

Ekspertyza dotycząca wykorzystania  
danych Zakładu Ubezpieczeń  
Społecznych na potrzeby ewaluacji  
wsparcia Europejskiego Funduszu  
Społecznego





Rzeczpospolita  
Polska



Unia Europejska  
Europejski Fundusz Społeczny



Ekspertyza współfinansowana ze środków Europejskiego Funduszu Społecznego

Autor:

Paweł Penszko

Warszawa, grudzień 2020

## Streszczenie

Niniejsze opracowanie jest rezultatem analizy dotychczasowych doświadczeń związanych z wykorzystaniem danych ZUS na potrzeby ewaluacji polityk publicznych i programów współfinansowanych z EFS. Doświadczenia te pochodzą z trzech badań ewaluacyjnych, w których wykorzystano wąski zakres danych indywidualnych ZUS do pomiaru osiągniętych wartości wskaźników rezultatu. W analizach uwzględniono również funkcjonujący już system monitoringu losów absolwentów szkół wyższych (ELA) i projektowany analogiczny system dotyczący absolwentów szkół ponadpodstawowych, a także wybrane przykłady zagraniczne.

Zasoby danych ZUS, o złożonej i częściowo rozproszonej strukturze, są bardzo obszerne i mogłyby być potencjalnie wykorzystane na potrzeby ewaluacji w wielu dziedzinach interwencji EFS (przede wszystkim: rynek pracy, integracja społeczna, ochrona zdrowia, edukacja) oraz znaleźć nowe zastosowania, np. w analizie trwałości efektów, do charakterystyki uczestników w okresie przed udziałem w projekcie lub do doboru prób o różnym przeznaczeniu badawczym. Do ich zalet należą: możliwość określenia sytuacji jednostki (np. statusu na runku pracy) w dowolnie wybranym momencie, wysoka jakość i precyzja danych, niski koszt regularnego ich pozyskiwania. Przy planowaniu wykorzystania danych ZUS należy wziąć pod uwagę obciążenia nakładane na kadry i system informatyczny tej instytucji.

Praktycznym warunkiem wykorzystania danych indywidualnych ZUS w ewaluacji EFS jest istnienie podstawy do tego w przepisach prawnych. W trakcie przygotowań nowego okresu programowania warto więc przewidzieć i zaplanować sposoby użycia tych danych i wprowadzić odpowiednie zmiany legislacyjne, które to ewentualnie umożliwią. Do rekomendowanych działań, które będą sprzyjały korzystaniu z tego źródła informacji, a przez to istotnemu powiększeniu zasobów wiedzy przydatnej przy prowadzeniu polityki publicznej, należą:

- Utrzymanie, a w miarę możliwości rozszerzenie dostępu do danych ZUS na potrzeby ewaluacji EFS w nowym okresie programowania;
- Uwzględnienie potrzeb ewaluacji polityki spójności przy wprowadzaniu systemu monitoringu losów absolwentów szkół ponadpodstawowych;
- Uwzględnienie potrzeb ewaluacji polityki spójności przy tworzeniu platformy integrującej zasoby danych publicznych;
- Uwzględnienie w systemie ewaluacji polityki spójności instytucji badawczych, posiadających dostęp do danych ZUS;
- Zbieranie w systemie realizacji polityki spójności danych identyfikacyjnych zarówno na poziomie uczestników, jak i na poziomie instytucji (np. pracodawców osób uczestniczących w projektach współfinansowanych z EFS);
- Wykorzystanie dostępnych danych zagregowanych.

## Spis treści

Streszczenie .....	3
1. Przedmiot i metodyka ekspertyzy .....	6
2. Zasoby danych ZUS.....	8
2.1. Opis systemów informatycznych i zasobów danych jednostkowych.....	8
2.2. Udostępnianie danych zagregowanych.....	13
3. Dotychczasowe wykorzystanie danych ZUS w badaniach współfinansowanych z EFS.....	15
3.1. Wykorzystanie danych indywidualnych w badaniach polskich.....	15
3.1.1. Metaanaliza wyników badań ewaluacyjnych dotyczących oceny wsparcia z Europejskiego Funduszu Społecznego .....	15
3.1.2. Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój .....	16
3.1.3. Badanie ewaluacyjne dotyczące wyliczenia wartości wskaźników rezultatu długoterminowego w ramach Europejskiego Funduszu Społecznego oraz Inicjatywy na rzecz zatrudnienia osób młodych .....	17
3.1.4. Monitoring losów absolwentów szkół wyższych.....	17
3.1.5. Monitoring losów absolwentów szkół ponadpodstawowych .....	19
3.2. Wykorzystanie danych zagregowanych udostępnianych przez ZUS .....	21
3.3. Zagraniczne przykłady wykorzystania danych ubezpieczeń społecznych zdrowotnych w ewaluacji EFS .....	21
3.3.1. Szwecja .....	21
3.3.2. Litwa .....	22
4. Specyfika danych ZUS.....	24
4.1. Kwestie związane z pozyskaniem danych.....	24
4.2. Kwestie związane z populacyjnym charakterem danych .....	28
4.3. Kwestie związane z zakresem i kompletnością danych.....	30
4.4. Kwestie związane z przetwarzaniem i analizą danych indywidualnych .....	31
5. Perspektywa wykorzystania danych ZUS w przyszłości.....	34
6. Wnioski.....	37
Bibliografia.....	41

## Słownik skrótów

CAS	Centralna Aplikacja Statystyczna
CEIDG	Centralna Ewidencja i Informacja o Działalności Gospodarczej
CeSAR	Centralny System Analityczno- Raportowy
CIE	Centrum Informatyczne Edukacji
CKE	Centralna Komisja Egzaminacyjna
CRP KEP	Centralny Rejestr Podmiotów - Krajowa Ewidencja Podatników
CSIZS	Centralny System Informatyczny Zabezpieczenia Społecznego
EFS	Europejski Fundusz Społeczny
ELA	Ogólnopolski system monitorowania Ekonomicznych Losów Absolwentów szkół wyższych
GUS	Główny Urząd Statystyczny
IBE	Instytut Badań Edukacyjnych
KRUS	Kasa Rolniczego Ubezpieczenia Społecznego
KSI ZUS	Kompleksowy System Informatyczny Zakładu Ubezpieczeń Społecznych
MLASZ	Monitorowanie losów absolwentów szkół zawodowych
MLEZAiMD	Monitorowanie Losów Edukacyjno-Zawodowych Absolwentów i Młodych Dorosłych
NFZ	Narodowy Fundusz Zdrowia
NIP	Numer identyfikacji podatkowej
nSIU	Nowy System Informatyczny Ubezpieczeń
OKE	Okręgowa Komisja Egzaminacyjna
OPI	Ośrodek Przetwarzania Informacji
PARP	Polska Agencja Rozwoju Przedsiębiorczości
PEJK UW	Pracownia Ewaluacji Jakości Kształcenia Uniwersytetu Warszawskiego
PESEL	Powszechny Elektroniczny System Ewidencji Ludności
PIE	Polski Instytut Ekonomiczny
PFR	Polski Fundusz Rozwoju
POKL	Program Operacyjny Kapitał Ludzki 2007-2013
POWER	Program Operacyjny Wiedza Edukacja Rozwój 2014-2020
PPMI	ang. Public Policy and Management Institute = Instytut Polityki Publicznej i Zarządzania
REGON	Rejestr Gospodarki Narodowej
RODO	Ogólne rozporządzenie o ochronie danych
RSPO	Rejestr Szkół i Placówek Oświatowych
SCB	szw. Statistiska centralbyrån = Centralny Urząd Statystyczny
SIO	System Informacji Oświatowej
UE	Unia Europejska
UNECE	United Nations Economic Commission for Europe
UW	Uniwersytet Warszawski
YEI	ang. Youth Employment Initiative = Inicjatywa na rzecz zatrudnienia ludzi młodych
ZETO	Zakład Elektronicznej Techniki Obliczeniowej
ZPA	Zintegrowana Platforma Analityczna
ZUS	Zakład Ubezpieczeń Społecznych

## 1. Przedmiot i metodyka ekspertyzy

Przedmiotem ekspertyzy jest analiza dotychczasowych doświadczeń związanych z wykorzystaniem danych ZUS na potrzeby ewaluacji polityk publicznych i programów współfinansowanych z EFS.

Niniejsza ekspertyza koncentruje się na danych indywidualnych, czyli na zawartych w zasobach ZUS informacjach o konkretnych osobach, płatnikach składek lub dokumentach. Dostęp do nich, również na potrzeby ewaluacji polityk publicznych, jest ograniczony, przede wszystkim w związku z tym, że znaczna ich część ma charakter danych osobowych, a niekiedy danych wrażliwych. Mniej uwagi poświęcono danym zagregowanym, ponieważ kwestia ich potencjalnego wykorzystania jest znacznie prostsza, a dostęp do nich nie rodzi problemów.

Zakres ekspertyzy obejmuje:

1. stwierdzone w praktyce - bazujące na rzeczywistych procesach badawczych i analitycznych, prowadzonych na potrzeby ewaluacji programów współfinansowanych z EFS w latach 2014-2020 (POWER, RPO):
  - a. korzyści wynikające z wykorzystania danych o charakterze populacyjnym,
  - b. uwarunkowania związane z pozyskaniem danych,
  - c. uwarunkowania związane z obróbką (przystosowaniem / przetworzeniem do wtórego wykorzystania) i analizą danych,
  - d. uwarunkowania związane z zakresem i kompletnością danych,
2. potencjalne, niewykorzystane dotychczas, możliwości związane z analizą danych ZUS na potrzeby ewaluacji polityk publicznych (w szczególności EFS),
3. wnioski dotyczące potencjału (w tym zakresu zastosowań) oraz ograniczeń wykorzystania przedmiotowych danych w przyszłych badaniach ewaluacyjnych.

W ramach ekspertyzy przeanalizowano treść aktów prawnych (ustaw, rozporządzeń krajowych, rozporządzeń unijnych, projektu ustawy), mających znaczenie dla omawianej problematyki, a także wytycznych związanych z realizacją polityki spójności w obszarze EFS. Wykorzystano również wcześniejsze opracowania, takie jak raporty, artykuły czy referaty - ich lista znajduje się w zamieszczonej na końcu bibliografii. Sięgnięto po publicznie dostępne źródła zamieszczone w Internecie, dotyczące na przykład systemów informatycznych wykorzystywanych w ZUS i innych instytucjach. Przeprowadzono ponadto wywiady z osobami zaangażowanymi w realizację dotychczasowych polskich badań współfinansowanych ze środków EFS i wykorzystujących dane ZUS, pozyskując w ten sposób wartościowe informacje na temat badań ewaluacyjnych oraz systemów monitoringu losów, opisanych w rozdziale 3.1. Rezultatem tych prac badawczych jest niniejsze opracowanie.

W jego początkowej części omówione są zasoby danych, którymi dysponuje ZUS. Następnie opisane są dotychczasowe badania współfinansowane z EFS, w których wykorzystano dane indywidualne ZUS, pod kątem wniosków z zebranych w nich doświadczeń. Wspomniana jest również kwestia wykorzystania danych zagregowanych, a także zagraniczne przykłady wykorzystania analogicznych danych, które mogą posłużyć jako inspiracja. Kolejny rozdział, opatrzony numerem 4 i bazujący na dotychczasowych doświadczeniach, poświęcony jest specyfice pracy z indywidualnymi danymi ZUS w ramach ewaluacji EFS. W rozdziale 5. nakreślone są perspektywy przyszłego wykorzystania danych ZUS w ewaluacji EFS, przy wzięciu pod uwagę ich zakresu opisanego w rozdziale 2., dotychczasowych

doświadczeń opisanych w rozdziale 3. i praktycznych aspektów opisanych w rozdziale 4. Ostatni rozdział zawiera wnioski i rekomendacje wypływające z wcześniejszych części.

## 2. Zasoby danych ZUS

### 2.1. Opis systemów informatycznych i zasobów danych jednostkowych

Podstawowym systemem informatycznym ZUS, w którym gromadzone są dane przydatne w ewaluacji EFS, jest Kompleksowy System Informatyczny ZUS (KSI ZUS). Funkcjonuje on od czasu reformy ubezpieczeń społecznych z 1999 roku. Z technicznego punktu widzenia jest on systemem scentralizowanym. Źródłami danych w nim zawartych są:

- a) Formularze i wnioski składane do ZUS przez płatników składek, ubezpieczonych, lekarzy itd.
- b) Postępowania prowadzone w ZUS, podejmowane w ramach funkcjonowania tej instytucji decyzje, transfery finansowe itd.
- c) Inne systemy informatyczne, w szczególności rejestry państwowe (PESEL, REGON, CRP KEP, CEiDG), które zasilają KSI ZUS danymi w ramach regularnych aktualizacji.

W KSI ZUS funkcjonują konta ubezpieczonych, a także konta płatników składek, na których gromadzone są informacje określone w Ustawie o systemie ubezpieczeń społecznych (Dz. U. 1998, poz. 887 z późn. zm.). W skład KSI ZUS wchodzi również zestaw rejestrów prowadzonych na podstawie art. 33 ust. 1 tej ustawy:

- Centralny Rejestr Ubezpieczonych,
- Centralny Rejestr Płatników Składek,
- Centralny Rejestr Członków Otwartych Funduszy Emerytalnych,
- Centralny Rejestr Członków Rodziny Ubezpieczonych Uprawnionych do Ubezpieczenia Zdrowotnego

Zakres danych określonych w powyższych rejestrach określa Rozporządzenie Ministra Pracy i Polityki Społecznej w sprawie szczegółowego zakresu danych zawartych w centralnych rejestrach prowadzonych przez Zakład Ubezpieczeń Społecznych (tekst jednolity: Dz. U. 2013, poz. 219).

Ponadto ZUS, na podstawie Ustawy z dnia 25 czerwca 1999 r. o świadczeniach pieniężnych z ubezpieczenia społecznego w razie choroby i macierzyństwa (tekst jednolity: Dz. U. 2020, poz. 870, z późn. zm.), prowadzi:

- Rejestr lekarzy, lekarzy dentyków, felczerów i starszych felczerów, którzy zgłosili wniosek w sprawie upoważnienia ich do wystawiania zaświadczeń lekarskich (art. 56 ust. 1),
- Rejestr zaświadczeń lekarskich (art. 55b ust. 1).

W KSI ZUS przechowywane są również:

- dane o orzeczeniach komisji lekarskich i lekarzy orzeczników ZUS o niezdolności do pracy,
- dane o orzeczeniach komisji lekarskich i lekarzy orzeczników ZUS o uszczerbku na zdrowiu,
- wyniki kontroli orzekania o czasowej niezdolności do pracy,
- wyniki kontroli wykorzystania zwolnień,



- dane o rehabilitacji leczniczej.

Oprócz KSI ZUS, w Zakładzie używane są dwa starsze systemy obsługi świadczeń emerytalno-rentowych: EMIR-SEKS AD oraz RENTIER-MANAGER AD. Zostały one stworzone w 1981 roku w ZETO Katowice i ZETO Bydgoszcz. Podlegały od tego czasu rozwojowi i modernizacjom. Funkcjonalności obu systemów są zbliżone<sup>1</sup>. Bazy danych tworzone w ramach tych systemów mają charakter rozproszony – każdy oddział ZUS i każda komórka realizująca umowy międzynarodowe dysponuje własną odrębną bazą w jednym z tych dwóch systemów.

Ponadto w ZUS funkcjonuje system finansowo-księgowy SAP ERP, służący do bieżącego zarządzania tą instytucją. Z punktu widzenia ewaluacji EFS nie wydaje się on mieć istotnego znaczenia. Informacje o głównych wymienionych zbiorach danych funkcjonujących w ZUS oraz zakresie przechowywanych danych zostały podsumowane w poniższej tabeli.

Wobec rozproszenia danych gromadzonych przez ZUS instrumentem mającym fundamentalne znaczenie dla możliwości ich wykorzystania w ewaluacji EFS jest system raportowania, umożliwiający wygenerowanie jednolitego raportu na podstawie danych zgromadzonych w różnych systemach i bazach. Instrument ten jest używany do przygotowania zbiorów danych służących następnie do wyliczenia osiągniętych wartości wskaźników rezultatu długoterminowego EFS i YEI.

---

<sup>1</sup> Zgodnie z informacją opublikowaną w lutym 2019 roku w ramach procedury zamówień publicznych 2019/S 032-072790, systemy te obsługiwały część świadczeń wypłacanych na podstawie przepisów obowiązujących do 31.12.1998 r., podczas gdy resztę tego rodzaju świadczeń oraz wszystkie świadczenia wypłacane na podstawie przepisów obowiązujących po 1.01.1999 r. obsługiwał już nowszy Kompleksowy System Informatyczny ZUS.

Tabela 1. Zasoby danych ZUS

Lp.	Nazwa zbioru danych	Zawarte dane
1	Centralny Rejestr Ubezpieczonych	<ol style="list-style-type: none"> <li>dane identyfikacyjne ubezpieczonego: numer PESEL, rodzaj, serię i numer dokumentu tożsamości, nazwisko, pierwsze i drugie imię, datę urodzenia, nazwisko rodowe, obywatelstwo, płeć;</li> <li>dane adresowe ubezpieczonego odpowiednio: zameldowania, zamieszkania, do korespondencji: kod pocztowy, miejscowość, gminę lub dzielnicę, ulicę, numer domu, numer lokalu, skrytkę pocztową, numer telefonu, adres poczty elektronicznej, symbol państwa wraz z zagranicznym kodem pocztowym, numer rachunku bankowego;</li> <li>dane dotyczące podlegania ubezpieczeniom: tytuł ubezpieczenia, rodzaj ubezpieczenia, datę objęcia ubezpieczeniem, datę wyrejestrowania z ubezpieczenia, informację za okres do dnia 31 grudnia 2008 r. o pracy szczególnej, wymiar czasu pracy;</li> <li>dane dotyczące stopnia niepełnosprawności;</li> <li>dane dotyczące niezdolności do pracy: całkowitej lub częściowej.</li> </ol>
2	Centralny Rejestr Płatników Składek	<ol style="list-style-type: none"> <li>dane identyfikacyjne płatnika składek: numer NIP, numer REGON, numer PESEL, nazwę skróconą, rodzaj, serię i numer dokumentu tożsamości, nazwisko, pierwsze i drugie imię, datę i miejsce urodzenia, obywatelstwo, nazwę lub firmę zgodną z aktem konstytuującym podmiotu, nazwę organu założycielskiego, rodzaj uprawnień, na podstawie którego jest prowadzona pozarolnicza działalność, numer uprawnień do prowadzenia działalności, nazwę organu rejestrowego lub ewidencyjnego wydającego uprawnienia do prowadzenia pozarolniczej działalności, datę wpisu do rejestru-ewidencji lub wydania uprawnienia do prowadzenia pozarolniczej działalności, numer wpisu do rejestru lub uprawnienia do prowadzenia pozarolniczej działalności, datę powstania obowiązku opłacania składek na ubezpieczenia społeczne, przyczynę wyrejestrowania, datę wyrejestrowania, numery rachunków bankowych płatnika składek;</li> <li>dane adresowe płatnika składek: typ adresu (adres siedziby, zamieszkania, do korespondencji) oraz kod pocztowy, miejscowość, gminę lub dzielnicę, ulicę, numer domu, numer lokalu, skrytkę pocztową, numer telefonu, adres poczty elektronicznej, symbol państwa wraz z zagranicznym kodem pocztowym, numer telefonu do teletransmisji;</li> <li>dane dotyczące biura rachunkowego płatnika składek: numer NIP, numer REGON, nazwę skróconą.</li> </ol>
3	Centralny Rejestr Członków Otwartych Funduszy Emerytalnych	<ol style="list-style-type: none"> <li>dane identyfikacyjne członka otwartego funduszu emerytalnego: numer PESEL, rodzaj i numer dokumentu tożsamości, nazwisko, pierwsze i drugie imię, datę urodzenia;</li> <li>dane dotyczące członkostwa w otwartym funduszu emerytalnym, zwanym dalej „OFE”: datę i godzinę zawarcia umowy z OFE, sposób zawarcia umowy z OFE, datę wyrejestrowania z OFE, identyfikator OFE.</li> </ol>
4	Centralny Rejestr Członków Rodziny Ubezpieczonych Uprawnionych do Ubezpieczenia Zdrowotnego	<ol style="list-style-type: none"> <li>dane identyfikacyjne członka rodziny ubezpieczonego i przebieg ubezpieczenia: numer PESEL, rodzaj, serię i numer dokumentu tożsamości, nazwisko, pierwsze imię, datę urodzenia, stopień pokrewieństwa lub powinowactwa z ubezpieczonym, stopień niepełnosprawności, pozostawanie we wspólnym gospodarstwie domowym z osobą ubezpieczoną, datę uzyskania i datę utraty uprawnień do ubezpieczenia zdrowotnego przez członka rodziny;</li> <li>dane adresowe członka rodziny ubezpieczonego: kod pocztowy, miejscowość, gminę lub dzielnicę, ulicę, numer domu, numer lokalu, symbol państwa wraz z zagranicznym kodem pocztowym, numer telefonu;</li> <li>dane dotyczące oddziału wojewódzkiego Narodowego Funduszu Zdrowia właściwego dla członka rodziny ubezpieczonego.</li> </ol>

5	Konta ubezpieczonych	<p>PESEL oraz informacje:</p> <ul style="list-style-type: none"> <li>▪ o zwaloryzowanej wysokości należnych/wpłaconych składek zgodnie z art. 40 i 40a ustawy;</li> <li>▪ o kwocie środków zewidencjonowanych na subkoncie, o którym mowa w art. 40a ustawy, na ostatni dzień miesiąca poprzedzającego miesiąc ustalenia prawa do emerytury z tytułu osiągnięcia wieku emerytalnego wynoszącego 65/60 lat przez osobę, która nie spełnia ustawowych warunków do ustalenia prawa do okresowej emerytury kapitałowej;</li> <li>▪ o zwaloryzowanej wysokości wpłaconych składek, o których mowa w art. 22 ust. 3 pkt 1 lit. b i pkt 2, wraz z wyegzekwowanymi od tych składek odsetkami za zwłokę i opłatą prolongacyjną, o których mowa w art. 23 ust. 2, oraz kwotę środków odpowiadających wartości umorzonych przez otwarty fundusz emerytalny jednostek rozrachunkowych;</li> <li>▪ wymienione w drukach: zgłoszenia do ubezpieczeń społecznych, imiennym raporcie miesięcznym, deklaracji rozliczeniowej i raporcie informacyjnym;</li> <li>▪ o członkostwie w otwartym funduszu emerytalnym i o terminach przekazania składek do tego funduszu;</li> <li>▪ o członkostwie w kasie chorych i o terminach przekazania składek do tej kasy;</li> <li>▪ o wysokości należnych i wpłaconych składek na ubezpieczenia emerytalne, rentowe, chorobowe, wypadkowe i zdrowotne oraz o wysokości należnych i odprowadzonych składek podlegających odprowadzeniu do otwartego funduszu emerytalnego;</li> <li>▪ o wartości środków wypłaconych z tytułu okresowej emerytury kapitałowej;</li> <li>▪ o faktach pozaubezpieczeniowych, mających wpływ na prawo do świadczeń z ubezpieczeń społecznych i na ich wysokość;</li> <li>▪ niezbędne do przyznania, ustalenia wysokości i wypłaty świadczeń z ubezpieczeń społecznych, a także świadczeń finansowanych z budżetu państwa oraz o dokonanych wypłatach;</li> <li>▪ niezbędne do ustalenia, ponownego ustalenia lub przeliczenia kapitału początkowego;</li> <li>▪ o kapitale początkowym oraz zwaloryzowanym kapitale początkowym;</li> <li>▪ niezbędne do realizacji przez Zakład zadań zleconych na podstawie odrębnych przepisów;</li> <li>▪ o okresach pracy w szczególnych warunkach lub o szczególnym charakterze;</li> <li>▪ o małżeńskich stosunkach majątkowych oraz o osobach fizycznych, na rzecz których ma nastąpić w razie śmierci wypłata środków zewidencjonowanych na subkoncie, o którym mowa w art. 40a, w przypadku osób niebędących członkami otwartego funduszu emerytalnego, dla których Zakład prowadzi to subkonto; o kwocie środków, o której mowa w art. 25 ust. 1 ustawy z dnia 17 grudnia 1998 r. o emeryturach i rentach z Funduszu Ubezpieczeń Społecznych zewidencjonowanych na subkoncie, o którym mowa w art. 40a;</li> <li>▪ o osobach uposażonych;</li> <li>▪ o oświadczeniach: a) o przekazywaniu do otwartego funduszu emerytalnego składki, o której mowa w art. 22 ust. 3 pkt 1 lit. a ustawy, począwszy od składki opłaconej za lipiec, b) o zewidencjonowaniu składki, o której mowa w art. 22 ust. 3 pkt 2, na subkoncie, o którym mowa w art. 40a ustawy, począwszy od składki opłaconej za miesiąc, w którym złożono wniosek.</li> </ul>
6	Konta płatników składek	<ol style="list-style-type: none"> <li>1. dane identyfikacyjne, numery rachunków składkowych, numer NIP i numer identyfikacyjny REGON, a jeżeli płatnikowi składek nie nadano tych numerów lub jednego z nich – numer PESEL lub serię i numer dowodu osobistego albo paszportu, nazwy i numery rachunków bankowych płatnika składek, dane informacyjne płatnika składek, w tym szczególną formę prawną, kod PKD oraz wszelkie inne dane konieczne do obsługi konta, a w szczególności do celów rozliczania należności z tytułu składek oraz do celów prowadzenia postępowania egzekucyjnego, w tym informacje dotyczące współników spółek cywilnych, jawnych i komandytowych, w zakresie rejestrowanym w Centralnym Rejestrze Podmiotów – Krajowej Ewidencji Podatników;</li> <li>2. liczba pracowników, za których jest opłacana składka na Fundusz Emerytur Pomostowych;</li> <li>3. informacje dotyczące zawartych umów o dzieło, o których mowa w art. 36 ust. 17 ustawy;</li> <li>4. rozliczenia należnych składek, wypłacanych przez płatnika zasiłków oraz zasiłków rodzinnych i pielęgnacyjnych podlegających zaliczeniu na poczet składek oraz innych składek pobieranych przez Zakład;</li> <li>5. dane niezbędne do realizacji przez Zakład zadań zleconych odrębnymi przepisami.</li> </ol>

7	Rejestr lekarzy, lekarzy dentystów, felczerów i starszych felczerów, którzy zgłosili wniosek w sprawie upoważnienia ich do wystawiania zaświadczeń lekarskich	<ol style="list-style-type: none"> <li>1. numer prawa wykonywania zawodu;</li> <li>2. imię i nazwisko;</li> <li>3. numer PESEL albo serię i numer paszportu, w przypadku gdy nie nadano numeru PESEL;</li> <li>4. rodzaj i stopień specjalizacji;</li> <li>5. adres miejsca udzielania świadczeń zdrowotnych;</li> <li>6. nazwa i siedziba właściwej izby lekarskiej;</li> <li>7. informacja o cofnięciu upoważnienia.</li> </ol>
8	Rejestr zaświadczeń lekarskich	<ol style="list-style-type: none"> <li>1. identyfikator i datę wystawienia zaświadczenia lekarskiego;</li> <li>2. dane ubezpieczonego: pierwsze imię, nazwisko, PESEL (albo serię i numer paszportu i datę urodzenia, jeżeli nie nadano numeru PESEL), adres miejsca pobytu ubezpieczonego w czasie trwania niezdolności do pracy;</li> <li>3. dane płatnika składek: NIP (lub PESEL albo serię i numer paszportu, jeżeli nie ma obowiązku posługiwania się NIP i nie nadano numeru PESEL), rodzaj identyfikatora płatnika składek;</li> <li>4. imię, nazwisko i numer prawa wykonywania zawodu wystawiającego zaświadczenie lekarskie, adres miejsca udzielania świadczeń zdrowotnych;</li> <li>5. okres orzeczonej czasowej niezdolności do pracy, w tym okres pobytu w szpitalu;</li> <li>6. informacje o okolicznościach mających wpływ na prawo do zasiłku chorobowego lub jego wysokość;</li> <li>7. wskazania lekarskie – odpowiednio: chory powinien leżeć albo chory może chodzić;</li> <li>8. okres zwolnienia od wykonywania pracy z powodu konieczności sprawowania osobistej opieki nad chorym członkiem rodziny, datę urodzenia tego członka rodziny i stopień jego pokrewieństwa lub powinowactwa z ubezpieczonym;</li> <li>9. numer statystyczny choroby ubezpieczonego;</li> <li>10. oznaczenie instytucji, w której ubezpieczony został zgłoszony do ubezpieczenia.</li> </ol>
9	Zbiory systemu obsługi świadczeń emerytalno-rentowych EMIR (EMIR-SEKS AD)	Brak szczegółowych informacji
10	Zbiory systemu obsługi świadczeń emerytalno-rentowych RENTIER (RENTIER-MANAGER AD)	Brak szczegółowych informacji
11	System finansowo-księgowy SAP ERP	Dokumentacja systemu na stronie internetowej producenta: <a href="https://www.sap.com/poland/index.html">https://www.sap.com/poland/index.html</a> .

*Źródło: opracowanie własne.*

## 2.2. Udostępnianie danych zagregowanych

Pod adresem internetowym <https://psz.zus.pl/> dostępny jest Portal Statystyczny ZUS, w którym publikowane są dane statystyczne dotyczące: płatników składek, ubezpieczonych, emerytur, rent z tytułu niezdolności do pracy, rent rodzinnych, zasiłków, absencji chorobowych, orzecznictwa lekarskiego, rehabilitacji leczniczej oraz wyników kontroli. Są to dane zagregowane, prezentowane w podziale na wybrane kategorie. Zawartość Portalu Statystycznego ZUS jest udostępniana także na portalu Otwarte Dane ([dane.gov.pl](https://dane.gov.pl)).

Na stronie internetowej ZUS, na podstronie „Baza wiedzy”, udostępniane są ponadto liczne publikacje i wydawnictwa Zakładu, zawierające wyniki analiz statystycznych oraz badań przeprowadzonych lub zleconych przez tę instytucję. Są to potencjalnie cenne źródła informacji z dziedzin rynku pracy, wykluczenia społecznego i ochrony zdrowia. Kilka z tych opracowań ma charakter ewaluacji efektów prowadzonych działań, z reguły programów prewencji rentowej i wypadkowej. W jednej z nich wykorzystano dane z systemów informatycznych ZUS<sup>2</sup>, pozostałe oparte są na własnych badaniach ankietowych i medycznych<sup>3</sup>. Wszystkie te ewaluacje dotyczą programów własnych Zakładu. Warto odnotować również Zieloną księgę i Białą księgę przeglądu systemu emerytalnego 2016, które zawierają diagnozę tego systemu oraz rekomendacje co do jego reformy. Na stronie internetowej ZUS nie są natomiast publikowane prace naukowe własnych i niezależnych ekspertów, wykorzystujących bazy danych Zakładu do analiz społeczno-ekonomicznych, w tym dotyczących efektów prowadzonych polityk publicznych, wdrażanych reform lub programów realizowanych przez inne instytucje publiczne. Zagranicznym przykładem tego rodzaju działalności jest projekt VisitINPS włoskiego odpowiednika ZUS – Istituto nazionale della previdenza sociale. W jego ramach opublikowano na przykład analizy wpływu podwyższenia wieku emerytalnego na zatrudnianie osób młodych (Boeri, Garibaldi, & Moen, 2016), oddziaływania regionalnych polityk publicznych na regionalne rynki pracy (Pellegrini & Cerqua, 2017) czy efektów reformy staży zawodowych (Albanese, Capellari, & Leonardi, 2017). Ukazuje to pewien dodatkowy potencjał wykorzystania danych ZUS do ewaluacji i kształtowania polityk publicznych, który w Polsce nie jest niestety wykorzystywany. Aktualnie obowiązujące akty prawne nie sprzyjają przekazywaniu danych ZUS przedstawicielom zewnętrznych instytucji naukowych, choć stwarzają pewne możliwości – przykładowo, Polski Instytut Ekonomiczny jest ustawowo umocowany do pozyskiwania z ZUS danych statystycznych na potrzeby realizacji analiz ekonomicznych i społecznych<sup>4</sup>.

Dane z Kompleksowego Systemu Informatycznego Zakładu Ubezpieczeń Społecznych oraz z systemów obsługi świadczeń emerytalno-rentowych EMIR I RENTIER są regularnie przekazywane do Głównego Urzędu Statystycznego i wykorzystywane do stworzenia operatów oraz w badaniach GUS dotyczących następujących tematów<sup>5</sup>:

- Produkt krajowy brutto i jego elementy w ujęciu regionalnym
- Rachunki narodowe niefinansowe według sektorów i podsektorów instytucjonalnych
- Pracujący w gospodarce narodowej

---

<sup>2</sup> (Sarżyńska-Długosz i in., 2018)

<sup>3</sup> (Piotrowicz i in., 2011; Pęciło i Najmiec, 2015; Centralny Instytut Ochrony Pracy – Państwowy Instytut Badawczy w Warszawie i Instytut Medycyny Pracy im. Prof. dr hab. med. J. Nofera w Łodzi, 2016)

<sup>4</sup> Art. 16 Ustawy z dnia 20 lipca 2018 r. o Polskim Instytucie Ekonomicznym, Dz. U. 2018 poz. 1735.

<sup>5</sup> (GUS, 2020).

- Osoby powyżej 50. roku życia na rynku pracy
- Zasoby migracyjne
- Wynagrodzenia w gospodarce narodowej
- Warunki pracy
- Wypadki przy pracy
- Świadczenia z ubezpieczeń społecznych i pozaubezpieczeniowe
- Roczne badanie działalności gospodarczej przedsiębiorstw
- Statystyka strukturalna przedsiębiorstw niefinansowych i jednostek zależnych
- Kwartalne rachunki niefinansowe sektora instytucji rządowych i samorządowych
- Narodowy Rachunek Zdrowia
- Stowarzyszenia, fundacje, samorząd gospodarczy i zawodowy oraz społeczne jednostki wyznaniowe
- Partie polityczne.

Natomiast dane z używanego w ZUS systemu księgowo-finansowego SAP ERP są wykorzystywane przez GUS w ramach badań statystycznych dotyczących następujących tematów:

- Kwartalne rachunki niefinansowe sektora instytucji rządowych i samorządowych
- Rachunki finansowe sektora instytucji rządowych i samorządowych
- Dane fiskalne na potrzeby nadzoru budżetowego UE.

Część danych pochodzących z systemów informatycznych ZUS jest więc upubliczniana w postaci zagregowanych wskaźników i statystyk w opracowaniach GUS i jako taka jest dostępna dla wykonawców badań ewaluacyjnych.

## 3. Dotychczasowe wykorzystanie danych ZUS w badaniach współfinansowanych z EFS

### 3.1. Wykorzystanie danych indywidualnych w badaniach polskich

#### 3.1.1. Metaanaliza wyników badań ewaluacyjnych dotyczących oceny wsparcia z Europejskiego Funduszu Społecznego

Pierwszym badaniem ewaluacyjnym w obszarze EFS, w którym wykorzystano indywidualne dane ZUS, była „Metaanaliza wyników badań ewaluacyjnych dotyczących oceny wsparcia z Europejskiego Funduszu Społecznego”. Stało się to możliwe dzięki zmianie Ustawy o systemie ubezpieczeń społecznych, wprowadzonej wraz z Ustawą z dnia 11 lipca 2014 r. o zasadach realizacji programów w zakresie polityki spójności finansowanych w perspektywie finansowej 2014–2020 (Dz.U. 2014 poz. 1146). Dodany wtedy w Ustawie o systemie ubezpieczeń społecznych art. 50 ust. 3a stanowił, że dane uczestników projektów EFS zgromadzone na kontach ubezpieczonych i kontach płatników składek mogą być udostępniane ministrowi właściwemu do spraw rozwoju regionalnego. Zakres i cel tego udostępnienia ograniczono do spełnienia wymogów załącznika I lub załącznika II do Rozporządzenia EFS<sup>6</sup>, czyli w praktyce do ustalenia wartości wspólnych wskaźników produktu i rezultatu EFS. Po kolejnej nowelizacji, z roku 2017 (Dz.U. 2017 poz. 1475), do potencjalnych odbiorców danych dołączył minister właściwy do spraw pracy oraz minister właściwy do spraw zabezpieczenia społecznego. Dodano również ust. 3c, pozwalający na wykorzystanie danych osób niebędących uczestnikami projektów EFS w celu spełnienia wymogów art. 56 ust. 3 Rozporządzenia ogólnego<sup>7</sup>, czyli ewaluacji skuteczności, efektywności i wpływu programu. Zakres tych danych ograniczono do informacji gromadzonych w centralnym systemie teleinformatycznym polityki spójności w Polsce, kodu tytułu zatrudnienia i wysokości składek odprowadzanych z tego tytułu. Czytelną intencją dodania tego ustępu jest stworzenie możliwości pozyskania danych ZUS dla grupy kontrolnej, pozwalającej wnioskować o efektach netto interwencji.

Po pierwszej z wymienionych nowelizacji nastąpił wielomiesięczny proces rozmów między ówczesnym Ministerstwem Infrastruktury i Rozwoju a ZUS, w rezultacie których ustalono zakres przekazywanych danych i procedurę ich udostępniania. Zaplanowany wtedy sposób wykorzystania danych ZUS do wyliczenia wartości wskaźników rezultatu długoterminowego EFS został wcielony w życie w ramach badania ewaluacyjnego o nazwie „Metaanaliza wyników badań ewaluacyjnych dotyczących oceny wsparcia z Europejskiego Funduszu Społecznego”, realizowanego w latach 2016-2019 przez Evalu sp. z o.o. Procedura ta wyglądała następująco:

- 1) Każdego roku Ministerstwo generowało zbiór danych uczestników EFS z właściwego okresu sprawozdawczego. Zbiór ten przekazywało wykonawcy badania ewaluacyjnego, a także –

---

<sup>6</sup> Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 1304/2013 z dnia 17 grudnia 2013 r. w sprawie Europejskiego Funduszu Społecznego i uchylającego rozporządzenie Rady (WE) nr 1081/2006 (Dz. Urz. UE L 347 z 20.12.2013, str. 470, z późn. zm.).

<sup>7</sup> Rozporządzenie Parlamentu Europejskiego i Rady (UE) nr 1303/2013 z dnia 17 grudnia 2013 r. ustanawiające wspólne przepisy dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego, Funduszu Spójności, Europejskiego Funduszu Rolnego na rzecz Rozwoju Obszarów Wiejskich oraz Europejskiego Funduszu Morskiego i Rybackiego oraz ustanawiające przepisy ogólne dotyczące Europejskiego Funduszu Rozwoju Regionalnego, Europejskiego Funduszu Społecznego, Funduszu Spójności i Europejskiego Funduszu Morskiego i Rybackiego oraz uchylającego rozporządzenie Rady (WE) nr 1083/2006 (Dz. Urz. UE L 347 z 20.12.2013, str. 320, z późn. zm.)

w postaci ograniczonej do numeru PESEL, imienia, nazwiska i daty zakończenia udziału w projekcie – Zakładowi Ubezpieczeń Społecznych.

- 2) Zakład Ubezpieczeń Społecznych w ciągu 30 dni kalendarzowych identyfikował uczestników EFS we własnych bazach danych i przekazywał do Ministerstwa ich listę uzupełnioną o informacje na temat tytułu ubezpieczenia i wysokości podstawy wymiaru składek w momencie pomiaru wartości wskaźnika (w przypadku większości wskaźników - 6 miesięcy po zakończeniu udziału w projekcie).
- 3) Ministerstwo przekazywało listę uzupełnioną przez ZUS wykonawcy badania ewaluacyjnego.
- 4) Wykonawca łączył dane wygenerowane przez ZUS ze zbiorem danych uczestników EFS.
- 5) Dysponując połączonym zbiorem danych, Wykonawca wyliczał osiągnięte wartości wskaźników rezultatu długoterminowego, odnoszące się do posiadania pracy i prowadzenia działalności gospodarczej, na populacjach określonych zgodnie z definicją wskaźników i w wielu szczegółowych przekrojach.
- 6) Po zakończeniu prac Wykonawca usuwał dane pozyskane z ZUS.

Sposób analizy danych ZUS był bardzo prosty – na podstawie informacji o tytule ubezpieczenia danej osoby określano, czy należy ona do kategorii osób pracujących oraz czy należy do kategorii osób prowadzących działalność gospodarczą.

W ostatnim etapie ewaluacji wartości wskaźników obliczone przy wykorzystaniu danych ZUS zostały wykorzystane w dodatkowych, pogłębionych analizach, dostarczających wniosków na temat efektywności kosztowej interwencji<sup>8</sup> i czynników wpływających na skuteczność wsparcia<sup>9</sup> (Penszko, 2019). Nie korzystano natomiast z danych ZUS o osobach niebędących uczestnikami projektów, czyli nie użyto ich w celu oszacowania efektu netto.

### **3.1.2. Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój**

Było to pierwsze polskie badanie ewaluacyjne EFS perspektywy finansowej 2014-2020, niemające charakteru ewaluacji ex-ante, z którego opublikowany został raport. Wiązało się to z wczesnym uruchomieniem Inicjatywy na rzecz zatrudnienia ludzi młodych (YEI) i potrzebą corocznego raportowania Komisji Europejskiej osiągniętych wartości wskaźników dla tej interwencji. Dlatego w ramach badania ewaluacyjnego, realizowanego przez konsorcjum składające się ze spółek Imapp i IQS oraz Instytutu Badań Strukturalnych, opracowano pięć raportów wskaźnikowych, z których pierwszy został przygotowany w kwietniu 2016 roku, a ostatni w listopadzie 2019 roku. W pierwszych dwóch z tych raportów oszacowanie wartości wskaźników rezultatu długoterminowego YEI i EFS

---

<sup>8</sup> Dane o nakładach finansowych przeznaczonych na wsparcie uczestników, pochodzące z publicznie dostępnych list projektów, zestawiono z danymi o osiągniętych rezultatach zatrudnieniowych, ustalonych na podstawie danych z systemów SL 2014 i KSI ZUS. W ten sposób obliczono, jaka wysokość nakładów przypada na jednostkę efektu w postaci jednego uczestnika pracującego 6 miesięcy po zakończeniu udziału w projekcie. Tak rozumiany wskaźnik efektywności kosztowej obliczono i przedstawiono w podziale na okresy zakończenia udziału w projekcie, priorytety inwestycyjne, programy operacyjne i ich części składowe (do poziomu poddziałania).

<sup>9</sup> Wskaźnik rezultatu długoterminowego, którego wartość ustalono na podstawie danych ZUS, został wykorzystany jako miernik skuteczności wsparcia. Za pomocą modeli regresji wielokrotnej, analizując zbiór danych uczestników projektów pochodzący, poza tym wskaźnikiem, z systemu SL 2014, zweryfikowano osiem hipotez dotyczących czynników, od których zależała tak mierzona skuteczność.



oparto jeszcze na badaniach kwestionariuszowych. Począwszy od trzeciego raportu wskaźnikowego, przygotowanego w marcu 2018 roku, w celu wyliczenia wartości trzech z czterech wskaźników korzystano z danych ZUS (Kalinowski, 2019).

Badanie ewaluacyjne było realizowane na zamówienie tej samej jednostki ewaluacyjnej, wchodzącej w skład Departamentu Zarządzania EFS (później Departamentu EFS) w Ministerstwie Infrastruktury i Rozwoju (później Ministerstwie Rozwoju, później Ministerstwie Inwestycji i Rozwoju), co opisana wcześniej „Metaanaliza”. Zastosowano więc w nim te same, wypracowane już zasady współpracy z ZUS, w tym opisaną wyżej procedurę przetwarzania danych. Również cel i sposób przeprowadzenia analiz był analogiczny – informacje o tytule ubezpieczenia wykorzystano do wyliczenia wartości wskaźników rezultatu długoterminowego. Przeprowadzone w tym badaniu analizy kontrfaktyczne nie wykorzystywały danych ZUS, lecz inne dane administracyjne, pochodzące z systemu CeSAR używanego przez Powiatowe Urzędy Pracy (Kalinowski i in., 2020). W rekomendacjach ewaluator zasugerował natomiast, że z punktu widzenia oszacowania efektu netto wskazana byłaby integracja danych z systemów CeSAR i KSI ZUS (Baran, Hardy i Kalinowski, 2017) .

### **3.1.3. Badanie ewaluacyjne dotyczące wyliczenia wartości wskaźników rezultatu długoterminowego w ramach Europejskiego Funduszu Społecznego oraz Inicjatywy na rzecz zatrudnienia osób młodych**

Badanie to, rozpoczęte w lutym 2020 roku, służy między innymi wyliczeniu wartości wskaźników rezultatu długoterminowego EFS i YEI, czyli kontynuacji analiz wykonywanych w dwóch projektach ewaluacyjnych opisanych wyżej. Realizowane przez tego samego zamawiającego i wykonawcę, co w przypadku „Metaanalizy”, w taki sam sposób wykorzystuje dane ZUS do pomiaru wskaźników. W przyszłych etapach badania, zaplanowanych na lata 2022 i 2024, analizy mają wykraczać poza wyliczenie wartości wskaźników – ich rezultaty mają posłużyć do uzyskania szerszego obrazu efektów wsparcia, w tym jego efektywności. Dane ZUS mają zostać wykorzystane do obliczenia stopy zwrotu poniesionych nakładów, a także oceny jakości miejsc pracy zajmowanych przez dawnych uczestników projektów (dzięki informacji o podstawie wymiaru składki).

### **3.1.4. Monitoring losów absolwentów szkół wyższych**

Funkcjonujący obecnie Ogólnopolski system monitorowania ekonomicznych losów absolwentów szkół wyższych (ELA) nie jest finansowany ze środków EFS, nie ma również charakteru badania ewaluacyjnego. Wyrasta jednak z doświadczeń przedsięwzięcia badawczego zrealizowanego w ramach Programu Operacyjnego Kapitał Ludzki. Ponadto można podać przykład wykorzystania wytworzonych w jego ramach danych w badaniu ewaluacyjnym EFS. Za omówieniem historii i sposobu organizacji tego monitoringu w niniejszej ekspertyzie przemawia jednak przede wszystkim to, że można na tej podstawie wyciągnąć bardzo ważne wnioski co do perspektyw wykorzystania danych ZUS w ewaluacji EFS.

Za początek rozwoju ogólnopolskiego systemu monitorowania losów absolwentów szkół wyższych można uznać rozpoczęcie w czerwcu 2011 roku prac nad projektem badawczym „Monitorowanie losów absolwentów uczelni wyższych z wykorzystaniem danych administracyjnych Zakładu Ubezpieczeń Społecznych”. Został on zlecony Pracowni Ewaluacji Jakości Kształcenia Uniwersytetu Warszawskiego (PEJK UW) przez Instytut Badań Edukacyjnych (IBE). Środki na jego realizację pochodziły z budżetu projektu systemowego Poddziałania 3.1.1 POKL „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego”. Przedsięwzięcie zostało zwieńczone raportem opublikowanym w styczniu 2014 roku (Bożykowski i in., 2014).

W projekcie tym zakładano wypracowanie metodyki monitoringu losów absolwentów (np. zestawu wskaźników, procedury bezpiecznego przekazywania danych), a następnie jej zastosowanie na przykładzie Uniwersytetu Warszawskiego, w tym połączenie danych ZUS z danymi z rejestru uczelni. Na potrzeby testowania i dopracowania zaprojektowanych rozwiązań uzyskano ze strony ZUS dane dotyczące losowej próbki studentów UW, nie zawierające informacji pozwalających ustalić tożsamość jednostek, czyli nie stanowiące danych osobowych. W kolejnym etapie projektu zamierzano połączyć dane osobowe ZUS o losach zawodowych absolwentów z danymi o rodzaju i przebiegu studiów, co pozwoliłoby na zbadanie istniejących między nimi zależności, w tym formułowanie wniosków o efektywności kształcenia. Operacja ta nie doszła jednak do skutku z powodu wątpliwości zgłaszanych przez Generalnego Inspektora Ochrony Danych Osobowych. Wątpliwości tych nie udało się przelamać mimo opracowania specjalnych procedur zapewniających bezpieczeństwo danych osobowych i uzyskania ekspertyzy prawnej potwierdzającej ich legalność. Badacze musieli więc poprzestać na analizie pozyskanych z ZUS danych zanonimizowanych, które w założeniu miały jedynie służyć testowaniu rozwiązań metodycznych.

Dorobek metodologiczny projektu nie został jednak zaprzepaszczony. Znalezione bowiem rozwiązanie prawne, które umożliwiło kontynuację budowy systemu monitoringu losów absolwentów szkół wyższych. Było nim wprowadzenie ustawowych zapisów o tym, że minister właściwy do spraw szkolnictwa wyższego prowadzi monitoring absolwentów, a ZUS regularnie udostępnia dane na potrzeby tego monitoringu. W ten sposób podłożono podwaliny pod system ELA. Obecnie podstawę prawną do prowadzenia monitoringu stanowi art. 352 ustawy Prawo o szkolnictwie wyższym i nauce<sup>10</sup>.

Pierwszym rocznikiem objętym monitoringiem byli absolwenci z roku 2014, a pierwsze jego wyniki ogłoszono w maju 2016 roku. System funkcjonuje do chwili obecnej. W jego ramach co roku w terminie do 31 stycznia do ZUS przekazywane są dane pochodzące z systemu POL-on, którym administruje Ośrodek Przetwarzania Informacji, obejmujące numer PESEL, płeć, rok urodzenia i informacje o odbytych studiach dotyczące absolwentów, studentów i doktorantów z określonych lat. ZUS, korzystając z numeru PESEL jako klucza identyfikacyjnego, dołącza do nich informacje, pochodzące z kont ubezpieczonych i kont płatników, na temat składek na ubezpieczenia społeczne i zdrowotne (w tym podstawy wymiaru) oraz dodatkowo charakteryzujące te osoby (np. identyfikujące powiat zamieszkania). W ciągu 3 miesięcy Zakład jest zobowiązany przekazać z powrotem połączone dane, ale nie zapisując w nich numeru PESEL. Dzięki połączeniu informacji o przebiegu studiów i płaconych po upływie pewnego czasu składkach możliwe jest obliczenie na poziomie kierunku studiów lub uczelni wartości wskaźników opisujących losy zawodowe absolwentów. Przykładowo, do wskaźników takich należy „Średni czas (w miesiącach) od uzyskania dyplomu do podjęcia pierwszej pracy po uzyskaniu dyplomu”, „Procent absolwentów, którzy mieli doświadczenie bycia bezrobotnym w drugim roku po uzyskaniu dyplomu” czy „Średnie miesięczne wynagrodzenie z tytułu umów o pracę po uzyskaniu dyplomu”. Informacje o wartości tych wskaźników, w postaci zbiorów danych i wizualizacji, są udostępniane przez Ośrodek Przetwarzania Informacji na stronie internetowej systemu monitoringu (<https://ela.nauka.gov.pl>). Z powodu tajemnicy statystycznej nie są jednak ujawniane dla kierunków studiów ani innych zbiorowości, jeśli te liczą mniej niż 10 osób.

Dane z systemu ELA znalazły zastosowanie w badaniu ewaluacyjnym EFS pt. „Analiza zapotrzebowania na kompetencje w gospodarce i na rynku pracy wraz z badaniem wartości docelowej wspólnego wskaźnika długoterminowego POWER w obszarze szkolnictwa wyższego”

---

<sup>10</sup> Ustawa z dnia 20 lipca 2018 - Prawo o szkolnictwie wyższym i nauce (tekst jednolity: Dz. U. 2020, poz. 85, z późn. zm.).

(CEAPP UJ i IDEA Instytut, 2019). Jego celem było udzielenie odpowiedzi na pytania: 1) Czy oferowane w ramach Programu Rozwoju Kompetencji wsparcie przyczynia się do zmniejszania luk pomiędzy ofertą edukacyjną uczelni a potrzebami rynku pracy? 2) Jakiego rodzaju kompetencje i w jaki sposób powinny być wzmacniane w ramach kolejnych programów, aby kompetencje absolwentów były lepiej dopasowane do potrzeb rynku pracy i pracodawców? Nie korzystano bezpośrednio z indywidualnych danych ZUS, lecz z wyników monitoringu ELA. Wskaźniki ELA na poziomie uczelni i regionu wykorzystano do ustalenia, jakie są charakterystyki szkoły wyższej różnicujące sytuację zawodową absolwentów. Analizy polegały w tym przypadku na porównaniu wartości wskaźników ELA, dotyczących czasu podjęcia pierwszej pracy, ryzyka bezrobocia i przeciętnego wynagrodzenia, między poszczególnymi regionami, typami uczelni i kierunkami studiów. Na tej podstawie sformułowano wnioski co do czynników różnicujących szanse sukcesu zawodowego absolwentów. Ponadto wartości wskaźników na poziomie kierunków studiów posłużyły jako dane źródłowe do analizy skupień, tak aby ustalić podobieństwo między kierunkami studiów i podzielić na tej podstawie kierunki studiów na kilka grup. Był to jeden z początkowych etapów w złożonej procedurze doboru próby absolwentów do badań.<sup>11</sup>

### 3.1.5. Monitoring losów absolwentów szkół ponadpodstawowych

W ramach projektów Działania 2.15 POWER od kilku lat budowany jest system monitoringu losów zawodowych i edukacyjnych absolwentów szkół ponadpodstawowych. Monitoring ten nie ma charakteru badania ewaluacyjnego, jednak doświadczenia zebrane przy jego przygotowaniu rzucają światło na perspektywy wykorzystania indywidualnych danych ZUS w ewaluacji EFS, dlatego zostaną tu omówione.

System był rozwijany w ramach czterech projektów współfinansowanych ze środków EFS: dwóch pozakonkursowych oraz dwóch konkursowych. Są to następujące projekty:

- Pierwszy projekt pozakonkursowy: „Monitorowanie Losów Edukacyjno-Zawodowych Absolwentów i Młodych Dorosłych” (MLEZAiMD), nr umowy POWR.02.15.00-00-0004/16, realizowany przez Instytut Badań Edukacyjnych od maja 2016 do maja 2019;
- Drugi projekt pozakonkursowy: „Monitorowanie Losów Absolwentów Szkół Zawodowych - etap II” (MLASZ), nr umowy POWR.02.15.00-00-0002/19, realizowany przez Instytut Badań Edukacyjnych od czerwca 2019 do 2022 roku;
- Pierwszy projekt konkursowy: „Monitorowanie losów absolwentów szkół zawodowych - pierwsza runda” (MLASZ), nr umowy POWR.02.15.00-00-3001/16, realizowany przez PBS sp. z o.o. od stycznia 2017 do września 2018;
- Drugi projekt konkursowy: „Monitorowanie losów absolwentów szkół zawodowych - II i III runda” (MLASZ), nr umowy POWR.02.15.00-00-3001/19, realizowany przez PBS sp. z o.o. i Danae sp. z o.o. od listopada 2019 do 2022 roku.

---

<sup>11</sup> Przy doborze próby kontrolnej służącej do analiz kontrfaktycznych założono mianowicie, że absolwenci „kontrolni” powinni mieć ukończony kierunek studiów z tej samej grupy, co odpowiadający im absolwenci, będący beneficjentami Programu. Przyjmując tę zasadę jako ważne kryterium, wyłoniono dwie próby absolwentów – beneficjentów oraz nie-beneficjentów – na których przeprowadzono badanie kwestionariuszowe. Dane z tego badania poddano następnie analizom kontrfaktycznym w celu oszacowaniu efektu netto interwencji. O losach zawodowych absolwentów wnioskowano w tym przypadku na podstawie odpowiedzi udzielanych przez nich w ankiecie, a nie na podstawie danych ZUS, co wynikało z braku dostępu do danych indywidualnych ZUS

Dane ZUS zamierzano wykorzystać już w pierwszym projekcie pozakonkursowym. Planowano ich połączenie z wynikami badań kwestionariuszowych (prowadzonymi w ramach projektów konkursowych), dlatego zbierano od respondentów zgody na wykorzystanie ich danych gromadzonych w KSI ZUS. Przeprowadzona później ekspertyza prawna wskazała jednak, że zgody te nie są wystarczającą przesłanką do przetwarzania danych indywidualnych ZUS na potrzeby monitoringu losów absolwentów, w związku z czym plany pozyskania tych danych zarzucono. Zebrane doświadczenia prowadziły do wniosku, że właściwą przesłanką do użycia danych administracyjnych w monitoringu losów absolwentów będzie nie zgoda osób, których te dane dotyczą, lecz obowiązek prawny lub interes publiczny. W związku z tym w czasie realizacji drugiego projektu pozakonkursowego plany wykorzystania danych ZUS oparto na opracowanym projekcie nowelizacji ustawy Prawo oświatowe, która będzie stanowić podstawę do wykorzystywania danych ZUS. Projekt taki w dniu 20 listopada 2020 roku został przekazany do uzgodnień międzyresortowych, konsultacji publicznych i zaopiniowania przez właściwe instytucje<sup>12</sup>.

Tworzony system monitoringu losów absolwentów szkół ponadpodstawowych jest wzorowany na systemie monitorowania losów absolwentów szkół wyższych ELA, opisanym w punkcie poprzednim. Ma być oparty na podobnych rozwiązaniach prawnych i korzystać z analogicznego zakresu danych ZUS. Inaczej natomiast zorganizowany ma być przepływ danych i podział zadań między instytucje: o ile w przypadku ELA kluczowy proces przetwarzania danych odbywa się w ZUS, o tyle na potrzeby monitoringu losów absolwentów szkół ponadpodstawowych za łączenie danych z różnych źródeł (ZUS, System Informacji Oświatowej, Okręgowe Komisje Egzaminacyjne, system POL-on) odpowiedzialne będzie Centrum Informatyczne Edukacji (CIE). Identyfikatorem osoby pozwalającym na dokonanie tej operacji jest numer PESEL. Po połączeniu danych z poszczególnych instytucji mają one zostać zanonimizowane i przekazane do Instytutu Badań Edukacyjnych (IBE), którego zadaniem będą analizy, w tym obliczenie wartości określonego zestawu wskaźników. Przedmiotem zainteresowania są zarówno losy zawodowe (np. podjęcie zatrudnienia), jak i edukacyjne (np. kontynuacja nauki na studiach). Planuje się, że powstałe w rezultacie tych działań raporty na poziomie poszczególnych szkół ponadpodstawowych będą udostępniane jedynie zainteresowanym szkołom za pośrednictwem Systemu Informacji Oświatowej. Natomiast wartości wskaźników na wyższym poziomie agregacji, prezentowane w wielu przekrojach, mają być publikowane przez IBE.

Podsumowując, dotychczas w ramach omawianych projektów monitoringu losów absolwentów szkół ponadpodstawowych nie doszło do rzeczywistego wykorzystania danych ZUS, niemniej zaprojektowano służące temu rozwiązania. Nie przewidziano w nich wykorzystania połączonych danych na potrzeby ewaluacji EFS. Aktualnie proponowane zapisy prawne stwarzają barierę dla użycia powstałych zbiorów danych osobowych do celów innych niż monitoring losów absolwentów szkół ponadpodstawowych. Warunkiem zastosowania zgromadzonych danych w ewaluacji EFS jest więc albo to, by wpisywało się w zakres takiego monitoringu, albo uznanie go za dalsze przetwarzanie danych osobowych do celów naukowych lub statystycznych, które zgodnie z RODO nie jest naruszeniem zasady ograniczenia celu przetwarzania. Mogą się jednak przy tym pojawić wątpliwości co do interpretacji aktów prawnych. Dlatego najlepszym rozwiązaniem byłoby zawarcie w planowanej nowelizacji zapisów, stwierdzających *explicite*, że w ramach prowadzonego monitoringu minister właściwy do spraw oświaty może prowadzić i zlecać analizy służące ewaluacji prowadzonych polityk publicznych, w tym unijnej polityki spójności. Otworzyłyby to pole do wykorzystania gromadzonych danych osobowych na potrzeby ewaluacji interwencji skierowanej do szkół ponadpodstawowych. Trudno jednak sobie wyobrazić, by dane te przekazywane były do prywatnego podmiotu będącego wykonawcą badania ewaluacyjnego. Ewentualne łączenie danych

---

<sup>12</sup> <https://legislacja.rcl.gov.pl/projekt/12340511/katalog/12740340#12740340> .

zbieranych w ramach monitoringu losów absolwentów szkół ponadpodstawowych z danymi uczestników projektów EFS (z wykorzystaniem numeru PESEL w charakterze klucza identyfikującego) musiałyby więc zapewne być wykonywane przez CIE.

Poza indywidualnymi danymi osobowymi, w ewaluacji EFS przydatne mogą okazać się również dane na poziomie poszczególnych szkół ponadpodstawowych. Przykładowo, ewaluator mógłby porównać szkoły objęte i nieobjęte wsparciem EFS pod względem losów absolwentów, aby na tej podstawie, stosując techniki kontrfaktyczne, wnioskować o efektach interwencji. Wymagałoby to rozluźnienia założenia, że wyniki analiz na poziomie szkoły są udostępniane tylko zainteresowanej szkole.

### 3.2. Wykorzystanie danych zagregowanych udostępnianych przez ZUS

Jak stwierdzono w rozdziale 2.2, opracowania statystyczne publikowane przez ZUS są źródłem wielu informacji, z których część mogłaby zostać wykorzystana w ewaluacji EFS, choćby w roli danych kontekstowych. Stopień wykorzystania tego typu materiałów w raportach ewaluacyjnych jest jednak znikomy. W ramach niniejszej ekspertyzy dokonano przeglądu dostępnych raportów uwzględnionych w Zestawieniu badań ewaluacyjnych w obszarze EFS w latach 2019 – 2020<sup>13</sup>. Ograniczono się przy tym do trzech obszarów: rynku pracy (14 raportów), włączenia społecznego (7 raportów) i zdrowia (5 raportów). Odwołania do zestawień danych publikowanych przez ZUS nie odnaleziono w żadnym z 26 uwzględnionych raportów. Pomijając badania opisane w poprzednim podrozdziale, ZUS jako źródło informacji wystąpił jedynie w dwóch raportach z obszaru rynku pracy, przy czym w jednym z tych przypadków skorzystano z zamieszczonych na jego stronie internetowej materiałów informacyjnych dotyczących wysokości świadczeń, a w drugim powołano się na wypowiedź prasową pracowniczki tej instytucji. Wydaje się, że udostępniane przez ZUS dane zagregowane nie są dostrzegane przez wykonawców badań ewaluacyjnych, choć mogłyby znaleźć pewne zastosowanie w ewaluacji EFS. Przykładowo, statystyki dotyczące cudzoziemców zgłoszonych do ubezpieczeń emerytalnych i rentowych (w podziale na kraj pochodzenia, województwo siedziby płatnika, sekcje PKD itd.) mogłyby stanowić dane kontekstowe w ewaluacji poruszającej kwestie trafności wsparcia skierowanego do kadr przedsiębiorstw z punktu widzenia potrzeb grup docelowych.

### 3.3. Zagraniczne przykłady wykorzystania danych ubezpieczeń społecznych i zdrowotnych w ewaluacji EFS

#### 3.3.1. Szwecja

Specyfiką krajów skandynawskich jest szerokie wykorzystanie danych administracyjnych w badaniach społecznych. W Szwecji instytucją pełniącą przy tym kluczową rolę jest Centralny Urząd Statystyczny (Statistiska centralbyrån, SCB). Gromadzi on dane z rejestrów administracyjnych regularnie przekazywane mu przez wiele instytucji publicznych, integruje dane z różnych źródeł, korzystając z wykorzystywanego w nich wspólnego identyfikatora osoby (odpowiednika numeru PESEL), a następnie udostępnia je na potrzeby badań (Jasiński, Bożykowski, Zając, Styczeń i Izdebski, 2015). Zanonimizowane dane indywidualne z takich dziedzin, jak edukacja, dochody i zatrudnienie, są udostępniane przedstawicielom świata nauki z zachowaniem procedur zapewniających bezpieczeństwo danych. W roku 2005 wprowadzono system informatyczny MONA, który umożliwia wygenerowanie i pobranie zbiorów danych indywidualnych za pośrednictwem Internetu (UNECE, 2007, pp. 34-35). Poczynając od roku 2011 możliwe jest również powiązanie danych dotyczących

<sup>13</sup> Zestawienie opublikowane 6 marca 2020 roku pod adresem <https://www.power.gov.pl/strony/o-programie/raporty/zestawienie-badan-ewaluacyjnych-w-obszarze-efs-w-latach-2019-2020/>.

konkretnej osoby z danymi o gospodarstwie domowym (Thörn, 2012). Wobec funkcjonowania takiego systemu wiele badań społecznych, które w innych krajach realizuje się z użyciem metod kwestionariuszowych, w Szwecji przeprowadza się na danych administracyjnych.

Nie inaczej jest w przypadku monitoringu i ewaluacji EFS. Przykładem może być opracowanie podsumowujące realizację interwencji EFS w latach 2008-2013 (Sävenstrand, 2014). Zawiera ono opis populacji beneficjentów i uczestników oraz sytuacji zawodowej i materialnej uczestników w trakcie udziału w projekcie oraz w ciągu roku po jego zakończeniu. Godne uwagi jest to, że rozdziały te oparte są wyłącznie na danych zastanych, a mianowicie na własnych danych Instytucji Zarządzającej, pochodzących przede wszystkim z formularzy wypełnianych przez beneficjentów, oraz na informacjach pochodzących z ośmiu rejestrów publicznych, w tym sześciu rejestrów Centralnego Urzędu Statystycznego. Znajduje się wśród nich baza danych LISA, w której gromadzone są m.in. dane związane z ubezpieczeniem zdrowotnym. Zaczepnięto z niej informacje na temat dochodów uczestników.

W analogiczny sposób dokonano pomiaru wartości wskaźników EFS w okresie programowania 2007-2013 (Sävenstrand, 2014, strony 69-71). Podobnie postąpiono w okresie programowania 2014-2020 – na tyle, na ile pozwolił na to rozbudowany system wskaźników. Zgodnie z przyjętym rozwiązaniem, do rejestru uczestników projektów EFS przekazywane były dane administracyjne z instytucji będącej szwedzkim odpowiednikiem ZUS, publicznych służb zatrudnienia oraz rejestru gromadzącego informacje z zakresu demografii, edukacji i rynku pracy. Instytucja zarządzająca, po połączeniu tych zbiorów danych, przekazywała je do Centralnego Urzędu Statystycznego, gdzie były zasilane dodatkowymi danymi o projektach przekazywanymi przez Instytucję zarządzającą oraz przez beneficjentów logujących się na stworzoną w tym celu platformę. Centralny Urząd Statystyczny, dysponujący stworzoną w ten sposób zintegrowaną bazą danych, był odpowiedzialny za wyliczenie osiągniętych wartości wskaźników EFS i YEI. W przypadku YEI dane administracyjne wykorzystano do ustalenia wartości wszystkich wskaźników produktu i części wskaźników rezultatu długoterminowego. Nie stosowano ich natomiast w przypadku wskaźników rezultatu bezpośredniego, czyli tych, których wartość jest mierzona bezpośrednio po zakończeniu przez uczestnika udziału w projekcie (Thörn, 2012).

### 3.3.2. Litwa

Interesujący przykład wykorzystania danych o ubezpieczeniach społecznych w ewaluacji EFS można odnaleźć również w bliższym nam geograficznie i kulturowo kraju Europy, jakim jest Litwa. Udało się w nim stworzyć ramy międzyresortowej współpracy instytucji publicznych, służącej integracji danych osobowych z różnych źródeł. Stanowiło to sposób na ominięcie barier wiążących się z przekazywaniem danych osobowych zewnętrznemu ewaluatorowi, w typowej sytuacji należącemu do sektora prywatnego.

Przykładem w ten sposób przeprowadzonej ewaluacji są analizy kontrfaktyczne wpływu aktywnych polityk rynku pracy finansowanych ze środków EFS na sytuację zawodową uczestników, zrealizowane w latach 2014-2015 na zlecenie litewskiego Ministerstwa Zabezpieczenia Społecznego i Pracy przez wileński Public Policy and Management Institute. Wykonawca otrzymał zanonimizowane dane uczestników projektów z lat 2008-2012 od publicznych służb zatrudnienia, które realizowały projekty EFS. Dane uczestników, zawierające tym razem zaszyfrowany kod identyfikacyjny osób, a także dane osób stanowiących pulę kontrolną zostały ponadto przekazane przez publiczne służby zatrudnienia do państwowej instytucji Sodra, odpowiedzialnej za zarządzanie ubezpieczeniami społecznymi. Sodra, dysponując tym samym zaszyfrowanym kodem identyfikacyjnym, dołączyła informacje o sytuacji zawodowej wskazanych osób po zakończeniu udziału w projekcie. Następnie przekazała połączone,

lecz już zanonimizowany zbiór danych, dotyczący około 100 tys. uczestników i 800 tys. osób stanowiących pulę kontrolną, do zespołu kierującego ewaluacją, składającego się z przedstawicieli PPMI i Ministerstwa. Korzystając z tak obszernego zbioru danych indywidualnych, ewaluator przeprowadził planowane analizy kontrfaktyczne (Jonavičienė, 2015) (Jonavičienė, 2016).



## 4. Specyfika danych ZUS

W kontekście możliwości wykorzystania w ewaluacji EFS danych ZUS, można wskazać na szereg ich specyficznych cech. Zostały one omówione poniżej.

### 4.1. Kwestie związane z pozyskaniem danych

#### Podstawy prawne

Dotychczasowe doświadczenia skłaniają do wniosku, że warunkiem wykorzystania indywidualnych danych pochodzących z systemów informatycznych ZUS jest wprowadzenie odpowiednich zapisów do aktów prawnych. Wszystkie omawiane w rozdziale 3.1 badania, w których doszło do połączenia danych ZUS z innymi zbiorami danych osobowych, stały się możliwe dzięki specjalnie w tym celu dokonanym zmianom legislacyjnym. Bardzo pouczające są niepowodzenia na początkowych etapach monitoringu losów absolwentów szkół wyższych oraz losów absolwentów szkół ponadpodstawowych. Wypływa z nich między innymi wniosek, że trudno oprzeć przetwarzanie danych osobowych przechowywanych przez ZUS na przesłance zgody osób, których dane dotyczą. Zastosowanie znajdują natomiast inne przesłanki wskazane w RODO, a mianowicie obowiązek prawny oraz realizacja zadań w interesie publicznym lub w ramach sprawowania władzy publicznej. Oznacza to ograniczenie możliwości użycia indywidualnych danych ZUS w badaniach ewaluacyjnych do przypadków przewidzianych i zapisanych w aktach prawnych.

#### Generowanie danych

Zasoby danych ZUS przechowywane są w wielu systemach informatycznych o złożonej, częściowo rozproszonej strukturze, co przybliży rozdział 2. Dodatkowo w chwili obecnej nie jest publicznie dostępna pełna informacja o zawartości i sposobie organizacji tych baz danych. Mogłoby to stanowić bardzo poważne utrudnienie dla wykorzystania danych ZUS w badaniach ewaluacyjnych. Zostały już jednak wypracowane narzędzia informatyczne, które pozwalają na migracje danych między systemami oraz na generowanie zbiorów danych pochodzących z wielu baz. Znaczne skomplikowanie zasobów danych ZUS należy jednak wziąć pod uwagę przy planowaniu nowych sposobów ich wykorzystania. W szczególności może to wymagać długotrwałych konsultacji z pracownikami ZUS w celu ustalenia zakresu przekazywanych danych, odpowiednio do potrzeb oraz możliwości technicznych i prawnych. Należy to uwzględnić w harmonogramie prowadzonych działań.

#### Organizacja przetwarzania danych i związane z nim obciążenia mocy przerobowych ZUS

Czasochłonne mogą być nie tylko przygotowania do pozyskania danych, ale również sam proces generowania właściwego zbioru przez ZUS. W tym kontekście istotne są regularne działania ZUS, które w dużym stopniu angażują jego personel oraz moc obliczeniową systemów. Należy do nich waloryzacja roczna i generowanie informacji o stanie kont ubezpieczonych. Procesy te przypadają na koniec I kwartału, II kwartał i początek III kwartału każdego roku. Wcześniej, bo do dnia 31 stycznia, do ZUS przekazywane są dane z systemu POL-on na potrzeby prowadzenia monitoringu losów absolwentów szkół wyższych. ZUS, zgodnie z Prawem o szkolnictwie wyższym i nauce, ma 90 dni na ich połączenie z informacjami zawartymi na kontach ubezpieczonych i płatników składek oraz odesłanie rozszerzonego w ten sposób zbioru danych. Ponadto w styczniu i lutym w najbliższych latach ZUS w podobny sposób generuje zbiory na potrzeby wyliczenia osiągniętych wartości wskaźników rezultatu długoterminowego EFS i YEI. W przypadku projektowania nowych zastosowań dla danych pochodzących z systemów informatycznych ZUS warto wziąć pod uwagę te okresowe obciążenia i planować pozyskiwanie danych na pozostałe miesiące.



Istotne znaczenie ma organizacja procesu przepływu danych i związane z tym przypisanie zadań poszczególnym instytucjom. W omówionych wyżej polskich badaniach zastosowano następujący model:

- Krok 1: instytucja uprawniona do skorzystania z danych generuje zbiór danych zawierający numer PESEL i przekazuje go do ZUS;
- Krok 2: ZUS dołącza do otrzymanego zbioru danych dane pochodzące z własnych zasobów danych;
- Krok 3: ZUS zwraca instytucji uprawnionej zbiór danych z dołączonymi informacjami.

Różnice między badaniami dotyczą zakresu danych. W badaniach ewaluacyjnych służących wyliczeniu osiągniętych wartości wskaźników rezultatu długoterminowego EFS i YEI zakres danych przekazywanych do ZUS w kroku 1. jest ograniczony do minimum, skromny jest również zakres danych dołączanych w kroku 2., natomiast zbiór danych zwracanych w kroku 3. zawiera numer PESEL, ponieważ potrzebne jest jego dalsze łączenie. W monitoringu losów absolwentów szkół wyższych ELA zakresy danych przekazywanych do ZUS w kroku 1. oraz dołączanych w kroku 2. są znacznie obszerniejsze, ale zbiór zwracany w kroku 3 nie zawiera numeru PESEL. Nakłada to na ZUS dodatkowe zadanie anonimizacji danych oraz obowiązek wykonania bardziej złożonych czynności związanych z łączeniem zbiorów, ale pozwala ograniczyć przekazywanie danych osobowych poza ZUS. Projektowany sposób organizacji przesyłu danych w monitoringu losów absolwentów szkół ponadpodstawowych można uznać w obszarze współpracy z ZUS za wariant pośredni. Do badań ewaluacyjnych zbliża go brak anonimizacji danych przez ZUS, natomiast zakres danych przetwarzanych w kroku 1 i 2 jest podobny jak w systemie ELA.

Wprowadzanie nowych zastosowań danych ZUS w ewaluacji EFS niesie ze sobą realne ryzyko przeciążenia zasobów kadrowych i technologicznych ZUS. Warto pamiętać, że KSI ZUS jest systemem scentralizowanym, w którym prawie cały zasób informacji jest gromadzony i przetwarzany w tzw. Komputerze Centralnym, w którego skład wchodzi (lub wchodziły kilka lat temu) trzy komputery mainframe IBM połączone ze sobą w układzie syspleksu równoległego (Asseco Poland S.A., 2020). Centralizacja systemu ułatwia dostęp do danych, ale zwiększa ryzyko osiągnięcia granic jego wydolności. Stąd dostępna moc obliczeniowa i wynikający z niej czas potrzeby na wygenerowanie danych są kwestiami, które również należy wziąć pod uwagę.

Jak widać, w obrębie stosowanego dotąd w Polsce modelu napotykamy na rozwiązania różniące się pod względem stopnia obciążenia zasobów ZUS. W porównaniu z ELA, organizacja przesyłu danych przyjęta w dotychczasowych badaniach ewaluacyjnych EFS zmniejsza to obciążenie kosztem przekazywania danych osobowych poza ZUS.

Wymieć można dwa alternatywne modele, minimalizujące zaangażowanie zasobów ZUS.

Pierwszy z nich polegałby na regularnym transferze danych indywidualnych poza ZUS, do wyspecjalizowanej instytucji udostępniającej dane na potrzeby badań ewaluacyjnych. Jest to model stosowany w Szwecji, w której Centralny Urząd Statystyczny gromadzi dane z rejestrów innych instytucji, np. Urzędu Skarbowego, i jest odpowiedzialny za ich połączenie i udostępnianie na potrzeby badań naukowych. Jak zauważają autorzy z PEJK UW, taki sposób organizacji umożliwia szerokie wykorzystanie danych z rejestrów administracyjnych, zapobiegając zablokowaniu instytucji je prowadzących nadmierną liczbą wniosków o udostępnienie informacji (Jasiński, Bożykowski, Zając, Styczeń i Izdebski, 2015). Dodatkową zaletą jest ogromne ułatwienie łącznego wykorzystania danych

z różnych źródeł, co mogłoby być niezwykle przydatne również w ewaluacji EFS (np. w integracji danych ZUS i KRUS).

Również w Polsce Główny Urząd Statystyczny gromadzi dane z wielu publicznych instytucji, w tym Zakładu Ubezpieczeń Społecznych. Wykorzystuje je następnie do realizacji badań statystycznych i opracowania zestawień na poziomie zagregowanym. Nie dokonuje jednak łączenia danych administracyjnych w celu dalszego udostępniania zbiorów danych indywidualnych. Potencjalnie GUS mógłby w polskim systemie ewaluacji polityk publicznych pełnić rolę centralnej instytucji zajmującej się integracją danych. W praktyce uczyniono jednak inny krok w stronę opisywanego tu alternatywnego modelu, a jest nim projekt Zintegrowanej Platformy Analitycznej (ZPA), realizowany obecnie w ramach Programu Operacyjnego Polska Cyfrowa (POPC.02.02.00-00-0015/18-00) przez partnerstwo, w skład którego wchodzi Kancelaria Prezesa Rady Ministrów (w miejsce dawnego Ministerstwa Cyfryzacji), wybrane ministerstwa i ośrodki akademickie. Celem prowadzonych prac jest między innymi udostępnienie danych administracji publicznej do badań naukowych wspierających podejmowanie decyzji przez twórców polityk publicznych. Aby dorobek projektu ZPA okazał się trwały, potrzebne wydaje się ustawowe uregulowanie funkcjonowania platformy po zakończeniu projektu i powierzenie jej prowadzenia określonej instytucji publicznej. Efekty prowadzonych obecnie działań służących integracji danych publicznych będą widoczne w przyszłości. Dopóki jednak trwają prace nad ZPA, nieefektywne wydaje się tworzenie alternatywnego systemu integracji danych w instytucjonalnych ramach GUS.

Opisywane rozwiązania eliminują ryzyko przeciążenia kadr ZUS, wymagając od tej instytucji jednokrotnego wygenerowania i przesłania danych na potrzeby wielu projektów badawczych, zgodnie z regularnym i przewidywalnym wzorem. Zmniejszyłyby również obciążenie systemu informatycznego ZUS – w tym większym stopniu, im rzadziej następowałaby migracja danych do platformy zewnętrznej (związana jest z tym oczywiście kwestia aktualności danych). Ciężar przetwarzania danych byłby w znacznej mierze przeniesiony na zewnętrzną, wyspecjalizowaną instytucję. Należy przy tym zastrzec, że ze względu na obszerność, złożoność i rozproszenie zasobów danych ZUS trudno sobie wyobrazić ich przekazywanie w całości. Warto wspomnieć, że sam czas potrzebny na operację wewnętrznego transferu danych o świadczeniach emerytalno-rentowych ze starszych systemów informatycznych do KSI ZUS był na początku roku 2019 szacowany na ponad 4 lata. Regularne przesyłanie danych na platformę zewnętrzną może więc dotyczyć jedynie uzgodnionego podzbioru danych ZUS. Kluczowe znaczenie ma to, czy wśród danych tych znajdują się informacje użyteczne przy ewaluacji EFS.

Inny alternatywny model zakłada udzielenie dostępu do danych ZUS osobom spoza tej instytucji. Można sobie na przykład wyobrazić sytuację, w której upoważnieni (wg ściśle określonego poziomu dostępu, celu, sposobu przetwarzania np. w odpowiednio zabezpieczonych systemach / środowisku pracy) pracownicy jednostek ewaluacyjnych mogliby za pomocą specjalnej aplikacji (np. typu webAPI) wygenerować określone dane z baz ZUS na potrzeby prowadzonych badań – podobnie jak obecnie generują dane z systemów obsługujących realizację programów operacyjnych. Byłoby to rozwiązanie minimalizujące zaangażowanie kadr ZUS. Obciążałoby jednak system informatyczny ZUS w sposób mało przewidywalny z punktu widzenia jego administratorów. Nie wymagałoby istnienia zewnętrznej platformy gromadzącej dane publiczne, która jeszcze nie funkcjonuje w Polsce, lecz prawdopodobnie konieczny byłby rozwój aplikacji zapewniających dostęp do baz danych ZUS. Potencjalną zaletą byłaby możliwość skorzystania z szerszego zakresu danych, nie ograniczonego do wąskiego i ściśle określonego zestawu informacji przekazywanych na zewnątrz. Warunkiem byłoby jednak uprzednie nabycie przez pracowników jednostki ewaluacyjnej orientacji w bardzo złożonej strukturze danych ZUS oraz umiejętności posługiwania się oprogramowaniem służącym do ich przetwarzania. Ciemną

stroną swobodniejszego dostępu do danych osobowych są ponadto problemy z zagwarantowaniem ich bezpieczeństwa. Trudno byłoby uzasadnić i zapewnić, że dostęp jednostek ewaluacyjnych do danych osobowych, a w szczególności danych wrażliwych, którymi dysponuje ZUS, nie jest nadmierny w stosunku do celów ich przetwarzania. Dlatego zastrzeżenia prawne, obok zastrzeżeń formułowanych z punktu widzenia zarządzaniem systemem informatycznym, stanowiłyby prawdopodobnie poważne bariery we wcieleniu w życie opisywanego modelu. Sprawiają one, że jego wdrożenie wydaje się mało realne.

Na koniec warto wspomnieć o jeszcze jednym modelu organizacji przetwarzania danych, który został z powodzeniem zastosowany w celu wykorzystania informacji gromadzonych przez GUS na potrzeby określonych ewaluacji wsparcia przedsiębiorstw. Dotyczyły one m.in. regionalnej pomocy inwestycyjnej w ramach Celu tematycznego 3. polityki spójności UE realizowanej w Polsce (Evalu, S-T-O-S i WiseEuropa, 2020) oraz funkcjonowania specjalnych stref ekonomicznych, a zatem interwencji publicznych finansowanych z innych środków niż EFS. Analizy przeprowadzono na danych zbieranych w sprawozdaniach SP i - w przypadku pierwszej ewaluacji - PNT-02, składanych przez przedsiębiorców do GUS. Ze względu na tajemnicę statystyczną nie było możliwe przekazanie danych indywidualnych poza GUS. Dlatego przyjęto rozwiązanie, zgodnie z którym analizy wykonywali pracownicy GUS, według algorytmów opracowanych przez zewnętrznego ewaluatora i przekazywanych w postaci skryptów pakietu statystycznego. Pliki wynikowe generowane po zastosowaniu otrzymanych skryptów pracownicy GUS przekazywali (po sprawdzeniu czy nie zawierają danych o indywidualnych przedsiębiorstwach) do ewaluatora, który na ich podstawie planował i podejmował dalsze działania. Model ten gdyby zastosować w przypadku danych ZUS, w porównaniu z poprzednio opisanymi, generowałby największe obciążenia po stronie Zakładu lub ewentualnie innej instytucji gromadzącej dane publiczne, w tym dane przekazywane jej z ZUS, a także po stronie zewnętrznego ewaluatora. Jednakże, model został tu opisany z tego względu, że pozwala uniknąć przekazywania danych indywidualnych poza dysponującą nimi instytucję. Z tego powodu może okazać się przydatny w ewaluacji EFS w obliczu problemów związanych z ochroną danych.

### **Koszty pozyskania danych**

Niemal całość kosztów pozyskania danych ZUS stanowią koszty pracy osób zaangażowanych w to przedsięwzięcie. W typowej sytuacji, gdy obowiązek wygenerowania i/lub łączenia danych został nałożony prawnie na ZUS, koszty pracy jego personelu są pokrywane ze środków na działalność bieżącą tej instytucji. Pozyskanie danych wiąże się również z kosztami po stronie instytucji zamawiającego i wykonawcy ewaluacji, które są nieznaczne na etapie odbioru danych, ale mogą być znaczne na etapie projektowania badań i analiz, gdy zachodzi potrzeba szczegółowego zaplanowania i przygotowania pozyskiwania danych, wypracowania odpowiednich uzgodnień z ZUS itp. Ponadto pewne nakłady pracy wiążą się z przygotowaniem zbiorów danych przekazywanych do ZUS (np. listy PESEL uczestników projektów EFS) w celu przyłączenia dodatkowych danych, a także dalszym przetwarzaniem pozyskanych danych – ich wielkość zależy od zakresu wymaganych czynności. Koszty po stronie wykonawcy ewaluacji zwykle obciążają budżet konkretnego badania ewaluacyjnego. Taka sytuacja może również dotyczyć kosztów po stronie zamawiającego, chyba że są one włączone do kosztów na ogólną działalność jednostki ewaluacyjnej. Niemniej w sytuacji, gdy zasady współpracy z ZUS są już uzgodnione, pozyskanie danych na tej podstawie wiąże się ze zdecydowanie mniejszymi kosztami niż zbieranie ich w dużych badaniach terenowych, zakładających kontakt telefoniczny lub osobisty z respondentem.

## 4.2. Kwestie związane z populacyjnym charakterem danych

Dane będące w posiadaniu ZUS można uznać za dane dotyczące populacji mieszkańców Polski, płatników składek, orzeczeń lekarskich itd. Jest to cecha typowa dla rejestrów administracyjnych, zdecydowanie odróżniająca je od danych pozyskiwanych w różnego typu badaniach kwestionariuszowych (w tym na potrzeby ewaluacji), które zazwyczaj są zbierane tylko dla dobranej próby, a nawet jeśli badaniu podlega cała populacja, nie są dostępne dla osób, które nie wypełniły kwestionariusza. Wyczerpujący charakter danych, rozumiany jako objęcie nimi całej populacji, jest ważną zaletą korzystania z danych ZUS. Pozwala uniknąć błędu statystycznego, związanego z losowaniem próby, a także, groźniejszego w skutkach, błędu systematycznego związanego z nielosowo rozłożoną niedostępnością niektórych jednostek dla badania. Dokładniej rzecz ujmując, również dane ZUS nie są całkowicie wolne od problemu niedostępności niektórych danych, z powodów, które zostaną omówione niżej. Problem ten ma jednak inne przyczyny i jest nieporównywalnie mniejszy niż w przypadku badań kwestionariuszowych, w których wynika przede wszystkim z braku możliwości nawiązania kontaktu z częścią osób z grupy docelowej oraz z udzielanych odmów udziału w badaniu. Wyniki analiz przeprowadzonych na danych administracyjnych są w związku z tym znacznie bardziej precyzyjne i wiarygodne. Można powiedzieć, że dzięki wykorzystaniu danych ZUS osiągnięte wartości wskaźników rezultatu długoterminowego EFS są wyliczane, a nie szacowane, jak w przypadku oparcia się na wynikach badań kwestionariuszowych.

W sytuacjach, gdy dane ZUS są wykorzystywane do ustalenia, czy dana osoba posiada w określonym momencie pracę – tak jak ma to obecnie miejsce przy wyliczaniu niektórych wartości wskaźników rezultatu długoterminowego EFS – populacyjny charakter danych nie oznacza, że dla każdej osoby pozyskiwane są dane z systemów informatycznych ZUS. W miejsce tego przyjmowane jest założenie, że brak danych ZUS dla danej osoby oznacza, że nie jest ona w danym momencie ubezpieczona z tytułu wskazującego na posiadanie pracy, a zatem nie posiada pracy. Jest to podejście racjonalne, ale niosące ze sobą ryzyko zaniżenia odsetka liczby pracujących o wartość odpowiadającą udziałowi osób, które posiadają pracę, lecz nie figurują w wykorzystywanym zbiorze danych ZUS. Warto zdawać sobie sprawę z możliwych przyczyn wystąpienia takiej sytuacji.

Po pierwsze należy zauważyć, że dana osoba nie zostanie zidentyfikowana w zbiorze danych ubezpieczonych, jeżeli nie zostanie zgłoszona do ubezpieczenia. W związku z tym w danych ZUS nie znajduje odzwierciedlenia praca, która nie wiąże się z obowiązkami ubezpieczeniowymi, na przykład oparta na umowie o dzieło nie zawieranej z pracodawcą (choć zachodzące w ostatnim okresie zmiany legislacyjne idą w kierunku obowiązku zgłaszania do ZUS również takich umów) czy zlecenie wykonywane przez ucznia szkoły ponadpodstawowej lub studenta, który nie ukończył 26 lat. Badaniu umyka również zatrudnienie nielegalne. Informacje na temat tego rodzaju zjawisk muszą więc być pozyskiwane z innych źródeł. Przykładowo, alternatywnym sposobem uzyskania danych pozwalających oszacować wartości wskaźników rezultatu długoterminowego są badania kwestionariuszowe, które pozwalają w pewnym, choć niedoskonałym stopniu uwzględnić pracę nierejestrowaną.

Po drugie trzeba pamiętać, że w Polsce obok pionu ubezpieczeń społecznych i zdrowotnych obsługiwanych przez ZUS istnieje osobny pion obsługiwany przez Kasę Rolniczego Ubezpieczenia Społecznego (KRUS). W danych ZUS nie ma informacji o ubezpieczonych rolnikach, członkach ich rodzin oraz rolnikach emerytowanych lub pobierających rentę, których liczba wynosi ponad 1,2 mln w przypadku ubezpieczeń społecznych i 2,3 mln w przypadku ubezpieczeń zdrowotnych<sup>14</sup>. Sposobem na uzupełnienie tego braku danych byłoby wykorzystanie, obok KSI ZUS, również systemu

<sup>14</sup> Wg Kwartalnej informacji statystycznej ZUS za II kwartał 2020 r.

informatycznego KRUS, czyli nSIU w przypadku ubezpieczeń społecznych i zdrowotnych. Ponieważ jednak znacznie komplikowałoby to proces pozyskania i analizy danych, w przypadku ewaluacji EFS takie przedsięwzięcie nie wydaje się celowe, z ewentualnym wyjątkiem interwencji skierowanych do środowisk rolniczych (np. do osób odchodzących z rolnictwa).

Trzecią możliwą przyczyną mylnego uznania braku informacji o ubezpieczeniu za dowód nieposiadania pracy są braki lub nieprawidłowości w zbiorze danych albo błędy w procesie ich przetwarzania. Ryzyko tego rodzaju sytuacji, choć trudne do oceny, wydaje się niewielkie. Z pewnością jednak przetwarzając dane ZUS należy zwracać na nie uwagę, przewidując, zapobiegając i wykrywając ewentualne zniekształcenia w zbiorze danych (np. usunięcie początkowych zer z numeru PESEL, które może prowadzić do nieprzyłączenia części danych).

Warto ponadto zauważyć, że osoby samozatrudnione nie odprowadzają składki na ubezpieczenia społeczne z tytułu prowadzonej działalności gospodarczej, jeżeli podlegają temu ubezpieczeniu z innego tytułu, czyli w typowej sytuacji z tytułu umowy o pracę. Płacą wtedy jedynie podwójną składkę na ubezpieczenie zdrowotne. Dlatego dane o ubezpieczeniach społecznych nie pozwalają na identyfikację wszystkich osób prowadzących działalność gospodarczą. Z tego powodu przy ustalaniu wartości wskaźników rezultatu długoterminowego EFS kierowano się informacją o składkach na ubezpieczenie zdrowotne.

Podsumowując, wykorzystując dane ZUS do ustalenia posiadania przez daną osobę pracy można oczekiwać systematycznego zaniżenia odsetka pracujących. Orientacyjnej informacji o skali tego niedoszacowania dostarcza przedstawione w poniższej tabeli porównanie osiągniętych wartości wskaźników rezultatu długoterminowego I Osi POWER wyliczonych na podstawie danych ZUS z wartościami tych samych wskaźników oszacowanymi na podstawie badań kwestionariuszowych na reprezentatywnej próbie. Zgodnie z przewidywaniami, odsetek osób pracujących ustalony na podstawie danych ZUS był niższy niż ten sam odsetek ustalony na podstawie badań kwestionariuszowych. Różnica ta nie przekraczała jednak 3,5 punktu procentowego (przy obu odsetkach przekraczających 70%). Co interesujące, odwrotna zależność – czyli większa wartość wskaźnika przy oparciu się na danych ZUS – miała miejsce w przypadku odsetka osób prowadzących działalność gospodarczą, co sugeruje, że wykorzystanie danych ZUS pozwala bardziej skutecznie zidentyfikować takie osoby niż badania kwestionariuszowe. Oczywiście, z powodów opisanych na początku niniejszego rozdziału, oszacowań uzyskanych w badaniu kwestionariuszowym nie sposób traktować jako dokładnych i w tym sensie stanowiących probierz wyliczeń opartych na danych ZUS. Wręcz przeciwnie, szacunki uzyskane na podstawie badania na próbie mogą być obarczone błędem statystycznym i wieloma specyficznymi rodzajami błędu systematycznego. Niewielka różnica między wynikami obu badań wskazuje jednak na to, że skala niedoszacowania odsetka pracujących w przypadku wykorzystania danych ZUS jest niewielka, rzędu co najwyżej kilku punktów procentowych.

*Tabela 2. Porównanie wartości wskaźników rezultatu długoterminowego w dwóch badaniach*

Kod	Nazwa wskaźnika	Badanie kwestionariuszowe	Dane ZUS	Różnica
KE32 (YEI)	Liczba osób w niekorzystnej sytuacji społecznej, pracujących, łącznie z prowadzącymi działalność na własny rachunek, sześć miesięcy po opuszczeniu programu	73,0%	70,9%	+2,1 pp.
YE11	Liczba osób pracujących 6 miesięcy po zakończeniu udziału w programie	76,4%	73,0%	+3,4 pp.

<b>YE12</b>	Liczba osób pracujących na własny rachunek 6 miesięcy po zakończeniu udziału w programie	42,5%	46,0%	-3,5 pp.
<b>PL02 (bez YEI)</b>	Liczba osób poniżej 30 lat, pracujących sześć miesięcy po opuszczeniu programu, w tym prowadzących działalność na własny rachunek	80,2%	77,1%	+3,1 pp.

*Źródło: (Penszko, 2016, strony 189-190).*

Przypuszczalnie zastrzeżenia co do wyczerpującego charakteru danych ZUS w mniejszym stopniu dotyczą takich zasobów, jak rejestr płatników składek, rejestr lekarzy upoważnionych do wystawiania zaświadczeń czy rejestr zaświadczeń lekarskich. Uwzględnienie w tym rejestrze może być bowiem właściwie utożsamione z przynależnością do populacji. Brak doświadczeń w wykorzystaniu wymienionych zasobów na potrzeby ewaluacji nakazuje jednak ostrożność w formułowaniu tego rodzaju twierdzeń.

### 4.3. Kwestie związane z zakresem i kompletnością danych

#### Zakres tematyczny danych

Zawartość zasobów danych ZUS jest oczywiście wyznaczona i w tym sensie ograniczona przez potrzeby informacyjne tej instytucji oraz zakres pełnionych przez nią zadań. Oznacza to jednak, że zasoby te mogą być źródłem cennych informacji na temat kilku obszarów ewaluacji EFS, a mianowicie rynku pracy, przedsiębiorczości, integracji społecznej i ochrony zdrowia. Ponieważ pozwalają ustalić status danej osoby na rynku pracy, mogą być również z powodzeniem wykorzystywane do badania efektów zatrudnieniowych interwencji EFS w obszarze edukacji, łącznie ze szkolnictwem wyższym. W pozostałych dziedzinach ewaluacji EFS dane ZUS nie wydają się mieć istotnego zastosowania. Potencjalne możliwości ich wykorzystania są szerzej omówione w rozdziale 5.

#### Kompletność danych

W dotychczasowych badaniach ewaluacyjnych EFS przeprowadzanych na indywidualnych danych ZUS stwierdzono, że w przypadku kilku lub kilkunastu procent osób (zależnie od fali badania), dla których podany jest kod tytułu ubezpieczenia, brakuje informacji o wysokości podstawy wymiaru składek.

W monitoringu losów absolwentów szkół wyższych zaobserwowano braki danych charakteryzujących osobę ubezpieczoną (np. roku urodzenia), jednak skala tego zjawiska była znikoma. Nie rodzi ono praktycznych problemów w wykorzystaniu danych.

Przypomnijmy, że brak danej jednostki w zbiorze danych ubezpieczonych jest interpretowany jako niepodleganie przez nią ubezpieczeniu, wynikające z nieposiadania pracy. Dlatego ewentualny przypadek niekompletności danych polegający na całkowitym braku informacji o osobie ubezpieczonej jest właściwie „niewidoczny” dla wykonawcy ewaluacji. Ryzyko tego rodzaju sytuacji wydaje się jednak niewielkie. Potwierdza to przedstawione w podrozdziale 4.2 porównanie osiągniętych wartości wskaźników rezultatu długoterminowego I Osi POWER wyliczonych na podstawie danych ZUS oraz tych samych wskaźników, oszacowanych w oparciu o badania kwestionariuszowe. Niewielkie różnice w wynikach obu analiz można wyjaśnić brakiem ubezpieczenia zdrowotnego w przypadku niektórych osób pracujących i ewentualnym błędem statystycznym. Gdybyśmy mieli do czynienia z istotną skalą braków danych w bazach ZUS, odsetek osób pracujących zidentyfikowanych na ich podstawie byłby wyraźnie niższy (w porównaniu do wyników badania kwestionariuszowego).

Ponieważ w ewaluacji EFS wykorzystywano dotąd jedynie mały wycinek z obszernych zasobów danych ZUS, brakuje doświadczeń z badań ewaluacyjnych mówiących o skali braków danych w innych, niewykorzystywanych dotąd strukturach zasobów danych ZUS.

### **Precyzja danych**

Dane liczbowe zbierane w systemach informatycznych ZUS są z reguły odnotowywane z dużą, często wręcz doskonałą precyzją. Jest to cecha typowa dla rejestrów administracyjnych, stanowiąca jedną z ich przewag – jako źródeł danych w badaniach społecznych – nad badaniami kwestionariuszowymi. Przykładem może być informacja o wysokości podstawy wymiaru składek, odnotowywana z dokładnością do jednego grosza. Warto zauważyć, że w oparciu o nią można określić wysokość wynagrodzenia pracownika. W kontekście ewaluacji EFS jest to cenna informacja, która może posłużyć np. do oceny jakości miejsca pracy lub szacowania społecznej stopy zwrotu z inwestycji. Należy jednak zastrzec, że nie pozwala ona ustalić dochodów osób prowadzących działalność gospodarczą, które mają swobodę w określaniu wysokości podstawy wymiaru składek, zwykle przyjmując jej minimalną wartość (odprowadzają najniższe, dopuszczalne prawem, składki do ZUS). Ponadto w przypadku umów zleceń relacja między odprowadzaniem składek a wysokością wynagrodzenia jest na tyle złożona, że praktycznie uniemożliwia obliczenie tej drugiej na podstawie danych ZUS.

## **4.4. Kwestie związane z przetwarzaniem i analizą danych indywidualnych**

### **Zakorzenie w prawno-instytucjonalnym systemie ubezpieczeń i świadczeń**

Decydujący wpływ na zawartość i formę danych ZUS mają prawne uregulowania dotyczące systemu ubezpieczeń społecznych i zdrowotnych oraz wypłacanych przez ZUS świadczeń. Akty prawne dokładnie wyznaczają zakres danych gromadzonych na kontach ubezpieczonych i kontach płatników, a także w niektórych prowadzonych przez ZUS rejestrach. Również właściwa interpretacja danych wymaga znajomości przepisów w odpowiedniej sferze działalności publicznej, a czasami również panującej w niej organizacji instytucjonalnej. Przykładowo, poszczególne kody tytułów ubezpieczenia są ściśle powiązane z konkretnymi przepisami prawnymi, które wyznaczają ich definicje.

Wnioskowanie o wysokości wynagrodzenia w oparciu o wysokość podstawy wymiaru składek wymaga znajomości regulacji określających tę podstawę. Zrozumienie, jakie podmioty znajdują się w rejestrze płatników, bazuje na wiedzy, kto w świetle prawa ma status płatnika. Tego rodzaju przykłady można mnożyć. W razie wątpliwości co do interpretacji danych zawartych w systemie informatycznym ZUS pomocny może być kontakt z personelem tej instytucji, który nim administruje lub wykorzystuje go w pracy.

Regulacje prawne dotyczące ubezpieczeń i świadczeń społecznych ulegają z czasem zmianom, co przekłada się na zakres zbieranych przez ZUS danych, sposób ich kodowania w danym okresie, ich interpretację itd. Ma to szczególne znaczenie przy wykorzystywaniu danych dotyczących różnych, oddalonych w czasie momentów.

Osadzenie w skomplikowanym i zmiennym systemie prawno-instytucjonalnym nie jest czymś nietypowym w obszarach życia społecznego, w których wykorzystywane są środki EFS. Biorąc za przykład dziedzinę edukacji, prowadzenie w niej ewaluacji, a w szczególności prawidłowa interpretacja danych zgromadzonych w Systemie Informacji Oświatowej, wymaga znajomości prawno-instytucjonalnego systemu oświaty - systemu o wysokim stopniu złożoności, zwielokrotnionym przez niedawną reformę strukturalną. Specyfiką danych ZUS jest więc nie tyle obecność uwarunkowań prawno-instytucjonalnych, co konkretny system, który stanowi ich kontekst,



a mianowicie system ubezpieczeń społecznych. Charakterystyczne dla niego jest to, że przekracza granice takich dziedzin życia społecznego, jak rynek pracy, integracja społeczna i ochrona zdrowia. Wykorzystanie danych ZUS wymaga więc orientacji w prawno-instytucjonalnej organizacji każdej z tych sfer.

### **Poprawność danych**

W trakcie analiz prowadzonych w ramach monitoringu losów absolwentów szkół wyższych stwierdzano występowanie pewnych niespójności w danych ZUS, jednak ich skala była niewielka. Nie zaobserwowano wewnętrznych niespójności logicznych lub wad technicznych w trakcie przetwarzania danych ZUS na potrzeby wyliczenia wartości wskaźników rezultatu długoterminowego EFS i YEI w badaniach ewaluacyjnych realizowanych przez Evalo sp. z o.o. Również to stwierdzenie należy opatrzyć zastrzeżeniem, że analizowany był bardzo wąski zakres danych spośród obszernych zasobów ZUS. Naturalnie wykonawcy badań nie dysponowali żadnymi środkami pozwalającymi ocenić rzetelność tych danych przez porównanie ze stanem faktycznym.

W artykule opublikowanym w 2015 roku przez pracowników Instytut Socjologii UW można znaleźć informację o tym, że w formularzach ZUS zbierane są dane o stanowisku pracy, lecz nie podlegają one kontroli poprawności i nie są później wykorzystywane przez tę instytucję. W rezultacie ich jakość jest zbyt niska, by mogły zostać poddane analizom statystycznym (Jasiński, Bożykowski, Zając, Styczeń i Izdebski, 2015, str. 56). W aktualnie stosowanych formularzach zgłoszenia do ubezpieczenia nie zbiera się informacji o stanowisku.

### **Badanie statusu na rynku pracy**

Bazy danych ZUS są dobrym źródłem informacji o sytuacji zawodowej konkretnych osób. Niemniej ze względu na swoją specyfikę ich wykorzystanie prowadzi do systematycznego zaniżenia odsetka osób pracujących. Przyczyny i skala tego zjawiska zostały omówione w rozdziale 4.2.

### **Możliwości analiz diachronicznych**

W Krajowym Systemie Informatycznym ZUS odnotowywana jest data, od której obowiązuje określony tytuł ubezpieczenia, a także data końca jego obowiązywania (jeżeli się zakończyło). Technicznie możliwe jest stwierdzenie, czy dana osoba była objęta ubezpieczeniem, w dowolnie wybranym przeszłym momencie objętym ewidencją. Można o tym wnioskować, jeżeli moment ten przypada na okres między początkową a końcową datą co najmniej jednego z tytułów ubezpieczenia lub moment ten następuje po początkowej dacie ubezpieczenia, a wartość daty końcowej jest nieokreślona (co sygnalizuje, że ubezpieczenie trwa nadal).

Taka możliwość określenia sytuacji jednostki w dowolnym momencie jest typową cechą rejestrów administracyjnych (Jasiński, Bożykowski, Zając, Styczeń i Izdebski, 2015, str. 49; European Commission, 2020). Otwiera wiele możliwości w ewaluacji EFS. Po pierwsze, stwarza swobodę w wyborze momentu weryfikacji sytuacji określonej osoby. Ma to duże znaczenie przy określaniu wartości wskaźników produktu i rezultatu, których właściwy czas pomiaru jest ściśle określony w wytycznych (np. sześć miesięcy po zakończeniu udziału w projekcie EFS). Pozwala również na sprawdzenie statusu na rynku pracy w chwili przystąpienia do projektu, co może być użyteczne w ewaluacji, a jeszcze bardziej w realizacji i kontroli interwencji. Możliwe jest też sięgnięcie jeszcze dalej w przeszłość, do okresu przed udziałem w projekcie, co może się okazać przydatne przy charakterystyce uczestników, badaniu ich trajektorii życiowych (sytuacji zawodowej, ekonomicznej) lub szacowaniu efektu netto interwencji z wykorzystaniem informacji z okresu ją poprzedzającego.



Informacja o początkowej i końcowej dacie obowiązywania tytułu ubezpieczenia stwarza pole do badania czasu występowania efektów wsparcia, na przykład w ramach oceny według ewaluacyjnego kryterium trwałości lub analizy występowania efektu lock-in (tj. czasowego obniżenia prawdopodobieństwa podjęcia pracy w czasie udziału w projekcie i krótko po jego zakończeniu). Tego rodzaju dane umożliwiają również zastosowanie technik analizy przeżycia, na przykład modelu regresji Coxa w celu ustalenia, jakie czynniki sprzyjają wcześniejszemu lub późniejszemu znalezieniu zatrudnienia po uczestnictwie w interwencji.

Trzeba jednak podkreślić, że aby te szerokie możliwości techniczne zostały wykorzystane, muszą im towarzyszyć również możliwości prawne. Dostęp do danych ZUS jest bowiem ograniczony. Przykładowo, obowiązujące przepisy prawne pozwalają na użycie danych ZUS do pomiaru wartości wskaźników produktu i rezultatu EFS. Takie określenie celu przetwarzania oznacza jednak, że na potrzeby ewaluacji możliwe jest pozyskanie jedynie danych dotyczących momentu pomiaru wskaźników, czyli w typowym przypadku sześć miesięcy po zakończeniu udziału w projekcie, a nie na przykład momentu przypadającego znacznie później (w celu oceny trwałości efektów) lub na pewien czas przed rozpoczęciem udziału. Potrzebę przeprowadzenia szerszych analiz należy zatem przewidzieć z dużym wyprzedzeniem, konstruując rozwiązania prawne, które na nie pozwolą.

## 5. Perspektywa wykorzystania danych ZUS w przyszłości

W dotychczasowej ewaluacji EFS w Polsce dane indywidualne ZUS były wykorzystywane jedynie do wyliczenia wartości wskaźników rezultatu długoterminowego oraz w dodatkowych analizach towarzyszących, dotyczących na przykład czynników wpływających na skuteczność wsparcia lub oceny efektywności kosztowej na podstawie porównania osiągniętych efektów z nakładami finansowymi. Zakres pozyskiwanych informacji był wąski – dotyczył tytułu ubezpieczenia społecznego i zdrowotnego w momencie pomiaru wartości wskaźnika oraz wysokości podstawy wymiaru składek. W związku z tym z tego źródła danych korzystano jedynie w obszarze rynku pracy, do ustalenia zatrudnieniowych efektów interwencji, w tym trwałości założonych działalności gospodarczych. Aktualne przepisy prawne dają podstawę do pozyskania informacji nie tylko o sytuacji zawodowej uczestników, ale również o grupie kontrolnej w celu oszacowania efektu netto. Dotychczas nie przeprowadzono jednak tego rodzaju analiz.

Oprócz dotychczasowej praktyki, dane indywidualne ZUS potencjalnie mogłyby zostać wykorzystane również do badania efektów w innych dziedzinach, a mianowicie integracji społecznej i ochrony zdrowia. Nie są co prawda tak obfite jak zasoby danych innych instytucji w tych obszarach – odpowiednio CSIZS z CAS i bazy NFZ – jednak zawierają szereg przydatnych informacji. Należą do nich dane dotyczące tytułu ubezpieczenia, wysokości podstawy wymiaru składek, zasiłków rodzinnych, niezdolności do pracy, niepełnosprawności, absencji, zaświadczeń lekarskich, orzeczeń lekarskich o niezdolności do pracy oraz rehabilitacji leczniczej. Tkwiący w nich potencjał nie był dotąd wykorzystywany.

Dane ZUS nie są właściwym źródłem informacji o efektach edukacyjnych. Jednak w przypadkach, gdy celem interwencji w obszarze oświaty lub szkolnictwa wyższego jest zwiększenie szans absolwentów na rynku pracy, mogą zostać użyte do ustalenia efektów zatrudnieniowych. W tym przypadku alternatywą wobec bezpośredniego pozyskania danych z ZUS jest wykorzystanie systemów monitoringu losów absolwentów. Zaletą takiego rozwiązania jest wyższy stopień przetworzenia danych, w szczególności połączenie danych z różnych źródeł oraz możliwość sięgnięcia po obliczone już wartości szerokiego zestawu wskaźników mierzących efektywność kształcenia.

Można sobie ponadto wyobrazić wykorzystanie informacji gromadzonych na kontach płatników składek i kontach ubezpieczonych do badania efektów interwencji w dziedzinie adaptacyjności przedsiębiorstw lub innych interwencji, których grupą docelową są osoby pracujące. Dane te nie wskazują co prawda bezpośrednio na lepsze przystosowania przedsiębiorstw do zmian zachodzących w gospodarce ani na podwyższenie kompetencji i kwalifikacji pracowników. Mogą jednak posłużyć do analizy zmian poziomu zatrudnienia i wysokości wynagrodzeń. Ich użycie w tym celu wydaje się mniej prawdopodobne niż we wcześniej wymienionych dziedzinach interwencji EFS ze względu na obecność alternatywnych, bogatszych źródeł danych, takich jak sprawozdania SP / SP-3 składane do GUS lub bazy Krajowej Administracji Skarbowej. Zasoby ZUS mogłyby jednak zostać wykorzystane, gdyby dostęp do tych alternatywnych źródeł był trudniejszy lub niemożliwy (a prawdopodobnie nie było jeszcze przypadków wykorzystania w ewaluacji funduszy europejskich w Polsce danych pochodzących z Krajowej Administracji Skarbowej) lub ich zakres byłby niewłaściwy z punktu widzenia celów ewaluacji (np. bardzo niewiele przedsiębiorstw zatrudniających mniej niż 10 pracowników składa sprawozdanie SP lub SP-3). Ponadto, na podstawie danych ZUS przypuszczalnie możliwe byłoby ustalenie poziomu rotacji kadr w przedsiębiorstwie (Pokorski, 2020, str. 96), co pozwoliłoby wnioskować na przykład o tym, jaka część pracowników przeszkolonych przy udziale środków EFS pozostaje w dotychczasowym przedsiębiorstwie, dzięki czemu wsparcie może przekładać się na jego adaptacyjność lub konkurencyjność.

Przytoczone przykłady pokazują, że choć ZUS zbiera niewiele danych na poziomie płatnika składek, to ich połączenie z posiadanymi przez Zakład danymi na poziomie osób ubezpieczonych i agregacji tych ostatnich na poziom płatnika (np. obliczenie odsetka pracowników w określonej kategorii wiekowej, średniego poziomu wynagrodzeń w przedsiębiorstwie czy odsetka osób ubezpieczonych wyrejestrowanych z ubezpieczeń opłacanych przez płatnika w danym okresie czasu) otwierałoby znacznie szersze możliwości prowadzenia analiz. Dalsza agregacja danych, na przykład do poziomu programów operacyjnych, działań, regionów, operatorów pomocy, przedsiębiorstw objętych i nieobjętych wsparciem itd. mogłaby być pomocna w poszukiwaniu odpowiedzi na szereg pytań badawczych. Dotychczas w ewaluacji EFS nie prowadzono analiz danych ZUS na poziomie płatnika, a mogłyby one znaleźć zastosowanie w obszarze przedsiębiorczości, ewentualnie w innych obszarach w sytuacji, gdy przedmiotem zainteresowania jest podmiot będący płatnikiem składek (niekoniecznie przedsiębiorstwo).

Innym potencjalnym sposobem rozszerzenia wykorzystania zasobów ZUS w stosunku do obecnej praktyki byłoby sięgnięcie po informacje pozwalające zbadać nie tylko to, jaka była sytuacja osób w momencie pomiaru wartości wskaźnika, ale również to, jak ta sytuacja się zmieniała w dłuższym okresie. Dzięki temu możliwe byłoby przeprowadzenie analiz dotyczących czasu występowania efektów, w tym np. ich trwałości czy występowania efektu lock-in. Na potrzeby oszacowania efektu netto bardzo przydatne byłyby z kolei dane dotyczące okresu przed otrzymaniem wsparcia. Mogłyby one zresztą posłużyć w ewaluacji EFS nie tylko do badania skuteczności i efektywności, ale również do charakterystyki odbiorców wsparcia, np. w celu stwierdzenia, czy interwencja jest dobrze ukierunkowana i dociera do właściwych grup docelowych. Ze względu na to, że zbiory danych ZUS powstają w wyniku prowadzonej na bieżąco ewidencji, ich analiza nie musiałaby być ograniczona do kilku wybranych momentów (jak to ma miejsce w przypadku badań kwestionariuszowych), lecz uwzględniać zmiany zachodzące w całym przedziale czasu, określonym odpowiednio do potrzeb i względem momentu dokonania interwencji publicznej.

Wypada wspomnieć o jeszcze jednym potencjalnym zastosowaniu rejestrów prowadzonych przez ZUS, a mianowicie wykorzystaniu ich jako operatów do doboru próby. Przykładowo, próba lekarzy, którzy zostaliby objęci badaniem w ramach ewaluacji wsparcia EFS w dziedzinie ochrony zdrowia, mogłaby pochodzić z Rejestru lekarzy, lekarzy dentyistów, felczerów i starszych felczerów, którzy zgłosili wniosek w sprawie upoważnienia ich do wystawiania zaświadczeń lekarskich. To, czy określony rejestr ZUS będzie operatem właściwym i najlepszym z dostępnych, zależy od celów, zakresu tematycznego i pytań badawczych konkretnego badania ewaluacyjnego.

Na koniec warto podkreślić, że wszystkie wymienione potencjalne zastosowania danych indywidualnych gromadzonych przez ZUS będą możliwe tylko pod warunkiem istnienia odpowiedniej podstawy prawnej. Zapewnienie tej podstawy jest więc wstępnym warunkiem wykorzystania bogatych zasobów informacji omawianej instytucji w ewaluacji Europejskiego Funduszu Społecznego. Nie ma natomiast podobnych przeszkód w korzystaniu z danych zagregowanych, udostępnianych na Portalu statystycznym ZUS oraz przez Główny Urząd Statystyczny. Dodatkową możliwość stwarza Ustawa o Polskim Instytucie Ekonomicznym<sup>15</sup> (PIE), która umocowuje tę instytucję do pozyskiwania „danych statystycznych” z ZUS. Co prawda zapis ten nie dotyczy danych osobowych. Można sobie jednak wyobrazić rozwiązanie, zgodnie z którym PIE, występując w roli wykonawcy ewaluacji przekazuje do ZUS dane identyfikujące jednostki objęte wsparciem EFS (np. listę przedsiębiorstw, których pracownicy brali udział w szkoleniach finansowanych z EFS), prosząc o zagregowane dane na

---

<sup>15</sup> Art. 16 Ustawy z dnia 20 lipca 2018 r. o Polskim Instytucie Ekonomicznym, Dz. U. 2018 poz. 1735.

temat tej grupy i ewentualnie grupy porównawczej (np. odsetek pracowników wyrejestrowanych z ubezpieczeń w wyznaczonym okresie jako wskaźnik rotacji kadr).

## 6. Wnioski

Dotychczas w ewaluacji EFS w Polsce dane indywidualne ZUS były wykorzystywane jedynie do badania zatrudnieniowych efektów wsparcia, w szczególności do pomiaru osiągniętych wartości wskaźników rezultatu długoterminowych i wykonywanych przy tej okazji dodatkowych analiz (dotyczących na przykład efektywności kosztowej lub czynników wpływających na skuteczność wsparcia). Sięgnięcie po to źródło informacji w okresie programowania 2014-2020 należy uznać za ważny sukces systemu ewaluacji w Polsce. Dzięki temu w bardziej wiarygodny i precyzyjny sposób ustalono rezultaty prowadzonej interwencji.

Potencjał drzemący w zasobach ZUS jest jednak o wiele większy. Pozyskanie z ZUS szerszego zakresu danych pozwoliłoby na skorzystanie z nich do oceny efektów na innych polach niż rynek pracy (w szczególności w dziedzinach integracji społecznej i ochrony zdrowia, ale także wsparcia przedsiębiorstw), wiarygodniejszego szacowania efektów netto, badania czasu występowania efektów, pełniejszej charakterystyki uczestników przed otrzymaniem wsparcia czy doboru próby. Byłby to cenny wkład w ewaluację EFS ze względu na zalety danych administracyjnych, takie jak pokrycie całej populacji, ciągłość rejestracji informacji w czasie, aktualność, dokładność, stosunkowo niski poziom błędów pomiaru czy relatywnie niskie koszty pozyskania. Oczywiście dane te posiadają również wady z punktu widzenia ewaluacji EFS, takie jak brak wpływu na zakres i sposób kodowania informacji (European Commission, 2020, p. 4; Jasiński, Bożykowski, Zajac, Styczeń i Izdebski, 2015). Niemniej zalety zdecydowanie przeważają i sprawiają, że wykorzystanie zasobów informacji, którymi dysponuje ZUS, pozwala na podwyższenie ilości i jakości wiedzy przydatnej przy prowadzeniu polityk publicznych.

Dotychczasowe doświadczenia wskazują na to, że trudno oprzeć dostęp do danych osobowych ZUS na przesłance zgody osoby zainteresowanej. Przesłankę, w rozumieniu RODO, może natomiast stanowić obowiązek prawny lub interes publiczny. Obie te potencjalne przesłanki opierają się na przepisach prawnych, określających cel, zakres i sposób przetwarzania (art. 6 RODO<sup>16</sup>). Oznacza to, że drogą do zapewnienia szerszych możliwości korzystania z danych ZUS w ewaluacji EFS jest wprowadzenie odpowiednich zmian w prawie.

Można wskazać na szereg działań, które mogą podjąć instytucje publiczne w celu pełniejszego wykorzystania zasobów ZUS w ewaluacji EFS:

### 1. Utrzymanie dostępu do danych ZUS w nadchodzącym okresie programowania.

Dotychczasowe wykorzystanie danych indywidualnych ZUS w ewaluacji EFS opierało się na przepisach wprowadzonych do Ustawy o systemie ubezpieczeń społecznych, wśród których kluczowe są art. 50 ust. 3a i ust. 3c. Ich treść odnosi się do Rozporządzenia EFS na okres programowania 2014-2020, a więc z końcem jego obowiązywania może stracić moc prawną. Dlatego należy zadbać o to, aby wraz z nową ustawą wdrożeniową artykuły te zostały zaktualizowane, zapewniając co najmniej takie same możliwości korzystania z danych ZUS, jak to miało miejsce dotychczas. Wskazane jest jednak rozszerzenie tych możliwości, którego potencjalne kierunki zostały opisane w niniejszej ekspertyzie.

---

<sup>16</sup> Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych) (Dz. Urz. UE L 119 z 04.05.2016, str. 1 oraz Dz. Urz. UE L 127 z 23.05.2018, str. 2).

## **2. Uwzględnienie ewaluacji polityki spójności przy wprowadzaniu monitoringu losów absolwentów szkół ponadpodstawowych.**

W trakcie opracowywania niniejszej ekspertyzy prowadzone są konsultacje projektu ustawy, która umożliwi integrację i wykorzystanie zasobów danych administracyjnych, zarządzanych przez takie instytucje, jak ZUS, CIE, OPI i OKE, w celu monitorowania losów absolwentów szkół ponadpodstawowych. Wskazane jest wprowadzenie w tym projekcie zmian umożliwiających wykorzystanie zgromadzonych danych w ewaluacji polityki spójności. Teoretycznie uwzględnienie potrzeb ewaluacji EFS może zachodzić na dwóch poziomach.

Pierwszy poziom dotyczyłby danych indywidualnych na etapie integracji danych z różnych źródeł, a przed ich anonimizacją, tj. przed usunięciem identyfikatorów PESEL. Tylko na tym etapie możliwa jest jeszcze identyfikacja uczestników projektów EFS wśród absolwentów objętych analizą. W związku z tym administrator systemu teleinformatycznego obsługującego realizację polityki spójności (obecnego SL 2014) powinien zostać dołączony w projekcie ustawy do instytucji przesyłających do CIE dane w celu ich integracji. CIE dołączałoby do przetwarzanych danych informację o udziale w projekcie EFS. Następnie do planowanego zestawu analiz i wskaźników monitoringu należałoby dołączyć takie, które pozwoliłyby wnioskować o efektach polityki spójności. Planuje się, że analizy danych z monitoringu będą wykonywane przez IBE. Działania analityczne służące ewaluacji polityki spójności mogłyby więc zostać zlecone IBE przez ministra właściwego do spraw oświaty i wychowania, występującego jako instytucja pośrednicząca w systemie realizacji polityki spójności.

Drugi poziom dotyczy obliczonych wartości wskaźników na poziomie szkół. Kwestia dostępu do tych informacji nie jest uregulowana w obecnym projekcie ustawy. Wstępnie planuje się, że byłyby dostępne wyłącznie dla szkół, których dane dotyczą, w Systemie Informacji Oświatowej. Z punktu widzenia ewaluacji EFS dobrze byłoby zagwarantować możliwość użycia ich przez ewaluatorów. Otworzyłyby to pole do badania wpływu objęcia interwencją poszczególnych szkół na losy ich absolwentów.

Porównując oba te poziomy warto zauważyć, że ewentualną przewagą pierwszego jest zawężenie badania efektów do konkretnych osób objętych wsparciem, a nie całych szkół. Ma to znaczenie wtedy, gdy wsparcie jest udzielane poszczególnym uczniom, a nie szkole jako całości – na przykład gdy ewaluacja dotyczy stypendiów, a nie programów rozwoju szkoły. Natomiast przewagą uwzględnienia potrzeb ewaluacji EFS na drugim poziomie jest większa elastyczność co do zakresu analiz. Możliwe jest na przykład ograniczenie ewaluacji do określonych działań prowadzonych w określonym czasie – o ile tylko dane pozwalają na identyfikację szkół objętych tymi działaniami. Tej elastyczności nie ma w pierwszym przypadku, w którym zbiór osób klasyfikowanych jako objęty wsparciem jest determinowany corocznie przez listę numerów PESEL przekazywanych do CIE i nie może być później zmieniany lub dostosowywany do potrzeb konkretnego badania ewaluacyjnego. Wynika to z założenia, że dane gromadzone w ramach monitoringu są anonimizowane niezwłocznie po ich połączeniu. Ze względu na ten brak elastyczności i na to, że znaczną część interwencji EFS należy traktować jako oddziałującą na całą społeczność szkolną, a nie na pojedynczych uczniów, uwzględnienie potrzeb ewaluacji EFS na drugim poziomie okazałoby się prawdopodobnie bardziej użyteczne niż na poziomie pierwszym.

## **3. Uwzględnienie potrzeb ewaluacji polityki spójności przy integracji zasobów danych publicznych.**

W miarę postępu działań służących stworzeniu platformy integrującej dane administracyjne z różnych źródeł, aktualnie prowadzonych w ramach projektu Zintegrowana Platforma Analityczna, warto zadbać o to, aby:

- zakres danych przekazywanych przez ZUS do platformy odpowiadał na potrzeby ewaluacji polityki spójności;
- stworzone zostały kanały dostępu do danych gromadzonych na platformie w celu prowadzenia ewaluacji polityki spójności.

#### **4. Uwzględnienie w systemie ewaluacji polityki spójności UE w Polsce instytucji badawczych, posiadających dostęp do danych ZUS**

Przy planowaniu ewaluacji EFS w okresie programowania 2021-2027 warto wziąć pod uwagę publiczne ośrodki badawcze, które mają dostęp do danych ZUS i innych danych administracyjnych. Jak wspomniano wyżej, należą do nich instytucje, które są lub mają być zaangażowane w prowadzenie monitoringu losów absolwentów – Ośrodek Przetwarzania Informacji i Instytut Badań Edukacyjnych. Ustawowo zagwarantowany dostęp do danych statystycznych ZUS i danych wielu innych organów władzy publicznej ma Polski Instytut Ekonomiczny. Szczególną pozycję zajmuje Główny Urząd Statystyczny, który już obecnie otrzymuje i opracowuje dane ZUS, a także ogromnej liczby innych podmiotów. Ponadto, GUS pełni ważną rolę w systemie realizacji polityki spójności UE w Polsce jako podmiot dostarczający informacji, a jego zadania na kolejny okres programowania zostaną prawdopodobnie określone przy projektowaniu działań stanowiących kontynuację systemu STRATEG i projektów „Statystyka dla polityki spójności”. Włączenie tego rodzaju instytucji do systemu ewaluacji stworzyłoby warunki prawno-instytucjonalne do wykorzystania w niej danych ZUS. Niniejsza rekomendacja nie stanowi propozycji konkretnych decyzji przy projektowaniu tego systemu, lecz jest wskazówką, że przy podejmowaniu tych decyzji warto pamiętać o potencjale wybranych instytucji badawczych, wynikającym z dostępu do danych administracyjnych.

#### **5. Zbieranie danych o jednostkach objętych wsparciem zarówno na poziomie uczestników, jak i na poziomie instytucji.**

Identyfikatorem osoby, który umożliwia przyłączenie danych indywidualnych ZUS, jest numer PESEL. Numer ten jest obecnie rejestrowany w systemie SL 2014, co otwiera szerokie możliwości w ewaluacji. Aby jednak skorzystać z danych o płatnikach składek, np. przedsiębiorstwach będących miejscem pracy uczestnika, potrzebnymi byłby dodatkowo numer NIP lub REGON.

W obszarze edukacji niezwykle przydatną informacją byłby identyfikator szkoły, do której uczęszcza lub w której pracuje uczestnik. Jego brak w SL 2014 jest stałą bolączką ewaluacji. W przypadku pracowników dydaktycznych można co prawda liczyć na informację o nazwie szkoły, która jednak często nie jest wystarczająca do jej identyfikacji (np. „Szkoła podstawowa nr 1”). Z kolei w dziedzinie edukacji przydatne byłoby przyłączenie nie danych o płatnikach, lecz wyników monitoringu losów absolwentów, które są przetworzoną postacią danych ZUS. W tym kontekście najlepszym identyfikatorem dla szkół i placówek systemu oświaty byłby numer RSPO. W razie jego braku rolę zastępczą mógłby pełnić numer REGON, notowany w SIO i RSPO. W systemie szkolnictwa wyższego brak podobnego identyfikatora.

Względy te warto mieć na uwadze przy konstruowaniu systemu informatycznego obsługującego realizację polityki spójności w nowym okresie programowania.

#### **6. Wykorzystanie dostępnych danych zagregowanych**

Zakład Ubezpieczeń Społecznych publikuje wiele danych statystycznych, związanych ze swoją działalnością i dotyczących wielu dziedzin interwencji EFS, w szczególności rynku pracy, integracji społecznej i ochrony zdrowia. Raporty, zestawienia i opracowania, w tym Rocznik statystyczny ubezpieczeń społecznych, są upubliczniane na Portalu statystycznym ZUS. Wyniki analiz danych

pochodzących z zasobów ZUS są również publikowane przez GUS oraz na stronie internetowej Ogólnopolskiego systemu monitorowania Ekonomicznych Losów Absolwentów szkół wyższych (ELA). Opracowania statystyczne ZUS nie są praktycznie wykorzystywane w ewaluacji EFS, choć stanowią dla niej potencjalnie przydatne i łatwo dostępne źródło informacji. Warto uwzględnić wymienione wyżej materiały przy analizie danych zastanych – oczywiście wtedy, gdy mają zastosowanie w przypadku danego badania ewaluacyjnego.

Realizacja tej rekomendacji spoczywa głównie w rękach ewaluatorów. Niemniej przyczynić się do jej wdrożenia mogą również jednostki ewaluacyjne zlecające badania, poprzez uwzględnianie opracowań ZUS na liście materiałów źródłowych, które ma wykorzystać ewaluator. Lista źródeł jest zwykle zamieszczana w opisie przedmiotu zamówienia na przeprowadzenie ewaluacji.



## Bibliografia

- Albanese, A., Capellari, L., & Leonardi, M. (2017). *Effetti delle riforme nel mercato del lavoro dei giovani: evidenza degli apprendistati italiani*. *WorkINPS Papers - numero 6*. Roma: INPS.
- Asseco Poland S.A. (2020, 12 20). *KSI ZUS. Kompleksowy system informatyczny ZUS*. Pobrano z lokalizacji <https://pl.asseco.com/case-study/42/kompleksowy-system-informatyczny-zus-96/>
- Boeri, T., Garibaldi, P., & Moen, E. (2016). *A clash of generations? Increase in Retirement Age and Labor Demand for Youth*. *WorkINPS Papers - numero 1*. Roma: INPS.
- Bożykowski, M. i in. (2014). *Monitorowanie losów absolwentów uczelni wyższych z wykorzystaniem danych administracyjnych Zakładu Ubezpieczeń Społecznych. Raport końcowy*. Warszawa: Instytut Badań Edukacyjnych.
- CEAPP UJ i IDEA Instytut. (2019). *Analiza zapotrzebowania na kompetencje w gospodarce i na rynku pracy wraz z badaniem wartości docelowej wspólnego wskaźnika długoterminowego POWER w obszarze szkolnictwa wyższego*. Warszawa: Uniwersytet Jagielloński - Centrum Ewaluacji i Analiz Polityk Publicznych oraz IDEA Instytut dla Narodowego Centrum Badań i Rozwoju.
- Centralny Instytut Ochrony Pracy – Państwowy Instytut Badawczy w Warszawie i Instytut Medycyny Pracy im. Prof. dr hab. med. J. Nofera w Łodzi. (2016). *Kompleksowy program interwencji profilaktycznej ukierunkowanej na zapobieganie chorobom układu krążenia. Raporty z etapu 2 i 3*. Warszawa: ZUS.
- European Commission. (2020). *How to use administrative data for European Social Funds counterfactual impact evaluations. A step-by-step guide for managing authorities*. Luxembourg: Publications Office of the European Union.
- GUS. (2020, 4 23). *PBSSP na 2020 - wersja interaktywna*. Pobrano 12 15, 2020 z lokalizacji <http://form.stat.gov.pl/formaty/pbssp.php?rok-pbssp=2020>
- ICF GHK Polska. (2013). *Raport z wykonania ekspertyzy na temat źródeł danych wykorzystywanych do realizacji badań kontryfaktycznych w ramach ewaluacji EFS*. Warszawa: ICF GHK Polska Sp. z o.o. dla Ministerstwa Rozwoju Regionalnego.
- Jasiński, M., Bożykowski, M., Zając, T., Styczeń, M. i Izdebski, A. (2015). *Dokładniej, rzetelniej, taniej. Badania oparte na rejestrach publicznych jako szansa dla badań społecznych w Polsce*. *Studia Socjologiczne*(1 (216)), strony 45-72.
- Jonavičienė, D. (2015). *Counterfactual Impact Evaluation of ESF-funded Active Labour Market Measures in Lithuania. Executive Summary*. PPMI dla Socialinės apsaugos ir darbo ministerija. Pobrano z lokalizacji <https://www.researchgate.net/publication/340256085>
- Jonavičienė, D. (2016). *Poster: Counterfactual Impact Evaluation of ESF-funded Active Labour*. PPMI dla Socialinės apsaugos ir darbo ministerija. Pobrano z lokalizacji <https://www.researchgate.net/publication/340255996>
- Kalinowski, H. (2019). *Badanie efektów wsparcia zrealizowanego na rzecz osób młodych w ramach Programu Operacyjnego Wiedza Edukacja Rozwój*. Warszawa: Imapp sp. z o.o. Instytut Badań Strukturalnych i IQS Sp. z o.o. dla Ministerstwa Inwestycji i Rozwoju.
- Pellegrini, G., & Cerqua, A. (2017). *Qual'è l'impatto delle politiche regionali sul mercato del lavoro locale?* *WorkINPS Papers - numero 5*. Roma: INPS.

- Penszko, P. (Red.). (2016). *Metaanaliza wyników badań ewaluacyjnych dotyczących oceny wsparcia z EFS. Raport cząstkowy 2016*. Warszawa: Agrotec Polska sp. z o.o. dla Ministerstwa Rozwoju.
- Penszko, P. (Red.). (2019). *Metaanaliza wyników badań ewaluacyjnych dotyczących oceny wsparcia z EFS. Raport końcowy*. Warszawa: Evalu sp. z o.o. dla Ministerstwa Inwestycji i Rozwoju.
- Pęciłto, M. i Najmiec, A. (2015). *Opracowanie programu kształtowania zachowań probezpiecznych w przedsiębiorstwach budowlanych*. Warszawa: ZUS.
- Piotrowicz, R. i i in. (2011). *Ocena wyników programu pilotażowego rehabilitacji leczniczej układu krążenia w systemie ambulatoryjnym z monitorowaną telemedycynie rehabilitacją kardiologiczną w warunkach domowych*. Warszawa: ZUS.
- Pokorski, J. (2020). Ewaluacja wsparcia kompetencji osób na rynku pracy i promocji zatrudnienia. Pełny obraz efektów z mikro- i makroperspektywy. W B. Ciężka i A. Rybińska (Redaktorzy), *Ewaluacja w działaniach edukacyjnych* (strony 88-106). Warszawa: Fundacja Rozwoju Systemu Edukacji.
- Sarzyńska-Długosz, I. i in. (2018). *Program rehabilitacji w ramach prewencji rentowej ZUS dla osób z uszkodzeniem OUN*. Warszawa: ZUS.
- Sävenstrand, A. (Red.). (2014). *The Social Fund in Figures 2014. Project participants and benefits*. Stockholm: Swedish ESF Council.
- Thörn, L. (2012). Use of administrative registers – the Swedish example. *Learning seminar on YEI. 12.03.2012*. Brussels: Swedish ESF Council.