The so-called Classical Directive of 2014 (Directive 2014/24/EU), which is one of the basic legal acts regulating public procurement at the EU level, continued to emphasise the need to shape the public procurement system in a way that makes it favour the participation of SMEs in the procedures and enables better use of public procurement in support of common societal goals. The significance of public procurement as a tool for the implementation of economic policies was also recognised in the Communication *Making Public Procurement work in and for Europe*, listing the priorities for the execution of public procurement in the EU (European Commission, 2017). It should be added that one of these priorities is the improvement of the quality of the available data on public procurement.

The above-described view on the role of public procurement in Poland is most fully reflected in the PZP of 2019 (Act, 2019). The explanatory note to the draft act emphasised that public procurement is *a tool for shaping the state's economy, including the pursuit of social, political and strategic goals, and supporting important market sectors. The procedure included in the Act (...) may be the most effective way to support economic and development objectives, and important social and labour market policies.* (MPiT, 2019, pp. 1-3). Accordingly, the PZP of 2019 included provisions on the state's procurement policy (Article 21), which was to indicate the desired directions of contracting authorities' actions with regard to the awarding of contracts as well as taking into account the goals and directions set out in the country's medium-term development strategy. The public procurement policy for 2022-2025 was adopted by the resolution of the Council of Ministers of 11 January 2022. One of its priorities is to develop the capacity of SMEs by taking account of the sector's specific needs and opportunities when awarding public procurement contracts. These measures are intended to increase the activity of the SMEs already present on the public procurement market, and encourage the SME entrepreneurs who have not applied for public procurement contracts so far to start operating on this market (Ministry of Development and Technology [MRiT], 2021, p. 22).

Principle of openness in public procurement

Due to the involvement of public funds, public procurement contracts must be awarded in accordance with certain principles: fair competition, equal treatment of contractors, transparency, and proportionality (Act, 2019, Article 16). Another basic principle is that of the procedure's openness¹⁰². Pursuant to Article 18(1) of the PZP of 2019, public procurement proceedings are open to the public¹⁰³. This means that the activities of the contracting authority and the contractor associated with the participation in the procedure are subject to disclosure at the request of interested parties (e.g., report on the procedure, tenders, requests), made public on the procedure's website (e.g., information from the tender-opening session), and published in official publications (Nowak, Winiarz, 2021, p. 161), depending on the value of the contract, European i.e., the Official Journal of the European Union (OJEU), or Polish i.e., the Public Procurement Bulletin (BZP). Since 2004, the PZP has distinguished between contracts with a value equal to or greater than the EU threshold and those with a value below the EU threshold. EU thresholds are defined by amounts¹⁰⁴ and derive from EU law (directives 2014/24/EU, 2014/25/EU and 2009/81/EC). When awarding contracts whose value is equal to, or exceeds, the EU thresholds, contracting authorities apply procedures standardised for the EU, fully corresponding to the requirements arising from the provisions of EU directives. These procedures are characterised by a higher formal rigour compared to national procedures governing the awarding of contracts with a value below the thresholds. Thus, the PZP identifies the BZP as the official publication to provide information on public procurement contracts below the thresholds. It also specifies the types of information (notices) about the procedures that the contracting authority is obliged to include in the BZP.

Publication of the BZP has been one of the principal competences of the president of the UZP since the office's 1994 establishment. Initially, the BZP was published in paper form and on on-line (first, on the Council of Ministers and later on the UZP's website). Notices to be published in the BZP were sent by contracting authorities on paper forms or, optionally,

¹⁰² All these principles were also present in the PZP of 2004 (Art. 7(1) and Art. 8(1)).

¹⁰³ The contracting authority may limit access to information related to the procurement procedure only in cases specified in the Act, primarily those related to corporate secrecy.

¹⁰⁴ The EU threshold is communicated by the president of the UZP by way of a notice (when the Act of 2004 was in force, it was communicated by the minister responsible for economy and the Prime Minister by way of an ordinance).

electronically. Since the 13 April 2007 amendment of the PZP, notices have been submitted by contracting authorities on interactive electronic forms provided by the UZP, and the BZP has been published on-line only.

During the last 26 years of the public procurement system's operation in Poland, an extensive database has been collected in the BZP. The size of the database is the result of the number of contracts awarded. For instance, between 2014 and 2020, approximately 270,000 notices per year were published in the BZP (UZP, 2014-2020; author's calculations). Most of these were contract notices and contract award notices (approximately 115,000 per year each). The amount of data on awarded contracts, published in the BZP, is substantial as well. Publication obligations apply to various stages of the procedure. Consequently, the database contains information on the procedure applied, the contract, and its contractor.

Along with information provided by contracting authorities in their annual reports on the awarded public contracts, data from the BZP are the basis for regular analyses of public procurement system's functioning, conducted by the UZP. Key findings of the analyses are presented in the annual *Reports on the functioning of the public procurement system*, referred to above. They are a source of vast amounts of valuable information and provide a picture of the market, concentrating mainly – in accordance with the scope of competence of the UZP's president – on the procedural and demand-related aspects. Data held by the UZP, including those collected in the BZP, can also be used for other purposes, including evaluation of public programmes and projects. They may fill the gaps in the knowledge about the public procurement market, especially after the scope of information collected was broadened as of 2021. Besides, they may help evaluate public procurement as a tool for implementing other public policies, to an extent greater than so far.

9.1. Information needs related to the public procurement market

Enterprises' participation in the public procurement market

SMEs' under-representation in the group of public procurement contractors and the need for legal solutions that would facilitate their participation in the market have featured in the debate on the role of public procurement for several years, at both EU and Member State level. Considering SMEs potential for creating jobs, innovation, and economic growth, the need for the sector's increased participation in the public procurement market was explicitly highlighted in the aforementioned Classical Directive of 2014 (Directive 2014/24/EU). The need was also emphasized to monitor the SME performance indicator (understood as the sector's share in the total value of contracts awarded) and analyse the outcomes of national initiatives implemented to increase SME participation in the public procurement market.

In Poland, these proposals were reflected in the way the PZP was drafted to facilitate SMEs' access to public procurement contracts. The *Concept for a new public procurement act* called attention to the fact that SMEs' share in the public procurement market was disproportionate considering its economic significance. SMEs' increased access to this market was considered one of the specific objectives of the new regulation (MPiT, UZP, 2018). Legislative changes were accompanied by training and advisory initiatives targeting SME representatives and contracting authorities.

However, analyses of the effectiveness of the implemented legislative and non-legislative measures were only small-scale. In its 2005-2012 *Reports on the functioning of the public procurement market*, the UZP estimated the number of contractors who were awarded contracts below the EU thresholds. According to the estimates, the number increased from around 25,000 in 2005 to around 82,000 in 2012¹⁰⁵. Monitoring of the outcomes of the new approach to public procurement has been carried out by means of the 2011-2013 surveys conducted by the Polish Agency for Enterprise Development (PARP) on representative

¹⁰⁵ It should be noted that in 2007 the lower threshold for the application of the Act was lowered from EUR 60,000 to EUR 14,000, which contributed to sharp increase in the number of contractors (from 21,000 in 2006 to 50,000 in 2007).

samples of companies from the SME sector (Kowalewska, Szut, 2011, 2012, 2013). The results of these surveys are now outdated. However, at that time, a systematic downward trend was observed in the percentage of companies active in the public procurement market, from 25% in 2011 to 14% in 2013 (Kowalewska, Szut, 2013, p. 39). As shown by the data presented in the next part of this chapter, it is likely that the trend has not been reversed.

Following the recommendation stemming from Directive 2014/24/EU, since 2017, the UZP has been monitoring the share of SMEs in the value of awarded contracts. It has also been presenting data regarding their share in the number of contracts and the tenders submitted. These data show that the percentage of tenders submitted by SMEs in procedures for contracts with values below the EU thresholds oscillated around 80% in 2017-2020 (UZP, 2017-2020). SMEs' share in the value of awarded contracts below the EU thresholds was slightly higher (85%-86% in the analysed period). The picture of the sector was less positive in the case of procurements with values above the EU thresholds. SME tenders accounted for around 65%¹⁰⁶ of all tenders submitted, while their share in the value of contracts awarded ranged from 40% to 49% in 2017-2020 (UZP, 2017-2020). Thus, the share of SMEs in the value of all contracts awarded under the PZP (both below and above the EU thresholds) ranged from 53% to 58% in 2017-2020 (UZP, 2017-2020; author's calculations) and was actually lower than the share of SMEs in gross value added generated by enterprises in Poland (67.7% in 2018), employment (67.4% of the total number of employees in the enterprise sector in 2019), and the total of enterprises (99.8% in 2019) (Skowrońska, Tarnawa, 2021).

This is unlikely to be due to SMEs lower tendering efficiency in procurement procedures as compared to large enterprises. In tenders for contracts below the EU thresholds, the proportion of tenders submitted by SME contractors is almost the same as the proportion of tenders in which an SME contractor's tender is selected as the best. In tenders conducted under the EU procedure (where the value of the contract is equal to or higher than the EU thresholds), SME contractors are slightly more likely to submit tenders than be awarded contracts, but the differences are up to 5 percentage points (UZP, 2019-2020). The difference in activity may therefore be due to SMEs' lower interest in the public procurement market compared to large enterprises.

¹⁰⁶ The data are for the years 2019-2020.

Reports on the functioning of the public procurement system suggest that this interest has further decreased over the last few years (UZP, 2013-2019)¹⁰⁷. This is because the average number of tenders submitted in procedures for contracts below the EU thresholds is decreasing (from about 3 in 2013 to 2.4 in 2019)¹⁰⁸. The same trend was observed in procedures for contracts above the EU thresholds (a decrease from 2.5 to 2.1 in the above-mentioned period). For years, in a significant proportion of procedures (approx. 40%-50%, depending on the year and contract value: below or above EU thresholds), only one tender has been submitted. As regards the percentage of procedures conducted under the EU procedure in which only one tender was submitted Poland ranks last in the EU¹⁰⁹ (European Commission, 2020).

Opinions about the waning attractiveness of the public procurement market were also expressed by the entrepreneurs who participated in the 2018 large-scale qualitative study conducted by PARP (Kowalewska, Głuszyński, 2018). According to the respondents, SME enterprises' low activity on the public procurement market is largely due to the perceived low business profitability of public procurement contracts. This obstacle is also highlighted in the *SME needs analysis* report based on a quantitative survey conducted by the European Commission (2021) on a sample that included Polish companies.

The phenomena that have been described above lead to reduced competitiveness on the public procurement market, which may have a negative impact on the quality of procured products (goods and services) and best value for money. Due to the lack of price competition, contractors may be inclined to overcharge, while relative ease of winning a contract (high probability of success) may make contractors submit tenders and win contracts that exceed their capacity, which in turn may negatively affect the quality of the completed contract. In the worst-case scenario, reduced competitiveness may result in contractors' excessive

¹⁰⁷ However, there are no data that would enable assessment to what extent this trend applies to SMEs compared to large enterprises.

¹⁰⁸ By comparison, in 2005-2006, there were around 4-5 non-rejectable tenders (the figures above also include rejectable tenders). In 2020, the average number of tenders increased to 2.78, but this trend was likely to have been influenced by the COVID-19 pandemic and the economic shutdown. Research results (Kowalewska, Głuszyński, 2018) demonstrate that the attractiveness of the public procurement market increases during economic downturns and slumps in the commercial market. The situation is similar as regards the attractiveness of the public sector as employer. During periods of economic downturns and crises, interest in employment in the public sector grows.

¹⁰⁹ In 2020, the proportion of tenders under the EU procedure in which only one tender was submitted ranged from 2% in Spain to 51% in Poland, while the EU average was 24%.

concentration on the public procurement market (where the same or a smaller number of contractors win more contracts). It should be noted that this is not necessarily reflected in the available indicators describing the functioning of the market, such as the average number of tenders submitted, or the share of SMEs in the value of contracts awarded.

In a situation where SMEs' insufficient participation in the public procurement market results not only from the sector's enterprises' lower potential for contract execution (financial, technical, administrative, etc.), but also from a specific attitude towards the market, it seems important to monitor the said attitude. Its most straightforward indicator may be the number of SMEs active on the market. Including this indicator in the set of indicators that have been used so far will enable a comprehensive evaluation of SMEs' presence on the public procurement market. The need to monitor how many enterprises from the sector are present on the public procurement market appears essential, as, without increasing the number of contractors, it can be very difficult to further increase the share of SMEs in the value of contracts awarded.

Impact of training and advisory programmes/projects on the activity of enterprises on the public procurement market

For many years, training schemes from the field of public procurement have been among the activities co-financed from EU funds. Currently, activities of this type aimed at improving the quality of human capital in enterprises are mainly implemented based on the principles of the demand model, under the so-called Entity Financing Systems (PSF). PSFs consist in the entrepreneur independently choosing a development service (training or advisory), and enable partial co-financing of the service's costs from the European Social Fund (ESF), under Regional Operational Programmes for 2014-2020. Subsidised services can be selected from the offer available in the Development Services Database (BUR) maintained by PARP¹¹⁰. Since the BUR was launched in the fourth quarter of 2015, 3,718 development services from the field of public procurement have been published in the Database, of which 157 have been implemented. They covered 681 participants from 279 enterprises (information obtained from the PARP Department of Human Resources Development in Enterprises, as of 13.10.2021).

¹¹⁰ The database is available at <https://uslugirozwojowe.parp.gov.pl/>.

The offer of development services available in the BUR is complemented by activities offered to enterprises under the supply model, financed from the Knowledge Education Development Programme for 2014-2020 (POWER). Under the PARP-implemented POWER Measure 2.2 *Support for strategic management of enterprises and building competitive advantage on the market*, training and advisory projects targeting SMEs are conducted, which concern public procurement. They aim to increase enterprise employees' competence in the area of public procurement and, ultimately, increase enterprises' activity on the public procurement market. According to the *POWER 2014-2020 Project list as of 8 June 2021*, 15 national public procurement contracts¹¹¹ (eight of which are completed projects) were funded under Measure 2.2. A total of 4,949 SMEs were to participate (author's calculations), including enterprises with no experience of the public procurement market. Participation in the project was supposed to make them active on this market.

In addition, under POWER Measure 2.9 *Social economy development* implemented by the Ministry of Family and Social Policy (MRiPS), similar projects targeting social economy entities, including social enterprises, are being executed. The *POWER 2014-2020 Project list* includes five such projects, three of which are being executed. The planned number of entities to be covered by training support is 1,614 (author's calculations).

Thus, within the framework of actions financed from the ESF (under ROP and POWER 2014-2020), educational support was directed to a group of about 6,900 entities (mainly enterprises from the SME sector) with the intention of activating them on the public procurement market. The scale and purpose of these initiatives justify the need to monitor the activity of support recipients on the public procurement market. A reliable evaluation of the effectiveness of this type of projects is also important in the context of new activities related to training human resources for the economy in the 2021-2027 Financial Perspective, under the European Social Fund (FERS) programme. According to the draft programme submitted for public consultation in June 2021, one of the objectives of FERS is to support the adaptation of employees, enterprises, and entrepreneurs to change through training and advisory activities, including further development of the BUR and strengthening of the quality of the development services provided (Ministry of Funds and Regional Policy [MFiPR], 2021, pp. 76-77). A continuation of training and advisory activities offered in the supply model is also possible. The importance of the public procurement market as a tool

¹¹¹ There is also a project aimed at activating Polish enterprises on the foreign public procurement market.

for achieving various development policy objectives suggests that these activities can be supported further.

It should be emphasised that the training and advisory projects described above are subject to evaluation, also in terms of their impact on the participants' activation on the public procurement market. So far, assessments of this impact have been based on surveys and participants' declarations (PARP, 2021). Measuring the change that occurred as a result of project participation is not an easy task. It requires two surveys (one conducted before, and one after project participation) ideally, covering all or almost all participants. Moreover, in order to reliably assess projects' impact on their participants activity on the public procurement market, it would be advisable to carry out a second measurement – at an equal interval after project completion, among all participants, as well as incorporating precise, preferably quantitative indicators of activity on the public procurement market. However, even then, relying on respondents' declarations regarding their activity in obtaining contracts may result in errors due to respondents' inaccurate memory. Using objective (factual) data when assessing a project's impact on the activity on the public procurement market would significantly increase the accuracy of conclusions and involve less financial and organisational effort than preparation and conducting of ex post research based on communication with participants (e.g., interviews, questionnaires) under a wider evaluation project. Data resources collected in the public procurement system, especially following the PZP of 2019 and its implementing regulations, open up such opportunities.

9.2. The scope and potential of data collected in the public procurement system¹¹²

This subsection presents the solutions and regulations introduced in 2021 (under the PZP of 2019), as well as those in force in 2016-2020 (under the PZP of 2004). The scope of data collected and the ways data is stored and made available differ, depending on whether we are dealing with the regime of the “old” or the “new” act. From the point of view of

¹¹² Design contests, which, like public procurement contracts, are also subject to publication in the BZP are excluded from this discussion. However, the specific nature of this instrument and its infrequent use by contracting authorities (in 2014-2020, on average, 75 tenders per year were published in the BZP) led to its exclusion from the analysis.

researchers, this circumstance is somewhat inconvenient as differences resulting from the different regulations need to be acknowledged and understood. Besides, data collected in two legal and organisational-technical systems have to be integrated in the analytical process. The decision to cover the two systems in this chapter was made due to the fact that the “new” act was in force for a relatively short period of time. Data collected during the period when the “old” act was in force may be potentially useful in evaluating the training and advisory activities described above and carried out in the 2018-2020 period. It may also help to assess the new regulations’ effectiveness in improving SMEs’ access to the market as compared to that of the old regulations. In the case of the “old” PZP, analysis is limited to the 2016-2020 period, when the regulations, including the provisions on the scope of information to be published in the BZP, were relatively stable. Therefore, the data from 2016-2020 describe a relatively homogeneous subject and have a uniform structure. Besides, such a time frame seems to be reasonable from the perspective of analysing the dynamics of participation in the public procurement market.

Data published in the BZP

Publication obligations at different stages of the public procurement contract

Despite the differences in the form of the two acts, the principles related to awarding public procurement contracts, the course of contract execution (including the procedure for its awarding), and the related publication obligations are essentially similar.

First of all, in the case of both acts, the obligation to adhere to them is applicable to contracts with a value exceeding a certain threshold. In 2016-2020, the threshold was EUR 30,000 (Act, 2014), while in the PZP of 2019 the threshold is set at PLN 130,000. As mentioned earlier, the BZP publishes information on contracts with values lower than the EU thresholds but higher than the indicated minimum thresholds for compliance with the act.

The scope of information to be provided in the BZP is determined by the type of contract-awarding procedure applied. The Act of 2004 provided for the following procedures: open procedure, restricted procedure, negotiated procedure with publication, competitive dialogue, negotiated procedure without publication, single-source procurement, price inquiry, innovation partnership and electronic auction. Each could be applied to contracts

with a value below the EU thresholds. The Act of 2019 identifies the following as appropriate for classic contracts¹¹³ with a value below the EU thresholds: basic procedure, innovation partnership, negotiated procedure without publication and single-source procurement.

The above procedures may be divided into open (competitive) and non-competitive. In open procedures, a contract notice is publicly announced (in the BZP or the Official Journal of the European Union). This enables an unlimited number of economic operators to participate in the procedure (provided that they meet the conditions set by the contracting authority). On the other hand, in non-competitive procedures, the contracting authority can select a group of entities to take part. In such a case, the contracting authority invites only a selected group of contractors to participate, with no public announcement of the contract. Under the PZP of 2004, competitive procedures include the open procedure, restricted procedure, negotiated procedure with publication, competitive dialogue, innovation partnership and electronic auction, while under the PZP of 2019, they include the basic procedure and innovation partnership. In both acts, the non-competitive procedures comprise the negotiated procedure without publication and single-source procurement, with the PZP of 2004 additionally listing the price inquiry.

The diagram below illustrates the public procurement contract process, detailing the types of notices published in the BZP at the particular stages and the modes to which they apply in light of the PZP of 2004 and 2019. Additionally, notices whose publication in the BZP is obligatory and those whose publication is optional have been marked.

The first stage of the public procurement contract process is the preparation of the procedure (description of the subject of the contract, definition of the conditions for participation in the procedure, and preliminary market consultations, if any). At this stage, there are no publication obligations for the contracting authority. Subsequently, the contracting authority commences the procedure to award the public procurement contract. In the case of open procedures, this is done by publishing a contract notice in the BZP, as mentioned above. If the content of the notice changes (e.g., the terms of reference or description of the needs and requirements), the contracting authority publishes a notice of amendment to the notice in the BZP. In the case of non-competitive procedures, the contract notice is not published in the BZP, but once the procedure is commenced, the contracting

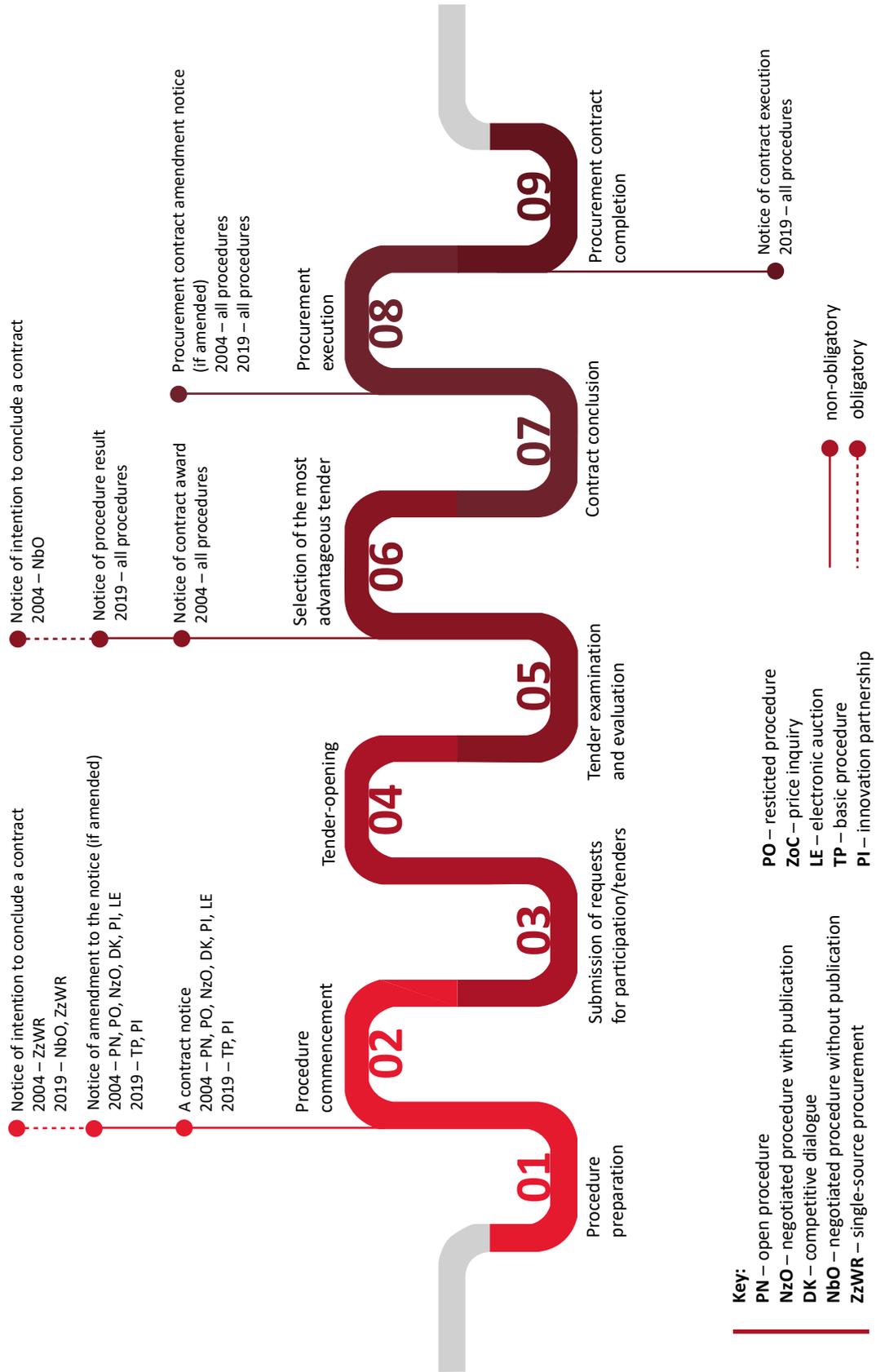
¹¹³ Classic contracts under the PZP of 2019 refer to contracts other than utilities contracts and defense and security contracts.

authority may publish a notice of the intended conclusion of the contract in the BZP. This increases contractors' control over the contracting authority's use of non-competitive procedures and brings certain benefits to the contracting authority if the PZP is violated in the course of awarding the contract, as the contract will not be invalidated (Nowak, Winiarz, 2021, p. 790). The next stage of the procedure involves interested contractors' submission of tenders, or requests to participate. Depending on the type of procedure, contractors submit tenders by the deadline set by the contracting authority (one-stage procedures), or requests to participate in the procedure. In the latter case, contractors qualified for the next stage are provided with further information and only they can submit tenders during the next stage of the procedure (two-stage procedures). Upon expiry of the deadline for submitting tenders, tenders are opened, examined, and evaluated against predetermined criteria. The evaluation results in selection of the best tender, following which a contract is concluded with the contractor who submitted the tender¹¹⁴. This stage of the procedure involves another publication obligation i.e., publishing, in the BZP, a notice informing about the result of the proceedings (in the Act of 2004, referred to as 'contract award notice'). Publication of a notice informing about the outcome of the proceedings (contract award) is and was obligatory, regardless of the mode selected for awarding the contract. Thereafter, the contract enters the execution phase which, as a rule, does not entail any publication obligations except for the situation where the contract is amended due to strictly defined circumstances provided for in the act¹¹⁵. If the contract is amended due to such circumstances, the contracting authority is obliged to publish a notice of contract amendment in the BZP. Upon execution of the contract, the contracting authority places a notice of contract execution in the BZP. This obligation was introduced by the Act of 2019.

¹¹⁴ During the period from the selection of the best tender to contract conclusion, legal protection measures may be filed. In cases provided for in the PZP; the procedure may also be annulled.

¹¹⁵ In the PZP of 2004, these are laid down in Article 144(1) p. 2 and 3, and in the PZP of 2019 – in Article 455(1) p. 3 and 4.

Diagram 9.1.1. Public procurement process and the obligation to publish notices in the BZP



Source: Own elaboration based on the PZP of 2004 and 2019

Thus, notices published in the BZP include¹¹⁶: the contract notice, notice of amendment to the notice, notice of intention to conclude a contract, notice of contract award/procedure result, notice of procurement contract amendment and, from 2021, notice of contract execution. Templates of these notices are determined by an ordinance of the minister responsible for the economy. When the Act of 2004 was in force, the content of the ordinances was amended several times. This is one of the reasons why, as mentioned above, the analysis is limited to the 2016-2020 period, when the same notice templates were used (Ordinance, 2016), with minimal amendments introduced in 2019 (Ordinance, 2019). However, at the beginning of 2021, along with the “new” PZP, the Ordinance of the Minister of Development, Labour and Technology of 23 December 2020 came into force and is the one currently applicable.

A detailed review of the notice templates included in the ordinances indicates that these are the data contained in the notice of contract award (PZP of 2004)/procedure results (PZP of 2019) that have greatest potential in an evaluation aimed at assessing the activity of SMEs on the public procurement market. They are published for all procedures, including those not publicly announced, which also provides value. The next section discusses the range of information contained in these notices highlighting the data with greatest utility potential. However, it should be noted that the data resources published in the BZP also provide a lot of other interesting information that can be used for research and evaluation purposes. In particular, this applies to contract notices which include a description of the subject of the contract, the applicable tender evaluation criteria, and other aspects of the contract awarding procedure.

Notice of contract award/procedure result

Notices provide basic information about the type of notice, contracting authority, subject of the contract, contract awarding procedure, number and prices of submitted tenders (lowest, highest, and price of the best tender), and the contractor to whom the contract was awarded. In analyses of enterprises’ activity on the public procurement market, most

¹¹⁶ The Act of 2019 also provides for a notice of compliance with circumstances referred to in Article 214(1) pp. 11-14 of the PZP concerning the so-called in-house orders. It is not discussed in this paper due to the limited usefulness of the information contained in this notice for the purposes of evaluation. For similar reasons, contract notices in the fields of defense and security and on subcontracting referred to in the Act of 2004 have been omitted as well.

important is the information about the contractor to whom the contract is awarded. In 2016-2020, notices provided the name of the contractor, their address, and information on their affiliation to the SME sector.

The introduction, as of 2021, of the obligation to provide the contractor's national identification number, which, in the case of Polish contractors conducting economic activity, is the REGON or NIP number, and in the case of foreign contractors, the VAT identification number, should be welcomed. The presence of such an identifier makes it possible to unequivocally determine the identity of the contractor. The contractor can also be identified by their name and address, but this is subject to greater risk of errors due to misspelling. It should be noted that according to the Single Market Scoreboard in the area of public procurement (concerning contracts published in the Official Journal of the European Union) Poland belongs to the group of countries with a high percentage of procedures where identification numbers of contracting authorities and contractors are missing (in 92% and 73% of procedures published in the Official Journal of the EU in 2019, respectively) (European Commission, 2020). As the authors of the ranking point out, these identifiers are crucial for understanding to whom and by whom public procurement contracts are awarded. Therefore, the requirement to provide data unambiguously identifying contractors (as well as contracting authorities¹¹⁷) introduced as of 2021 is a ground-breaking change that significantly increases the possibility of analysing the public procurement market from contractors' side.

Another positive change introduced in 2021 is the replacement of information on the contractor's affiliation to the SME sector with the *enterprise size* variable, taking into account the *micro*, *small* or *medium* categories. This variable is of great value in the system of information on the functioning of public procurement as it makes it possible to analyse SMEs' share in public procurement based on the three size categories. Although the problem of SMEs' under-participation in public procurement is described in general terms, by referring to the SME sector as a whole, it can be assumed that micro-enterprises are particularly affected.

Also potentially useful are variables describing the type of contracting authority, subject of its activity (from 2021 onwards), mode of contract award, and the prices of tenders (the lowest, highest, and the best tender). They enable contractors' characterisation in terms of

¹¹⁷ However, information on contracting authorities is of lesser importance for the subject of this study.

their operating strategy on the public procurement market, including their pricing strategy¹¹⁸. It is also possible to describe contractors' specialising in a certain type of contracts, in general (construction works, supplies, services), and in detail, using the Common Procurement Vocabulary (CPV) codes. The CPV is an extensive, multi-level classification system used in public procurement to describe the subject of a contract.

The variables listed above make it possible to narrow down the set of analysed procurement contracts to a certain group (e.g., contracts awarded by local authorities, contracts for services or services of a certain type, contracts for the supply of a specific good/product, contracts of a specific value, etc.). Variables informing whether the contract is financed with EU funds (and, if so, which project or programme) and, as of 2021, whether it is related to social and other specific services may serve the same purpose. This category of procurement contracts was introduced into the Polish PZP in 2016. Under the PZP, social and other specific services are those listed in Articles 74-77 of the Classical Directive (Directive 2014/24/EU). They include, i.a. health, social, community, other municipal, social and personal services. Considering the subject matter, social economy entities, including social enterprises, to which training and advisory support in the field of public procurement was directed under POWER, are naturally predisposed to their execution (MRiPS, 2019).

Table 9.1. Key information in notices of contract award/procedure result

	Notice of contract award (PZP of 2004)	Notice of procedure result (PZP of 2019)
General information		
Information on procedure result (contract conclusion or annulment)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Contract award (PZP of 2004)/contract conclusion (PZP of 2019) date	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Was the contract preceded by a contract notice published in the BZP; if so: contract notice number	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Information on the division of the contract into lots	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Does the public procurement concern a project or programme co-financed by European funds; if so: name of the project/programme	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Does the notice concern a contract for social and other specific services		<input checked="" type="checkbox"/>
Information on the contracting authority		
Type of contracting authority	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Subject of activity of the contracting authority		<input checked="" type="checkbox"/>

¹¹⁸ Information on tender evaluation criteria included in contract notices can also be used for the same purpose.

	Notice of contract award (PZP of 2004)	Notice of procedure result (PZP of 2019)
Procedure		
Mode of contract award	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Subject of the public procurement		
Public procurement name assigned by the contracting authority	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Type of procurement (construction works, supplies, services)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Short description of the subject of the procurement	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
CPV code (Common Procurement Vocabulary) – main and secondary	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Number of tenders		
Number of tenders received	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Number of tenders received from SMEs	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Number of rejected tenders	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Prices of tenders		
Price or cost of the tender of the contractor who was awarded the contract	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Price of the tender with the lowest price/cost	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Price of the tender with the highest price/cost	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Information on the contractor who was awarded the contract		
Was the contract awarded to contractors jointly applying for the award of the contract	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Name and address of the contractor who was awarded the contract; in the case of contractors applying jointly: names of all the contractors (PZP of 2004, PZP of 2019), proxy contact details (PZP of 2004)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
If the contractor conducts business activity, national identification number (REGON or NIP in the case of Polish contractors); in the case of contractors applying jointly: proxy's national identification number		<input checked="" type="checkbox"/>
Information on the contractor's enterprise size (micro, small, or medium); in the case of contractors applying jointly: information on the proxy's enterprise size		<input checked="" type="checkbox"/>
Is the contractor a small/medium enterprise (yes/no) to be provided for all contractors in the case of contractors applying jointly	<input checked="" type="checkbox"/>	
Does the contractor intend to subcontract parts of the contract to subcontractors	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
If so: names of subcontractors, if known		<input checked="" type="checkbox"/>

Source: Author's own study based on model templates (ordinances of 2016, 2019, and 2020).

The data listed above make it possible to meet the main information needs related to the activity of enterprises on the public procurement market understood as successful winning of contracts.

In particular, they allow to:

- determine the number of enterprises that were awarded a public contract (from 2021, also by size);
- analyse the number of companies that were awarded contracts in dynamic terms (from 2021, also by size), including the timeline related to the entry into force of the PZP of 2019;
- assess the impact of training and advisory support in the area of public procurement contracts on support's recipients' winning of contracts on this market, including:
 - verification of the presence on the market of entities which, prior to receiving support, were not present, or tried to win contracts but failed;
 - comparison of the level of market activity (measured by the number or value of contracts won) before and after receiving support among entities that had previously won public procurement contracts;
- analyse the activity of the total number of enterprises/supported enterprises/ other group of enterprises (depending on the objective of the study) on the public procurement market:
 - by the characteristics described above (e.g., number or value of contracts won, market operation strategy, including pricing strategy, specialisation, etc.)
 - or within a specific type of public procurement.

However, information about contractors contained in notices presents certain limitations. First of all, the BZP publishes data on contracts awarded under the PZP, while a significant portion of the market consists of contracts awarded under exemptions to the PZP. In 2016-2020, their share in the total value of contracts awarded oscillated between 30% and 40%, of which 12% to 20% were contracts with values below the threshold for the application of the Act (UzP, 2016-2020). Thus, the existing public procurement information system lacks information on contractors winning such contracts of lower value, or for other reasons awarded outside the PZP regime. Data on contractors applying jointly (i.e., with another contractor), but not acting as a proxy, are also incomplete. When the "old" Act was in force, their names and addresses were provided. However, since 2021, only the names have been provided. Information on enterprises participating in the public procurement market as subcontractors is limited as well. The "old" Act did not require their inclusion in the notices.

Since 2021, they have been mentioned by name, if known to the contracting authority. However, data available in the BZP primarily concerns contractors who were awarded a public procurement contract, omitting those who participated in the procedure but were unsuccessful. The lack of information on unsuccessful tenderers is a significant drawback as regards information on enterprises active on the public procurement market. The pool of unsuccessful tenderers could potentially include parties that participated in the training and advisory programmes described above, especially those taking their first steps on the public procurement market.

Access to data on unsuccessful entities would also enable analyses of companies' effectiveness winning public procurement contracts, by size (e.g., proportion of enterprises from a given size category that are awarded contracts against all enterprises from that size category participating in the procedures).

Article 81(1) of the PZP of 2019 provides an opportunity to fill in the information gap described above. It imposes a new obligation on contracting authorities, to provide the president of the UZP with information on requests to participate and tenders submitted during the public procurement procedure. This obligation applies to procedures where tenders are submitted, i.e., excluding contracts awarded under the single-source procurement procedure¹¹⁹, regardless of type of contract awarded. In line with the provisions of the act, this information is to serve as an ongoing analysis of the functioning of the procurement system.

The contracting authority prepares information on requests and tenders electronically, and delivers it to the president of the UZP using the form available on the UZP website. The scope of data to be delivered is laid down in the ordinance of the minister responsible for the economy. Under the current ordinance, these are: contracting authority data, basic information on the procedure and contract (subject, mode), and data of contractors participating in the procedure, essential from the point of view of their usability in the evaluation (Ordinance on information on submitted applications or tenders, 2020). These include the name, address, type of contractor¹²⁰, and national identification number. In the case of contractors applying jointly, all the aforementioned data are to be provided for the

¹¹⁹ This mode consists in awarding contracts following negotiations with only one contractor.

¹²⁰ Selection options: micro enterprise, small enterprise, medium enterprise, self-employed, natural person not conducting business activity, other.

proxy. As regards the remaining contractors, only their names and national identification numbers are to be provided. In addition, information on the prices of tenders submitted by the particular contractors is to be provided, as well as the contract notice number, if it was published in the BZP. Therefore, the number may serve as a variable linking information on requests or tenders with data contained in the relevant notice of procedure result.

Data from the two above sources i.e., the BZP and information provided to the president of the UZP about all contractors participating in the procedure, provide full knowledge about companies active on the public procurement market as contractors acting independently or jointly with another entity/other entities, regardless of the outcome of this activity. It should be noted that the value of both sets of data stems from the fact that they are generated dynamically, in the course of awarding public procurement contracts (almost in real time). The obligation to provide the president of the UZP with information on contractors participating in the procedure must be fulfilled no later than within seven days from the date of tender-opening session, or date of procedure annulment. On the other hand, notices of procedure result must be published in the BZP within 30 days from procedure completion.

9.3. Ways of accessing data collected in the public procurement system

There are various ways of accessing the data that is described above. In the case of data collected in the BZP, the current degree of their “readiness” for being made available also varies.

The “old” BZP

Data covering the period of the “old” PZP, i.e., for procedures initiated before 1 January 2021, can be found in the so-called “old” BZP at <https://bzp.uzp.gov.pl/>. These are data for the period beginning in May 2017. Data from the earlier period, collected from the 2007 introduction of the principle of electronic publication of notices in the BZP, can be found on the office’s ftp website: <ftp://ftp.uzp.gov.pl/bzp2/>.

Access to notices published in the “old” BZP is possible via a web service available at http://websrv.bzp.uzp.gov.pl/BZP_PublicWebService.asmx. Data sharing and exchange between IT systems is possible thanks to the Web Service. It enables secure access to data stored in the BZP and their automatic downloading by external systems. The service makes it possible to search for notices by selected criteria. The searched data are returned in the xml or json formats.

The “new” BZP

Notices regarding procedures initiated as of 2021 under the PZP of 2019 are published in the so-called “new” BZP on the e-Zamówienia (*e-Procurement*) platform at <https://ezamowienia.gov.pl/pl/>. The platform’s design is the main part of the *e-Zamówienia – electronic public procurement* project¹²¹ implemented by the UZP in partnership with the Ministry of Digitalisation (currently KPRM Cyfryzacja). The design meets the obligation of digitisation of the process of awarding public procurement contracts introduced by the PZP. The platform is to enable electronic communication and exchange of information at all stages of the public procurement procedure, including submission of tenders, requests, statements and other documents. It is also intended to enable electronic publication of the BZP and support contracting authorities in meeting their reporting obligations under the PZP.

Platform developer was selected in the procedure entitled Development of e-Zamówienia platform and contract engineer services (part I – Design, development, and implementation of the e-Zamówienia platform including author’s supervision over the operation of the e-Zamówienia platform in the stabilisation period and maintenance service for 12 months)¹²² conducted by the UZP. The contract was concluded on 31 March 2020, and the expected delivery date is 30 June 2022. The platform will comprise a number of functionalities grouped into modules and components.

The “new” BZP was implemented at the beginning of 2021 as part of the notices module. Unlike the previously applicable BZP, the “new” BZP, in addition to national notices, includes

¹²¹ The project is co-financed from EU funds under the Digital Poland Operational Programme 2014-2020, **priority axis 2 E-administration and open government, measure 2.1 High availability and quality of public e-services**.

¹²² In the description of the e-procurement platform concept, description of the subject of the listed procedure was used.

also EU notices. These are imported after their publication on the Tenders Electronic Daily website (supplement to the Official Journal of the EU). Thanks to this option, the “new” BZP functions as a central database of all notices (both national and EU) published by Polish contracting authorities, with uniform data structure maintained. Thus, it offers the possibility of analysing the activity of contractors on the market, covering both contracts below the EU thresholds, and those with values above the thresholds, although, as indicated above, the obligation of BZP publication applies only to the former. Data from notices published in the “old” BZP will also be migrated and made available in the “new” BZP.

Use of data published in the “new” BZP will be possible upon the launch of the monitoring and analysis module, currently under development (as of 15 November 2021). According to information from the Report on the material and financial progress of the IT project for the first quarter of 2021, the module is to be made available in August 2022 (UZP, 2021). It will enable ongoing tracking of information from the entire procurement lifecycle, and analyses of the public procurement market. Thanks to advanced tools, access will be provided to specific participants’ activities on the public procurement market, on either individual or aggregated basis, according to selected criteria. Access to data will be possible through an advanced business intelligence tool enabling data visualisation and generation of reports according to user-selected criteria. The tool is expected to use all data sources collected on the e-Zamówienia platform. In the future, it should also be able to use external data sources (e.g., KRS, KIO, Electronic Invoicing Platform, etc.). In line with the public data opening trend, data will be made available via an application programming interface (API) enabling the connection and interchange of data between computer programs or data communication systems. The module will allow data to be downloaded for export for further processing (for example in csv, xls(x), xml, json formats).

Information on requests to participate and tenders submitted under the PZP of 2019

The e-Zamówienia platform also enables transfer of information about the tenders and requests submitted in the procedures after the contracting authority logging in using the form of information about submitted tenders and applications to the president of the UZP. This information is not subject to mandatory publication under the provisions of the PZP.

However, implementation of public data openness policy and trends from the area of re-use of public information (Act, 2021; Open data programme for 2021-2027) suggest that this data is likely to be made available. Due to the ongoing works on the e-Zamówienia platform, the method for making them available has not yet been determined (information obtained from the UZP). Regardless of the decisions that will be taken in this respect, it should be emphasised that even if it turns out that they will not be made available in a non-applicable manner, as a rule, they are available on request, as public information, in accordance with the Act on Access to Public Information (2001, Article 10(1)). Contracting authorities are obliged to make public information available under the provisions of this Act, considering the specific provisions of the PZP (Bogdanowicz, 2021, pp. 1-2).

Summary

The results of the analysis show that the BZP, a national publication, holds a rich resource of data on public procurement contracts awarded in Poland. Positive changes in terms of the scope and quality of data collected in the BZP, especially those concerning contractors, were brought about by the PZP of 2019. The 2021 introduction of the requirement to include contractors' NIP or REGON numbers in notices of procedure result, clearly identifying contractors, is a new improvement in the public procurement information system. The same applies to the potential of data on contractors who submitted requests to participate in the procedure or tenders. The possibility to import notices published in the Official Journal of the European Union to the BZP is also a significant improvement. Thanks to this, information on national public procurements below the EU thresholds and procurements with values higher than the thresholds, which was previously dispersed, will now be located in one data repository. This will extend analyses of the functioning of the system and open up the possibility of re-using this data for other studies, as existing data that can be further analysed, but also as a database of enterprises operating in the particular market areas (e.g., defined by CPV codes).

The chapter focused on the possibility of using data on public procurement in the evaluation of projects aimed at improving enterprises' knowledge in this area and, ultimately, increasing their activity on the public procurement market. However, the scope of data collected by the UZP also enables their use in analyses regarding different topics. The key issue here seems to be the evaluation of public procurement as a tool for the implementation of other (non-

procurement) strategic objectives of government policy (such as increasing competitiveness, increasing SMEs' market share, increasing the number of innovative, "green," social procurement) and other specific objectives included in the country's procurement policy adopted in January 2022. As recommended by the OECD (2015), use of the public procurement system to achieve such additional objectives should be evaluated. However, the OECD cautions that the burden of monitoring progress towards these objectives should be minimised. The use of data collected in the procurement process in evaluations, mainly in connection with the implementation of the principle of openness and transparency, seems to be fully in line with the postulate of cost effectiveness.

It is also worth considering the validity and applicability of data collected in the public procurement system in evaluations of EU operational programmes implemented in Poland. Since about 50% of the funds under these programmes (European Commission, 2015) are spent in the public procurement system, data from the system (including unstructured data from contract award notices, e.g., descriptions of the subject of the contracts, selection criteria applied) could support final evaluation of the programmes.

References

- Act of 15 February 1933 on supplies and works for the State Treasury, Local Government and Public Law Institutions [*Ustawa z 15.02.1933 r. o dostawach i robotach na rzecz Skarbu Państwa, samorządu oraz instytucji prawa publicznego*]. Reference of the Polish Journal of Laws: Dz. U. z 19 r. poz. 127., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU19330190127> (access: 19.04.2022)
- Act of 10 June 1994 on Public Procurement [*Ustawa z 10.06.1994 r. o zamówieniach publicznych*]. Reference of the Polish Journal of Laws: Dz. U. z 76 r. poz. 344., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU19940760344> (access: 19.04.2022)
- Act of 6 September 2001 on Access to Public Information [*Ustawa z 6.09.2001 r. o dostępie do informacji publicznej*]. Reference of the Polish Journal of Laws: Dz. U. z 2020 r. poz. 2176, as amended., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20200002176> (access: 19.04.2022)
- Act of 29 January 2004 on Public Procurement [*Ustawa z 29.01.2004 r. – Prawo zamówień publicznych*]. Reference of the Polish Journal of Laws: Dz. U. z 2019 r. poz. 1843., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190001843> (access: 19.04.2022)
- Act of 13 April 2007 amending the Act – Act on Public Procurement and some other acts [*Ustawa z 13.04.2007 r. o zmianie ustawy – Prawo zamówień publicznych oraz niektórych innych ustaw*]. Reference of the Polish Journal of Laws: Dz. U. nr 82, poz. 560., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20070820560> (access: 19.04.2022)

- Act of 14 March 2014 amending the Act – Act on Public Procurement and some other acts [*Ustawa z 14.03.2014 r. o zmianie ustawy – Prawo zamówień publicznych oraz niektórych innych ustaw*]. Reference of the Polish Journal of Laws: Dz. U. poz. 423., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20140000423> (access: 19.04.2022)
- Act of 11 September 2019 on Public Procurement [*Ustawa z 11.09.2019 r. – Prawo zamówień publicznych*]. Reference of the Polish Journal of Laws: Dz. U. z 2021 r., poz. 1129, as amended., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20210001129> (access: 19.04.2022)
- Act of 11 August 2021 on open data and reuse of public sector information [*Ustawa z 11.08.2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego*]. Reference of the Polish Journal of Laws: Dz. U. poz. 1641. <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20210001641> (access: 19.04.2022)
- Bogdanowicz, P. (2021). *Dostęp do informacji w zamówieniach publicznych w prawie polskim i jego główne ograniczenia*, https://www.batory.org.pl/wp-content/uploads/2021/08/P_Bogdanowicz_Dostep-do-informacji-w-zamowieniach_Analiza.pdf (access: 21.10.2021).
- Dolecki, J. (2009). *Zamówienia publiczne – efektywny instrument prawny czy biurokratyczna procedura*, p. 26-35. In: *XV-lecie systemu zamówień publicznych w Polsce*, Nowicki, H., Sadowy, J. (ed.). Toruń-Warsaw: UZP.
- *Directive 2009/81/EC of the European Parliament and of the Council of 13 July 2009 on the coordination of procedures for the award of certain works contracts, supply contracts and service contracts by contracting authorities or entities in the fields of defence and security and amending Directives 2004/17/EC and 2004/18/EC*. Official Journal of the European Union L 216, p. 76, as amended., <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0081&from=EN> (access: 19.04.2022)
- *Directive 2014/24/EU of the European Parliament and of the Council of 26 February 2014 on public procurement*. Official Journal of the European Union L 94, p. 65, as amended., <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0024&from=EN> (access: 19.04.2022)
- *Directive 2014/25/EU of the European Parliament and of the Council of 26 February 2014 on procurement by entities operating in the water, energy, transport and postal services sectors*. Official Journal of the European Union L 94, p. 243, as amended., <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0025&from=EN> (access: 19.04.2022)
- European Commission. (2010). *Communication of 3 March 2010 – Europe 2020: A strategy for smart, sustainable and inclusive growth*. COM (2010) 2020.
- European Commission. (2015). *Public Procurement – Guidance for practitioners on the avoidance of the most common errors in projects funded by the European Structural and Investment Funds*. https://ec.europa.eu/regional_policy/sources/docgener/informat/2014/guidance_public_proc_en.pdf (access: 30.10.2021).
- European Commission. (2017). *Communication of 3 October 2017 – Making Public Procurement work in and for Europe*. COM (2017) 572 final.
- European Commission. (2020). *Single Market Scoreboard, Public Procurement*. https://ec.europa.eu/internal_market/scoreboard/performance_per_policy_area/public_procurement/index_en.htm (access: 21.10.2021).

- European Commission. (2021). *SME Needs Analysis in Public Procurement. Final Report*. <https://ec.europa.eu/docsroom/documents/46111> (access: 24.10.2021).
- GUS, Macroeconomic Data Bank, <https://bdm.stat.gov.pl/> (access: 28.10.2021).
- Kowalewska A., Szut J. (2011). *Nowe podejście do zamówień publicznych. Raport z badań 2010/2011*. Warsaw: PARP.
- Kowalewska A., Szut J. (2012). *Nowe podejście do zamówień publicznych. Raport z badań 2012*. Warsaw: PARP.
- Kowalewska A., Szut J. (2013). *Nowe podejście do zamówień publicznych. Raport z badań 2013*. Warsaw: PARP.
- Kowalewska, A., Głuszyński J. (2018). *Raport z wyników badania. Ocena gotowości polskich przedsiębiorstw do realizacji zadań publicznych*. https://www.parp.gov.pl/storage/publications/pdf/Ewaluacja_Zamowienia_publ_PPP_Raport_-01_02_2019_FINAL_www.pdf (access: 21.10.2021).
- MFiPR. (2021). Lista projektów POWER 2014-2020 z 8.06.2021 r. <https://www.power.gov.pl/strony/o-programie/projekty/lista-projektow/> (access: 22.10.2021).
- MPiT, UZP. (2018). *Koncepcja nowego prawa zamówień publicznych*. https://www.uzp.gov.pl/__data/assets/pdf_file/0029/36875/Koncepcja_nowego_prawa_zamowien_publicznych.pdf (access: 19.09.2021).
- MPiT. (2019). *Uzasadnienie projektu ustawy – Prawo zamówień publicznych skierowanego do Sejmu RP w dniu 12.07.2019 [The explanatory note to the draft act – Public Procurement Law submitted to the Polish Sejm on 12.07.2019]*. [http://orka.sejm.gov.pl/Druki8ka.nsf/Projekty/8-020-1389-2019/\\$file/8-020-1389-2019.pdf](http://orka.sejm.gov.pl/Druki8ka.nsf/Projekty/8-020-1389-2019/$file/8-020-1389-2019.pdf) (access: 03.11.2021).
- MRiT. (2021). *Projekt polityki zakupowej państwa [Draft state procurement policy]*. <https://www.gov.pl/web/rozwoj-technologie/konsultacje-publiczne-polityki-zakupowej-panstwa> (access: 31.01.2022).
- MRPiPS. (2019). *Krajowy Program Rozwoju Ekonomii Społecznej do 2023 roku. Ekonomia Solidarności Społecznej*. https://www.ekonomiaspoleczna.gov.pl/download/files/EKONOMIA_SPOLECZNA/KPRES_2019.pdf (access: 21.10.2021).
- Nowak, H., Winiarz, M. (ed). (2021). *Prawo zamówień publicznych. Komentarz*. https://www.uzp.gov.pl/__data/assets/pdf_file/0028/49078/Komentarz-do-Prawa-Zamowien-Publicznych-wersja-uzytkowa.pdf (accessed: 10/10/2021).
- OECD. (2015). *Recommendation On Public Procurement*. <https://www.oecd.org/gov/public-procurement/OECD-Recommendation-on-Public-Procurement.pdf> (access: 30.10.2021).
- Ołdak–Bułanowska, K. (2014). *Aspekty społeczne w zamówieniach publicznych – Podręcznik*. https://www.uzp.gov.pl/__data/assets/pdf_file/0013/24700/UZP_aspekty_spoleczne.pdf (access: 25.10.2021).
- Ordinance of the Minister of Development of 26 July 2007 on template notices published in the Public Procurement Bulletin [*Rozporządzenie Ministra Rozwoju z 26.07.2016 r. w sprawie wzorów ogłoszeń zamieszczanych w Biuletynie Zamówień Publicznych*]. Reference of the Polish Journal of Laws: Dz. U. poz. 1127., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20160001127> (access: 19.04.2022)
- Ordinance of the Minister of Development of 16 December 2019 amending the Ordinance on template notices published in the Public Procurement Bulletin [*Rozporządzenie Ministra Rozwoju z 16.12.2019 r. zmieniające rozporządzenie w sprawie wzorów ogłoszeń zamieszczanych w Biuletynie Zamówień*

Publicznych]. Reference of the Polish Journal of Laws: Dz. U. poz. 2469., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20190002469> (access: 19.04.2022)

- Ordinance of the Minister of Development, Labour and Technology of 21 December 2020 on information on submitted requests to participate in procedures or tenders transmitted to the President of the Public Procurement Office [*Rozporządzenie Ministra Rozwoju, Pracy i Technologii z 21.12.2020 r. w sprawie informacji o złożonych wnioskach o dopuszczenie do udziału w postępowaniu lub ofertach przekazywanej Prezesowi Urzędu Zamówień Publicznych*]. Reference of the Polish Journal of Laws: Dz. U. poz. 2406., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20200002406> (dostęp: 19.04.2022)
- Ordinance of the Minister of Development, Labour and Technology of 23 December 2020 on notices published in the Public Procurement Bulletin [*Rozporządzenie Ministra Rozwoju, Pracy i Technologii z 23.12.2020 r. w sprawie ogłoszeń zamieszczanych w Biuletynie Zamówień Publicznych*]. Reference of the Polish Journal of Laws: Dz. U. poz. 2439., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20200002439&SessionID=07A732157F7C91A9B7EC217000E54AB44AC54346> (access: 19.04.2022)
- PARP. (2021). *Badania CAWI oraz IDI w ramach ewaluacji on-going działań 2.2 i 2.21 Programu Operacyjnego Wiedza Edukacja Rozwój, 2014-2020. Raport końcowy z czterech edycji badania*. https://www.parp.gov.pl/storage/publications/pdf/PO-WER_raport-Danae-dla-PARP_kocowy_ost1.pdf (access: 24.10.2021).
- Skowrońska, A., Tarnawa, A. (ed.). (2021). *Raport o stanie sektora małych i średnich przedsiębiorstw w Polsce z 2021 r.* https://www.parp.gov.pl/storage/publications/pdf/PARP-26_Raport-2021-07-22_WCAG_210726.pdf (access: 24.10.2021).
- Resolution No. 28 of the Council of Ministers of 18 February 2021 on the Data Opening Programme for 2021-2027 [*Uchwała nr 28 Rady Ministrów 18.02.2021 r. w sprawie Programu otwierania danych na lata 2021–2027*]. Reference to the Official Journal of the Republic of Poland: M.P. poz. 290., <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WMP20210000290> (access: 19.04.2022)
- UZP. (1995-2020). *Sprawozdania o funkcjonowaniu systemu zamówień publicznych za lata od 1995 r. do 2020 r.* <https://www.uzp.gov.pl/baza-wiedzy/analizy-systemowe/sprawozdania-o-funkcjonowaniu-systemu-zamowien-publicznych> (access: 21.10.2021).
- UZP. (2019). *Budowa Platformy e-Zamówienia oraz usługi Inżyniera Kontraktu*. <https://uzp.bip.gov.pl/publiccontracts/view/20623> (accessed: 21.10.2021).
- UZP. (2021). *e-Zamówienia – elektroniczne zamówienia publiczne. Raport z postępu rzeczowo-finansowego projektu informatycznego za I kwartał 2021 roku*. <https://www.gov.pl/web/krmc/e-zamowienia--elektroniczne-zamowienia-publiczne5> (access: 30.10.2021).
- Włodarczyk, A. (ed.). (2010). *Nowe podejście do zamówień publicznych – szkolenia i doradztwo. Projekt systemowy Polskiej Agencji Rozwoju Przedsiębiorczości*. <https://en.parp.gov.pl/storage/publications/pdf/9174.pdf> (accessed: 12.10.2021).

10. Using company financial statements in evaluating the impact of economic development programmes¹²³

Andrzej Jędrzejowski

Department of analysis and strategy

Polish Agency for Enterprise Development (PARP)

Jacek Pokorski

Department of analysis and strategy

Polish Agency for Enterprise Development (PARP)

Introduction

This chapter presents the concept of evaluation studies using administrative micro data as the main source of information for identifying causal (net) effects and assessing the impact of public interventions. It also includes a cross-sectional discussion of the analytical potential of a source of information that was previously underestimated in evaluations i.e., a set of enterprises' financial statements from the repository of documents of the National Court Register (KRS) made publicly available by the Minister of Justice. The chapter elaborates on the rationality of such an approach in evaluations of the impact of economic development programmes, and refers to an example of its practical application within the framework of *Counterfactual analysis of the impact of the Operational Programme Eastern Poland 2014-2020 (...)* (POPW) (IDEA Institute and IDEA Development Foundation commissioned by PARP, 2021) identifying the analytical possibilities offered by open KRS data.

¹²³ All conclusions formulated in this chapter are the opinion of the authors and do not express the official stance of Statistics Poland, the Ministry of Justice, or the National Court Register and have not been verified by the above-mentioned institutions.

Conclusions from longstanding conferences and international evaluation community's debates¹²⁴ on the role of evaluation in evidence-based policy making, the directions of the EU *evaluation policy*¹²⁵ reflected in the guidelines for this process's implementation at different levels of public funds' use (horizontal, sectoral, community, national), and the growing possibilities of processing and opening large public data sets confirm this trend as leading for the future development of evaluation practices. The use of data recorded in public systems in evaluation processes, almost in real time, becomes an even more urgent challenge in the dynamically changing world where, in many respects, it seems unreasonable to prepare and conduct research *as before*. Undoubtedly, the catalyst for the said trend was the coronavirus pandemic, which radically changed the way of obtaining data in socio-economic research (including evaluation). On the one hand, remote research techniques practically dominated the primary data acquisition methodology. On the other hand, even more attention was devoted to the necessity of secondary use of the already existing extensive resources, i.e., objective, free, up-to-date data extracted from public registers. The latter perspective applied not only to evaluation units, but extended to wider public service structures responsible for managing the particular development policies (including diagnostics, international benchmarking, preparation, monitoring, and evaluation of interventions) in the turbulent socio-economic circumstances of the pandemic.

The chapter takes detailed inventory of the selected source of administrative information, i.e. the set of companies' financial data reported to the KRS. This is publicly available (free of charge, standardised), micro, non-anonymised data from annual cycles (end of the accounting period), presenting the basic characteristics and numerous indicators of the financial condition of domestic enterprises. Since 2018, the statements have been submitted in electronic form only, and made available through the Financial Document Browser (EKRS) of the <https://ekrs.ms.gov.pl/> portal. Inventory of this resource was taken to present the example of the evaluation of the impact of the Operational Programme Eastern Poland 2014-2020 (POPW) referred to in the chapter. It enabled the assessment of the feasibility of this study using data from the EKRS.

¹²⁴ Cf. Conference on the Evaluation of the Cohesion Policy of the European Union organised by the European Commission since 1998, https://ec.europa.eu/regional_policy/en/policy/evaluations/conferences/; International Evaluation Conference in Poland organised by the minister responsible for regional development and PARP since 2005; <https://www.ewaluacja.gov.pl/strony/archiwalne-strony-konferencji-ewaluacyjnej/>.

¹²⁵ Cf. UNDP <http://web.undp.org/evaluation/policy.shtml>; OECD <https://www.oecd.org/gov/policy-monitoring-evaluation.htm>

As part of the inventory, characteristics of the data resource made public through the Browser were discussed. The following were presented: the entity's scope (entities whose data are made available were identified), content of financial statements, time range of the available data, specific character of access to data, and detailed characteristics of the particular elements of the statements. Lists of the available financial indicators were also discussed, along with data coverage of the sample population of enterprises (applicants for selected support measures of the POPW subject to analysis). Inventory results may also be used for future research conducted on the basis of this source of information.

Most important results of analyses presenting causal effects of enterprise support under the POPW, estimated on the basis of the KRS data in question, are cited as well, for illustration purposes. In this case, the so-called counterfactual approach was applied, which is particularly demanding with regard to micro data. Its application usually involves different types of statistical matching techniques based on a pre-estimated propensity score value. Preparation of such matching models and conducting comparative analyses of entities affected (treated) and not affected (untreated) by the impact of the intervention requires both the provision of data on relevant characteristics of these entities, and values of economic indicators regarding the entities, which can be used when estimating the possible effects of the intervention. Importantly, such data can be downloaded from the EKRS Browser, for specific entities (e.g., beneficiaries and non-beneficiaries). Comparative analyses which, in addition to beneficiaries' data, use data of selected control entities, make it possible to model the economic standing of beneficiaries (counterfactual), if the entities had not received co-financing, thus enabling an estimation of the programme's net effects.

10.1. Data... the fuel of evaluation?¹²⁶

A technological trend that has been identified and described by analysts for several years is the growing role of data. Analysis and interpretation of increasingly rich data sets is an element of the ongoing digital transformation (Hatałska-Woźniak, 2020)¹²⁷. This trend

¹²⁶ Cf. *Data... the oil of evaluation?* (Walsh, 2021) – speech by John Walsh (European Commission, Directorate General for Regional and Urban Policy, Evaluation and the European Semester), during the debate *The use of data for evaluation. Data sources for evaluation and the links among them* at the 9th Conference on Evaluation of EU Cohesion Policy (Porto, Portugal 16-17.09.2021).

¹²⁷ Cf. *Data Is the New Black* (Hatałska-Woźniak, 2020).

and the opportunities provided by data collected in countless public information systems cannot be ignored also when designing various types of research and analyses conducted in the public sector (including programme evaluations). Considering the needs of evaluation studies and the relatively new approach based on a counterfactual standing enabling the estimation of interventions' causal effects, in 2012-2013, the ministry responsible for regional development prepared an expert opinion on the sources of institutional data that can be used in such evaluations. This expert opinion "identified and described (...) 253 institutional databases available to 112 companies and institutions, including ministries, their agencies, and other central entities, provincial marshal offices, provincial offices, provincial labour offices, Statistics Poland, the Social Insurance Institution, district examination boards, non-governmental institutions, scientific entities, and research companies." (GHK Polska commissioned by the Ministry of Regional Development, 2013). Unfortunately, since then, there have been few examples of successful use of administrative data in counterfactual impact evaluations, and of institutions that regularly turn to this data source in this type of processes.

The successive financial perspectives on the use of European Funds result in new guidelines for the evaluation process. In both the 2007-2013 and 2014-2020 perspectives, European Commission services' guidance and national documents of this type (Ministry of Development Funds and Regional Policy, 2021)¹²⁸ identified provision of data for evaluations as one of the key criteria for feasibility (in particular of counterfactual evaluations). "Effective monitoring of the intervention and accurate collection and processing of data are crucial for ensuring the quality of the evaluation. As soon as the aid scheme is approved, a mechanism should therefore be put in place to monitor the intervention and to collect and process the appropriate data. (...) The evaluation may need to draw on existing data sources, such as administrative data sources (e.g. the tax office, the companies register, innovation surveys and the patent office). The evaluation plan therefore needs to review the existing data sources, decide whether they provide sufficient information for the evaluation and ensure that access to them will be possible within the relevant timeframes." (European Commission, 2014 a, pp. 9-11).

¹²⁸ "While designing the evaluation process, consideration should be given to (...) planning how to provide the data necessary for the implementation of the evaluation, taking particular account of the needs of counterfactual evaluations" – cf. Ministry of Development Funds and Regional Policy, 2021, p. 18.

For many years, the data issue has also been discussed by evaluation professionals, at various meetings and conferences. On the one hand, their participants are aware that, in the process of developing evaluation plans, besides information obtained as a result of the evaluation, identification of data sources is also crucial. If resources of the existing administrative systems are to be used, it is necessary to plan this, and arrange access to data for the planned study with the hosts. In practice, it is not uncommon for the issue of data provision to be addressed as late as the initiation stage of the research process. In the case of data for counterfactual impact evaluations, the issue is particularly complicated because analytical procedures (e.g. statistical unit matching) require access to micro data covering the target group of the programme (including beneficiaries, and potentially similar entities eligible for support but ultimately not covered). Requirements regarding confidentiality and protection of data subjects' rights often prove to be an insurmountable barrier, especially if the data subjects (natural or legal persons) have not consented to the use of their data for evaluation purposes, which is usually the case among non-participating entities. At the same time, as previously indicated, data requirements imposed by the counterfactual analysis methodology are relatively demanding. As a rule, such analyses are conducted on micro data and involve a number of other conditions for their feasibility, in particular:

- the set must contain potentially relevant characteristics of individuals (metrics: age, gender, educational attainment, city size, region, institutional characteristics, etc.) as well as indicators of change in their condition over time; in the case of programme participants, these indicators will potentially show the intervention's effect and other effects of unrelated events; in the case of individuals not subject to support, the indicators will only show the effects of events unrelated to the intervention;
- the set must enable separation of programme participants from *the rest* (e.g. on the basis of PESEL – personal ID number, NIP/REGON/KRS identifiers – business ID numbers, a local government unit designation) and must be compatible with other supplementary sets (e.g. databases of applicants, databases of ultimate programme participants, etc.), i.e. it must be prepared according to a similar scheme, making it possible to combine the sets,
- *the rest* set (potential control pool) should preferably be multiple times the number of beneficiaries.

A few years ago, the data issue in Poland was overcome to some extent, as the 2014-2020 financial perspective for the use of EU funds approached, given the information needs related to the ex-post evaluations of the programmes of the previous perspective (2007-2013) and, above all, in connection with the stringent requirements for the evaluation of EU

State aid after 2014 (under the provisions of the GBER regulation¹²⁹) (European Commission, 2014 b). Research approaches and analysis methodologies on micro data extracted from public statistics resources for counterfactual impact evaluations of selected programmes were developed (GUS¹³⁰ Statistical Research and Education Centre, 2015, 2017, 2018). This was the result of the involvement, in the evaluation process, of experts (GUS Statistical Research and Education Centre, 2013) and stakeholders from the public sector, including the management of Statistics Poland (GUS) and representatives of selected evaluation units. In a sense, this approach was supposed to “combine water with fire”, i.e. enable analyses on micro data collected in the public system while protecting statistical confidentiality of individuals. To put it simply, in practice, the designed analyses of the data collected in the GUS were carried out by evaluators (in cooperation with evaluation units) using the *multi-iteration in the dark* method, i.e. they did not have access to micro data, but used queries (specific sets of commands in the scripting language) to the sets prepared by the GUS. Thus, the data remained safe in the statistical office as its employees tested and then ran the analysis algorithm programmed by the evaluator. In this way, the GUS obtained the matches and effect estimates assumed by the evaluator. They no longer constituted micro data, but aggregated indicators for a set group of units, which could be made available to the evaluators (GUS Statistical Research and Education Centre, 2018). Despite its procedural complexity, this approach proved to be very effective. On the basis of data collected in the public system, it enabled implementation of several groundbreaking projects for the output of the evaluation of the EU cohesion policy implemented in Poland (IDEA Institute, IDEA Development Foundation, and Jagiellonian University commissioned by PARP, 2020, 2021); Evalu and Ecorys Polska commissioned by the Ministry of Enterprise and Technology, 2020; Evalu, STOS Association for Labour Market Development and WiseEuropa Foundation commissioned by the Ministry of Development Funds and Regional Policy, 2020)¹³¹ and other research, e.g. conducted by the World Bank under the ROI project (Return on investment of public support to SMEs and innovation in Poland) (World Bank, 2019)¹³².

¹²⁹ *General Block Exemption Regulation, GBER.*

¹³⁰ GUS – Statistics Poland – Poland’s chief government executive agency responsible for collecting and publishing statistics related to the country’s economy, population, and society, at the national and local levels.

¹³¹ Examples of evaluations in which one of the authors of the chapter was directly involved.

¹³² More on the cited approach in impact evaluations using data from public statistics is presented in the chapter *Past experience and future perspectives of the GUS on the use of administrative data in enterprise surveys.*

An example of a study that carried out counterfactual analyses on GUS data was the *Mid-term evaluation of PARP state aid under the Operational Programme Smart Growth 2014-2020 (POIR)* (IDEA Institute, IDEA Development Foundation, and Jagiellonian University commissioned by PARP, 2020). The study was conducted using i.a. data collected in the Annual Survey of Enterprises (SP GUS form). The structure of this questionnaire includes, i.a. Part II entitled *Balance sheet and profit and loss account*, which has been adapted to the financial statements template (Accounting Act, 1994, Appendix 1). The questionnaire is filled in by enterprises (excluding entities employing up to nine persons) and the obtained micro data are used by GUS, as a rule, for the implementation of statutory tasks (i.a. Public Statistics Research Programme). Pursuant to the legislation in force “no identifiable micro data obtained in statistical surveys may be published or made available” (Act on Public Statistics, 1995, art. 38, par. 1). Consequently, within the POIR evaluation given as an example, all calculations on micro data had to be carried out by the GUS according to the previously mentioned model of cooperation with the evaluator, i.e.:

- GUS was tasked with calculating the values of specific indicators for a set group of enterprises (beneficiaries, and enterprises from the potential control pool) on the basis of data from the annual reporting for public statistics (e.g. Annual Survey of Enterprises);
- GUS verified the availability (coverage) of data for individual entities, and carried out the calculation procedure on the enterprises for which data were available (e.g. calculation of similarity index – propensity score, matching of the control sample, calculation of average values of outcome indicators, according to the order, etc.), using for this purpose specific queries to the reporting database (script) provided by the ordering party/evaluator;
- GUS made the results of its calculations available to the commissioning party in aggregated form – without the possibility to review micro data.

Such a solution makes it possible to conduct counterfactual analyses, but the lack of direct access to micro data results in a number of limitations. It remains a statistical mystery for which units such calculations are made¹³³.

¹³³ In another analytical project of this type, public statistical services decided to calculate the value of effect indicators on a sample drawn from a given group of enterprises. Thus, the principle of confidentiality was extended to those included in the sample as a result of the draw, as well as those included in the draw, in connection with the classified coverage of the set group of enterprises of the GUS reporting data resources (Pokorski, 2020; Koniewski, Krupnik, 2017).

Another serious limitation that should be mentioned is the fact that analyses of this type can be conducted on historical data only. Considering the above-mentioned observations of the POIR evaluation, public statistics' readiness to carry out analyses using data from the Annual Survey of Enterprises occurs up to two years after the end of the reporting period to which the data relates. In this case, analysis of the effects of POIR support up to and including 2018 (mid-term) was possible in the first quarter of 2020 at the earliest. Given the dynamics of socio-economic phenomena (including the coronavirus pandemic or the global financial crisis), the above approach and data source can be considered as far from optimal.

While addressing the above limitations, without elaborating on other restrictions that may apply to the public statistics data and, at the same time, bearing in mind their ultimate value, it is worth noting that the GUS is not the only provider of financial data reported by enterprises in Poland. As previously indicated, data reported in the Annual Survey of Enterprises are relatively consistent with financial statements filed by enterprises under the Accounting Act (Accounting Act, 1994). Depending on the type of entity, financial statements are submitted to the National Revenue Administration and the National Court Register (KRS). Both entities collect detailed and non-aggregated data on the financial situation of economic entities in Poland, i.e. situation of those entities which are obliged by the law to submit financial statements. The publicly available KRS resources are particularly interesting in the context of evaluations. Any interested party has access to the individual, non-aggregated and non-anonymised corporate financial statements collected by the KRS. This data has been available for a long time. Since 2018, financial statements have been filed only in electronic form and made available through the free and publicly accessible Financial Document Browser, hosted by the Ministry of Justice at <https://ekrs.ms.gov.pl/> (hereinafter: EKRS Financial Document Browser).

In the case of the aforementioned GUS data source (Annual Survey of Enterprises), in 2021, it would have been impossible to carry out analyses for the purpose of programme evaluation including data for 2020, a period affected by various types of pandemic-related restrictions in the functioning of the economy. However, if KRS data were used, analyses turned out to be possible as early as October 2021.¹³⁴

¹³⁴ In principle, this would have been possible even earlier, however, due to the pandemic mitigation regulations, businesses were given the opportunity to submit their 2020 financial statements at a later date than in the previous years.

10.2. Description of the EKRS Financial Document Browser

Entity scope

The EKRS Financial Document Browser provides access to financial statements of all entities in Poland entered in the Register of Entrepreneurs of the National Court Register (in accordance with the provisions of the Act of 1997 on the National Court Register) which are obliged to prepare financial statements in accordance with national accounting standards (in accordance with the provisions of the Accounting Act of 1994). A detailed list of the types of entities required to be entered in the Register of Entrepreneurs is contained in Article 36 of the Act on the National Court Register. These include:

- general partnerships;
- limited liability partnerships;
- limited partnerships;
- partnerships limited by shares;
- limited liability companies;
- joint-stock companies.

As a rule, the obligation to prepare financial statements in accordance with national accounting standards covers all entities entered in the Register of Entrepreneurs of the National Court Register. The exceptions are as follows:

- general partnerships and limited liability partnerships with low turnover whose net revenues from sales of goods, products, and financial operations for the previous financial year amounted to less than EUR 2 million;
- entities preparing financial statements in accordance with International Accounting Standards (IAS) – this applies to issuers of securities and members of capital groups (if

the parent company prepares financial statements in accordance with IAS), and branches of foreign entities (if the entity prepares financial statements in accordance with IAS).¹³⁵

Content of financial statements

The content of financial statements varies depending on enterprise size. This applies to both the elements of financial statements required under the Accounting Act, and their level of detail. The basic elements of financial statements required from all enterprises listed in the previous section, are:

- profit and loss account;
- balance sheet;
- additional information (including, i.a. introduction, additional information, and explanations).¹³⁶

Additionally, joint-stock companies and entities that meet at least two of the three following conditions:

- 1) entities with average annual employment of at least 50 persons¹³⁷ (converted into full time equivalents),
- 2) entities with total balance sheet assets at the end of the financial year of at least EUR 2.5 million,
- 3) entities with net revenue for the financial year of at least EUR 5 million,

must include: (4) a statement of changes in equity (fund) and (5) a cash flow statement.

The required level of detail for the profit and loss account, balance sheet, and additional information is different for micro, small, and other enterprises. The basic level of detail of the

¹³⁵ Review of the substantive scope and assessment of the analytical potential of data collected on the basis of IAS-compliant statements (for entities operating in Poland e.g. as part of a foreign group) is beyond the scope of this chapter. However, this type of reporting data can be a valuable source of information (e.g. in relation to analyses of the issue of internationalisation of the economy) and is worth bearing in mind in future research for the management processes of economic development policies (including diagnosis, monitoring, evaluation).

¹³⁶ Due to the scope of this element, it will not be analysed further.

¹³⁷ The adopted headcount is a criterion that would suggest the classification of these entities as medium-sized enterprises, however, the remaining financial criteria (turnover and assets) remain incompatible (lower) with the classification for this size of enterprise, according to the Enterprise Law (Law of 6 March 2018) and Appendix 1 to Commission Regulation (EU) No 651/2014 of 17 June 2014 (European Commission, 2014).

listed elements was defined in Appendix 1 to the Accounting Act. Micro and small enterprises may prepare them in a simplified form, whose scope is specified in Appendix 4 for micro enterprises, and Appendix 5 for small enterprises.¹³⁸

Simplified rules for micro and small enterprises may be applied by entities that, in the financial year for which they prepare financial statements, met at least two of the following three conditions (see Table 1). In addition, the Act allows for the possibility of applying simplifications if the entity fulfilled a minimum of one of the following conditions and applied simplified principles in its previous financial statements.

It should be pointed out that the definition of micro and small enterprises differs between the Accounting Act and the Enterprise Law (see Table 2). Classification is based on the same indicators, however, in the Enterprise Law, significantly higher values were determined for total assets and net income.¹³⁹ From the analytical point of view, the lack of harmonization, between the Accounting Act and the Enterprise Law, of currency units (PLN vs. EUR) in which the conditions for classification to a given category of enterprises were determined, is also significant. On the other hand, from the point of view of monitoring and evaluation processes, which are usually carried out for specific support programmes, the conditions for classifying SMEs, resulting from the EU state aid rules, are of greater importance (European Commission, 2014 b). In the case of classification criteria for enterprise size, the Enterprise Law is to some extent¹⁴⁰ harmonised with them (including using the EUR as currency unit).

Table 1. Conditions for categorizing an enterprise as micro or small – Accounting Act

Indicator	condition for micro	condition for small
balance sheet total at the end of the financial year	not more than PLN 1.5 million	not more than PLN 25.5 million
net revenue for the financial year	not more than PLN 3 million	not more than PLN 51 million
average annual employment (converted into full time equivalents)	not more than 10	not more than 50

Source: own study based on the Accounting Act of 29 September 1994.

¹³⁸ The Act also defines the scope for banks (Appendix 2), insurance and reinsurance companies (Appendix 3), and public benefit organisations (Appendix 6). Considering the subject matter of the chapter, discussion of these scopes is omitted.

¹³⁹ In both the Accounting Act and the Enterprise Law, a company must meet a minimum of two of the three designated conditions, with the employment condition being mandatory in the case of Enterprise Law.

¹⁴⁰ Cf. the category of *affiliated enterprises* (European Commission, 2014 b).

Table 2. Conditions for categorizing an enterprise as micro or small – Enterprise Law

Indicator	condition for micro	condition for small
balance sheet total at the end of the financial year	not more than EUR 2 million	not more than EUR 10 million
net revenue for the financial year	not more than EUR 2 million	not more than EUR 10 million
average annual employment (converted into full time equivalents)	not more than 10	not more than 50

Source: own study based on the Enterprise Law of 6 March 2018.

It should be noted that a company's use of the simplifications provided for in the Act is optional. In practice, some small and micro enterprises report without using the simplifications.

A detailed description of the availability and content of the particular elements of financial statements will be provided in the next section. At this point, implied correctness of the data contained in all elements of financial statements should be indicated as a common feature. Both the Accounting Act and the Act on the National Court Register require entrepreneurs to ensure that the data shown in their financial statements is factually correct.

Time scope

According to the provisions of the Accounting Act, financial statements must be filed with the KRS within 15 days of their approval by the enterprise. This means that if financial statements cover the entire calendar year, they must be submitted to the National Court Register no later than by 15 July (financial statements must be approved no later than by 30 June of the year following the reporting year). Exceptions to this rule are 2020 and 2021. Due to the coronavirus pandemic, the deadline was extended to 15 October (Ministry of Finance, 2020).

A financial statement contains data for two years. Within each element (e.g. balance sheet or profit and loss account), data for the current and previous financial year are presented.¹⁴¹ In

¹⁴¹ If the reporting period does not overlap with the calendar year: for the profit and loss account and cash flow statement, data are presented for the current reporting period and the corresponding reporting period of the previous financial year; for the balance sheet and statement of changes in equity, data are presented for the current reporting period (as of the end of the reporting period) and the previous financial year.

other words, a statement filed on 15 October 2021 presents financial data as of 31 December 2020 and 31 December 2019.

The obligation to prepare financial statements in electronic form only has been in force since 1 October 2018. As a result, data resources currently stored in the EKRS Financial Document Browser include data for 2017-2020.

Access to resources

Pursuant to the provisions of the Act on the National Court Register (Article 8 and Article 9a), information contained in the Register of Entrepreneurs (including financial statements) is public and access to it is unlimited, unrestricted, and free of charge. Financial statements are made available through the EKRS Financial Documents Browser where, for a given KRS number of a particular company, one can search and download selected documents. Importantly, the system does not automatically enable the download of data for a given set of enterprises. If data for multiple enterprises are required, the query must be run independently for every enterprise.

Financial statements are made available in XML format files – structurally organized text files. A single file contains whole financial statements – all the information contained in the profit and loss account, the balance sheet, and other required elements. The system does not enable downloading the value of a selected indicator (e.g. salaries from the profit and loss account) for a given group of enterprises. In order to analyse data for a group of companies, it is necessary to download financial statements for all the analysed companies (as separate files), search for the required indicators, and then combine them into a table containing the required data. It is worth noting that the process can be automated using Python (Selenium and BeautifulSoup libraries).

10.3. Characteristics of the particular elements of financial statements

This section will present the scope of data available in the profit and loss account, the balance sheet, the statement of changes in equity, and the cash flow statement. The lists of

required indicators¹⁴² will be presented in the form of diagrams illustrating the dependencies between them. Within each element of financial statements, variants permitted by the Act will be analysed – differing in the list of indicators required from the entrepreneur. As stated above, the level of detail may vary, depending i.a. on enterprise size:

- basic scope shall be understood as the scope defined in Appendix 1 to the Accounting Act;
- simplified scope for small companies shall be understood as the scope defined in Appendix 5 to the Accounting Act;
- simplified scope for micro companies shall be understood as the scope defined in Appendix 4 to the Accounting Act.

Moreover, availability of particular elements of financial statements will be analysed. An example of a point of reference will be data of applicants, i.e. all applications for co-financing submitted by enterprises until August 2021 under the assistance measures of PARP as part of the Operational Programme Eastern Poland 2014-2020 – POPW.01.02, POPW.01.03.01, and POPW.01.04 (3,308 in total, including 1,433 successful applications where a co-financing agreement was concluded, and 1,875 unsuccessful applications).

The unit of analysis will be the application for co-financing¹⁴³. Information will be presented on the number of applications for co-financing submitted by enterprises for which data from financial statements for the selected year were obtained in November 2021 using the EKRS Financial Document Browser (the year 2020 was discussed as an example)¹⁴⁴. The analysis below also presents the scope of data obtainable from the discussed source, potentially available for research concerning other topics – e.g. other aid programmes for entrepreneurs.

¹⁴² The lists of indicators required in the respective elements of financial statements are defined in Appendices 1, 4, and 5 to the Accounting Act.

¹⁴³ One company could submit several applications, including under different measures.

¹⁴⁴ For the purposes of the analysis, all available annual financial statements (for all available years) were searched and downloaded from the EKRS Financial Document Browser for beneficiaries with a KRS number. Data were collected for 970 companies. These were derived from 2,680 financial statements. Data specific to individual elements of financial statements were extracted from the downloaded files. Separate databases were created containing information from: (1) profit and loss accounts; (2) balance sheets; (3) statements of changes in equity; (4) cash flow statements. The records in the databases were single financial values – value of a given indicator declared by one enterprise in a given year. The profit and loss account database contains a total of 198,055 records; balance sheets – 535,198; statements of changes in equity – 72,541; cash flow statements – 84,280.

Profit and loss account

Data availability – based on the example of POPW applicants

The EKRS Financial Document Browser provides access to the 2020 profit and loss accounts for enterprises that submitted 44% of the analysed applications for co-financing (1,439 out of 3,308). It is important to bear in mind that the point of reference here are all the applications for co-financing submitted by both companies covered and not covered by the obligation to prepare financial statements and submit them to the National Court Register¹⁴⁵.

Regardless of the level of detail in the profit and loss account, which depends on company size and is described above, under the Accounting Act, profit and loss account can be prepared in either the comparative or calculation variant¹⁴⁶. An exception is the simplified variant for micro companies, which may be prepared in the comparative variant only. This means that the available profit and loss accounts were prepared in a total of 5 variants:

- basic scope in the calculation variant;
- basic scope in the comparative variant;
- simplified scope for small companies in the calculation variant;
- simplified scope for small companies in the comparative variant;
- simplified scope for micro companies in the comparative variant.

Detailed structure of the available profit and loss accounts for POPW applicants, including both their scope and variant, is presented in the table below.

¹⁴⁵ This obligation does not apply to natural persons conducting business activity, and partnerships. Such enterprises submitted a total of 1,283 applications for co-financing. In addition, some enterprises are excluded from the said obligation on the basis of exceptions provided for in the Accounting Act (see Entity scope) – e.g. due to low turnover (unfortunately, the available data do not make it possible to determine the number of enterprises that are subject to exemption due to the level of turnover).

¹⁴⁶ The variants differ in the first part of the study, which concerns determination of the profit on sales. The comparative variant groups costs by their type (e.g. consumption of materials and energy, remuneration), while the calculation variant groups costs by their place of origin (e.g. sales costs, overheads).

Table 3. Availability of 2020 profit and loss account data for applications for co-financing (WoD) submitted under selected POPW¹⁴⁷

Measure	Total WoD	Number of WoD for which profit and loss account data are available	basic scope			simplified scope for small companies			simplified scope for micro companies		
			calculation variant	comparative variant	Total	calculation variant	comparative variant	total	calculation variant	comparative variant	total
POPW.01.02.00	1093	455 (42%)	38	265	303	13	105	118	0	34	34
agreement concluded	424	196 (46%)	16	122	138	3	43	46	0	12	12
agreement not concluded	669	259 (39%)	22	143	165	10	62	72	0	22	22
POPW.01.03.01	505	250 (50%)	24	155	179	8	57	65	0	6	6
agreement concluded	175	92 (53%)	11	62	73	3	15	18	0	1	1
agreement not concluded	330	158 (48%)	13	93	106	5	42	47	0	5	5
POPW.01.04.00	1.710	734 (43%)	86	432	518	36	159	195	0	21	21
agreement concluded	834	377 (45%)	56	227	283	17	70	87	0	7	7
agreement not concluded	876	357 (41%)	30	205	235	19	89	108	0	14	14
Total	3,308	1439 (44%)	148	852	1000	57	321	378	0	61	61

Source: own study based on data from LSI PARP and EKRS Financial Document Browser

Data scope

Due to the large number of profit and loss account variants, comparability of data reported by different entities is limited. Figure 1 presents a comparison of indicators required in each variant. The indicators common to all enterprises (regardless of the reporting option) are:

- net revenues from sales;
- operating expenses¹⁴⁸;

¹⁴⁷ An application for co-financing is the unit of analysis – one company could submit several applications, including under different measures.

¹⁴⁸ Assuming that operating costs in the comparative variant are equal to the value of products, goods and materials sold, sales costs and overheads in the calculation variant.

- profit (loss) on sales;
- other revenue;
- other costs;
- income tax;
- net profit (loss).

If statements prepared using the calculation variant are excluded from the analysis, the list of common indicators is extended to include:

- depreciation and amortisation;
- consumption of materials and energy;
- remuneration, social security, and other benefits;
- other (operating) costs.

It should be noted that among the analysed enterprises, profit and loss accounts using the comparative variant are by far predominant in the statements. As a result, the above-mentioned four additional indicators are available for enterprises that submitted 37% of the analysed applications for POPW co-financing (1,234 out of 3,308).

Legend to Figure 1

Green – the indicator is present in all variants (in the basic scope in the calculation and comparative variant, in the simplified scope for small companies in the calculation and comparative variant and in the simplified scope for micro companies in the comparative variant). Exception: indicators included in Operating expenses are only available in the profit and loss account prepared under the comparative variant.

 – indicators surrounded by a dashed line in the simplified variant for micro companies have been combined into one

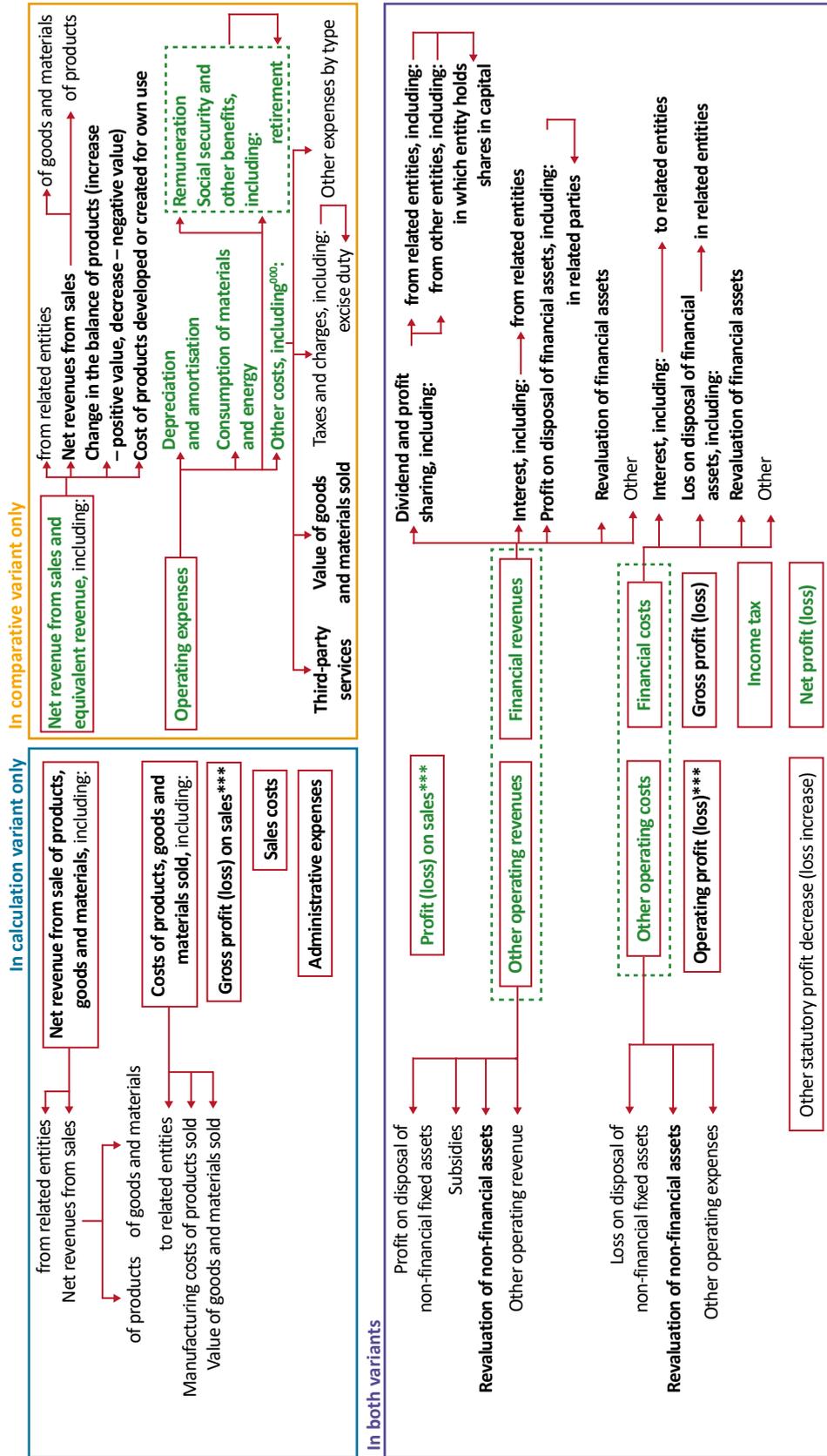
Bold – the indicator is present in the simplified scope for small companies and in the basic scope

Normal font – the indicator is present only in the basic scope

*** – the indicator has not been listed literally in the simplified scope (for micro or small companies) but its value can be calculated on the basis of other required indicators

°°° – aggregating indicator added to maintain comparability between basic and simplified scopes

Figure 1. Profit and loss account information set – basic and simplified scope for small and micro enterprises



Source: own study based on Appendices 1, 4 and 5 to the Accounting Act of 29 September 1994

Balance sheet

Data availability

The availability of balance sheet data is analogous to availability of profit and loss account data. The EKRS Financial Document Browser enabled access to balance sheet data of enterprises responsible for 44% (1,439 out of 3,308) of the analysed applications for co-financing. Balance sheets were prepared in three variants – basic, simplified for small enterprises, and simplified for micro enterprises. It should be noted that, as in the case of profit and loss account, the majority of balance sheets were prepared using the basic variant. They constituted 68% (990 out of 1,439) of the available balance sheets. The table below presents the detailed structure of the available balance sheets.

Table 4. Availability of 2020 balance sheet data for applications for co-financing (WoD) submitted under selected POPW¹⁴⁹

Measure	Total WoD	Number of WoD for which balance sheet data are available	basic scope	simplified scope for small companies	simplified scope for micro companies
POPW.01.02.00	1093	455 (42%)	294	127	34
agreement concluded	424	196 (46%)	136	48	12
agreement not concluded	669	259 (39%)	158	79	22
POPW.01.03.01	505	250 (50%)	179	65	6
agreement concluded	175	92 (53%)	73	18	1
agreement not concluded	330	158 (48%)	106	47	5
POPW.01.04.00	1710	734 (43%)	517	196	21
agreement concluded	834	377 (45%)	283	87	7
agreement not concluded	876	357 (41%)	234	109	14
Total	3308	1439 (44%)	990	388	61

Source: own study based on data from LSI PARP and EKRS Financial Document Browser

¹⁴⁹ An application for co-financing is the unit of analysis – one company could submit several applications, including under different measures.

Data scope

The balance sheet consists of two parts – assets and liabilities. Figure 2 and Figure 3 present detailed lists of indicators contained in both parts of the balance sheet (covering the basic and simplified variants).

The list of indicators available in all balance sheet types (basic and simplified scopes) is limited. The available indicators are aggregated values of: (1) fixed assets, (2) current assets (including itemised values of inventories and short-term receivables), (3) own shares (equities), (4) called up share capital, (5) equity (including share capital), and (6) liabilities and provisions for liabilities (including provisions for liabilities and itemised values of credit facilities and loans).

However, it is worth noting that in the basic and simplified scope for small enterprises, the list of available indicators is much more extensive. This applies in particular to the basic scope, under which entrepreneurs are required to disclose, i.a.:

- costs of completed development works;
- (value of) technical devices and machines;
- (value of) real estate, including land and buildings;
- (value of) goods;
- short-term credit facilities and loans;
- long-term credit facilities and loans.

As noted earlier, among the financial statements obtained from the EKRS Financial Document Browser for the given sub-set of companies (POPW applicants), balance sheets in the basic variant prevail. As a result, values of the six indicators listed above are available for companies responsible for as much as 30% of applications for co-financing submitted under the analysed POPW measures.

Legend to Figure 2 and Figure 3

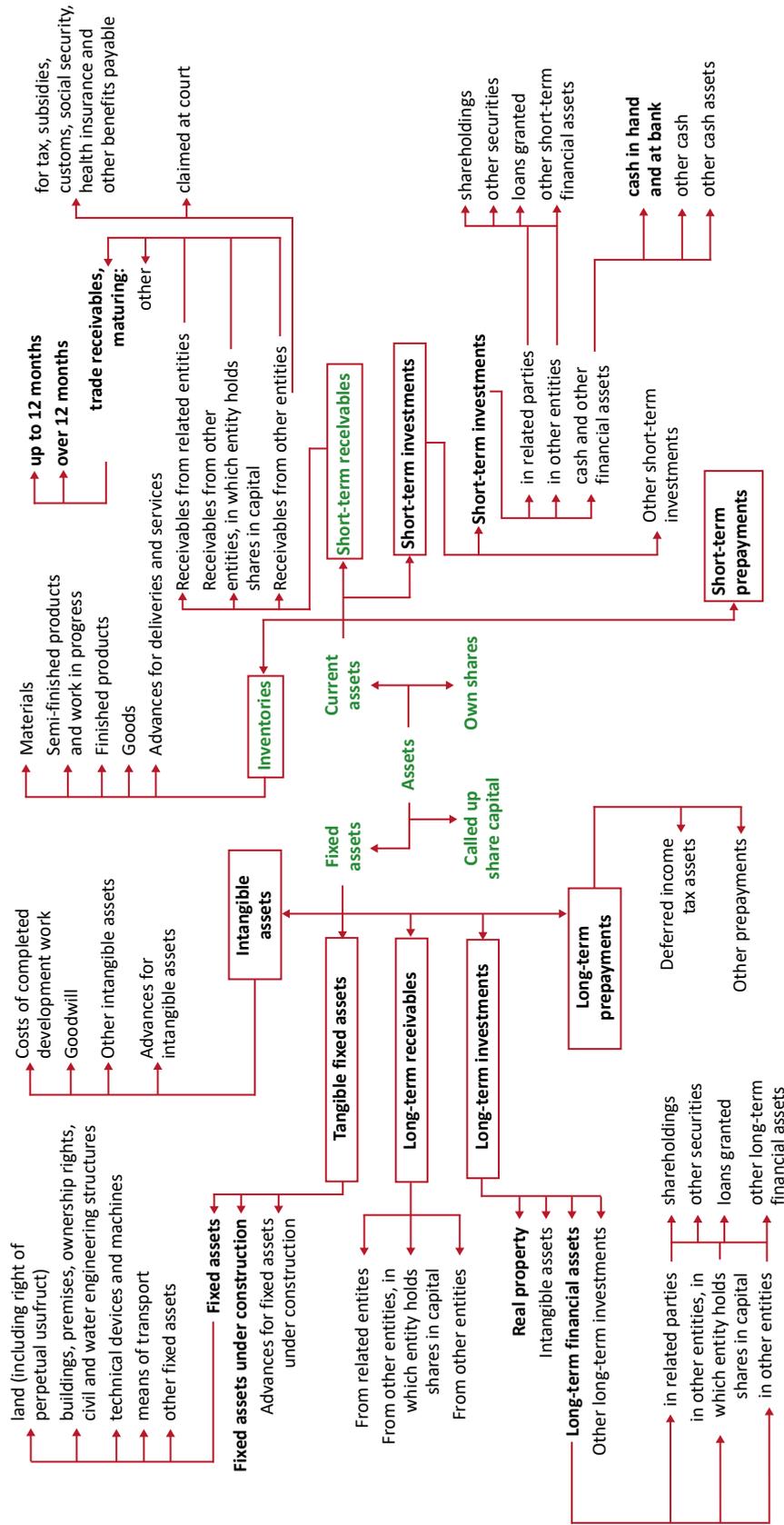
Green – the indicator is present in all variants (in the basic scope, simplified scope for small companies, and simplified scope for micro companies).

 – indicators surrounded by a dashed line in the simplified variant for micro companies have been combined into one

Bold – the indicator is present in the simplified scope for small companies and in the basic scope

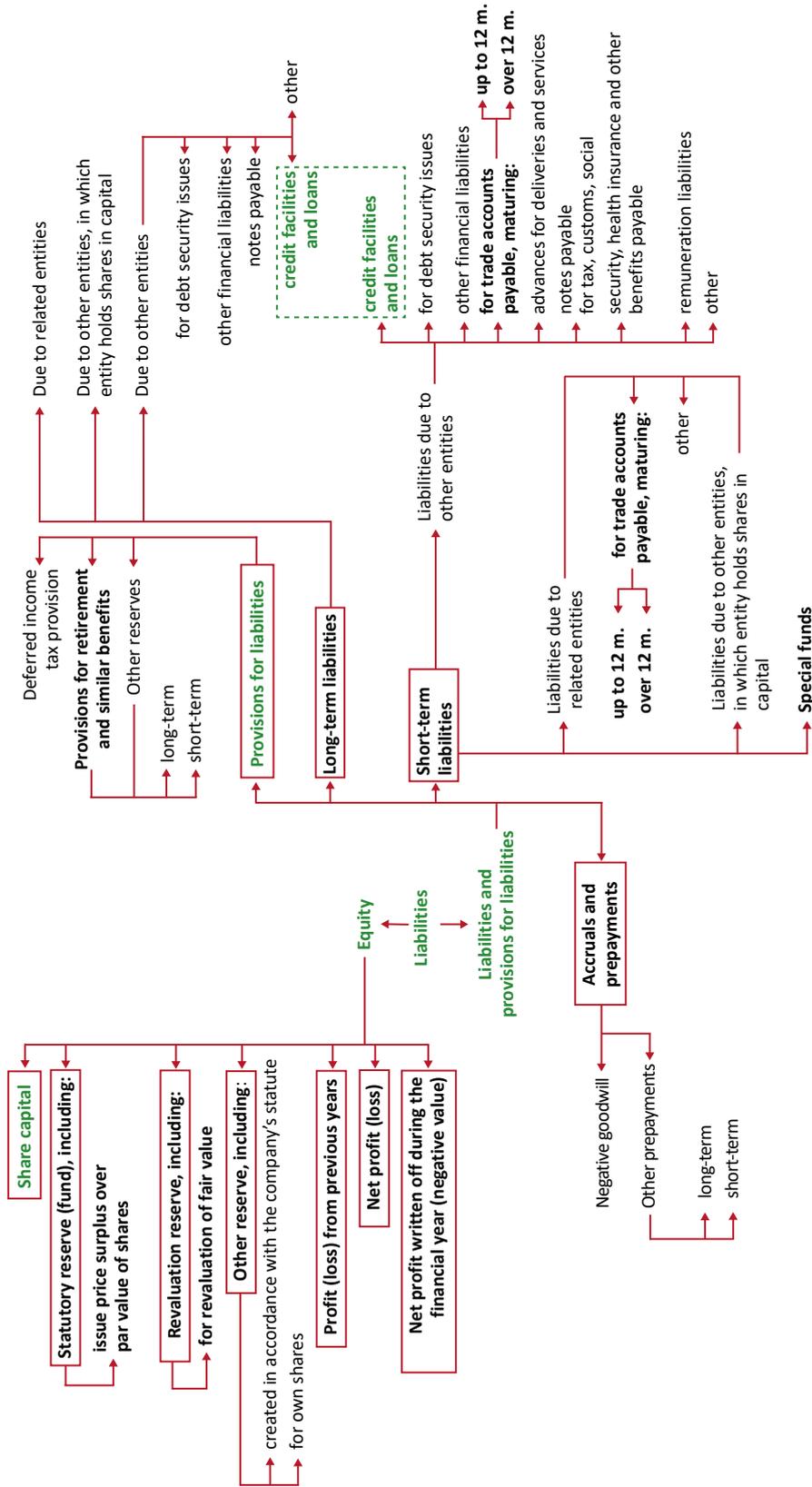
Normal font – the indicator is present only in the basic scope

Figure 2. Balance sheet information set, assets – basic and simplified scope for small and micro enterprises



Source: own study based on Appendices 1, 4, and 5 to the Accounting Act of 29 September 1994

Figure 3. Balance sheet information set, liabilities – basic and simplified scope for small and micro enterprises



Source: own study based on Appendices 1, 4, and 5 to the Accounting Act of 29 September 1994

Statement of changes in equity

Data availability

The obligation to prepare a statement of changes in equity was imposed on a relatively small group of enterprises (as described above). This results in a significantly lower availability (compared to that of profit and loss account and balance sheet) and, consequently, lower usefulness of data for evaluation processes. The EKRS Financial Document Browser provides data from the 2020 statement of changes in equity for enterprises responsible for 16% (536 out of 3,308) of the analysed applications for co-financing. It should be noted that this statement is only submitted in one variant and no simplifications can be applied. As a result, all the available statements of changes in equity are fully comparable. The table below presents the detailed structure of the available statements.

Table 5. Availability of 2020 statement of changes in equity data for applications for co-financing (WoD) submitted under selected POPW¹⁵⁰

Measure	Total WoD	Number of WoD for which statements of changes in equity data are available
POPW.01.02.00	1093	129 (12%)
agreement concluded	424	60 (14%)
agreement not concluded	669	69 (10%)
POPW.01.03.01	505	111 (22%)
agreement concluded	175	52 (30%)
agreement not concluded	330	59 (18%)
POPW.01.04.00	1710	296 (17%)
agreement concluded	834	177 (21%)
agreement not concluded	876	119 (14%)
Total	3308	536 (16%)

Source: own study based on LSI PARP and EKRS Financial Document Browser

Data scope

In line with the above observation, all statements of changes in equity follow a uniform scheme. The figure below shows the list of the available indicators. As the name suggests, the scope of data within this element of financial statements concerns changes in a company's equity.

¹⁵⁰ An application for co-financing is the unit of analysis – one company could submit several applications, including under different measures.

Cash flow statement

Data availability

The level of availability of data from cash flow statements is the same as in the case of statements of changes in equity. The EKRS Financial Document Browser provides data from cash flow statements from 2020 for entrepreneurs who submitted 16% (508 out of 3,308) of the analysed applications for co-financing. The Act does not provide for simplifications based on company size, while allowing for two variants differing with regard to the analysis of cash flows from operating activities. It should be noted that among the analysed financial statements, cash flow statements were prepared almost exclusively using the so-called intermediate variant. The table below presents the detailed structure of the available cash flow statements for a given subset of enterprises.

Table 6. Availability of 2020 cash flow statements data for applications for co-financing (WoD) submitted under selected POPW¹⁵¹

Measure	Total WoD	Number of WoD for which cash flow statements data are available	direct variant	intermediate variant
POPW.01.02.00	1093	128 (12%)	4	124
agreement concluded	424	59 (14%)	2	57
agreement not concluded	669	69 (10%)	2	67
POPW.01.03.01	505	106 (21%)	1	105
agreement concluded	175	50 (29%)	0	50
agreement not concluded	330	56 (17%)	1	55
POPW.01.04.00	1710	279 (16%)	0	279
agreement concluded	834	167 (20%)	0	167
agreement not concluded	876	112 (13%)	0	112
Total	3308	513 (16%)	5	508

Source: own study based on LSI PARP and EKRS Financial Document Browser

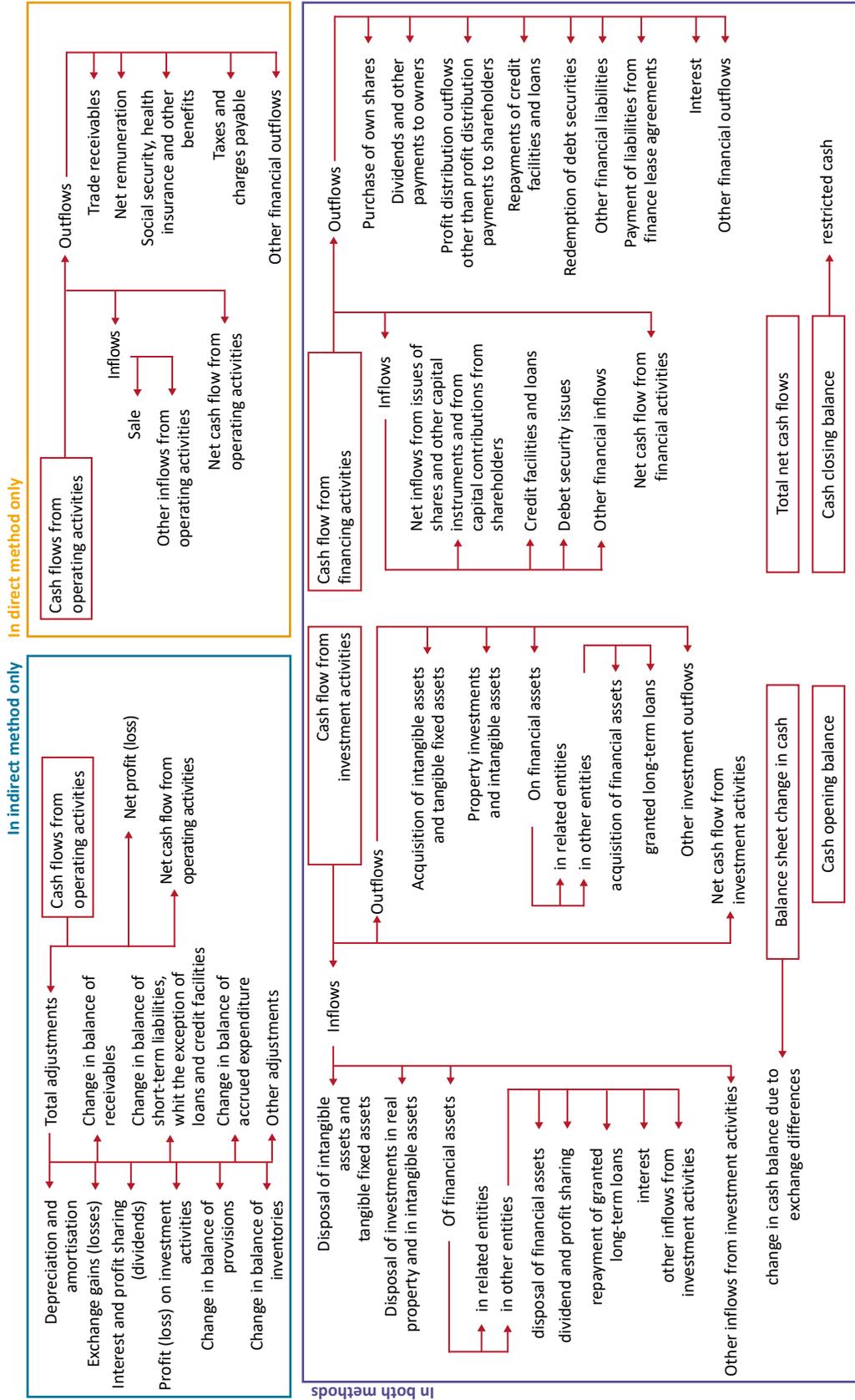
¹⁵¹ An application for co-financing is the unit of analysis – one company could submit several applications, including under different measures.

Data scope

Cash flow statements provide information on enterprises' inflows and outflows during the reporting period – for cash flows arising from both operating and other activities. Figure 5 presents the full list of indicators included in the two variants permitted under the Act. Due to the fact that most enterprises prepare cash flow statements using the indirect method, the available scope does not include detailed data on operating expenses – including, e.g. expenses on supplies and services, remuneration, and social security (data required under the direct method). Nevertheless, it is worth noting a number of indicators potentially useful for evaluations, required in both variants of the remaining part of the study, including i.a.:

- expenditure on the acquisition of intangible and tangible fixed assets;
- expenditure on property investments and intangible assets;
- proceeds from credit facilities and loans;
- expenditure on repayment of credit facilities and loans.

Figure 5. Set of information included in the cash flow statement



Source: own study based on Appendix 1 to the Accounting Act of 29 September 1994

10.4. Use of KRS data in impact evaluation practices

Assumptions for the analysis

There are several reasons why PARP reached for data from companies' financial statements in the evaluation of the impact of the POPW. On the one hand, this resulted from the need to obtain quick access to enterprises' up-to-date financial data. On the other hand, it was aimed at ensuring triangulation of sources of information (Nowińska) in relation to earlier analyses conducted by PARP in a similar scope, where other data were used (Annual Survey of Enterprises – SP GUS).

In connection with the process of programming of the new perspective for the use of EU funds in Poland, in particular the successor programme Eastern Poland for 2021-2027, the European Commission and the POPW Managing Authority expected that conclusions of such counterfactual analysis would be available for the process. This requirement was to be fulfilled before the formal start of negotiations of the new programme. In relation to the study conducted on GUS data, the analysis was intended to constitute “a complementary source of information and empirical evidence indicating whether the companies-beneficiaries of selected POPW measures would have made investments in the areas covered by the above-mentioned measures if they had not received a grant from the funds of POPW 2014-2020. The conclusions (...) could serve as one of the premises for the continuation of analogous co-financing support for companies of the macro-region under the new POPW for 2021-2027” (PARP, 2021, p. 2).

One of the key tasks of evaluation is to determine the effects of the implemented programmes. In this case, this task was to focus on selected support measures of the POPW for micro, small, and medium-sized enterprises. Its performance was to help answer the general question whether the existing support provided under the Programme in the Eastern Poland macro-region really works (i.e. provides an incentive for investing in innovations¹⁵²,

¹⁵² “The key research question of this analysis (...) concerns verification of the so-called *incentive effect* of the support, i.e. answer to the question whether the beneficiaries would have made investments in the supported areas if they had not received POPW 2014-2020 co-financing (in other words, whether public support was a vital factor in entrepreneurs' decision to implement the said investments)” (PARP, 2021, p. 8).

is effective and solves the diagnosed socio-economic problems, etc.). The answer required the estimation of the so-called causal effect of the programme (net), i.e. taking into account causal dependencies of the impact of public intervention on changes in the situation of beneficiaries. Identifying the existence and strength of causal relations is crucial from the point of view of monitoring and improving public interventions financed under the EU cohesion policy, which was also emphasised by the Commission services in relation to PARP interventions under the POPW implemented in the macro-region.

Modern methods used to estimate the impact of interventions focus on the notion of causality, defined using the so-called concept of counterfactual states. Programme effects are estimated by comparing the situation of beneficiaries with the hypothetical situation which would have occurred if they had not received support. The hypothetical situation is impossible to observe in reality, but may be estimated thanks to the application of various statistical methods and techniques, and the use of entities which did not receive co-financing (control group). In general, at the level of research methodology, the counterfactual approach boils down to comparisons of the situation of beneficiaries of a given support with an appropriately selected control group. On the one hand, implementation of counterfactual analyses in measuring effects arises from the contemporary trends in the methodology dedicated to identifying causal dependencies. On the other hand, it has its source in the expectations of the European Commission towards all Member States spending EU funds and implementing support programmes. In this case, the analysis was intended to provide knowledge for more effective implementation of public policies for the development of entrepreneurship in the Eastern Poland macro-region (PARP, 2017, 2021).

While the first analysis on SP GUS data provided valuable observations and conclusions, it did not cover a period in which the recorded trends could have been completely reversed due to the constraints regarding the functioning of the economy during the pandemic. Repeated analyses on more recent GUS data were also limited by the time in which the survey had to be conducted. Therefore, as an experiment, an attempt was made to use KRS data, which could provide data including for the last financial year – first year of the pandemic (2020). Furthermore, company data extracted from the EKRS Financial Document Browser covered micro-enterprises, which, as a rule, were not among entities reporting under the SP GUS form. In the case of analyses based on data from public statistics, an important negative factor was also the fact that, based on the observed coverage of applicants' data in the GUS sets, in the case of the POPW, it was not possible to use the so-called unsuccessful applicants

as a control group (as a rule, entities significantly similar to entities that received support – i.e. companies from the macro-region interested in investing in a particular subject area in the period around Programme beneficiaries' investments). The control group consisted of entities that were similar (according to statistical criteria), but completely unrelated to the Programme. Besides, they did not necessarily originate from the Eastern Poland macro-region. When KRS data were used, counterfactual analyses within the group of applicants (successful vs. unsuccessful) proved possible. In addition, sub-group analyses were carried out, by type of POPW intervention (pro-innovation measures POPW 1.3.1 and 1.4 vs. pro-export measure POPW 1.2). "The control group consisted of unsuccessful applicants, i.e. entities which applied for support under the analysed instruments but did not receive it. (...) Each time, the entities were selected statistically, taking into account a set of specific characteristics. Additionally, the effects were calculated in the form of double difference (difference in differences, DiD), i.e. estimation of change in year A+2 and A+3, in relation to the year preceding the submission of the application (A-1) in the group of beneficiaries and comparing it with an analogous estimation made in the control group" (IDEA Institute and IDEA Development Foundation commissioned by PARP, (2021), p. 5). Ultimately, the analysis covered entities that submitted applications for co-financing between 2015 and 2018. Balance sheet data and data from profit and loss accounts of applicants covering the period before their availability in the EKRS database (2015-2016) were taken from the PARP Local IT System (LSI). As a rule, to the application for co-financing submitted in a given year, the entrepreneur also attached financial statements for the previous years¹⁵³.

¹⁵³ It is worth emphasising that, along with historical statements (balance sheet, profit and loss account), applicants also submitted a financial forecast referring to indicators from these statements, which included values of indicators expected by applicants following project implementation. During the analyses in question, based on KRS data, these forecasts, which entrepreneurs formulated ex ante for the following years, were compared with the actual results achieved in these years. The approach was also innovative. It had never been used before in evaluation studies in Poland, however, presenting it in more detail is beyond the scope of this chapter. More on this subject can be found in the report (IDEA Institute and IDEA Development Foundation commissioned by PARP, 2021).

Table 7. Analyses of relative effects of POPW from the year before the application for co-financing (A-1) to the second and third year after (A+2, A+3)

Application year	Reporting period						
	2014	2015	2016	2017	2018	2019	2020
2015	Minus 1	Application	Plus 1	Plus 2	Plus 3	Plus 4	Plus 5
2016	n/a	Minus 1	Application	Plus 1	Plus 2	Plus 3	Plus 4
2017	n/a	n/a	Minus 1	Application	Plus 1	Plus 2	Plus 3

Source: IDEA Institute and IDEA Development Foundation study based on PARP data

It was unreasonable to include in the analysis applications from later years, because, for more recent applications, it would be impossible to analyse the effects using financial statements from the KRS. For instance, in order to maintain a uniform approach, for applications submitted in 2019 it would be necessary to provide financial data of companies from the period 2018-2022 (from A-1 to A+3) or 2018-2021 (up to A+2), while the analysis was carried out between November and December using data extracted in October 2021 (most recent data was for 2020).

Table 8. Distribution of numbers of applications and beneficiaries of the analysed POPW measures

Measure	Total applications ¹⁵⁴	Including number of applications submitted by 2018	Including number of agreements (not terminated)	Including unique beneficiaries
1.2	1054	661	172	136
1.3.1	478	288	88	77
1.4 (phase I)	1079	830	379	292
1.4 (phase II)	558	186	80	77
Total	3169	1965	719	582 (444)*

Source: IDEA Institute and IDEA Development Foundation study based on PARP data

*Some entities signed agreements under more than one measure. A unique number of entities was provided in brackets, irrespective of the measure from which the company benefited

¹⁵⁴ As of the end of August 2021, 139 applications for which co-financing agreements had been terminated were excluded from the analyses (these entities could have constituted specific cases and therefore should not be included in the group of the so-called unsuccessful applicants out of which the control group was to be built). At the same time, applications that did not pass the formal assessment were admitted. In the latter case, it was acknowledged that further reduction of the potential control pool (of unsuccessful applicants) could make it impossible to carry out the analysis.

It should be noted that it was a significant challenge to automate the extraction of data from the EKRS Financial Document Browser and then merge the retrieved XML sets into a system that could be analysed by the evaluators. However, this labour-intensive stage produced a set of data of excellent quality, free from the various anomalies that usually occur in sets derived from questionnaires (incorrect/illogical answers, missing data, etc.). Another challenge was to reconstruct the logic of data included in financial statements and their substantive relevance in the context of formulating similarity and causal effect indicators. Experience gained from previous projects carried out in cooperation with the GUS (similar sets of balance sheet and profit and loss account data, developed sets of indicators and their interpretations within the framework of the developed methodology (GUS Statistical Research and Education Centre, 2018) proved invaluable in this respect. However, KRS data resources required a thorough assessment with respect to legal regulations, document templates for the particular types of entities, and the selected form of reporting, and with respect to the actual content of the sets (continuity, consistency, completeness). Conclusions from this work were presented in the previous sections. Without them, later analyses would not have been possible in terms of substance. This type of research is necessary for any attempt of using the available administrative sets of data for purposes other than their original intended use.

Main outcomes of the analysis

The analyses carried out confirmed the causal effects of POPW pro-innovation investments in the supported enterprises. The average value of beneficiaries' assets turned out to be significantly higher than the value of assets of the unsuccessful applicants (control group). In this respect, the average net effect per company (DiD) was between PLN 6 and 9 million (depending on the sub-measure). "In the third year from applying for co-financing, the average value of beneficiaries' fixed assets would have been lower by the stated amount, had they not received support (...). This corresponds to approximately 70% of the value of assets held in the period prior to applying for co-financing and proves the importance of the implemented investment processes" (IDEA Institute and IDEA Development Foundation commissioned by PARP, 2021, p. 6). In addition, a significant level of effects was confirmed with respect to the growth of assets of the supported companies (total assets index), growth of production capacity (expressed in significantly increased values of the energy and materials consumption index against the control group), and growth of employment

(this effect, in turn, was indirectly expressed in the values of the remuneration index, social security, and other benefits – average increase from PLN 1 million to 1.8 million in relation to changes that occurred in the control group in the corresponding period).

Table 9. Average impact of support for the analysed POPW instruments in the third year from submitting the application for co-financing according to selected balance sheet indicators and profit and loss account indicators

Indicator (in PLN thousands)	Measure 1.2	Submeasure 1.3.1	Measure 1.4a
Fixed assets	2369	8920	6025
Balance sheet total	6437	13618	9556
Consumption of materials and energy ^b	118	4161	3418
Remuneration, social security, and other benefits ^b	-81	1800	946

Source: IDEA Institute and IDEA Development Foundation based on PARP and EKRS data.

*** statistical validity of 0.01; ** of 0.05; * of 0.1

a – effects estimated for beneficiaries of measure 1.4 benefiting from both phases of support,

b – effects estimated for the sub-group of applicants which report profit and loss accounts in the comparative variant

It is worth noting that the estimates based on KRS data up to 2020 turned out to be consistent with the results of analogous analyses of the POPW effects up to 2018 carried out on the basis of SP GUS data. This provides solid grounds for the conclusion that the intervention was effective and that “it can be stated with a high degree of probability that without the support of the Programme the beneficiaries would not have made the investments in question. Therefore, the expected incentive effect of the investment aid materialised” (IDEA Institute and IDEA Development Foundation commissioned by PARP, 2021, p. 6).

The analyses also confirmed the presence of causal effects of the POPW during the COVID-19 pandemic, with regard to the revenues indicator of beneficiary companies (in the 2019-2020 period, the percentages of entities that increased their revenues were significantly higher among beneficiaries than among unsuccessful applicants; in the case of measure 1.4, supporting design innovations, a reduction in the share of beneficiaries that recorded a decrease in revenues was observed in the aforementioned period). Analogous results were not confirmed for beneficiaries of the pro-export instrument (negative impact of the pandemic on the value of achieved sales was experienced by beneficiaries of measure 1.2).

However, as far as financial indicators are concerned (profit/loss, sales revenue), the analysed data do not provide clear conclusions for the POPW (the analysed investments were completed only recently, or were in the finalisation phase, which means that they could not demonstrate a return in terms of financial results). Doubts regarding the effects (for most indicators) are also present in the case of the pro-export intervention. However, in this POPW measure, the KRS data used may not have been sufficiently sensitive or appropriate to capture the causal effects of the relatively small subsidies associated with exports promotion.

Conclusions

The EKRS Financial Document Browser provides universal, free access to financial statements of a significant proportion of Polish enterprises, including the applicants and beneficiaries of European Funds. This source provides access to annual (as of 2017) balance sheet data and profit and loss account data for enterprises responsible for more than 2/5 of applications for co-financing – the sample subpopulation of 3,308 enterprises-applicants of the POPW, analysed under this study.

For the purpose of evaluation of various types of economic development programmes (especially counterfactual impact evaluations that require micro data analyses), availability of data on enterprises from the EKRS Financial Document Browser is essential. However, it should be emphasised that it covers only a proportion of enterprises (applicants, beneficiaries, suppliers of goods and services to beneficiaries in projects, etc.) i.e. entities obliged to submit financial statements to the National Court Register (KRS). Data extracted from this source enable conclusions about enterprises organised as companies (including, i.a. general partnerships, limited partnerships, joint stock companies, limited liability companies). Financial data of companies operating under other legal forms (e.g. natural persons conducting business activity) are not available in the Browser.

Data from profit and loss accounts and balance sheets have particularly high potential for use in the evaluation of public programmes. Along with data on profits, revenues and costs, these studies provide extensive access to data on, i.a. remuneration, consumption of materials and energy, costs of completed development works, and the value of technical devices and machines. Both the Accounting Act and the Act on the National Court Register require entrepreneurs to ensure that the data shown in their financial statements is factually correct.

Data provided by the EKRS Financial Document Browser are considerably up-to-date. As a rule, data as at the end of a given calendar year must be made available in the EKRS at the beginning of the third quarter of the following year¹⁵⁵, which makes the period of waiting for access to data significantly shorter than in the case of analyses conducted through public statistics, using enterprise data extracted from annual statistical forms (e.g. SP GUS). In addition, financial statements are submitted by companies on an annual basis enabling access to both historical and new data, with no need to conduct additional rounds of measurements (which would have to be done in the case of statistical surveys).

Theoretically, access to these EKRS data is unrestricted. The Browser does not require a login, and allows to download any number of financial statements. However, it should be borne in mind that the Browser has not been created for analytical purposes. This makes it somewhat difficult for analysts to make secondary use of the published data. Files containing financial statements are made available in XML format i.e. as structurally organized text files. One file contains all financial statements (including the profit and loss account and the balance sheet) from a given year for a selected enterprise. The Browser makes it possible to search for all financial statements available for a given KRS number, and then download them (each file is downloaded independently, one by one). As a result, the Browser does not make it possible to download values of only selected indicators for specific enterprises or reporting years, or download all files, in bulk, for a specific list of KRS numbers. To analyse the data, all files must be downloaded manually (particular statements submitted by particular companies) and combined into a coherent table. It should be noted that the above process can be automated using Python (Selenium and BeautifulSoup libraries).

In conclusion, it should be pointed out that data from the EKRS Financial Document Browser has not previously been used to such a significant extent in impact evaluations of any program known to the authors. 2021 was the first year when PARP extracted and prepared sets of such company data from the EKRS for the purpose of impact evaluation. PARP also designed relevant analyses of counterfactual effects of the POPW, and carried them out (in cooperation with an independent evaluator) using individual financial data of applicant companies. This initiative also confirms the effectiveness of using the discussed source of administrative data in evaluations of the impact of economic development programmes.

¹⁵⁵ The exceptions in this regard are years 2020 and 2021, for which the deadline was extended to the fourth quarter of the following year.

References

- Accounting Act of 29 September 1994 [*Ustawa z dnia 29 września 1994 r. o rachunkowości*]. <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=wdu19941210591>
- Act on Public Statistics of 1995 [*Ustawa o statystyce publicznej z 29 czerwca 1995 r.*]. <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=wdu19950880439>
- Act on the National Court Register of 20 August 1997 [*Ustawa z dnia 20 sierpnia 1997 r. o Krajowym Rejestrze Sądowym*] <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU19971210769>
- Enterprise Law of 6 March 2018 [*Ustawa Prawo przedsiębiorców z dnia 6 marca 2018 r.*] <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20180000646>
- European Commission. (2014 a). *Commission staff working document. Common methodology for State aid evaluation*, Brussels, https://ec.europa.eu/competition/state_aid/modernisation/state_aid_evaluation_methodology_en.pdf.
- European Commission. (2014 b). *Commission Regulation (EU) No 651/2014 of 17 June 2014 declaring certain categories of aid compatible with the internal market in application of Articles 107 and 108 of the Treaty*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0651&from=EN>
- Evalu and Ecorys Polska commissioned by Ministry of Entrepreneurship and Technology. (2020). *Ewaluacja funkcjonowania programu pomocy regionalnej dla podmiotów gospodarczych prowadzących działalność gospodarczą w Specjalnych Strefach Ekonomicznych, na podstawie zezwoleń wydanych między 25 grudnia 2014 a 31 grudnia 2020 r., Raport końcowy*, Warsaw.
- Evalu, Stowarzyszenie na rzecz rozwoju Rynku Pracy STOS and WiseEuropa Foundation commissioned by Ministry of Development Funds and Regional Policy. (2020). *Ewaluacja regionalnej pomocy inwestycyjnej w zakresie celu tematycznego 3 w obszarze wzmocnienia konkurencyjności mikroprzedsiębiorstw oraz małych i średnich przedsiębiorstw w ramach regionalnych programów operacyjnych na lata 2014-2020*, Warsaw.
- GHK Polska commissioned by the Ministry of Regional Development. (2015). *Ekspertyza na temat źródeł danych wykorzystywanych do realizacji badań kontryfaktycznych w ramach ewaluacji Europejskiego Funduszu Społecznego*, Warsaw.
- GUS Statistical Research and Education Centre. (2013). *Ekspertyza: Wykorzystanie danych jednostkowych dla celów prowadzenia ewaluacji kontryfaktycznych interwencji publicznych – analiza wstępna*, Jachranka.
- GUS Statistical Research and Education Centre (2015). *Ustalenie wartości wybranych wskaźników ekonomicznych dla beneficjentów Regionalnych Programów Operacyjnych, Programu Operacyjnego Innowacyjna Gospodarka oraz dla dobranych grup kontrolnych*, Jachranka. https://www.ewaluacja.gov.pl/media/28765/Raport_koncowy_RPO_POIG_GUS_final.pdf.
- GUS Statistical Research and Education Centre. (2017). *Ustalenie wartości wybranych wskaźników ekonomicznych dla beneficjentów Programu Operacyjnego Kapitał Ludzki i dla dobranych grup kontrolnych*. Jachranka. <https://stat.gov.pl/statystyka-regionalna/statystyka-dla-polityki-spojnosci/statystyka-dla-polityki-spojnosci-2016-2018/badania/badania-kontryfaktyczne>.
- GUS Statistical Research and Education Centre. (2018). *Ustalenie wartości wybranych wskaźników ekonomicznych dla odbiorców pomocy, udzielanej za pośrednictwem PARP w ramach POIR i POPW, oraz dla dobranych grup kontrolnych – Studium wykonalności*. Jachranka. <https://stat.gov.pl/statystyka->

regionalna/statystyka-dla-polityki-spojnosci/statystyka-dla-polityki-spojnosci-2016-2018/badania/badania-kontrfaktyczne.

- Hatalaska-Woźniak, N. et al. (2020). *Mapa Trendów 2020*. Infuture Institute. <https://hatalaska.com/wp-content/uploads/2020/01/MapaTrendow2020.pdf>.
- IDEA Institute, IDEA Development Foundation and Jagiellonian University commissioned by PARP. (2020). *Ewaluacja mid-term pomocy publicznej PARP w ramach Programu Operacyjnego Inteligentny Rozwój 2014-2020. Raport końcowy, Warsaw.*
- IDEA Institute and IDEA Development Foundation commissioned by PARP. (2021). *Analizy kontrfaktyczne efektów POPW 2014-2020. Analiza ekspercka na potrzeby programowania następcy POPW na lata 2021-2027. Raport końcowy, Warsaw.*
- Koniewski, M., Krupnik S. (2017). *How to better understand the results of counterfactual analyses? Selection of beneficiaries for qualitative research using statistical matching, illustrated by an example of the impact evaluation of the Operational Programme Innovative Economy 2007-2013.* in: *Theory-based evaluation in complex environments*, ed. Pokorski, J., et al., PARP.
- MCM Institute, IDEA Institute and IDEA Development Foundation commissioned by PARP. (2021). *Efekty bieżące wsparcia udzielonego w ramach wybranych instrumentów POPW 2014-2020. Raport z analiz ad hoc w projekcie Barometr innowacyjności. Warsaw.*
- Ministry of Development Funds and Regional Policy. (2021). *Wytyczne w zakresie ewaluacji polityki spójności na lata 2014-2020, Warsaw.* <https://www.funduszeuropejskie.gov.pl/strony/o-funduszach/dokumenty/wytyczne-w-zakresie-ewaluacji-polityki-spojnosci-na-lata-2014-2020/>
- Ministry of Finance. (2020). *Rozporządzenie Ministra Finansów z dnia 31 marca 2020 r. w sprawie określenia innych terminów wypełniania obowiązków w zakresie ewidencji oraz w zakresie sporządzenia, zatwierdzenia, udostępnienia i przekazania do właściwego rejestru, jednostki lub organu sprawozdań lub informacji.* <https://isap.sejm.gov.pl/isap.nsf/DocDetails.xsp?id=WDU20200000570>
- Nowińska, K. *Czym jest triangulacja w badaniach społecznych?* Warsaw University of Technology. <https://www.cziitt.pw.edu.pl/czym-jest-triangulacja-w-badaniach-spoecznych/>.
- PARP. (2017). *Fizyka pracy badawczej PARP-GUS, służącej analizom kontrfaktycznym na potrzeby ewaluacji programu pomocowego PARP.*
- PARP. (2021). *Analizy kontrfaktyczne efektów POPW 2014-2020. Analiza ekspercka na potrzeby programowania następcy POPW na lata 2021-2027. Opis przedmiotu zamówienia. Warsaw.* <https://www.een.org.pl/component/content/article/1046-bip-zamowienia/74641-oopracowanie-analiz-kontrfaktycznych-efektow-popw-2014-2020?Itemid=907>.
- Pokorski, J. (2020). *Ewaluacja wsparcia kompetencji osób na rynku pracy i promocji zatrudnienia. Pełny obraz efektów z mikro- i makroperspektywy, in: Ewaluacja w działaniach edukacyjnych, ed. Ciężka, B., Rybińska, A., Fundacja Rozwoju Systemu Edukacji, Warsaw.*
- Walsh, J. (2021). *Data... the oil of evaluation?* European Commission. https://ec.europa.eu/regional_policy/en/policy/evaluations/conferences/cohesioneval2021/.
- World Bank. (2019). *Return on investment of public support to SMEs and innovation in Poland.* <https://www.worldbank.org/pl/country/poland/publication/poland-return-on-investment>.

List of abbreviations

- AGRI – Directorate-General for Agriculture and Rural Development
- BR – Better Regulation – Communication of the European Commission on Better Lawmaking
- BUDG – Directorate-General for Budget
- BUSC – Civil Registry Office register
- BZP – Public Procurement Bulletin (pl. Biuletyn Zamówień Publicznych)
- CAS – Central Statistical Application
- CC BY-NC-ND 4.0 – Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International
- CC-ME – Competence Centre on Microeconomic Evaluation at Joint Research Centre of the European Commission
- CF – Cohesion Fund
- CIT – Corporate Income Tax
- CNECT – Directorate-General for Communications Networks, Content and Technology
- COMP – Directorate-General for Competition
- COMPIE – Counterfactual Methods for Policy Impact Evaluation – series of conferences organized by CC-ME
- DALY – Disability-Adjusted Life Years
- DEASP – Irish Department of Employment Affairs and Social Protection
- DG – Directorate General of the European Commission
- DiD – Difference-In-Differences
- EAC – Directorate-General for Education, Youth, Sport and Culture
- EAT – Earnings After Taxes
- EC – European Commission
- EKRS – website of the Ministry of Justice that provides information from the National Court Register (<https://ekrs.ms.gov.pl/>)
- ELA – Polish system for monitoring the economic situation of graduates
- EMIS – Emerging Markets Information Service – website containing economic, financial and political information about Poland and the countries of Central and Eastern Europe
- EMPL – Directorate-General for Employment, Social Affairs and Inclusion
- ERDF – European Regional Development Fund
- ESF – European Social Fund
- Espacenet – international database of patent documents provided by the European Patent Office (<https://worldwide.espacenet.com>)

- EU – European Union
- EUR – euro
- EUROSTAT – European Statistical Office
- GBD – Global Burden of Diseases – project conducted by the University of Washington in Seattle
- GDP – Gross Domestic Product
- GUS – Statistics Poland
- IEFPP – Portuguese public employment service ((pt. Instituto de Emprego e Formação Profissional)
- iMSiG – website Court and Commercial Journal Online – <https://www.imsig.pl/> (pl. Internetowy Monitor Sądowy i Gospodarczy)
- JobsPlus – Irish scheme of hiring incentive named launched in 2012
- JPK – VAT Standard Audit File for Tax (pl. Jednolity Plik Kontrolny)
- JST – local government units
- JTF – Just Transition Fund
- KACHNA – insurance registration system of the Agricultural Social Insurance Fund
- KAS – National Revenue Administration (pl. Krajowa Administracja Skarbowa)
- KEP – National Register of Taxpayers (pl. Krajowa Ewidencja Podatników)
- KJE – National Evaluation Unit (pl. Krajowa Jednostka Ewaluacji)
- KRDN – website National Register of Insolvent Debtors – <https://krdn.pl/> (pl. Krajowy Rejestr Dłużników Niewypłacalnych)
- KRN – Polish National Cancer Registry (pl. Krajowy Rejestr Nowotworów)
- KRS – National Court Register (pl. Krajowy Rejestr Sądowy)
- LISA – Swedish integrated base of longitudinal data on health, labour market and education (se. Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier)
- MONA – Microdata Online Access – system in the Swedish system of official statistics ensuring secure internet access to individual data
- MRPiPS – Ministry of Family, Labour and Social Policy of the Republic of Poland (pl. Ministerstwo Rodziny, Pracy i Polityki Społecznej)
- NBP – National Bank of Poland (pl. Narodowy Bank Polski)
- NCBiR – National Center for Research and Development (pl. Narodowe Centrum Badań i Rozwoju)
- NFZ – National Health Fund (pl. Narodowy Fundusz Zdrowia)
- NIK – Supreme Audit Office (pl. Najwyższa Izba Kontroli)
- NIP – Taxpayer Identification Number in Poland
- nSIU – central system of the Agricultural Social Insurance Fund
- **NTS – The Nomenclature of Territorial Units for Statistical Purposes**

- OECD – Organisation for Economic Cooperation and Development
- PARP – Polish Agency for Enterprise Development (pl. Polska Agencja Rozwoju Przedsiębiorczości)
- PATSTAT – Worldwide Patent Statistical Database
- PESEL – Universal Electronic System for Registration of the Population
- PIE – Polish Economic Institute (pl. Polski Instytut Ekonomiczny)
- PIT – Personal Income Tax
- PKD – Polish classification of the enterprise’s activity
- PNT-01 – Report on research and development (R&D) activities; GUS Survey
- PNT-02 – Industrial Innovation Report; GUS Survey
- POL-on – Integrated Information System on Higher Education and Science
- R&D – research & development
- REGIOGMINA – project “Placing instruments of support for Small and Medium Enterprises (SMEs) at the level of local governments, operating on the basis of multi-level region management model (REGIOGMINA)”, carried out under the initiative of the National Centre for Research and Development entitled “GOSPOSTRATEG”
- REGON – National Official Register of Business Entities
- RTD – Directorate-General for Research and Innovation
- SBA – Small Business Act for Europe – program of the European Commission setting out priorities for promoting entrepreneurship in the EU and facilitating development of small businesses
- SME – Small and medium-sized enterprises
- SOM – Self-Organizing Kohonen Maps
- SP – Annual Enterprise Survey; GUS Survey
- SS – Social Security in Portugal (pt. Segurança Social)
- UPRP – Patent Office of the Republic of Poland (pl. Urząd Patentowy Rzeczypospolitej Polskiej)
- VAT – Value-Added Tax
- WIPO – World Intellectual Property Organization
- YEI – Portuguese Youth Employment Initiative
- ZPA – Integrated Analytical Platform (pl. ZIntegrowana Platforma Analityczna)
- ZPA CK – Centre of Competence of Integrated Analytical Platform
- ZUS – Social Insurance Institution (pl. Zakład Ubezpieczeń Społecznych)



www.parp.gov.pl