



**Dane publiczne –
nowy impuls
do rozwoju Polski**

Warszawa, sierpień 2019 r.

Autorzy: Marek Bożykowski, Agnieszka Chłoń-Domińczak, Mikołaj Jasiński, Tomasz Zając

Współpraca: Jacek Grzeszak, Ignacy Świącicki

Redakcja: Małgorzata Wieteska, Jakub Nowak

Projekt graficzny: Anna Olczak

Współpraca graficzna: Grzegorz Piechnik, Liliana Gałązka, Tomasz Gałązka

Polski Instytut Ekonomiczny

Al. Jerozolimskie 87

02-001 Warszawa

© Copyright by Polski Instytut Ekonomiczny

ISBN 978-83-66306-33-2

Spis treści

Najważniejsze liczby	4
Kluczowe wnioski	5
Wstęp	7
Dane administracyjne jako podstawa efektywnych i rzetelnych analiz	10
Otwartość danych publicznych jako kreator innovacyjnego państwa	12
Jak udostępnia się dane administracyjne do badań w innych krajach? Dobre praktyki z zagranicy	16
Wykorzystanie danych administracyjnych dla badań i oceny polityk publicznych w Polsce	20
Potencjał danych administracyjnych w Polsce	22
Opis koncepcji integracji zasobów informacyjnych państwa oraz mechanizmu udostępniania danych	26
Podsumowanie i rekomendacje	31
Bibliografia	33

Najważniejsze liczby

>700 liczba rejestrów, w których przechowywane są dane administracyjne w Polsce

2 liczba rejestrów wykorzystywanych w projekcie monitorowania ekonomicznych losów absolwentów ELA

12 liczba rejestrów, z których dane są dostępne w australijskim systemie MADIP

42 liczba projektów badawczych przeprowadzonych przy wykorzystaniu australijskiego systemu MADIP w latach 2017-2019

50 tys. USD tylko tyle wynosi roczny koszt systematycznie prowadzonych badań aktywności zawodowej i edukacyjnej wszystkich obywateli Szwecji, z czego połowa to koszty druku

ok. 70 proc. brytyjskich firm opierających swój model biznesowy na przetwarzaniu danych wykorzystywało dane przygotowane przez administrację publiczną, według danych przedstawionych przez ODI w 2015 r.

Dane publiczne jako całość niosą bardzo dużą wartość dla gospodarki:

26,7 mld PLN szacowana wartość gospodarki Polski opartej na danych w 2020 r. (IDC, 2019)

51 mld PLN szacowana wartość gospodarki Polski opartej na danych w 2025 r. w scenariuszu szybkiego wzrostu (IDC, 2019)

205 mln PLN tyle szacunkowo oszczędności może przynieść Polsce otwarcie danych publicznych, według szacunków z 2015 r. dla lat 2016-2020 (European Commission, 2015)

Kluczowe wnioski

- 1.** Dane w coraz większym stopniu stają się zasobem dla gospodarki przyszłości. Współcześni giganci technologiczni zbudowali swoje potęgi korzystając z danych, które zgromadzili. Właścicielem znaczących zasobów informacyjnych jest również państwo i jego instytucje. Dane gromadzone w różnego typu rejestrach administracyjnych w celu dostarczania usług publicznych mogą być używane również do analityki, niezbędnej do wspierania kształtowania polityki gospodarczej i społecznej opartej na faktach. Dane odpowiednio przetworzone mogą wspierać rozwój gospodarczy, społeczny i naukowy kraju. Należy pamiętać, że wykorzystanie zasobów rejestrów publicznych wymaga zapewnienia ich bezpieczeństwa oraz ochrony prywatności osób objętych rejestrami.
- 2.** Doświadczenia międzynarodowe pokazują, że umiejętne korzystanie z zasobów informacyjnych gromadzonych przez państwo przynosi znaczące i wymierne korzyści w postaci zmniejszenia kosztów opieki zdrowotnej, usprawnienia funkcjonowania rynku pracy czy poprawy jakości systemu edukacji. Dane administracyjne są również bezcennym źródłem dla badań naukowych, które wspierają realizację polityk publicznych. Odpowiednio udostępnione dane publiczne stanowią impuls do rozwoju nowych firm technologicznych.
- 3.** W wielu krajach powstają specjalne systemy integrujące dane publiczne i tworzące ramy administracyjne do ich udostępniania i wykorzystywania. MONA i LISA w Szwecji, MADIP w Australii czy ADRP w Wielkiej Brytanii to przykłady inicjatyw, które pozwalają na diagnozowanie priorytetów polityki gospodarczej, społecznej i środowiskowej oraz poprawienie usług publicznych. Dotyczy to między innymi działań wspierających przedsiębiorstwa, w tym w sektorze MSP, opieki zdrowotnej czy wspierania rozwoju małych dzieci.
- 4.** Obecnie w Polsce dane administracyjne wykorzystywane są m.in. w analizach prowadzonych przez Ministerstwo Finansów, Ministerstwo Infrastruktury i Rozwoju – do ewaluacji polityk i instrumentów wsparcia oraz przez Ministerstwo Nauki i Szkolnictwa Wyższego – do monitorowania ekonomicznych losów absolwentów szkół wyższych (System ELA). Stworzony system monitorowania absolwentów gwarantuje zachowanie prywatności absolwentów, a jednocześnie dostarcza informacji różnorodnym instytucjom.

5. Integracja zasobów informacyjnych państwa, przez odpowiednie regulacje ustawowe, może znacząco wzmocnić ich potencjał dla administracji, gospodarki i nauki. Dotychczasowe doświadczenia oraz realizacja projektu „Zintegrowana Platforma Analityczna” są kolejnymi etapami procesu dochodzenia do integracji zasobów informacyjnych państwa. Wzmocnienie administracji publicznej przez partnerstwo ze środowiskiem badawczym może w jeszcze większym stopniu przyczynić się do wykorzystania danych jako kapitału rozwojowego w przyszłości.

Korzyści z takiej integracji i wykorzystania danych mogą dotyczyć m.in. polityki rynku pracy (analizy przepływów, losów absolwentów i skuteczności poszczególnych działań), ochrony zdrowia (skuteczniejsza profilaktyka i niższe koszty) czy wspierania przedsiębiorczości (nowe usługi wykorzystujące dane z rejestrów, lepsze planowanie polityki kadrowej czy lokalizacji nowych przedsięwzięć).



Wstęp

Oznaczeniu danych we współczesnym świecie świadczą pojawiające się wielokrotnie w prasie (np. "The Economist", 2017a; 2017b; Toonders, 2014; Vanian, 2016) porównania znaczenia tych zasobów do zasobów ropy naftowej. Dane mają być podstawowym surowcem dla gospodarki przyszłości, czyli pełnić taką rolę, jaką ropa pełniła w ostatnich kilku dekadach. O znaczeniu danych ma świadczyć m.in. rozwój gigantów technologicznych, takich jak Alphabet (firma, w której skład wchodzi Google) czy Facebook, które specjalizują się w zbieraniu i przetwarzaniu danych oraz wycena firm takich jak Uber, których wartość ma wynikać w znacznym stopniu ze zgromadzonych przez nią danych o pasażerach i kierowcach oraz odbywanych przez nich podróżach (mimo że wciąż brakuje standardów wyceny danych posiadanych przez firmy). Porównanie danych do ropy, jak zauważają krytycy tego pomysłu (Thorp, 2012; Rajan, 2017; Marr, 2018) choć użyteczne z punktu widzenia marketingu i ukazywania wartości danych, jest niedokładne. Znaczną różnicą między ropą i danymi jest ich trwałość – dane się nie zużywają, użyte raz mogą być wykorzystane ponownie.

Właścicielem ogromnej ilości danych, które mogą być użyte ponownie, jest państwo i jego instytucje. Państwo realizując swoje zadania, określone w różnych aktach prawnych, gromadzi niezbędne informacje w rejestrach publicznych. Dane administracyjne są definiowane jako informacje pochodzące z działania systemów administracyjnych, prowadzonych zazwyczaj przez agencje sektora publicznego (Elias, 2015, s. 47-48). Woollard (2014) określa dane administracyjne jako informacje zebrane w celu rejestracji, prowadzenia transakcji i prowadzenia rejestrów, a dane administracyjne są często

związane z dostarczaniem usług. Dane te mogą pochodzić z rejestrów administracyjnych, wykorzystywanych w edukacji, opiece zdrowotnej, podatkach, mieszkalnictwie czy licencjonowaniu pojazdów. Dane administracyjne obejmują również informacje z rejestrów, takich jak powiadomienia o urodzeniach, zgonach i małżeństwach, spisy wyborców i narodowe spisy powszechne. Zebrane dane są zatem wykorzystywane przede wszystkim przy podejmowaniu indywidualnych decyzji administracyjnych oraz realizacji zadań ustawowych; są źródłem dużych i złożonych informacji ilościowych i znajdują się w nich informacje generowane głównie w celu innym niż badania (Connelly i in., 2016). W niektórych krajach, takich jak Norwegia, Finlandia i Szwecja, zasoby danych administracyjnych są dostępne dla naukowców od wielu lat (ONZ, 2007). W innych częściach świata, zwłaszcza w Wielkiej Brytanii i Stanach Zjednoczonych, ostatnio zwiększona dostępność danych administracyjnych do badań stanowi skokową zmianę w infrastrukturze danych nauk społecznych. Odpowiednio przetworzone informacje mogą posłużyć do udoskonalenia polityk publicznych oraz wspierać rozwój naukowy i gospodarczy kraju. Część zasobów danych zbieranych przez sektor publiczny jest udostępniana na mocy istniejących przepisów, jak chociażby Dyrektywy w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego (Dyrektywa, 2019) jednak naszym zdaniem należy skupić się na poszerzeniu dostępu do danych pozostających poza jej zakresem lub też tych, dla których dostęp jest ograniczony ze względu chociażby na przepisy Ogólnego rozporządzenia o ochronie danych osobowych.

Celem niniejszego opracowania jest przedstawienie koncepcji integracji zasobów

informacyjnych państwa jako sposobu wsparcia rozwoju innowacyjności Polski, zarówno jej gospodarki, jak i administracji oraz nauki. Doświadczenia badawcze oraz związane ze współpracą z administracją publiczną przekonały autorów tej publikacji o palącej potrzebie stworzenia w Polsce stabilnego systemu pozwalającego na łączenie i analizę informacji zawartych w różnych rejestrach publicznych oraz udostępniającego wyniki tych analiz. Utwierdziły autorów w przekonaniu o konieczności kompleksowego ustawowego uregulowania procesu przetwarzania danych administracyjnych jako wsparcia analityki publicznej.

Kluczową kwestią w wykorzystywaniu do analiz informacji pochodzących z rejestrów publicznych jest bezpieczeństwo danych i ochrona prywatności osób objętych rejestrem.

Bezpieczeństwo cyfrowe jest jednym z najważniejszych aspektów funkcjonowania współczesnego państwa.

Funkcjonujący od 2016 r. w Polsce *Ogólnopolski system monitorowania ekonomicznych losów absolwentów szkół wyższych* (ELA) jest przykładem sprawnego i bezpiecznego funkcjonowania rozwiązania organizacyjno-analitycznego bazującego na łączonych danych z odrębnych rejestrów administracyjnych. Przy tworzeniu kompleksowych rozwiązań wykorzystujących dane administracyjne niezwykle istotne jest bowiem zagwarantowanie bezpieczeństwa danych i ochrona prywatności obywateli. Analogiczne rozwiązania na szerszą skalę są z powodzeniem stosowane np. w Szwecji. Bezpieczeństwu cyfrowemu służą zarówno precyzyjnie określone procedury przetwarzania danych w systemie, jak i statystyczny charakter wyników analiz – wnioski z analiz odnoszą się zawsze do grup, nie zaś do indywidualów, jak decyzje administracyjne. Dodatkowym efektem wykorzystania

danych administracyjnych jako źródła badań jest ukierunkowanie publicznej debaty na temat bezpieczeństwa cyfrowego danych podlegających administracji publicznej. Stworzy ono okazję do upowszechnienia dobrych praktyk związanych z bezpieczeństwem cyfrowym w administracji.

Stworzenie kanału dostępu do zasobów informacyjnych zawartych w rejestrach państwowych uruchomi mechanizmy wykorzystania ich potencjału przez zainteresowane podmioty. Zasoby te były dotychczas udostępniane jedynie w postaci prostych informacji zbiorczych w ramach statystyki publicznej. Przykładem są zarówno wyniki udostępniane przez Główny Urząd Statystyczny, jak i portal „Otwarte dane” (<https://dane.gov.pl>). W obu przypadkach są to informacje na temat wielu ważnych aspektów życia społecznego oraz funkcjonowania państwa. Udostępniane dane mają charakter zagregowany, co nie pozwala na prowadzenie analiz na poziomie indywidualnym, a jedynie na zbiorowościach. Ponadto prezentowane wyniki pochodzą z pojedynczych źródeł bez możliwości łączenia informacji z różnych rejestrów publicznych. Podmioty zainteresowane bardziej złożonymi analizami lub zagadnieniami pominiętymi w publikowanych danych podejmują próby ich pozyskania na własną rękę, np. w drodze badań terenowych. Jest to jednak kosztowny i pracochłonny proces, na który nie wszyscy mogą sobie pozwolić.

Stabilny mechanizm tworzenia, przetwarzania i analizy informacji na podstawie danych administracyjnych stwarza szansę na demokratyzację dostępu do danych i zwiększenie transparentności. Rzeczony mechanizm obejmowałby nie tylko publikację wybranych danych zbiorczych, lecz także pośrednictwo między zapotrzebowaniem analitycznym zgłaszanym przez podmioty a zabezpieczonymi danymi indywidualnymi. Otwarcie danych na użytek analiz będzie stanowić wsparcie dla biznesu, administracji państwowej i twórców polityk publicznych, a także dla nauki.

Ważne jest również stworzenie mechanizmów systemowego udostępniania danych administracyjnych. Korzystanie z danych administracyjnych odbywa się na zasadzie jednostkowych inicjatyw współpracy między instytucjami udostępniającymi dane i analitykami (w tym badaczami), często na podstawie wypracowywanych przez lata mechanizmów współpracy i rozwiązań prawnych. Taka sytuacja jest nieefektywna i ogranicza rozwój możliwości wsparcia przy kształtowaniu polityk publicznych w ramach badań prowadzonych w naukach społecznych. Dlatego też, na co wskazują praktyki międzynarodowe, pojawiają się instytucje pośredniczące w udostępnianiu danych w formie modelu pozwalającego na korzystanie z zasobów w sposób bardziej efektywny przez szerszą grupę badawczy, przy jednoczesnym zapewnieniu bezpieczeństwa danych z perspektywy ich gestorów (King, 2018).

W pierwszej kolejności w opracowaniu omówiono korzyści płynące z korzystania z danych administracyjnych do analiz, szczególnie wspierających kształtowanie polityk publicznych. W kolejnej części zaprezentowano wpływ wykorzystania danych publicznych na rozwój innowacyjnego państwa. W trzeciej części przedstawiono dobre praktyki budowania systemów umożliwiających korzystanie z danych administracyjnych na świecie, a w czwartej – doświadczenia polskie. W piątej części omówiono potencjalne obszary, w których gromadzone w Polsce dane administracyjne mogą być wykorzystane dla efektywnych i rzetelnych analiz wspierających tworzenie innowacyjnego państwa. W szóstej części przedstawiono rekomendacje dotyczące tworzenia systemu wsparcia analizy danych administracyjnych w Polsce.



Dane administracyjne jako podstawa efektywnych i rzetelnych analiz

Dane administracyjne są szczególną kategorią danych użytecznych publicznie. Coraz częściej bywają one wykorzystywane do bieżących celów sprawozdawczych przez podmioty prowadzące rejestry. Zauważa się potencjał zgromadzonych danych jako zasobu analitycznego, chociaż nie jest on jeszcze w pełni wykorzystywany. Zasoby każdego rejestru z osobna mogą być źródłem bardzo cennej wiedzy dotyczącej obszaru tematycznego związanego z tym rejestrem, jednak obecnie takie analizy prowadzone są głównie incydentalnie. Często uwadze analityków umyka wartość dodana płynąca z możliwości łączenia informacji pochodzących z różnych rejestrów. Powiązanie wielu zbiorów danych niesie za sobą szereg korzyści.

Najważniejszym pożytkiem z wykorzystania danych administracyjnych z różnych źródeł jest możliwość prowadzenia przekrojowych i pogłębionych analiz szerokiego spektrum zjawisk. Z jednej strony włączenie dodatkowych źródeł pozwala na pogłębienie analiz, umożliwia uwzględnienie szerszego zestawu czynników kształtujących badane zjawisko. Z drugiej – pozwala na wyodrębnienie nowych obszarów problemowych (grup społecznych, problemów społecznych i gospodarczych), których różne wymiary funkcjonowania są rejestrowane niezależnie przez różne instytucje. Przykładowo, do badania funkcjonowania absolwentów na rynku pracy potrzeba zarówno informacji, kto, kiedy i jaki kierunek studiów ukończył (w Polsce obecnych w *Zintegrowanym Systemie Informacji o Nauce i Szkolnictwie Wyższym POL-on*), jak i informacji o aktywności tych osób na rynku

pracy (np. na podstawie danych ZUS). Żadne z tych źródeł z osobna nie jest wystarczające – POL-on nie zawiera informacji na temat pracy, zarobków czy bezrobocia, zaś rejestry ZUS nie zawierają informacji o wykształceniu ubezpieczonych i dacie uzyskania dyplomu. Dopiero połączenie informacji z obu źródeł pozwala na opisanie sytuacji absolwentów na rynku pracy. Ponadto połączenie informacji z dwóch rejestrów umożliwia badanie interakcji między trajektoriami kształcenia a dynamiką aktywności na rynku pracy. Analizy takie są w Polsce prowadzone systemowo (w postaci zautomatyzowanej oraz ekspertyz autorskich) w ramach *Ogólnopolskiego systemu monitorowania ekonomicznych losów absolwentów szkół wyższych* (ELA).

Na cechy wyróżniające badania z użyciem danych administracyjnych ma wpływ specyfika wykorzystywanych zasobów. Po pierwsze badania te są wyjątkowo efektywne kosztowo i organizacyjnie. Użytkowane dane zebrano w ramach standardowych prac instytucji publicznych przy realizacji ich ustawowych zadań. Należy w tym kontekście odróżnić dwa pojęcia:

1. Dane administracyjne, zawierające dane osobowe, są podstawą realizacji ustawowych zadań instytucji publicznych przetwarzających je na użytek swoich zadań regulowanych prawnie; stanowią z reguły podstawę do podejmowania decyzji wobec indywidualu;
2. Dane analityczne, pozbawione danych osobowych, skonstruowane na podstawie danych publicznych, w tym danych administracyjnych, służą analizom statystycznym na potrzeby polityk publicznych, nauki i biznesu;

są podstawą do formułowania wniosków o charakterze statystycznym dotyczących zbiorowości.

Do przygotowania danych do analiz z zasady wystarcza niewielki – jak na skalę przedsięwzięcia – zespół badawczy oraz przygotowani urzędnicy publiczni. W przypadku tych danych nie występują bardzo wysokie koszty pozyskania, typowe dla badań społecznych i statystycznych prowadzonych metodami wywiadów. Realizowane prace eksperckie wymagają oczywiście znacznego wysiłku, odpowiedniej koordynacji pracy i zaangażowania specjalistycznego oprogramowania, jednakże sam proces gromadzenia informacji z punktu widzenia organizatora badania jest względnie prosty. Należy przy tym pamiętać, że przygotowanie zbiorów analitycznych na podstawie danych administracyjnych i ich wykorzystanie w dalszych analizach wymaga zastosowania specyficznej metodologii badań opartych na informacjach pochodzących z rejestrów administracyjnych (Jasiński i in., 2015, s. 45-72; Wallgren, Wallgren, 2007).

Dane znajdujące się w rejestrach zbiera się i udostępnia obligatoryjnie, czyli podmioty wskazane w przepisach nie mogą odmówić właścicielowi rejestru dzielenia się danymi. Prowadzi to do drugiej bardzo istotnej cechy danych rejestrowych – możliwości analizowania całych populacji – wszystkich osób i zdarzeń objętych rejestrami administracyjnymi. Trzecią cechą badań wykorzystujących dane obecne w rejestrach jest znacznie dokładniejszy pomiar badanych zjawisk – nieobciążony niedoskonałością pamięci badanych, jak np. w przypadku badań sondażowych. Po czwarte pojęcia używane w badaniach odwołujących się do informacji z rejestrów administracyjnych są konsekwencją precyzyjnych definicji urzędowych zawartych w zapisach prawnych regulujących funkcjonowanie rejestrów. Ogranicza to wprawdzie nieco swobodę w definiowaniu pojęć, jednakże w zamian zapewnia ich jednoznaczność i stabilność.

Ważnym aspektem jest prawny kontekst wykorzystywania rejestrów publicznych jako

źródła danych analitycznych. Rejestry prowadzone są na podstawie precyzyjnych przepisów prawa i możliwości ich wykorzystywania z zasady regulowane są ustawowo. Z jednej strony zmniejsza to w oczywisty sposób swobodę stawiania pytań badawczych pojawiających się w trakcie prac analitycznych, czy też możliwości odpowiedzi na pytania stawiane ad hoc przez administrację publiczną. Z drugiej jednak strony, jednoznaczność zapisów i procedur usprawnia komunikację między zespołami eksperckimi a gestorami rejestrów oraz twórcami polityk publicznych, chroni przed ewentualnymi zarzutami nadużywania informacji do celów pozaustawowych, wreszcie – sprzyja planowym, systemowym działaniom analitycznym skoordynowanym z uzgodnionymi potrzebami kluczowych odbiorców wyników analiz.

Pożytkiem płynącym z wykorzystania danych administracyjnych do analiz jest możliwość poprawiania jakości danych gromadzonych w rejestrach publicznych. Te same informacje bywają zapisywane przez różne instytucje. Zestawienie zapisów dotyczących tych samych faktów pozwala na weryfikację, a w konsekwencji sanację rejestrów. Stanowi to odpowiedź na stawiany niekiedy zarzut występowania błędów w rejestrach administracyjnych. Doświadczenia autorów tego raportu wskazują, że wykorzystanie danych administracyjnych do kompleksowych i systematycznych analiz jest kluczowym narzędziem wspomagającym proces diagnozy błędów w rejestrach. Wykorzystanie w analizach informacji z więcej niż jednego rejestru umożliwi wypełnienie ewentualnych luk i poprawienie błędów występujących w zasobach zaangażowanych w procedurę analityczną.

Wszystkie te cechy danych z rejestrów publicznych uzasadniają duże korzyści z możliwości prowadzenia badań. Fundacja NESTA (2019) wskazuje dane administracyjne jako mające największy potencjał osiągnięcia korzyści społecznych spośród wszystkich grup analizowanych danych.

Otwartość danych publicznych jako kreator innowacyjnego państwa

W wykorzystanie danych administracyjnych do prowadzenia analiz nie jest pomysłem nowym. Upowszechnienie tego typu przedsięwzięć nastąpiło w ostatnich latach wraz z gwałtownym rozwojem technologii IT i ucyfrowieniem administracji publicznej.

Potencjał danych pochodzących z zasobów informacyjnych państwa jest najbardziej widoczny w trzech obszarach:

1. Wsparcie dla rozwoju i ewaluacji polityk publicznych.
2. Wsparcie dla nauki.
3. Wsparcie dla biznesu i rozwoju przedsiębiorczości.

Poniżej przedstawiono przykłady wykorzystania danych administracyjnych w tych trzech obszarach.

Wsparcie dla rozwoju i ewaluacji polityk publicznych

Odpowiedzialne zarządzanie państwem wymaga równoczesnego uwzględniania wielu wzajemnie powiązanych czynników. Powinno się więc odbywać na podstawie najbardziej wiarygodnych i aktualnych danych, z wykorzystaniem narzędzi pozwalających na analizę tych współzależności i w sposób gwarantujący minimalizację ryzyka podjęcia złych decyzji.

Przykłady wykorzystania danych administracyjnych obejmują analizy efektywności polityk publicznych, w tym polityki społecznej, zdrowia czy polityki edukacyjnej. W szczególności pozwalają one na identyfikację

możliwości wprowadzania oszczędności budżetowych.

Oszacowano, że w USA koszty poniesione na koordynację opieki społecznej przyczynią się w przyszłości do zaoszczędzenia trzykrotnie większej kwoty na wydatki zdrowotne (Doar, Gibbs, 2017). Inne analizy przeprowadzone w Stanach Zjednoczonych wykazały, że szpitale stosujące krótszy czas hospitalizacji pooperacyjnej w przypadku kolektomii, pomostowania aortalno-wieńcowego (*bypassy*) oraz całkowitej endoprotezoplastyki stawu biodrowego redukowały całkowity koszt operacji o 6 proc. do 12 proc., nie generując przy tym większych kosztów związanych z opieką poszpitalną czy ewentualną readmisją (Regenbogen i in., 2017). Przedstawione wyżej przykłady stanowią jedynie małą część przypadków oszczędności budżetowych dzięki zastosowaniu analiz bazujących na informacjach pochodzących z rejestrów administracyjnych.

Innym przykładem są szwedzkie badania interakcji edukacji i rynku pracy. Program takich badań pozwolił na dogłębne zdiagnozowanie mechanizmów kryzysu rynku pracy, który Szwecja przechodziła we wczesnych latach 90. XX w. Kryzys jest z zasady najbardziej dolegliwy dla osób wchodzących na rynek pracy. Okazało się, że jego skutki były znacznie mniej odczuwalne dla absolwentów uniwersytetów niż dla osób kończących szkoły średnie. Wyniki studium, niezwykle cenne dla kształtowania polityki rynku pracy, uzyskano przy bardzo niskich kosztach. Roczny koszt systematycznie

prorowadzonych badań aktywności zawodowej i edukacyjnej wszystkich obywateli Szwecji wynosi około 50 000 USD, z czego połowę pochłania druk (Wallgren, Wallgren, 2007, s. 8-9).

Interesującym obszarem analiz, w których dane administracyjne są intensywnie wykorzystywane przy wspieraniu diagnozowania polityki publicznej, jest edukacja. Systemy monitorowania procesów edukacyjnych oraz aktywności zawodowej absolwentów funkcjonują w szeregu państw europejskich, m.in.: w Austrii, w Belgii, na Litwie, w Szwecji. Za najnowocześniejszy w Europie uznaje się *Ogólnopolski system monitorowania ekonomicznych losów absolwentów szkół wyższych* (ELA). Obszerne opisy wymienionych wyżej rozwiązań dostępne są w ekspertyzie opublikowanej przez Instytut Badań Edukacyjnych poświęconej możliwościom wykorzystania w rejestrach kwalifikacji informacji z systemów monitorowania losów absolwentów szkół wyższych (Bożykowski i in., 2019).

Wsparcie dla nauki

W miarę poprawy dostępności dane administracyjne zyskują także coraz większą rolę w badaniach społecznych, w tym w naukach ekonomicznych (Einav, Levin, 2014). Przykładem szeroko diskutowanych w ostatnich latach badań prowadzonych z wykorzystaniem danych administracyjnych są prace (Chetty i in., 2017, s. 94), w szczególności analiza międzypokoleniowej mobilności społecznej w Stanach Zjednoczonych i roli jaką w owej ruchliwości odgrywa wykształcenie wyższe. Badacze pokazali wyraźne zróżnicowanie szans dostania się na prestiżowe uczelnie należące do Ligi Bluszczowej w zależności od dochodów rodziców. Pokazali także, że uczelnie są bardzo zróżnicowane pod względem tworzonych przez nie szans na awans społeczny.

Dane administracyjne są szeroko wykorzystywane przez badaczy zajmujących się ścieżkami edukacyjnymi i zawodowymi. Kwestie

poruszane w badaniach to m.in. trajektorie w ramach systemu edukacji, zjawiska odsiewu i odpadu na studiach, różnice w sytuacji na rynku pracy między absolwentami różnych obszarów studiów, typem uczelni czy miejscem zamieszkania po studiach, dopasowania umiejętności do wymagań rynku pracy i zjawiska *over-education*, zwrot z edukacji, nierówności w dostępie do edukacji ze względu na płeć i stan zdrowia.

Dane administracyjne służą także prowadzeniu badań dotyczących medycyny i funkcjonowania służby zdrowia. W USA stwierdzono, że w regionach, w których częściej stosowano tomografię komputerową, wzrosła częstotliwość przeprowadzenia nefrektomii (usunięcia nerki). Wiąże się to z przypadkowym wykrywaniem guzów nerek podczas tomografii jamy brzusznej i klatki piersiowej (Welch i in., 2018, s. 221-227). Badanie przeprowadzone na populacji osób starszych, które doznały złamania kości, pozwoliło na estymację ryzyka ponownego złamania w ciągu roku od pierwszego wypadku i na wskazanie czynników wpływających na to ryzyko (Bynum i in., 2016, s. 2207-2215).

Łączenie danych z różnych rejestrów otwiera zupełnie nowe możliwości badawcze.

Jako przykład mogą posłużyć szwedzkie badania populacyjne dotyczące popełniania przestępstw z użyciem przemocy przez osoby cierpiące na depresję (Fazel i in., 2015, s. 224-232) oraz popełniania samobójstw, przestępstw z użyciem przemocy, i przedwczesnej śmierci wśród osób cierpiących na schizofrenię i pokrewne zaburzenia (Fazel i in., 2014, s. 44-54). Analizy te wykonano na podstawie danych z krajowego rejestru pacjentów oraz krajowego rejestru przestępstw.

Powyższa lista badań i publikacji stanowi jedynie wybór przykładów korzystania z danych administracyjnych w badaniach. Przedstawione

przykłady wskazują na ważną cechę badań naukowych wykorzystujących dane administracyjne. Publikacje akademickie wykorzystujące dane administracyjne mogą być często użyteczne także przy analizie efektów wdrażanych polityk. Aby tego typu badania były możliwe konieczne jest zapewnienie naukowcom dostępu do odpowiednich danych. Dostęp do danych administracyjnych pozwala naukowcom skutecznie konkurować na arenie międzynarodowej (Card i in., 2010, s. 1-4; Frank, 2000, s. 2398-2399).

Wsparcie dla biznesu i rozwoju przedsiębiorczości

Dane administracyjne są również ważne dla sektora przedsiębiorstw. Po pierwsze podmioty biznesowe powszechnie pozyskują i analizują dane, które służą optymalizacji szeregu decyzji podejmowanych przez przedsiębiorstwa. W części z nich pomocne mogą być informacje statystyczne wygenerowane na podstawie danych zawartych w rejestrach administracyjnych. Po drugie podmioty te mogą oferować usługi związane z analizą danych, w tym danych administracyjnych.

Dostęp do informacji na temat potencjalnego rynku, dostępnej infrastruktury czy instrumentów wsparcia może wspomagać podejmowanie decyzji o uruchomieniu i rozwoju działalności gospodarczej na danym terenie.

Dane administracyjne są źródłem informacji o populacji przedsiębiorstw, co jest kluczowe dla oceny istniejącego środowiska biznesowego – w tym potencjalnych kooperantów i konkurentów. W szczególności mogą one dostarczyć wiedzy o liczbie przedsiębiorstw, branżach działalności oraz ich aktywności, w tym np. dynamice rozwoju kadr (Pope, 2018).

Informacje pochodzące z rejestrów są przydatne ponadto w optymalizacji zarządzania kadrami. Znajomość rozkładu faktycznych płac osób o określonym zestawie cech wspiera polityki płacowe firm. Poza tym upowszechnienie

wiedzy (tzw. wiedza wspólna) na temat sytuacji na rynku pracy osób o wskazanych charakterystykach wyznacza punkt odniesienia w ustalaniu warunków zatrudnienia, w tym wysokości płac. Udostępnienie podobnej informacji także potencjalnym pracownikom służy uporządkowaniu relacji i urealnieniu wzajemnych oczekiwań pracodawców i potencjalnych pracowników.

Udostępnianie danych przez administrację rządową lub samorządową oraz przedsiębiorstwa coraz częściej postrzegane jest jako szansa na rozwój nowych firm technologicznych. Na świecie istnieją już organizacje, które mają promować ideę otwartości i dostępności danych (*open data*) oraz wspierać przedsiębiorców w wykorzystaniu danych w budowaniu ich biznesu, takie jak ODINE Open Data Incubator Europe, Open Data Institute (ODI). Zbiory przygotowane przez administrację stanowią jeden z podstawowych elementów służących do tworzenia nowych produktów i usług. Według danych przedstawionych przez ODI w 2015 r. ok. 70 proc. firm w Zjednoczonym Królestwie, opierających swój model biznesowy na przetwarzaniu danych, wykorzystywało informacje przygotowane przez administrację publiczną (Open Data Institute, 2015).

Doskonałym przykładem są powszechnie dostępne aplikacje korzystające z danych dotyczących komunikacji publicznej. W 2017 r. Deloitte (2017) przygotowało dla Transport for London (TfL), agencji odpowiedzialnej za transport w Londynie, w tym komunikację publiczną, raport podsumowujący korzyści płynące z udostępniania danych. Dane dotyczące sieci komunikacyjnej i przejazdów udostępniane przez TfL są wykorzystywane przez 600 różnych aplikacji stworzonych m.in. przez Waze, Twitter, Google, Apple, Citymapper, Bus Checker, Bus Times and Mapway, co pomaga użytkownikom sprawniej przemieszczać się w obrębie metropolii. Poza korzyściami dla TfL, np. w postaci oszczędności związanych z brakiem konieczności budowy własnej aplikacji czy mniejszym obciążeniem centrum

komunikacji z pasażerami oraz korzyściami dla pasażerów zyskujących łatwy dostęp do informacji oraz oszczędność czasu, autorzy raportu wymieniają także korzyści dla gospodarki regionu. Znaczna część firm korzystających z danych położona jest w okolicy Londynu, a wynikający z wykorzystania danych wzrost wartości dodanej

brutto w ciągu roku szacowany jest na 12 do 15 mln GBP. Firmy korzystające z danych tworzą też setki miejsc pracy w sektorze kojarzonym z wysoką produktywnością. Dodatkową korzyścią jest budowanie marki miasta jako wiodącego ośrodka, w którym prowadzi się działalność związaną z przetwarzaniem danych.



Jak udostępnia się dane administracyjne do badań w innych krajach? Dobre praktyki z zagranicy

W niniejszym opracowaniu zwracano już wielokrotnie uwagę na korzyści płynące z posługiwania się danymi pochodzącymi z wielu rejestrów administracyjnych, w szczególności, gdy możliwe jest wykorzystanie w analizach danych indywidualnych. Czynnikiem istotnie ułatwiającym wykorzystanie tego typu zasobów i znacznie poprawiającym bezpieczeństwo danych jest istnienie systemów integrujących dane i tworzących ramy organizacyjne do ich udostępniania. Poniżej przedstawiono rozwiązania funkcjonujące w Szwecji, Wielkiej Brytanii i Australii.

MONA i LISA w Szwecji

Dane gromadzone w publicznych rejestrach w Szwecji są udostępniane za pomocą dwóch systemów: MONA i LISA. Systemy te są prowadzone przez Statistics Sweden, stanowiąc w przybliżeniu strukturalny odpowiednik polskiego Głównego Urzędu Statystycznego.

System MONA: *Microdata Online Access* zapewnia bezpieczny dostęp internetowy do mikrodanych (danych jednostkowych) w szwedzkim systemie statystyki publicznej. Platforma dostarcza narzędzi do przetwarzania i analizy w postaci gotowego zestawu aplikacji.

Celem działania systemu jest zapewnienie bezpiecznego dostępu do systemu statystyki Szwecji, w którym żadne dane jednostkowe nie mogą opuścić systemu (platformę opuszczają wyłącznie dane zagregowane i przetworzone);

łatwego w obsłudze interfejsu dla użytkowników końcowych, opartego na dobrze znanych standardach technicznych i komponentach oprogramowania; dostępu do aktualizacji danych bez potrzeby tworzenia nowych zestawów dysków lub taśm do redystrybucji dla użytkowników zewnętrznych; kompletnego systemu z wydajnymi serwerami i bogatym zestawem aplikacji niewymagającego kosztownych urządzeń i ponoszenia kosztów oprogramowania przez użytkowników końcowych.

Ze względów bezpieczeństwa użytkownicy nie mają możliwości kopiowania danych między swoim terminalem a systemem MONA. Zamiast tego każdy użytkownik ma przypisaną swoją własną przestrzeń do przechowywania wyników, bibliotek kodów i innych utworzonych przez siebie obiektów. W celu pobrania tych plików należy skorzystać z indywidualnej skrzynki wychodzącej, skąd pliki są automatycznie wysyłane na wyznaczone konto pocztowe użytkownika.

Drugim systemem pozwalającym na systemowe wykorzystanie danych administracyjnych w Szwecji jest system **LISA (*Longitudinell integrationsdatabas för sjukförsäkrings- och arbetsmarknadsstudier*)**. To zintegrowana baza danych wzdłużnych dotyczących ochrony zdrowia oraz badań rynku pracy i edukacji.

W bazie LISA przechowuje się dane wzdłużne od 1990 r. dotyczące osób od 16. roku życia. Zawierają zmienne dotyczące zatrudnienia,

dochodu z pracy lub rekompensaty z tytułu zatrudnienia, branży zatrudnienia, dochodu rozporządzonego, kraju urodzenia (w tym również kraju urodzenia rodziców), ostatniego roku imigracji (jeżeli dotyczy), miejsca zamieszkania i zatrudnienia, poziomu wykształcenia.

Dostępne są również dane o przedsiębiorstwach: lokalizacja, typ produkcji, branża, liczba pracowników według płci i poziomu wykształcenia, zmiany zatrudnienia, wynagrodzenie, kluczowe wskaźniki i podstawowe dane ekonomiczne.

Dane LISA stanowią podstawę dla wzdłużnych analiz statystycznych oraz badań na całych populacjach lub w wyodrębnionych grupach. Możliwa tematyka analiz to zagadnienia związane z aktywnością ekonomiczną ludności: pracą zarobkową i alternatywnymi aktywnościami (studia, urlop rodzicielski, bezrobocie itp.) oraz chorobami.

Dane statystyczne z systemu LISA dostarczane są w postaci zagregowanej lub jako zanonimizowane dane jednostkowe w postaci gotowej do dalszych analiz. W tym drugim przypadku zakłada się, że materiał zostanie wykorzystany do celów badawczych lub statystycznych.

W większości przypadków w systemach MONA i LISA przeprowadza się specjalny przegląd etyczny w celu uzyskania dostępu do danych do celów badawczych. W przypadku każdego udostępniania danych jednostkowych Statistics Sweden przeprowadza specjalną procedurę zabezpieczającą zwaną „klauzulą tajności”. Jeśli projekt badawczy zawiera poufne informacje o osobach lub dane osobowe dotyczące przestępstw przeciwko prawu, niezbędne jest uzyskanie zgody Komisji etycznej. Warto zaznaczyć, że przez cały okres funkcjonowania tych systemów nie doszło do wycieku danych.

Korzystanie z systemów wymaga zarejestrowania się w Statistics Sweden za pośrednictwem

instytucji badawczej. Dane udostępniane są wyłącznie w celach badawczych, w szczególności nie ma możliwości udostępnienia ich służbom policyjnym w celach śledczych. Dostęp do danych jest odpłatny, kwotę wlicza się w wysokości wystarczającej na pokrycie kosztów pracy osób dokonujących przeglądu etycznego i eksportu danych.

Administrative Data Research Partnership w Wielkiej Brytanii

Przykładem systemowych rozwiązań wspierających wykorzystanie danych administracyjnych dla badań naukowych w Wielkiej Brytanii jest inicjatywa **Administrative Data Research Partnership (ADRP)**: ADRP współpracuje z Office for National Statistics (ONS) w celu strategicznego pozyskiwania i zarządzania danymi administracyjnymi oraz udostępniania ich badaczom w bezpieczny sposób, w celu prowadzenia wysokiej jakości badań naukowych o i dostarczaniu analiz na potrzeby priorytetowych tematów polityk publicznych w Wielkiej Brytanii.

Celem ADRP jest zapewnienie możliwości bezpiecznego dostępu do baz danych wytworzonych na podstawie danych administracyjnych departamentów rządowych, władz lokalnych i organów ochrony zdrowia. Projekty badawcze wykorzystujące te dane mają na celu wsparcie przy rozwiązywaniu głównych wyzwań społecznych i przy poprawie świadczenia usług publicznych w wielu dziedzinach, w tym w edukacji, opiece zdrowotnej i walce z przestępczością. Finansowanie tej inicjatywy pochodzi z Narodowego Funduszu Innowacji Produktywności (NPIF) i zapewnia funkcjonowanie partnerstwa w latach 2018-2021.

Przykładowe inicjatywy badawcze wspierane przez ADRP:

→ umożliwienie pomiaru różnych poziomów potrzeb u dzieci oraz ujawnienie

¹ <https://escr.ukri.org/research/our-research/administrative-data-research-partnership/> [dostęp: 12.07.2019].

różnych doświadczeń dzieci w kontekście ich rodzin, gospodarstw domowych, szkół i społeczności;

- stabilność i bezpieczeństwo finansowe osób starszych, w tym planujących emeryturę;
- lepsze zrozumienie produktywności firm, rolę przywództwa i wpływ inwestycji na wydajność w różnych sektorach gospodarki;
- badanie czynników wpływających na wchodzenie i wychodzenie osób z rynku pracy przez cały okres aktywności zawodowej.

Mechanizm udostępniania danych przewiduje specjalną autoryzację badaczy przez UK Statistics Authority. Zakłada się ich udostępnianie po przeprowadzeniu pseudonimizacji, aby uniknąć przekazywania danych osobowych.

MADIP w Australii

Kolejnym przykładem dotyczącym tworzenia systemowych rozwiązań dla korzystania z danych administracyjnych jest **Data Integration Partnership for Australia (DIPA)**. To rządowa inicjatywa mająca na celu lepsze wykorzystanie istniejących danych publicznych. Celem DIPA jest informowanie o rozwoju pojawiających się priorytetów polityki społecznej, gospodarczej i środowiskowej oraz poprawa świadczenia usług publicznych. Publiczne agencje wnoszą do partnerstwa uzupełniające się kompetencje: jedne zapewniają umiejętności analizy danych i polityk publicznych, podczas gdy inne – wiedzę techniczną i infrastrukturę.

Kluczowy komponent partnerstwa DIPA stanowi **Multi-Agency Data Integration Project (MADIP)**. MADIP jest porozumieniem sześciu australijskich agencji rządowych, służy łączeniu informacji pochodzących z rejestrów publicznych. Obejmuje szeroki zakres polityk publicznych, w tym m.in.: opieki zdrowotnej, edukacji, świadczeń społecznych, dochodów i podatków, ruchliwości społecznej, polityki rodzinnej.

Przeprowadzona ewaluacja projektu wykazała, że MADIP zwiększa potencjał potencjalnych informacji w celu lepszego ukierunkowania usług, takich jak opieka zdrowotna i usługi dotyczące wspierania rozwoju małych dzieci.

W ramach programu MADIP Australijskie Biuro Statystyczne (ABS) opracowało bazę danych *Expanded Analytical Business Longitudinal Database* (EABLD). Wykorzystując ABS Business Register (ABSBR) jako „kręgosłup” integrujący różne bazy danych, w ramach EABLD łączy dane administracyjne na poziomie przedsiębiorstw, takie jak zanonimizowane deklaracje aktywności biznesowej, oświadczenia podatkowe przedsiębiorstw z Australian Taxation Office (ATO) z danymi statystycznymi zbieranymi bezpośrednio przez ABS. Umożliwiło to naukowcom analizowanie zależności między cechami biznesowymi, innowacyjnością, produktywnością, tworzeniem miejsc pracy i przedsiębiorczością.

W badaniu oceniano rolę firm o różnej wielkości i wieku pod kątem ich wkładu we wzrost zatrudnienia. Metodologia pozwala na dokonanie porównań międzynarodowych. Prace te wykazały, że młode MSP i bardzo młode firmy konsekwentnie tworzą większość miejsc pracy w wielu krajach. Dane z EABLD wykazały, że duże firmy ograniczały zatrudnienie i były źródłem destrukcji miejsc pracy. Świadomość tego zjawiska nie była przedmiotem debaty politycznej w Australii, wręcz przeciwnie kolejne rządy mówiły o tym, że wielki biznes jest motorem tworzenia miejsc pracy. EABLD pomógł rządowi w weryfikacji hipotezy, że start-upy są ważnym źródłem generowania wzrostu zatrudnienia, ważniejszym niż większe, starsze firmy.

Dzięki EABLD zwiększono możliwości porównywania wyników przedsiębiorstw wspieranych przez politykę i programy rządowe z tymi, które nie otrzymały wsparcia, co jest przydatne w ocenie skuteczności i wartości interwencji rządowych. Baza EABLD pozwala również

na porównywanie efektów osiągniętych przez firmy o różnej wielkości, długości funkcjonowania i sektorach działalności, a docelowo również w układzie regionalnym. Pozwala to odpowiadać na pytania dotyczące czynników wsparcia rozwoju regionów, przedsiębiorstw, a także identyfikacji rozwojowych sektorów.

Zbiór danych i metodyka EABLD pozwalają na trwałe budowanie bazy danych przy niewielkich nakładach finansowych i organizacyjnych, co daje możliwość trwałego korzystania z tego narzędzia do diagnozowania i kształtowania polityki publicznej w kolejnych latach. Zwiększanie świadomości istnienia EABLD wśród instytucji publicznych oznacza również,

że więcej instytucji będzie korzystać z bazy w celu diagnozowania swoich potrzeb, co stwarza przestrzeń do dalszego rozwoju EABLD.

W systemie MADIP istnieją silne zabezpieczenia prawne w celu ochrony prywatności, w tym ustawa o prywatności z 1988 r. Dostęp do danych można uzyskać tylko za pośrednictwem bezpiecznych systemów, przez certyfikowanych naukowców i analityków. MADIP służy wyłącznie do analizy polityki, badań i celów statystycznych, nie można go używać do innych celów. Certyfikowani naukowcy poszukujący wzorców i trendów w danych w MADIP mają dostęp do zanonimizowanych informacji, nie ma możliwości identyfikacji pojedynczych osób.



Wykorzystanie danych administracyjnych do badań i oceny polityk publicznych w Polsce

W Polsce nie istnieje obecnie platforma udostępniania danych administracyjnych z rejestrów do celów analitycznych. Nie oznacza to jednak, że te dane nie są w ogóle wykorzystywane do prowadzenia badań i analiz. Analizy takie są prowadzone w pewnym zakresie również na łączonych danych pochodzących z więcej niż jednego źródła. Jest to wykorzystywane m.in. w analizach prowadzonych w Ministerstwie Finansów. Budowane tam modele mikrosymulacyjne wykorzystują potężne informacje z rejestrów podatkowych oraz Zakładu Ubezpieczeń Społecznych, co pozwala na kompleksowe analizowanie populacji osób płacących podatki oraz składki na ubezpieczenia społeczne, w tym na przykład osób zatrudnionych na podstawie umów cywilnoprawnych czy prowadzących jednoosobową działalność na własny rachunek.

Innym przykładem jest możliwość wykorzystania danych pochodzących z ZUS do ewaluacji efektów programów wspierania beneficjentów środków publicznych. Na podstawie tych danych Ministerstwo Infrastruktury i Rozwoju szacuje wskaźniki związane z realizacją projektów finansowanych z Europejskiego Funduszu Społecznego w Regionalnych Programach Operacyjnych (Ministerstwo Inwestycji i Rozwoju, 2016). Analizy te prowadzi się w ramach administracji publicznej i ich efekty nie są szeroko publikowane.

System ELA

Przykładem wykorzystania danych publicznych wraz z udostępnieniem wyników jest wspomniany wcześniej *Ogólnopolski system monitorowania ekonomicznych losów absolwentów szkół wyższych*, znany jako system ELA (<http://ela.nauka.gov.pl/>). System bazuje na zapisach obecnych w art. 352 ustawy Prawo o szkolnictwie wyższym i nauce (Ustawa..., 2018). W ustawie określono źródła danych wykorzystywanych w ELA, ich zakres oraz procedury przetwarzania informacji. System korzysta zasadniczo z dwóch źródeł: z rejestrów Systemu POL-on oraz Zakładu Ubezpieczeń Społecznych (rejestry ubezpieczonych i płatników składek). W ustawie w szczególności wskazane są wszystkie zmienne w zbiorach danych objętych systemem ELA.

Procedury przewidziane w ustawie gwarantują zachowanie pełnej anonimowości osób objętych badaniem. W analizowanym zbiorze brak jakichkolwiek danych osobowych, zaś absolwenci są odróżniani na podstawie losowego numeru, który nie daje możliwości identyfikacji osób. Dzięki brakowi konieczności kontaktów z absolwentami w celu przeprowadzenia monitoringu nie narusza się prywatności badanych. Ponadto w zapisach ustawowych regulujących funkcjonowanie systemu ELA wprowadzono jednoznaczne zasady chroniące prywatność badanych i uniemożliwiające identyfikacje osób.

Jednoznaczność zapisów i procedur usprawnia komunikację między zespołem

eksperckim ELA a ZUS, chroni przed ewentualnymi zarzutami nadużywania informacji do celów pozaustawowych oraz sprzyja planowym i systemowym działaniom analitycznym skoordynowanym z uzgodnionymi potrzebami kluczowych odbiorców wyników systemu.

System ELA dostarcza informacji o funkcjonowaniu absolwentów szkół wyższych na rynku pracy. Dostęp do wyników jest powszechny i bezpłatny. Analizy obejmują wszystkich absolwentów w Polsce. Informacje udostępnia się na poziomach: kierunku studiów, uczelni i ogólnopolskim, w języku polskim i angielskim. W ramach publikowanych raportów charakteryzowane są trzy merytoryczne aspekty sytuacji na rynku pracy:

- czas poszukiwania pracy,
- stabilność pracy,
- wynagrodzenia.

Podstawową formą prezentacji wyników w ramach ELA są automatycznie generowane

raporty opisujące sytuację zawodową absolwentów, według jednolitych wzorców.

System ELA należy do najbardziej innowacyjnych tego typu rozwiązań w Europie. The Central Asia Education Platform (CAEP) określa to rozwiązanie jako „prawdopodobnie najbardziej zaawansowany system monitorowania losów absolwentów w Unii Europejskiej” (CAEP, 2019). Stanowi przykład ilustrujący olbrzymi potencjał wykorzystania danych administracyjnych do celów analiz, zarówno jako system monitoringu wspierającego evidence-based policy, jak i zasób do pogłębionych analiz naukowych. Przypadek ELA wskazuje, że tego typu przedsięwzięcia mogą być prowadzone w sposób bezpieczny i gwarantujący ochronę prywatności osób objętych badaniem. Warto podkreślić, że system korzysta jedynie z dwóch źródeł danych rejestrowych: z POL-on i ZUS; wykorzystanie szerszego zestawu danych znacznie zwiększyłoby możliwości prowadzenia analiz.



Potencjał danych administracyjnych w Polsce

Caość gospodarki opartej na danych ma coraz większy udział w PKB poszczególnych krajów. Liderem w Europie jest Estonia (4,3 proc.) oraz Wielka Brytania (3,5 proc.). W scenariuszu szybkiego wzrostu wartości te mogą wzrosnąć w 2025 r. odpowiednio do nawet 12,1 proc. i 8,9 proc., przy średniej dla całej Unii rosnącej z 2,6 proc. do 6,3 proc.

W Polsce wartość gospodarki opartej na danych szacowana jest obecnie na 6,2 mld EUR (ok. 26,5 mld PLN), a w 2025 r. może to być – w zależności od scenariusza wzrostu – między 7,9 mld EUR a 12 mld EUR (odpowiednio 33,7 mld PLN i 51,2 mld PLN). W odniesieniu do PKB z obecnego 1,2 proc. PKB można osiągnąć nawet 2 proc. (wszystkie dane za IDC, 2019). Umiejętne udostępnianie danych administracyjnych może w istotny sposób wpłynąć na powodzenie realizacji scenariusza optymistycznego, zwiększając pulę dostępnych zasobów, a tym samym przyczyniając się do zwiększenia liczby firm opierających swój biznes na danych.

Otwieranie danych publicznych to jednak nie tylko korzyści dla firm i nowe przedsiębiorstwa – to również wymierne korzyści dla samej administracji. Według szacunków Komisji Europejskiej z 2015 r., polski sektor publiczny mógł zaoszczędzić ponad 200 mln PLN w latach 2015-2020 dzięki udostępnieniu posiadanych danych (European Commission, 2015).

Dane administracyjne w rejestrach publicznych w Polsce

Zasoby informacyjne gromadzone w rejestrach publicznych w Polsce stwarzają różnorodne możliwości analizy, zarówno z wykorzystaniem

informacji w poszczególnych rejestrach, ale także informacji pochodzących z kilku rejestrów. Powszechne wykorzystanie numeru PESEL jako identyfikatora osoby pozwala na jednoznaczne łączenie danych pochodzących z różnych rejestrów. Jednocześnie wykorzystanie odpowiednich kluczy służących do anonimizacji danych pozwala na bezpieczne łączenie danych pochodzących z różnych źródeł przy jednoczesnym zagwarantowaniu ich bezpieczeństwa.

W konsekwencji możliwe jest wykorzystanie łącznego potencjału (ze znacznym efektem synergii) danych administracyjnych. W Polsce administracja państwowa prowadzi się ponad 700 rejestrów i systemów informacyjnych. Bazując na doświadczeniach międzynarodowych oraz zasobie informacyjnym danych administracyjnych dostępnych w Polsce, obszary, w których można wykorzystać dane administracyjne do prowadzenia analiz, obejmują m.in. edukację i rynek pracy, warunki życia, opiekę zdrowotną czy bezpieczeństwo i resocjalizację. Ponadto stworzenie mechanizmów integracji informacji obecnych w zasobach państwa spowoduje radykalne zwiększenie potencjału analitycznego tych danych i umożliwi prowadzenie analiz o charakterze ponadresortowym.

Dane administracyjne jako potencjalne źródło wspierania kształtowania i realizacji polityk publicznych

Dane gromadzone w rejestrach publicznych mogą być wykorzystane przy projektowaniu i ewaluacji polityk publicznych, co stanowi wsparcie dla administracji na poziomie centralnym. Informacje zgromadzone w rejestrach administracyjnych

mogą stanowić wsparcie dla oceny skutków regulacji (OSR). Należy pamiętać, że skutki dziś podjętej decyzji często widoczne są po wielu latach, wykorzystanie danych gromadzonych w rejestrach umożliwia projektowanie rozwiązań, które pozwolą na podstawie danych bieżących oraz historycznych, oceniać faktyczne i możliwe następstwa podejmowanych decyzji.

Zarządzający, aby móc podejmować właściwe decyzje, oprócz dostępu do aktualnych danych będących w dyspozycji jednostek odpowiedzialnych za konkretny obszar, powinni mieć do dyspozycji narzędzia, które wykorzystując modele stworzone przez kompetentny zespół analityków, pozwolą na prowadzenie wielowymiarowych analiz następstw decyzji oraz predykcję zdarzeń. Wykorzystanie zintegrowanych danych z rejestrów administracyjnych pozwoli ponadto na precyzyjną diagnozę skali analizowanych zjawisk, jak też na rozpoznanie ich społecznych uwarunkowań.

Poniżej przedstawiono wybrane przykłady potencjalnego wykorzystania danych administracyjnych w Polsce do celów analiz na użytek polityk publicznych, jednak możliwości, które daje zintegrowanie zasobów informacyjnych Polski są znacznie szersze.

Wsparcie dla polityki zdrowotnej

Pierwszym polem działania, na którym mogą być wykorzystane te dane, jest polityka publiczna w obszarze ochrony zdrowia. W Polsce obserwuje się brak koordynacji działań na poziomie instytucji odpowiedzialnych za finansowanie ze środków publicznych świadczeń medycznych oraz świadczeń pieniężnych w razie choroby. Jest to konsekwencją istnienia prawnego podziału zarządzania ryzykiem choroby między mechanizmem powszechnego ubezpieczenia społecznego i ubezpieczenia zdrowotnego. Sytuacja taka prowadzi do potencjalnej nieefektywności wydatkowania środków w ramach bazowej części systemu zabezpieczenia społecznego. Obecnie nie

prowadzi się systematycznych (i systemowych) analiz łączących obie perspektywy, co uniemożliwia podejmowanie decyzji w zakresie obu systemów rozpatrywanych łącznie.

Wykorzystanie zintegrowanej informacji zawartej w rejestrach pozwoli na systemowe wsparcie szacowania pełnych kosztów terapii a także ewaluacji skuteczności i jakości leczenia oraz rehabilitacji leczniczej, ruchowej, społecznej i zawodowej w kontekście jednolitego systemu orzecznictwa z tytułu niezdolności do pracy i niepełnosprawności. Ponadto proponowane rozwiązanie zapewni wsparcie przy wypracowywaniu optymalnych metod wychodzenia z chorób oraz środków profilaktyki. Możliwe będzie wskazanie technologii medycznych efektywnych nie tylko klinicznie, ale też społecznie, tj. takich, które umożliwiają pacjentom powrót na rynek pracy, przynosząc potencjalny przyrost PKB.

Szersze korzystanie z danych administracyjnych daje możliwości do wytworzenia mechanizmu wykorzystywania zintegrowanych informacji o świadczeniach opieki zdrowotnej, rehabilitacji leczniczej, społecznej i zawodowej przy kształtowaniu polityk publicznych, w szczególności zdrowotnej i społecznej oraz przy ustalaniu cen świadczeń medycznych i rehabilitacyjnych. Zapewni też wsparcie rozwoju statystyki publicznej w zakresie udzielanych świadczeń.

Wsparcie dla polityki edukacyjnej i rynku pracy

Drugim ważnym polem działań państwa, zyskującym na dostępie do zintegrowanej informacji zawartej w rejestrach administracyjnych, jest edukacja.

Jak pokazuje przykład systemu ELA w Polsce dane publiczne dają możliwość systematycznego monitorowania losów absolwentów szkół i uczelni. Na przykład gromadzone informacje pochodzące również z Systemu Informacji Oświatowej pozwalają na rozszerzenie zakresu monitorowania o szkoły ponadpodsta-

wowe, w tym zawodowe, których absolwenci rozpoczynają swoją aktywność na rynku pracy. Dane pochodzące np. z Badania Aktywności Ekonomicznej Ludności pozwalają stwierdzić, że ryzyko bezrobocia jest szczególnie wysokie wśród absolwentów zasadniczych szkół zawodowych. Obecnie nie ma ogólnopolskich badań, które pozwalają na sformułowanie oceny zawodów i rynków pracy, w których ryzyko bezrobocia wśród tej grupy jest szczególnie wysokie. Badanie takie prowadzono w województwie małopolskim w latach 2011-2015, ale nie jest obecnie kontynuowane, ze względu na koszty związane z koniecznością prowadzenia ankiet wśród szkół i absolwentów. Wykorzystanie danych rejestrowych pozwoliłoby na uzyskanie informacji o absolwentach na poziomie ogólnopolskim, bez dodatkowych kosztów. Założenia takiego systemu są obecnie wypracowywane w ramach projektów realizowanych w Ministerstwie Edukacji Narodowej oraz Instytucie Badań Edukacyjnych. Podstawową przeszkodą dla ich realizacji jest brak możliwości skorzystania z zasobów danych administracyjnych.

Wsparcie dla polityki społecznej, w tym świadczeń społecznych

Gromadzone w rejestrach publicznych dane dotyczące korzystania ze świadczeń społecznych oraz płacenia podatków pozwalają m.in. na poszerzenie wiedzy dotyczącej sytuacji społecznej osób korzystających z różnego typu świadczeń społecznych, w tym kumulacji świadczeń w wybranych grupach gospodarstw domowych. W szczególności analizy takie mogą pozwolić na identyfikację luk w systemie zabezpieczenia społecznego i grup wymagających nowego wsparcia. Obecnie informacje o korzystaniu ze świadczeń społecznych (m.in. pomocy społecznej, świadczeń rodzinnych, świadczenia wychowawczego, świadczeń z systemów ubezpieczeń społecznych) oraz płacenia danin publicznych (podatków, składek na ubezpieczenia

społeczne i zdrowotne) gromadzi się w odrębnych rejestrach.

Zintegrowana informacja może pozwolić na pozyskanie wiedzy dotyczącej korzystania ze świadczeń oraz osiągania dochodów z pracy, a także ocenę, czy potencjalni beneficjenci wykorzystują dostępne wsparcie, czy też na przykład z powodu braku wiedzy o istniejących możliwościach, nie korzystają z pomocy.

Wsparcie dla polityki lokalnej

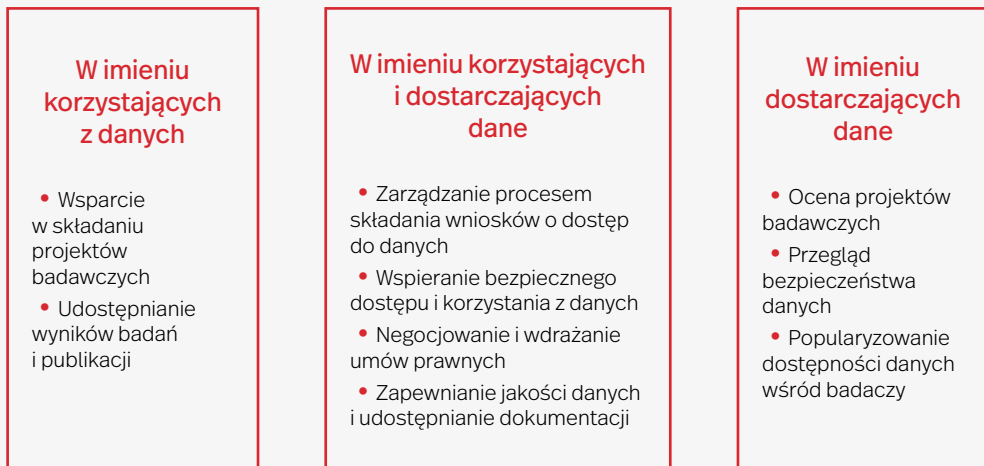
Wykorzystanie danych administracyjnych daje również możliwości wspierania prowadzenia polityki publicznej na poziomie samorządowym. Większość dostępnych danych statystycznych pochodzących z badań nie pozwala na diagnozowanie zjawisk na poziomie gminnym lub powiatowym. Z danych gromadzonych w rejestrach publicznych można pozyskać informacje o wielu obszarach polityki publicznej na poziomie lokalnym. Dotyczy to w szczególności obszarów wspierania rozwoju przedsiębiorczości na rynku lokalnym, przez gromadzenie i udostępnianie informacji dotyczących otoczenia sektora biznesu, w tym informacji o infrastrukturze, dostępnych usługach publicznych itp.

Wsparcie dla polityki naukowej

Jednym z ważnych celów polityki państwa jest wspieranie jakości prowadzonych badań naukowych. Umożliwienie autoryzowanym badaczom dostępu do danych publicznych oraz inicjowanie projektów badawczych wykorzystujących takie dane jest ważnym sposobem wsparcia badań naukowych. Utworzenie instytucji pośredniczących w dostępie do danych oraz powierzenie im zadań związanych z korzystaniem z danych publicznych jest istotnym narzędziem wspierania, na co wskazują również opisane praktyki międzynarodowe. Efekty takich badań mają również wysoki potencjał ich publikowania w piśmiennictwie międzynarodowym.

Funkcje instytucji pośredniczących mogą wspierać zarówno tych którzy z danych korzystają, jak i tych którzy danych dostarczają (King, 2018). Przedstawiono to na rysunku 1.

➤ **Rysunek 1. Funkcje pośredników danych: co mogą robić i w czym imieniu?**



Uwaga: Rysunek przetłumaczony i zaadaptowany na podstawie licencji: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). Tłumaczenie i adaptacja dokonane przez autorów raportu.

Źródło: King (2018).

Wspieranie badań naukowych z wykorzystaniem danych administracyjnych może być promowane w ramach programów dofinansowania badań podstawowych realizowanych ze środków Narodowego Centrum Nauki, jak również badań aplikacyjnych, np. finansowanych przez Narodowe Centrum Badań i Rozwoju. Pozwoli to na przeznaczenie znacznie większej puli dostępnych środków na prowadzenie badań, przez ograniczenie znacznych kosztów

ponoszonych na zbieranie informacji o faktach, gromadzonych w ramach finansowanych ze środków przeznaczonych na naukę badań ankietowych, które duplikują informacje zbierane w rejestrach administracyjnych. Badania takie mogą być również zlecane przez zainteresowane podmioty publiczne. Proces aplikowania o takie projekty może być również wspierany przez instytucje zarządzające dostępem do danych publicznych.

Opis koncepcji integracji zasobów informacyjnych państwa oraz mechanizmu udostępniania danych

Niniejsza część raportu stanowi rodzaj podsumowania przedstawionych wyżej rozważań. Autorzy zarysowali potencjał koncepcji wtórnego wykorzystania danych administracyjnych jako wsparcia dla gospodarki, administracji publicznej oraz nauki. Opisali również wybrane przypadki zastosowania tego podejścia w różnych krajach. Jednym z pozytywnych przykładów jest polski system ELA, którego sukces, jako wsparcia dla MNiSW, społeczności akademickiej oraz podmiotów rynku pracy i stabilność jego funkcjonowania wynika m.in. z możliwości jakie zespołowi eksperckiemu ELA stworzyły zapisy zawarte w ustawie Prawo o szkolnictwie wyższym i nauce (wcześniej w ustawie Prawo o szkolnictwie wyższym).

ELA jest przykładem efektywnego podejścia do analityki opartej na danych administracyjnych. Jak napisano wyżej, aby skutecznie i bezpiecznie przetwarzać dane pochodzące z rejestrów administracyjnych na użytek analiz, potrzeba jasno sformułowanych, jednoznacznych przepisów precyzyjnie regulujących niezbędne procedury. Wymaga to również odpowiedniego zaplecza instytucjonalnego realizującego i nadzorującego te procesy, zarówno pod względem merytorycznym, jak i administracyjnym.

Efektywne korzystanie z zasobów informacyjnych państwa wymaga jasnej i precyzyjnej koordynacji funkcjonowania podmiotów zaangażowanych w złożony proces integracji i analizy

zasobów informacyjnych na użytek analiz prowadzonych na rzecz rozwoju społecznego i gospodarczego, w tym prowadzenia polityk publicznych. Potrzebne jest zatem określenie:

- procedur analizy zgłoszeń nowych przedsięwzięć analitycznych i badawczych,
- zasad wskazywania i diagnozowania niezbędnych zasobów informacyjnych,
- mechanizmów ułatwiających angażowanie kompetentnych zespołów analitycznych w analitykę państwa i zapewniających ciągłą współpracę administracji publicznej z tymi zespołami,
- mechanizmów monitorowania i rozwijania procedur bezpiecznego przetwarzania, analizy i anonimizacji danych oraz udostępniania wyników w postaci użytecznej dla twórców polityk publicznych.

Określenie tych kwestii, a także celów i zasad integracji zasobów informacyjnych państwa (ZIP), rozumianych jako informacje gromadzone w rejestrach prowadzonych przez różnych gestorów do realizacji ich zadań publicznych wymaga odpowiednich regulacji – w postaci specjalnych **zapisów ustawowych dotyczących integracji zasobów informacyjnych państwa**. Powinny one w szczególności określić zasady korzystania z tych zasobów do prowadzenia analiz strategicznych w ramach polityki publicznej, w tym zakresu i sposobu integracji ZIP, zarządzania zintegrowanymi zasobami oraz narzędzi wspierających integrowanie i wykorzystywanie

tych zasobów. Niezbędne jest również wskazanie instytucji pełniącej rolę operatora ZIP, którego zadaniem będzie zarządzanie systemem gromadzenia i udostępniania zasobów do prowadzenia analiz.

Ustalenie jasnych i jednoznacznych reguł dostępu do informacji pochodzących z rejestrów publicznych będzie zarazem stanowiło mechanizm chroniący ZIP przed próbami niekontrolowanego pozyskania danych administracyjnych do nieautoryzowanych celów. W tym sensie w postępującej cyfryzacji kolejnych sfer życia utworzenie regulowanego kanału dostępu do ZIP zapewni bezpieczeństwo tych zasobów.

Podstawowym celem integracji ZIP jest wykorzystanie danych gromadzonych w rejestrach publicznych do prowadzenia zintegrowanych analiz o charakterze ponadresortowym oraz optymalizacja dostępu do informacji

gromadzonych w tych rejestrach, w tym na potrzeby zadań publicznych realizowanych przez administrację rządową i samorządową w celu realizacji polityki opartej na faktach oraz wsparcie tych zadań przez badania naukowe.

Zgodnie z modelem zaproponowanym przez King (2018) na podstawie analizy doświadczeń międzynarodowych utworzenie takich rozwiązań pozwala na w pełni efektywne korzystanie z dostępnych zasobów informacyjnych. Obecnie w Polsce sposób korzystania z danych administracyjnych jest na drugim etapie tego modelu. Podejmowane działania w ramach projektu „Zintegrowana Platforma Analityczna” (opisane w dalszej części opracowania) stanowią trzeci etap procesu. Ostatnim, czwartym etapem jest instytucjonalizacja procesu gromadzenia i udostępniania danych do analiz i badań zasobów informacyjnych państwa (rysunek 2).

➤ Rysunek 2. Model procesu dochodzenia do integracji zasobów informacyjnych państwa.



Uwaga: Rysunek przetłumaczony i zaadaptowany na podstawie licencji: Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). Tłumaczenie i adaptacja dokonane przez autorów raportu.

Źródło: King (2018).

Doświadczenia autorów niniejszego opracowania przy projektowaniu i rozwijaniu Systemu ELA oraz podczas prac nad przygotowaniem projektu „Zintegrowana Platforma Analityczna” stanowią podstawę do sformułowania propozycji rozwiązań prawno-instytucjonalnych, które pozwolą rozwinąć w Polsce system analityczny wspierający całą administrację publiczną, a zarazem stworzyć efektywny mechanizm umożliwiający systematyczną współpracę administracji publicznej i środowisk naukowych.

Przedstawione niżej propozycje opierają się na funkcjonowaniu trzech elementów:

- **Operatora zasobów informacyjnych państwa** (OP-ZIP) odpowiedzialnego za całość procesów związanych z przygotowaniem danych do prowadzenia analiz,
- **„Zintegrowanej Platformy Analitycznej”** (ZPA) stanowiącej narzędzie do prowadzenia analiz strategicznych,
- **Środowiska badawczego** stanowiącego zaplecze kompetencyjne analityki publicznej.

Integracja zasobów informacyjnych państwa do prowadzenia analiz

Efektywny dostęp do zasobów informacyjnych państwa wymaga wskazania instytucji będącej kluczowym operatorem tych danych. Instytucja taka, określana dalej jako Operator ZIP (OP-ZIP), powinna realizować następujące zadania:

- nadzorowanie procedur łączenia, przetwarzania centralnych danych publicznych oraz udostępniania przetworzonych zasobów i wyników analiz opartych na tych danych,
- nadzorowanie funkcjonowania „Zintegrowanej Platformy Analitycznej” oraz określanie jej zadań,
- powoływanie i nadzorowanie funkcjonowania Środowiska badawczego ZIP

odpowiedzialnego za prowadzenie badań na użytek administracji publicznej.

Zakres ZIP zaangażowanych w proces integracji obejmuje wszystkie zasoby niezbędne do realizacji zgłoszonych badań, niemniej jednak specyfika Polski wymaga szczegółowego określenia potencjalnego zakresu danych w odpowiednich przepisach. Oznacza to, że zakres zasobów udostępnianych w ramach proponowanego rozwiązania może być stopniowo rozszerzany, w miarę pojawiania się zapotrzebowania na analizy z nowych obszarów.

Szczególną rolę OP-ZIP pełni w procesie zapewnienia bezpieczeństwa dostępu do danych. Wymaga to precyzyjnego zdefiniowania zasad łączenia, zabezpieczania i analizy danych w ramach integracji ZIP oraz procedur analiz prowadzonych w ramach systemu na zamówienie jednostek administracji publicznej oraz zgłaszanych przez zespoły badawcze.

Możliwość prowadzenia analiz *offline*

Bardzo istotnym składnikiem proponowanego rozwiązania jest mechanizm umożliwiający realizację analiz *offline* na podstawie ZIP, z wyłączeniem bezpośredniego dostępu do danych. Przykładowo, system mógłby działać w następujący sposób: użytkownik przesyła instrukcje realizacji analiz do OP-ZIP, który po weryfikacji przesłanego kodu realizuje zamówioną analizę i przesyła użytkownikowi wyniki w formie zregulowanej. Środowisko badawcze wspiera OP-ZIP w realizacji tego procesu. Szczegółowa procedura pozwala na pogodzenie sprawności funkcjonowania systemu z bezpieczeństwem analiz danych i udostępniania wyników.

Udostępnienie możliwości prowadzenia analiz *offline* jest odpowiedzią na dwie powiązane ze sobą potrzeby. Z jednej strony tworzy możliwości szerszego korzystania przez instytucje badawcze i inne podmioty z zasobów informacyjnych państwa i będzie silnym bodźcem dla rozwoju badań naukowych. Z drugiej strony

ułatwia administracji publicznej korzystanie z potencjału kompetencyjnego różnorodnych wyspecjalizowanych środowisk naukowych za pośrednictwem Środowiska badawczego ZIP. Pozwala to na większe zaangażowanie tych środowisk w działalność na rzecz analityki publicznej.

„Zintegrowana Platforma Analityczna”

Drugim elementem realizacji wizji integracji zasobów informacyjnych państwa jest stworzenie narzędzia do prowadzenia analiz opartych na danych zintegrowanych. Takie narzędzie wypracowywane jest obecnie w partnerskim projekcie „Zintegrowana Platforma Analityczna” (ZPA) z udziałem Ministerstwa Cyfryzacji (lider projektu), instytucji administracji centralnej (Ministerstwa Nauki i Szkolnictwa Wyższego, Ministerstwa Rodziny, Pracy i Polityki Społecznej, Ministerstwa Zdrowia) oraz instytucji ze środowiska naukowego (Uniwersytetu Warszawskiego i Szkoły Głównej Handlowej w Warszawie). **Celem** ZPA jest stworzenie informatycznej i analitycznej infrastruktury służącej łączeniu, przetwarzaniu i analizie zasobów informacyjnych państwa oraz zintegrowanie środowisk naukowych wokół analityki państwa. Stworzone rozwiązanie powinno umożliwić prowadzenie analiz wspierających prowadzenie polityk wszystkim resortom. Ma charakter horyzontalny w dwóch znaczeniach: zarówno jako platforma wspierająca współpracę międzyresortową przy prowadzeniu polityk publicznych, jak i ze względu na wykorzystanie zasobów pochodzących z rejestrów podlegających różnym resortom.

Wśród zakładanych **efektów** ZPA należy wymienić przede wszystkim ułatwienie twórcom polityk publicznych dostępu do zasobów przygotowanych do prowadzenia analiz wspierających ich działania (jako ewaluacja realizowanych działań oraz jako podstawa do planowania przyszłych polityk i prognozowania ich skutków).

Oznacza to stworzenie narzędzi i procedur ułatwiających wykorzystanie analityki w obszarze strategicznego zarządzania państwem.

Szczególną cechą ZPA jest model funkcjonowania zorganizowany według **dziedzinowych rozwiązań analitycznych** – problemów społecznych wskazanych przez twórców polityk publicznych. Pozwala to na zoptymalizowanie mechanizmu łączenia, przetwarzania i analizowania danych, nie na zasadzie łączenia „wszystkiego ze wszystkim” lecz według logiki budowania zasobów zawierających wskaźniki przeznaczone do obrazowania konkretnych problemów społecznych, posiadających jasne i weryfikowalne interpretacje, użytecznych dla wsparcia podejmowania decyzji. To podejście odróżnia ZPA od typowych badań *big data*.

Na etapie tworzenia ZPA będzie koncentrować się na sześciu zagadnieniach dziedzinowych wskazanych przez partnerów projektu, które mają dotyczyć obszarów na styku edukacji i rynku pracy oraz ochrony zdrowia i pomocy społecznej.

Infrastruktura informatyczno-analityczna oraz baza wiedzy powstała w ramach projektu ZPA będzie użyteczna przy realizacji przedsięwzięć analitycznych również w innych obszarach, w miarę wzrostu zapotrzebowania i zgłaszania kolejnych wniosków.

Stworzenie narzędzi służących systematycznej, precyzyjnej i trafnej analizie problemów wymienionych wyżej oraz kolejnych pojawiających się w obszarze zainteresowań twórców polityk publicznych, wymaga stałej współpracy administracji publicznej i środowisk naukowych. Dlatego ZPA stworzy techniczne i organizacyjne możliwości współpracy tych środowisk w rozwoju analityki państwa i podnoszeniu efektywności rozwiązywania problemów społecznych.

Docelowo „Zintegrowana Platforma Analityczna” powinna być narzędziem służącym **integracji zasobów informacyjnych państwa** i może spełniać rolę hurtowni danych oraz technicznego zaplecza analitycznego OP-ZIP.

Środowisko badawcze

Dopełnieniem proponowanego systemu integracji i udostępniania danych z rejestrów jest zapewnienie **interakcji między administracją publiczną a światem nauki**, który stanowi naturalne zaplecze do dostarczania wiedzy opartej na faktach, wynikających z prowadzonych badań krajowych oraz międzynarodowych.

W tym celu powstanie tzw. Środowisko badawcze (trwają prace w ramach projektu ZPA), zorganizowana struktura zaangażowana w wymianę doświadczeń na poziomie krajowym i międzynarodowym. Będzie to platforma integracji administracji publicznej, instytucji odpowiedzialnych za statystykę publiczną (GUS) i środowiska naukowego, w tym badaczy i analityków zajmujących się politykami publicznymi.

Systematyczna współpraca przedstawicieli środowiska naukowego oraz administracji publicznej ma umożliwić wzajemne uzupełnianie się kompetencji tych grup oraz wzajemną kontrolę podejmowanych działań. Pozwoli to na zapewnienie wysokiego poziomu całego przedsięwzięcia, w szczególności wysokich standardów metodologicznych i użyteczności rezultatów badań, zarówno dla podniesienia efektywności prowadzonych polityk publicznych, jak i dla celów analiz naukowych. Analizy te będą skutkować identyfikacją kolejnych występujących problemów, a w konsekwencji pomogą w formułowaniu i realizacji celów polityk publicznych. Efektem prac będzie również systematyczne doskonalenie mechanizmów zapewniania bezpieczeństwa danych oraz ochrony prywatności obywateli.

Środowisko badawcze ma stanowić kompetentne wsparcie podczas nawiązywania dialogu środowiska naukowego z administracją publiczną przy rozbudowie systemu ewaluacji jakości polityk publicznych w kraju, wyciąganiu wniosków z badań i formułowaniu oceny skutków regulacji.

Oprócz tego Środowisko badawcze umożliwi administracji publicznej dotarcie do szerokiego grona badaczy i analityków zajmujących się podejmowanymi zagadnieniami interwencji publicznej. Będzie stanowiło ponadto naturalne zaplecze Operatora zasobów informacyjnych państwa (OP-ZIP) jako:

- bank środowisk analitycznych gotowych do współpracy przy przedsięwzięciach analitycznych OP-ZIP;
- bank rozwiązań analitycznych potencjalnie użytecznych w rozwijaniu analityki państwa.

Środowisko badawcze może mieć nadany status formalny publicznego instytutu badawczego lub agencji rządowej, może również wykorzystywać doświadczenia współpracy z uczelniami z projektu „Zintegrowanej Platformy Analitycznej”, w dłuższym okresie rozszerzane o kolejne instytucje naukowo-badawcze. Finansowanie tworzonej instytucji, podobnie jak pozostałych elementów systemu integracji i udostępniania danych z rejestrów państwowych, może pochodzić ze środków wnoszonych za realizację poszczególnych badań i innych opłat od użytkowników systemu oraz z budżetu państwa. Należy podkreślić, że poprawa jakości polityk publicznych, osiągnięta dzięki oparciu ich na rzetelnych analizach, przyniesie korzyści wielokrotnie przekraczające poniesione koszty.

Podsumowanie i rekomendacje

Dane nazywane „nową ropą” są kluczowym zasobem gospodarczym. Jego umiejętne wykorzystanie daje szansę na osiągnięcie poziomu efektywności państwa niedostępnego bez tego zasobu. Szczególnie ważnym rodzajem są tu dane administracyjne. W niniejszym opracowaniu przedstawiono ich potencjał w wielu obszarach: funkcjonowaniu administracji, gospodarce, polityce społecznej, ochronie zdrowia, edukacji i nauce, biznesie oraz szeregu innych sektorów.

Efektywne wykorzystanie informacji pochodzących z rejestrów administracyjnych wymaga dwóch komponentów: metodologii dostosowanej do ich specyfiki oraz odpowiednich procedur. Z technologicznego punktu widzenia doprowadzenie do integracji zasobów jest przedsięwzięciem dość prostym. Zaprojektowano już i przetestowano metodologię badań opartych na danych administracyjnych i sprawdzono zasady bezpiecznego łączenia informacji. Kwestią kluczową jest natomiast przygotowanie procedur prawnych i organizacyjnych przedsięwzięcia. Jest to niezbędne do systemowego i bezpiecznego uruchomienia potencjału informacji przechowywanych w zasobach administracyjnych państwa. W niniejszym opracowaniu przedstawiono założenia proponowanego rozwiązania. Jego efektem będzie stworzenie w Polsce jednolitego, zintegrowanego mechanizmu analitycznego wspierającego budowę innowacyjnego państwa zasilanego rzetelnymi danymi. Pozwoli to nie tylko na znaczne oszczędności, ale również na tworzenie bardziej efektywnych rozwiązań w różnych sektorach aktywności państwa, społeczeństwa i gospodarki.

- Dane administracyjne są cennym zasobem gospodarczym kraju. Ich umiejętne wykorzystanie jest niezbędne do optymalizacji polityk publicznych, utrzymania wzrostu gospodarczego oraz budowy innowacyjnego państwa.
- Liczne międzynarodowe (i pierwsze polskie) doświadczenia wykorzystania danych administracyjnych jako systemowego wsparcia polityk publicznych oraz badań naukowych dowodzą, że rozwiązanie to może łączyć w sobie nieosiągalną w inny sposób efektywność, bezpieczeństwo i niski koszt przedsięwzięcia.
- Polska w swoich rejestrach publicznych posiada bardzo cenne zasoby informacyjne, jednak ich potencjał jako wsparcie polityk opartych na faktach wykorzystuje w niewielkim stopniu. Niezbędne wydaje się wobec tego stworzenie systemowego mechanizmu wykorzystywania zasobów informacyjnych państwa jako wsparcia dla analityki publicznej i badań naukowych.
- Systemowe rozwiązanie powinno być oparte na trzech elementach:
 1. **Operator zasobów informacyjnych państwa** (OP-ZIP) będący instytucją odpowiedzialną za całość procesów związanych z przygotowaniem danych do prowadzenia analiz – obecnie najbardziej zbliżone kompetencje posiada GUS.
 2. **„Zintegrowana Platforma Analityczna”** (ZPA) stanowiąca narzędzie do prowadzenia analiz strategicznych – obecnie tworzy się pilotażowe rozwiązania w tym zakresie w ramach projektu finansowanego z POPC.

3. **Środowisko badawcze** stanowiące zaplecze kompetencyjne analityki publicznej – warto rozważyć powołanie odrębnego instytutu bądź agencji skupiającego naukowców i analityków z różnych środowisk.
- Budowa nowych rozwiązań, które mogą czerpać z doświadczeń wypracowywanych w ramach projektu ZPA (zagadnienia informatyczne; współpraca środowiska naukowego; dobór danych pod kątem konkretnych problemów) oraz z doświadczeń projektu Ogólnopolskiego systemu monitorowania ekonomicznych losów absolwentów szkół wyższych (ELA) pod kątem podstawy prawnej, prezentacji danych, sposobu dokonywania analiz.
 - Do osiągnięcia przedstawionych celów niezbędna jest ustawa pozwalająca na systemową i bezpieczną integrację zasobów informacyjnych państwa, która stworzy ramy do prowadzenia analiz bazujących na danych z rejestrów publicznych.



Bibliografia

- Bożykowski, M., Chtóń-Domińczak, A., Jasiński, M. (2019), *Możliwości wykorzystywania informacji z systemów monitorowania ekonomicznych losów absolwentów szkół wyższych w rejestrach kwalifikacji w wybranych krajach europejskich w oparciu o dane administracyjne – ekspertyza*, Instytut Badań Edukacyjnych, Warszawa.
- Bynum, J.P.W., Bell, J.E., Cantu, R.V., Wang, Q., McDonough, C.M., Carmichael, D., Tosteson, T.D., Tosteson, A.N.A. (2016), *Second fractures among older adults in the year following hip, shoulder, or wrist fracture*, "Osteoporosis International", No. 27(7).
- Card, D., Chetty, R., Feldstein, M., Saez, E. (2010), *Expanding Access to Administrative Data for Research in the United States*, (w:) Schultze, C.L., Newlon, D.H. (Eds.), *Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas*, American Economic Association, Nashville.
- Central Asia Educational Platform (CAEP) (2019), Informacja prasowa, <https://www.caep-project.org/final-working-group-meeting-focuses-on-graduate-tracer-studies/> [dostęp: 12.07.2019].
- Chetty, R., Friedman, J.N., Saez, E., Turner, N., Treasury, U.S., Yagan, D. (2017), *Mobility Report Cards: The Role of Colleges in Intergenerational Mobility*, retrieved from The Equality of Opportunity Project website: http://www.equality-of-opportunity.org/assets/documents/coll_mrc_paper.pdf [dostęp: 15.07.2019].
- Connelly, R., Playford, C.J., Gayle, V. (2016), *The role of administrative data in the big data revolution in social science research*, "Social Science Research", No. 59.
- Deloitte (2017), *Assessing the value of TfL's open data and digital partnerships*, <http://content.tfl.gov.uk/deloitte-report-tfl-open-data.pdf> [dostęp: 12.07.2019].
- Deloitte, European Commission (2018), *Study to support the review of Directive 2003/98/EC on the re-use of public sector information*.
- Doar, R., Gibbs, L. (2017), *Unleashing the Power of Administrative Data. A Guide for Federal, State, and Local Policymakers*, American Enterprise Institute, Washington.
- Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/1024 z dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego, Dziennik Urzędowy Unii Europejskiej L 172.
- Einav, L., Levin, J. (2014), *Economics in the age of big data*, "Science", No. 346(6210), <https://doi.org/10.1126/science.1243089> [dostęp: 15.07.2019].
- Elias, P. (2015), *Administrative data*, (w:) Duża, A., Nelle, D., Stock, G., Wagner, G. (Eds.), *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences*, SCIVERO, Berlin.
- European Commission (2015), *Creating Value Through Open Data, Study on the Impact of Re-use of Public Sector Resources*, Brussels.
- Fazel, S., Wolf, A., Chang, Z., Larsson, H., Goodwin, G. M., Lichtenstein, P. (2015), *Depression and violence: a Swedish population study*, "Lancet Psychiatry", No. 2(3).

- Fazel, S., Wolf, A., Palm, C., Lichtenstein, P. (2014), *Violent crime, suicide, and premature mortality in patients with schizophrenia and related disorders: a 38-year total population study in Sweden*. "Lancet Psychiatry", No. 1(1).
- Frank, L. (2000), *When an Entire Country Is a Cohort*, "Science", No. 287(5462).
- IDC (2019), *Second Report on Facts and Figures: Dataset*, http://datalandscape.eu/sites/default/files/report/EDM_D2_4_Data_Set_20032019.xlsx [dostęp: 15.07.2019].
- Jasiński, M., Bożykowski, M., Zając, T., Styczeń, M., Izdebski, A. (2015), *Dokładniej, rzetelniej, taniej. Badania oparte na rejestrach publicznych jako szansa dla badań społecznych w Polsce*, „Studia Socjologiczne”, nr 1(216).
- King, M. (2018), *Data Sharing Governance and Management*. ADRF Network Working Group Reports, <https://texaserc.utexas.edu/wp-content/uploads/2018/11/Data-Sharing-Governance-and-Management.pdf> [dostęp: 12.07.2019].
- Marr, B. (2018), *Here's Why Data Is Not The New Oil*, <https://www.forbes.com/sites/bernard-marr/2018/03/05/heres-why-data-is-not-the-new-oil/#2d8450c3aa96> [dostęp: 12.07.2019].
- Ministerstwo Inwestycji i Rozwoju (2016), *Załącznik 6. Sposób pomiaru wskaźników rezultatu długoterminowego EFS, dla których źródłem danych jest badanie ewaluacyjne, (w:) Wytyczne w zakresie monitorowania postępu rzeczowego realizacji programów operacyjnych na lata 2014-2020*, Warszawa, https://www.funduszeuropejskie.gov.pl/media/2279/wytyczne_nr_13_220415_zal_6.pdf [dostęp: 12.07.2019].
- Mulgan, G., Straub, V., (2019), *The new ecosystem of trust* NESTA, <https://www.nesta.org.uk/blog/new-ecosystem-trust/> [dostęp: 15.07.2019].
- Open Data Institute (2015), *Open data means business: UK innovation across sectors and regions*, London, <http://theodi.org/open-data-means-business-uk-innovation-sectors-regions> [dostęp: 15.07.2019].
- Pope, T. (2018), *A picture of business owners in administrative data*, Institute for Fiscal Studies, <https://www.ifs.org.uk/uploads/Selfemployment%20slides%20Tom.pdf> [dostęp: 15.07.2019].
- Rajan, A. (2017), *Data is not the new oil*, <https://www.bbc.com/news/entertainment-arts-41559076>, [dostęp: 12.07.2019].
- Regenbogen, S.E., Cain-Nielsen., A.H., Norton, E.C., Chen, L.M., Birkmeyer, J.D., Skinner, J.S. (2017), *Costs and Consequences of Early Hospital Discharge After Major Inpatient Surgery in Older Adults*, "JAMA Surgery", No. 152(5).
- "The Economist" (2017a), *Data is giving rise to a new economy*, <https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy> [dostęp: 12.07.2019].
- "The Economist" (2017b), *The world's most valuable resource I no longer oil but data*, <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data> [dostęp: 12.07.2019].
- Thorp, J. (2012), *Big Data Is Not the New Oil*, <https://hbr.org/2012/11/data-humans-and-the-new-oil> [dostęp: 12.07.2019].
- Toonders, J. (2014), *Data Is The New Oil Of The Digital Economy*, <https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/> [dostęp: 12.07.2019].

- Ustawa z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. 2018 poz. 1668 z późn. zm.).
- Vanian, J. (2016), *Why Data Is The new Oil*, <http://fortune.com/2016/07/11/data-oil-brainstorm-tech/> [dostęp: 12.07.2019].
- Wallgren, A., Wallgren, B. (2007), *Register-based Statistics. Administrative Data for Statistical Purposes*, John Wiley & Sons, Chichester.
- Welch, H.G., Skinner, J.S., Schroeck, F.R., Zhou, W., Black, W.C. (2018), *Regional Variation of Computed Tomographic Imaging in the United States and the Risk of Nephrectomy*, "JAMA Internal Medicine", No. 178(2).
- Woollard, W. (2014), *Administrative data: problems and benefits. A perspective from the United Kingdom* (w:) Duşa, A., Nelle, D., Stock, G., Wagner, G. (Eds.), *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences*, SCIVERO, Berlin.

Polski Instytut Ekonomiczny

Polski Instytut Ekonomiczny to publiczny *think tank* gospodarczy, którego historia sięga 1928 roku. Obszary badawcze Polskiego Instytutu Ekonomicznego to przede wszystkim handel zagraniczny, makroekonomia, energetyka i gospodarka cyfrowa oraz analizy strategiczne dotyczące kluczowych obszarów życia społecznego i publicznego Polski. Instytut zajmuje się dostarczaniem analiz i ekspertyz do realizacji Strategii na Rzecz Odpowiedzialnego Rozwoju, a także popularyzacją polskich badań naukowych z zakresu nauk ekonomicznych i społecznych w kraju oraz za granicą.