

Warszawa, dnia 23 marca 2021 r.

Poz. 290

**UCHWAŁA NR 28
RADY MINISTRÓW**

z dnia 18 lutego 2021 r.

w sprawie Programu otwierania danych na lata 2021–2027

Na podstawie art. 9e ust. 3 ustawy z dnia 6 września 2001 r. o dostępie do informacji publicznej (Dz. U. z 2020 r. poz. 2176) Rada Ministrów uchwala, co następuje:

§ 1. Przyjmuje się Program otwierania danych na lata 2021–2027, zwany dalej „Programem”, stanowiący załącznik do uchwały.

§ 2. 1. Program realizują członkowie Rady Ministrów oraz Prezes Głównego Urzędu Statystycznego.

2. Minister właściwy do spraw informatyzacji koordynuje i monitoruje realizację Programu.

3. Członkowie Rady Ministrów oraz Prezes Głównego Urzędu Statystycznego, realizując Program, współpracują z ministrem właściwym do spraw informatyzacji.

§ 3. 1. Minister właściwy do spraw informatyzacji przedstawia Radzie Ministrów, w terminie do dnia 31 lipca danego roku, informację o realizacji Programu.

2. Pierwszą informację o realizacji Programu minister właściwy do spraw informatyzacji przedstawi Radzie Ministrów w terminie do dnia 30 listopada 2021 r.

§ 4. Uchwała wchodzi w życie z dniem ogłoszenia.

Prezes Rady Ministrów: *M. Morawiecki*

Załącznik do uchwały nr 28 Rady Ministrów
z dnia 18 lutego 2021 r. (poz. 290)



Program

Otwierania Danych

na lata 2021 – 2027

oprac. urząd obsługujący ministra
właściwego do spraw informatyzacji

PROGRAM OTWIERANIA DANYCH NA LATA 2021–2027

Spis treści

Spis treści	2
Wykaz użytych skrótów i pojęć	6
1. Wstęp	10
2. Krajowy i europejski kontekst strategiczny	12
Strategie europejskie.....	12
Strategie krajowe	12
3. Diagnoza	15
3.1. Obowiązujący stan prawny	15
3.2. Użyteczność portalu otwartych danych	17
3.3. Stan danych udostępnionych w portalu Dane.gov.pl	17
3.4. Przegląd prowadzonej polityki otwartych danych	20
Ewaluacja PODP	20
Europejska Strategia w zakresie danych	21
Podsumowanie ankiety diagnozującej potrzeby użytkowników portalu Dane.gov.pl..	22
Badanie Głównego Urzędu Statystycznego	23
3.5. Gospodarczy wpływ otwartych danych.....	24
3.6. Przykłady wykorzystania danych przemysłowych i komercyjnych	26
4. Cel główny Programu otwierania danych na lata 2021–2027	29
Cel Programu	29
Korzyści z realizacji Programu.....	30
5. Cele szczegółowe Programu	32
5.1. Cel pierwszy – Zwiększenie dostępności danych w portalu Dane.gov.pl	32
5.1.1. Dostosowanie otoczenia regulacyjnego do potrzeb rynku ponownego wykorzystywania	32
5.1.2. Zwiększenie dystrybucji danych w portalu Dane.gov.pl	35
5.1.3. Udostępnianie danych przez API	38
5.1.4. Zasady tworzenia danych.....	40
5.1.5. Konsultacje z interesariuszami	42
ZALECENIA.....	43
5.2. Cel drugi – Poprawa interoperacyjności i jakości danych	44
5.2.1. Poprawa jakości danych	44

5.2.2. Działania wspierające dostęp do danych przez interfejs programowania aplikacji (API).....	50
5.2.3. Interoperacyjność danych.....	51
5.2.4. Interoperacyjność wymiany danych między europejskimi, krajowymi i lokalnymi systemami otwartych danych.....	53
ZALECENIA.....	54
5.3. Cel trzeci – Wzrost wykorzystywania i wymiany danych	55
5.3.1. Wymiana danych między różnymi dostawcami danych	55
5.3.2. Likwidacja barier utrudniających dostęp i wymianę danych.....	58
5.3.3. System zachęt do stymulowania rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych.....	65
5.3.4. Bezpieczeństwo danych	67
ZALECENIA.....	70
5.4. Cel czwarty – Stymulowanie rynku ponownego wykorzystywania zasobów kultury i danych naukowych	71
5.4.1. Zwiększanie dostępności zasobów kultury	71
5.4.2. Zwiększenie dostępności zasobów nauki	71
ZALECENIE.....	73
5.5. Cel piąty – Wzmocnienie współpracy z krajowymi i zagranicznymi interesariuszami danych	74
5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	74
5.5.2. Otwarcie portalu Dane.gov.pl na nowych dostawców danych.....	78
5.5.3. Wzmocnienie współpracy z europejskimi interesariuszami danych oraz prowadzenie aktywnej polityki otwartych danych na arenie międzynarodowej.....	79
ZALECENIA.....	80
5.6. Cel szósty – Podnoszenie wiedzy i umiejętności pracowników administracji publicznej w zakresie otwierania i zarządzania danymi oraz zwiększanie świadomości społecznej na temat potencjału otwartych danych	81
5.6.1. Zwiększanie umiejętności pracowników administracji publicznej w celu poprawy jakości polityk publicznych, dostarczania danych zarządczych i wspierania w podejmowaniu decyzji.	81
5.6.2. Upowszechnianie informacji na temat otwierania i ponownego wykorzystywania danych publicznych oraz podnoszenie świadomości społecznej w zakresie korzyści płynących z otwartych danych	82
ZALECENIA.....	83
6. System koordynacji i monitorowania Programu	85
6.1. Wskaźniki określone dla Programu	85
6.2. System realizacji Programu	87
6.3. Mechanizm monitorowania realizacji Programu.....	90

7. Finansowanie	92
Załącznik nr 1. Plan działań na rzecz realizacji Programu	94
Tabela 1. Plan działań.....	94
Tabela 2. Plan dobudowy API.....	114
Załącznik nr 2. Standard techniczny	118
1. Wstęp	118
2. Formaty plików.....	120
3. Formatowanie danych	122
4. Szczegółowe wymagania dla zasobów w formacie CSV.....	125
5. Szczegółowe wymagania dla zasobów w formacie JSON	126
6. Szczegółowe wymagania dla zasobów w formacie XML.....	126
7. Szczegółowe wymagania dla zasobów na 4. (****) i 5. (*****) poziomie otwartości (RDF/XML, Turtle, N3, JSON-LD, N-tripels, RDFa)	127
8. Profil aplikacji DCAT-AP	129
9. Lokalizacja źródeł danych publicznych.....	130
10. Procedura publikacji danych w portalu Dane.gov.pl.....	130
11. Podsumowanie	132
Załącznik nr 3. Standard API	134
1. Kategorie obecnych i potencjalnych odbiorców danych wykorzystujących API.	134
2. Architektura usług sieciowych REST.....	135
3. Reguły nazewnictwa zasobów oraz parametry usług API	136
4. Opis składni oraz elementów URI.....	137
5. Opis dostępu do zasobów powiązanych za pomocą URI	137
6. Manipulowanie zasobami – filtrowanie, sortowanie, stronicowanie	137
7. Metadane.....	138
8. Mechanizm linków do nawigacji.....	139
9. Wersjonowanie	139
10. Obsługa zdarzeń HTTP.....	139
11. Uwierzytelnienie i autoryzacja	141
12. Określenie reguł wprowadzania limitów danych per IP i per API-Key	142
13. Opis standardu dokumentacji usług API.....	142
14. Odstępstwa od standardu JSON API.....	143
Załącznik nr 1 do „Standardu API”	143
Załącznik nr 2 do „Standardu API”	145
Załącznik nr 3 do „Standardu API”	147
Załącznik nr 4 do „Standardu API”	148
Załącznik nr 5 do „Standardu API”	150

Załącznik nr 6 do „Standard API” 152
Załącznik nr 7 do „Standardu API” 153
Załącznik nr 8 do „Standardu API” 154
Załącznik nr 9 do „Standardu API” 155



Wykaz użytych skrótów i pojęć

Skrót/pojęcie	Objaśnienie
AI	(ang. Artificial Intelligence) sztuczna inteligencja – szeroki obszar technologii obliczeniowych, umożliwiających zautomatyzowane rozumowanie, jej głównymi elementami są dane i algorytmy ¹⁾
API	zbiór technicznych funkcji umożliwiających połączenie i wzajemną wymianę danych lub metadanych między programami komputerowymi lub systemami teleinformatycznymi
dane badawcze	informacje sektora publicznego utrwalone w postaci elektronicznej, inne niż publikacje naukowe, które zostały wytworzone lub zgromadzone w ramach działalności naukowej w rozumieniu art. 4 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2021 r. poz. 478) i są wykorzystywane jako dowody w procesie badawczym lub służą weryfikacji poprawności ustaleń i wyników badań
dane dynamiczne	informacje sektora publicznego utrwalone w postaci elektronicznej podlegające częstym aktualizacjom lub aktualizacjom w czasie rzeczywistym, w szczególności ze względu na ich zmienność lub szybką dezaktualizację; za dane dynamiczne w szczególności uznaje się dane wygenerowane przez czujniki ²⁾
dane otwarte (open data)	informacje sektora publicznego utrwalone w postaci elektronicznej, udostępnione w otwartym formacie do ponownego wykorzystywania, spełniające filary otwartości danych
dane prywatne	każda treść lub jej część, niezależnie od sposobu utrwalenia, w szczególności w postaci papierowej, elektronicznej, dźwiękowej, wizualnej lub audiowizualnej, będąca w posiadaniu podmiotu innego niż podmiot zobowiązany i przez niego wytworzona, z wyjątkiem danych osobowych
dane publiczne	liczby i pojedyncze wydarzenia lub obiekty na możliwie najniższym poziomie agregacji, które nie zostały poddane przez administrację publiczną przetworzeniu do postaci raportów, wykresów itp. oraz nie został im nadany odpowiedni kontekst lub interpretacja
dane o wysokiej wartości	informacje sektora publicznego, których ponowne wykorzystywanie wiąże się z istotnymi korzyściami dla społeczeństwa, środowiska i gospodarki, w szczególności ze względu na ich przydatność do tworzenia produktów, usług i zastosowań opartych na wykorzystaniu tych danych ³⁾
DESI	(ang. The Digital Economy and Society Index) Indeks Gospodarki Cyfrowej i Społeczeństwa Cyfrowego
dostawca danych	podmiot udostępniający dane w portalu Dane.gov.pl
dyrektywa reuse/dyrektywa	dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/1024 z dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego

¹⁾ Projekt Polityki dla rozwoju Sztucznej Inteligencji w Polsce od roku 2020, <https://bip.kprm.gov.pl/kpr/bip-rady-ministrow/prace-legislacyjne-rm-i/prace-legislacyjne-rady/wykaz-prac-legislacyjny/r1758446092,Projekt-uchwaly-Rady-Ministrow-w-sprawie-ustanowienia-Polityki-dla-rozwoju-Sztuc.html>.

²⁾ Art. 2 pkt 8 dyrektywy reuse.

³⁾ Art. 2 pkt 10 dyrektywy reuse.

2019/1024/UE	wykorzystywania informacji sektora publicznego (Dz. Urz. UE L 172 z 26.06.2019, str. 56)
EDP	(ang. European Data Portal) Europejski Portal Danych
ESD	(ang. A European strategy for data) Europejska strategia w zakresie danych, komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów (COM (2020)66)
filary otwartości danych	wytyczne dotyczące udostępniania danych, których realizacja nadaje danym przymiot otwartości, opisane w Standardzie prawnym otwartości danych opracowanym przez Ministerstwo Cyfryzacji ⁴⁾
format odczytywany maszynowo	format pliku ustrukturyzowany tak, aby aplikacje mogły łatwo zidentyfikować, rozpoznać i pozyskać określone dane, w tym poszczególne stwierdzenia faktów, i ich wewnętrzną strukturę ⁵⁾
format otwarty	format pliku, który jest niezależny od systemu, udostępniany bez żadnych ograniczeń, które utrudniałyby ponowne wykorzystywanie
informacja sektora publicznego	każda treść lub jej część, niezależnie od sposobu utrwalenia, w szczególności w postaci papierowej, elektronicznej, dźwiękowej, wizualnej lub audiowizualnej, będąca w posiadaniu podmiotu zobowiązanego w rozumieniu upwisp ⁶⁾
IoT	(ang. Internet of Things) Internet rzeczy – sieć łącząca przewodowo lub bezprzewodowo urządzenia charakteryzujące się autonomicznym (niewymagającym zaangażowania człowieka) działaniem w zakresie pozyskiwania, udostępniania, przetwarzania danych lub wchodzenia w interakcje z otoczeniem pod wpływem tych danych ⁷⁾
KRI	rozporządzenie Rady Ministrów z dnia 12 kwietnia 2012 r. w sprawie Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych (Dz. U. z 2017 r. poz. 2247)
metadane	ustrukturyzowane informacje opisujące, tłumaczące, lokalizujące i ułatwiające odnalezienie, wykorzystanie lub zarządzanie informacjami, na które składają się w szczególności elementy, o których mowa w rozporządzeniu Rady Ministrów z dnia 12 marca 2014 r. w sprawie Centralnego Repozytorium Informacji Publicznej (Dz. U. poz. 361, z późn. zm.)
OECD	(ang. Organisation for Economic Co-operation and Development) Organizacja Współpracy Gospodarczej i Rozwoju
udip	ustawa z dnia 6 września 2001 r. o dostępie do informacji publicznej (Dz. U. z 2020 r. poz. 2176)
uksc	ustawa z dnia 5 lipca 2018 r. o krajowym systemie cyberbezpieczeństwa (Dz.U. z 2020 r. poz. 1369)
upwisp	ustawa z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego (Dz. U. z 2019 r. poz. 1446)

⁴⁾ Zniesione z dniem 6 października 2020 r. na podstawie rozporządzenia Rady Ministrów z dnia 7 października 2020 r. w sprawie zniesienia Ministerstwa Cyfryzacji (Dz. U. poz. 1730).

⁵⁾ Art. 2 pkt 13 dyrektywy reuse.

⁶⁾ Art. 2 ust. 1 upwisp.

⁷⁾ Raport Internet Rzeczy w polskiej gospodarce, Raport grupy roboczej do spraw Internetu rzeczy przy Ministerstwie Cyfryzacji 2019 r. <https://www.gov.pl/web/cyfryzacja/polska-przyszlosci-to-polska-z-internetem-rzeczy>.

POD	Program otwierania danych na lata 2021–2027
PODP	Program otwierania danych publicznych stanowiący załącznik do uchwały nr 107/2016 Rady Ministrów z dnia 20 września 2016 r. w sprawie ustanowienia Programu otwierania danych publicznych na lata 2016–2020
ponowne wykorzystywanie	wykorzystywanie przez osoby fizyczne, osoby prawne i jednostki organizacyjne nieposiadające osobowości prawnej informacji sektora publicznego w celach komercyjnych lub niekomercyjnych innych niż pierwotny cel, dla którego informacja została wytworzona ⁸⁾
portal Dane.gov.pl/ portal/ portal otwartych danych	serwis realizujący cel Centralnego Repozytorium Informacji Publicznej, wskazanego w ustawie z dnia 6 września 2001 r. o dostępie do informacji publicznej (Dz. U. z 2020 r. poz. 2176). Stanowi bezpłatny dostęp do danych z różnych kategorii, takich jak np. oświata, środowisko, budżet i finanse, umożliwiając ich ponowne wykorzystywanie
RODO	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, z późn. zm.)
SIST	System Inwentaryzacji Systemów Teleinformatycznych
WRF 2021–2027	wieloletnie ramy finansowe, zwane też perspektywą finansową, ustanawiające maksymalne roczne kwoty („pułapy”), które UE może wydać w różnych obszarach polityki („działach”) przez okres nie krótszy niż 5 lat (art. 312 Traktatu o funkcjonowaniu Unii Europejskiej)
zasób danych	1. element zbioru danych, typu: plik, strona internetowa, API; 2. informacja sektora publicznego o szczególnym znaczeniu dla rozwoju innowacyjności w państwie i rozwoju społeczeństwa informacyjnego, udostępniana w centralnym repozytorium informacji publicznej
zbiór danych	informacje sektora publicznego o szczególnym znaczeniu dla rozwoju innowacyjności w państwie i rozwoju społeczeństwa informacyjnego, które ze względu na sposób przechowywania i udostępniania pozwalają na ich ponowne wykorzystywanie w rozumieniu przepisów ustawy z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego, udostępniane w centralnym repozytorium informacji publicznej



⁸⁾ Art. 2 ust. 2 upwisp.

The background of the page is a deep blue gradient, overlaid with a complex, semi-transparent pattern of white and light blue lines and circles, resembling a circuit board or a network diagram. The word "Wstęp" is centered in a clean, white, sans-serif font.

Wstęp

1. Wstęp

Dane zmieniają nasze życie i tworzą nowe możliwości rozwoju. Dostępność ogromnej liczby danych, z których wiele pochodzi od maszyn i czujników pomiarowych, ma wpływ na nas wszystkich. W istocie mało jest obszarów naszego życia, na które nie ma jeszcze wpływu trwająca rewolucja w dziedzinie danych⁹⁾. Jak wskazuje Komisja Europejska (dalej także „KE”), liczba danych generowanych na świecie gwałtownie rośnie – z 33 zettabajtów w 2018 r. do prognozowanych 175 zettabajtów w 2025 r.¹⁰⁾. Dane są fundamentem cyfrowej transformacji. To one kształtują sposób, w jaki produkujemy, konsumujemy i żyjemy. Dostęp do stale rosnącej liczby danych oraz możliwość korzystania z nich mają pierwszorzędne znaczenie dla innowacji. Innowacyjne rozwiązania wykorzystujące potencjał danych mogą przynieść ważne, konkretne korzyści¹¹⁾. Bez wartościowych danych dostępnych w otwartych formatach umożliwiających maszynowy odczyt i przetwarzanie rozwój sztucznej inteligencji i innych innowacyjnych zastosowań cyfrowych (jak np. IoT) nie jest możliwy¹²⁾.

Sektor publiczny posiada ogromną liczbę cennych zasobów informacji w wielu obszarach, m.in. dane o mobilności, dane meteorologiczne, ekonomiczne czy finansowe, które mogą być następnie wykorzystywane do tworzenia nowych, innowacyjnych produktów i usług z korzyścią dla społeczeństwa i gospodarki. Nieustające zwiększanie liczby danych sektora publicznego dostępnych do ponownego wykorzystania, w tym zwiększenie podaży danych wartościowych, będzie miało szereg pozytywnych skutków dla innowacyjności gospodarki i jakości życia społeczeństwa.

Europejskie kraje wspierają ideę otwartej administracji/otwartości danych przez ustawodawstwo oraz tworzenie portali otwartych danych. Stopień zaawansowania inicjatyw oraz świadomość kwestii związanych z otwartymi danymi są jednak zróżnicowane w poszczególnych państwach członkowskich. Polska dąży do tego, aby wykorzystać szanse, jakie niosą otwarte dane, i nieprzerwanie stać na czele państw europejskich, w których funkcjonują dobrze ugruntowane zasady dotyczące otwartych danych i zarządzania nimi.

W Polsce od lat podejmowane są inicjatywy sprzyjające otwieraniu i wykorzystywaniu danych. Strategiczny dokument poświęcony otwieraniu danych publicznych – Program otwierania danych publicznych utracił swoją moc w grudniu 2020 r. Z uwagi na konieczność zachowania ciągłości działań i kontynuowanie realizowanej polityki otwartości danych dostosowanej do rozwoju technologii informacyjno-komunikacyjnych uchwała się niniejszy Program otwierania danych na lata 2021–2027. Dokument określa priorytety administracji rządowej w otwieraniu dostępu do danych. Zawarta w POD strategia otwartych danych ma na celu rozwój ekosystemu, w którym duża liczba dobrej jakości danych pozwoli na rozwój innowacyjnych rozwiązań, zwiększy przejrzystość działań organów administracji oraz zwiększy jakość realizowanych badań, przyspieszając postęp naukowy.

⁹⁾ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, komunikat „W kierunku wspólnej europejskiej przestrzeni danych”, COM (2018) 232.

¹⁰⁾ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Europejska strategia w zakresie danych, za IDC, 2018.

¹¹⁾ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_pl.

¹²⁾ Stanowisko Rządu do Komunikatu Komisji do Parlamentu Europejskiego i Rady: Biała Księga „Sztuczna Inteligencja - Europejskie podejście do doskonałości i zaufania”, COM (2020)65, dnia 27 kwietnia 2020 r.



Krajowy i europejski kontekst strategiczny

2. Krajowy i europejski kontekst strategiczny

Strategie europejskie

Komisja Europejska opublikowała 19 lutego 2020 r. „**Europejską strategię w zakresie danych**” (COM(2020) 66), (dalej zwaną „Strategią”). Celem Strategii jest stworzenie europejskiej przestrzeni danych, jednolitego rynku danych umożliwiającego swobodny przepływ danych w Unii Europejskiej (dalej także „UE”) i między sektorami, z korzyścią dla przedsiębiorstw, naukowców i administracji publicznej. Obywatele, przedsiębiorstwa i organizacje uzyskają możliwość podejmowania lepszych decyzji na podstawie analizy „uwolnionych” danych, które powinny być dostępne dla wszystkich niezależnie od statusu, wielkości czy innych cech. W Strategii wskazano również działania mające na celu opracowanie przestrzeni danych sektorowych w strategicznych obszarach, takich jak: produkcja, rolnictwo, zdrowie i mobilność. Zwiększenie dostępnego wolumenu danych jest postrzegane przez twórców Strategii jako szczególnie istotne z punktu widzenia nowych, opartych na danych technologiach jak AI czy IoT. Stąd wynika też powiązanie *Europejskiej strategii w zakresie danych* z Białą Księgą w sprawie sztucznej inteligencji (COM (2020)65). Oba komplementarne wobec siebie dokumenty tworzą pierwszy filar szerszej europejskiej polityki cyfrowej.

W komunikacie KE „BIAŁA KSIĘGA w sprawie sztucznej inteligencji. Europejskie podejście do doskonałości i zaufania” (COM(2020) 65) Komisja Europejska zaakcentowała poprawę dostępu do danych i zarządzania nimi jako kwestie o zasadniczym znaczeniu. Bez danych rozwój sztucznej inteligencji i innych zastosowań cyfrowych nie jest możliwy. Propagowanie odpowiedzialnych praktyk w zakresie zarządzania danymi i zgodności danych z zasadami FAIR¹³ przyczyni się do budowania zaufania i zapewnienia możliwości ponownego wykorzystywania danych.

W komunikacie KE „W kierunku wspólnej europejskiej przestrzeni danych” (COM(2018)232) Komisja Europejska przedstawiła środki, które ułatwią przedsiębiorstwom i sektorowi publicznemu dostęp do danych pochodzących z różnych źródeł, sektorów i dyscyplin w UE oraz ich ponowne wykorzystywanie. Dokument podkreśla konieczność działania w celu poprawy efektywnego wykorzystania danych w całej UE.

W komunikacie KE „Europejskie ramy interoperacyjności – strategia wdrażania” (COM(2017)134) Komisja Europejska przedstawiła zaktualizowane zalecenia dotyczące interoperacyjności danych, które powinny pomóc w zarządzaniu danymi, aby ułatwić ich publikowanie w portalach oraz ich łączenie, udostępnianie i ponowne wykorzystywanie.

Niniejszy program jest spójny z prowadzonymi działaniami Komisji Europejskiej.

Strategie krajowe

Program Zintegrowanej Informatyzacji Państwa, w kierunku interwencji 5.1. *Reorientacja administracji publicznej na usługi zorientowane wokół potrzeb obywatela*, wskazano na zasady niezbędne do uwzględnienia w praktyce inwestowania w cyfrowe usługi publiczne w Polsce, tj. otwartość i przejrzystość oraz domyślną interoperacyjność. Organy administracji publicznej powinny wymieniać się informacjami i danymi oraz zapewnić obywatelom

¹³) Zasady FAIR, zgodnie z którymi dane muszą być możliwe do znalezienia (ang. Findability), dostępne (ang. Accessibility), interoperacyjne (ang. Interoperability) i możliwe do ponownego wykorzystania (ang. Reusability), <https://www.force11.org/group/fairgroup/fairprinciples>.

i przedsiębiorcom dostęp do danych, kontrolę nad nimi i możliwość ich korekty, powinny też umożliwić użytkownikom monitorowanie procesów administracyjnych, które ich dotyczą. Ponadto cyfrowe usługi publiczne powinny być opracowywane i udostępniane w taki sposób, aby funkcjonowały bez ograniczeń w ramach jednolitego rynku i między sztywnymi strukturami organizacyjnymi, w oparciu o swobodny przepływ danych i usług cyfrowych w UE. Interoperacyjność w warstwie danych dotyczy ich wymiaru semantycznego i syntaktycznego, co ma istotny wpływ na ponowne wykorzystanie udostępnianych danych.

Strategia Sprawne i Nowoczesne Państwo 2030¹⁴⁾, w ramach celu III *Podniesienie sprawności realizacji zadań państwa poprzez wykorzystanie technologii cyfrowych* strategia przewiduje działania związane z udostępnianiem danych publicznych, ich ponownym wykorzystywaniem i promocją. Działania te umieszczono w kierunku interwencji pt. Otwarty rząd i wspieranie rozwoju społeczeństwa informacyjnego.

Strategia Produktywności 2030¹⁵⁾, jeden z głównych działów dokumentu o nazwie „Dane” jest poświęcony tematyce danych publicznych oraz prywatnych, którego celem jest szybki rozwój algorytmicznej gospodarki opartej na danych. Strategia ta podkreśla znaczenie danych, które stają się zasobem produkcyjnym tworzącym wartość.

Program rozwoju kompetencji cyfrowych¹⁶⁾, zapewniający osobom już korzystającym z technologii cyfrowych możliwość pogłębiania umiejętności w tym zakresie na każdym etapie życia, zaś osobom wykluczonym cyfrowo – pomoc w rozpoczęciu korzystania z narzędzi cyfrowych. Jednym z zadań programu jest podniesienie kompetencji cyfrowych pracowników instytucji publicznych, w tym w zakresie otwierania i wykorzystywania danych publicznych. Narzędziem zaplanowanym do wykonania zadania jest Akademia Otwartych Danych, realizowana w projekcie „Otwarte dane plus”, finansowanym w ramach poddziałania 2.3.1. Programu Operacyjnego Polska Cyfrowa na lata 2014–2020.

Polityka dla rozwoju Sztucznej Inteligencji w Polsce od roku 2020¹⁷⁾ kładzie nacisk na ekosystem AI oraz strategiczne czynniki budowania potencjału polskiej AI, czyli: dane, wiedza i kompetencje, infrastruktura oraz inwestycje i finansowanie. Strategia ta wskazuje jako jeden z priorytetów do kontynuacji – otwieranie zbiorów danych administracji publicznej oraz zachęcanie podmiotów prywatnych do dzielenia się danymi z innych źródeł, w tym z przemysłu, w formie budowania klubów rozproszonych repozytoriów danych, jak np. wirtualne składnice danych, powierników danych jak data trusts.



¹⁴⁾ Projekt uchwały Rady Ministrów w sprawie przyjęcia strategii „Sprawne i Nowoczesne Państwo 2030” został wpisany do Wykazu prac legislacyjnych i programowych Rady Ministrów pod numerem ID63.

¹⁵⁾ Projekt Strategii Produktywności 2030, <https://bip.kprm.gov.pl/kpr/bip-rady-ministrow/prace-legislacyjne-rm-i/prace-legislacyjne-rady/wykaz-prac-legislacyjny/r3762474>, Projekt uchwały Rady Ministrów w sprawie Strategii produktywności 2030.html.

¹⁶⁾ Projekt Programu Rozwoju Kompetencji Cyfrowych.

¹⁷⁾ Uchwała nr 196 Rady Ministrów z dnia 28 grudnia 2020 r. w sprawie ustanowienia „Polityki dla rozwoju Sztucznej Inteligencji w Polsce od roku 2020” (M.P. z 2021 r. poz. 23), <https://bip.kprm.gov.pl/kpr/bip-rady-ministrow/prace-legislacyjne-rm-i/prace-legislacyjne-rady/wykaz-prac-legislacyjny/r1758446092>, Projekt uchwały Rady Ministrów w sprawie ustanowienia Polityki dla rozwoju Sztuc.html.



Diagnoza

3. Diagnoza

3.1. Obowiązujący stan prawny

Ekosystem otwartych danych publicznych tworzą ramy ustawowe, dokumenty strategiczne i komplementarne wobec nich projekty współfinansowane ze środków UE w ramach Programu Operacyjnego Polska Cyfrowa.

Ramy prawne ponownego wykorzystywania informacji sektora publicznego w Polsce wyznacza obecnie ustawa z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego¹⁸⁾, dokonująca w zakresie swojej regulacji wdrożenia postanowień dyrektywy 2003/98/WE Parlamentu Europejskiego i Rady z dnia 17 listopada 2003 r. w sprawie ponownego wykorzystywania informacji sektora publicznego¹⁹⁾, zmienionej dyrektywą 2013/37/UE Parlamentu Europejskiego i Rady z dnia 26 czerwca 2013 r.²⁰⁾ Komisja Europejska w następstwie przeprowadzonego przeglądu dotychczasowych przepisów regulujących ponowne wykorzystywanie informacji sektora publicznego zainicjowała w 2018 r. prace legislacyjne²¹⁾, których efektem było przyjęcie, w miejsce obowiązującej dyrektywy 2003/98/WE Parlamentu Europejskiego i Rady z dnia 17 listopada 2003 r. w sprawie ponownego wykorzystywania informacji sektora publicznego, nowych regulacji w zakresie otwartych danych – dyrektywy Parlamentu Europejskiego i Rady 2019/1024/UE z dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego²²⁾ (dalej „dyrektywy 2019/1024/UE” lub „dyrektywy reuse”). Państwa członkowskie zostały zobowiązane do implementacji postanowień nowej dyrektywy w terminie do dnia 17 lipca 2021 r. W związku z koniecznością implementacji dyrektywy 2019/1024/UE urząd obsługujący ministra właściwego do spraw informatyzacji pracuje nad nowym aktem prawnym w obszarze otwartych danych – ustawą o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego, która zastąpi obecnie obowiązującą ustawę z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego.

W Polsce informacje publiczne są udostępniane zgodnie z ustawą z dnia 6 września 2001 r. o dostępie do informacji publicznej²³⁾. Istnieją dwa podstawowe tryby udostępniania informacji: wnioskowy i bezwnioskowy. Centralne Repozytorium Informacji Publicznej zapewnia szybki dostęp do danych publicznych, bez potrzeby składania wniosków do urzędu, bez ograniczeń terytorialnych i licencyjnych. Informacje publiczne o szczególnym znaczeniu dla rozwoju innowacyjności w państwie i rozwoju społeczeństwa informacyjnego, które ze względu na sposób przechowywania i udostępniania pozwalają na ich ponowne wykorzystywanie, są udostępniane w portalu Dane.gov.pl. Portal stanowi centralny punkt dostępu do otwartych danych publicznych w Polsce, a docelowo również danych gromadzonych i udostępnianych przez sektor prywatny. Dostęp do zasobów informacyjnych oraz metadanych, w tym komunikowanie się z centralnym repozytorium, jest możliwy przez portal Dane.gov.pl. Portal został uruchomiony w maju 2014 r., w oparciu o ustawę z dnia 6 września 2001 r. o dostępie do informacji publicznej.

¹⁸⁾ Dz. U. z 2019 r. poz. 1446, z późn. zm.

¹⁹⁾ Dz. Urz. UE L 345 z 31.12.2003, str. 90.

²⁰⁾ Dyrektywa Parlamentu Europejskiego i Rady 2013/37/UE z dnia 26 czerwca 2013 r. zmieniająca dyrektywę 2003/98/WE w sprawie ponownego wykorzystywania informacji sektora publicznego (Dz. Urz. UE L 175 z 27.06.2013, str. 1).

²¹⁾ COM/2018/234 Wniosek Dyrektywa Parlamentu Europejskiego i Rady w sprawie ponownego wykorzystywania informacji sektora publicznego (wersja przekształcona), COM (2018) 234.

²²⁾ Dz. Urz. UE L 172 z 26.06.2019, str. 56.

²³⁾ Dz. U. z 2020 r. poz. 2176.

Głównym dokumentem strategicznym określającym kierunki realizacji polityki na rzecz otwierania danych do 2020 r. jest przyjęty przez Radę Ministrów we wrześniu 2016 r. Program otwierania danych publicznych. Wśród głównych zadań wyznaczonych w Programie znalazła się poprawa jakości zbieranych i udostępnianych danych publicznych niezależnie od miejsca przechowywania, zgodnie z filarami otwartości i standardem udostępniania danych w portalu otwartych danych – Dane.gov.pl.

Poza obszarem regulacji prawnych, krajowych i unijnych ważnymi elementami polityki otwartych danych są projekty realizowane przez urząd obsługujący ministra właściwego do spraw informatyzacji, finansowane ze środków Programu Operacyjnego Polska Cyfrowa, tj. Otwarte dane: dostęp, standard, edukacja (okres realizacji 2017.06.–2020.05.) i Otwarte dane plus (okres realizacji 2019.03.–2022.05.) oraz Kronik@ – Krajowe Repozytorium Obiektów Kultury i Nauki (okres realizacji 2018.10.–2021.08.). Projekty dostarczają systemowych rozwiązań zwiększających dostępność i jakość danych publicznych oraz ich ponowne wykorzystywanie i są praktycznymi narzędziami służącymi realizacji celów Programu otwierania danych publicznych (na lata 2016–2020) (dalej także „PODP”). Ważnym działaniem w ramach projektu Otwarte dane: dostęp, standard, edukacja było stworzenie standardów otwartości danych w 4 wymiarach: regulacji prawnych, bezpieczeństwa, technicznym i API. Istotnym komponentem projektów jest sukcesywne otwieranie za pomocą interfejsu programistycznego API rejestrów publicznych o dużym potencjale gospodarczym i społecznym oraz znacznym potencjale ponownego wykorzystania.

Minimalne wytyczne w zakresie formatów danych określa ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne²⁴⁾ oraz wydane na jej podstawie rozporządzenie Rady Ministrów z dnia 12 kwietnia 2012 r. w sprawie Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych²⁵⁾.

Dynamika działań realizowanych w ramach przedstawionych dokumentów i przedsięwzięć w obszarze otwierania danych publicznych przyczynia się do stopniowego wzrostu efektywności realizacji polityki otwartych danych, jak również użyteczności portalu Dane.gov.pl. Od 2015 r. Polska poprawiła swoją pozycję w międzynarodowych rankingach dotyczących otwierania danych publicznych. W rankingu Organizacji Współpracy Gospodarczej i Rozwoju (dalej „OECD”) „Open, Useful, Reusable Government Data”, opublikowanym w raporcie Government at a Glance 2019, Polska znalazła się na 14. miejscu na 32 badane państwa – biorąc pod uwagę dostęp i otwartość danych oraz rządowe wsparcie dla możliwości ponownego wykorzystywania informacji sektora publicznego. Jest to awans o 6 miejsc w stosunku do rankingu z 2017 r.

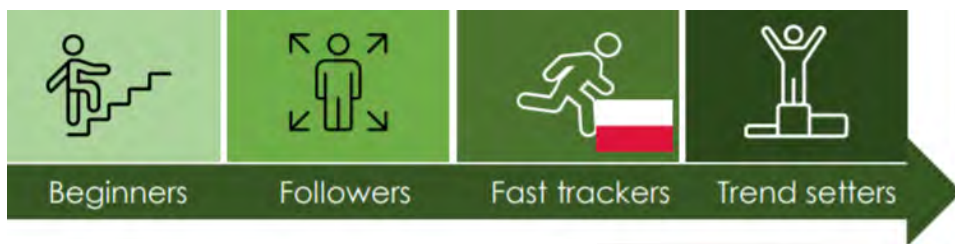
W rankingu Open Data Maturity in Europe 2019, sporządzonym na zlecenie Komisji Europejskiej, badającym poziom otwartości danych w poszczególnych krajach UE, Polska zajęła 7. miejsce. Jest to duży awans z 14. miejsca w 2018 r. i 22. miejsca w 2017 r. Należymy do tzw. grupy „fast – trackers” – szybko awansujących²⁶⁾. W czołówce krajów EU znajdują się Irlandia, Hiszpania, Francja i Cypr.

²⁴⁾ Dz. U. z 2020 r. poz. 346, z późn. zm.

²⁵⁾ Dz. U. z 2017 r. poz. 2247.

²⁶⁾ Tłumaczenie własne.

Rysunek 1. Grupy wg rankingu Open Data Maturity



Źródło: Raport Open Data Maturity 2019 r.

W 2020 r. Polska zajęła 7. miejsce również w rankingu DESI dotyczącym indeksu gospodarki cyfrowej i społeczeństwa cyfrowego, w kategorii otwartych danych. Jest to awans w stosunku do 13. miejsca uzyskanego w 2019 r. Polska wykazuje dojrzałość, jeżeli chodzi o dostępność otwartych danych, co jest wynikiem znacznie powyżej średniej dla UE.

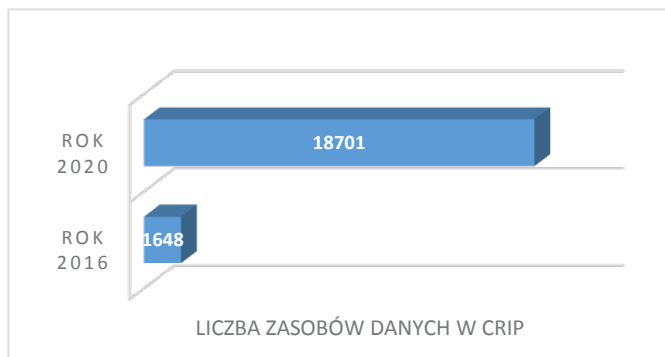
Mimo awansów w rankingach, działania Polski w obszarze otwartości danych wymagają ciągłej intensyfikacji, usystematyzowania i zbliżania się do międzynarodowych, najwyższych standardów otwartości danych.

3.2. Użyteczność portalu otwartych danych

Europejskie portale otwartych danych były wczesnym i koniecznym krokiem w rozwoju narracji otwartych danych, ale ich celem nie może być wyłącznie publikowanie danych – ostatecznie są one środkiem ułatwiającym szerokie wykorzystanie i generowanie wpływowych zmian²⁷⁾. Obecnie portal Dane.gov.pl zapewnia odwiedzającym oraz dostawcom danych wiele zaawansowanych funkcjonalności, m.in. wizualizację danych, podgląd danych, automatyzację dodawania zbiorów danych i metadanych. Funkcje portalu pomagają również dostawcom danych w śledzeniu popularności danych oraz w rozumieniu zapotrzebowania na dane. Portal z bogatymi możliwościami może przynieść wiele korzyści. Dlatego też nieustający rozwój użyteczności portalu jest niezwykle istotną kwestią. Na tle europejskich portali otwartych danych polski portal w 2019 r. uplasował się na 11. miejscu w badaniu Komisji Europejskiej „Open Data Maturity 2019”. Jednakże w celu dążenia do najwyższych standardów portal powinien ewoluować, wymaga dalszych prac, nieustającego rozwoju funkcjonalności, które będą sprzyjać ponownemu wykorzystywaniu danych.

3.3. Stan danych udostępnionych w portalu Dane.gov.pl

Wykres 1. Liczba zasobów danych w portalu Dane.gov.pl w latach 2016–2020

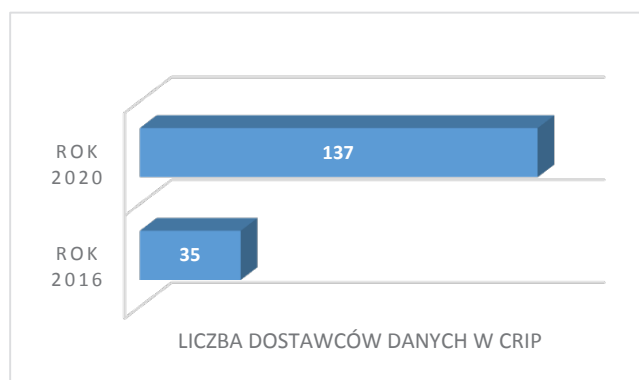


Źródło: Opracowanie własne

²⁷⁾ Analytical Report 8: The Future of Open Data Portals, EDP, Lipiec 2020 r., <https://www.europeandataportal.eu/pl/highlights/practical-guide-building-future-proof-open-data-portals>.

Wykres przedstawia liczbę zasobów danych udostępnionych w portalu Dane.gov.pl w 2016 r., w okresie przyjęcia Programu Otwierania Danych Publicznych (na lata 2016–2020) oraz liczbę zasobów danych udostępnianych w ostatnim roku obowiązywania ww. programu – w 2020 r. (stan na grudzień 2020 r.). Biorąc pod uwagę przyrost danych w latach obowiązywania Programu Otwierania Danych Publicznych, należy podkreślić, że nastąpił duży postęp w kierunku zwiększenia dostępu do danych. Jednakże, jak wskazują dane statystyczne innych krajów, dostęp do danych w Polsce jest nadal niewystarczający. Dla porównania, liczba zasobów danych upublicznionych w najwyższej ocenionych europejskich portalach otwartych danych (według rankingu Open Data Maturity 2019) wynosi: portal hiszpański – 137 385 danych, portal francuski – 193 792. Istotne jest zwiększenie wysiłków zarówno po stronie podmiotów publicznych w kierunku zwiększenia dostępu do danych publicznych, w tym danych należących do jednostek samorządu terytorialnego, jak i po stronie podmiotów prywatnych, które są często w posiadaniu danych o dużym znaczeniu dla realizacji zadań w interesie publicznym (np. w dziedzinie transportu, dostarczania i wykorzystania energii, ochrony środowiska czy edukacji). Liczba publikowanych zasobów danych z jednostek samorządu terytorialnego nieznacznie wzrosła w 5-letnim okresie – z 0 do 640 zasobów danych.

Wykres 2. Liczba dostawców danych w portalu Dane.gov.pl w latach 2016–2020



Źródło: Opracowanie własne

Na przestrzeni 5 lat czterokrotnie wzrosła liczba dostawców danych w portalu Dane.gov.pl. Liczba dostawców danych z administracji samorządowej w 2016 r. wynosiła 0, a w 2020 r. (stan na grudzień 2020 r.) – 45. Zainteresowanie podmiotów samorządowych udostępnianiem otwartych danych do ponownego wykorzystania jest znikome. Od maja 2020 r. możliwość udostępniania danych w portalu Dane.gov.pl posiadają również podmioty prywatne. Jednakże obecnie żaden z podmiotów prywatnych nie udostępnia na nim swoich danych. Z uwagi na obecną sytuację jest pożądane zintensyfikowanie współpracy z jednostkami samorządu terytorialnego oraz zainicjowanie współpracy między administracją rządową a sektorem prywatnym, organizacjami non profit, organizacjami branżowymi i pozarządowymi oraz środowiskami naukowymi.

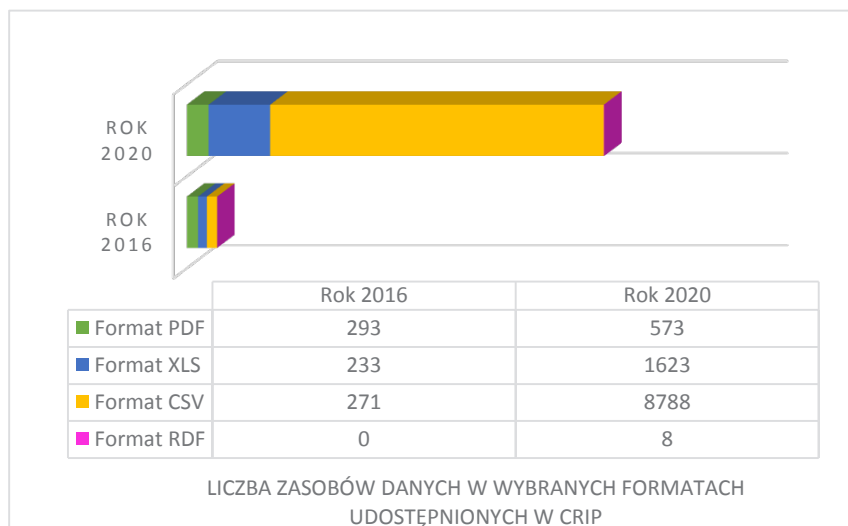
Wykres 3. Liczba zasobów danych udostępnianych w portalu Dane.gov.pl przez API w latach 2016–2020



Źródło: Opracowanie własne

Analizując informacje przedstawione na wykresie, można zauważyć duży wzrost danych publikowanych przy wykorzystaniu zaawansowanych technicznie narzędzi, jakim jest API – wzrost z 4 API do 305 API (stan na grudzień 2020 r.). Mimo to, na tle innych państw członkowskich UE, stan ten nie jest zadowalający. Irlandzki portal publikuje 509 zasobów danych z wykorzystaniem interfejsu programistycznego aplikacji, co jest znacznie większą liczbą w stosunku do danych udostępnianych przez API w portalu Dane.gov.pl. Budowanie przez organy sektora publicznego interfejsów programowania aplikacji do baz/systemów jest niewystarczające i nie jest powszechne. Powoduje to nieoptymalne wykorzystanie danych dynamicznych z sektora publicznego do celów tworzenia usług o wartości dodanej. Wspieranie wykorzystania interfejsów API na rzecz prostszego i bardziej zautomatyzowanego dostępu do zbiorów danych przyczyni się do większego ponownego ich wykorzystywania.

Wykres 4. Liczba wybranych formatów zasobów danych udostępnianych w portalu Dane.gov.pl w latach 2016–2020



Źródło: Opracowanie własne

Jak pokazują statystyki na powyższym wykresie, liczba zasobów danych udostępniona w formacie o niskim stopniu otwartości jest wciąż nadmiernie stosowana w portalu Dane.gov.pl, w szczególności format PDF – 573 zasobów danych (stan na grudzień 2020 r.). Porównanie

popularności poszczególnych formatów wskazuje format CSV jako najczęściej używany przez dostawców danych – 8788 zasobów danych (stan na grudzień 2020 r.). W zderzeniu z innymi portalami UE, liczba zasobów danych w dobrze ustrukturyzowanych plikach, takich jak CSV na hiszpańskim portalu, wynosi ok. 11 427. W tym kontekście warto nadmienić, że irlandzki portal otwartych danych publikuje dane wyłącznie od 3 stopnia otwartości (czyli CSV, JSON, XML itp.). Ponadto małą popularnością w portalu Dane.gov.pl cieszą się formaty danych w najbardziej zaawansowanej formie, na najwyższych stopniach otwartości, jak np. RDF – 8 zasobów danych w CRIP.

3.4. Przegląd prowadzonej polityki otwartych danych

Ewaluacja PODP

W 2019 r. na zlecenie Ministra Cyfryzacji przeprowadzono badanie jakościowe wśród realizatorów i beneficjentów Programu otwierania danych publicznych (na lata 2016–2020) w celu oceny jakości realizacji oraz dotychczasowych jego efektów. Badaniem objęci byli pełnomocnicy do spraw otwartości danych, przedsiębiorcy, przedstawiciele organizacji pozarządowych i środowiska naukowego. Analiza zgromadzonego materiału badawczego pozwala stwierdzić, że założone cele Programu zostały osiągnięte – zarówno cel główny, jak i cele szczegółowe. W czasie trwania Programu zauważalne było zarówno zwiększenie liczby dostępnych danych, jak i poprawa ich jakości. Jednocześnie badanie ewaluacyjne zdiagnozowało zbyt niską wiedzę na temat otwartych danych wśród ich potencjalnych odbiorców, a także zbyt niską świadomość o korzyściach z otwierania danych wśród urzędników, co utrudnia i przedłuża proces upubliczniania kolejnych zbiorów danych. Diagnoza pozwoliła też na wskazanie tych rozwiązań POPD, które dowiodły swojej skuteczności i powinny być kontynuowane:

- a) szkolenia, w ramach których pełnomocnicy do spraw otwartości danych oraz inni pracownicy ministerstw i podległych jednostek, pracujący z danymi, mogli zyskać i poszerzyć swoją wiedzę i umiejętności techniczne oraz prawne, związane z otwieraniem danych publicznych. Szkolenia cieszą się bardzo dużym zainteresowaniem i są pozytywnie oceniane przez uczestników, zarówno w zakresie ich użyteczności, jak i trafności,
- b) sposób koordynacji PODP przez urząd obsługujący ministra właściwego do spraw informatyzacji – zarówno organizacji Programu, jak i działań podejmowanych w celu jego realizacji.

Drugą kategorią rozwiązań wyodrębnioną w wyniku ewaluacji POPD są te działania, które wymagają szczególnej uwagi:

- a) jakość, kompletność oraz aktualność zasobów danych – nie wszystkie publikowane dane są odpowiedniej jakości. Zasoby danych są publikowane w formatach otwartych, ale na najniższym poziomie otwartości. Kwestiami, które należy poprawić poza dążeniem do publikowania, w możliwie najwyższych poziomach otwartości, jest wciąż kompletność oraz aktualność otwieranych danych. Konieczne jest tu wsparcie urzędów centralnych w kwestiach technicznych związanych z przygotowywaniem danych w formatach otwartych,
- b) promocja idei otwierania danych, która jest jednym z głównych czynników wpływających na skuteczność i użyteczność PODP. Działania promocyjne powinny koncentrować się na podkreślaniu roli otwartych danych w rozwoju gospodarczym; powinny prezentować przykłady dobrych praktyk z krajów, które są liderami w otwieraniu danych. Promocja powinna być prowadzona głównie przez spotkania,

warsztaty, konferencje skierowane nie tyle do pełnomocników do spraw otwartości danych i ich współpracowników, ale do osób zajmujących wyższe stanowiska, podejmujących ostateczne decyzje w kwestii danych, które mogą zostać upublicznione. Ponadto istotne jest również prowadzenie kampanii informacyjno-promocyjnych, których celem będzie dotarcie do potencjalnych odbiorców danych i stałe podnoszenie poziomu świadomości dotyczącej korzyści płynących z otwartych danych, głównie wśród przedsiębiorców,

- c) zmiany w dotychczasowym zakresie pełnienia funkcji pełnomocnika do spraw otwartości danych – miejsca, roli i znaczenia w strukturze.

Raport Ministerstwa Rozwoju

Według raportu Ministerstwa Rozwoju „Nowy surowiec danych, otwarte zasoby danych dla polskiej gospodarki”²⁸⁾, mimo bardzo dużych wysiłków i rozwiązań również w prawodawstwie, czy to krajowym czy międzynarodowym, dane ciągle nie są udostępniane przez rządy w stopniu zadowalającym. W oparciu o badania „Barometru Otwartych Danych”²⁹⁾ jako główny problem wskazano, że tylko 1 na 10 zbiorów rządowych jest otwarty. Kolejnym czynnikiem do wzięcia pod uwagę jest jakość danych – dane rządowe są często niekompletne i słabej jakości. Portale danych wprawdzie już istnieją, ale proces ich aktualizacji jest ręczny. Raport również wskazuje, że w szczególności dla przedsiębiorców dużo istotniejszy jest dostęp do danych dynamicznych. Oznacza to, że z punktu widzenia przedsiębiorcy cenniejsze jest źródło, które udostępnia dane bieżące przez API, niż źródło, które udostępnia dane archiwalne w postaci pliku. Przedsiębiorcy powinni nie tylko otrzymywać dostęp do danych, ale również konserwować swoje dane przez dołączanie danych skojarzonych, jak np. koncepcja *linked data*. Ponadto raport przedstawia rekomendacje dla administracji, do których należy m.in. rozszerzenie serwisu Dane.gov.pl o katalog API i usług udostępnianych dynamicznie³⁰⁾.

Europejska Strategia w zakresie danych

Jak wskazuje Europejska Strategia w zakresie danych (dalej „ESD”), zwiększenie wymiany różnego rodzajów danych między różnymi interesariuszami jest czynnikiem niezbędnym do tworzenia nowych produktów i usług, poprawy świadczonych usług publicznych oraz budowania konkurencyjnej gospodarki i poprawy jakości życia obywateli. Kluczowym czynnikiem jest nie tylko intensyfikacja wymiany danych, ale też włączenie do tego procesu jak największej liczby różnorodnych podmiotów. Wartościowe wyniki osiąga się, analizując dane różnorodne, ale również te odnoszące się do tej samej tematyki, ale gromadzone przez kilka podmiotów. Dlatego jest konieczne wzmocnienie działań w obszarze wymiany danych na linii Administracja – Biznes (government to business data sharing, G2B) oraz Administracja – Administracja (government to government data sharing, G2G) oraz podjęcie aktywności w celu stworzenia linii współpracy na linii Biznes – Administracja (business to government data shearing B2G) oraz Biznes – Biznes (business to business data sharing, B2B).

ESD dostrzega również wartość danych naukowych, wyodrębniając ten temat jako osobną przestrzeń danych, która ma na celu zapewnienie dostępu do danych naukowych w celu ich ponownego wykorzystywania.

²⁸⁾ Nowy surowiec danych, otwarte zasoby danych dla polskiej gospodarki, raport Ministerstwa Rozwoju, 2019 r.

²⁹⁾ <https://opendatabarometer.org/4thedition/report/>.

³⁰⁾ Nowy surowiec danych, otwarte zasoby danych dla polskiej gospodarki, raport Ministerstwa Rozwoju, 2019 r.

Podsumowanie ankiety diagnozującej potrzeby użytkowników portalu Dane.gov.pl

W maju 2020 r. Ministerstwo Cyfryzacji w portalu Dane.gov.pl udostępniło ankietę diagnozującą potrzeby i oczekiwania społeczne w zakresie otwierania danych. Analizowane dane pochodzą z badania on-line, przeprowadzonego w dniach 22 maja – 30 września 2020 r. W okresie tym ankietę umieszczoną w portalu Dane.gov.pl wypełniło 108 użytkowników.

W zakresie celu korzystania z portalu: Najczęstszym powodem, dla którego użytkownicy sięgają po dane publiczne, jest chęć zdobycia lub pogłębienia wiedzy. Czynnikiem ten wskazało dwie piąte respondentów. Niemal o połowę mniej badanych stwierdziło, że powodem, dla którego wykorzystują dane publiczne, jest zwykła ciekawość i chęć sprawdzenia, jakie możliwości drzemą w tych danych. Ponad jedna dziesiąta użytkowników wykorzystuje dane publiczne do tworzenia aplikacji, zaś blisko jedna dziesiąta – do badań naukowych.

Na pytania dotyczące zawartości portalu: Za najbardziej przydatne dla siebie użytkownicy portalu uznają znajdujące się na nim dane dotyczące biznesu i gospodarki, zdrowia, środowiska, a także budżetu i finansów publicznych. Dla ponad jednej trzeciej respondentów najbardziej przydatnymi są dane dotyczące nauki i oświaty, a dla blisko jednej trzeciej – dane z obszaru transportu. Ponad jedna piąta badanych wskazała dane z dziedziny kultury, zaś niemal jedna piąta – dane dotyczące sportu i turystyki – dane z obszaru kultury. Podobny poziom wskazań uzyskały dane z zakresu sportu i turystyki. Wśród innych danych badani wymieniali np. dane z obszaru demografii, informatyki, energetyki, sektora budowlanego, motoryzacyjnego czy też nieruchomości.

W odpowiedzi na pytanie o formaty danych: Za najbardziej przydatne formaty danych użytkownicy uznają XLS, XLSX, CSV, PDF. Blisko jedna trzecia badanych wskazała format JSON. Ponad jedna czwarta – XML, a jedna czwarta wymieniła jako przydatne formaty: DOC, DOCX oraz JPEG, PNG. Nieco ponad jedną dziesiątą wskazań uzyskał format HTML. Co dwudziesty badany za najbardziej przydatny format uznał RDF. Podobny poziom wskazań uzyskał ODS. Inne wskazane formaty GML, SWDE, SHP, TIFF.

Na pytanie o znaczenie udostępniania danych przez interfejs API: Blisko dwie trzecie użytkowników portalu przypisuje duże znaczenie udostępnianiu danych przez interfejs API, uznając je za ważne lub bardzo ważne. Co ósmy badany jest zdania, że udostępnianie danych przez API jest mało istotne, choć jednocześnie przyznaje, że zdarza mu się pobierać dane w taki sposób. Dla jednej czwartej respondentów udostępnianie danych przez API w ogóle nie jest ważne, ponieważ nie pobierają oni danych w ten sposób.

Ankieta badała również oczekiwania użytkowników w odniesieniu do danych, których nie ma w portalu Dane.gov.pl, będących w posiadaniu instytucji publicznych. Odpowiedzi koncentrowały się przede wszystkim na następujących obszarach: *zdrowie, transport publiczny i komunikacja, środowisko, dane geodezyjno-kartograficzne, lokalizacyjne, dane społeczno-demograficzne, nauka i oświata, jak również dane będące w dyspozycji jednostek samorządu terytorialnego (dalej „JST”), badania ze statystyki publicznej z gospodarki paliwowo-materiałowej oraz górnictwa węgla kamiennego; dane o podmiotach według sekcji i działów PKD; dane z Ministerstwa Sprawiedliwości dotyczące wpływu załatwień i liczby kadry orzeczniczej w poszczególnych sądach rejonowych dla wybranych repozytoriów – dla badania efektywności załatwiania spraw.*

Oczekiwania użytkowników w odniesieniu do danych, których nie ma w portalu Dane.gov.pl, będących w posiadaniu sektora prywatnego dotyczyły m.in. danych z taksówek (na podobieństwo Nowego Jorku), hulajnog elektrycznych/skuterów do wypożyczenia.

W zakresie funkcjonowania portalu: Respondenci postulują o kwestie rozbudowywania zasobów danych oraz ich aktualizacji.

Kategorie danych, jakimi dysponuje administracja rządowa

Wśród danych administracji rządowej, w tym jednostek podległych i nadzorowanych, wyróżnia się w szczególności następujące kategorie: dane z obszaru budżetu i finansów, biznesu i gospodarki, bezpieczeństwa, administracji publicznej, kultury, nauki i oświaty, pracy i pomocy społecznej, miast i regionów, rolnictwa, społeczeństwa, sportu i turystyki, środowiska oraz zdrowia. Wśród tych danych szczególną kategorię stanowią dane o wysokiej wartości, do których to zbiorów należą: dane geoprzestrzenne, dane dotyczące obserwacji Ziemi i środowiska, dane meteorologiczne, dane statystyczne, dane dotyczące przedsiębiorstw i ich własności oraz dane dotyczące mobilności.

Badanie Głównego Urzędu Statystycznego

Jak pokazuje badanie GUS³¹⁾, w roku 2018 firmy, które wykorzystywały dane publiczne stanowiły 1/6 ogółu przedsiębiorstw w Polsce (16,9%). Ze względu na wielkość, najczęściej użytkownikami danych publicznych były firmy duże (38,4%), rzadziej średnie (25,6%), a zdecydowanie najmniejszą grupą były firmy małe (14,3%). W porównaniu do roku 2017 odsetek podmiotów ogółem korzystających z danych publicznych wzrósł jedynie nieznacznie – o 0,5 punktu procentowego (p.p.), przy czym wzrost ten dotyczył przede wszystkim firm dużych (+7,6 p.p.). W przypadku firm średnich jest on nieznaczny (+2,1 p.p.), zaś wśród podmiotów małych notujemy nawet minimalny spadek (-0,1 p.p.).

Warto dodać, że 2,3% firm podkreśla, że są takie dane publiczne, które mogłyby być wykorzystane w celach biznesowych, lecz wciąż nie zostały udostępnione w ramach otwartych danych publicznych. Największy popyt na dane, które nie zostały jeszcze otwarte, obserwujemy w firmach dużych (5,9%).

Analiza stopnia wykorzystywania danych publicznych ze względu na branże pokazuje, że najczęściej z danych tych korzystały firmy działające w obszarach: wytwarzania i zaopatrywania w energię elektryczną, gaz, parę wodną i gorącą wodę (40,6%), informacji i komunikacji (32,4%) oraz dostaw wody, gospodarowania ściekami i odpadami (29,1%). W dalszej kolejności są to firmy prowadzące działalność w obszarze nauki i techniki (23,1%), naprawy i konserwacji komputerów i sprzętu komunikacyjnego (22,2%), a także zajmujące się administrowaniem i działalnością wspierającą (19,8%) oraz obsługą rynku nieruchomości (19,7%). Korzystanie z otwartych danych publicznych jest najmniej popularne wśród podmiotów działających w branży zakwaterowania i gastronomii (9,0%).

Warto odnotować, że branże, w których obserwujemy obecnie najwyższy wskaźnik wykorzystywania otwartych danych publicznych – czyli wytwarzanie i zaopatrywanie w energię elektryczną, gaz, parę wodną i gorącą wodę oraz informacja i komunikacja – są jednocześnie branżami, w których notujemy największą dynamikę wzrostu korzystania z tych danych. W przypadku pierwszej z nich notujemy wzrost o 10,2 p.p. w stosunku do roku 2017, natomiast w przypadku drugiej jest to wzrost o 8,8 p.p. w tym samym okresie.

Największym zainteresowaniem wśród przedsiębiorstw cieszą się otwarte dane publiczne z zakresu gospodarki i finansów (15,1%). Wyraźnie rzadziej firmy korzystają z danych przestrzennych (5,9%), transportowych (5,8%) oraz na temat środowiska (5,8%).

³¹⁾ Społeczeństwo informacyjne w Polsce. Wyniki badań statystycznych z lat 2015–2019, GUS, Warszawa, Szczecin 2019 r., str. 133–135.

Zdecydowanie najmniejszą popularnością cieszą się dane na temat kultury, sportu i rekreacji (wykorzystuje je 2,3% firm).

Analiza wyników badania ze względu na województwa pokazuje, że wskaźnik wykorzystywania otwartych danych publicznych przez przedsiębiorstwa jest wyższy niż średnia krajowa w sześciu województwach: dolnośląskim, kujawsko-pomorskim, lubuskim, małopolskim, mazowieckim i podlaskim. Liderem jest tu województwo mazowieckie, w którym z otwartych danych publicznych korzysta jedna piąta firm.

3.5. Gospodarczy wpływ otwartych danych

Kontekst europejski

Wartość rynku otwartych danych



- Wartość rynku otwartych danych 184,45 mld euro w 2019 r.
- Wartość rynku otwartych danych 199,51–334,20 mld euro – prognoza na 2025 r.³²⁾

Źródło: Raport Europejskiego Portalu Otwartych Danych

Zatrudnienie w oparciu o otwarte dane



- 1,09 mln pracowników dzięki otwartym danym w 2019 r.
- 1,12–1,97 mln pracowników otwartych danych – prognoza na 2025 r.³³⁾

Źródło: Raport Europejskiego Portalu Otwartych Danych

Potencjał otwartych danych w sektorach



- Oczekiwany wzrost o 15,7% w przypadku sektorów o dużym wpływie i dużym potencjale³⁴⁾
- Wysoki wpływ:³⁵⁾
- Wysoki potencjał:³⁶⁾



Źródło: Raport Europejskiego Portalu Otwartych Danych

³²⁾ The Economy Impact of Open Data: Opportunities for value creation in Europe. Raport Europejskiego portalu otwartych danych, 02.2020 r. <https://www.europeandataportal.eu/en/doc/economic-impact-open-data-opportunities-value-creation-europe>.

³³⁾ Tamże.

³⁴⁾ Tamże.

³⁵⁾ Administracja publiczna, sektory profesjonalne naukowe i techniczne, technologie informacyjno-komunikacyjne, transport i przechowywanie.

³⁶⁾ Rolnictwo, finanse i ubezpieczenia, zdrowie, edukacja, handel, działania w obszarze nieruchomości.

Wzrost wydajności



- *Ratowanie życia, np. uratowano 54–202 tys. istnień ludzkich dzięki szybkiej reakcji w nagłych przypadkach*
- *Oszczędność czasu, np. 27 mln godzin zaoszczędzonych w publicznym transporcie*
- *Doskonalenie usług językowych za pomocą otwartych danych, np. poprzez zwiększenie tłumaczenia maszynowego³⁷⁾*

Źródło: Raport Europejskiego Portalu Otwartych Danych

Oszczędności kosztów



- *Oszczędność kosztów pracy, np. 13,7–20 mld euro dzięki skróceniu czasu spędzanego w ruchu*
- *Oszczędność kosztów na rachunkach za energię, np. 79,6 mld euro dzięki większej produkcji energii słonecznej*
- *Oszczędność kosztów sektora publicznego, np. 1,1 mld euro dzięki niższym kosztom tłumaczenia³⁸⁾*

Źródło: Raport Europejskiego Portalu Otwartych Danych

Otwarte dane w organizacjach (instytucjach publicznych, NGO, sektorze prywatnym)



- *49% danych wykorzystywanych przez badane organizacje to dane otwarte, a 77% organizacji planuje wykorzystać więcej danych*
- *Otwarte dane mają wpływ na 46% przychodów organizacji, a 73% organizacji spodziewa się wzrostu tego wpływu*
- *70% badanych organizacji tworzy dane wewnętrznie, z czego 58% publikuje niektóre z nich jako otwarte dane³⁹⁾*

Źródło: Raport Europejskiego Portalu Otwartych Danych

Kontekst polski

Eksperci Polskiego Instytutu Ekonomicznego (PIE) – powołując się na dane firmy analitycznej International Data Corporation (IDC) z roku 2019 – szacują, że:



- *Wartość gospodarki opartej na danych w roku 2020 wyniesie w Polsce ok. 26,5 mld zł, a w ciągu kolejnych pięciu lat może ona wzrosnąć i wynosić od 33,7 mld zł do 51,2 mld zł*
- *W odniesieniu do PKB może więc nastąpić wzrost z 1,2 proc. do nawet 2 proc. PKB w 2025 r.*

Jak wskazuje raport PIE „*umiejętne udostępnianie danych administracyjnych może w istotny sposób wpłynąć na powodzenie realizacji scenariusza optymistycznego, zwiększając pulę*

³⁷⁾ Tamże, The Economy Impact of Open Data, Raport Europejskiego Portalu Otwartych Danych.

³⁸⁾ Tamże.

³⁹⁾ Tamże.

dostępnych zasobów, a tym samym przyczyniając się do zwiększenia liczby firm opierających swój biznes na danych⁴⁰⁾.

Otwieranie danych publicznych ma jednak wymiar nie tylko biznesowy ale również przynosi korzyści administracji. Z szacunków Komisji Europejskiej z 2015 r. wynika, że:



➤ W latach 2015–2020 sektor publiczny w Polsce mógł zaoszczędzić ponad 200 mln zł dzięki udostępnianiu danych będących w jego posiadaniu⁴¹⁾

Jak wskazuje raport Organizacji Narodów Zjednoczonych (ONZ)⁴²⁾, podsumowujący stan globalnej gospodarki cyfrowej w 2019 r., dane stanowią jedną z cech charakterystycznych gospodarki cyfrowej, określane są mianem nowej waluty. Dawniej wartość ekonomiczna była ściśle powiązana z wytwarzaniem dóbr i usług. Obecnie wartość tę mogą generować dane: osobowe, nieosobowe, prywatne, publiczne, komercyjne, należące do administracji, dobrowolne, wrażliwe i niewrażliwe. Liczba danych nieustannie rośnie. Stanowią one cenne źródło, które może przynieść zysk, jeżeli zostanie odpowiednio zmonetyzowane. Rozwój gospodarki opartej na danych tworzy wiele nowych możliwości. Wykorzystanie danych wpływa na wzrost produktywności przedsiębiorstw, obniżenie kosztów, sprawniejszą transformację cyfrową wszystkich sektorów rynku, a nawet może przyczynić się do rozwiązania problemów natury ekonomicznej czy społecznej. Co więcej cyfryzacja dokonuje zmiany łańcucha wartości, otwierając możliwości dla nowych obszarów biznesowych. Jednak pozytywny wpływ cyfryzacji nie jest automatyczny, a powyższe korzyści nie będą dystrybuowane po równo. Osoby prywatne, przedsiębiorstwa, a wreszcie kraje będą zmuszone dołączyć do wyścigu w gospodarce opartej na danych⁴³⁾.

3.6. Przykłady wykorzystania danych przemysłowych i komercyjnych

- ❖ silniki odrzutowe są wyposażone w tysiące czujników, które zbierają i przekazują dane, aby zapewnić ich skuteczne funkcjonowanie,
- ❖ dane przemysłowe wykorzystuje się na farmach wiatrowych, aby zmniejszać ich wpływ na krajobraz i maksymalnie wykorzystać siłę wiatru⁴⁴⁾,
- ❖ w sektorze rolnym analiza aktualnych danych dotyczących pogody lub zawartości wody w glebie może pomóc zmaksymalizować produkcję roślinną. W produkcji dane z czujników pomiarowych w czasie rzeczywistym umożliwiają przeprowadzenie z wyprzedzeniem odpowiedniej konserwacji maszyn,
- ❖ wysokorozdzielcze dane satelitarne z satelitów Sentinel programu Copernicus przyczyniają się do monitorowania, w czasie rzeczywistym, naturalnych zasobów wodnych w celu zapobiegania suszom lub zanieczyszczeniu. Takie dane przynoszą

⁴⁰⁾ Dane publiczne – nowy impuls do rozwoju Polski, POLICY PAPER 8/2019, Polski Instytut Ekonomiczny, Warszawa 2019 r., str.4, 22, <https://pie.net.pl/wp-content/uploads/2019/09/Policy8-08-08-WEB1.pdf>.

⁴¹⁾ Tamże, str. 22.

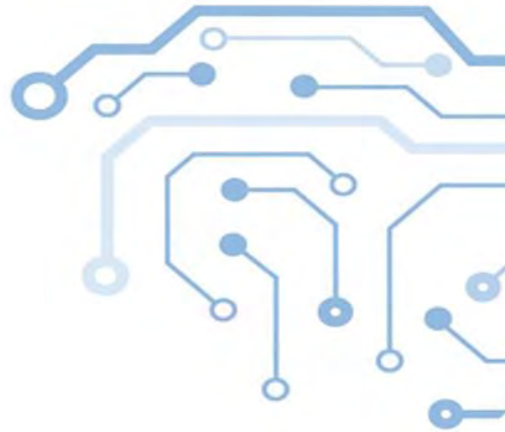
⁴²⁾ Digital Economy Report 2019, Raport ONZ <https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=2466>.

⁴³⁾ Digital Economy Report 2019, Portal CyberPolicy NASK, <https://cyberpolicy.nask.pl/digital-economy-report-2019/>.

⁴⁴⁾ https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_pl#relatedlinks.

znaczne korzyści organom publicznym, naukowcom i przedsiębiorstwom prywatnym, umożliwiając świadczenie innowacyjnych usług⁴⁵⁾.

Inteligentne wykorzystanie danych wpływa na przekształcenie wszystkich sektorów gospodarki i sektora publicznego⁴⁶⁾.



⁴⁵⁾ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, komunikat „W kierunku wspólnej europejskiej przestrzeni danych”, COM (2018) 232.

⁴⁶⁾ Tamże.



Cel główny Programu

4. Cel główny Programu otwierania danych na lata 2021–2027

Cel Programu

Celem głównym Programu otwierania danych na lata 2021–2027 (dalej także „Programu”) jest:

- ❖ zwiększanie podaży i poprawa jakości danych dostępnych w portalu Dane.gov.pl dla każdego użytkownika w celu ponownego wykorzystywania,
- ❖ stworzenie środowiska, w którym korzyści gospodarcze i społeczne wynikające z otwartych danych są dostrzegane i przekładają się na budowę nowych usług, produktów, tworzenia modeli biznesowych i nowych miejsc pracy oraz wspierają podejmowanie decyzji strategicznych,
- ❖ wejście do grupy liderów w Europie w zakresie jak najpełniejszego wykorzystania potencjału otwartych danych.

Cel główny Programu zostanie osiągnięty dzięki realizacji wymienionych niżej komplementarnych celów szczegółowych, które koncentrują się wokół kluczowych elementów polityki otwartości danych, zarówno w zakresie sektora publicznego, jak i sektora prywatnego.

Cele sformułowane w Programie wynikają z oceny dotychczasowych efektów podejmowanych działań, a także tempa zmian zachodzących w otoczeniu tematyki otwartych danych wynikających z rozwoju technologii cyfrowych oraz postępów europejskich portali otwartych danych. Program otwierania danych na lata 2021–2027 jest kontynuacją i rozszerzeniem działań podejmowanych przez administrację rządową w ramach Programu otwierania danych publicznych na lata 2016–2020 r.

Cele nakreślone w Programie będą osiągnane przez realizację szeregu działań podejmowanych przez realizatorów Programu, jednak kluczowe z nich, konieczne do przeprowadzenia w pierwszej kolejności, zostały przełożone na priorytetowe kierunki interwencji Programu.

Program jest skierowany do organów administracji rządowej oraz jednostek organizacyjnych im podległych lub przez nie nadzorowanych. Urząd obsługujący ministra właściwego do spraw informatyzacji będzie zachęcać do włączenia się w jego realizację również przez inne podmioty, które tworzą lub przechowują dane o dużym potencjale ponownego wykorzystywania, w szczególności jednostki samorządu terytorialnego oraz podmioty prywatne⁴⁷⁾. Skoordynowany plan działania, angażujący wiele organów, podmiotów, może zapewnić niezbędny zakres i skalę wymaganych działań, które kompleksowo przyczynią się do realizacji głównego celu. Beneficjentem – dzięki większej liczbie lepszej jakości danych – będzie natomiast ogół społeczeństwa, w szczególności: obywatele, biznes i administracja publiczna.

Program otwierania danych na lata 2021–2027 wprowadzony w drodze uchwały Rady Ministrów oddziałuje w sposób bezpośredni na podmioty administracji rządowej, a w sposób pośredni na pozostałe podmioty publiczne, biznes i obywateli.

Cele szczegółowe:

Cel pierwszy – Zwiększenie dostępności danych w portalu Dane.gov.pl.

Cel drugi – Poprawa interoperacyjności i jakości danych.

Cel trzeci – Wzrost wykorzystywania i wymiany danych.

⁴⁷⁾ Tematyka otwierania danych przez sektor prywatny została również uwzględniona w dokumencie: „Gospodarka oparta na danych Przemysł Plus”, <https://www.gov.pl/web/cyfryzacja/gospodarka-oparta-o-dane-przemysl->.

Cel czwarty – Stymulowanie rynku ponownego wykorzystywania zasobów kultury i danych naukowych.

Cel piąty – Współpraca z krajowymi i międzynarodowymi interesariuszami danych.

Cel szósty – Podnoszenie wiedzy i umiejętności pracowników administracji publicznej w zakresie otwierania i zarządzania danymi oraz zwiększanie świadomości społecznej na temat potencjału otwartych danych.

Załącznikiem nr 1 do Programu jest Plan działań, który uwzględnia przedsięwzięcia wskazane w tekście Programu, przypisane do konkretnego celu szczegółowego. Każdy cel szczegółowy zakończony jest zaleceniami stanowiącymi zwięzłe wytyczne do realizacji z opisanego obszaru tematycznego. Są one skierowane do realizatorów Programu oraz na zasadzie fakultatywności do pozostałych dostawców danych.

Korzyści z realizacji Programu

W efekcie przeprowadzenia zaplanowanych w Programie otwierania danych działań nastąpi wzrost podaży danych (w tym dynamicznych, o wysokiej wartości i badawczych) udostępnionych w portalu Dane.gov.pl (również przez API), gdzie w łatwy, szybki i bezpłatny sposób można je pozyskać i ponownie wykorzystywać bez konieczności składania wniosku. Nastąpi również poprawa jakości udostępnianych danych, a tym samym zwiększy się ich użyteczność dla wtórnego zastosowania. Korzyści z realizacji Programu będą odczuwalne zarówno dla przedsiębiorców (wesprze on innowacyjne firmy i przyczyni się do powstania innowacyjnych produktów, w tym na potrzeby nowych technologii), jak i obywateli (zwiększy ich udział we współdecydowaniu o sprawach publicznych oraz poprawi standard życia dzięki stworzeniu nowych usług i aplikacji wykorzystujących dane gromadzone przez podmioty publiczne).

Rysunek 2. Korzyści z realizacji Programu



Źródło: Opracowanie własne



Cele szczegółowe Programu

5. Cele szczegółowe Programu

5.1. Cel pierwszy – Zwiększenie dostępności danych w portalu Dane.gov.pl

5.1.1. Dostosowanie otoczenia regulacyjnego do potrzeb rynku ponownego wykorzystywania

Prawodawca unijny, przyjmując nową dyrektywę w sprawie otwartych danych, wskazał na konieczność zaktualizowania dotychczasowych ram ustawodawczych ponownego wykorzystywania informacji sektora publicznego w celu ich dostosowania do postępujących zmian w dziedzinie technologii cyfrowych oraz zapewnienia pełnego wykorzystania potencjału informacji sektora publicznego dla europejskiej gospodarki i europejskiego społeczeństwa. Jednocześnie zasadniczym celem zmian wprowadzonych w dyrektywie 2019/1024/UE jest przede wszystkim zwiększenie liczby danych dostępnych do ponownego wykorzystywania z korzyścią dla innowacyjności gospodarki i jakości życia społeczeństwa, aby w większym stopniu wykorzystać potencjał informacji sektora publicznego.

Ustawa o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego

W związku z koniecznością implementacji dyrektywy 2019/1024/UE jest niezbędne przygotowanie projektu nowego aktu prawnego w obszarze otwartych danych – ustawy o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego. Zastąpi on obecnie obowiązującą ustawę z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego. Regulacje ponownego wykorzystywania wymagają kompleksowego podejścia do tematyki otwierania danych w myśl wyrażonej w dyrektywie 2019/1024/UE zasady „otwarty jak to najbardziej możliwe, zamknięty jak to konieczne (*ang.* as open as possible, as closed as necessary). Główne elementy konstrukcyjne nowego aktu prawnego powinny poprawić dostępność i innowacyjne wykorzystywanie danych publicznych, a tym samym przyczynić się do rozwoju technologii wymagających dużej liczby danych, takich jak AI.

a) dane o wysokiej wartości

Szczególne znaczenie dla tworzenia innowacyjnych rozwiązań mają dane o wysokiej wartości. Są to informacje, których ponowne wykorzystywanie wiąże się z istotnymi korzyściami dla społeczeństwa, środowiska i gospodarki, w szczególności ze względu na ich przydatność do tworzenia usług i zastosowań o wartości dodanej oraz nowych godziwych miejsc pracy, a także ze względu na liczbę potencjalnych beneficjentów usług i zastosowań o wartości dodanej opartych na tych zbiorach danych. Do kategorii zbiorów danych o wysokiej wartości zalicza się:

- a) dane geoprzestrzenne (kody pocztowe, mapy krajowe i lokalne),
- b) dane dotyczące obserwacji Ziemi i środowiska (zużycie energii i obrazy satelitarne),
- c) dane meteorologiczne (dane in situ z instrumentów i prognozy pogody),
- d) dane statystyczne (wskaźniki demograficzne i gospodarcze),
- e) dane dotyczące przedsiębiorstw i ich własności (rejesty przedsiębiorstw i numery identyfikacyjne),
- f) dane dotyczące mobilności (znaki drogowe i śródlądowe drogi wodne).

Ten zakres danych jest szczególnie istotny dla ponownego wykorzystywania ze względu na ich charakter i możliwości zastosowania w projektowanych aplikacjach i usługach tworzonych zarówno przez sektor prywatny, jak i przez administrację publiczną. Konieczne jest podjęcie działań legislacyjnych, które zobowiązywałyby dostawców tych kategorii danych do ich szerszego udostępniania. Zwiększenie podaży danych o wysokiej wartości, z uwagi na ich potencjał, należy uznać z jeden z priorytetów polityki otwierania danych. Jednocześnie w celu zapewnienia odpowiednich warunków sprzyjających ponownemu wykorzystywaniu danych o wysokiej wartości ważne jest zaimplementowanie do przepisów prawa szczególnych wymogów dotyczących publikacji i ponownego wykorzystywania tego rodzaju informacji, które powinny być udostępniane:

- a) co do zasady bezpłatnie,
- b) w formacie nadającym się do odczytu maszynowego,
- c) za pośrednictwem interfejsów API, oraz
- d) w odpowiednich przypadkach do zbiorczego pobrania⁴⁸⁾.

b) dane dynamiczne

Należy zwiększyć dostępność danych dynamicznych. Są to informacje w postaci cyfrowej, podlegające częstym aktualizacjom lub aktualizacjom w czasie rzeczywistym, w szczególności ze względu na ich zmienność lub szybką dezaktualizację. Za dane, które spełniają mogą kryteria danych dynamicznych, uznaje się np. dane środowiskowe, dane o ruchu, dane satelitarne, dane meteorologiczne, a także dane wygenerowane przez czujniki. Wartość ekonomiczna tego rodzaju danych zależy od natychmiastowej dostępności informacji oraz regularnego ich aktualizowania. W związku z powyższym w celu zagwarantowania możliwości pełnego i efektywnego wykorzystania potencjału danych dynamicznych, w tym aby ułatwić rozwój bazujących na takich danych aplikacji internetowych i mobilnych oraz aplikacji korzystających z chmury obliczeniowej, jest konieczne wprowadzenie rozwiązań legislacyjnych zapewniających szczególne warunki, na jakich dane dynamiczne powinny być udostępniane do ponownego wykorzystywania, tj. zapewnienie, aby dane takie były udostępniane:

- a) niezwłocznie (natychmiast) po ich zgromadzeniu,
- b) co do zasady bezpłatnie,
- c) za pośrednictwem API, oraz
- d) w odpowiednich przypadkach do zbiorczego pobrania.

c) dane badawcze

Zakres projektowanej ustawy wymaga objęcia regulacją ponownego wykorzystywania kategorii danych badawczych, czyli informacji sektora publicznego utrwalonych w postaci elektronicznej, innych niż publikacje naukowe, które zostały wytworzone lub zgromadzone w ramach działalności naukowej w rozumieniu art. 4 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce, i wykorzystywanych jako dowody w procesie badawczym lub służących weryfikacji poprawności ustaleń i wyników badań. Dostawcy tego rodzaju danych powinni udostępniać je do ponownego wykorzystywania w celach komercyjnych lub niekomercyjnych (z uwzględnieniem uzasadnionych interesów handlowych, działań związanych z transferem wiedzy oraz istniejących już praw własności intelektualnej),

⁴⁸⁾ Przez udostępnianie danych do zbiorczego pobrania należy rozumieć zapewnienie (o ile to możliwe) przez podmiot zobowiązany funkcjonalności systemu teleinformatycznego, dzięki której użytkownicy będą mogli pobrać określony zasób danych w całości za pomocą funkcji w rodzaju „pobierz wszystko”. Definicja ujęta w uzasadnieniu do projektu ustawy o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego <https://legislacja.rcl.gov.pl/projekt/12337400/katalog/12711402#12711402>.

w zakresie, w jakim dane te są finansowane ze środków publicznych i zostały już publicznie udostępnione przez naukowców, organizacje prowadzące badania naukowe lub organizacje finansujące badania naukowe za pośrednictwem repozytorium instytucjonalnego lub tematycznego⁴⁹⁾. Potencjał tego rodzaju danych jest bardzo istotny dla ponownego wykorzystywania, dlatego jest konieczne zwiększenie ich dostępności dla użytkowników. Służą do weryfikacji innych badań, stanowią argument za prawdziwością prezentowanych tez, a także są źródłem wiedzy, m.in. dla innych ośrodków naukowych.

d) dane przedsiębiorstw publicznych

Interwencji legislacyjnej wymaga regulacja dostępu do danych będących w posiadaniu przedsiębiorstw publicznych, w szczególności tych, które prowadzą działalność w kluczowych sektorach gospodarki, np. obszarze usług pocztowych, transportu, dostaw wody, gazu, energii elektrycznej. To ważny krok, który należy wykonać na drodze do zwiększenia liczby wartościowych danych dostępnych, np. na potrzeby nowych technologii.

e) dodatkowe rozwiązania usprawniające otwieranie danych oraz ponowne wykorzystywanie informacji sektora publicznego wykraczające poza minimum wyznaczone dyrektywą 2019/1024/UE

Optymalizacja otoczenia regulacyjnego dla portalu otwartych danych

- zaktualizowanie i przeniesienie do nowej ustawy regulacji dot. portalu Dane.gov.pl. (CRIP) znajdującej się obecnie w udiip.

Ustawowe uregulowanie statusu i zadań pełnomocników do spraw otwartości danych

- uregulowanie statusu i zadań pełnomocników do spraw otwartości danych działających w ministerstwach i innych urzędach.
- umożliwienie powoływania pełnomocników w innych instytucjach⁵⁰⁾.

Wprowadzenie do ustawy „filarów otwartości” zdefiniowanych w standardach otwartości danych i PODP

Nowy katalog zasad udostępniania i przekazywania informacji sektora publicznego (ISP)

- wprowadzenie katalogu zasad obejmującego już zdefiniowane oraz nowo sformułowane zasady, takie jak np. zasada odpowiedzialności podmiotu zobowiązanego za dane.

Zasoby bibliotek, archiwów i muzeów (BAM)

- uproszczenie zasad ponownego wykorzystywania (warunki, ograniczenia) informacji sektora publicznego będących w posiadaniu BAM.

Umożliwienie stosowania licencji standardowych (np. Creative Commons) zamiast warunków ponownego wykorzystywania informacji sektora publicznego

Kod źródłowy

- wprowadzenie możliwości fakultatywnego udostępniania do ponownego wykorzystywania kodu źródłowego programu komputerowego.

Wdrożenie aktu wykonawczego dotyczącego danych o wysokiej wartości

W celu wdrożenia dyrektywy 2019/1024/UE i wsparcia ponownego wykorzystywania danych jest konieczne opracowanie wykazu szczególnych rodzajów zbiorów danych o wysokiej

⁴⁹⁾ Art. 10 ust. 2 dyrektywy 2019/1024/UE.

⁵⁰⁾ W instytucjach, które nie są realizatorami POD, a realizują zadania publiczne.

wartości. W tym celu zostanie przeprowadzona identyfikacja danych w oparciu o ocenę ich potencjału w zakresie:

- a) generowania istotnych korzyści społeczno-ekonomicznych lub środowiskowych oraz innowacyjnych usług,
- b) przydatności dla dużej liczby użytkowników, w szczególności małych i średnich przedsiębiorstw,
- c) pomagania w generowaniu dochodów,
- d) łączenia ich z innymi zbiorami danych.

Planowane stworzenie otoczenia regulacyjnego udostępniania wskazanych kategorii informacji wpłynie pozytywnie na zwiększenie dostępności wartościowych danych, które będą mogły być wykorzystywane przez użytkowników bezpłatnie do dowolnych celów, w tym również jako fundament rozwoju nowych technologii.

Działania:

1. Uchwalenie ustawy o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego wraz z aktami wykonawczymi.
2. Bieżący monitoring przepisów mających wpływ na udostępnianie danych.

5.1.2. Zwiększenie dystrybucji danych w portalu Dane.gov.pl

Portal Dane.gov.pl jest centralnym punktem dostępu do wiedzy na temat danych tworzonych i przechowywanych przez organy administracji rządowej, jednostki organizacyjne podległe lub nadzorowane przez te organy, administracji samorządowej, fundusze celowe, Zakład Ubezpieczeń Społecznych, Kasę Rolniczego Ubezpieczenia Społecznego, Narodowy Fundusz Zdrowia, państwowe instytuty badawcze, państwowe osoby prawne utworzone na podstawie odrębnych ustaw w celu wykonywania zadań publicznych. Na zasadzie dobrowolności w portalu Dane.gov.pl mogą być również udostępnione dane innych podmiotów, w tym podmiotów sektora prywatnego.

Organy sektora publicznego, w tym jednostki samorządu terytorialnego, wytwarzają i gromadzą ogromne liczby danych, które stanowią wyjątkowo cenne źródło do rozwoju innowacyjnych usług cyfrowych i lepszego kształtowania polityki. Zwiększenie dostępu do posiadanych przez administrację publiczną dużych zbiorów danych w celu ich swobodnego wykorzystywania może wspierać powstawanie pożytecznych społecznie rozwiązań, poprawić jakość życia obywateli, a zarazem przyczynić się do budowy konkurencyjności gospodarki. Dopuszczenie do ponownego wykorzystywania dokumentów będących w posiadaniu organów sektora publicznego tworzy wartość dodaną dla podmiotów ponownie wykorzystujących te dokumenty, dla użytkowników końcowych, dla ogółu społeczeństwa, a w wielu przypadkach dla samego organu sektora publicznego, wspierając przejrzystość i odpowiedzialność oraz zapewniając informację zwrotną od podmiotów ponownie wykorzystujących dokumenty i użytkowników końcowych, która pozwala zainteresowanemu

organowi sektora publicznego na poprawę jakości gromadzonych informacji oraz na realizację jego zadań⁵¹⁾.

Dobłą praktyką, już realizowaną przez niektórych dostawców danych, jest publikowanie w portalu Dane.gov.pl rejestru umów zamówień publicznych. W związku z tym rekomenduje się kontynuację stosowanej praktyki oraz włączanie się do jej realizacji innych dysponentów danych. W przypadku udostępniania ww. zasobów danych publikacja w portalu powinna być dokonywana przez poszczególnych dysponentów danych, każdy w swoim zakresie (np. jednostki podległe i nadzorowane przez ministrów publikują samodzielnie przedmiotowy rejestr). Zaleca się udostępnianie rejestru umów co najmniej od drugiego poziomu otwartości danych (szerzej o stopniach otwartości danych w rozdziale 5.2.) i jego aktualizację. Rekomenduje się, aby udostępniany rejestr zawierał co najmniej: numer umowy, nazwę wykonawcy, datę zawarcia umowy, datę zakończenia umowy, przedmiot umowy albo opis, wartość.

Wiele danych, które mogą mieć istotne znaczenie dla realizacji zadań w interesie publicznym (np. w dziedzinie transportu, dostarczania i wykorzystania energii, ochrony środowiska, edukacji itp.) i stanowią wysoką wartość dla dobra wspólnego, jest gromadzonych przez podmioty prywatne. Dane znajdujące się w posiadaniu przedsiębiorstw, takich jak operatorzy telekomunikacyjni, platformy internetowe, producenci samochodów, sprzedawcy detaliczni lub media społecznościowe, mają duże znaczenie. Ich wykorzystywanie może na przykład prowadzić do bardziej ukierunkowanej reakcji na epidemie, lepszego planowania przestrzennego miast, poprawy bezpieczeństwa ruchu drogowego i zarządzania ruchem, a także lepszej ochrony środowiska, monitorowania rynku i ochrony konsumentów⁵²⁾. Dane płynące z sektora prywatnego bezpośrednio wpływają na rozwój innowacyjnych produktów, rozwój sztucznej inteligencji oraz innych pożądaných zastosowań i narzędzi cyfrowych. Jest pożądanę, aby tego typu dane, ze względu na ich znaczenie i wysoką wartość dla społeczeństwa, podlegały udostępnianiu przy jednoczesnym zapewnieniu możliwości ich ponownego wykorzystania dla dobra ogólnego.

Wiele danych jest już obecnie wygenerowanych i przetwarzanych przez sektor publiczny i prywatny. Jak wskazuje Strategia Produktywności⁵³⁾ urzędy, ale także przedsiębiorcy, muszą być świadomi, że wytworzone, lecz nieużyte dane stanowią marnotrawstwo i wpływają na popadanie gospodarki w dług innowacyjny. Dotyczy to również danych gromadzonych przez sektor prywatny, szczególnie danych nieosobowych.

Należy zwiększyć liczbę danych (np. z obszaru zdrowia, opieki, transportu, bezpieczeństwa itp.), które mogłyby zostać udostępniane, a są szczególnie pożądanę, np. w czasie stanu epidemicznego czy innego rodzaju zagrożenia wpływającego na społeczeństwo i gospodarkę. Stanowią cenne źródło informacji, a dla przedsiębiorców w trudnym dla nich okresie spowolnienia gospodarczego inspirację dla tworzenia mobilnych aplikacji związanych np. ze zdrowiem i usług wspomagających walkę z występującym zagrożeniem, np. z COVID-19. Rzetelne udostępnianie danych oraz ich swobodny przepływ pozwoli na opracowanie mechanizmów i instrumentów przeciwdziałania rozprzestrzenianiu się określonego zagrożenia i walki z jego skutkami. Na tę kwestie zwrócono uwagę w komunikacie Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, komunikat „W kierunku wspólnej

⁵¹⁾ Dyrektywa 2019/1024/UE.

⁵²⁾ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, komunikat „W kierunku wspólnej europejskiej przestrzeni danych”, COM (2018) 232.

⁵³⁾ Projekt Strategii Produktywności 2030, s. 112.

europiejskiej przestrzeni danych”. Wskazano, że udostępnianie danych z badań naukowych na temat pojawiania się ognisk epidemii może znacznie przyspieszyć odpowiednie badania i przyczynić się do szybszego reagowania.

Dane z obszaru zdrowia, opieki, transportu, bezpieczeństwa, rolnictwa umożliwiają skuteczne działanie *ad hoc*, często w warunkach sytuacji kryzysowych. W każdej fazie zarządzania kryzysowego, tj. zapobiegania, przygotowania, reagowania i odbudowy, kluczowe znaczenie ma udostępnianie i wymiana danych. Niezwykle istotny etap to faza reagowania na zaistniałe zdarzenie lub zagrożenie. Celem tego etapu jest uruchomienie działań zapobiegających lub minimalizujących możliwość zniszczeń, a po ich wystąpieniu podjęcie akcji ratowniczej celem dostarczenia pomocy poszkodowanym i ograniczenia wtórnych szkód i strat. Tu szczególne znaczenie ma przepływ informacji i wymiana danych o istniejącej sytuacji, aby uruchomić odpowiednie procedury. Jako przykład może posłużyć stan klęski żywiołowej w postaci susz, które dotyczą rolnictwo. Niezakłócony przepływ informacji o rozmiarach strat w rolnictwie na skutek suszy wraz z informacją o miejscu jej wystąpienia daje podstawę do rzetelnej analizy i opracowania odpowiednich programów wsparcia. Dane takie mogą również stanowić podstawę do przygotowania aplikacji do monitoringu tego typu sytuacji, które będą stanowić doskonałe narzędzie⁵⁴⁾ do planowania upraw przez rolników albo narzędzie do poszukiwania pomocy w przypadku wystąpienia sytuacji mającej wpływ na realizację upraw i hodowli zwierząt. Dane są główną składową relacji między poszczególnymi poziomami zarządzania kryzysowego. Stanowią integralny element tego procesu, którego celem jest zmniejszenie stanu niepewności. Zatem należy uznać, że obieg informacji stoi u podstaw prawidłowego funkcjonowania Systemu Zarządzania Kryzysowego⁵⁵⁾. W przypadku udostępniania danych w powyższych sytuacjach należy wskazać cel możliwego wykorzystywania danych. Dane są podstawą dla podejmowania decyzji prowadzących do eliminacji lub redukcji możliwości zaistnienia sytuacji kryzysowej, a w przypadku jej wystąpienia do zapobiegania lub minimalizacji skutków i pomocy poszkodowanym. Innowacje wykorzystujące potencjał danych mogą pomóc w zarządzaniu kryzysowym⁵⁶⁾.

Istotne znaczenie dla rozwoju innowacyjnej gospodarki oraz nowoczesnego systemu ochrony zdrowia, które umożliwią tworzenie nowych rozwiązań z zakresu AI, IoT oraz innych nowych technologii, mają dane z obszaru zdrowia. To ważne, strategiczne dane, które powinny być szerzej wykorzystywane z korzyścią zarówno dla pacjentów, jak i gospodarki, uwzględniając konieczność poszanowania praw pacjenta. Obecnie, w stanie epidemii, znaczenie danych z obszaru zdrowia stało się wyraźnie zauważalne. Innowacyjne rozwiązania wykorzystujące zbiory big data pozwalają coraz skuteczniej walczyć z zagrożeniem epidemicznym.

Mając na względzie powyższe oraz dążenie do stałego wzrostu liczby zasobów danych publikowanych w portalu Dane.gov.pl, zaleca się zwiększanie liczby zasobów danych dostępnych do ponownego wykorzystywania z korzyścią dla innowacyjności gospodarki i jakości życia społeczeństwa.

⁵⁴⁾ <https://aplikacje.gov.pl/app/susza/>.

⁵⁵⁾ <http://www.nowastrategia.org.pl/obieg-informacji-w-ramach-systemu-zarządzania-kryzysowego-w-gminie-powiecie-i-województwie/>.

⁵⁶⁾ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, komunikat „W kierunku wspólnej europejskiej przestrzeni danych”.

Działania:

1. Udostępnianie przez dostawców danych publicznych (administracja rządowa) dwa razy w roku kalendarzowym przynajmniej dwóch nowych zasobów danych lub podniesienie jakości dotychczas udostępnianych danych w portalu Dane.gov.pl lub udostępnianie jednego zbioru danych przez API, na podstawie tzw. harmonogramu udostępniania danych.
2. Zwiększenie w portalu Dane.gov.pl liczby zasobów danych, szczególnie z obszaru zdrowia, opieki, transportu, bezpieczeństwa, pożądaných w sytuacjach różnego rodzaju zagrożeń, np. stanu epidemii.
3. Zachęcanie administracji samorządowej do udostępniania (co najmniej raz w roku kalendarzowym) nowych zasobów danych lub podniesienia jakości dotychczas udostępnianych danych w portalu Dane.gov.pl lub jednego zasobu danych przez API, na podstawie harmonogramu udostępniania danych (przez działania informacyjne, promocyjne i szkoleniowe, w tym np. warsztaty).
4. Prowadzenie działań zachęcających podmioty prywatne do udostępniania danych w portalu Dane.gov.pl w porozumieniu z ministrem właściwym do spraw informatyzacji (przez działania informacyjne, promocyjne, w szczególności np. konferencje, media społecznościowe, rozpowszechnianie materiałów edukacyjnych).
5. Przeprowadzenie analizy SIST pod kątem danych spełniających cechy otwartości w celu ich publikowania w portalu Dane.gov.pl
6. Poszerzenie katalogu podmiotów obligatoryjnie lub dobrowolnie publikujących dane w portalu Dane.gov.pl.

5.1.3. Udostępnianie danych przez API

Priorytetem jest zwiększenie dostępności danych publicznych przez API – interfejs programistyczny aplikacji, w tym danych o wysokiej wartości oraz danych dynamicznych. Udostępnianie danych przez API umożliwi ich dowolne wykorzystywanie w wielowymiarowych analizach, produktach, aplikacjach czy usługach i w sposób oczywisty przełoży się na zwiększenie dostępności danych publicznych i skalę ich wtórnego wykorzystania. Ponowne wykorzystywanie danych i ich wymiana przez odpowiednie wykorzystanie interfejsów API przyniesie powszechne korzyści, zapewniając otwarty, bezkosztowy i automatyczny dostęp dla każdego zainteresowanego do aktualnych danych, ułatwi twórcom oprogramowania i przedsiębiorstwom typu start-up tworzenie nowych usług i produktów, przeprowadzanie analiz na potrzeby podejmowania decyzji biznesowych i tworzenie oceny skutków regulacji. Odgrywa to również ważną rolę w tworzeniu cennych ekosystemów wokół zasobów danych, które często pozostają niewykorzystane⁵⁷⁾.

W wyniku przeprowadzonej analizy działań planowanych do wykonania w perspektywie najbliższych 7 lat zakłada się budowę interfejsów programistycznych aplikacji do 40 baz

⁵⁷⁾ Dyrektywa 2019/1024/UE.

danych/rejestrów będących w dyspozycji resortów oraz jednostek podległych lub nadzorowanych przez właściwych ministrów. Dzięki budowie API do wspomnianych rejestrów zostanie zapewniony otwarty dostęp dla każdego zainteresowanego do aktualnych danych możliwych do dalszego automatycznego przetwarzania. Spis poszczególnych baz danych/rejestrów, do których nastąpi dobudowanie API, ujęto w planie działań stanowiącym załącznik nr 1 do niniejszego Programu.

Dane o wysokiej wartości

Wprowadzenie regulacji ustalających wymóg udostępniania danych o wysokiej wartości za pośrednictwem interfejsu programistycznego aplikacji przyczyni się do pełnego wykorzystania ich potencjału oraz uzyskania maksymalnych korzyści społeczno-ekonomicznych lub środowiskowych, przy jednoczesnym zapewnieniu dostępu do tego rodzaju zasobów jak najszerszej liczbie użytkowników. Aby MŚP mogły rozszerzyć swoją działalność transgraniczną, muszą być w stanie ponownie wykorzystywać te same dane w różnych krajach w standardowych i interoperacyjnych formatach oraz z równoważnym poziomem jakości i szczegółowości. Standaryzacja danych i metadanych ponad granicami jest konieczna, aby zwiększyć interoperacyjność i umożliwić lokalne, krajowe i europejskie wykorzystywanie danych wysokiej wartości⁵⁸⁾.

Jako zasadnicze wsparcie dla działań prowadzących do udostępniania danych wysokiej wartości przez API planuje się przygotowanie przez urząd obsługujący ministra właściwego do spraw informatyzacji projektu, którego jednym z głównych zadań będzie budowa API do wybranych baz danych, należących do kategorii wysokiej wartości. Projekt ten będzie m.in. służyć jako finansowe narzędzie wsparcia budowy baz danych, które nie zostały jeszcze otwarte i udostępnione przez API, a zostały wytypowane do otwarcia jako bazy gromadzące dane o wysokiej wartości.

Dane dynamiczne

Dane dostępne powszechnie w sposób dynamiczny z różnego rodzaju interfejsów sieciowych API pozwalają na budowanie nowych, innowacyjnych usług. Ich wartość ekonomiczna zależy od natychmiastowej dostępności oraz regularnego aktualizowania takich danych w czasie rzeczywistym. W szczególności dla przedsiębiorców bardzo istotny jest dostęp do danych dynamicznych⁵⁹⁾. Zapewnianie do nich dostępu za pośrednictwem interfejsów programowania aplikacji jest szczególnie ważne, ponieważ wspiera otwarty ekosystem danych, oszczędza czas i koszty dzięki automatyzacji procesu pobierania danych, a także znacznie ułatwia ponowne wykorzystywanie danych na potrzeby szerokiej gamy nowych produktów i usług⁶⁰⁾. Sprawia to, że potencjał gospodarczy i społeczny danych dynamicznych zostanie w pełni wykorzystany.

Wdrożenie dyrektywy 2019/1024/UE nałoży na podmioty sektora publicznego obowiązek udostępniania zbiorów danych, spełniających kryteria danych dynamicznych, niezwłocznie (natychmiast) po ich zgromadzeniu, za pośrednictwem API oraz – jeżeli będzie to możliwe – do zbiorczego pobrania.

W celu pełnego i efektywnego wykorzystania potencjału danych dynamicznych, jest pożądane kreowanie polityki mającej na celu wzrost udostępniania danych dynamicznych w portalu Dane.gov.pl.

⁵⁸⁾ Analytical Report 15 high value datasets, European Data Portal.

https://www.europeandataportal.eu/sites/default/files/analytical_report_15_high_value_datasets.pdf.

⁵⁹⁾ Nowy surowiec danych, otwarte zasoby danych dla polskiej gospodarki, raport Ministerstwa Rozwoju, 2019 r.

⁶⁰⁾ Komunikat Parlamentu Europejskiego „W kierunku wspólnej europejskiej przestrzeni danych”.

Mając na względzie obowiązki nałożone dyrektywą 2019/1024/UE, konieczne jest dokonanie przez dostawców danych przeglądu danych opublikowanych w portalu pod kątem możliwości spełniania kryteriów danych dynamicznych, a następnie dokonanie ich przekształcenia – dostosowanie do warunków nałożonych dyrektywą (API, zbiorcze pobranie). Jeżeli jest to niemożliwe ze względu na ograniczenia techniczne lub finansowe, podmioty sektora publicznego powinny udostępniać (i aktualizować) tego rodzaju dokumenty w terminie umożliwiającym pełne wykorzystanie ich potencjału ekonomicznego.

Ponadto, mając na względzie wartość danych dynamicznych, zasadne jest dążenie do ciągłego podnoszenia częstotliwości aktualizacji obecnie publikowanych informacji i tym samym stopniowego przekształcania takich zasobów do miana/postaci danych dynamicznych (przy uwzględnieniu właściwych regulacji prawnych dotyczących danych). Do tego celu będzie wykorzystywane narzędzie, jakim jest harmonogram publikowania danych publicznych, o którym mowa w celu 1 pkt 2. Dostawcy danych wraz z propozycjami nowych zbiorów danych będą dokonywać dwa razy do roku przeglądu udostępnianych danych w celu ciągłego podnoszenia częstotliwości ich aktualizacji.

Działania:

1. Dobudowa API do baz danych/rejestrów będących w dyspozycji resortów oraz jednostek podległych i nadzorowanych przez właściwych ministrów, o których mowa w załączniku nr 1, w tabeli nr 2 Plan dobudowy API.
2. Udostępnianie przez API w portalu Dane.gov.pl danych spełniających kryteria danych dynamicznych oraz danych o wysokiej wartości.
3. Realizacja projektów dotyczących otwartych danych finansowanych z WRF 2021–2027.
4. Rozbudowa portalu Dane.gov.pl w zakresie udostępniania danych dynamicznych i danych o wysokiej wartości.
5. Identyfikacja potencjalnych dostawców danych dynamicznych oraz nawiązanie współpracy w celu publikacji w portalu Dane.gov.pl.
6. Zwiększanie częstotliwości aktualizacji danych opublikowanych w portalu Dane.gov.pl, uwzględniając możliwości dostawców danych.

5.1.4. Zasady tworzenia danych

Przy projektowaniu struktur instytucjonalnych, obowiązków i umiejętności, ale przede wszystkim przy budowaniu i modernizowaniu infrastruktury, należy stosować zasady „otwartości w fazie projektowania” (ang. open by design) oraz „otwartości domyślnej” (ang. open by default). Na istotność tych reguł wskazuje prawodawca unijny w art. 5 ust. 2 oraz motywie 16 dyrektywy reuse.

Zasada „otwartości w fazie projektowania” wymaga od dostawców danych uwzględnienia otwartości danych już na etapie projektowym, a następnie przez cały cykl istnienia systemu, usługi, produktu lub procesu. Oznacza to, że w każdy projekt bazujący na danych powinna być „wbudowana” ta zasada w taki sposób, aby od samego początku jego istnienia otwartość danych stanowiła jego część składową. Dla prawidłowej jej implementacji zalecane byłoby

przeanalizowanie polityk, procesów, praktyk lub strategii, ale także realizowanych projektów obowiązujących w instytucjach, które mają wpływ na otwartość danych, oraz przemyślenie, w jaki sposób niektóre produkty i usługi zostaną zaprojektowane. W konsekwencji tego procesu możliwe byłoby podjęcie decyzji o wprowadzeniu modyfikacji w realizowanych działaniach, uwzględniających zagadnienia otwartości, np. w kontekście zwiększenia liczby danych możliwych do udostępniania.

Zasada „otwartości domyślnej” nakazuje, aby wszystkie dane były otwarte, czyli dostępne bez ograniczeń terytorialnych i technologicznych, których sposób przechowywania i udostępniania umożliwi łatwe wyszukanie i swobodne wykorzystywanie do dowolnych celów, w tym również komercyjnych. Otwartość domyślna powinna stanowić integralny element każdej polityki otwartych danych. Zasada ta doznaje pewnych ograniczeń wynikających np. z zasad bezpieczeństwa, ochrony własności intelektualnej, zachowania tajemnicy statystycznej czy ochrony prywatności. Musi być to jednak ograniczenie prawnie uzasadnione, ponieważ każde wyłączenie danych z możliwości ich udostępnienia przeczy idei „otwartości domyślnej”.

Jej realizacja w przypadku organów administracji publicznej wymaga podjęcia działań polegających m.in. na wprowadzaniu klauzul umownych przesądzających o własności danych powstałych w wyniku realizacji kontraktu na usługi i inne produkty (np. raporty, wyniki badań), jakie są zawierane z przedstawicielami sektora prywatnego. Przede wszystkim dotyczy to tzw. danych surowych, czyli nieprzetworzonych do postaci raportów, wykresów, nieopatrzonej odpowiednim kontekstem czy też interpretacją. Ich wartość dla ponownego wykorzystywania jest niepodważalna i często o wiele większa niż produkt powstały na ich bazie.

Implementacja zasady „otwartości w fazie projektowania” oraz „otwartości domyślnej” wspiera tworzenie danych, które mogą zostać udostępnione do ponownego ich wykorzystywania w portalu Dane.gov.pl. Kierując się wytycznymi, jakie wynikają ze wskazanych reguł, dostawcy danych będą mogli już na bardzo wczesnym etapie realizacji swoich projektów odpowiednio zaplanować i zrealizować politykę otwartości w swojej instytucji, dbając m.in. o większą podaż danych, które udostępnione do ponownego wykorzystania pozwolą na sprawniejsze nimi zarządzanie i tworzenie wartości dodanej. Realizacja wskazanych działań powinna uwzględniać obowiązujące przepisy prawa, m.in. przepisy ustawy z dnia 5 sierpnia 2010 r. o ochronie informacji niejawnych.

Działania:

1. Przygotowanie projektu rekomendacji Komitetowi Rady Ministrów do spraw Cyfryzacji w sprawie stosowania przez organy administracji centralnej zasady „otwartości w fazie projektowania” oraz „otwartości domyślnej”.
2. Podjęcie działań zmierzających do włączenia zasady otwartości w fazie projektowania i otwartości domyślnej w opisach założeń projektów informatycznych oraz list kontrolnych osiągnięcia interoperacyjności przedkładanych do zaopiniowania Komitetowi Rady Ministrów do spraw Cyfryzacji.

5.1.5. Konsultacje z interesariuszami

W każdym przypadku otwierania danych ważne jest prowadzenie przez administrację publiczną stałych konsultacji z potencjalnymi użytkownikami w celu określenia najbardziej pożądanej grupy danych, które mają najwyższy potencjał do ponownego wykorzystywania.

Skutecznym mechanizmem monitoringu zapotrzebowania na dane jest przeprowadzanie badań analizujących trendy rynkowe, m.in. w zakresie, na jakie dane, z jakich obszarów tematycznych jest i w najbliższych latach może być największy popyt i dlaczego, a także na jakie branże oraz rozwiązania biznesowe może się to przełożyć w największym stopniu. Urząd obsługujący ministra właściwego do spraw informatyzacji zakłada przeprowadzanie cyklicznych badań rynku ponownego wykorzystywania danych publicznych w Polsce, w szczególności skali wykorzystywania danych publicznych przez przedsiębiorców oraz ich potrzeb dla budowania modeli biznesowych w oparciu o dane. Ponadto zamierzeniem jest również przeprowadzenie okresowych badań ankietowych użytkowników portalu Dane.gov.pl. Ankiety te będą przeprowadzane za pomocą portalu, minimum raz w roku. Wyniki ankiet pomogą w określeniu najbardziej pożądanej grupy danych oraz w dostosowaniu portalu do potrzeb użytkowników. Mając na względzie oszacowanie popytu na poszczególne kategorie danych publicznych wśród szerokiej grupy użytkowników, ankiety będą również przeprowadzane w jednostkach samorządu terytorialnego.

Dane mogą być wykorzystywane na różne sposoby i do różnych celów, a publikowanie otwartych danych powinno to umożliwiać. Użytkownicy mogą mieć uwagi do zbiorów danych, do ich jakości oraz mogą zaproponować nowe zbiory danych. Informacje zwrotne pozwolą ocenić, w jaki sposób zbiory danych są wykorzystywane, jak ulepszyć ich publikowanie, oraz mogą stanowić kluczowy czynnik dla procesu otwierania danych publicznych. Informacja zwrotna umożliwi korektę danych i może w ten sposób zwiększyć jakość i wartość danych. Działaniem wpisującym się w ww. założenia, wzorem jednego z najlepszych portali otwartych danych, jakim jest francuski portal, jest dodanie forum doradczego dla użytkowników portalu Dane.gov.pl. Narzędzie to umożliwi użytkownikom komentowanie i ocenianie zestawów danych oraz przedstawianie ich opinii szerokiemu gronu i wymianie pomysłów. Narzędzie to posłuży również wymianie informacji i poglądów między użytkownikami portalu oraz dostawcami danych.

Dodatkowo w ramach rozwoju portalu Dane.gov.pl system zostanie wzbogacony o nowe funkcje poprawiające użyteczność portalu i wspierające transfer wiedzy między administratorem portalu a ich użytkownikami. Utworzona zostanie funkcja pomoc/chatbox, pozwalająca na szybką i łatwą komunikację między użytkownikami portalu oraz administratorem. Dzięki tej funkcji użytkownicy będą mogli „na żywo” komentować zamieszczane zbiory danych, wymieniać poglądy oraz zgłaszać zapotrzebowanie na dane. Powyższe działania pozwolą na stały monitoring popytu na dane przez różne grupy społeczne, a w następstwie na dostosowanie działań ułatwiających udostępnianie najbardziej pożądanej grupy danych.

Działania:

1. Przeprowadzanie:
 - cyklicznych badań rynku ponownego wykorzystywania danych publicznych w Polsce, w szczególności skali wykorzystywania danych publicznych przez przedsiębiorców oraz ich potrzeb dla budowania modeli biznesowych w oparciu o dane,
 - minimum raz w roku badania ankietowego w jednostkach samorządu terytorialnego w celu identyfikacji zasobów danych, na które jest zapotrzebowanie,
 - minimum raz w roku badania ankietowego użytkowników portalu Dane.gov.pl w celu oszacowania popytu na poszczególne kategorie danych.
2. Rozbudowanie funkcjonalności portalu Dane.gov.pl usprawniającej komunikację z jego użytkownikami.

ZALECENIA

- Zwiększanie dostępności zasobów danych w portalu Dane.gov.pl przez dostawców danych na potrzeby ich ponownego wykorzystania, w tym w nowych technologiach.
- Zwiększanie liczby zasobów danych udostępnianych przez API w portalu Dane.gov.pl.
- Zwiększanie liczby zasobów danych dynamicznych udostępnianych w portalu Dane.gov.pl przez API.
- Uwzględnianie przy budowie i rozwoju systemów teleinformatycznych wymagań dla projektowania interfejsów komunikacji przez API, projektowanie rozwiązań z myślą o przyszłym udostępnianiu danych.
- Uwzględnianie zapisów w dokumentacji przetargowej na budowę, wsparcie i rozwój systemów teleinformatycznych w zakresie udostępniania danych przez interfejs programistyczny aplikacji.
- Uwzględnianie przez wszystkie zainteresowane podmioty koncepcji interoperacyjności danych w fazie projektowania systemów teleinformatycznych lub w procesie ich rozbudowy.
- Uwzględnianie zasad tworzenia danych w dokumentacji postępowań o udzielenie zamówień publicznych.



5.2. Cel drugi – Poprawa interoperacyjności i jakości danych

5.2.1. Poprawa jakości danych

Ocena jakości danych według modelu 5-Star Open

Pięciogwiazdkowy schemat otwartych danych⁶¹⁾ to model jakości danych, który umożliwia ocenianie danych na podstawie 5-stopniowej skali. Im większa liczba gwiazdek, tym więcej danych można ponownie wykorzystać i łatwiej połączyć dane. Uzyskanie 5 gwiazdek otwartości składa się z etapów⁶²⁾.

1. Stopień otwartości danych: Publikowanie danych

Stopień ten można osiągnąć, publikując swoje dane w sieci Internet, na podstawie otwartej licencji (np. typu Creative Commons).

2. Stopień otwartości danych: Udostępnianie danych ustrukturyzowanych

Aby otrzymać 2 gwiazdki, dane muszą kwalifikować się do 1 gwiazdki, a także muszą mieć strukturę i możliwość odczytu maszynowego. Dane zawarte w obrazach, zwykłym tekście, dokumentach Microsoft Word i Portable Document Files (PDF) nie kwalifikują się, ale formaty odczytywalne maszynowo, takie jak pliki Microsoft Excel, spełniają warunek 2 gwiazdki.

Formaty danych można podzielić na dwie kategorie: dane strukturalne (czytelne dla maszyn i ludzi) oraz dane nieustrukturyzowane (czytelne dla człowieka). Dane w formacie pliku ustrukturyzowanym są łatwo identyfikowalne przez aplikacje, do dalszego pozyskania⁶³⁾.

3. Stopień otwartości danych: Korzystanie z niezastrzeżonych formatów

Trzecia gwiazdka jest przyznawana, gdy dane są dostępne na licencji otwartej, w ustrukturyzowanym i powszechnie stosowanym formacie. Niezastrzeżony format oznacza, że nie jest on związany z konkretnym oprogramowaniem, a jego charakterystyka jest jawna i dobrze opisana (udokumentowana).

4. Stopień otwartości danych: Używanie URI do oznaczenia zasobów

Stopień ten można osiągnąć, spełniając 3 stopień oraz używając identyfikatorów URI⁶⁴⁾ do oznaczenia zasobów – aby możliwe było ich wyszukiwanie. Oznacza to, że dane nie tylko są odczytywane przez maszyny, ale także zawierają metadane, które pozwalają na ich zrozumienie przez maszyny. To pierwszy krok w kierunku danych połączonych⁶⁵⁾.

5. Stopień otwartości danych: Łączenie danych z innymi danymi, aby zapewnić kontekst to bardzo zaawansowany etap otwartych danych, na którym etapie dane są powiązane z innymi danymi w celu zapewnienia kontekstu. Działania te doprowadzą do interoperacyjnych i łatwych do odkrycia danych⁶⁶⁾.

W celu ułatwienia ponownego wykorzystywania dane i metadane powinny być udostępnione w portalu Dane.gov.pl w formacie przeznaczonym do odczytu maszynowego i w otwartych formatach; w możliwie najwyższej skali otwartości danych.

⁶¹⁾ Model otwartych danych pod względem jakości opracował Tim Berners-Lee, twórcy sieci WWW i prekursora Linked Data.

⁶²⁾ <https://5stardata.info/en/>.

⁶³⁾ Open data goldbook for Data Managers and Data Holders, EDP, 2018 r.

⁶⁴⁾ URI (z ang. Uniform Resource Identifier) jest unikalnym identyfikatorem zasobów, <https://www.ietf.org/rfc/rfc3986.txt>.

⁶⁵⁾ Open data goldbook for Data Managers and Data Holders.

⁶⁶⁾ Tamże.

Rysunek 3. Pięciogwiazdkowy schemat otwartych danych



Źródło: <http://5stardata.info/en/>

Oznaczenia skrótów:

OL – otwarta licencja OF – format otwarty LD – dane połączone
 RE – odczyt maszynowy URI – jednolity identyfikator zasobu.

Szczegółowe objaśnienia formatów danych wraz z ich przyporządkowaniem do odpowiedniego stopnia otwartości danych znajdują się w załączniku nr 2 do niniejszego Programu o nazwie „Standard techniczny”.

Ponadto wśród czynników wpływających na jakość danych wyróżnia się: spełnienie filarów otwartości danych⁶⁷⁾, wiarygodność, jednorodność, brak redundancji (nadmiarowości/powtórzeń), naturalny język danych (gdy ma znaczenie).

KRI wymienia rodzaje dopuszczalnych standardów danych w procesach wymiany danych między systemami teleinformatycznymi podmiotów realizujących zadania publiczne. Zgodnie z art. 18 przedmiotowego rozporządzenia systemy udostępniają zasoby informacyjne co najmniej w jednym z formatów danych określonych w załączniku nr 2 do rozporządzenia. Spośród ok. 50 formatów wymienionych w KRI jedynie nieliczne posiadają cechy umożliwiające wykorzystywanie ich w systemach udostępniania danych publicznych, np. w portalu Dane.gov.pl. Rozporządzenie nie uwzględnia formatów danych znajdujących się na najwyższych stopniach otwartości, które są fundamentem budowania otwartych danych.

W celu uwzględnienia najnowszych standardów dotyczących otwartych formatów jest pożądane znowelizowanie KRI.

Działania:

1. Publikowanie przez dostawców w portalu Dane.gov.pl nowych zasobów danych w możliwie najwyższym poziomie otwartości zgodnie z pięciogwiazdkowym schematem otwartych danych, w celach ułatwienia ponownego wykorzystywania danych oraz dalszego przetwarzania maszynowego.
2. Dokonywanie przez dostawców danych bieżących przeglądów zasobów danych opublikowanych w portalu w celu poprawy ich jakości (wzrostu poziomu otwartości) i zwiększenia ich wykorzystania w usługach i produktach.

⁶⁷⁾ Patrz rozdział 5.3.2.

3. Dostosowanie funkcjonalności portalu Dane.gov.pl do:
 - publikowania zasobów danych w wielu formatach, polegającego na udostępnianiu kilku plików danych w różnych formatach danych w jednym zasobie; lub udostępnianiu zasobu danych z materiałami współtowarzyszającymi, które mogą być istotne dla tego zasoby danych (np. zbiór danych w formacie csv, a jako materiały uzupełniające plik PDF lub XLSX),
 - konwersji zasobów danych do alternatywnych formatów o wyższych stopniach otwartości – konwersja plików CSV do formatu JSON.
4. Poszerzenie katalogu formatów danych wskazanych w KRI o formaty na 4 i 5 stopniu otwartości danych.

Zwiększenie interoperacyjności danych – dane połączone (Linked Data)

Koncepcja danych połączonych zwiększa interoperacyjność i wykrywalność zbiorów danych. Połączone dane są sposobem publikowania otwartych danych jako danych powiązanych lub wzbogacania zbiorów danych o połączone metadane. Istnieje szereg zalet tej koncepcji, m.in. służą one do:

- a) upublicznienia i ułatwienia dostępu do danych i metadanych w sieci Internet,
- b) połączenia danych i metadanych z innymi, aby: nadać im treść i znaczenie, umieścić je w kontekście, wzbogacić je o metadane, umożliwić dalsze wyszukiwanie⁶⁸⁾.

Podstawowe elementy danych połączonych to m.in. OWL, RDF, RDF Schema (inaczej RDFS), URI, SPARQL. Elementy te warunkują interoperacyjność zbiorów danych tworzonych przez różnych dostawców danych. Szczegółowe objaśnienia formatów danych na najwyższych stopniach otwartości oraz zasady projektowania danych połączonych znajdują się w załączniku nr 2 do niniejszego Programu.

Koncepcja danych połączonych wymaga zainwestowania środków finansowych, technologii, kompetencji i czasu. Mimo to założenia Linked Data są wykorzystywane przez coraz większą liczbę europejskich portali otwartych danych. Mając na względzie ww. zalety danych połączonych, zaleca się, o ile to możliwe, publikowanie danych połączonych w portalu Dane.gov.pl.

Działania:

1. Zwiększanie liczby zasobów danych o najwyższym stopniu otwartości (Linked Data) udostępnianych w portalu Dane.gov.pl (przez prowadzenie działań informacyjnych w ramach sieci pełnomocników do spraw otwartości danych oraz w ramach mediów społecznościowych, publikacji newsletterów).

⁶⁸⁾ <https://www.w3.org/DesignIssues/LinkedData.html>,
https://www.europeandataportal.eu/sites/default/files/d2.1.2_training_module_1.2_introduction_to_linked_data_pl_edp.pdf.

2. Szkolenia dostawców danych w zakresie tworzenia i publikowania danych połączonych.
3. Rozbudowa funkcjonalności portalu Dane.gov.pl o wyszukiwarkę danych połączonych SPARQL.

Zapewnienie wysokiej jakości metadanych

Jednym z głównych wyzwań w inicjatywach dotyczących otwartych danych jest ich wykrywalność. Wykrywalność otwartych danych jest związana z jakością metadanych, które nie zawsze są kompletne lub dokładne. Publikowanie metadanych dobrej jakości ma fundamentalne znaczenie dla poprawy ponownego wykorzystania, znajdowania i katalogowania danych, a także tworzenia powiązań i relacji między zbiorami danych.⁶⁹⁾

Metadane są danymi służącymi do opisu innych danych⁷⁰⁾. Definiowane są również jako informacje strukturalne, opisujące, wyjaśniające, lokalizujące lub w inny sposób ułatwiające odnalezienie i wykorzystanie informacji⁷¹⁾. Wiąże się to z tworzeniem zwięzłego i usystematyzowanego zestawu informacji odsyłającej, który służy do efektywnego wyszukiwania danych potrzebnych użytkownikowi⁷²⁾. Metadane powinny być zrozumiałe dla człowieka i możliwe do odczytu przez maszyny.

W 2011 r., stosując podobne podejście do pięciogwiazdkowego schematu Open Data Tima Bernersa-Lee, Komisja Europejska stworzyła pięciopoziomowy schemat dojrzałości do zarządzania metadanymi, aby pomóc państwom członkowskim określić kierunki poprawy ich własnych polityk zarządzania metadanymi⁷³⁾. Publikowanie metadanych na zasadzie Linked Data zwiększa wykrywalność i interoperacyjność danych. Metadane służą nie tylko do opisu i odkrywania, ale także stają się niezbędne dla zakresu kontekstualizacji – trafności, jakości, ograniczeń (warunków), a także do zespolenia użytkowników i oprogramowania do danych dostępnych w Internecie⁷⁴⁾. Metadane mają duży wpływ na możliwość ponownego wykorzystania otwartych danych. W związku z tym istotne jest zapewnienie wysokiej jakości metadanych, które przyczynią się do użyteczności danych.

Zestaw elementów metadanych zasobów udostępnianych w portalu Dane.gov.pl określony został w rozporządzeniu Rady Ministrów z dnia 12 marca 2014 r. w sprawie Centralnego Repozytorium Informacji Publicznej (dalej „rozporządzenie CRIP”)⁷⁵⁾. Część metadanych jest wpisywana do systemu automatycznie w momencie dodawania zasobu. Jednakże szczególną rolę odgrywa właściwie dobrany i bezbłędny zestaw metadanych określanych przez dostawcę, gdyż są to informacje najczęściej wykorzystywane do precyzyjnego wyszukiwania zasobów.

Mając na uwadze powyższe, zaleca się dostarczanie właściwie dobranych i bezbłędnych zestawów metadanych określanych przez dostawcę danych oraz publikowanie metadanych w otwartych formatach.

⁶⁹⁾ Sustainability of (Open) Data Portals Infrastructures reports, Report European Data Portal, 2020 r. s. 14 <https://www.europeandataportal.eu/en/highlights/sustainability-open-data-portals-infrastructures-reports-pt-1>.

⁷⁰⁾ Informatyka: Słownik encyklopedyczny, Z. Płoski, Wrocław 1999 r.

⁷¹⁾ National Information Standards Organization, Understanding Metadata, 2017 r.

⁷²⁾ Metadane w pracy statystyka, Jolanta Szutkowska, Wiadomości statystyczne 2010 r., nr 3, str. 76-91, <https://bazekon.uek.krakow.pl/en/gospodarka/164715378> [dostęp: 12.2019 r.].

⁷³⁾ Sustainability of (Open) Data Portals Infrastructures reports, Raport European Data Portal, 2020 r. str. 14.

⁷⁴⁾ Open data goldbook for Data Managers and Data Holders, EDP, 2018 r.

⁷⁵⁾ Dz. U. z 2014 r. poz. 361, z późn. zm.

Jakość metadanych jest jednym z trzech istotnych komponentów podczas procesu poszukiwania danych. Pozostałe dwa to: trafność i użyteczność. Wzorując się na praktykach najlepszych europejskich portali (francuski i irlandzki), zakłada się dodanie funkcjonalności portalu umożliwiającej dodanie przez dostawcę danych lub użytkownika portalu przykładów ponownego wykorzystania zbioru danych, tzw. użyteczności danych. Funkcjonalność ta ma na celu zwiększenie precyzyjnego wyszukiwania zasobów oraz możliwość wymiany umiejętności i wiedzy w celu szerszego wykorzystania danych.

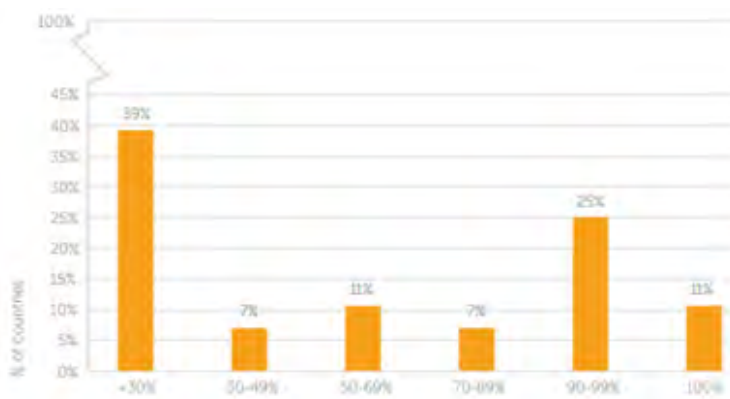
Działania:

1. Dostarczanie właściwie dobranych i kompletnych zestawów metadanych określanych przez dostawcę danych.
2. Publikowanie przez dostawców metadanych w otwartych formatach.
3. Rozbudowanie funkcjonalności portalu Dane.gov.pl polegającej na automatycznym udostępnianiu przez oprogramowanie portalu Dane.gov.pl metadanych w formacie rdf/xml. opisujących strukturę zbiorów danych.
4. Dodanie nowej funkcjonalności portalu Dane.gov.pl umożliwiającej dostawcom danych, użytkownikom portalu uzupełnienie przykładów ponownego wykorzystania zbioru danych, tzw. użyteczności danych.

Automatyzacja procesu pobierania danych i metadanych

Możliwość aktualizowania metadanych zależy między innymi od stopnia, w jakim metadane są uzyskiwane automatycznie z zewnętrznych źródeł danych. Poniższy rysunek pokazuje „odsetek” metadanych pozyskiwanych automatycznie z zewnętrznych źródeł danych:

Wykres 5. „Odsetek” metadanych pozyskiwanych automatycznie z zewnętrznych źródeł danych



Źródło: Open Data Maturity 2019

Z danych na wykresie wynika, że tylko w 3 krajach (11%) – Belgii, Danii i Szwecji wszystkie metadane są przesyłane w sposób zautomatyzowany do centralnego portalu krajowego. Ponadto 7 krajów (25%) uzyskuje od 90% do 99% metadane ze źródła automatycznego. Tylko 11 krajów (39%) uzyskują mniej niż 30% metadanych ze źródła automatycznego.

Oznacza to, że większość metadanych w krajach UE jest edytowana ręcznie. Manualna edycja metadanych jest czasochłonna i z łatwością pozwala na błędy i pomyłki⁷⁶⁾.

Krajowe portale otwartych danych często działają jako agregator mniejszych portali otwartych danych. Regularnie sprawdzają nowe dane w mniejszych portalach z otwartymi danymi i kopiuje znalezione metadane do swoich repozytoriów. Ten proces nazywa się z ang. harvesting.

Polski ustawodawca w udip przewiduje przekazywanie metadanych, w przypadku kiedy dane są umieszczane przez podmiot w zewnętrznych repozytoriach powszechnie dostępnych w sieci teleinformatycznej. W przypadku posiadania interfejsów API w repozytoriach danych dostępnych powszechnie w sieci teleinformatycznej (np. lokalne repozytoria) oraz w celu uniknięcia błędów, jak i minimalizacji czasochłonności, jest pożądane wykorzystywanie procesu automatycznego przekazania metadanych do portalu Dane.gov.pl. W przypadku pobierania metadanych z innych portali za pomocą harvestera zaleca się wysoką częstotliwość pobierania metadanych.

Z uwagi na zalety stosowania automatyzacji procesów rekomenduje się wykorzystywanie przez dostawców danych narzędzi umożliwiających automatyczne dodawane, aktualizowanie lub usuwanie zbiorów danych oraz zasobów. Proces ten składa się z dwóch elementów – importu metadanych z pliku XML oraz walidacji danych zasobów w portalu.Dane.gov.pl.

Działania:

1. Zwiększanie liczby automatycznego pobierania metadanych do portalu Dane.gov.pl z innych repozytoriów danych posiadających dobudowane API (przez działania organizacyjne, promocyjne, informacyjne, np. spotkania z pełnomocnikami, warsztaty, media społecznościowe, newsletter).
2. Zwiększanie liczby podmiotów zaangażowanych w realizację procesu automatycznego dodawania i aktualizowania zbiorów i zasobów danych w portalu Dane.gov.pl (na podstawie pliku XML) przez działania organizacyjne, promocyjne, informacyjne, np. spotkania z pełnomocnikami, warsztaty, media społecznościowe, newsletter)

Monitoring jakości metadanych

Większość z 28 krajów UE (23 kraje, 82%) monitoruje jakość metadanych dostępnych w krajowych portalach otwartych danych. Metadane często poddawane są kontroli jakości, zanim zostaną opublikowane w portalu⁷⁷⁾. Proces kontroli jakości różni się w zależności od kraju. W Polsce rozporządzenie CRIP określa minimalny zestaw elementów metadanych opisujących strukturę zasobów udostępnianych w portalu Dane.gov.pl. Część obligatoryjnych metadanych jest automatycznie dodawana do systemu w momencie publikowania zasobu. W celu zapewnienia kompletności minimalnych elementów metadanych oraz formatów danych, portal Dane.gov.pl, na podstawie § 8 ust. 1 pkt 1 rozporządzenia CRIP,

⁷⁶⁾ Open Data Maturity, Raport 2019 r.

⁷⁷⁾ Tamże.

automatycznie dokonuje ich weryfikacji. Istotne jest wzmocnienie monitoringu oraz zautomatyzowanie procesów kontroli jakości danych i metadanych udostępnianych w repozytoriach danych, w tym w portalu Dane.gov.pl.

Działania:

Utworzenie w portalu Dane.gov.pl narzędzia:

- wspomagającego kontrolę metadanych przez dostawcę, generującego automatyczne powiadomianie do dostawcy danych, np. o zbliżającej się dacie aktualizacji danych (lub o ewentualnych problemach, weryfikacji opisów przed wygaśnięciem ważności danych),
- regularnie sprawdzającego poprawność linków do danych źródłowych umieszczonych w repozytoriach zewnętrznych oraz generującego do dostawców danych informacje o konieczności aktualizacji uszkodzonych linków,
- umożliwiającego „testowanie” przez dostawców danych zbiorów danych przed ich publikacją w portalu (testowanie polegające na weryfikacji odpowiednich formatów danych zgodnych ze „Standardem technicznym” (stanowiącym załącznik nr 2 do Programu), linków zewnętrznych dla danych źródłowych, typów danych w plikach, uzyskania podglądu danych, map, wykresów).

Standard udostępniania danych oraz formatowania zasobów danych w portalu Dane.gov.pl

W celu ułatwienia ponownego wykorzystywania dane i metadane powinny być przygotowane zgodnie z załącznikiem nr 2 do niniejszego Programu pt. Standard techniczny.

Stosowanie standardu przyczyni się do tworzenia danych zgodnie z wymaganiami skali otwartości danych, jednolitych standardów oraz prawidłowego formatowania określonych typów danych.

Działanie:

Weryfikacja Standardu technicznego (dokonywana w ramach ewaluacji bieżącej (on-going)) w celu zapewnienia aktualności zalecanych norm, standardów ujętych w tym standardzie.

5.2.2. Działania wspierające dostęp do danych przez interfejs programowania aplikacji (API)

Jak wskazuje raport techniczny KE⁷⁸⁾, interfejsy programowania aplikacji (API) stały się fundamentalnym technologicznym elementem nowoczesnej architektury cyfrowej,

⁷⁸⁾ Web Application Programming Interfaces (APIs): general-purpose standards, terms and European Commission initiatives, raport Joint Research Centre, Luksemburg 2019 r.

wpływającej na każdy sektor globalnej gospodarki. Według raportu, ze względu na zmienność i szybką ewolucję technologii informatycznych i internetowych, warto polegać na określonych standardach i normach dostarczanych przez organy normalizacyjne.

Udostępnianie danych za pomocą prawidłowego i bezpiecznego interfejsu programowania aplikacji może przynieść znaczącą wartość dodaną różnym podmiotom. Może również przyczynić się do tworzenia cennych ekosystemów wokół zasobów danych, których potencjał pozostaje często niewykorzystany przez posiadaczy danych⁷⁹⁾.

Zgodnie z motywem 32 dyrektywy 2019/1024/UE API jest zestawem funkcji, procedur, definicji i protokołów do komunikacji maszyna-maszyna oraz płynnej wymiany danych. Ustawodawstwo unijne określa podstawę konfiguracji i stosowania API, którą stanowi kilka zasad: dostępność, stabilność, utrzymanie przez cały cykl użytkowania, jednolitość sposobu korzystania i standardów, łatwość użytkowania oraz bezpieczeństwo. Podkreśla ponadto, aby w miarę możliwości wykorzystywać otwarte interfejsy API, stosować unijne lub uznane na szczeblu międzynarodowym standardowe protokoły i, w stosownych przypadkach, wykorzystywać międzynarodowe normy dotyczące zbiorów danych. Interfejsy API mogą różnić się poziomem złożoności i oznaczać zwykły odnośnik do bazy danych umożliwiający pobranie konkretnych zbiorów danych, interfejs internetowy lub bardziej złożoną konfigurację.

Niniejszy Program określa w załączniku nr 3 „Standard API” uwarunkowania techniczne interfejsu API. Standard określa minimalne rekomendowane zalecenia dotyczące interfejsu programistycznego aplikacji dostępu do baz danych, które przechowują dane publiczne. Udostępnienie danych zgodnie ze standardem ułatwi wykorzystywanie danych i łączenie ich z różnych źródeł, a w konsekwencji tworzenie innowacyjnych dóbr, usług i produktów przez przedsiębiorców, organizacje pozarządowe, programistów.

Działanie:

Weryfikacja Standardu API (dokonywana w ramach ewaluacji bieżącej (on-going)) w celu zapewnienia aktualności zalecanych norm, standardów oraz wytycznych ujętym w standardzie.

5.2.3. Interoperacyjność danych

Interoperacyjność stanowi kluczowy czynnik umożliwiający transformację cyfrową i wpływający na nią. Pozwala podmiotom na elektroniczną wymianę – między sobą oraz z obywatelami i przedsiębiorstwami – istotnych informacji w sposób zrozumiały dla wszystkich stron.

Przy stale rosnącej liczbie danych (często nazywanej „dużymi zbiorami danych” / „big data”), w tym danych otwartych, interoperacyjność staje się kluczowym zagadnieniem, zwłaszcza w kontekście wdrażania sztucznej inteligencji (AI). Standaryzacja na różnych poziomach (takich jak schematy metadanych, formaty reprezentacji danych i warunki licencjonowania otwartych danych) jest niezbędna, aby umożliwić szeroką integrację danych, wymianę danych i interoperacyjność z ogólnym celem wspierania innowacji opartych na danych. Odnosi się to do wszystkich rodzajów danych (wielojęzycznych), w tym zarówno danych ustrukturyzowanych, jak i danych nieustrukturyzowanych oraz danych z różnorodnych

⁷⁹⁾ Komunikat Parlamentu Europejskiego, „W kierunku wspólnej europejskiej przestrzeni danych”.

dziedzin, takich jak informacje sektora publicznego, dane geoprzestrzenne, dane statystyczne, dane pogodowe i dane badawcze⁸⁰⁾.

Wszystkie sektory gospodarki w coraz większym stopniu opierają się na technologiach cyfrowych, które zmieniają się coraz szybciej, często znacznie przekraczając tempo zmian w tradycyjnych sektorach i gałęziach przemysłu. Wartość systemów cyfrowych w coraz większym stopniu zależy od zastosowań, danych i konwergencji technologii, które wykraczają poza zasięg jednego sektora. Oprócz tego konwergencja świata fizycznego i cyfrowego zaciera granice między tradycyjnymi sektorami i gałęziami przemysłu, między produktami a usługami, między konsumpcją a produkcją, między online a offline – przez co jest wyzwaniem dla procesu opracowania norm. Rozwiązania interoperacyjne oparte na otwartych systemach i interfejsach zapewniają otwartość rynków i możliwość przenoszenia usług oraz pobudzają innowacyjność na jednolitym rynku cyfrowym⁸¹⁾.

Według informacji wskazanych w ESD podmioty generujące dane i użytkownicy danych wskazali na istnienie poważnych problemów w zakresie interoperacyjności, które utrudniają łączenie danych z różnych źródeł w obrębie sektorów, a tym bardziej między sektorami. Mając na względzie zalecenia ESD, jest pożądane stosowanie standardowych i wspólnych kompatybilnych formatów i protokołów na potrzeby gromadzenia i przetwarzania danych pochodzących z różnych źródeł w spójny i interoperacyjny sposób we wszystkich sektorach i na wszystkich rynkach przez plan normalizacji ICT⁸²⁾ oraz – w odniesieniu do usług publicznych – wzmocnione europejskie ramy interoperacyjności (EIF)^{83), 84)}. W ramach prac nad Architekturą Danych Państwa (ADP), która będzie stanowiła część Architektury Informacyjnej Państwa, zostanie położony nacisk na interoperacyjność danych, tj. na identyfikację i analizę obiektów danych przechowywanych w poszczególnych zasobach informacyjnych państwa (ograniczonych do zasobów będących w gestii szeroko rozumianej administracji publicznej), przepływy danych między systemami, opracowanie standardowych modeli najważniejszych, najczęściej wykorzystywanych obiektów danych oraz na określenie referencyjności dla poszczególnych obiektów i ich własności, a także na identyfikację zagadnień i problemów związanych z interoperacyjnością (także semantyczną), ich analizę i opracowanie wytycznych, rekomendacji i standardów.

Działania związane z normalizacją i standaryzacją powinny zmaterializować się w warstwie interoperacyjności semantycznej i technicznej, w celu ułatwienia publikowania danych w portalach otwartych danych do ponownego ich wykorzystywania, łączenia (przetwarzania) oraz zapewnienia płynnej, bezkolizyjnej wymiany danych⁸⁵⁾.

Działanie:

1. Opracowanie Architektury Informacyjnej Państwa, w tym architektury danych publicznych.

⁸⁰⁾ „The Rolling Plan on ICT Standardisation, broszura Komisji Europejskiej 05.2018; <https://ec.europa.eu/digital-single-market/en/news/rolling-plan-ict-standardisation>.

⁸¹⁾ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Priorytety w normalizacji ICT na jednolitym rynku cyfrowym, COM(2016) 178.

⁸²⁾ <https://ec.europa.eu/digital-single-market/en/news/rolling-plan-ict-standardisation>.

⁸³⁾ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów, Europejskie ramy interoperacyjności – strategia wdrażania, COM(2017) 134, https://ec.europa.eu/isa2/eif_pl.

⁸⁴⁾ W ramach EIF promuje się możliwość ponownego wykorzystania interoperacyjności jako siły napędowej, która obejmuje takie elementy jak: świadczenie usług zintegrowanych; ponowne wykorzystanie danych i usług, katalogi opisujące usługi wielokrotnego użytku, zarządzanie; bezpieczeństwo i prywatność.

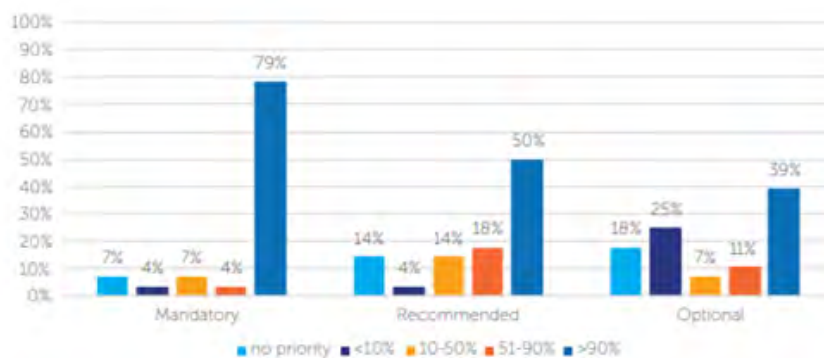
⁸⁵⁾ Komunikat Komisji do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów „Europejskie ramy interoperacyjności – strategia wdrażania”, https://ec.europa.eu/isa2/eif_pl.

5.2.4. Interoperacyjność wymiany danych między europejskimi, krajowymi i lokalnymi systemami otwartych danych

Dobrze zdefiniowane standardy umożliwiają stronom wspólne zrozumienie zagadnienia. W przypadku danych oznacza to, że pojęcia, które są istotne w dziedzinie danych, sposób ich nazywania, ich atrybuty i powiązania z innymi pojęciami są zdefiniowane i uzgodnione w ramach wspólnoty praktyków⁸⁶⁾.

Standaryzacja i normy stają się niezbędne w kontekście metadanych, ponieważ zapewniają wspólną strukturę danych, pomagają zapewnić interoperacyjność systemów oraz usprawnić wykrywanie i dostęp do danych. Zapewnienie jakościowych metadanych jest złożoną, ale niezbędną praktyką. Kluczowa jest interoperacyjność z innymi portalami danych. Jednym ze stosowanych i zalecanych standardów metadanych dla portali otwartych danych jest DCAT-AP (Application Profile)⁸⁷⁾. specyfikacja opisująca zbiory danych sektora publicznego w Europie na podstawie słownika W3C Data Catalog (DCAT). Specyfikacja DCAT-AP, definiująca schemat metadanych dla portali danych w Europie, określa podział klas i właściwości (wykorzystywanych do opisu schematu metadanych) na „wymagane”, „rekomendowane” i „opcjonalne”. Poniższy wykres pokazuje odsetek metadanych w krajowych portalach otwartych danych w Europie zgodnych z DCAT-AP, zapewniających wymagane, zalecane i opcjonalne klasy.

Wykres 6. Odsetek metadanych w krajowych portalach otwartych danych w Europie zgodnych z DCAT-AP



Źródło: Open Data Maturity 2019 r.

Dane pokazują, że w 22 krajach (79%) ponad 90% metadanych jest zgodnych z obowiązkowymi klasami (agent, katalog, zestaw danych, literał, zasób). W 14 krajach (50%), ponad 90% metadanych spełnia standardy zalecanych klas (kategoria, schemat kategorii, dystrybucja, dokument licencyjny). Natomiast w 11 krajach (39%) ponad 90% metadanych spełnia standardy klas opcjonalnych (katalog rekord, suma kontrolna, dokument, częstotliwość). Z raportu „Open Data Maturity 019 r.” wynika, że w 2019 r. polski portal Dane.gov.pl w 59% spełniał wymagania specyfikacji DCAT-AP⁸⁸⁾.

Udostępnianie otwartych danych między różnymi zainteresowanymi stronami w procesie publikacji danych jest ułatwione dzięki zastosowaniu wspólnego słownictwa. To sprawia, że proces jest bardziej wydajny i lepiej zrozumiały dla wszystkich zaangażowanych podmiotów. Profil aplikacji DCAT (DCAT-AP) to wspólny standard metadanych do

⁸⁶⁾ Sustainability of (Open) Data Portals Infrastructures, Raport EDP, 08.2020 r. <https://www.europeandataportal.eu/en/highlights/sustainability-open-data-portals-infrastructures-reports-pt-1>.

⁸⁷⁾ <https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/about>.

⁸⁸⁾ Open Data Maturity, Report 2019 r., EDP.

opisywania zbiorów danych sektora publicznego w Europie. Z tej perspektywy pożądanym rozwiązaniem jest stosowanie wymienionego standardu przez miasta, posiadające lokalne repozytoria danych, w szczególności podczas automatycznego pobierania danych przez portal Dane.gov.pl z lokalnych repozytoriów, które mają API (na zasadzie harwestera).

Działania:

1. Dostosowanie opisów zasobów dostępnych w portalu Dane.gov.pl do standardu DCAT-AP.
2. Rekomendowanie wdrażania standardu DCAT-AP w lokalnych portalach otwartych danych (przez prowadzenie działań informacyjnych, edukacyjnych).

ZALECENIA

- Udostępnianie w portalu Dane.gov.pl zasobów danych i metadanych w otwartych formatach przeznaczonych do odczytu maszynowego, umożliwiających ponowne wykorzystywanie, na możliwie najwyższym poziomie, otwartości w skali otwartości danych.
- Publikowanie w portalu Dane.gov.pl danych od drugiego poziomu otwartości.
- Publikowanie w portalu Dane.gov.pl danych w formacie pdf jest dopuszczalne, jeżeli towarzyszą mu pliki z danymi w formacie edytowalnym np. xlsx/csv.
- Przygotowywanie zasobów danych i metadanych zgodnie ze Standardem technicznym stanowiącym załącznik nr 2 do niniejszego Programu
- Zwiększanie liczby zasobów danych o najwyższym stopniu otwartości (linked data), udostępnianych w portalu Dane.gov.pl.
- Udostępnianie danych przez API według jednolitego standardu API stanowiącego załącznik nr 3 do niniejszego Programu.
- Wykorzystywanie procesu automatycznego przekazania metadanych do portalu Dane.gov.pl w przypadku posiadania API w repozytoriach danych dostępnych powszechnie w sieci teleinformatycznej.
- Wykorzystywanie w portalu Dane.gov.pl narzędzi umożliwiających automatyczne dodawanie, aktualizowanie lub usuwanie zbiorów danych oraz zasobów.
- Dostosowanie standardowych i wspólnych (kompatybilnych) formatów i protokołów na potrzeby gromadzenia i przetwarzania danych pochodzących z różnych źródeł w spójny i interoperacyjny sposób we wszystkich sektorach i na wszystkich rynkach dzięki planowi normalizacji ICT⁸⁹⁾ oraz – w odniesieniu do usług publicznych – wzmocnionym europejskim ramom interoperacyjności⁹⁰⁾.
- Stosowanie opisów zbiorów danych zgodnie ze standardem DCAT-AP.
- Uwzględnianie standardów przygotowania i udostępniania danych opisanych w Programie w obecnie wdrażanych i planowanych projektach informatycznych zarówno na poziomie legislacyjnym, jak i specyfikacji technicznej.

⁸⁹⁾ The 2018 Rolling Plan on ICT Standardisation, Komisja Europejska, <https://ec.europa.eu/digital-single-market/en/news/rolling-plan-ict-standardisation>.

⁹⁰⁾ Komunikat Komisji do do Parlamentu Europejskiego, Rady, Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów „Europejskie ramy interoperacyjności – strategia wdrażania”, https://ec.europa.eu/isa2/eif_pl.

5.3. Cel trzeci – Wzrost wykorzystywania i wymiany danych

5.3.1. Wymiana danych między różnymi dostawcami danych

Zgodnie z ESD wartość danych wynika z możliwości ich (ponownego) wykorzystania. Choć często pisze się o dużej liczbie dostępnych danych, to wciąż nie są one odpowiednio zarządzane i wystarczające na potrzeby ich ponownego wykorzystywania w technologiach cyfrowych, np. AI czy IoT. Zwiększenie wymiany danych między różnymi ich dostawcami jest czynnikiem niezbędnym do tworzenia oraz poprawy produktów i usług, a także budowania konkurencyjnej gospodarki i podniesienia poziomu jakości życia obywateli. Kluczowym czynnikiem jest nie tylko intensyfikacja wymiany danych, ale też włączenie do tego procesu jak największej liczby różnorodnych podmiotów. Wartościowe wyniki osiąga się, analizując dane pochodzące z różnych gałęzi gospodarki, ale również analizując dane dotyczące określonej tematyki, przy czym gromadzone przez kilka podmiotów. Dlatego jest konieczne wzmocnienie działań w obszarze wymiany danych na linii administracja – biznes (government to business data sharing, G2B) oraz administracja – administracja (government to government data sharing, G2G), a także podjęcie aktywności w celu stworzenia linii współpracy na linii biznes – administracja (business to government data sharing B2G) oraz biznes – biznes (business to business data sharing, B2B).

Rysunek 4. Relacje wymiany danych



Źródło: Opracowanie własne

Administracja – biznes (G2B) i administracja – administracja (G2G)

Spośród czterech głównych płaszczyzn relacji dostawców danych najbardziej utrwalona jest praktyka wykorzystywania informacji sektora publicznego przez podmioty prywatne (wymiana G2B). Wymiana danych znajdujących się w posiadaniu administracji rządowej jest prowadzoną od dawna polityką publiczną, zarówno na szczeblu krajowym, jak i unijnym. Podstawą takiej wymiany są dyrektywa reuse oraz upwisp. Zwiększenie intensyfikacji wymiany G2B powiększa wolumen dostępnych danych, które mogą być wykorzystywane z korzyścią dla różnych grup interesariuszy.

Odbiorcą otwartych danych publicznych jest również sama administracja. Dostęp do wysokiej jakości danych udostępnianych przez administrację innym instytucjom wspiera tworzenie lepszych usług publicznych, pozwala na sprawniejsze planowanie działań oraz przeciwdziałanie negatywnym zjawiskom społecznym. Pozwala też na usprawnienie

działania samej administracji przez zmniejszenie jej pracochłonności, a w konsekwencji obniżanie kosztów funkcjonowania. Analiza danych ułatwia optymalizację procesów i decyzji oraz wspomaga przewidywanie zdarzeń. Oparcie decyzji na analizie danych pozwala na bardziej skuteczne kształtowanie polityki gospodarczej i społecznej oraz projektowanie usług publicznych w taki sposób, aby były one w większym stopniu dostosowane do potrzeb obywateli.

Ważne jest zintensyfikowanie wymiany danych między jednostkami publicznymi. W tym celu konieczne jest wprowadzenie działań w zakresie:

- 1) zapewnienia dostępności i wysokiej jakości danych;
- 2) stworzenia rozwiązań ułatwiających dzielenie się danymi, co wiąże się z przełamaniem niechęci organizacyjnej w tym zakresie, znajdowaniem skutecznych sposobów komunikowania się i usuwaniem barier dla interoperacyjności;
- 3) budowania umiejętności i zdolności do jak najbardziej skutecznego wykorzystania danych.

Równie ważna jest wymiana danych między administracjami różnych państw. Istotnym aspektem takiego wykorzystywania danych udostępnianych przez administracje innych państw jest wspieranie funkcjonowania przedsiębiorców na jednolitym rynku (np. przez skuteczniejsze stosowanie zasady jednorazowości i zmniejszenie tym samym obciążeń administracyjnych) czy zmniejszenie skali negatywnych zjawisk i nieprawidłowości w życiu społeczno-gospodarczym (np. zmniejszenie przestępczości), ale również możliwość wykorzystywania danych pochodzących z zagranicznych podmiotów publicznych przez krajowe administracje w ramach wykonywania zadań publicznych.

Szerokie zaangażowanie administracji rządowej w otwieranie danych publicznych oraz w zwiększenie ich wymiany w relacjach G2B i G2G może mieć pozytywny wpływ na prywatnych dostawców danych. Organy publiczne, dzięki podejmowanym przez siebie działaniom na rzecz otwartości danych, kształtują oraz upowszechniają kulturę wymiany danych. Dzięki własnym doświadczeniom mogą skuteczniej wypracowywać i rozpowszechniać modele zarządzania danymi oraz standardy dotyczące danych, które zagwarantują dostępność danych i ich interoperacyjność.

Biznes – biznes (B2B) i biznes – administracja (B2G)

Wymiana danych między podmiotami prywatnymi stanowi dobrowolne udostępnianie danych innym przedsiębiorcom lub uzyskiwanie dostępu do danych posiadanych przez inne firmy w celach biznesowych. Jej celem jest zwiększenie możliwości biznesowych przedsiębiorstwa, a także poprawienie jego wewnętrznej wydajności. Istnieją różne modele udostępniania i ponownego wykorzystania danych w relacjach B2B⁹¹⁾:

- a) monetyzacja danych – przedsiębiorcy sprzedają dane, które posiadają; w tym modelu dane mogą być udostępniane odpłatnie jako usługa oferowana chętnym podmiotom,
- b) rynki danych (ang. data marketplaces) – platformy internetowe, które działają jako zaufani pośrednicy, umożliwiając wymianę danych między zainteresowanymi podmiotami,
- c) przemysłowe platformy danych (ang. industrial data platforms) – przykład wymiany danych opartej na współpracy – platformy przemysłowe są otwarte dla ograniczonej, zamkniętej grupy podmiotów, dane są wymieniane z reguły nieodpłatnie, a wykorzystywane przede wszystkim do wspierania rozwoju nowych produktów oraz poprawy wydajności i produktywności,

⁹¹⁾ <https://op.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en>.

- d) rozwiązania techniczne – istnieją firmy oferujące innym firmom wyspecjalizowane rozwiązania techniczne służące wymianie danych, a zarabiają na ich projektowaniu, dostosowaniu do potrzeb klienta oraz administrowaniu wdrożonym systemem; ten mechanizm wymiany danych jest podobny do platform wymiany danych opisanych wyżej, jednak w ich przypadku platformy zarabiają, pobierając prowizję od każdej przeprowadzonej za ich pośrednictwem transakcji sprzedaży danych,
- e) polityka otwartych danych – ten model wymiany danych zakłada bezpłatne udostępnienie danych przez firmę chętnym podmiotom w celu ulepszenia oferowanych produktów lub tworzenia nowych rozwiązań. To model często stosowany w firmach, które są prawnie zobowiązane do udostępniania danych.

Dla wymiany danych pochodzących od podmiotów sektora prywatnego brakuje regulacji na poziomie zarówno europejskim, jak i krajowym, ustanawiających ogólne zasady udostępniania tych danych i ich ponownego wykorzystywania, zarówno innym podmiotom prywatnym, jak i administracji. Z tego względu dostęp do dużej liczby danych, które mają istotne znaczenie dla realizacji zadań w interesie publicznym i wysoką wartość dla dobra wspólnego (ale które są generowane przez podmioty prywatne), jest zamknięty lub istotnie ograniczony. Potrzeba zapewnienia szerszego dostępu do danych generowanych przez podmioty prywatne jest tym większa, że wiele danych, które mogą mieć istotne znaczenie dla realizacji zadań w interesie publicznym (np. w dziedzinie transportu, dostarczania i wykorzystania energii, ochrony środowiska, edukacji itp.), jest świadczonych przez podmioty prywatne.

Udostępnianie danych prywatnych do ponownego wykorzystywania przez podmioty publiczne jest obecnie mocno ograniczone. Większość rozwiązań dzielenia się danymi przez podmioty prywatne z podmiotami publicznymi czy udostępniania danych będących w posiadaniu podmiotów prywatnych w sferze publicznej ma charakter dobrowolny. Wyjątkiem są sektory regulowane, gdzie dzielenie się danymi przez podmioty prywatne z podmiotami publicznymi już się odbywa bądź może zostać wprowadzone (np. dostęp organów państwa do danych będących w posiadaniu operatorów telekomunikacyjnych). Jest pożądane to, aby podmioty prywatne udostępniały te dane, które powstały w wyniku realizacji zadania publicznego. Dane są uznawane przez podmioty prywatne za ich przewagę konkurencyjną i często podlegają ochronie z tytułu ochrony konkurencji, tajemnicy przedsiębiorstwa, ochrony praw własności intelektualnej czy ochrony danych osobowych. Również wymiana danych między przedsiębiorcami regulowana jest jedynie nielicznymi przepisami sektorowymi, np. w sektorze finansowym, motoryzacyjnym czy w odniesieniu do baz danych. Istniejące platformy wymiany danych B2B funkcjonują na podstawie postanowień umownych. Nie gwarantują one dynamiki i wolumenu wymienianych danych odpowiednich do potrzeb cyfrowej gospodarki.

Istnieje wiele barier⁹²⁾ dotyczących wymiany danych między podmiotami prywatnymi oraz między nimi a instytucjami publicznymi, które powodują, że mimo potencjału gospodarczego tych danych przedsiębiorstwa niechętnie dzielą się nimi. Konieczne jest podjęcie działań, które odpowiadałyby na istniejące w tym zakresie potrzeby, niwelując bariery w udostępnianiu i ponownym wykorzystywaniu danych, a tym samym zwiększyłyby dynamikę wymiany zasobów informacji pozostających w dyspozycji podmiotów prywatnych, ale również administracji.

⁹²⁾ Patrz rozdział 5.3.2.

W ramach szeroko pojętej unijnej polityki cyfrowej Komisja Europejska podejmuje działania w celu ożywienia wymiany danych na rynku europejskim oraz w wymiarze międzynarodowym. Polska musi być aktywnym uczestnikiem w projektowanych inicjatywach dotyczących wymiany danych w ramach europejskiej przestrzeni danych.

Działania:

1. Aktywne uczestnictwo Rzeczypospolitej Polskiej w tworzeniu unijnych aktów prawnych oraz mechanizmów i rozwiązań pozalegislacyjnych dotyczących wymiany danych.
2. Aktywne promowanie w kraju przyjętych rozwiązań i mechanizmów mających na celu upowszechnienie wymiany i ponownego wykorzystywania danych.

5.3.2. Likwidacja barier utrudniających dostęp i wymianę danych

Filary otwartości

Liczba danych generowanych na świecie gwałtownie rośnie, oddziałując na wszystkie sektory gospodarki i sfery życia człowieka. Dane są siłą napędową nowych technologii, cennym zasobem dla powstania innowacyjnych produktów i usług, źródłem zachodzących zmian gospodarczych i społecznych, inspiracją dla środowisk naukowych, narzędziem służącym do usprawnienia realizowanych procesów. Aby było to możliwe, należy udostępniać dane w sposób otwarty, co należy rozumieć jako spełnienie następujących filarów otwartości:

a) dostępne

Udostępnione bez żadnych ograniczeń szerokiemu gronu użytkowników (osobie fizycznej, osobie prawnej, jednostce organizacyjnej nieposiadającej osobowości prawnej) do dowolnych celów (komercyjnych i niekomercyjnych). Tak udostępnione dane można np. pobierać, indeksować, przeszukiwać.

b) upublicznione w wersji źródłowej

Dostępne w oryginalnej i niezmienionej formie, a nie w postaci np. analiz, podsumowań, skrótów czy streszczeń, tak aby możliwe było m.in. łączenie danych z różnych źródeł. Chodzi zatem o dostęp do danych w tzw. postaci surowej, a nie tylko zagregowanej lub zmodyfikowanej. Udostępnienie danych w wersji źródłowej nie stanowi przeszkody do udostępniania materiałów (analiz, podsumowań itp.) powstałych na ich podstawie.

c) kompletne

Udostępnione w całości, bez wyłączenia poszczególnych części danych i zgodnie ze stanem faktycznym, chyba że wyłączenie wynika z przepisów prawa (np. ochrona danych osobowych zawartych w danych publicznych czy zachowanie tajemnicy statystycznej). Wybiórcze udostępnianie danych publicznych (np. wybranych dokumentów czy dokumentacji pozbawionej załączników) przeczy idei dostępu do danych publicznych i może zniweczyć użyteczność tych danych.

d) aktualne

Niezwłoczne udostępnianie danych, które dostawca danych sam wytworzył lub pozyskał. Istotne jest udostępnianie danych najnowszych, a tym samym najbardziej przydatnych w działalności gospodarczej, naukowej lub badawczej. Udostępnianie danych po kilku miesiącach od ich uzyskania przez dostawcę danych może prowadzić do znacznego

obniżenia ich wartości. Udostępnianie danych najnowszych nie wpływa na możliwość udostępniania danych historycznych, które są przydatne do prac analitycznych.

e) odczytywalne maszynowo

Udostępnione w formatach umożliwiającym automatyczne odczytywanie przez przeglądarkę lub system komputerowy. Przykładami takich formatów są: XML, JSON, RDF, CSV. Formaty te ułatwiają dostęp i umożliwiają bardziej zaawansowane analizy dużej liczby informacji. Korzystanie w sposób zautomatyzowany z danych udostępnianych w formacie PDF czy w formie pliku tekstowego jest utrudnione, ponieważ wymaga przetworzenia do ustrukturyzowanego formatu otwartego.

f) udostępnione niedyskryminująco

Dostępne dla każdego bez konieczności rejestracji, posiadania konta internetowego, weryfikacji tożsamości przez podawanie hasła, loginu czy podpisywania jakichkolwiek umów. Konieczność rejestracji uniemożliwia również maszynowy odczyt danych i może prowadzić do ograniczenia kręgu potencjalnych użytkowników.

g) dostępne bez ograniczeń licencyjnych

Udostępniane dane nie powinny być przedmiotem praw autorskich, patentów, znaków towarowych lub tajemnicy handlowej. Mogą być wykorzystywane w dowolnych celach bez konieczności ubiegania się o jakąkolwiek zgodę na ich używanie, z zachowaniem obowiązujących przepisów prawa.

h) niezastrzeżone

Dostępne w formacie powszechnie stosowanym, który nie jest kontrolowany przez żaden podmiot. Chodzi o formaty, których charakterystyka jest jawna i dobrze opisana (udokumentowana), a zatem w formacie otwartym.

Realizacja wszystkich opisanych wytycznych pozwoli przypisać udostępnianym danym przymiotu otwartości, co stanowi podstawę dostępności i jakości udostępnianych danych, a te cechy przekładają się na ich użyteczność i możliwość dalszego wykorzystania w produktach i usługach. Pożądana byłaby ich implementacja w przepisach prawa, aby stały się drogowskazem, jak udostępniać dane, zachowując ich wartość informacyjną.

Rysunek 5. Filary otwartości



Źródło: Opracowanie własne

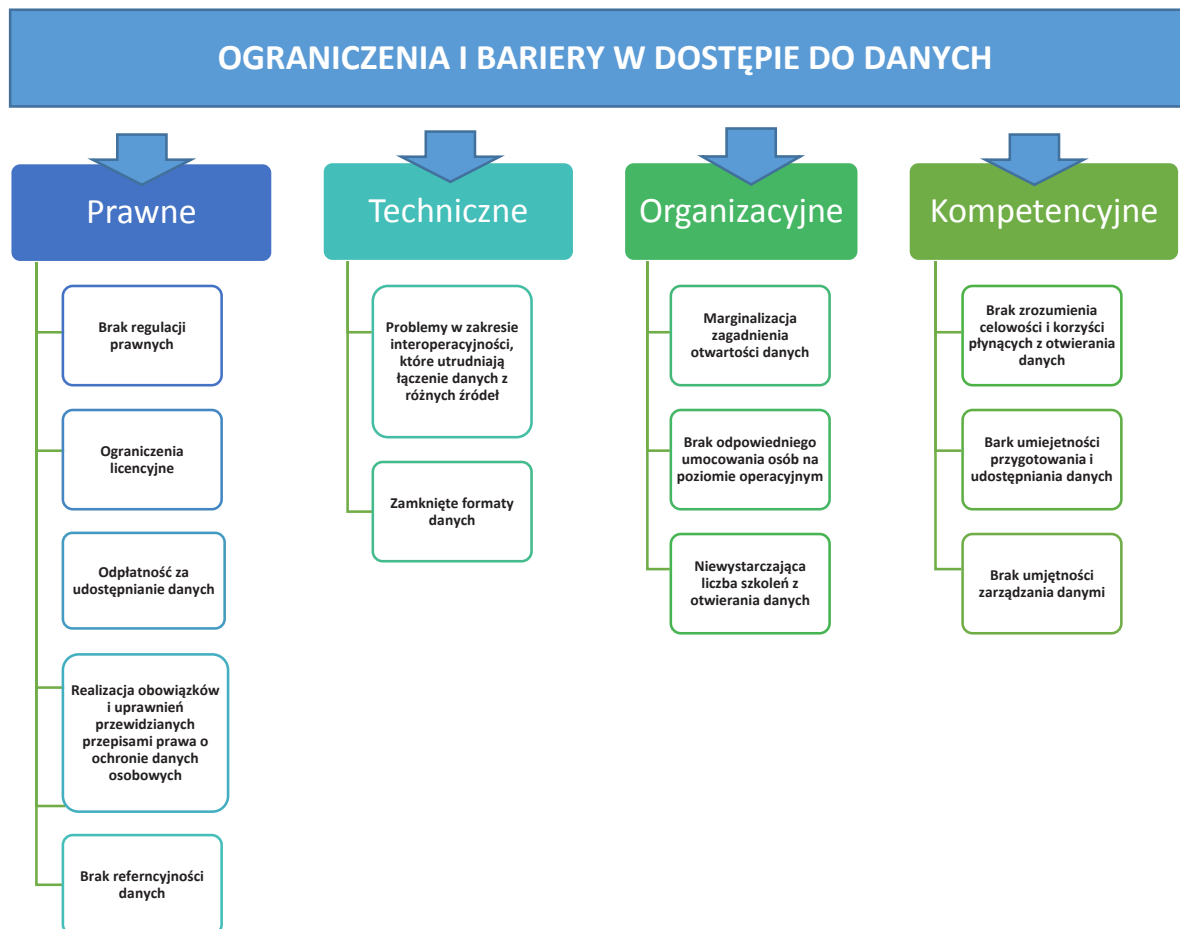
W celu weryfikacji otwartości udostępnianych danych Ministerstwo Cyfryzacji opracowało, w ramach standardu prawnego otwartości danych, tzw. listę sprawdzającą⁹³⁾.

Ograniczenia i bariery w dostępie do danych

Istnieje szereg czynników utrudniających dostęp do danych i ich wymianę między różnymi grupami dostawców danych, w tym również przedstawicielami sektora prywatnego.

Bariery znacząco uniemożliwiające dostęp do danych można podzielić na cztery kategorie: prawne, techniczne, organizacyjne oraz kompetencyjne.

Rysunek 6. Ograniczenia i bariery w dostępie do danych uniemożliwiające ponowne ich wykorzystywanie



Źródło: Opracowanie własne

Wśród barier prawnych⁹⁴⁾ wyróżnia się:

a) brak regulacji prawnych

Brak odpowiednich regulacji prawnych dotyczących otwartości danych, czyli ich zgodności ze wszystkimi filarami otwartości, co prowadzi do wielu komplikacji po stronie potencjalnego użytkownika tych danych, który znajduje się w stanie niepewności prawnej i ponosi ryzyko naruszenia zasad wynikających m.in. z przepisów dotyczących praw własności intelektualnej czy ochrony danych osobowych,

b) ograniczenia licencyjne

⁹³⁾ <https://dane.gov.pl/media/ckeditor/2020/05/29/standard-prawny.pdf>.

⁹⁴⁾ Szerzej standard prawny <https://dane.gov.pl/media/ckeditor/2020/05/29/standard-prawny.pdf>.

Umowa licencyjna może stanowić źródło ograniczeń w wykorzystywaniu udostępnionych danych przez wskazanie zakresu operacji na danych, jakie użytkownik może realizować, w szczególności ich utrwalania, zmieniania, udostępniania. Często praktyką w tym zakresie jest również stosowanie sankcji za naruszenie licencji oraz zawieranie licencji wyłącznych, w których licencjodawca udziela konkretnemu podmiotowi tylko prawo do korzystania z danego utworu na określonym polu eksploatacji. Może to prowadzić do ograniczenia tworzenia produktów konkurencyjnych (komercyjnych lub niekomercyjnych), wykorzystujących przedmiotowe informacje przez innych zainteresowanych. W efekcie społeczna wartość uzyskiwana z danych publicznych byłaby ograniczona do korzyści jednego podmiotu, co godziłoby w istotę ponownego wykorzystywania,

c) odpłatność za udostępnianie danych

Kolejną barierą jest pobieranie opłat za udostępnianie danych, przy czym dotyczy to płatności rozumianej jako zysk z udostępniania danych. Zatem jako barierę należy identyfikować działania polegające na pobieraniu płatności w celach zarobkowych, stanowiących dochód dostawcy danych. Nie dotyczy to opłat uwzględniających koszty przygotowania lub przekazania danych, jeżeli przygotowanie lub przekazanie danych w sposób lub w formie wskazanej we wniosku wymaga poniesienia dodatkowych kosztów, np. kosztów nośnika, na którym mają zostać udostępnione dane, kosztów tłumaczenia na język obcy czy kosztów skorzystania z usługi zewnętrznego podmiotu w celu utrwalania informacji w sposób wskazany we wniosku (np. druk wielkoformatowy),

d) realizacja obowiązków i uprawnień przewidzianych przepisami prawa o ochronie danych osobowych

Przepisy o ochronie danych osobowych oraz o ponownym wykorzystywaniu informacji sektora publicznego krzyżują się i mają wspólne obszary regulacji. Ich współstosowanie może jednak rodzić dla praktyki pewne trudności. Przy udostępnianiu danych do ponownego wykorzystywania konieczne jest uwzględnienie prawa do ochrony danych osobowych. Wątpliwości dysponentów danych budzić może spełnienie obowiązków wynikających z istniejących regulacji, a dotyczących w szczególności spełnienia obowiązku informacyjnego (art. 13, art. 14 i art. 19 RODO), wymogu testu zgodności celów (art. 6 ust. 4 RODO), przeprowadzenia oceny skutków dla ochrony danych osobowych (zgodnie z art. 35 RODO) oraz możliwości nałożenia warunków ponownego wykorzystywania informacji sektora publicznego zawierającej dane osobowe,

e) brak wskazania referencyjności danych i określenia jednego źródła danych

Wielość rejestrów i aktów prawnych, na podstawie których są one prowadzone, różnorodność celów ich powstania oraz funkcji, jakie pełnią, powoduje, że następuje powielanie publikowania danych tego samego rodzaju i w tym zakresie w różnych rejestrach. Brakuje jasnego wskazania rejestrów bazowych prowadzonych przez podmioty publiczne, które zawierają dane referencyjne i stanowią wiarygodne źródło danych dla ich użytkowników. Sytuacja ta utrudnia wykorzystywanie danych przez użytkowników, ponieważ między danymi tego samego rodzaju pobranymi z różnych rejestrów mogą występować różnice. Bariera ta negatywnie wpływa na spójność danych, a jej konsekwencją są niezgodności w danych.

Do barier technicznych zalicza się:

a) dostępność danych w formatach zamkniętych lub zastrzeżonych

Istotną barierą w korzystaniu z danych jest ich publikowanie w formatach uniemożliwiających maszynowy odczyt lub wymagających odpłatnego oprogramowania. Oznacza to, że

specyfikacja formatów danych jest własnością firm je tworzących, a możliwość ich przeszukiwania, analizowania i ponownego wykorzystywania wymaga znacznych nakładów osobowych i finansowych (np. aby korzystać z danych, konieczny jest zakup specjalistycznego oprogramowania),

b) problemy z interoperacyjnością danych

Według informacji wskazanych w ESD podmioty generujące dane i użytkownicy danych wskazali na istnienie poważnych problemów w zakresie interoperacyjności, które utrudniają łączenie danych z różnych źródeł w obrębie sektorów, a tym bardziej między sektorami. Podobne zastrzeżenia zostały zgłoszone przez sektor prywatny (biznes).

W ramach barier kompetencyjnych występują:

a) brak zrozumienia celowości i korzyści płynących z otwierania danych

Wciąż spora część pracowników administracji posiada małą wiedzę, czym są otwarte dane i jakie dają możliwości. Niska jest świadomość w zakresie korzyści, jakie wynikają z udostępniania danych do ponownego wykorzystywania, w tym również w nowych technologiach. Sprowadza się to do podejmowania mocno zachowawczych działań w obszarze dostępu do danych, które opierają na wypełnianiu niezbędnego minimum, wynikającego z obowiązujących poszczególnie jednostki regulacji,

b) brak umiejętności przygotowania i udostępniania danych

Poprawy wymagają umiejętności urzędników w zakresie odpowiedniego przygotowania zbiorów danych do udostępniania, w szczególności w zakresie strukturyzacji plików danych, formatów czy podnoszenia jakości udostępnianych już danych, aby udostępniane dane były użyteczne i mogły zostać wykorzystane zarówno przez sektor prywatny, jak i przez sektor publiczny,

c) brak umiejętności zarządzania danymi

Do niedawna dane były postrzegane jako produkt uboczny realizowanych procesów, obecnie stanowią wartość samą w sobie, o ile są dobrze zarządzane. Wymaga to od dostawców danych odpowiedniego przygotowania swojej jednostki, w tym również zdobycia wiedzy, jak właściwie np. gromadzić, archiwizować, udostępniać i wyszukiwać potrzebne dane oraz jak należy prowadzić ich analizę, aby przynosiła ona wymierne korzyści dla danej instytucji.

W ramach barier organizacyjnych występuje:

a) marginalizacja zagadnienia otwartości danych

Choć dane są źródłem powstania nowych produktów i usług, a wielkie zbiory danych są kluczowe dla rozwoju nowych technologii, to zagadnienie otwartości danych nadal nie jest priorytetem, w szczególności dla administracji publicznej. Brak odpowiedniej kultury dostępu i wymiany danych hamuje rozwój wykorzystania potencjału zawartego w gromadzonych informacjach,

b) brak odpowiedniego umocowania osób na poziomie operacyjnym

Choć w praktyce działa sieć pełnomocników do spraw otwartości danych, to brak ich umocowania ustawowego oraz brak nadania tej funkcji odpowiedniego znaczenia w danej organizacji uniemożliwia prawidłową realizację procesu otwierania danych. W konsekwencji osoby realizujące te zadania nie mają przełożenia na jednostki organizacyjne w swojej instytucji, co w sposób znaczący utrudnia realizację polityki otwartego dostępu do danych w celu ponownego wykorzystywania,

c) niewystarczająca liczba szkoleń z otwierania danych

Mimo wielu działań szkoleniowych prowadzonych w obszarze otwartych danych nadal istnieje zapotrzebowanie na ich kontynuację w odniesieniu do nowych interesariuszy, np. jednostek samorządu terytorialnego.

Wszystkie opisane zagadnienia utrudniają dostęp do danych i ich ponowne wykorzystywanie, hamując potencjał informacyjny w nich zawarty. Tym samym negatywnie wpływają na możliwość wtórnego użycia danych w produktach i usługach oraz nowych technologiach. Dlatego priorytetowym zadaniem jest stworzenie przyjaznego otoczenia regulacyjnego oraz użycie wszelkich środków usuwających te bariery. Pomocą dla dostawców danych będą działania informacyjne pokazujące sposoby przeciwdziałania występującym zagadnieniom, które zakłócają użytkownikom dostęp do danych i możliwość ich wykorzystania. Może się to wiązać ze zmianą prowadzonej polityki w zakresie otwartości danych, która musi zostać ukierunkowana na potrzeby użytkowników, co powinno prowadzić, w miarę możliwości, dostawców danych do opisanych poniżej działań.

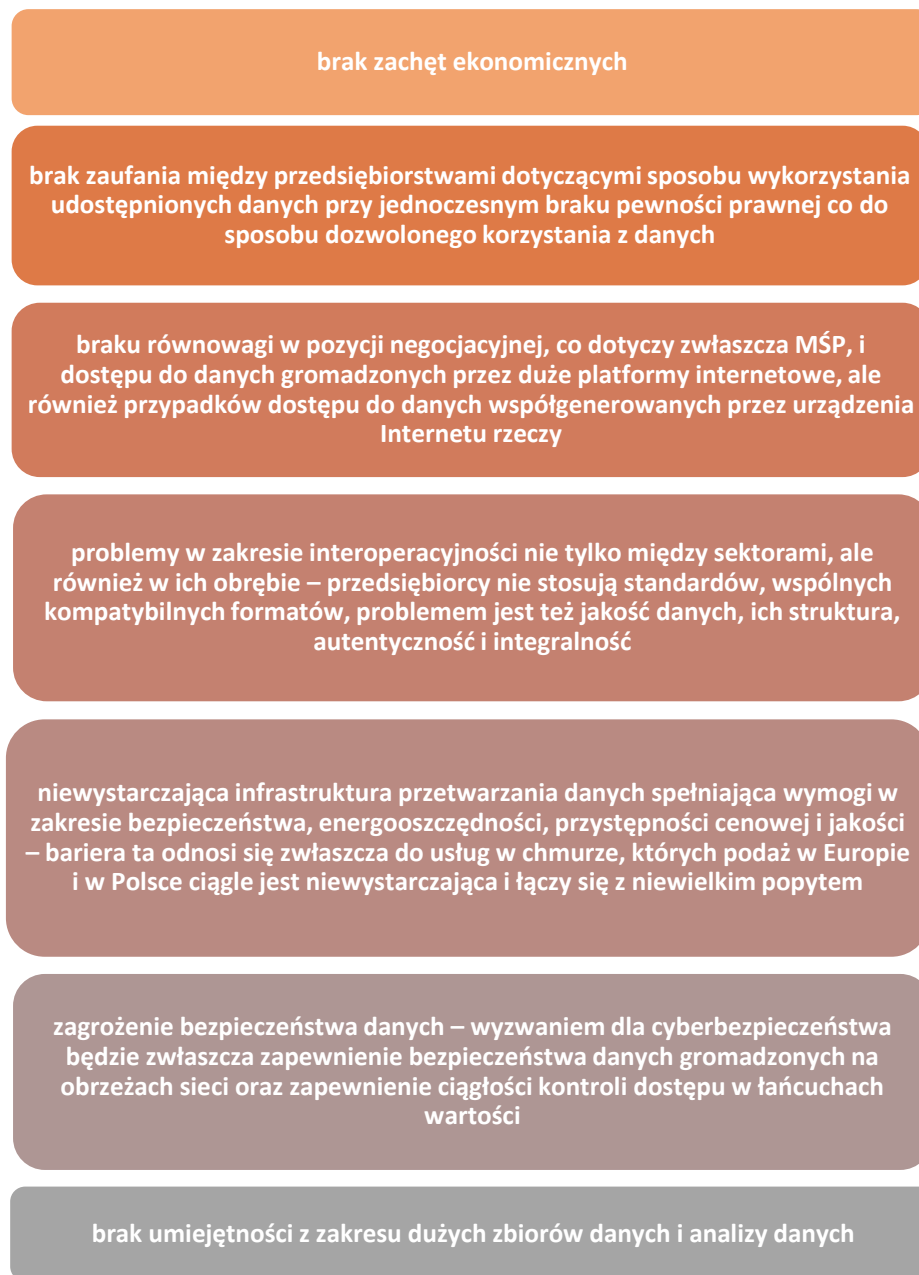
Działania usuwające bariery w dostępie do danych i ich ponownym wykorzystaniu:

- a) udostępnianie danych w sposób otwarty, tj. zgodnie ze wszystkimi filarami otwartości,
- b) podejmowanie działań legislacyjnych w celu ustalenia jasnych reguł dostępu i wykorzystywania danych,
- c) ograniczenie pobierania opłat za udostępniane dane,
- d) unikanie licencjonowania dostępu do danych,
- e) udostępnianie danych z baz/rejestrów finansowanych lub współfinansowanych ze środków publicznych w sposób otwarty, w szczególności przez API,
- f) oznaczanie danych i metadanych będących przedmiotem praw własności intelektualnej,
- g) udostępnianie danych w formatach otwartych, których charakterystyka jest jawna i dobrze opisana, albo umożliwianie w prosty sposób wyeksportowania danych do otwartego formatu pliku,
- h) wskazywanie w aktach prawnych danych referencyjnych stanowiących wiarygodne źródło informacji,
- i) nieużywanie zasad ochrony danych osobowych jako argumentu do odmowy udostępnienia danych do ponownego wykorzystania, lecz podjęcie działań prowadzących do uniemożliwienia identyfikacji osób fizycznych, których dane dotyczą, przy zachowaniu wartości informacyjnej określonego zasobu,
- j) prowadzenie działań informacyjno-promocyjnych pokazujących wartość i znaczenie otwartych danych,
- k) organizowanie szkoleń dla dostawców danych podnoszących kompetencje osób odpowiedzialnych za udostępnianie danych oraz właściwe nimi zarządzanie,
- l) umocowanie funkcji pełnomocnika do spraw otwartości danych w akcie prawnym rangi ustawowej,
- m) prowadzenie kampanii informacyjnych dotyczących potencjału otwartych danych.

Bariery utrudniające wymianę danych

Oprócz barier w dostępie do danych istnieją, poza brakiem ogólnych ram regulacyjnych, liczne bariery hamujące wzrost wymiany danych między przedsiębiorstwami oraz w obszarze udostępniania przez nich danych administracji publicznej, na które zwraca uwagę Komisja Europejska w ESD. Oprócz faktu, że przedsiębiorstwa obawiają się utraty przewagi konkurencyjnej, której źródła upatrują w gromadzonych danych, wskazuje się również na inne czynniki ograniczające wymianę danych.

Rysunek 7. Bariery utrudniające wymianę danych



Źródło: Opracowanie własne

Kluczowe jest zaprojektowanie i wdrożenie rozwiązań eliminujących przeszkody stojące na drodze do swobodnej wymiany danych między przedstawicielami sektora prywatnego oraz sektorem prywatnym a administracją. Efektem tych działań będzie stworzenie przyjaznego ekosystemu do wymiany danych opartego na uregulowaniach z obszaru prawno-legislacyjnego, infrastrukturalnego oraz kompetencji. W pierwszej kolejności działania powinny skupiać się na eliminacji barier odnoszących się do wymiany danych i na opracowaniu standardów dzielenia się danymi, aby dostawcy danych nie obawiali się zakłócenia konkurencji na rynku, na którym działają.

Działania:

1. Implementacja filarów otwartości do przepisów ustawowych.
2. Prowadzenie działań informacyjnych dla dostawców danych dotyczących sposobów postępowania z barierami utrudniającymi dostęp do danych.
3. Opracowanie standardów wymiany danych.

5.3.3. System zachęt do stymulowania rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych

W ramach tego kierunku interwencji będą prowadzone działania skierowane do administracji rządowej, samorządowej oraz przedsiębiorców, obejmujące wsparcie finansowe, doradcze i promocyjne. Przedstawienie poniższego systemu zachęt i narzędzi wsparcia ma na celu stymulowanie rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych.

Finansowanie z UE w perspektywie 2021–2027

Państwa członkowskie i organy sektora publicznego w ramach podejmowanych starań o zapewnienie łatwego dostępu do danych na potrzeby ponownego ich wykorzystywania mogą czerpać korzyści i uzyskiwać stosowne wsparcie finansowe z odpowiednich funduszy i programów Unii zapewniające upowszechnienie technologii cyfrowych lub cyfrową transformację administracji i służby publicznej.

Obecnie trwają prace przygotowawcze do nowej perspektywy finansowej. Urząd obsługujący ministra właściwego do spraw informatyzacji prowadzi aktywne działania w celu włączenia kwestii związanych z otwieraniem danych publicznych w nowym programie operacyjnym (następca Programu Operacyjnego Polska Cyfrowa). W programie tym przewidziano działania zwiększające dostępność ISP przez np.:

- a) poprawę jakości danych,
- b) digitalizację zasobów kultury, administracji i nauki (m.in. dane z badań naukowych oraz zasobów edukacyjnych finansowanych ze środków publicznych),
- c) opracowywanie standardów oraz budowę API, w tym narzędzi umożliwiających dostęp w czasie rzeczywistym do danych dynamicznych oraz danych o wysokiej wartości,
- d) dostosowanie danych do standardów, w tym tych opracowanych przez ministra właściwego do spraw informatyzacji,
- e) budowę i rozbudowę infrastruktury do przechowywania i udostępniania ISP, w tym zapewnienie interoperacyjności systemów udostępniających dane do ponownego wykorzystania, integrację i harmonizację danych,
- f) zwiększenie wymiany danych G2G w celu lepszego wykorzystania potencjału danych przez administrację publiczną,
- g) tworzenie i udostępnianie narzędzi do analityki danych,
- h) upowszechnianie wiedzy i informacji na temat otwierania i ponownego wykorzystywania otwartych zasobów.

Ponadto będzie kontynuowana realizacja 16 regionalnych programów operacyjnych będących następcami obecnych zarządzanych i wdrażanych przez samorządy województw.

Kolejnym wsparciem jest skorzystanie z Instrumentu na rzecz Odbudowy i Zwiększania Odporności z Mechanizmu Odbudowy i Odporności, pakietu nowych instrumentów, mechanizmów i programów, przygotowanego przez Komisję Europejską na odbudowę europejskiej gospodarki dotkniętej kryzysem w wyniku pandemii Covid-19. Jest to narzędzie tymczasowe – trwające do końca 2024 r. O wsparcie finansowania będą mogły ubiegać się projekty, których celem będzie przeciwdziałanie skutkom pandemii, także gospodarczym, np. przez budowanie usług i produktów przy wykorzystaniu danych.

Cyfrowa Piaskownica Administracji

Celem projektu jest m.in. umożliwienie testowania inicjatyw budujących nowe rozwiązania biznesowe w oparciu o usługi udostępniane przez platformę CPA, udostępnienie z jednego miejsca dostępu do testowych usług API oraz skrócenie czasu i kosztów związanych z integracją podmiotów zewnętrznych za pomocą usług API z nowymi systemami administracji publicznej. Obecnie w piaskownicy są dostępne różne typy usług API: takie, z których już można korzystać produkcyjnie, usługi w trakcie realizacji oraz prototypy nowych usług, które zostaną zaproponowane przez społeczność CPA. Działania w postaci inicjatyw, wydarzeń i otwartego laboratorium umożliwiają członkom społeczności CPA ocenienie przydatności propozycji nowych usług API oraz budowę własnych rozwiązań wykorzystujących prototypy udostępnione w piaskownicy. Rozwój Cyfrowej Piaskownicy Administracji wskazano w projekcie „Polityki dla rozwoju Sztucznej Inteligencji w Polsce od roku 2020” jako narzędzie do podejmowania działań w określonych obszarach związanych z rozwojem Sztucznej Inteligencji, w szczególności w celu wydajnego i szybkiego dostępu do danych i ich wykorzystywania przez wszystkich uczestników życia gospodarczego bez względu na wielkość instytucji.

Wyzwania organizowane przez GOV TECH Polska

Formuła GovTech polega na dialogu między administracją publiczną a innowatorami: przedsiębiorcami, start-upami, środowiskiem naukowym oraz samymi obywatelami, wspierającymi swoimi zasobami inicjatywy, które wykorzystując posiadane przez instytucje dane, zakładają użycie nowych technologii do ich analizy celem dostarczenia nowych informacji lub usług. Program ten oferuje budowę nowoczesnych rozwiązań technologicznych i informatycznych z myślą o rozwiązywaniu problemów, z jakimi mierzy się sektor publiczny.

Hackathony otwartych danych w ramach administracji rządowej

Hackathony, czyli maratony programowania, to sprawdzone narzędzie wspierające przedsiębiorców i programistów, którego celem jest wykorzystywanie otwartych danych publicznych do tworzenia prototypów innowacyjnych produktów i aplikacji. Hackathony to także rodzaj zachęty dla urzędów i instytucji do udostępniania danych.

Laboratorium Otwartych Danych

Raporty opracowane przez Laboratorium Otwartych Danych i rekomendacje powstałe w ich wyniku stanowią element kompleksowego wsparcia eksperckiego dla dostawców danych w zakresie analitycznym, prawnym, technicznym i bezpieczeństwa udostępniania informacji sektora publicznego. Funkcjonowanie Laboratorium Otwartych Danych jest nakierowane na problemy administracji rządowej i stanowi jej wsparcie doradcze. Przygotowane raporty mogą stanowić źródło informacji w zakresie udostępniania zasobów publicznych do ponownego wykorzystywania, nie tylko dla instytucji, które zdecydowały się na skorzystanie ze wsparcia Laboratorium Otwartych Danych, ale także dla innych podmiotów publicznych, w tym również jednostek samorządu terytorialnego, zaangażowanych w proces otwierania danych publicznych w Polsce.

Zintegrowana Platforma Analityczna

Głównym celem projektu jest podniesienie skuteczności działań administracji w wybranych obszarach problemów społecznych i gospodarczych, przez wsparcie procesów decyzyjnych za pomocą informacji analitycznej wysokiej jakości. W kontekście danych cel ten zostanie zrealizowany przez:

- a) opracowanie wzorców dostępu do danych oraz analiz dla kluczowych obszarów społecznych i gospodarczych,
- b) usystematyzowanie metod technicznych i zasad dostępu do kluczowych zasobów informacyjnych państwa dla celów analitycznych, wraz z wdrożeniem spójnego systemu bezpieczeństwa, metod dostępu do danych, udostępniania do celów analitycznych,
- c) udostępnienie danych administracji publicznej do badań naukowych wspierających podejmowanie decyzji przez twórców polityk publicznych przy równoczesnym zapewnieniu właściwej ochrony.

Jakość oraz dostępność (w tym przypadku niezwiązana z publicznością danych) jest szczególnie istotna w kwestii udostępnienia w jednym miejscu narzędzi analitycznych wspierających procesy decyzyjne w administracji publicznej, a także stworzenia technicznych i organizacyjnych możliwości dialogu społecznego i współpracy środowisk naukowych w rozwoju analityki państwa i podnoszeniu efektywności rozwiązywania problemów społecznych.

Publikowanie aplikacji w portalu Dane.gov.pl

Narzędziem rozpowszechniania informacji o aplikacjach i usługach bazujących na danych stworzonych przez administrację rządową, samorządową, sektor prywatny, organizacje pozarządowe jest ich publikowanie w portalu Dane.gov.pl⁹⁵. W zakładce „Aplikacje” publikowane są przykłady wykorzystania danych publicznych w aplikacjach, produktach i usługach. Strona ta ma na celu upowszechnianie informacji o wybranych projektach wykorzystujących dane.

Działania:

1. Realizacja projektów dotyczących zwiększania dostępności i jakości danych w ramach WRF 2021–2027.
2. Realizacja projektu Budowa Zintegrowanej Platformy Analitycznej.
3. Realizacja projektu Cyfrowa Piaskownica Administracji.
4. Realizacja hackathonów wspierających startupy wykorzystujące dane do realizacji usług, produktów.
5. Wypracowywanie rekomendacji Laboratorium Otwartych Danych ułatwiających administracji rządowej i samorządowej otwieranie danych lub podnoszenie ich jakości.
6. Rozpowszechnianie aplikacji i usług bazujących na danych stworzonych przez administrację rządową, samorządową, sektor prywatny w portalu Dane.gov.pl.

5.3.4. Bezpieczeństwo danych

Rozwój społeczny i gospodarczy w coraz większym stopniu jest zależny od szybkiego i nieskrępowanego dostępu do informacji oraz od jej wykorzystania w zarządzaniu, produkcji,

⁹⁵) <https://dane.gov.pl/application>.

sektorze usług oraz sektorze publicznym. W ramach realizacji tego procesu szczególnym priorytetem należy objąć zagadnienie bezpieczeństwa danych udostępnianych przy wykorzystaniu systemów teleinformatycznych. Kompleksowa ochrona danych jest kluczowa dla każdego rodzaju organizacji, zarówno z sektora publicznego, jak i sektora prywatnego. Przechowywanie i udostępnianie danych w sposób nieodpowiedni, bez zachowania wymogów obowiązujących regulacji, może mieć negatywny wpływ na zachowanie ciągłości działania organizacji i jej bezpieczeństwa. W szczególności może to doprowadzić do utraty zaufania interesariuszy, których prawa i wolności mogą zostać naruszone. Jest to zatem zagadnienie bardzo istotne, które powinno rozpatrywać się na dwóch płaszczyznach: ochrony danych osobowych i zapewnienia poufności, integralności, dostępności i autentyczności danych.

Ochrona danych osobowych

Otwieranie danych do ponownego wykorzystywania często wiąże się z przetwarzaniem danych osobowych zawartych w zasobach publicznych i może rodzić ryzyko wkroczenia w sferę autonomii informacyjnej jednostek. Udostępniając dane do ponownego wykorzystywania, ich dostawca powinien dokonać weryfikacji zakresu udostępnianych danych i rozważyć, czy istnieje konieczność wprowadzenia dodatkowych mechanizmów ograniczenia ryzyka dla praw lub wolności osób, którego źródłem może być naruszenie zasady minimalizacji danych. Takim mechanizmem może być np. określenie warunków ponownego wykorzystania informacji sektora publicznego. Kluczowe jest przeprowadzenie oceny ryzyka już w fazie projektowania przetwarzania danych osobowych (art. 25 RODO) oraz oceny skutków dla ochrony danych osobowych (art. 35 RODO). Na bieżąco należy wdrożyć odpowiednie środki techniczne i organizacyjne, aby zapewnić stopień bezpieczeństwa odpowiadający temu ryzyku. Są to wymogi, które wynikają z RODO. Dotyczą administratorów zarówno sektora prywatnego, jak i sektora publicznego, którymi mogą być zarówno podmioty udostępniające dane, jak i użytkownicy danych. Na dostawcy danych spoczywa obowiązek pogodzenia zachowania wartości informacyjnej zbioru danych i ich potencjału do ponownego wykorzystywania z zapewnieniem ochrony prywatności. Konieczność zapewnienia ochrony danych osobowych nie powinna stać na przeszkodzie otwieraniu danych. Proces otwierania danych powinien odbywać się z uwzględnieniem przepisów o ochronie danych osobowych. Podjęcie opisanych działań oraz zastosowanie odpowiednich technik anonimizacji i pseudonimizacji danych pozwolą na pogodzenie obu wartości, tj. prywatności osób fizycznych i prawa do ponownego wykorzystywania danych. Wymaga to jednak od dostawców danych oraz użytkowników posiadania odpowiedniej wiedzy o rodzajach technik anonimizacji i pseudonimizacji oraz o zasadach wynikających z RODO. W tym celu rekomenduje się stosowanie standardu bezpieczeństwa⁹⁶⁾, który jest zbiorem zaleceń w zakresie zapewnienia bezpieczeństwa prywatności w procesie otwierania danych publicznych. Może on również służyć jako wytyczna w procesach udostępniania danych z zasobów dostawców danych spoza sektora publicznego.

Cyberbezpieczeństwo a otwieranie danych

Relacje społeczne są kształtowane z wykorzystaniem technologii cyfrowych tworzących cyberprzestrzeń, a usługi w sieci Internet stały się narzędziem do wpływania na zachowania grup społecznych, a także oddziaływania w sferze politycznej⁹⁷⁾. Niezakłócone działania

⁹⁶⁾ Standard bezpieczeństwa. Standardy otwartości danych opracowane przez Ministerstwo Cyfryzacji <https://dane.gov.pl/pl/knowledgebase/useful-materials/standardy-otwartosci-danych>.

⁹⁷⁾ Uchwała nr 125 Rady Ministrów z dnia 22 października 2019 r. w sprawie Strategii Cyberbezpieczeństwa Rzeczypospolitej Polskiej na lata 2019–2024 (M.P. poz. 1037).

cyberprzestrzeni stały się jednym z podstawowych celów strategicznych w obszarze bezpieczeństwa każdego państwa. Odporność systemów informacyjnych na działania naruszające poufność, integralność, dostępność i autentyczność przetwarzanych danych lub związanych z nimi usług oferowanych przez te systemy to istota cyberbezpieczeństwa⁹⁸⁾.

Do najważniejszych korzyści płynących z bezpieczeństwa danych publikowanych w repozytoriach danych w systemach teleinformatycznych lub lokalnych serwisach internetowych należy przede wszystkim wysoki poziom bezpieczeństwa pobieranych danych/informacji, gwarantujący stabilny dostęp do wiarygodnych i niezmiennych danych od niezafałszowanych dostawców.

W celu zapewnienia bezpieczeństwa danych w systemach podmiotów publicznych jest konieczne stosowanie się do podstawowych zasad cyberbezpieczeństwa, tj.:

- a) serwisy internetowe, lokalne repozytoria danych publicznych powszechnie dostępne w sieci teleinformatycznej udostępniające informacje są zobowiązane do ujęcia tych serwisów, repozytoriów w politykach, zasadach i procedurach oraz instrukcjach stanowiskowych tworzonych w ramach systemu zarządzania bezpieczeństwem informacji wymaganego przepisem § 20 KRI oraz zastosować zabezpieczenia adekwatne do oszacowanego poziomu ryzyka związanego z udostępnieniem. Niezbędne jest wypracowanie zasad doboru zabezpieczeń przed zagrożeniami bezpieczeństwa danych i systemów teleinformatycznych,
- b) przy szacowaniu ryzyka, o którym mowa w lit. a, należy uwzględnić m.in. następujące zagrożenia:
 - atak typu DDoS i z nim związane utrudnienia dostępu do danych,
 - podszycie się pod podmiot udostępniający, np. przez atak na serwer DNS albo metodą Man in the Middle,
 - zmiana (podmiana) informacji/danych podlegających udostępnieniu,
 - atak na wewnętrzne struktury sieci podmiotu udostępniającego,
 - inne metody ataku, możliwe do przeprowadzenia w konkretnym środowisku informatycznym podmiotu udostępniającego dane,
- c) w celu zapewnienia autentyczności miejsca pochodzenia oraz integralności danych podczas transmisji API dostawca danych powinien posługiwać się certyfikatem klasy EV (Extended Validation Certificate), który informuje o bezpieczeństwie i gwarantuje autentyczność serwisu. Takim samym certyfikatem powinna posługiwać się strona internetowa podmiotu, przez którą udostępniane są pliki,
- d) architektura sieci podmiotu publicznego powinna zawierać strefę DMZ (DeMilitarized Zone), z której świadczone jest udostępnianie danych. Jest to wydzielony obszar w strukturze sieci komputerowej danego podmiotu izolowany zaporami sieciowymi (firewall) zarówno od wnętrza sieci podmiotu, jak i od Internetu. Ze strefy DMZ powinny być na zewnątrz świadczone usługi typu WEB. DMZ zapewnia z jednej strony bezpieczeństwo informacji serwowanych przez usługi WEB, a z drugiej strony bezpieczeństwo istotnym zasobom informacyjnym znajdującym się wewnątrz sieci podmiotu,
- e) dane udostępniane przez API nie powinny pochodzić z produkcyjnej bazy danych podmiotu, a z pewnego rodzaju hurtowni danych, w której użytkownik zewnętrzny nie ma możliwości dokonywania jakichkolwiek zapisów.

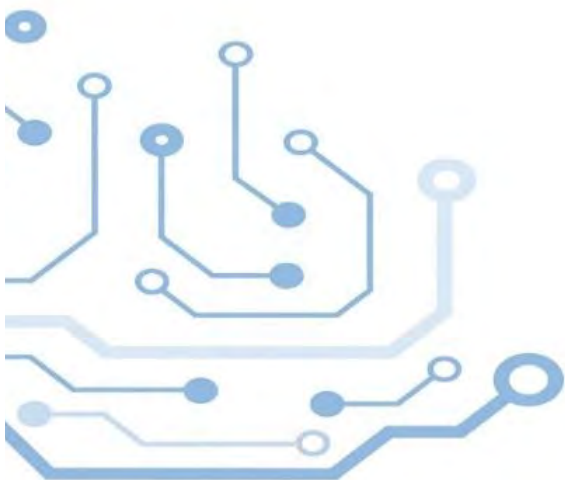
⁹⁸⁾ Definicja ujęta w ustawie z dnia 5 lipca 2018 r. o krajowym systemie cyberbezpieczeństwa (Dz.U. z 2020 r. poz. 1369).

Działania:

1. Stymulowanie badań i rozwoju w obszarze cyberbezpieczeństwa przez kontynuację dedykowanych programów badawczych wspólnie z Narodowym Centrum Badań i Rozwoju.
2. Rozwijanie bazy wiedzy o tematyce cyberbezpieczeństwa dla różnych grup użytkowników technologii informacyjnych w portalu GOV.pl <https://www.gov.pl/web/baza-wiedzy/cyberbezpieczenstwo>.
3. Prowadzenie działań informacyjno-edukacyjnych w celu zwiększenia kompetencji dostawców danych w portalu Dane.gov.pl z zakresu cyberbezpieczeństwa i ochrony danych.
4. Upowszechnianie standardu bezpieczeństwa.
5. Rozwijanie bazy wiedzy oraz promowanie dobrych praktyk w portalu Dane.gov.pl w zakresie ochrony danych osobowych, ze szczególnym uwzględnieniem zagadnień dotyczących zapewnienia tej ochrony w procesie otwierania danych.

ZALECENIA

- Udostępnianie w portalu Dane.gov.pl danych w sposób otwarty, zgodnie z filarami otwartości danych.
- Stosowanie standardu bezpieczeństwa opracowanego przez Ministerstwo Cyfryzacji.
- Stosowanie podstawowych zasad cyberbezpieczeństwa.



5.4. Cel czwarty – Stymulowanie rynku ponownego wykorzystywania zasobów kultury i danych naukowych

5.4.1. Zwiększanie dostępności zasobów kultury

Postępujący rozwój nowych technologii zmienia gospodarkę, społeczeństwo i oddziałuje na wszystkie sektory działalności, w tym również na obszar kultury i dziedzictwa narodowego. Zasadniczym elementem tych zmian są dane, których liczba sukcesywnie powinna wzrastać. Innowacyjne rozwiązania oparte na dostępnych zasobach przynoszą obywatelom korzyści. Coraz powszechniejszy dostęp do zasobów kultury znacząco zwiększa możliwości aktywnego w niej uczestnictwa. Narzędzia, jakie daje współczesna technologia, pozwalają zainteresowanym na stały dostęp do zasobów, danych i wytworów dziedzictwa kultury. Coraz częściej możemy odwiedzać wirtualne muzea, galerie, a także biblioteki, bez wychodzenia z domu. Jest to szczególnie istotne dla tych, którzy z różnych względów nie mogą aktywnie korzystać z dóbr kultury i dziedzictwa narodowego, np. ze względu na ograniczenia wynikające ze stanu epidemii.

Mając na względzie powyższe, rekomenduje się podejmowanie działań zwiększających dostępność zasobów kultury w portalu Dane.gov.pl oraz możliwość ich ponownego wykorzystywania.

Inicjatywą realizującą powyższe cele jest np. projekt Kronik@ – Krajowe Repozytorium Obiektów Kultury i Nauki (instytucje współpracujące: Ministerstwo Edukacji i Nauki, Ministerstwo Kultury i Dziedzictwa Narodowego). Projekt umożliwi podmiotom publicznym udostępnianie ich zasobów z obszaru nauki i kultury w sposób otwarty z uwzględnieniem obowiązujących przepisów prawa. Projekt zakłada zapewnienie interoperacyjności z istniejącymi krajowymi i zagranicznymi systemami, będącymi repozytoriami zasobów kultury i nauki.

W celu ułatwienia wyszukiwania obiektów kultury i nauki z repozytorium Kronik@ oraz zasobów danych z portalu Dane.gov.pl planuje się integrację wyszukiwarek tych portali. Działanie to potencjalnie zwiększy wykorzystywanie danych i obiektów z ww. portali.

Działania:

1. Realizacja projektu Kronik@ – Krajowego Repozytorium Obiektów Kultury i Nauki.
2. Integracja wyszukiwarki portalu Dane.gov.pl z wyszukiwarką repozytorium KRONIK@.
3. Integracja wyszukiwarki repozytorium KRONIK@ z wyszukiwarką serwisu Szukaj w Archiwach.

5.4.2. Zwiększenie dostępności zasobów nauki

Krajowa polityka otwartego dostępu

Otwarty dostęp do publikacji naukowych i otwarta nauka zyskują coraz większe znaczenie w kontekście polityki naukowej na świecie. Polityki otwartego dostępu (ang. open access) przyjęło wiele wiodących uniwersytetów, instytucji badawczych oraz podmiotów finansujących badania naukowe. Otwarty dostęp jest obowiązkowym wymogiem w europejskim programie Horyzont 2020 finansowanym przez UE. W coraz większym

stopniu otwierane i współdzielone są również dane badawcze. Główne korzyści modelu otwartego dostępu dla społeczeństwa to m.in.: lepszy obieg cyfrowych treści naukowych (publikacje, dane badawcze), podnoszenie jakości badań (przeciwdziałanie plagiatom przez kontrolę ze strony środowiska naukowego), promocja nauki w społeczeństwie, efektywniejsze wydatkowanie środków publicznych.

Ministerstwo Nauki i Szkolnictwa Wyższego opublikowało dnia 23 października 2015 r. dokument pt. „Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce”. Dokument ma charakter zaleceń dotyczących wprowadzania otwartego dostępu przez podmioty finansujące badania (NCN, NCBR, MNiSW), jednostki naukowe, uczelnie oraz wydawców, w tym wydawców czasopism naukowych.

Najważniejsze rekomendacje zawarte w „Kierunkach rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce” to:

- a) opracowanie i przyjęcie polityk w zakresie otwartego dostępu przez podmioty tworzące system szkolnictwa wyższego i nauki, o których mowa w art. 7 ust. 1 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce,
- b) wyznaczenie przez kierowników jednostek naukowych i uczelni pełnomocników do spraw otwartości danych,
- c) przechodzenie czasopism naukowych do modeli otwartych,
- d) udostępnienie rozpraw doktorskich w otwartych repozytoriach,
- e) monitorowanie i raportowanie do MNiSW (od 1 stycznia 2021 r. do MEiN) postępów we wdrażaniu OD, w tym systematyczne analizowanie liczby publikacji powstających w danej jednostce naukowej lub uczelni w celu określenia proporcji publikacji w OD w stosunku do wszystkich publikacji,
- f) organizowanie szkoleń w zakresie OD dla pracowników naukowych i doktorantów danej jednostki naukowej lub uczelni,
- g) uwzględnienie doświadczeń i potencjału bibliotek naukowych, które często koordynują proces redakcji i deponowania publikacji naukowych w repozytoriach.

Ze względu na rosnące znaczenie wiedzy, nauki i danych dla rozwoju społeczeństwa i gospodarki niezbędna są kontynuacja i intensyfikacja działań w zakresie otwartego dostępu, w szczególności do danych nieprzetworzonych do postaci analiz, raportów, prowadzonych przez MNiSW, a od 1 stycznia 2021 r. przez MEiN, agencje finansujące badania, a także uczelnie i jednostki naukowe oraz wszystkich naukowców, którzy realizują projekty badawcze, gromadzą dane badawcze i publikują artykuły naukowe dzięki finansowaniu ze środków publicznych. Pierwszym krokiem w kierunku określenia zasad otwartego dostępu powinno być przyjęcie przez podmioty tworzące system szkolnictwa wyższego i nauki własnej instytucjonalnej polityki otwartego dostępu. W celu wsparcia działań związanych z otwartym dostępem Ministerstwo Edukacji i Nauki udostępnia na swoich stronach internetowych materiały i publikacje dotyczące OD, w tym również przykładowy wzór instytucjonalnej polityki otwartego dostępu, który może być wykorzystany i dostosowany do własnych potrzeb przez ww. podmioty.

Mając na uwadze wspieranie przez MEiN polityki otwartych danych oraz biorąc pod uwagę sytuację epidemiczną na terenie Rzeczypospolitej Polskiej, która stała się motorem do jeszcze bardziej dynamicznego rozwoju szeroko rozumianych zasobów cyfrowych, MEiN planuje przeprowadzić badanie w zakresie istnienia potrzeby aktualizacji obecnie obowiązujących zaleceń MEiN w obszarze otwartych danych.

Udział Polski w pracach na rzecz utworzenia European Open Science Cloud

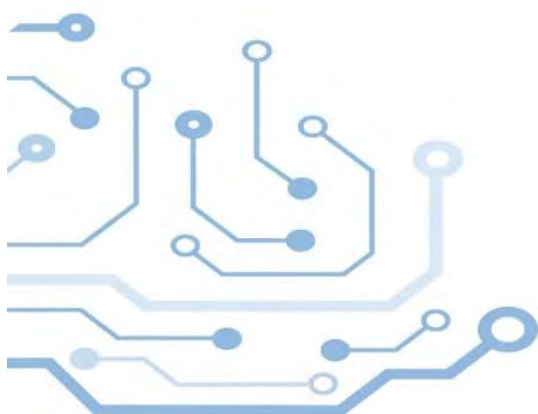
European Open Science Cloud (dalej „EOSC”) jest unijną inicjatywą, której celem jest zapewnienie Unii Europejskiej roli światowego lidera w zarządzaniu danymi badawczymi oraz zapewnienie, że europejscy naukowcy będą czerpać pełne korzyści z nauki opartej na danych. EOSC ma zaoferować europejskim naukowcom i specjalistom wirtualne środowisko umożliwiające przechowywanie, udostępnianie i ponowne wykorzystywanie danych badawczych ponad granicami i dyscyplinami naukowymi oraz oferujące bogaty wachlarz powiązanych usług. Usługi zapewniane przez EOSC mają być dostępne niezależnie od granic i dyscyplin naukowych. Będą one oparte na federowaniu istniejącej infrastruktury danych naukowych. W dniu 29 lipca 2020 r. powołana została asocjacja EOSC AISBL i Polska przystąpiła do partnerstwa EOSC, wskazując swojego reprezentanta – Narodowe Centrum Nauki. Dodatkowe zgłoszenia przystąpienia do asocjacji EOSC wystosowało wiele krajowych uczelni, instytutów badawczych i instytutów PAN.

Działania:

1. Aktualizacja krajowej polityki otwartego dostępu do danych i publikacji naukowych.
2. Udział Polski w pracach na rzecz utworzenia European Open Science Cloud.

ZALECENIE

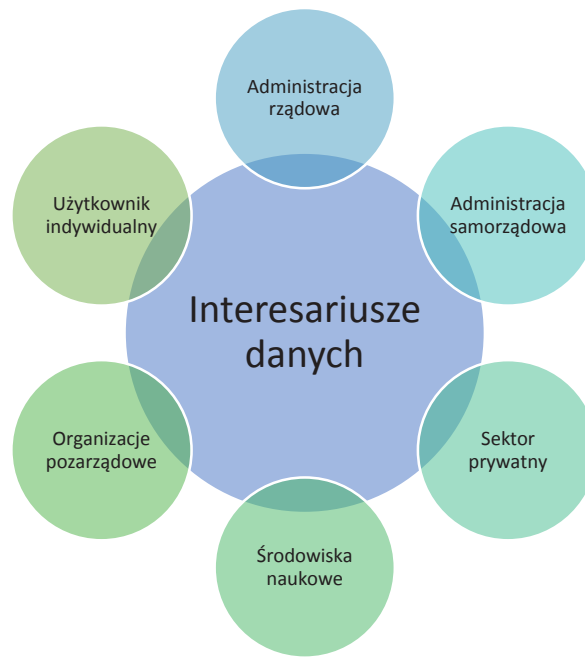
- Stosowanie zaktualizowanych zaleceń zawartych w dokumencie nt. otwartego dostępu do danych naukowych opracowanych przez Ministerstwo Edukacji i Nauki



5.5. Cel piąty – Wzmocnienie współpracy z krajowymi i zagranicznymi interesariuszami danych

5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami

Rysunek 8. Interesariusze danych



Źródło: Opracowanie własne

Krajobraz interesariuszy otwartych danych nie jest jednorodny. Składa się z różnych podmiotów, o potencjalnie różnych zainteresowaniach.

Jednostki samorządu terytorialnego

Nowelizacja ustawy z dnia 6 września 2001 r. o dostępie do informacji publicznej poszerzyła zakres podmiotów zobowiązanych do udostępniania posiadanych zasobów informacyjnych w portalu Dane.gov.pl o jednostki samorządu terytorialnego. Dane będące w posiadaniu samorządów mają ogromną wartość informacyjną i mogą stanowić cenne źródło informacji dla różnych grup interesariuszy.

Analiza polskich lokalnych repozytoriów otwartych danych prowadzi do wniosku, że w coraz większym stopniu lokalne portale wspierane są solidnymi zasobami otwartych danych. Miasta korzystają z otwartych danych w celu stawienia czoła typowym wyzwaniom miejskim, takim jak zatłoczenie i zanieczyszczenie, a także w celu poprawy jakości miejskich usług publicznych oraz interakcji między samorządem lokalnym a obywatelami. Miasta i obszary miejskie stały się zagorzałymi użytkownikami i źródłami danych, głównie pod ogólnym pojęciem Smart City (inteligentnych miast). Otwarte dane są uważane za element definiujący inteligentne miasta, ponieważ dane są niezbędne dla miast i ich mieszkańców⁹⁹⁾. Jak wynika z raportu Komisji Europejskiej pt. „Open Data in Cities”¹⁰⁰⁾, duże miasta w Unii Europejskiej

⁹⁹⁾ The Open Rural Data Gap, EDP, 05.2020 r., <https://www.europeandataportal.eu/en/highlights/open-rural-data-gap>.

¹⁰⁰⁾ Analytical Report 4: Open Data in Cities, EDP 07.2020, https://www.europeandataportal.eu/sites/default/files/edp_analytical_report_n4_-_open_data_in_cities_v1.0_final.pdf.

mają mechanizm koordynacji z poziomem krajowym. Jest to ważne, ponieważ ułatwia interoperacyjność różnych systemów, a dzieląc się najlepszymi praktykami i doświadczeniami łatwiej pokonać występujące bariery. Bariery napotymane przez portale są bardzo podobne do barier napotypanych na poziomie krajowym, przy czym bariera techniczna jest najbardziej trwała. Koordynacja na szczeblu krajowym jest ważna dla powodzenia strategii otwartych danych, ponieważ na poziomie krajowym można wprowadzić krajowe wytyczne i wspólne podejścia stosowane przez inne szczeble administracji publicznej. Pozwala to nie tylko na płynniejszą współpracę różnych systemów (interoperacyjności różnych systemów), ale może także pomóc mniejszym społecznościom z mniejszą liczbą zasobów w ich podróży po otwartych danych¹⁰¹⁾.

W związku z powyższym należy nawiązać współpracę między lokalnymi portalami otwartych danych a portalem krajowym, jakim jest portal Dane.gov.pl. Nawiązanie dialogu ułatwi w przyszłości integrację portali lokalnych z portalem Dane.gov.pl. W dobie powszechnego stosowania zaawansowanych technologii jest zasadne automatyczne zbieranie danych/metadanych przez portal Dane.gov.pl z lokalnych repozytoriów, które posiadają API (na zasadzie tzw. harvestera). Zaletą takiej integracji danych jest oferowanie odwiedzającym jednego punktu dostępu do danych w Polsce, należących do administracji rządowej i samorządowej. Aby zapewnić sukces krajowej inicjatywy otwartych danych, zasadne jest udostępnianie danych/metadanych lokalnych w portalu Dane.gov.pl, którego dane są publikowane również na zasadzie automatycznego udostępniania w EDP. Obecnie portal Dane.gov.pl pobiera automatycznie metadane tylko z jednego portalu miejskiego otwartych danych, z Gdańska¹⁰²⁾.

Jak pokazuje raport analityczny z 2019 r. EDP zatytułowany „Umożliwienie inteligentnego rozwoju obszarów wiejskich: luka w otwartych danych”¹⁰³⁾, istnieje luka odnośnie do danych na obszarach wiejskich. Jeżeli jakieś samorządy wiejskie dostarczają otwarte dane, to i tak niewiele wiadomo na temat tego, w jaki sposób otwarte dane wpływają na społeczności wiejskie. Raport zaznacza, iż potrzeba instytucjonalizacji i odpowiednich umiejętności w zakresie otwartych danych pozostaje kluczową barierą w publikowaniu i wykorzystywaniu otwartych danych na obszarach wiejskich. Bardzo ważne jest, aby ludność wiejska i sektor rolny nie były wyłączone z korzyści płynących z gospodarki opartej na danych. Aby tak się stało, inteligentna wieś musi stać się czymś bardziej realnym niż kilka odizolowanych projektów i niezadowolająca replika inteligentnych miast. Największa szansa leży w inteligentnych regionach, które tworzą niezbędne powiązania między miastem a wsią¹⁰⁴⁾. Istotna jest również infrastruktura internetu szerokopasmowego oraz podnoszenie cyfrowych umiejętności mieszkańców obszarów wiejskich, w tym rolników i administracji samorządowej. W związku z obowiązkiem publikowania danych przez administrację samorządową oraz mając na względzie rekomendacje ujęte w raporcie, jest pożądane m.in.:

- a) podkreślanie znaczenia zestawów danych rolniczych o wysokiej wartości (np. statystyki handlu rolnego, budżet na rozwój rolnictwa i obszarów wiejskich),
- b) wspieranie lokalnych liderów w ustanawianiu i instytucjonalizacji procesów i praktyk w zakresie otwartych danych w administracji wiejskiej,
- c) tworzenie hackatonów lub wyzwań ukierunkowanych konkretnie na problemy obszarów wiejskich i rolnictwa¹⁰⁵⁾.

¹⁰¹⁾ Analytical Report 4: Open Data in Cities, EDP.

¹⁰²⁾ Portal otwartych danych Gdańska, <https://dane.gov.pl/institution/154,ztm-gdansk>, <https://ckan.multimediagdansk.pl/>.

¹⁰³⁾ Enabling Smart Rural: The Open Data Gap, EDP, 12.2019 r. <https://www.europeandataportal.eu/en/highlights/enabling-smart-rural-open-data-gap>.

¹⁰⁴⁾ „Analytical Report 14: Enabling Smart Rural: The Open Data Gap”, EDP, 07. 2020.

¹⁰⁵⁾ Tamże.

Konsekwencje przepaści cyfrowej i luki w danych na obszarach wiejskich są jeszcze bardziej uderzające podczas pandemii COVID-19. Brak otwartych zestawów danych i infrastruktury danych (takich jak portale regionalne) utrudnia ekspertom do spraw zdrowia i administracji rządowej mapowanie rozprzestrzeniania się COVID-19 na obszarach wiejskich¹⁰⁶⁾.

Zagadnienie dotyczące rolnictwa zostało również dostrzeżone w ESD.

Obecnie w portalu Dane.gov.pl widnieje niska liczba dostawców danych z jednostek samorządu terytorialnego, tj. ok. 45 instytucji. Z uwagi na obecną sytuację jest pożądane nawiązanie bliższej współpracy między administracją rządową a jednostkami samorządu terytorialnego w celu udostępniania większej liczby zasobów danych w portalu Dane.gov.pl. W związku powyższym zakłada się podjęcie szeregu działań inicjujących i wspierających publikowanie otwartych danych przez administrację samorządową w portalu Dane.gov.pl.

Podmioty prywatne

Zmiany wprowadzone w udip uzupełniły krąg podmiotów, które mogą udostępniać dane, o podmioty spoza sektora publicznego. Wprowadzając możliwość dobrowolnego publikowania danych prywatnych w portalu Dane.gov.pl, w ten sposób ustawodawca podjął aktywność w celu stworzenia linii współpracy m.in. na w relacji biznes – administracja (B2G). Obecnie w portalu Dane.gov.pl żaden z podmiotów z sektora prywatnego nie publikuje swoich danych. W związku z powyższym jest pożądane nawiązanie i rozwój współpracy między administracją rządową a sektorem prywatnym.

Organizacje pozarządowe

Współpraca z organizacjami pozarządowymi w obszarze polityki zarządzania dostępem do informacji publicznej i ponownego wykorzystywania informacji sektora publicznego oraz stosowania otwartych standardów dostępu będzie prowadzona przez wszystkich realizatorów Programu. Będzie polegała na:

- a) wymianie doświadczeń, m.in. w zakresie zidentyfikowania najbardziej potrzebnych zestawów danych oraz w ich przygotowaniu do publikacji, rekomendacji najbardziej użytecznych form przygotowania i prezentacji danych,
- b) wzajemnym informowaniu o planowanych kierunkach działalności przez organy administracji rządowej i organizacje pozarządowe,
- c) organizacji wspólnych konferencji i seminariów.

Środowisko naukowe

Warto wykorzystać aktywność środowiska naukowego w procesie otwierania danych z uwagi na jego kompetencje w zakresie analizy otwierania danych, udostępniania i przechowywania danych. Pożądane byłoby włączenie środowiska naukowego w udostępnianie szerszego zakresu danych badawczych niż tylko te, na które wskazuje dyrektywa 2019/1024/UE.

Użytkownik indywidualny

Analizując praktyki europejskich portali otwartych danych, można wskazać, że istnieją różne metody dotarcia do obywateli lub nawiązania z nimi kontaktu. Portal Dane.gov.pl korzysta z takich metod i narzędzi jak: media społecznościowe, dystrybucja cyklicznych biuletynów informacyjnych (ang. newsletter), organizowanie hackathonów. Ponadto, aby zwiększyć świadomość tego, co można dokonać z danymi, portal przedstawia przykłady aplikacji bazujących na danych oraz wizualizuje dane na mapach oraz wykresach. Portal zawiera także wiele funkcji mających na celu kontakt z użytkownikami, przekazywanie uwag i opinii,

¹⁰⁶⁾ Tamże.

a także publiczne konsultacje online, zbierające informacje, w jaki sposób można uczynić portal bardziej przyjaznym dla użytkownika.

Konieczne jest stosowanie wszystkich powyższych metod przez zespół odpowiedzialny za portal Dane.gov.pl w celu wzmocnienia współpracy z użytkownikami portalu. Chcąc dotrzeć do jeszcze większej liczby użytkowników, zasadne jest organizowanie wydarzeń w celu poznania potrzeb i oczekiwań potencjalnych użytkowników.

Administracja publiczna

Obecnie funkcjonuje już tzw. sieć pełnomocników do spraw otwartości danych, powołanych w każdym resorcie. Natomiast jest ona ograniczona wyłącznie do poziomu administracji rządowej. Potrzeba poszerzenia sieci pełnomocników na wielu poziomach instytucjonalnych pojawiła się na bazie doświadczeń z wdrażania PODP i wymaga ustawowego umocowania. Zakłada się jej poszerzenie na zasadzie fakultatywności o pełnomocników do spraw otwartości danych powołanych w innych instytucjach niż podmioty zobowiązane. Działanie to pozwoli na ujednoczenie działań w zakresie wdrażania polityki otwartości na różnych szczeblach instytucjonalnych państwa i przyczyni się do zwiększenia w Polsce stopnia otwartych danych możliwych do ponownego wykorzystywania oraz poprawy ich jakości.

Działania:

1. Poszerzenie sieci pełnomocników na wielu poziomach instytucjonalnych (przez powoływanie fakultatywnych pełnomocników do spraw otwartości danych).
2. Wzmocnienie współpracy między portalem Dane.gov.pl a lokalnymi portalami otwartych danych (przez realizację działań informacyjnych, edukacyjnych, np. warsztaty, udostępnianie materiałów edukacyjnych).
3. Organizacja inicjatyw na rzecz wzmocnienia współpracy administracji rządowej z JST w zakresie otwierania danych (przez działania informacyjne, szkoleniowe, promocyjne, np. warsztaty, szkolenia, konferencje, materiały edukacyjne).
4. Realizacja działań na rzecz obszarów wiejskich z zakresu otwierania danych.
5. Realizacja „Dekalogu wyzwań cyfrowych na lata 2020–2023” z zakresu otwartych danych.
6. Nawiązanie i rozwój współpracy między administracją rządową a sektorem prywatnym w zakresie otwartych danych (przez działania informacyjne, promocyjne, np. konferencje, hackathony, media społecznościowe, newsletter, materiały edukacyjne).
7. Kontynuacja współpracy z organizacjami pozarządowymi w obszarze otwierania danych (przez działania informacyjne, promocyjne, konferencje, warsztaty, materiały edukacyjne).

8. Nawiązanie współpracy ze środowiskiem naukowym w zakresie publikowania danych na Dane.gov.pl (przez działania informacyjne, np. warsztaty, wymianę doświadczeń).
9. Włączenie użytkowników indywidualnych we współzarządzanie danymi w portalu Dane.gov.pl (przez udostępnianie w portalu aplikacji, publikowanie artykułów w bazie wiedzy i nowych zbiorów danych zaproponowanych przez użytkowników).

5.5.2. Otwarcie portalu Dane.gov.pl na nowych dostawców danych

Portal otwartych danych pełni funkcję centralnego repozytorium danych publicznych, które jest powszechnie dostępne w sieci teleinformatycznej, dlatego jest pożądane rozszerzenie obecnego zakresu podmiotów zobowiązanych do udostępniania w nim posiadanych informacji sektora publicznego. Chodzi o objęcie podmiotów wskazanych w art. 3 upwisp obowiązkiem publikowania danych w portalu Dane.gov.pl. W następstwie tego działania katalog podmiotów, które są zobligowane do udostępniania w portalu otwartych danych, byłby tożsamy z katalogiem podmiotów zobligowanych do udostępniania lub przekazywania informacji sektora publicznego w celu ich ponownego wykorzystywania. Portal Dane.gov.pl jest również przestrzenią do udostępniania danych prywatnych.

Dane będące w posiadaniu ww. podmiotów mają ogromną wartość informacyjną i mogą stanowić cenne źródło informacji, na bazie których powstaje coraz więcej nowoczesnych produktów. Sam portal wzbogaciłby się o nowe zasoby stanowiące jeden z fundamentów innowacji i niosące potencjał ich wykorzystywania w usługach, aplikacjach, modelach biznesowych i nowych technologiach.

Na podstawie otwartych danych udostępnionych przez sektor publiczny zbudowano wiele usług, aplikacji. Niektóre z nich są opublikowane w portalu Dane.gov.pl w zakładce „Aplikacje”. Gromadzenie tego typu produktów, aplikacji ma na celu informowanie społeczeństwa, przedsiębiorców o korzyściach, jakie można uzyskać dzięki danym, oraz inspirowanie do kolejnych pomysłów. Taka funkcja pokazuje, co można zrobić z poszczególnymi zestawami danych. Jest pożądane, aby portal Dane.gov.pl pełnił rolę agregatora wszystkich aplikacji wykorzystujących otwarte dane opracowanych przez przedsiębiorców, administrację rządową oraz samorządową.

Ponadto zakładka „Baza wiedzy” jest przestrzenią, w której dostawcy danych mogą dzielić się z innymi użytkownikami informacjami dotyczącymi otwierania i wymiany danych.

Działanie:

1. Zachęcanie nowych podmiotów, wskazanych w rozdziale 5.5.1., do publikowania danych w portalu Dane.gov.pl.
2. Zachęcanie do udostępniania w portalu Dane.gov.pl aplikacji opracowanych na podstawie otwartych zbiorów danych.

5.5.3. Wzmocnienie współpracy z europejskimi interesariuszami danych oraz prowadzenie aktywnej polityki otwartych danych na arenie międzynarodowej

Polska aktywnie uczestniczy w pracach na rzecz otwierania danych na poziomie Unii Europejskiej i na arenie międzynarodowej, w szczególności w ramach OECD. We współpracę zaangażowani są interesariusze otwierania danych w celu konsultacji różnego rodzaju dokumentów, uzyskania wsparcia merytorycznego i dalszej dystrybucji informacji. Kluczową rolę w tym aspekcie odgrywa sieć pełnomocników do spraw otwartości danych.

Unia Europejska

a) Public Sector Information Group

To grupa ekspercka działająca przy Komisji Europejskiej. Zadaniem grupy roboczej jest wymiana informacji i dobrych praktyk w zakresie ponownego wykorzystywania danych, portali otwartych danych oraz wdrażania europejskich instrumentów, w tym dyrektyw o ponownym wykorzystywaniu informacji sektora publicznego,

Istotnym przedmiotem prac jest cykliczne badanie „Open Data Maturity”¹⁰⁷⁾.

b) Komitet do spraw Otwartych Danych i Ponownego Wykorzystywania Informacji Sektora Publicznego

Zadaniem Komitetu jest wsparcie Komisji Europejskiej w przygotowaniu projektu aktu wykonawczego dotyczącego listy danych o wysokiej wartości (zgodnie z dyrektywą 2019/1024/UE),

c) Grupa Robocza do spraw Telekomunikacji i Społeczeństwa Informacyjnego, działająca przy Radzie Unii Europejskiej

Rzeczpospolita Polska aktywnie wspiera prace prowadzone przez Stałe Przedstawicielstwo Rzeczypospolitej Polskiej przy Unii Europejskiej w ramach Grupy Roboczej do spraw Telekomunikacji i Społeczeństwa Informacyjnego. W zakresie problematyki dotyczącej danych istotny przedmiot prac Grupy będą stanowić w najbliższych latach inicjatywy podejmowane w obszarach wskazanych jako priorytetowe w ESD. Strategia zawiera plany działań, których celem jest budowa jednolitego rynku danych, który umożliwiając rozwój innowacyjnych procesów, produktów i usług, ma się przyczynić do zwiększenia konkurencyjności Europy na świecie.

OECD

a) Open Government Data

Jest to grupa ekspercka działająca przy OECD. Celem grupy jest wymiana opinii i doświadczeń w zakresie otwierania danych przez administracje publiczne państw członkowskich OECD.

Istotnym efektem prac grupy jest przygotowywany przez OECD co dwa lata OURdata Index¹⁰⁸⁾, czyli ranking otwartości danych. Urząd obsługujący ministra właściwego do spraw informatyzacji bierze udział w OECD Open Government Data Survey, na podstawie którego opracowywany jest ranking,

b) Grupa e-leaderów do spraw sektora publicznego opartego na danych

¹⁰⁷⁾ <https://www.europeandataportal.eu/pl/news/open-data-maturity-report-2019>.

¹⁰⁸⁾ <https://www.oecd.org/gov/digital-government/open-government-data.htm>.

Grupa opracowuje wspólne wytyczne w zakresie etyki danych w sektorze publicznym do stosowania przez państwa członkowskie. Ich celem jest podniesienie świadomości dostawców danych i podmiotów przetwarzających dane w zakresie obowiązków, jakie na nich spoczywają, i ryzyka, jakie niesie za sobą udostępnianie i przetwarzanie danych. Wytyczne mają wspierać szersze działania OECD w zakresie zarządzania danymi, w tym projekt zalecenia OECD w sprawie lepszego dostępu do danych i ich udostępniania.

Współpraca bilateralna

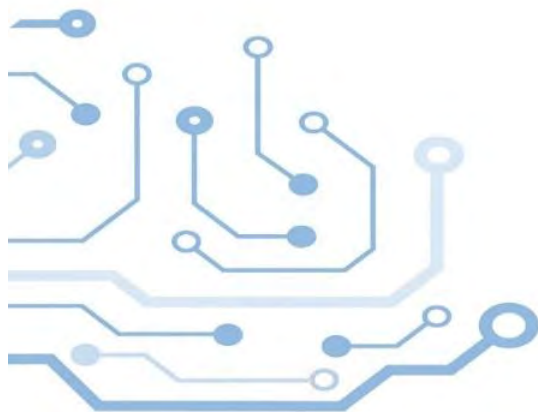
Poza współpracą wynikającą z uczestnictwa Polski w organizacjach międzynarodowych można wymienić również współpracę bilateralną z różnymi państwami członkowskimi UE mającą na celu wymianę dobrych praktyk w zakresie otwierania danych w obszarze budowy niezbędnej infrastruktury, kreowania przyjaznego otoczenia prawno-organizacyjnego i wzmacniania kompetencji. Jako przykład takiej współpracy można wymienić organizację wizyt studyjnych w krajach uznawanych – według wyżej wymienionych rankingów – za liderów w zakresie otwartości, czyli we Francji, Hiszpanii oraz Irlandii. Nawiązane podczas wizyt studyjnych kontakty z przedstawicielami miejscowych instytucji są podtrzymywane i rozwijane m.in. przez konsultacje eksperckie, współpracę w organizacji międzynarodowych wydarzeń, w obszarze otwartych danych.

Działanie:

1. Promowanie polskich rozwiązań na arenie międzynarodowej.
2. Wymiana doświadczeń i wiedzy mogących stanowić inspirację dla krajowych rozwiązań oraz dystrybucji uzyskanych informacji wśród interesariuszy Programu.

ZALECENIA

- Wdrażanie polityki otwartości danych.
- Wzmocnienie współpracy między interesariuszami danych a koordynatorem Programu.



5.6. Cel szósty – Podnoszenie wiedzy i umiejętności pracowników administracji publicznej w zakresie otwierania i zarządzania danymi oraz zwiększanie świadomości społecznej na temat potencjału otwartych danych

5.6.1. Zwiększanie umiejętności pracowników administracji publicznej w celu poprawy jakości polityk publicznych, dostarczania danych zarządczych i wspierania w podejmowaniu decyzji

Działania, które będą realizowane w tym obszarze, zostały zaprojektowane na podstawie wniosków z ewaluacji działań szkoleniowych, promocyjnych i sieciujących w obszarze otwartych danych realizowanych w okresie 2017–2020. Mają na celu niwelowanie barier kompetencyjnych zdefiniowanych w rozdziale 5.3.2 niniejszego Programu.

Z badania oceniającego realizację Programu otwierania danych publicznych (na lata 2016–2020) wśród jego realizatorów wynika, że jakość, kompletność oraz aktualność danych publicznych wciąż wymagają poprawy. Konieczne jest zatem wsparcie urzędów i ich pracowników przez działania rozwojowe i szkoleniowe w kwestiach technicznych związanych z przygotowywaniem zasobów danych w formatach otwartych.

Kontynuowane będą kursy sprofilowane pod kątem poziomu wiedzy i umiejętności, a także specyficznych aspektów tematyki otwartych danych. W takiej postaci – cyklicznie organizowanych do końca 2021 r. kursów – realizowana będzie Akademia Otwartych Danych, tj. 74-godzinne kursy w modułach prawnym i technicznym dla grup 20-osobowych. Udział w kursie poprzedzony jest badaniem poziomu wiedzy i zakończony egzaminem.

Równoległe kontynuowane będą działania szkoleniowe:

- a) na poziomie podstawowym tzw. wstęp do otwartych danych, które obejmą podstawy aspektów prawnych i technicznych oraz warsztaty na temat korzyści z otwierania i ponownego wykorzystywania danych. Skierowane będą do osób zapoznających się z tematyką,
- b) specjalistyczne z zakresu wykorzystywania otwartych danych do usprawnienia procesu decyzyjnego opartego na dowodach (przy opracowywaniu i wdrażaniu polityki publicznej), a także zaawansowanej analityki zbiorów danych oraz narzędzi typu BI w administracji publicznej. Szkolenia pracowników w obszarze inteligentnych technologii informacyjnych zapewnią instytucjom wykwalifikowane kadry, które będą umiejętnie wykorzystywały swoją wiedzę w procesach analizy danych.

Uzupełnieniem szkoleń będzie przewodnik obejmujący metody i sposoby analizy danych na potrzeby poprawy jakości polityk publicznych. Szkolenia będą skierowane do osób posiadających zaawansowane umiejętności z zakresu pracy z danymi.

Dla wszystkich rodzajów i poziomów rozwijane będą formy kształcenia on-line, docelowo w formule e-learningu. Promowane będą szeroko (także wśród administracji samorządowej wszystkich szczebli) elektroniczne materiały edukacyjne i ćwiczenia zamieszczone w „Bazie wiedzy” w portalu Dane.gov.pl.

Grupa docelowa: pracownicy administracji rządowej i samorządowej.

Rozwijane będą działania sieciujące pracowników administracji rządowej zaangażowanych w otwieranie danych publicznych, w szczególności pełnomocników do spraw otwartości danych.

Należą do nich cykliczne spotkania pełnomocników do spraw otwartości danych organizowane przez urząd obsługujący ministra właściwego do spraw informatyzacji, służące wymianie wiedzy, doświadczeń i dobrych praktyk, a także dyskusji na temat wyzwań i barier w realizacji zadań. Szczególną formą spotkań pełnomocników do spraw otwartości danych mogą być wizyty studyjne w państwach UE – liderach polityki otwartości (badaniach UE i OECD). Służą one wprowadzeniu wielu rozwiązań, które mogą usprawnić pracę sieci pełnomocników do spraw otwartości danych, a także wzmocnieniu kontaktów między ich uczestnikami i ich zaangażowaniu w realizację zadań.

Grupa docelowa: pełnomocnicy do spraw otwartości danych w ministerstwach i urzędach centralnych.

Kontynuowana będzie organizacja warsztatów dotyczących korzyści z otwierania danych publicznych, z uwzględnieniem udziału i potrzeb różnych grup docelowych. Pilotażowe, zrealizowane dotąd warsztaty, były poświęcone tematyce korzyści i potrzeb jednostek samorządu terytorialnego w kontekście otwierania danych publicznych. Forma ich prowadzenia jest oparta m.in. na studiach przypadków, poddanych moderowanej dyskusji, oraz pracy w grupach nad prototypowymi rozwiązaniami.

Grupa docelowa: pracownicy administracji rządowej i samorządowej, NGO, przedsiębiorcy.

Działania:

1. Realizacja działań szkoleniowych z obszaru zarządzania danymi i otwierania danych publicznych dla pracowników administracji rządowej i samorządowej.
2. Organizacja warsztatów z korzyści otwierania danych publicznych sprofilowanych pod kątem różnych grup odbiorców.

5.6.2. Upowszechnianie informacji na temat otwierania i ponownego wykorzystywania danych publicznych oraz podnoszenie świadomości społecznej w zakresie korzyści płynących z otwartych danych

Inicjatywy informacyjno-promocyjne w zakresie otwartości danych pozwalają na uświadomienie dostawcom danych i ich potencjalnym użytkownikom roli otwartych danych dla rozwoju gospodarczego oraz na wskazanie ich strategicznego charakteru, jako siły napędowej dla całej gospodarki.

Działania informacyjno-promocyjne w zakresie otwartości danych powinny być w pierwszej kolejności kierowane do następujących grup:

- a) obywatele – zarówno profesjonalisci wykorzystujący dane publiczne w swojej działalności zawodowej, jak i nieprofesjonalisci zainteresowani otwartymi danymi, szukający informacji w celu pogłębienia swojej wiedzy na temat aktywności administracji publicznej w danym obszarze czy korzystający z aplikacji i usług bazujących na danych udostępnionych przez API i w portalu Dane.gov.pl,

- b) przedsiębiorcy, którzy rozwijają i tworzą nowe modele biznesowe, nowe produkty i usługi w oparciu o otwarte dane publiczne,
- c) organizacje pozarządowe zajmujące się prawem do dostępu do informacji publicznej, otwartymi danymi, otwartym rządem, jawnością, transparentnością, prowadzące działalność analityczną i badawczą,
- d) środowiska naukowe, w tym przedstawiciele świata nauki wykorzystujący dane do badań i analiz,
- e) administracja – urzędnicy korzystający z danych publicznych w związku z realizacją zadań publicznych do przygotowywania analiz, dokumentów strategicznych czy oceny skutków regulacji, tworzenia usług dla obywateli, komunikowania efektów działań zarówno wewnątrz administracji, jak i w kontaktach z obywatelami.

Rozpowszechnianie korzyści płynących z ponownego wykorzystywania otwartych danych publicznych przez kampanie informacyjno-promocyjne to szansa na wykorzystanie danych w nowych dobrach i usługach. Konieczna jest realizacja działań informacyjno-promocyjnych, skierowanych do przedsiębiorców. Działania bieżące powinny być komunikowane przez media społecznościowe, umożliwiające interakcję z odbiorcą.

Należy zachęcać do korzystania z otwartych danych przez organizację maratonów programowania, zgromadzonych wokół tematu otwartych danych oraz udział w wydarzeniach o podobnym charakterze w formach partnerstwa. Wsparciem będzie tu realizacja kampanii internetowych oraz innych działań w mediach, ze szczególnym uwzględnieniem Internetu, np. krótkich filmów, infografik i materiałów prasowych, promujących otwartość danych i portal oraz innych materiałów informacyjno-promocyjnych.

Działania związane z otwartymi danymi wzmacniają przejrzystość i transparentność administracji publicznej oraz zwiększają udział obywateli w życiu publicznym. Generują ponadto oszczędności, w związku z mniejszą liczbą wniosków o dostęp i ponowne wykorzystywanie informacji sektora publicznego, i umożliwiają korzystanie wzajemnie ze swoich doświadczeń. Dlatego tak ważne jest informowanie o korzyściach z otwierania danych wewnątrz urzędów. Należy realizować wewnętrzne akcje informacyjne z wykorzystaniem ogólnodostępnych kanałów komunikacyjnych typu poczta elektroniczna, platformy służące elektronicznemu obiegowi dokumentów czy intranet, aby dotrzeć bezpośrednio do wszystkich urzędników, którzy gromadzą publiczne zasoby. Szansą na wymianę doświadczeń wśród ekspertów z różnych środowisk, w tym m.in. z administracji samorządowej, biznesu czy organizacji pozarządowych, mogą być konferencje otwartych danych. Pozwalają one zebrać najlepsze praktyki z kraju i z zagranicy, w zakresie otwierania i ponownego wykorzystywania danych.

Działania:

1. Realizacja kampanii internetowych promujących otwarte dane.
2. Organizacja konferencji/spotkań/warsztatów służących wymianie doświadczeń.

ZALECENIA

- Realizacja szkoleń zwiększających wiedzę na temat korzyści z otwierania danych i sposobów ich udostępniania.
- Organizacja działań informacyjno-promocyjnych prezentujących potencjał otwartych danych.



System koordynacji i

monitorowania Programu

6. System koordynacji i monitorowania Programu

6.1. Wskaźniki określone dla Programu

Nazwa wskaźnika	Źródło danych	Wartość (szt.)		Definicja wskaźnika
		2021 r.	2027 r. (wartości docelowe)	
Cel 1. Zwiększenie dostępności danych w portalu Dane.gov.pl na potrzeby ich ponownego wykorzystania, w tym w nowych technologiach				
Liczba zasobów danych udostępnionych w portalu Dane.gov.pl	portal Dane.gov.pl	19 000	55 000	wskaźnik wskazujący na przyrost zasobów danych dostępnych w portalu Dane.gov.pl do ponownego wykorzystywania
Liczba zasobów danych dynamicznych udostępnionych w portalu Dane.gov.pl	portal Dane.gov.pl	1750	3700	liczba zasobów danych podlegających częstym aktualizacjom lub aktualizacjom w czasie rzeczywistym
Liczba zasobów danych o wysokiej wartości udostępnionych w portalu Dane.gov.pl	portal Dane.gov.pl	2800	8000	liczba zasobów danych, których ponowne wykorzystywanie wiąże się z istotnymi korzyściami dla społeczeństwa, środowiska i gospodarki
Cel 2. Poprawa interoperacyjności i jakości danych jako czynniki kluczowe dla pełnego wykorzystania ich potencjału				

Liczba zasobów danych o zwiększonym stopniu otwartości lub poprawionej jakości zasobów danych dostępnych w portalu Dane.gov.pl	portal Dane.gov.pl	6000	22 000	zasoby danych o zwiększonym stopniu otwartości według 5-stopniowej skali otwartości Tima Bernersa-Lee i poprawionej jakości danych ułatwiających ponowne wykorzystywanie
Liczba zasobów danych udostępnionych przez API w portalu Dane.gov.pl	portal Dane.gov.pl	360	1000	liczba zasobów danych generowanych w portalu Dane.gov.pl za pomocą interfejsu programistycznego aplikacji
Cel 3. Zwiększenie intensywności wykorzystania i wymiany danych na potrzeby nowych technologii				
Liczba działań prowadzących do usuwania lub ograniczenia barier w wymianie danych	urząd obsługujący ministra właściwego do spraw informatyzacji /realizatorzy Programu	0	35	liczba podjętych inicjatyw prowadzących do ograniczenia lub usunięcia barier w udostępnianiu danych (np. działania legislacyjne, informacyjne, infrastrukturalne, organizacyjne itp.)
Cel 4. Stymulowanie rynku ponownego wykorzystywania zasobów kultury i danych naukowych w celach komercyjnych i niekomercyjnych				
Liczba zasobów kultury i nauki udostępnionych w portalu Dane.gov.pl	portal Dane.gov.pl	57 (stan na 2020.07.30)	120	liczba dokumentów z obszaru nauki i kultury w formie cyfrowej udostępnianych w portalu Dane.gov.pl
Cel 5. Wzmocnienie współpracy z interesariuszami danych				

Liczba podmiotów udostępniających swoje dane w portalu Dane.gov.pl	portal Dane.gov.pl	137 (stan na 2020.07.30)	500	podmioty publikujące swoje dane w portalu Dane.gov.pl
Cel 6. Podnoszenie wiedzy i umiejętności pracowników urzędów administracji publicznej w zakresie otwierania danych i zarządzania nimi oraz zwiększanie świadomości społecznej na temat potencjału otwartych danych				
Liczba osób objętych szkoleniami	urząd obsługujący ministra właściwego do spraw informatyzacji	60	400	liczba osób objętych szkoleniami
Liczba odbiorców działań informacyjno-promocyjnym (typu konferencje, maratony programowania, kampania w mediach)	urząd obsługujący ministra właściwego do spraw informatyzacji	600 000	1 500 000	liczba odbiorców działań realizowanych dla wszystkich kategorii

6.2. System realizacji Programu

Program otwierania danych na lata 2021–2027 jest uchwalany na okres 7 lat.

Osiągnięcie założonych celów dokumentu wymaga zapewnienia istnienia skutecznych struktur zarządzania wdrożeniem Programu, ale przede wszystkim ścisłej współpracy wszystkich jego realizatorów. Kooperacja przewidziana jest na poziomie strategicznym w ramach prac zespołu zadaniowego do spraw „Programu otwierania danych”, przy Komitecie Rady Ministrów do Spraw Cyfryzacji. Realizatorami Programu są członkowie Rady Ministrów oraz Prezes Głównego Urzędu Statystycznego, wspierani przez pełnomocników do spraw otwartości danych (działających w ramach poziomu operacyjnego). Organy i jednostki organizacyjne podległe lub nadzorowane przez członków Rady Ministrów biorą udział w realizacji Programu. Program może być realizowany przez inne niż wymienione, obligatoryjne podmioty (takie jak JST, podmioty prywatne, NGO).

Zespół zadaniowy do spraw „Programu otwierania danych”

Kluczową rolę w ocenie działań służących realizacji założeń Programu odgrywa zespół zadaniowy do spraw „Programu otwierania danych”, który pełni funkcję opiniodawczo-doradczą Komitetu Rady Ministrów do spraw Cyfryzacji. W skład zespołu wchodzi wyznaczeni przez poszczególnych ministrów sekretarze lub podsekretarze stanu, Prezes Głównego Urzędu Statystycznego oraz sekretarz stanu albo podsekretarz stanu z Kancelarii Prezesa Rady Ministrów. Członkowie zespołu kreują politykę otwierania danych. W ramach swoich zadań m.in. opiniują roczne sprawozdania z realizacji Programu, harmonogramy udostępniania danych w portalu Dane.gov.pl opracowane przez realizatorów Programu przy współpracy z pełnomocnikami do spraw otwartości danych, opiniują projekty dokumentów w zakresie związanym z otwieraniem danych, jak również przygotowują projekty rekomendacji dla Komitetu Rady Ministrów do spraw Cyfryzacji dotyczących inicjatyw związanych z otwieraniem danych.

Realizatorzy Programu oraz Pełnomocnicy do spraw otwartości danych

Zapewnienie prawidłowej i skutecznej realizacji Programu spoczywa na członkach Rady Ministrów oraz Prezesie Głównego Urzędu Statystycznego. Osobami wspierającymi realizatorów Programu są pełnomocnicy do spraw otwartości danych, tworząc tzw. sieć pełnomocników do spraw otwartości danych na poziomie administracji rządowej (w każdym ministerstwie, KPRM, GUS). Docelowo zakłada się poszerzenie sieci na zasadzie fakultatywności o pełnomocników powołanych z innych instytucji.

Do zadań realizatora Programu należy w szczególności:

- a) zapewnienie należytego wdrażania Programu,
- b) współpraca z koordynatorem Programu w zakresie wdrażania Programu,
- c) przedstawianie w każdym półroczu kalendarzowym koordynatorowi Programu harmonogramu udostępniania danych, tj. przynajmniej dwóch nowych zestawów zasobów danych lub podniesienie jakości dotychczas udostępnianych danych w portalu Dane.gov.pl lub udostępnianie jednego zestawu zbioru danych przez API,
- d) składanie koordynatorowi Programu rocznego sprawozdania z realizacji Programu w terminie do dnia 31 marca roku następującego po roku, którego dotyczy sprawozdanie.

Do zadań pełnomocników do spraw otwartości danych należy w szczególności:

- a) pozyskiwanie, nowych zasobów danych możliwych do udostępniania w portalu, będących w posiadaniu jednostki organizacyjnej, w której jest zatrudniony pełnomocnik, oraz, o ile dotyczy, w organach i jednostkach organizacyjnych podległych lub nadzorowanych przez właściwego ministra albo Prezesa Rady Ministrów,
- b) wspieranie, o ile dotyczy, realizatora Programu w wykonywaniu zadań związanych z realizacją Programu,
- c) ocena informacji sektora publicznego będących w posiadaniu jednostki organizacyjnej, w której jest zatrudniony pełnomocnik, oraz, o ile dotyczy, w organach i jednostkach organizacyjnych podległych lub nadzorowanych przez właściwego ministra albo Prezesa Rady Ministrów, pod względem jakości oraz użyteczności, do ich ponownego wykorzystywania,
- d) monitorowanie jakości, aktualności zasobów danych oraz metadanych,
- e) współpraca, w zakresie poprawy jakości informacji sektora publicznego udostępnianych w portalu, z komórkami organizacyjnymi urzędu oraz, o ile dotyczy,

- z organami i jednostkami organizacyjnymi podległymi lub nadzorowanymi przez właściwego ministra albo Prezesa Rady Ministrów,
- f) podejmowanie działań, w zakresie otwartości danych, mających na celu podnoszenie świadomości urzędu oraz, o ile dotyczy, organów i jednostek organizacyjnych podległych lub nadzorowanych przez właściwego ministra albo Prezesa Rady Ministrów.

Koordinator wdrażania Programu

Kooperatorem wdrażania Programu jest minister właściwy do spraw informatyzacji. Działalność kooperatora Programu zapewnia spójność kierunków w procesie zwiększenia liczby oraz poprawy jakości danych oraz metadanych. Do zadań kooperatora Programu należy w szczególności:

- a) promowanie jednolitej polityki otwartego dostępu do danych,
- b) wskazywanie kierunków rozwoju w zakresie otwartości danych oraz inicjowanie działań innowacyjnych w zakresie wspierania polityki otwartego dostępu do danych,
- c) harmonizacja procesów wynikających z wdrażania Programu,
- d) współpraca z realizatorami Programu,
- e) wsparcie realizatorów Programu w udostępnianiu zasobów danych i metadanych w portalu Dane.gov.pl,
- f) monitorowanie postępów w realizacji Programu,
- g) zaplanowanie i skoordynowanie działań edukacyjnych, podnoszących kompetencje cyfrowe administracji w zakresie otwartości danych,
- h) dokonywanie przeglądu Programu co dwa, cztery lata i w ostatnim roku obowiązywania Programu.

W celu zapewnienia komplementarności rozwiązań wdrażanych w ramach realizacji Programu do zadań kooperatora Programu należy również bieżąca współpraca z podmiotami administracji rządowej, z administracją samorządową, organizacjami pozarządowymi, sektorem prywatnym oraz z innymi podmiotami uczestniczącymi w działaniach w zakresie polityki otwierania danych.

Zespół Otwartych Danych

Kooperatorem Programu w realizacji bieżących zadań dotyczących wdrażania Programu wspiera zespół otwartych danych – grupa pracowników urzędu obsługującego ministra właściwego do spraw informatyzacji. Głównym zadaniem zespołu jest skoordynowanie działań związanych z funkcjonowaniem i rozwojem portalu Dane.gov.pl, stanowiącego centralny punkt dostępu do danych publicznych w Polsce, opracowywanie i upowszechnianie standardów danych, skoordynowanie polityki w zakresie ponownego wykorzystywania informacji sektora publicznego i wymiany danych wewnątrz administracji publicznej oraz z podmiotami zewnętrznymi na potrzeby rozwoju technologii służących do analizy wykorzystywania i przetwarzania danych, realizacja badań i analiz w zakresie rynku wykorzystywania i przepływu danych, upowszechnianie wiedzy w zakresie otwierania, ponownego wykorzystywania i przepływu danych oraz wzmacnianie kompetencji administracji i przedsiębiorców w zakresie dostępu do danych. Na potrzeby zwiększenia cyfrowej dostępności i użyteczności informacji sektora publicznego zespół będzie realizował projekty dofinansowane ze środków UE.

6.3. Mechanizm monitorowania realizacji Programu

System monitorowania Programu zakłada, że działania w nim opisane będą podlegać stałej i usystematyzowanej ocenie i analizie w zakresie stopnia osiągnięcia założonych wskaźników. Przyjęty sposób oceny realizacji dokumentu ma charakter dwupoziomowy. W ramach tzw. poziomu operacyjnego będą realizowane indywidualne oraz cykliczne, kolegalne spotkania z pełnomocnikami do spraw otwartości danych. Podczas nich weryfikacji będzie podlegać realizacja dokumentu w zakresie terminowości i zgodności z przyjętymi założeniami przez realizatorów Programu oraz jednostki im podległe i nadzorowane. Ponadto każdy z realizatorów Programu do dnia 31 marca każdego roku, począwszy od roku następującego po roku ustanowienia Programu, poinformuje ministra właściwego do spraw informatyzacji o rezultatach wdrożenia dokumentu w swojej instytucji, w tym planu działań na rzecz realizacji Programu otwierania danych na lata 2021–2027, który stanowi załącznik nr 1 do Programu. W pierwszym roku realizacji Programu każdy z realizatorów Programu poinformuje ministra właściwego do spraw informatyzacji o rezultatach wdrożenia dokumentu w swojej instytucji w terminie do dnia 31 sierpnia 2021 r. Informacje te posłużą do sporządzenia corocznego Raportu na temat rezultatów wdrażania Programu otwierania danych. Informacje zebrane w ramach raportu wspierać będą bieżące zarządzenia przez dostarczanie decydom aktualnych informacji na temat postępów prac oraz związanych z nimi ryzyk. Przygotowany dokument będzie opiniowany przez zespół zadaniowy do spraw „Programu otwierania danych” utworzony decyzją Przewodniczącego Komitetu Rady Ministrów do spraw Cyfryzacji, a następnie koordynator będzie przedkładał raport Radzie Ministrów w terminie do dnia 31 lipca, przy czym w pierwszym roku realizacji Programu w terminie do dnia 30 listopada 2021 r.

Po dwóch i czterech latach od obowiązywania dokumentu oraz w siódmym roku dokument będzie podlegać przeglądowi i ocenie efektów jego oddziaływania w formie ewaluacji bieżącej (on-going). Na podstawie wyników przeglądu minister właściwy do spraw informatyzacji może opracować propozycję działań korygujących lub projekt dokumentu na kolejny okres. W przypadku wystąpienia uzasadnionych okoliczności Program może być aktualizowany w innych terminach niż te, o których mowa powyżej.





Finansowanie

7. Finansowanie

Planowane koszty ogólne realizacji POD w znacznej części zostaną pokryte ze źródeł zagranicznych, w tym z Programu Operacyjnego Polska Cyfrowa na lata 2014–2020 oraz w ramach nowej perspektywy finansowej na lata 2021–2027 – Wieloletnie ramy finansowe UE. Obecnie przygotowujemy zakres krajowego programu operacyjnego zakłada finansowanie działań w zakresie zwiększenia dostępności i poprawy jakości danych w ramach celu „Cyfrowa dostępność i ponowne wykorzystanie informacji”.

Finansowanie działań w ramach realizacji POD zostało zaplanowane z projektów: Otwarte dane plus (POPC.02.03.01 00 0064/18 00) i Kronik@ Krajowe Repozytorium Obiektów Kultury i Nauki (POPC.02.03.01 IP.01 00 003/16), dofinansowanych ze środków Programu Operacyjnego Polska Cyfrowa na lata 2014–2020 w ramach II osi E – administracja i otwarty rząd, poddziałanie 2.3.1 „Cyfrowe udostępnienie informacji sektora publicznego ze źródeł administracyjnych i zasobów nauki”.

Finansowanie ze środków krajowych zostało zaplanowane głównie z części 27 budżetu. Urząd obsługujący ministra właściwego do spraw informatyzacji jako koordynator realizacji Programu w pierwszym roku jego obowiązywania przewidział w swoim budżecie środki w wysokości 140 tys. zł. Do połowy 2022 r. część działań wynikająca z Programu będzie realizowana za pomocą projektu „Otwarte dane plus”, w ramach II Osi „E-administracja i otwarty rząd” Programu Operacyjnego Polska Cyfrowa 2014–2020 (POPC).

Źródłami finansowania realizacji działań opisanych w Programie na latach 2021–2027 będą ponadto plany finansowe poszczególnych jednostek zaangażowanych we wdrażanie dokumentu – tj. realizatorów programu (członków Rady Ministrów oraz GUS). Skutki finansowe, które wynikają z przedmiotowego projektu, zostaną pokryte w ramach limitu wydatków w odpowiedniej części budżetu państwa bez konieczności ubiegania się o dodatkowe środki ponad limit dysponenta.

Szczegółowe wielkość i struktura kosztów poszczególnych przedsięwzięć będą określone w procesie inicjowania konkretnych projektów. Planowane do realizacji działania wskazane zostały w Planie działań na rzecz realizacji Programu otwierania danych na lata 2021–2027, który stanowi źródło informacji na temat planowanych inicjatyw. Z uwagi na siedmioletnią perspektywę realizacji Programu oraz zaplanowane działania, które będą zrealizowane w ramach zadań statutowych poszczególnych realizatorów Programu, nie jest możliwe wskazanie szacunkowych kosztów w rozbiciu na poszczególne działania. W ramach mechanizmu monitorowania Programu otwierania danych jest przewidziana ewaluacja określonych działań i wówczas będzie możliwe uzupełnienie Programu szczegółowymi, dodatkowymi informacjami.

Ponadto przygotowywana obecnie perspektywa finansowa UE na lata 2021–2027 zakłada powołanie nowego instrumentu finansowego, który w kolejnych latach będzie wspierał obszar cyfryzacji w państwach członkowskich. Instrumentem tym będzie Program Cyfrowa Europa (ang. Digital Europe Programme: DEP) na lata 2021–2027, który zakłada zapewnienie środków finansowych dla podmiotów publicznych i prywatnych oraz instytucji badawczych na potrzeby rozwoju technologii cyfrowych oraz innowacji, zarówno w sektorze prywatnym, jak i publicznym. DEP przewiduje wsparcie finansowe projektów z obszarów: obliczenia superkomputerowe, sztuczna inteligencja, cyberbezpieczeństwo, zaawansowane umiejętności cyfrowe, zapewnienie szerokiego wykorzystania technologii cyfrowych w gospodarce i społeczeństwie.

Załącznik nr 1

Plan działań na rzecz
realizacji Programu

Załącznik nr 1. Plan działań na rzecz realizacji Programu

Tabela 1. Plan działań

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
1	5.1.1. Dostosowanie otoczenia regulacyjnego do potrzeb rynku ponownego wykorzystywania	Uchwalenie ustawy o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego wraz z aktami wykonawczymi	legislacyjne	prace w toku	2021	urząd obsługujący ministra właściwego do spraw informatyzacji
2	5.1.1. Dostosowanie otoczenia regulacyjnego do potrzeb rynku ponownego wykorzystywania	Bieżący monitoring przepisów mających wpływ na udostępnianie danych	legislacyjne, informacyjne	2020	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
3	5.1.2. Zwiększenie dystrybucji danych w portalu Dane.gov.pl	Udostępnianie przez dostawców danych (administracja rządowa) dwa razy w roku kalendarzowym przynajmniej dwóch nowych zasobów danych lub podniesienie jakości dotychczas udostępnianych danych w portalu Dane.gov.pl lub jednego zasobu danych przez API, na podstawie tzw. harmonogramu udostępniania danych	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe, dwa razy w każdym roku	realizatorzy Programu
4	5.1.2. Zwiększenie dystrybucji danych w portalu Dane.gov.pl	Zwiększenie w portalu Dane.gov.pl liczby zestawów danych szczególnie z obszaru zdrowia, opieki, transportu, bezpieczeństwa, połączonych w sytuacjach	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe	realizatorzy Programu

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działania: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
		różnego rodzaju zagrożeń, np. stanu epidemii				
5	5.1.2. Zwiększenie dystrybucji danych w portalu Dane.gov.pl	Zachęcanie administracji samorządowej do udostępniania co najmniej raz w roku kalendarzowym nowych zasobów danych lub podniesienia jakości dotychczas udostępnionych danych w portalu Dane.gov.pl lub udostępniania jednego zasobu danych przez API, na podstawie tzw. harmonogramu udostępniania danych (przez działania informacyjne, promocyjne i szkoleniowe, np. warsztaty, szkolenia)	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe, raz w roku	urząd obsługujący ministra właściwego do spraw informatyzacji
6	5.1.2. Zwiększenie dystrybucji danych w portalu Dane.gov.pl	Prowadzenie działań zachęcających podmioty prywatne do udostępniania w portalu Dane.gov.pl danych w porozumieniu z ministrem właściwym do spraw informatyzacji (przez działania informacyjno-promocyjne, w szczególności np. konferencje, media społecznościowe, rozpowszechnianie materiałów edukacyjnych).	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
7	5.1.2. Zwiększenie dystrybucji danych w portalu Dane.gov.pl	Przeprowadzenie analizy SIST pod kątem danych spełniających cechy otwartości w celu ich publikowania w portalu Dane.gov.pl	organizacyjne, informacyjne	2021	2022	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działania: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
8	5.1.2. Zwiększenie dystrybucji danych w portalu Dane.gov.pl	Poszerzenie katalogu podmiotów obowiązkowo lub dobrowolnie publikujących dane w portalu Dane.gov.pl	informacyjne, promocyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
9	5.1.3. Udostępnianie danych przez API	Dobudowa API do baz danych/rejestrów będących w dyspozycji resortów oraz jednostek podległych i nadzorowanych przez właściwych ministrów, o których mowa w załączniku nr 1, w tabeli 2 „Plan dobudowy API”	techniczne	2021	2027	szczegóły w tabeli 2 „Plan dobudowy API”
10	5.1.3. Udostępnianie danych przez API	Udostępnianie przez API portalu Dane.gov.pl zasobów danych spełniających kryteria danych dynamicznych oraz danych wysokiej wartości	technologiczne, organizacyjne, informacyjne	2021	zadanie ciągłe	dysponenci danych dynamicznych i danych HVD
11	5.1.3. Udostępnianie danych przez API	Realizowanie projektów dotyczących otwartych danych finansowanych z WRF 2021–2027	organizacyjne	2021	2027	realizatorzy programu
12	5.1.3. Udostępnianie danych przez API	Rozbudowa portalu Dane.gov.pl w zakresie udostępniania danych dynamicznych i danych o wysokiej wartości	organizacyjne, technologiczne	2021	2022	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
13	5.1.3. Udostępnianie danych przez API	Identyfikacja potencjalnych dostawców danych dynamicznych z administracji rządowej oraz nawiązanie współpracy w celu publikacji w portalu Dane.gov.pl	organizacyjne	2021	zadanie ciągłe	realizatorzy Programu
14	5.1.3. Udostępnianie danych przez API	Zwiększenie częstotliwości aktualizacji danych opublikowanych w portalu Dane.gov.pl, uwzględniając możliwości dostawców danych	organizacyjne	2021	zadanie ciągłe	realizatorzy Programu
15	5.1.4. Zasady tworzenia danych	Przygotowanie projektu rekomendacji Komitetowi Rady Ministrów do spraw Cyfryzacji w sprawie stosowania przez organy administracji centralnej zasady „otwartości z samego założenia” oraz zasady „otwartości domyślnej”	organizacyjne	2021	2022	urząd obsługujący ministra właściwego do spraw informatyzacji
16	5.1.4. Zasady tworzenia danych	Podjęcie działań zmierzających do włączenia zasady otwartości w fazie projektowania i zasady otwartości domyślnej do fiszek projektów informatycznych oraz list kontrolnych osiągania interoperacyjności przedkładanych do zaopiniowania Komitetowi Rady Ministrów do spraw Cyfryzacji	organizacyjne	2021	2022	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
17	5.1.5. Konsultacje z interesariuszami	Prowadzenie: - cyklicznych badań rynku ponownego wykorzystywania danych publicznych w Polsce, w szczególności skali wykorzystywania danych publicznych przez przedsiębiorców oraz ich potrzeb dotyczących budowania modeli biznesowych w oparciu o dane, - minimum raz w roku badania ankietowego w jednostkach samorządu terytorialnego w celu identyfikacji zasobów danych, na które jest zapotrzebowanie, - minimum raz w roku badania ankietowego użytkowników portalu Dane.gov.pl. w celu oszacowania popytu na poszczególne kategorie danych	organizacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
18	5.1.5. Konsultacje z interesariuszami	Rozbudowanie funkcjonalności portalu Dane.gov.pl usprawniającej komunikację z jego użytkownikami	technologiczne	2021	2021	urząd obsługujący ministra właściwego do spraw informatyzacji
19	5.2.1. Podnoszenie jakości danych	Publikowanie przez dostawców portalu Dane.gov.pl nowych zasobów danych w możliwie najwyższym poziomie otwartości zgodnie z pięciogwiazdkowym schematem otwartych danych, w celach ułatwienia ponownego wykorzystywania danych oraz	organizacyjne	2021	zadanie ciągłe	realizatorzy programu

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
		dalszego przetwarzania maszynowego				
20	5.2.1. Podnoszenie jakości danych	Dokonywanie przez dostawców danych bieżących przeglądów zasobów danych opublikowanych w portalu w celu poprawy ich jakości i zwiększenia ich wykorzystania w usługach i produktach	organizacyjne	2021	zadanie ciągłe	realizatorzy Programu
21	5.2.1. Podnoszenie jakości danych	Dostosowanie funkcjonalności portalu Dane.gov.pl do: publikowania zasobów danych w wielu formatach, polegającego na udostępnianiu kilku plików danych w różnych formatach danych w jednym zasobie lub udostępnianiu zasobu danych z materiałami współtowarzyszącymi, które mogą być istotne dla tego zasobu danych (np. zbiór danych w formacie csv, a jako materiały uzupełniające plik PDF lub XLSX) - konwersji danych do alternatywnych formatów o wyższych stopniach otwartości, - konwersji plików CSV do formatu JSON	technologiczne	2021	2023	urząd obsługujący ministra właściwego do spraw informatyzacji
22	5.2.1. Podnoszenie jakości danych	Zwiększenie liczby zasobów danych o najwyższym stopniu otwartości (linked data) udostępnianych w portalu Dane.gov.pl (przez prowadzenie działań informacyjnych w ramach sieci peñnomocników do spraw otwartości danych oraz w ramach mediów	organizacyjne, informacyjne, technologiczne	2021	zadanie ciągłe	realizatorzy Programu

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
		społecznościowych, newslettera)				
23	5.2.1. Podnoszenie jakości danych	Poszerzenie katalogu formatów danych wskazanych w KRI o formaty na i 5 stopniu otwartości danych	legislacyjne	2021	2023	urząd obsługujący ministra właściwego do spraw informatyzacji
24	5.2.1. Podnoszenie jakości danych	Szkolenia dostawców danych w zakresie tworzenia i publikowania danych połączonych	edukacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
25	5.2.1. Podnoszenie jakości danych	Rozbudowa funkcjonalności portalu Dane.gov.pl o wyszukiwarkę danych połączonych SPARQL	technologiczne	2021	2025	urząd obsługujący ministra właściwego do spraw informatyzacji
26	5.2.1. Podnoszenie jakości danych	Dostarczanie właściwie dobranych i kompletnych zestawów metadanych określanych przez dostawcę danych	organizacyjne, informacyjne, edukacyjne	2021	zadanie ciągłe	realizatorzy Programu

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
27	5.2.1. Podnoszenie jakości danych	Publikowanie przez dostawców metadanych w otwartych formatach	organizacyjne, Informacyjne, edukacyjne	2021	zadanie ciągłe	realizatorzy Programu
28	5.2.1. Podnoszenie jakości danych	Rozbudowanie funkcjonalności portalu Dane.gov.pl polegającej na automatycznym udostępnianiu przez oprogramowanie portalu Dane.gov.pl metadanych w formacie rdf/xml. opisujących strukturę zbiorów danych	technologiczne	2021	2021	urząd obsługujący ministra właściwego do spraw informatyzacji
29	5.2.1. Podnoszenie jakości danych	Dodanie nowej funkcjonalności portalu Dane.gov.pl umożliwiającej dostawcom danych, użytkownikom portalu przykładów ponownego wykorzystania zbioru danych tzw. użyteczność danych	technologiczne	2022	2025	urząd obsługujący ministra właściwego do spraw informatyzacji
30	5.2.1. Podnoszenie jakości danych	Zwiększenie liczby automatycznego pobierania metadanych do portalu Dane.gov.pl z innych repozytoriów danych posiadających dobudowane API (przez działania organizacyjne, promocyjne, informacyjne, np. spotkania z pehonomocnikami, warsztaty, media społecznościowe, newsletter, warsztaty)	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
31	5.2.1. Podnoszenie jakości danych	Zwiększenie liczby podmiotów zaangażowanych w realizację procesu automatycznego dodawania i aktualizowania zbiorów i zasobów danych w portalu Dane.gov.pl (na podstawie pliku XML) (przez działania organizacyjne, promocyjne, informacyjne, np. spotkania z pełnomocnikami, warsztaty, media społecznościowe, newsletter, warsztaty)	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe	realizatorzy Programu
32	5.2.1. Podnoszenie jakości danych	Utworzenie w portalu Dane.gov.pl narzędzia: <ul style="list-style-type: none"> • wspomagającego kontrolę metadanych przez dostawcę, generującego automatyczne powiadomienie do dostawcy danych np. o zbliżającej się dacie aktualizacji danych (lub o ewentualnych problemach, weryfikacji opisów przed wygaśnięciem ważności danych), • regularnie sprawdzającego poprawność linków do danych źródłowych umieszczonych w repozytoriach zewnętrznych oraz generującego do dostawców danych informacje o konieczności aktualizacji uszkodzonych linków, • umożliwiającego „testowanie” przez dostawców danych zbiorów danych przed ich publikacją w portalu (testowanie polegające na weryfikacji odpowiednich formatów danych zgodnych ze standardem 	technologiczne	2022	2025	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działania: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
		technicznym (stanowiącym załącznik nr 2 do Programu), linków zewnętrznych dla danych źródłowych, typów danych w plikach, uzyskania podglądu danych, map, wykresów)				
33	5.2.1. Podnoszenie jakości danych	Weryfikacja Standardu technicznego dokonywana w ramach ewaluacji bieżącej (on-going) (stanowiącego załącznik nr 2 do Programu) w celu zapewnienia aktualności zalecanych norm, standardów ujętych w ww. Standardzie	informacyjne, organizacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
34	5.2.2. Działania wspierające dostęp do danych przez interfejs programowania aplikacji (API)	Weryfikacja Standardu API dokonywana w ramach ewaluacji bieżącej (on-going), stanowiącego załącznik nr 3 do Programu, w celu zapewnienia aktualności zalecanym normom, standardom oraz wytycznym ujętych w ww. Standardzie	informacyjne, organizacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
35	5.2.3. Interoperacyjność danych	Opracowanie Architektury Informatycznej Państwa, w tym architektury danych	organizacyjne, technologiczne	prace w trakcie	2027	urząd obsługujący ministra właściwego do spraw informatyzacji
36	5.2.4. Zapewnienia interoperacyjności wymiany danych między europejskimi, krajowymi i lokalnymi systemami	Dostosowanie opisów zasobów dostępnych w portalu Dane.gov.pl do standardu DCAT-AP	organizacyjne, technologiczne	2021	2021	urząd obsługujący ministra właściwego do

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
	otwartych danych					spraw informatyzacji
37	5.2.4. Zapewnienia interoperacyjności wymiany danych między europejskimi, krajowymi i lokalnymi systemami otwartych danych	Rekomendowanie wdrażania standardu DCAT-AP w lokalnych portalach otwartych danych (przez prowadzone działania informacyjne, edukacyjne)	informacyjne, organizacyjne	2021	ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
38	5.3.1. Wymiana danych między różnymi dostawcami danych	Aktywne uczestnictwo Rzeczypospolitej Polskiej w tworzeniu unijnych aktów prawnych oraz mechanizmów i rozwiązań pozalegisacyjnych dotyczących wymiany danych	organizacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
39	5.3.1. Wymiana danych między różnymi dostawcami danych	Aktywne promowanie w kraju przyjętych rozwiązań i mechanizmów mających na celu upowszechnienie wymiany i ponownego wykorzystywania danych	organizacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
40	5.3.2. Likwidacja barier utrudniających dostęp i wymianę danych	Implementacja filarów otwartości do przepisów ustawowych	legislacyjne	2021	2021	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działania: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
41	5.3.2. Likwidacja barier utrudniających dostęp i wymianę danych	Prowadzenie działań informacyjnych dla dostawców danych dotyczących sposobów postępowania z barierami utrudniającymi dostęp do danych	informacyjne, edukacyjne, promocyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
42	5.3.2. Likwidacja barier utrudniających dostęp i wymianę danych	Opracowanie standardów wymiany danych	organizacyjne	2021	2027	urząd obsługujący ministra właściwego do spraw informatyzacji
43	5.3.3. System zachęt dla stymulowania rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych	Realizacja projektów dotyczących zwiększenia dostępności i jakości danych w ramach WRF 2021–2027	organizacyjne	2021	2027	realizatorzy programu
44	5.3.3. System zachęt dla stymulowania rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych	Realizacja projektu Budowy Zintegrowanej Platformy Analitycznej	organizacyjne, technologiczne	prace w toku	2022	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
45	5.3.3. System zachęt dla stymulowania rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych	Realizacja projektu Cyfrowa Płaskownica Administracji	organizacyjne, technologiczne	prace w toku	2021	urząd obsługujący ministra właściwego do spraw informatyzacji
46	5.3.3. System zachęt dla stymulowania rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych	Realizacja hackathonów wspierających startupy wykorzystujące dane do realizacji usług, produktów	organizacyjne	2021	2027	urząd obsługujący ministra właściwego do spraw informatyzacji
47	5.3.3. System zachęt dla stymulowania rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych	Wypracowywanie rekomendacji LOD ułatwiających administracji rządowej i samorządowej otwieranie danych lub podnoszenie ich jakości	organizacyjne	prace w toku	2021	urząd obsługujący ministra właściwego do spraw informatyzacji
48	5.3.3. System zachęt dla stymulowania rozwoju rynku rozwiązań cyfrowych, aplikacji i usług bazujących na danych	Rozpowszechnianie aplikacji i usług bazujących na danych stworzonych przez administrację rządową, samorządową, sektor prywatny w portalu Dane.gov.pl	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe	realizatorzy programu
49	5.3.4. Bezpieczeństwo danych	Stymulowanie badań i rozwoju w obszarze cyberbezpieczeństwa przez kontynuację dedykowanych cyberbezpieczeństwu programów badawczych wspólnie z	edukacyjne, organizacyjne	prace w toku	2024	urząd obsługujący ministra właściwego do spraw

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
		Narodowym Centrum Badań i Rozwoju				informatyzacji
50	5.3.4. Bezpieczeństwo danych	Rozwijanie bazy wiedzy o tematyce cyberbezpieczeństwa dla różnych grup użytkowników technologii informacyjnych w portalu GOV.pl https://www.gov.pl/web/baza-wiedzy/cyberbezpieczenstwo	edukacyjne	prace w toku	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
51	5.3.4. Bezpieczeństwo danych	Prowadzenie działań informacyjno-edukacyjnych w celu zwiększenia kompetencji dostawców danych w portalu Dane.gov.pl z zakresu cyberbezpieczeństwa i ochrony danych	informacyjne, edukacyjne	prace w toku	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
52	5.3.4. Bezpieczeństwo danych	Upowszechnianie standardu bezpieczeństwa	informacyjne, edukacyjne, promocyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
53	5.3.4. Bezpieczeństwo danych	Rozwijanie bazy wiedzy oraz promowanie dobrych praktyk w portalu Dane.gov.pl w zakresie ochrony danych osobowych, ze szczególnym uwzględnieniem zagadnień dotyczących zapewnienia tej ochrony w procesie otwierania danych	informacyjne, edukacyjne, promocyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działania: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
54	5.4.1. Zapewnienie powszechnej dostępności zasobów kultury	Realizacja projektu Kronik@ – Krajowego Repozytorium Obiektów Kultury i Nauki	organizacyjne	prace w toku	2021	urząd obsługujący ministra właściwego do spraw informatyzacji
55	5.4.1. Zapewnienie powszechnej dostępności zasobów kultury	Integracja wyszukiwarki portalu Dane.gov.pl z wyszukiwarką repozytorium KRONIK@	organizacyjne, technologiczne	prace w toku	2021	urząd obsługujący ministra właściwego do spraw informatyzacji
56	5.4.1. Zapewnienie powszechnej dostępności zasobów kultury	Integracja wyszukiwarki repozytorium KRONIK@ z wyszukiwarką serwisu Szukaj w Archiwach	organizacyjne, technologiczne	prace w toku	2021	urząd obsługujący ministra właściwego do spraw informatyzacji
57	5.4.2. Zwiększenie dostępności zasobów nauki	Aktualizacja krajowej polityki w zakresie otwartego dostępu do danych i publikacji naukowych	organizacyjne	prace w toku	2022	urząd obsługujący ministra właściwego do spraw szkolnictwa wyższego i nauki

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
58	5.4.2. Zwiększenie dostępności zasobów nauki	Udział Polski w pracach na rzecz utworzenia European Open Science Cloud	organizacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw szkolnictwa wyższego i nauki
59	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Poszerzenie tzw. sieci peñnomocników na wielu poziomach instytucjonalnych (przez powołanie fakultatywnych peñnomocników do spraw otwartości danych)	organizacyjne	2021	zadanie ciągłe	zainteresowane podmioty
60	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Wzmocnienie współpracy między portalem Dane.gov.pl a portalami lokalnymi otwartych danych (przez realizację działań informacyjnych, edukacyjnych, np. warsztaty, udostępnianie materiałów edukacyjnych)	organizacyjne, informacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
61	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Organizacja inicjatyw na rzecz wzmocnienia współpracy administracji rządowej z JST w zakresie otwierania danych (przez działania informacyjne, szkoleniowe, promocyjne, np. warsztaty, szkolenia, konferencje, materiały edukacyjne)	organizacyjne, edukacyjne, informacyjne	2021	2023	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działania: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
62	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Realizacja działań na rzecz obszarów wiejskich z zakresu otwierania danych (przez działania informacyjne, promocyjne, organizacyjne, np. organizacja hackathonów, warsztaty, materiały edukacyjne, ankiety podaży danych)	organizacyjne, edukacyjne, informacyjne	2021	2027	urząd obsługujący ministra właściwego do spraw informatyzacji
63	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Realizacja „Dekalogu wyzwań cyfrowych na lata 2020–2023” z zakresu otwartych danych	organizacyjne	2021	2023	urząd obsługujący ministra właściwego do spraw informatyzacji
64	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Nawiązanie i rozwój współpracy między administracją rządową a sektorem prywatnym w zakresie otwartych danych (przez działania informacyjne, promocyjne, np. konferencje, hackathony, media społecznościowe, newsletter, materiały edukacyjne)	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
65	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Kontynuacja współpracy z organizacjami pozarządowymi w obszarze otwierania danych (przez działania informacyjne, promocyjne, np. konferencje, warsztaty, materiały edukacyjne)	organizacyjne, informacyjne, promocyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działania: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
66	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Nawiązanie współpracy ze środowiskiem naukowym w zakresie publikowania danych na Dane.gov.pl (przez działania informacyjne, np. warsztaty, wymianę doświadczeń)	organizacyjne, informacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
67	5.5.1. Inicjonowanie i wspieranie przedsięwzięć dotyczących współpracy z interesariuszami	Włączenie użytkowników indywidualnych we współzarządzanie danymi portalu Dane.gov.pl (przez udostępnianie w portalu aplikacji, publikowanie artykułów w bazie wiedzy i nowych zbiorów danych zaproponowanych przez użytkowników)	organizacyjne, edukacyjne, informacyjne	2021	2027	urząd obsługujący ministra właściwego do spraw informatyzacji
68	5.5.2. Otwarcie portalu Dane.gov.pl na nowych dostawców danych	Zachęcanie nowych podmiotów, wskazanych w rozdziale 5.5.1. do publikowania danych w portalu Dane.gov.pl	organizacyjne, edukacyjne, informacyjne	2021	2027	urząd obsługujący ministra właściwego do spraw informatyzacji
69	5.5.2. Otwarcie portalu Dane.gov.pl na nowych dostawców danych	Zachęcanie do udostępniania w portalu Dane.gov.pl aplikacji opracowanych na podstawie otwartych zbiorów danych	organizacyjne, edukacyjne, informacyjne	2021	2027	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działanie: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
70	5.5.3. Wzmocnienie współpracy z europejskimi interesariuszami danych oraz prowadzenie aktywnej polityki otwartych danych na arenie międzynarodowej	Promowanie polskich rozwiązań na arenie międzynarodowej	informacyjne, organizacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
71	5.5.3. Wzmocnienie współpracy z europejskimi interesariuszami danych oraz prowadzenie aktywnej polityki otwartych danych na arenie międzynarodowej	Wymiana doświadczeń i wiedzy mogącej stanowić inspirację dla krajowych rozwiązań oraz dystrybucji uzyskanych informacji wśród interesariuszy Programu	informacyjne, organizacyjne	2021	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
72	5.6.1. Zwiększanie umiejętności pracowników administracji publicznej w celu poprawy jakości polityk publicznych, dostarczania danych zarządzających i wspierania w podejmowaniu decyzji	Realizacja działań szkoleniowych z obszaru zarządzania danymi i otwierania danych publicznych dla pracowników administracji rządowej i samorządowej	organizacyjne, edukacyjne, informacyjne	prace w trakcie	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
73	5.6.1. Zwiększanie umiejętności pracowników administracji publicznej w celu poprawy jakości polityk publicznych, dostarczania danych zarządzających i wspierania w podejmowaniu decyzji	Organizacja warsztatów z korzyści otwierania danych publicznych sprofilowanych pod kątem różnych grup odbiorców	organizacyjne, edukacyjne, informacyjne	prace w trakcie	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji

Lp.	Numer i nazwa kierunku interwencji	Nazwa działania	Typ działania – działania: legislacyjne, organizacyjne, technologiczne, edukacyjne, informacyjne, promocyjne, inne	Termin rozpoczęcia	Termin zakończenia	Organ wiodący /Organy wiodące
74	5.6.2. Upowszechnianie informacji na temat otwierania i ponownego wykorzystywania danych publicznych oraz podnoszenie świadomości społecznej w zakresie korzyści płynących z otwartych danych	Realizacja kampanii internetowych promujących otwarte dane	organizacyjne, edukacyjne, informacyjne	prace w trakcie	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji
75	5.6.2. Upowszechnianie informacji nt. otwierania i ponownego wykorzystywania danych publicznych oraz podnoszenie świadomości społecznej w zakresie korzyści płynących z otwartych danych	Organizacja konferencji służących wymianie doświadczeń	organizacyjne, edukacyjne, informacyjne	prace w trakcie	zadanie ciągłe	urząd obsługujący ministra właściwego do spraw informatyzacji

Tabela 2. Plan dobowy API

Lp.	Nazwa rejestru/bazy, do których zostanie dobowane API	Przewidywany termin rozpoczęcia prac (harmonogram na lata 2021–2027)	Przewidywany termin zakończenia prac (harmonogram na lata 2021–2027)	Nazwa instytucji, będącej dysponentem rejestru/bazy, do których zostanie dobowane API
1	System Informacyjny o Instalacjach wytwarzających Promieniowanie ElektroMagnetyczne (SI2PEM)	2021	2023	urząd obsługujący ministra właściwego do spraw informatyzacji
2	Dziedziczne Bazy Wiedzy	prace w toku	2022	Główny Urząd Statystyczny
3	Baza danych Systemu Monitorowania Usług Publicznych	prace w toku	2021	Główny Urząd Statystyczny
4	Składnica Metadanych Statystycznych	2020	2022	Główny Urząd Statystyczny
5	Składnica Danych Publikacyjnych	2020	2022	Główny Urząd Statystyczny
6	Informatyczny System Obsługi Budżetu Państwa TREZOR	prace w toku	2021	Ministerstwo Finansów
7	Rejestr Należności Publicznych	2021	2023	Ministerstwo Finansów
8	EUREKA	2020	2021	Ministerstwo Finansów
9	Hydrologia rejonu Bałtyku Południowego	prace w toku	2022	Ministerstwo Rolnictwa i Rozwoju Wsi
10	Meteorologia z rejonu Bałtyku Południowego	prace w toku	2022	Ministerstwo Rolnictwa i Rozwoju Wsi
11	Biogeny z rejonu Bałtyku Południowego	prace w toku	2022	Ministerstwo Rolnictwa i Rozwoju Wsi
12	Wirtualny Informator Rzeczny	2020	2023	Ministerstwo Infrastruktury
13	System Monitorowania Kosztów Leczenia ("SMoKL")	2020	2023	Ministerstwo Zdrowia
14	Rejestr szpitali akredytowanych	2021	2021	Ministerstwo Zdrowia
15	Rejestr zapytań o dostęp do informacji publicznej	2021	2021	Ministerstwo Zdrowia
16	Rejestr produktów objętych powiadomieniem o pierwszym wprowadzeniu do obrotu	2020	2023	Ministerstwo Zdrowia

17	Digital Brain – cyfrowe zasoby Instytutu Psychiatrii i Neurologii w Warszawie	prace w toku	2021	Ministerstwo Zdrowia
18	ProfiBaza	prace w toku	2021	Ministerstwo Zdrowia
19	BackOffice	2021	2023	Ministerstwo Zdrowia
20	Rejestr projektu „InterScienceCloud – Zintegrowana platforma informacji o działalności naukowej Uniwersytetu Medycznego w Łodzi”	2021	2023	Ministerstwo Zdrowia
21	System dla Politransplant	2022	2026	Ministerstwo Zdrowia
22	Rejestr Aptek	2022	2026	Ministerstwo Zdrowia
23	Rejestr Ośrodków Medycznie Wspomaganej Prokreacji i Banków Komórek Rozrodczych i Zarodków	2022	2026	Ministerstwo Zdrowia
24	Rejestr Hurtowni Farmaceutycznych	2022	2026	Ministerstwo Zdrowia
25	Rejestr Jednostek Wspomagających Państwowe Ratownictwo Medyczne	2022	2026	Ministerstwo Zdrowia
26	Rejestr Produktów Leczniczych	2022	2026	Ministerstwo Zdrowia
27	Ewidencja szpitalnych oddziałów ratunkowych, centrów urazowych, centrów urazowych dla dzieci i jednostek organizacyjnych szpitali udzielających świadczeń w zakresie ratownictwa medycznego	2022	2026	Ministerstwo Zdrowia
28	Rejestr Podmiotów Wykonujących Działalność Leczniczą	2022	2026	Ministerstwo Zdrowia
29	System Obsługi Importu Docelowego	2022	2026	Ministerstwo Zdrowia
30	System Obsługi List Refundacyjnych	2022	2026	Ministerstwo Zdrowia
31	Rejestry Publiczne w Turystyce	prace w toku	2021	Ministerstwo Rozwoju, Pracy i Technologii
32	Elektroniczny Dziennik Budowy	prace w toku	2021	Ministerstwo Rozwoju, Pracy i Technologii
33	System Nadzoru Rynku WYROBÓW BUDOWLANYCH	2021	2021	Ministerstwo Rozwoju, Pracy i Technologii
34	Ewidencja eksploatowanych urządzeń technicznych	2022	2023	Ministerstwo Rozwoju, Pracy i Technologii
35	Platforma orzecznictwa urzędu patentowego Rzeczypospolitej Polskiej	2020	2022	Ministerstwo Rozwoju, Pracy i Technologii

36	Doskonalenie i Rozbudowa Zintegrowanego Systemu Informatycznego ZSI – ULC	prace w toku	2021	Ministerstwo Infrastruktury
37	RAD-on – dane zintegrowane, z wielu systemów źródłowych (rozwój API)	prace w toku	2027	Ministerstwo Edukacji i Nauki
38	POL-on – Zintegrowany System Informacji o Szkolnictwie Wyższym i Nauce POL-on, dane publiczne udostępniane w ramach API RAD-on (rozwój API)	prace w toku	2027	Ministerstwo Edukacji i Nauki
39	POL-on – Zintegrowany System Informacji o Szkolnictwie Wyższym i Nauce POL-on, usługi edycji danych (rozwój API)	prace w toku	2027	Ministerstwo Edukacji i Nauki
40	KRS	2020	2021	Ministerstwo Sprawiedliwości

Załącznik nr 2

Standard techniczny

Załącznik nr 2. Standard techniczny

1. Wstęp

Dokument definiuje zestaw minimalnych wymagań technicznych, które muszą spełniać dane, w szczególności udostępniane w portalu Dane.gov.pl.

Użyte w dokumencie sformułowania wymaga się, zaleca się, dopuszcza się, nie zaleca się, nie dopuszcza się oznaczają odpowiednio:

- **wymaga się** – aby zachować zgodność ze standardem, trzeba bezwarunkowo spełnić opisane kryteria/zasady,
- **zaleca się** – w uzasadnionych przypadkach można nie spełnić danego kryterium, pod warunkiem że opisująca zasoby dokumentacja (np. dokumentacja API, opisana w „Standardzie API”¹⁰⁹⁾) zawiera uzasadnienie decyzji o odstępstwie,
- **dopuszcza się** – można zastosować opisane kryterium lub go nie zastosować, a dokumentacja tej decyzji nie jest wymagana (choć można ją zamieścić),
- **nie zaleca się** – w wyjątkowych i uzasadnionych przypadkach można postąpić w opisany sposób, pod warunkiem że opisująca zasoby dokumentacja (np. dokumentacja API, opisana w „Standardzie API”) zawiera uzasadnienie decyzji o odstępstwie od danego zalecenia,
- **nie dopuszcza się** – aby zachować zgodność ze standardem, nie wolno postąpić w opisany sposób.

Poziomy otwartości danych

Schemat pięciu poziomów otwartości danych, zaproponowany przez Tima Bernersa-Lee, to użyteczne narzędzie klasyfikowania zasobów pod kątem ich otwartości. Model otwartości opisany tym schematem został przedstawiony w POD. Więcej informacji o tym schemacie można znaleźć na stronach W3C¹¹⁰⁾.

Im wyższy poziom otwartości danych, tym dane są lepiej przygotowane do dalszego przetwarzania. Wszystkie otwarte dane są udostępniane bez żadnych ograniczeń do dowolnych celów komercyjnych i niekomercyjnych. Schemat pięciu poziomów otwartości danych jest opisany w tabeli 1.

UWAGI:

Dane na poziomach otwartości 3 (*) i wyższych zaleca się udostępniać przez API, co pozwala na ich maszynowe przetwarzanie i opatrzenie dodatkowymi metadanymi.**

Nie zaleca się publikowania danych na pierwszym poziomie otwartości.

¹⁰⁹⁾ Stanowiący załącznik nr 3 do POD.

¹¹⁰⁾ <https://www.w3.org/DesignIssues/LinkedData.html>.

Tabela 3. Schemat pięciu poziomów otwartości danych

Poziom	Oznaczenie	Charakterystyka poziomu	Dodatkowe informacje	Opis
1	*	Dane w dowolnym formacie są udostępniane bez ograniczeń licencyjnych ¹¹¹⁾	Dopuszczalne jest publikowanie danych w formie plików graficznych, skanów z obrazem danych i tekstu, plików tekstowych	Mogą to być pliki JPEG z zeskanowanymi dokumentami, wygenerowane z różnych programów pliki PDF lub pliki tekstowe zawierające dane nieustrukturyzowane albo o strukturach niejednorodnych. Większość ludzi może je odczytać, ale jakiegolwiek dalsze ich wykorzystywanie wymaga dodatkowej pracy, polegającej na zidentyfikowaniu, odczytaniu i przeniesieniu danych (często ręcznie bądź ze wspomaganiami programów typu OCR)
2	**	Dane są udostępniane w postaci ustrukturyzowanej	Dane publikowane są w postaci sformatowanego arkusza kalkulacyjnego	Dane mają już określoną strukturę, którą można odczytać komputerowo, np. przez zastosowanie pliku w formacie arkusza kalkulacyjnego lub edytora tekstu. Nie są to zeskanowane obrazy w niedającej się przeszukiwać formie. Dane te są w formacie zamkniętym (własnościowym), dla którego stosowanie w oprogramowaniu jest ograniczone przez restrykcje patentowe, licencyjne lub podobne
3	***	Używanie formatów otwartych	Dane są publikowane w otwartym ustrukturyzowanym formacie, np. CSV lub JSON	Dane są w otwartym formacie, ale ich zrozumienie na potrzeby przetwarzania maszynowego wymaga każdorazowej analizy danych i ustalenia bądź odnalezienia w dokumentacji, jeżeli taka istnieje, znaczeń poszczególnych pól. Czy „nazwisko” oznacza samo nazwisko, czy imię i nazwisko? Czy „kod” to kod pocztowy, czy terytorialny? Czy „odległość” jest podana w metrach, czy w kilometrach? Czy „1/12/2018” to pierwszy grudnia, czy dwunasty stycznia? Dane udostępniane są w niezastrzeżonym formacie
4	****	Używanie URI do identyfikacji obiektów zgodnie z RDF	Dane publikowane są w formacie umożliwiający oznaczenie ich struktury znaczeniowej	Wejście na czwarty poziom otwartości pozwala jednoznacznie określić znaczenie udostępnianych danych. Technicznym sposobem wyrażania takiego znaczenia w sposób zrozumiały dla maszyn jest identyfikacja konkretnych właściwości danych za pomocą zrozumiałych dla maszyny URI zgodnie z modelem RDF (model opisu danych). Dane udostępniane są w niezastrzeżonym formacie
5	*****	Łączenie danych z innymi danymi za pomocą linków – budowanie kontekstu	Dane zawierają połączenia strukturalne online do innych zbiorów informacji	Piąty poziom dodatkowo ułatwia przetwarzanie – jawnie wskazuje relacje między danymi w formie linków. Dzięki temu możliwe jest odnajdywanie połączeń między różnymi zbiorami danych. Dane udostępniane są w niezastrzeżonym formacie

¹¹¹⁾ Więcej na ten temat w: „Standard prawny: Standardy otwartości danych”, Ministerstwo Cyfryzacji, <https://dane.gov.pl/knowledge-base/education/1219,standardy-otwartosci-danych>.

2. Formaty plików

2.1. Formaty JPEG i PNG

Formaty graficzne są dopuszczalne jako formaty udostępniania otwartych danych wyłącznie na 1. (*) poziomie otwartości, ponieważ co do zasady nie nadają się do automatycznego przetwarzania zamieszczonych w nich danych. Wyjątkiem jest sytuacja, w której same obrazy (a nie przedstawiona na nich zawartość, np. tekst) są udostępnianymi danymi. Wówczas poziom otwartości zależy od poziomu otwartości metadanych, którymi opatrzony jest obraz, przygotowanymi zgodnie z zaleceniami niniejszego dokumentu. Możliwe jest przetworzenie zasobów do postaci zgodnej z wyższymi poziomami otwartości przy użyciu dostępnych na rynku narzędzi do konwersji danych.

2.2. Format PDF

Format PDF jest dopuszczalny wyłącznie na 1. (*) poziomie otwartości, ponieważ co do zasady nie nadaje się do automatycznego przetwarzania. Możliwe jest przetworzenie zasobów do postaci zgodnej z wyższymi poziomami otwartości przy użyciu dostępnych na rynku narzędzi do konwersji danych.

2.3. Format TXT

Format TXT, pozbawiony struktury, jest dopuszczalny na 1. (*) poziomie otwartości. Zależnie od faktycznej struktury danych dane powinny być udostępniane z oznaczeniem innego, właściwego typu pliku (np. CSV).

2.4. Pliki archiwów

Dopuszczalne jest udostępnianie danych w postaci skompresowanej. **Na wszystkich poziomach otwartości dopuszcza się otwarte formaty kompresji, takie jak ZIP (bez własnościowych rozszerzeń), 7z, Gzip czy Bzip2**. Na poziomie otwartości do 2. (**) dopuszcza się także popularne formaty własnościowe: RAR, rozszerzony ZIP.

2.5. Formaty MS Office (DOC, XLS)

Są to własnościowe formaty dokumentów, **dopuszczalne na poziomie otwartości do 2. (**)**. Do arkusza danych w formacie XLS wymaga się stosowania szczegółowych wymagań dla zasobów w formacie CSV na 3. (***) poziomie otwartości określonych w pkt 4.1 i 4.2. Wobec tego zaleca się zastąpienie formatu XLS formatem CSV dla uzyskania 3. (***) poziomu otwartości. Możliwe jest przetworzenie zasobów do postaci zgodnej z wyższymi poziomami otwartości przy użyciu dostępnych na rynku narzędzi do konwersji danych.

2.6. Formaty z rodziny Office Open XML (XLSX, DOCX)

Formaty Office Open XML to otwarte formaty udostępniania dokumentów (DOCX) i ustrukturyzowanych danych (XLSX). Jednakże ze względów technicznych formaty te **nie są zalecane jako formaty udostępniania otwartych danych powyżej 2. (**)** poziomu. Możliwe jest przetworzenie zasobów do postaci zgodnej z wyższymi poziomami otwartości przy użyciu dostępnych na rynku narzędzi do konwersji danych.

2.7. Formaty dBase (dbf)

Format bazy danych dBase jest dopuszczalny do 3. (*) poziomu otwartości**. Format ten jest stosowany do przechowywania danych ustrukturyzowanych. Możliwe jest przetworzenie zasobów do postaci zgodnej z wyższymi poziomami otwartości przy użyciu dostępnych na rynku narzędzi do konwersji danych.

2.8. Format HTML

Format HTML jest otwartym formatem, pozwalającym na udostępnianie ustrukturyzowanych danych. **Nie zaleca się stosowania formatu HTML na poziomach otwartości powyżej 3. (***)**. W przypadku udostępniania danych nieustrukturyzowanych w formacie HTML jest on dopuszczalny na 1. (*) poziomie otwartości. Możliwa jest transformacja formatu HTML do wyższych poziomów otwartości.

2.9. Formaty OpenDocument Spreadsheet (ODS)

Format ODS jest otwartym formatem udostępniania ustrukturyzowanych danych i w związku z tym **jest dopuszczalny do poziomu 3. (***) otwartości**. Możliwe jest przetworzenie zasobów do postaci zgodnej z wyższymi poziomami otwartości przy użyciu narzędzi dostępnych na rynku do konwersji danych.

2.10. Format CSV

CSV to prosty format przechowywania danych tabelarycznych. **Format CSV jest zalecany do udostępniania otwartych danych na 3. (***) poziomie otwartości**. Zastosowanie dodatkowych standardów, takich jak Model for Tabular Data and Metadata on the Web¹¹²⁾, pozwala na przekształcenie CSV do formatów na najwyższych poziomach otwartości. Szczegółowe wymagania związane z formatem CSV znajdują się w pkt 4. „Szczegółowe wymagania dla zasobów w formacie CSV”.

2.11. Format JSON

JSON to prosty format udostępniania ustrukturyzowanych danych. **Format JSON jest zalecany do udostępniania otwartych danych na 3. (***) poziomie otwartości**. Zastosowanie dodatkowych standardów, takich jak JSON-LD¹¹³⁾, pozwala na właściwe wyrażenie metadanych również na wyższych poziomach otwartości. Szczegółowe wymagania techniczne związane z formatem JSON znajdują się w pkt 5. „Szczegółowe wymagania dla zasobów w formacie JSON”.

2.12. Format XML

Format XML jest dopuszczalny do udostępniania danych na 3. (*) poziomie otwartości**. Format ten po przekształceniu można wyrazić na najwyższych poziomach otwartości. Szczegółowe wymagania techniczne związane z formatem XML znajdują się w pkt 6. „Szczegółowe wymagania dla zasobów w formacie XML”.

2.13. Formaty danych na poziomach otwartości 4. (****) i 5. (****) (RDF/XML, Turtle, N3, JSON-LD, N-tripels, RDFa)

Formaty plików takie jak RDF/XML, Turtle, Notation3 (N3), JSON-LD, N-tripels, RDFa zaleca się do udostępniania otwartych danych na najwyższych stopniach otwartości. Szczegółowe wymagania techniczne związane z ww. formatami znajdują się w pkt 7. „Szczegółowe wymagania dla zasobów na 4. (****) i 5. (****) poziomie otwartości” (RDF/ XML, Turtle, N3, JSON-LD, N-tripels, RDFa”).

2.14. Formaty danych przestrzennych

Dopuszczalne jest stosowanie formatów danych przestrzennych, w szczególności:

- ✓ dane przestrzenne (rastrowe) bez georeferencji **są dopuszczalne na 1. (*) poziomie otwartości**, np. format JPG/JPEG, PNG, TIFF, BMP,
- ✓ dane przestrzenne (rastrowe) z georeferencją **są dopuszczalne na 3. (***) poziomie otwartości**, np. GeoTIFF,
- ✓ dane przestrzenne (wektorowe) **są dopuszczalne na 3. (***) poziomie otwartości**, np. format GML, SHP, KML, GeoJSON, GPX,
- ✓ dane przestrzenne zaopatrzone w kontekst pozwalający na automatyczne przetworzenie do postaci zgodnej z RDF **są dopuszczalne na wszystkich poziomach otwartości**. Udostępniając takie dane, należy zachować dobre praktyki opisane w dokumencie „Spatial Data on the Web Best Practices”¹¹⁴⁾.

Dane przestrzenne udostępniane są przez usługi umożliwiające m.in. wyszukiwanie, przeglądanie oraz pobieranie danych (np. usługi WMS, WMTS, WFS, WCS, ATOM). Udostępnianie zbiorów danych przestrzennych odbywa się na podstawie odpowiadających im standardów i zestawów metadanych uregulowanych w dyrektywie Parlamentu Europejskiego i Rady nr 2007/2/WE z dnia 14 marca 2007 r. ustanawiającej infrastrukturę informacji

¹¹²⁾ <https://www.w3.org/TR/tabular-data-model/>.

¹¹³⁾ <https://json-ld.org/spec/latest/json-ld/>, <https://w3c.github.io/json-ld-syntax/>.

¹¹⁴⁾ <https://www.w3.org/TR/sdw-bp/>.

przestrzennej we Wspólnocie Europejskiej (INSPIRE), rozporządzeń wykonawczych do dyrektywy oraz ustawy o infrastrukturze informacji przestrzennej¹¹⁵⁾

3. Formatowanie danych

Niezależnie od wyboru formatu pliku danych wymaga się stosowania właściwego formatowania, w szczególności dla danych typu: liczba, data, godzina, wartość logiczna.

Na poziomach otwartości 4. (***) i 5. (****) każda właściwość ma określony przez URI¹¹⁶⁾ typ, który wyznacza oczekiwany format wartości (zob. szczegółowe opisy poniżej w punktach dotyczących CSV, JSON, XML). Przykładowo, jeżeli mamy do czynienia z właściwością `schema:datePublished`, to jednoznacznie ustala typ wartości na `schema:Date`, który z kolei jest datą w formacie zgodnym z normą ISO 8601¹¹⁷⁾.

Na tych poziomach wymagane jest określenie typu danych w ramach definicji właściwości.

Na poziomach otwartości do 3. (***) zaleca się również stosowanie zapisu daty i czasu w formacie zgodnym z normą ISO 8601 oraz odpowiednio podstawowych typów danych ze standardu XML Schema¹¹⁸⁾. Oznacza to w szczególności:

- ✓ stosowanie literalnych wartości `true`–`false` jako wartości logicznych,
- ✓ stosowanie kropki dziesiętnej (a nie przecinka) w zapisie ułamków dziesiętnych, bez żadnych dodatkowych separatorów (np. oddzielających tysiące); dopuszczalny jest tzw. zapis naukowy,
- ✓ zapis dat w postaci `yyyy-mm-dd`, a łącznie dat i godzin w postaci `yyyy-mm-ddThh:mm` lub `yyyy-mm-dd hh:mm:ss` (spacja między datą a czasem).

Dane zaleca się udostępniać w możliwie najwyższym stopniu granulacji (rozdrobienia), tzn. nie należy łączyć kilku danych w jednym polu.

3.1. Przygotowanie arkusza ustrukturyzowanych danych w formacie xls lub xlsx

1. Arkusz musi mieć jeden wiersz nagłówka, który zawiera opisy wszystkich kolumn z danymi.

Niedopuszczalne jest:

- a) scalanie komórek w ramach zarówno wiersza, jak i kolumny – zaburza to strukturę arkusza danych,
- b) umieszczanie treści dotyczących jednego rekordu w kilku wierszach, np. kolejnych danych adresu (adresat, ulica, miejscowość itp.),
- c) wprowadzanie dodatkowych tytułów, opisów, przypisów i komentarzy poza obszarem danych tabeli,
- d) używanie formuł agregujących dane, np. funkcji SUMA, lub funkcji filtrowania danych,
- e) używanie formuł pobierających dane spoza źródła danych, które będą niedostępne po przeniesieniu arkusza na inną platformę.

2. Każda kolumna tabeli, poza wierszem nagłówka, musi zawierać dane jednego rodzaju, tzn. liczbę, tekst lub ciąg cyfrowo-tekstowy o określonej strukturze i określonym znaczeniu, np. numer telefonu, kod pocztowy.

3. Każda kolumna tabeli powinna zawierać dane dotyczące pojedynczej wielkości (elementarnej), np. numer kodu, a nie numer kodu i miejscowość; ułatwia to sortowanie i łączenie danych z wielu tabel.

4. Zapis w każdym polu w kolumnie musi być zgodny z założonym formatem i modelem – maską, właściwą dla danego typu zawartości.

5. Nie należy stosować formatowania tekstu i niektórych znaków interpunkcyjnych właściwych dla formatowania zaawansowanego, które w pewnych kontekstach mogą imitować operatory

¹¹⁵⁾ Ustawa z dnia 4 marca 2010 r. o infrastrukturze informacji przestrzennej (Dz. U. z 2021 r. poz. 214).

¹¹⁶⁾ <https://tools.ietf.org/html/rfc3986>.

¹¹⁷⁾ <https://www.iso.org/iso-8601-date-and-time-format.html>.

¹¹⁸⁾ <https://www.w3.org/TR/xmlschema/>.

arytmetyczne lub zmieniać wartość liczby, np. łącznik (dywiz) może prowadzić do utraty danych, generując w arkuszu wynikowym błąd formuły lub zmieniać wartość liczby na ujemną.

6. Jeżeli powyższych warunków nie można spełnić ze względu na złożoność danych zawartych w tabeli, należy tabelę podzielić w sposób, który umożliwi ich spełnienie.

7. Zaleca się opracowywanie danych w postaci pojedynczych arkuszy, a nie tzw. skoroszytów.

3.2. Rekomendacje w zakresie formatowania danych

REGON – wymaga się zapisywania identyfikatora podmiotu jako ciągu 9 lub 14 cyfr, bez spacji lub łączników.

NIP – wymaga się zapisywania numeru identyfikacji podatkowej jako ciągu 10 cyfr, bez spacji lub łączników.

KRS – wymaga się zapisywania numeru jako ciągu 10 cyfr, bez spacji lub łączników.

Numer telefonu – w polskiej strefie numeracyjnej zaleca się zapisywanie numeru telefonu jako ciągu 9 cyfr, bez wyróżniania tzw. numeru kierunkowego miejscowości i bez spacji lub łączników. Dopuszczalne jest stosowanie prefiksu międzynarodowego, poprzedzonego znakiem plus „+”.

Tabela 4. Prawidłowe formatowanie danych

Data	Liczba	Wartość logiczna
1970-01-01	42.01	true
2018-12-31	1.05e3	false

Tabela 5. Nieprawidłowe formatowanie danych

Data	Liczba	Wartość logiczna
1/1/1970	45,01	T
1.1.1970	1,234.56	0
1 1970	1 234,5	NIE

3.3. Dane adresowe

Wymaga się, aby adres pocztowy był udostępniany jako szereg danych opisowych: nazwa ulicy, numer budynku, kod pocztowy, miejscowość – umieszczonych w odrębnych polach, a nie jako jedno pole tekstowe, zawierające wszystkie wymienione dane.

Tabela 6. Prawidłowe formatowanie danych

Nazwa	Kod PNA	Miejscowość	Cecha	Nazwa ulicy	Nr budynku	Nr lokalu	Nr telefonu
CSIOZ	00-184	Warszawa	ul.	Stanisława Dubois	5A		225970927
KPRM	00-583	Warszawa	al.	Aleje Ujazdowskie	1/3		226946000
Biuro	51-152	Wrocław	pl.	Marsz. Józefa Piłsudskiego	4	75	712558765

Tabela 7. Nieprawidłowe formatowanie danych

Nazwa	Adres	Nr telefonu
CSIOZ	00-184 Warszawa, ul. Dubois 5A	+48 22 597-09-27
KPRM	00-583 Warszawa, Aleje Ujazdowskie 1/3	(22) 694-60-00 (cent.)
Biuro	51-152 Wrocław, J. Piłsudskiego 4/75	71-25-58-765

Wymaga się, aby dane opisowe (tekstowe) składające się na adres były literalnie zgodne z ich zapisem w źródle ich wytworzenia. Sposób udostępniania nie może wpływać na ich wartość, tzn. nie mogą one ulegać zmianom w procesie przekazywania do miejsca publikacji.

Zbiorowym miejscem publikacji danych, takich jak nazwy ulic, po zrealizowaniu procesu uchwalenia (nazwy ulic są nadawane przez Radę Gminy) oraz po ogłoszeniu przez wojewodę w wojewódzkim dzienniku urzędowym, jest rejestr TERYT. W tym rejestrze do konkretnych nazw są przypisane kody – identyfikatory ze zbiorów ULIC¹¹⁹⁾ oraz SIMC. W nim też znajdują się kody terytorialne TERC. Kody terytorialne mogą zmieniać się w zależności od aktualnego podziału terytorialnego kraju, natomiast identyfikatory miejscowości, a nawet ich części, są nadawane ostatecznie. Z uwagi na potrzeby związane z poprawną adresacją – z lokalizacją punktów adresowych – utworzono pojęcie **adresu uniwersalnego**.

Adres uniwersalny to adres zapisany jako ciąg kodów – kodu pocztowego (tzw. PNA – 5 cyfr, bez myślnika), kodu terytorialnego składającego się z kodów województwa, powiatu, gminy (6 cyfr), identyfikatora miejscowości podstawowej (7 cyfr), identyfikatora miejscowości (7 cyfr), identyfikatora ulicy (5 cyfr, w przypadku braku ulic ten fragment adresu składa się z ciągu „00000”), współrzędnych geodezyjnych x, y danego punktu adresowego oraz numeru budynku – oddzielonych separatorem „|”.

Zaleca się stosowanie adresu uniwersalnego. Konstrukcja adresu uniwersalnego umożliwia jego maszynowe (automatyczne) rozkodowanie pozwalające na uzyskanie nazwy województwa, powiatu, gminy, miejscowości i nazwy ulicy oraz wyświetlenie mapy z lokalizacją danego punktu adresowego. Adres uniwersalny nie służy do odczytu przez człowieka, lecz przez system informatyczny, którego zadaniem jest wizualizacja danego punktu adresowego na mapie.

Aplikacje służące do wprowadzania danych adresowych do różnego rodzaju rejestrów czy też systemów teleinformatycznych powinny mieć zabezpieczenia przed wprowadzaniem adresów nieistniejących. Zatem systemy te powinny umożliwiać dokonywanie wyboru adresu ze słowników rzeczywiście istniejących obiektów. Przykładem może być aplikacja służąca do weryfikacji adresów, znajdująca się w portalu Dane.gov.pl¹²⁰⁾. Aplikacja ta wykorzystuje rozwijane listy utworzone na podstawie słowników TERYT oraz danych zawartych w Państwowym Rejestrze Granic. Ma ona również konwerter zakodowanych adresów uniwersalnych na adresy pocztowe punktów adresowych.

Przykład: adres uniwersalny Centrum e-Zdrowia przy ul. Stanisława Dubois 5A w Warszawie wygląda następująco:

00184|146501|0918123|0918123|04337|489147.9218|636045.6562|5A|

Poprawne adresy uniwersalne znajdują się w kolumnie AdresCSIOZ w plikach CSV dostępnych na stronie <http://integracja.gugik.gov.pl/daneadresowe/>. W tych samych plikach, a także plikach GML¹²¹⁾ ¹²²⁾, znajdują się szczegółowe dane adresowe, tworzone bezpośrednio w gminach. Są to dane źródłowe i najbardziej wiarygodne. Jednocześnie, fakt otwarcia danych adresowych

¹¹⁹⁾ http://eteryt.stat.gov.pl/eTeryt/.../pliki_pelne_struktury.aspx.

¹²⁰⁾ <https://dane.gov.pl/application/1216,aplikacja-do-sprawdzania-adresu-uniwersalnego>.

¹²¹⁾ Ustawa z dnia 17 maja 1989 r. – Prawo geodezyjne i kartograficzne (Dz. U. z 2020 r. poz. 2052).

¹²²⁾ Rozporządzenie Ministra Administracji i Cyfryzacji z dnia 9 stycznia 2012 r. w sprawie ewidencji miejscowości, ulic i adresów (Dz. U. poz. 125).

oraz ściśle określone zasady tworzenia zapewniają interoperacyjność we wszelkich zastosowaniach.

4. Szczegółowe wymagania dla zasobów w formacie CSV

4.1. Format pliku CSV

W praktyce występują pliki określane jako „CSV”, które różnią się szczegółami technicznymi. Różne są znaki nowej linii, kodowania znaków, używane są różne separatory pól: przecinek, średnik, tabulator.

Na 3. (***) poziomie otwartości zaleca się, stosowanie formatu CSV określonego w RFC 4180¹²³⁾, a na 4. (****) poziomie – wymaga się tego ściśle. W szczególności oznacza to:

- ✓ stosowanie znaków końca wiersza w formacie CRLF,
- ✓ używanie przecinka, jako separatora pól,
- ✓ używanie cudzysłowów do otaczania znaków specjalnych (takich jak przecinki czy znaki nowej linii wewnątrz pól),
- ✓ stosowanie domyślnego kodowania znaków UTF-8 bądź prawidłowe oznaczenie kodowania w nagłówku Content-type, jeżeli stosowane jest inne kodowanie znaków.

4.2. CSV na 3. (***) poziomie otwartości

Dla każdego pliku w formacie CSV wymagany jest dokładnie jeden wiersz nagłówka, opisujący zawartość tabeli w sposób możliwie najbardziej zrozumiały dla użytkownika. Jak powiedziano wcześniej: na tym poziomie kluczowa jest rola człowieka, który musi nadać sens danym przeznaczonym do dalszego automatycznego przetwarzania.

Wymaga się, aby wiersz nagłówka zawierał wyłącznie nagłówki poszczególnych kolumn, a wszystkie pozostałe wiersze zawierały wyłącznie dane kolejnych rekordów. Niedopuszczalne jest, przykładowo, wyodrębnianie fragmentu arkusza jako „legenda”.

Niedopuszczalna jest obecność kolumn bez nagłówka lub z pustym (składającym się tylko z białych znaków) nagłówkiem. Niedopuszczalne są także nagłówki powtarzające się w różnych kolumnach.

Wymaga się, aby wszystkie wiersze tabeli, łącznie z nagłówkiem, miały taką samą liczbę komórek.

Niedopuszczalna jest obecność całkowicie pustych kolumn bądź wierszy.

Wymaga się, aby wszystkie wiersze (poza wierszem nagłówka) były traktowane jako niezależne od siebie. Niedopuszczalne jest w szczególności pozostawianie w tabeli wolnych miejsc w celu uniknięcia powtórzeń danej wartości w kolejnych wierszach. Jeżeli dana wartość jest identyczna dla kolejnych rekordów, to nadal musi być osobno określona dla każdego rekordu.

4.3. CSV na 4. (****) i 5. (****) poziomie otwartości

Na tych poziomach otwartości kluczowe jest zapewnienie możliwości maszynowego odczytu wszystkich danych z uwzględnieniem ich kontekstu semantycznego. W tym celu wymagane jest stosowanie standardu Model for Tabular Data and Metadata on the Web¹²⁴⁾, pozwalającego na przekształcenie plików CSV do formatu RDF oraz JSON-LD. Dodatkowe szczegóły na temat transformacji formatu CSV do wyższych poziomów otwartości znajdują się również w innych standardach W3C¹²⁵⁾.

Dane na 4. (****) i 5. (****) stopniu otwartości szczegółowo opisano w rozdziale 7.

¹²³⁾ <https://tools.ietf.org/html/rfc4180>.

¹²⁴⁾ <https://www.w3.org/TR/tabular-data-model/>.

¹²⁵⁾ <https://www.w3.org/TR/csv2json/>, <https://www.w3.org/TR/csv2rdf/>.

5. Szczegółowe wymagania dla zasobów w formacie JSON

Format JSON jest zalecany do udostępniania otwartych danych na trzecim poziomie otwartości. Jest to prosty format, nadający się do przetwarzania ogólnie dostępnymi narzędziami znajdującymi się w standardowych bibliotekach wielu języków programowania. Format ten jest oparty na podzbiórce standardu języka programowania JavaScript. JSON jest formatem tekstowym, który jest niezależny od konkretnego języka.

Doprecyzowania wymaga udostępnianie w tym formacie dużych kolekcji (stronicowanie), określanie semantyki na wyższych poziomach otwartości, a także określanie dodatkowych funkcji związanych z udostępnianiem danych w API, takich jak sortowanie, filtrowanie, wyszukiwanie i wybieranie podzbiorów zwracanych pól z wyników zapytania.

5.1. JSON na 3. (***) poziomie otwartości

Na tym poziomie **wymagane jest** stosowanie prawidłowego, tj. zgodnego z RFC 8259¹²⁶⁾, formatu JSON.

W kontekście udostępniania danych przez API najważniejsze jest zorganizowanie danych, nazw właściwości, pól służących do nawigacji między stronami kolekcji, a także mechanizmów do wyszukiwania i filtrowania, w sposób jak najbardziej zrozumiały dla użytkownika – programisty, który będzie miał za zadanie skorzystać z udostępnionych danych. W tym celu należy przyjąć standard JSON API.

Zaleca się przy tym organizowanie danych w strukturę, która pozwoli w łatwy sposób rozszerzyć ją o semantyczny kontekst, wymagany na wyższych poziomach otwartości danych. Użycie standardów wskazywanych w wyższych poziomach otwartości nie może kolidować z organizacją danych wynikającą ze standardu JSON API. W celu osiągnięcia kolejnych poziomów otwartości należy stworzyć dane kontekstowe przekazywane w nagłówku HTTP.

5.2. JSON na 4. (****) i 5. (*****) poziomie otwartości

Na 4. (****) poziomie otwartości wymagane jest stosowanie standardu JSON-LD i opatrzenie danych pełnym kontekstem semantycznym. To oznacza mapowanie wszystkich właściwości udostępnianych danych odpowiednim URI w modelu RDF. Użycie standardu JSON-LD nie może kolidować z organizacją danych wynikającą ze standardu JSON API. Dane kontekstowe należy przekazywać w nagłówku http. 5. (*****) poziom otwartości wymaga dodania do zasobów połączeń z innymi danymi. Szczegóły dotyczące 4. i 5. stopnia wraz z przykładami opisano w rozdziale 7.

6. Szczegółowe wymagania dla zasobów w formacie XML

XML to uniwersalny język znaczników, pozwalający na reprezentowanie i przechowywanie różnych danych w ustrukturyzowany sposób. XML pozwala na identyfikowanie znaczników jako URI, co pozwala stosować go do udostępniania danych na wszystkich poziomach otwartości. Obecnie format XML w wielu kontekstach zastępowany jest prostszymi formatami. Dopuszcza się stosowanie formatu XML, jednocześnie zalecając stosowanie formatów JSON lub CSV. Dane w formacie XML udostępnia się w sposób analogiczny do udostępniania w formacie JSON-LD.

6.1. XML na 3. (***) poziomie otwartości

Na tym poziomie wymagane jest stosowanie dowolnej poprawnej struktury XML, w której właściwości obiektów są reprezentowane przez elementy XML.

¹²⁶⁾ <https://tools.ietf.org/html/rfc8259>.

6.2. XML na 4. (****) i 5. (*****) poziomie otwartości

Wymaga się, aby na tych poziomach była określona semantyka (identyfikacja przez URI) wszystkich elementów za pomocą mechanizmu XML Namespaces¹²⁷⁾ i dedykowanych dla tych danych schematów.

7. Szczegółowe wymagania dla zasobów na 4. (****) i 5. (*****) poziomie otwartości (RDF/XML, Turtle, N3, JSON-LD, N-tripels, RDFa)

7.1. Model danych RDF

RDF (ang. Resource Data Framework) jest językiem opisującym model danych dla metadanych oparty na trójelementowych wyrażeniach (ang. triples). Stanowi standard wymiany metadanych przez sieć. RDF ma wiele sposobów serializacji (formaty plików) umożliwiających rozpowszechnianie danych między ludźmi i co ważniejsze ich automatyczne przetwarzanie. Zgodnie z propozycją Tima Bernersa-Lee semantykę Internetu można opisać za pomocą trójelementowych wyrażen dotyczących zasobów, składających się z podmiotu, orzeczenia i dopełnienia (ang. subject, predicate, object). Podmiot oznacza zasób, a orzeczenie oznacza własność zasobu i wyraża związek między podmiotem a obiektem¹²⁸⁾.

Podstawowy model danych składa się z następujących obiektów:

Zasoby

a) zasoby (ang. resources)

Wszystkie obiekty opisane w wyrażeniach RDF nazywane są zasobami. Zasobem może być cała strona internetowa, np. dokument HTML. Zasób może być częścią strony internetowej, np. określony element HTML lub XML w źródle dokumentu. Zasobem może być również cały zbiór stron, np. cała witryna internetowa. Zasób może być także obiektem, który nie jest bezpośrednio dostępny za pośrednictwem sieci, np. drukowana książka. Zasoby są zawsze identyfikowane przez URI. Rozszerzalność URI pozwala na wprowadzenie identyfikatorów dla każdego obiektu znajdującego się w sieci internetowej (URL) lub poza nią,

b) właściwości (ang. properties)

Właściwość – to określony aspekt, cecha, atrybut lub relacja używane do opisu zasobu. Każda właściwość ma przypisane znaczenie, określające jej dozwolone wartości, typy zasobów, które może opisać, oraz jej związek z innymi właściwościami. Sposób wyrażania cech właściwości jest definiowany przez schemat RDF (RDFS),

c) wyrażenia (ang. statements)

Określony zasób wraz z nazwaną właściwością oraz wartość tej właściwości dla tego zasobu jest wyrażeniem RDF. Wartość właściwości może być innym zasobem (określonym przez URI) lub literałem, czyli prostym ciągiem znaków lub innym typem pierwotnym (np. data, liczba całkowita).

7.2. Serializacja modelu RDF

Model RDF ma wiele sposobów serializacji w celu wymiany danych przez sieć oraz automatycznego przetwarzania danych między usługami oraz aplikacjami. Formaty plików (RDF/XML, Turtle, N3, JSON-LD, N-tripels, RDFa i inne), za pomocą których zapisuje się model RDF, są równoważne. Oznacza to, że ich zawartość można jednoznacznie transformować z jednego formatu na drugi. Transformację można wykonać za pomocą darmowych konwerterów, które są dostępne w sieci. Poniżej krótki opis każdego formatu.

¹²⁷⁾ <https://www.w3.org/TR/xml-names/>.

¹²⁸⁾ <https://www.w3.org/TR/rdf-concepts/>.

Tabela 8. Formaty plików, za pomocą których zapisuje się model RDF

Format	Opis
RDF/XML	RDF/XML to format zdefiniowany przez W3C, do wyrażania modelu metadanych RDF jako dokumentu XML. RDF/XML często jest błędnie określany jako RDF, ponieważ został wprowadzony wśród innych specyfikacji W3C definiujących RDF i był historycznie pierwszym standardowym formatem serializacji RDF W3C ¹²⁹⁾
Turtle	Jest to skrót od ang. Terse RDF Triple Language. Składnia tego formatu jest bardzo podobna do języka zapytań SPARQL. Turtle został zdefiniowany przez Dave'a Becketta jako podzbiór języka Notation3 (N3) oraz nadzbiór minimalnego formatu N-Triples. Turtle może serializować tylko prawidłowe grafy RDF i jest alternatywą dla RDF/XML. W przeciwieństwie do RDF/XML, Turtle nie bazuje na XML. Jest ogólnie uznawany za czytelniejszy i łatwiejszy do edycji ręcznej niż jego odpowiednik RDF/XML ¹³⁰⁾
Notation3 (N3)	Notacja3 lub N3 jest formatem znacznie czytelniejszym dla ludzi niż notacja RDF/XML. Format został opracowywany przez Tima Bernersa-Lee i innych członków społeczności semantycznej. N3 ma kilka cech wykraczających poza serializację modeli RDF, takich jak możliwość definiowania reguł wnioskowania ¹³¹⁾
N-Triples	N-Triples to liniowy, szeregowy format serializacji dla grafów RDF. Stanowi on podzbiór formatu Turtle. N-Triples został opracowany przez Dave'a Becketta z University of Bristol i Art Barstow z World Wide Web Consortium (W3C). N-Triples został zaprojektowany jako prostszy format niż Turtle, a zatem łatwiejszy do analizy i generowania przez oprogramowanie. Ze względu na brak niektórych skrótów zapewnianych przez inne serializacje RDF (takie jak CURIE i zasoby zagnieżdżone) uciążliwe może być ręczne wpisywanie dużej liczby danych ¹³²⁾
JSON-LD	JSON-LD został zaprojektowany w oparciu o koncepcję „kontekstu”, aby zapewnić dodatkowe odwzorowania z JSON na model RDF. Kontekst łączy właściwości obiektu w dokumencie JSON z pojęciami w ontologii. Aby zmapować składnię JSON-LD na RDF, JSON-LD umożliwia rzutowanie wartości do określonego typu lub ich otagowanie. Kontekst można osadzić bezpośrednio w dokumencie JSON-LD lub umieścić w osobnym pliku i odwoływać się do niego z różnych dokumentów (z tradycyjnych dokumentów JSON za pośrednictwem nagłówka HTTP Link) ¹³³⁾
RDFa	Format RDFa to rekomendacja W3C, który dodaje zestaw rozszerzeń na poziomie atrybutów do HTML, XHTML i różnych typów dokumentów opartych na XML do osadzania różnorodnych metadanych w dokumentach internetowych. RDFa umożliwia osadzanie wyrażen trójelementowych RDF w dokumentach XHTML ¹³⁴⁾

Składnia każdego formatu serializacji RDF składa się z dwóch części. W pierwszej znajdują się deklaracje przestrzeni nazw (słowników), które określają semantykę modelu RDF. Druga część jest opisowa i prezentuje zasoby (Identyfikator URI), cechy (słowniki) oraz ich wartości w postaci zasobów lub literalów.

¹²⁹⁾ <https://www.w3.org/TR/rdf-syntax-grammar/>.

¹³⁰⁾ <https://www.w3.org/TR/turtle/>.

¹³¹⁾ <https://www.w3.org/TeamSubmission/n3/>.

¹³²⁾ <https://www.w3.org/TR/n-triples/>.

¹³³⁾ <https://www.w3.org/TR/json-ld11/>.

¹³⁴⁾ <https://www.w3.org/TR/rdfa-core/>.

7.3. Schemat RDF

RDFS (ang. Recourse Data Framework Schema) jest językiem opisującym semantycznie model metadanych w grafie RDF. Zbiór klas odnoszący się do określonego celu lub określonej dziedziny stanowi schemat RDF i nadaje znaczenie pojęciowe obiektom występujących w grafie RDF. Klasy są zorganizowane w hierarchię i można je rozszerzać przez klasy podrzędne. Do klasy może należeć dowolna liczba wierzchołków z grafu RDF. Wierzchołek grafu RDF może być przydzielony do dowolnej liczby klas. W programowaniu obiektowym to definicja klasy determinuje obiekty należące do tej klasy. W RDFS to zbiór wierzchołków stanowi o tym, czym jest klasa. Schemat RDF opisuje się językiem RDF.

RDF Schema stanowi bazę dla języka OWL, pozwalającego na opisywanie zasobów w postaci ontologii. Struktury zapisane za pomocą RDFS są nazywane słownikami RDF (ang. RDF vocabularies). Język RDF Schema stanowi rekomendację W3C¹³⁵⁾.

7.4. Dane połączone

Dane połączone (ang. linked data) to najlepsze praktyki dotyczące tworzenia w sieci powiązań między danymi pochodzącymi z różnych źródeł. Działania mają na celu umożliwienie eksploracji przez człowieka oraz maszyny sieci danych. Dzięki połączonym ze sobą danym można znaleźć inne, powiązane dane¹³⁶⁾. Tim Berners-Lee w 2006 r. przedstawił zestaw zasad (ang. linked data principles) publikowania danych w sieci w taki sposób, że wszystkie publikowane dane stają się częścią jednej globalnej przestrzeni danych:

- ✓ Stosuj URI jako identyfikator zasobów,
- ✓ Stosuj schemat http dla URI, tak aby ludzie mogli wyszukiwać zasoby; gdy ktoś wyszukuje URI, dostarcz użytecznych informacji,
- ✓ Zamieszczaj powiązania do innych URI, tak aby można było znaleźć więcej informacji¹³⁷⁾.

Od strony technicznej połączone dane to dane publikowane w sieci opisane w standardzie RDF co umożliwi ich automatyczny odczyt oraz jednoznaczne powiązanie za pomocą URI kolejnych zasobów. Takie podejście pozwala na budowanie powiązań między zasobami, a tych z kolejnymi, co zwiększa jednoznaczny opis semantyczny zasobów w sieci wykorzystując do tego reguły wnioskowania.

Mając na względzie zalecenia przedstawione w niniejszym standardzie oraz konieczność poniesienia kosztów finansowych podczas budowy danych na najwyższych stopniach otwartości, rekomenduje się przeprowadzenie analizy kosztów i korzyści tworzenia danych połączonych.

8. Profil aplikacji DCAT-AP

Profil aplikacji DCAT-AP (ang. Application Profile) jest specyfikacją opartą na słowniku katalogu danych (DCAT), opracowanym przez W3C. Został on przygotowany w celu opisanego katalogów danych sektora publicznego, zestawów danych oraz promowania specyfikacji do wykorzystania przez portale w całej Europie¹³⁸⁾. DCAT-AP pozwala na ujednoczenie metadanych zbiorów dla portali danych w Europie. Zastosowanie wspólnego profilu i promowanie go wśród państw członkowskich znacznie poprawi interoperacyjność między katalogami danych oraz wymianę danych między państwami członkowskimi. Wdrożone zasady płynące z DCAT-AP powinny prowadzić do harmonizacji wymiany danych dzięki ich łatwemu przeszukiwaniu i wykorzystywaniu. DCAT-AP jest specyfikacją stosowaną przez Europejski Portal Danych oraz wiele innych europejskich portali otwartych danych. Zaleca się stosowanie specyfikacji do opisywania zbiorów danych sektora publicznego i umożliwienia wymiany opisów zbiorów danych między portalem Dane.gov.pl a innymi krajowymi i europejskimi portalami udostępniającymi publiczne dane.

¹³⁵⁾ <https://www.w3.org/TR/rdf-schema/>.

¹³⁶⁾ <https://www.w3.org/DesignIssues/LinkedData.html>.

¹³⁷⁾ Tamże.

¹³⁸⁾ <https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe/about>.

Specyfikacja DCAT-AP, definiująca profil aplikacyjny dla portali danych w Europie, określa podział klas i właściwości (wykorzystywanych do opisu schematu metadanych) na „wymagane”, „rekomendowane” i „opcjonalne”. W celu zapewnienia minimalnego poziomu interoperacyjności standard DCAT-AP definiuje słowniki kontrolowane, które pozwalają na opisywanie wybranych metadanych w wystandaryzowany sposób. Ponadto standard DCAT-AP rekomenduje opracowanie i publikację własnych otwartych słowników umożliwiających strukturyzowanie metadanych zgodnie z określonym schematem¹³⁹⁾.

9. Lokalizacja źródeł danych publicznych

Jednym z etapów udostępniania danych jest wybór lokalizacji źródła danych publicznych:

- portal Dane.gov.pl – opcja o wysokim stopniu pewności dostępu i bezpieczeństwa danych, ale zależna od dysponentów danych publikujących swoje zasoby w Dane.gov.pl,
- lokalny system teleinformatyczny, w tym lokalne repozytorium danych publicznych powszechnie dostępne w sieci teleinformatycznej – opcja obciążona ryzykiem utraty dostępu do danych wskutek zmian organizacyjnych podmiotu.

Udostępnianie zbiorów danych w portalu Dane.gov.pl jest możliwe przez:

- a) dodanie pliku z lokalnego komputera dostawcy danych, który jest następnie zapisywany i przechowywany w bazie danych centralnego repozytorium, albo
- b) wprowadzenie adresu URL strony internetowej repozytorium, w ramach której udostępniono zbiór danych (zewnętrzny link do źródła danych).

Dostawca danych decyduje o wyborze lokalizacji źródła danych, mając na względzie:

- posiadaną infrastrukturę teleinformatyczną umożliwiającą gromadzenie w niej informacji publicznych (zbiorów danych) oraz ich udostępnianie,
- możliwości techniczne przechowywania informacji w portalu Dane.gov.pl.

Dostawcom danych, udostępniającym zbiory danych w portalu Dane.gov.pl z zewnętrznych lokalizacji, zaleca się zapewnienie trwałej dostępności każdego opublikowanego zasobu. W przypadku np. usunięcia danych na zewnętrznej lokalizacji lub zmian organizacyjnych podmiotu zaleca się dostawcom danych przeniesienie i przechowywanie danych źródłowych w portalu Dane.gov.pl.

Zalecenia dotyczące organizacji i lokalizacji źródeł danych:

1. Publikowanie każdego zasobu (jednostki informacji publicznej) w odrębnym pliku.
2. Stosowanie wersjonowania zasobów, zamiast zastępowania/nadpisywania (każdy nowy plik pod nowym adresem URL).
3. Utrzymywanie niezmiennego adresu URL każdego zasobu (pliku).
4. Zapewnienie trwałej dostępności każdego, już opublikowanego zasobu.
5. Tworzenie lokalnych repozytoriów danych publicznych lub wykorzystywanie systemów scentralizowanych, w celu ochrony udostępnionych zasobów i ich URL przed zmianami organizacyjnymi podmiotów.

10. Procedura publikacji danych w portalu Dane.gov.pl

Etap wyboru zasobu informacyjnego:

1. Analiza posiadanych praw do zasobu umożliwiających jego udostępnienie do ponownego wykorzystywania i analiza jego dostępności (zasób w dyspozycji podmiotu).

¹³⁹⁾ Ekspertyza sporządzona na zlecenie Ministerstwa Cyfryzacji pt.: „Zapewnienie zgodności portalu dane.gov.pl z DCAT-AP. Ekspertyza dotycząca określenia zasad opisujących zbiory danych publicznych zgodnych z DCAT-AP do zaimplementowania w portalu Dane.gov.pl”.

2. Zapewnienie sobie praw do tzw. danych surowych (nieprzetworzonych do postaci np. raportów, wykresów i nieopatrzonych kontekstem lub interpretacją) lub zamawianie danych na otwartych licencjach (o ile w ogóle są konieczne) i w formatach umożliwiających powtórne wykorzystywanie (o odpowiednio wysokim stopniu otwartości).
3. Możliwość udostępniania w ramach domeny publicznej.
4. Ocena merytoryczna potencjału do powtórnego wykorzystywania (identyfikacja popytu na dane).

Etap analizy formatu danych i dostosowania do zasad otwartości i jakości:

1. Ocena dostępnych formatów danych.
2. Ocena możliwości modyfikacji zasobów do formatu o wyższym stopniu otwartości, np. przez strukturyzowanie danych.
3. Podział złożonych zasobów na zasoby o prostej strukturze, np. podział skroszytów na pojedyncze arkusze.
4. Posiadanie możliwości technicznych i organizacyjnych do niezbędnej przebudowy struktury i modyfikacji zawartości plików (np. usuwanie zakłóceń/agregacji struktury arkusza, usuwanie formuł oraz odniesień do danych poza źródłem, łączenie treści podzielonych między polami itp.) oraz poprawy jakości danych (czyszczenie) w celu zwiększenia użyteczności zasobów przeznaczonych do udostępniania.

Etap publikowania danych (formaty, miejsce publikacji, opis i metadane):

1. Opracowanie/przygotowanie danych w możliwie najwyższym stopniu otwartości.
2. Wybór lokalizacji źródła danych przeznaczonych do udostępniania:
 - a) lokalny system teleinformatyczny, w tym lokalne repozytorium – opcja obciążona ryzykiem utraty dostępu do danych wskutek zmian organizacyjnych podmiotu,
 - b) portal Dane.gov.pl – opcja o wysokim stopniu pewności dostępu i bezpieczeństwa danych.
3. Sporządzenie opisu zasobu, precyzyjnie charakteryzującego zawartość oraz istotne cechy wyróżniające i wartości brzegowe (np. data/okres, obszar, sektor gospodarczy itp.).
4. Dobór słów kluczowych (postulowane rozwiązanie: wybór ze słownika).

Etap nadzoru nad aktualizacją zasobu, trwałości i dostępności danych:

Wypracowanie wewnętrznego systemu udostępniania danych, w szczególności określenie:

1. źródeł danych,
2. tymczasowej i docelowej alokacji zasobów,
3. zasad opracowania zasobów, formatów plików danych, sposobu sporządzania opisów i stosowania słów kluczowych,
4. osób upoważnionych do przetwarzania i publikowania zasobów oraz pełniących nadzór nad aktualnością danych,
5. osób odpowiedzialnych za udostępnianie,
6. procedury obiegu dokumentów i danych.

Etap oceny wykorzystania i informacja zwrotna od użytkowników, analiza efektów udostępnienia danych:

1. Publikowanie:
 - a) liczby odsłon,
 - b) liczby użytkowników obserwujących zasób,
 - c) liczby pobrań,
 - d) liczby abonentów zasobu,
 - e) opinii na temat zasobów;
2. Tworzenie rankingów:
 - a) otwartości danych,
 - b) popularności zasobów,

- c) jakości danych,
- d) największej liczby zastosowań w ramach ponownego wykorzystywania.

11. Podsumowanie

Poniższa tabela zawiera zalecenia dotyczące udostępniania danych na określonych poziomach otwartości. Jak wskazano w pkt 1, **nie zaleca się publikowania danych na pierwszym poziomie otwartości.**

Tabela 9. Zalecenia

Rozwiązanie/standard	Poziom 2 **	Poziom 3 ***	Poziom 4 ****	Poziom 5 *****
Udostępnianie przez API	–	zalecane	zalecane	zalecane
Format CSV	zalecany	zalecany	dopuszczalny	dopuszczalny
Format JSON	zalecany	zalecany	dopuszczalny	dopuszczalny
Format XML	dopuszczalny	dopuszczalny	dopuszczalny	dopuszczalny
Format RDF/XML	dopuszczalny	dopuszczalny	zalecany	zalecany
Format Turtle	dopuszczalny	dopuszczalny	zalecany	zalecany
Format Notation3 (N3)	dopuszczalny	dopuszczalny	zalecany	zalecany
Format N-Triples	dopuszczalny	dopuszczalny	zalecany	zalecany
Format JSON-LD	dopuszczalny	dopuszczalny	zalecany	zalecany
Format RDFa	dopuszczalny	dopuszczalny	zalecany	zalecany
Format HTML	dopuszczalny	dopuszczalny	niezalecany	niezalecany
Format ODS	dopuszczalny	dopuszczalny	niedopuszczalny	dopuszczalny
Format dBase	dopuszczalny	dopuszczalny	niedopuszczalny	niedopuszczalny
Formaty Office Open XML (XLSX, DOCX)	dopuszczalne	niedopuszczalne	niedopuszczalne	niedopuszczalne
Formaty DOC, XLS, RTF	dopuszczalne	niedopuszczalne	niedopuszczalne	niedopuszczalne
Format PDF	niedopuszczalny	niedopuszczalny	niedopuszczalny	niedopuszczalny
Formaty JPEG, PNG	niedopuszczalne	niedopuszczalne	niedopuszczalne	niedopuszczalne
Format TXT	niedopuszczalny	niedopuszczalny	niedopuszczalny	niedopuszczalny
Formaty ZIP, 7z, Gzip, Bzip2	dopuszczalne	dopuszczalne	dopuszczalne	dopuszczalne
Własnościowe rozszerzenia ZIP	dopuszczalne	niedopuszczalne	niedopuszczalne	niedopuszczalne
Format RAR	dopuszczalny	niedopuszczalny	niedopuszczalny	niedopuszczalny
Dane przestrzenne	dopuszczalne	dopuszczalne	dopuszczalne	dopuszczalne
Typy danych z XML Schema	zalecane	zalecane	–	–
XML Namespaces	–	–	wymagane	wymagane

Załącznik nr 3

Standard API
(interfejsu
programistycznego
aplikacji)

Załącznik nr 3. Standard API

1. Kategorie obecnych i potencjalnych odbiorców danych wykorzystujących API

Odbiorcami korzystającymi z danych publicznych (obecnie i docelowo) są zarówno osoby fizyczne, jak i przedsiębiorstwa, jednostki administracji publicznej (centralnej i samorządu terytorialnego), organizacje pozarządowe, podmioty kultury, podmioty tworzące system szkolnictwa wyższego i nauki, o których mowa w art. 7 ust. 1 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce i inni interesariusze. Ponieważ dane publiczne mogą opisywać rzeczywistość w dowolnym aspekcie, to również interesariusze mogą przetwarzać te dane na wiele sposobów. To kształtuje bezpośrednio charakterystykę korzystania z API oraz wpływa na wymagania przed nim stawiane.

Większość użytkowników API będzie korzystała z produktów pochodnych opracowanych na podstawie danych pochodzących z API (usług, aplikacji komputerowych, informacji przetworzonych, know-how) często nieświadomie. Jest to tzw. grupa użytkowników pośrednich, która bezpośrednio nie korzysta z API, ale czerpie z niego korzyści i może mieć wpływ na pożądany kierunek jego ewolucji.

Pozostałych, bezpośrednich użytkowników można podzielić na dwie grupy: użytkownicy okazjonalni (*ad hoc*), użytkownicy profesjonalni.

Szczegółowa charakterystyka grup użytkowników znajduje się w załączniku nr 1 do „Standardu API”.

1.1. Rekomendacje szczegółowe ze względu na potrzeby użytkowników okazjonalnych

Rekomenduje się następujące wymagania w zakresie API ze względu na potrzeby użytkowników okazjonalnych:

1. API niweluje technologiczną barierę wejścia, rozumianą jako konieczność posiadania przygotowania technicznego do wykonywania podstawowych interakcji z niego.
2. API wymaga niskich nakładów czasu i pracy oraz nauki do korzystania z API.
3. API pozwala na korzystanie z zasobów w sposób okazjonalny bez uwierzytelniania.
4. API ma prosty w obsłudze interfejs webowy, który umożliwia korzystanie z API bez konieczności wpisywania rozbudowanego kodu, lecz przez składanie gotowych jego elementów, i natychmiastowe uzyskanie wyników zapytania oraz wykonywanie na nich podstawowych działań: filtrowania, sortowania, przeszukiwania i linkowania do poszczególnych URI.
5. API umożliwia użytkownikom zachowanie anonimowości, a do korzystania z kreatora nie wymaga rejestracji, zakładania konta itp.

1.2. Rekomendacje szczegółowe ze względu na potrzeby użytkowników profesjonalnych

Rekomenduje się następujące wymagania w zakresie API ze względu na potrzeby użytkowników profesjonalnych:

1. API dostarcza interfejs zgodny z uznanymi standardami, ze spójnie nazwanymi zasobami.
2. API jest gotowe do natychmiastowego korzystania oraz ma niską barierę wejścia.
3. API dostarcza dane z kompletem metadanych.
4. API pokazuje korzyści z jego używania w kontekście konkretnych zastosowań (np. przez przykłady zawarte w dokumentacji).
5. API zaspokaja zróżnicowane potrzeby deweloperów.
6. API dostosowuje się do zmian i nowych potrzeb (zbiera propozycje od użytkowników).
7. API zapewnia podstawową dokumentację usług oraz udostępnia pełną informację o strukturze danych.
8. API od początku przewiduje obsługę wersjonowania API.

9. API obsługuje manipulację danych: sortowanie, filtrowanie, przeszukiwanie.
10. API zapewnia stabilność, wydajność, skalowalność oraz bezpieczeństwo.
11. API jest traktowane jako produkt wspierający innowacyjność. W praktyce oznaczać to będzie np. nacisk na sprawne rozwiązywanie problemów i obsługę pytań deweloperów przez zespół obsługujący API lub zespół związany z centralnym repozytorium.
12. API zapewnia sprawną rejestrację nowych aplikacji.
13. API zezwala na intensywne wykorzystanie i realizowanie gwałtownych wzrostów obciążeń
14. API dostarcza dane umożliwiające przesłanie (np. e-mailem) pytań, uwag, zgłoszeń błędów itp. bezpośrednio do osoby / komórki odpowiedzialnej za zasób API.

1.3. Rekomendacje ogólne w kontekście użytkowników API

W kontekście użytkowników API formułuje się następujące rekomendacje standaryzujące API:

1. API ma spójny interfejs, korzystający z uznanych standardów światowych.
2. API zapewnia dokumentację usług, w szczególności pełną informację o strukturze udostępnianych danych.
3. API wymaga niskich nakładów pracy do korzystania z niego w podstawowym zakresie za pomocą prostego w obsłudze interfejsu webowego. Interfejs musi umożliwiać skorzystanie z API bez konieczności wpisywania rozbudowanego kodu, lecz przez składanie gotowych elementów zapytań. Interfejs musi dostarczać szybko rezultatów – danych prezentowanych w przystępny dla odbiorcy sposób z możliwością interakcji: filtrowania, sortowania i przeszukiwania pod kątem więcej niż jednego warunku. Musi pozwalać na korzystanie z zasobów bez uwierzytelniania.
4. API zapewnia gotowość do natychmiastowego korzystania (niska bariera wejścia).
5. API zapewnia obsługę wersjonowania kolejnych odsłon API.
6. API obsługuje manipulację danych: sortowanie, filtrowanie, przeszukiwanie po więcej niż jednym warunku.
7. API zapewnia skalowalność, stabilność, wydajność oraz bezpieczeństwo.
8. API określa zasady limitów wykorzystania, jeżeli takie są wprowadzone.
9. W przypadku możliwości korzystania z kluczy API, umożliwia natychmiastowe wygenerowanie odpowiednich API-Key.

Mając na względzie zalecenia przedstawione w niniejszym standardzie oraz konieczność poniesienia kosztów finansowych podczas budowy i utrzymania właściwego API, rekomenduje się przeprowadzenie analizy kosztów i korzyści budowy interfejsu API.

2. Architektura usług sieciowych REST

Standard API jest oparty na standardzie REST. Ponadto referencyjnie wykorzystuje format reprezentacji danych zdefiniowany przez standard JavaScript Object Notation, czyli JSON¹⁴⁰⁾. Jako standard notacji JSON przyjmuje się JSON Schema, a jako standard komunikatów (zapytania i odpowiedzi między użytkownikiem a systemem) przyjmuje się JSON API¹⁴¹⁾. Kontekst ułatwiający korzystanie z danych w sposób automatyczny jest generowany zgodnie ze standardem JSON-LD¹⁴²⁾.

Opis przyjętej architektury znajduje się w załączniku nr 2 do „Standardu API”.

Rekomendacje dotyczące architektury usług sieciowych REST

¹⁴⁰⁾ RFC 7159, <https://tools.ietf.org/html/rfc7159>.

¹⁴¹⁾ JSON API – Specification <http://jsonapi.org/format/>.

¹⁴²⁾ JSON for Linking Data <https://json-ld.org/spec/latest/json-ld/>.

W kontekście stosowania standardu REST formułuje się następujące rekomendacje standaryzujące:

1. Budowa API jest oparta na zasobach.
2. Manipulacja zasobami powinna być oparta na reprezentacji zasobów.
3. API wspiera obsługę buforowania, w tym zapewnia obsługę polecenia HEAD.
4. API jest wyposażone w jednolity interfejs z unikalną identyfikacją zasobów, dopuszczalne jest istnienie więcej niż jednego URI wskazującego zasób w uzasadnionych przypadkach (wersjonowanie API, różne wersje językowe itp.).
5. API funkcjonuje w modelu klient – serwer, w którym klient komunikuje się z serwerem tylko przez interfejs API zgodny ze standardem API.
6. API zapewnia możliwość wykorzystania samoopisujących się komunikatów.
7. API zapewnia informację o strukturze danych.
8. API zapewnia dostęp do zasobów w formacie JSON; opcjonalne jest udostępnianie danych również w formacie XML klientom oczekującym tego za pomocą nagłówka „Accept” lub parametru „&format=” w URL.
9. Jako standard notacji JSON przyjmuje się JSON Schema.
10. Dla danych spełniających założenia „trzech gwiazdek”¹⁴³⁾ otwartości – wymagane jest stosowanie podstawowego standardu JSON.
11. Dla danych spełniających założenia „czterech gwiazdek”¹⁴⁴⁾ otwartości – wymagane jest dodanie kontekstu obejmującego wszystkie atrybuty zgodnie ze standardem JSON-LD.
12. Dla danych spełniających założenia „pięciu gwiazdek”¹⁴⁵⁾ otwartości – wymagane jest stosowanie linkowania między zasobami.
13. Interakcja z API jest możliwa wyłącznie przy użyciu szyfrowanego połączenia HTTPS.

3. Reguły nazewnictwa zasobów oraz parametry usług API

REST wykorzystuje identyfikatory zasobu do jednoznacznego określania konkretnego zasobu. Dla zachowania spójności zalecane jest stosowanie jednolitych reguł nazewnictwa, gdyż mają one istotny wpływ na łatwość użycia, prostotę i zrozumiałość funkcjonowania API opartego na REST.

Rekomendacje dotyczące reguł nazewnictwa zasobów oraz parametrów usług API

W kontekście nazewnictwa zasobów API formułuje się następujące rekomendacje standaryzujące:

1. Należy stosować proste i objaśniające nazewnictwo.
2. Należy oznaczać zasoby rzeczownikami (np. „czujniki”, „samochody”).
3. Należy stosować rzeczowniki w liczbie mnogiej (np. „formularze” zamiast „formularz”, „czujniki” zamiast „czujnik” itp.).
4. Nie wolno używać czasowników lub sformułowań opisujących czynności (np. „pobierz formularze” lub „wyswietla listę czujników”).
5. Należy stosować nazewnictwo konkretnie wskazujące na zasób zamiast nazewnictwa, które wymaga znajomości kontekstu do zrozumienia jego znaczenia (np. „przeglady-rejestracyjne” zamiast „przeglady”).
6. Nie wolno stosować polskich znaków w polach będących identyfikatorem zasobów (np.: „usługi”), zwłaszcza gdy są one wykorzystywane w URI zasobu.

¹⁴³⁾ Zgodnie z klasyfikacją zaproponowaną przez Tima Bernersa-Lee, <http://5stardata.info>.

¹⁴⁴⁾ Tamże.

¹⁴⁵⁾ Tamże.

7. Należy oznaczać zasoby za pomocą małych liter¹⁴⁶⁾ (np. „pojazdy-zarejestrowane” zamiast „Pojazdy-Zarejestrowane”).
8. Należy używać łączników („-”) do łączenia fraz złożonych (np. „pojazdy-zarejestrowane” zamiast „pojazdyzarejestrowane”).
9. Nie wolno używać znaków podkreślenia („_”) do łączenia fraz złożonych (np. „pojazdy-zarejestrowane” zamiast „pojazdy_zarejestrowane”) lub innych znaków specjalnych.

4. Opis składni oraz elementów URI

URI (ang. Uniform Resource Identifier) jest unikalnym identyfikatorem zasobów, stosowanym zgodnie z RFC 1736¹⁴⁷⁾, RFC 1737¹⁴⁸⁾ oraz ustandaryzowanym zgodnie z RFC 3986¹⁴⁹⁾.

Definicje, szczegółowy opis składni oraz przykłady zastosowań znajdują się w załączniku nr 3 do Standardu API.

Rekomendacje dotyczące składni oraz elementów URI

W kontekście składni URI formułuje się następujące rekomendacje standaryzujące:

1. Domena w URI musi jednoznacznie wskazywać na właściciela API. Domena ta musi udostępniać funkcjonalność API przez subdomenę „api”, chyba że nie jest to możliwe z uzasadnionego powodu.
2. API może być wyposażone w portal dla deweloperów, który zawiera dokumentację API, informacje kontaktowe, generator kluczy API, przykładowy kod itp. Portal powinien być integralną częścią API i być dostępny przez subdomenę „dev”. W przypadku braku osobnego portalu, na głównej stronie API muszą znaleźć się podstawowe informacje pozwalające na korzystanie z niego – w tym informacje o limitach oraz o możliwości uzyskania klucza API, jeżeli te opcje są dostępne.
3. Znak „/” (ukośnik, ang. forward slash) służy oznaczeniu zależności hierarchicznej między zasobami.
4. Stosowanie znaku „/” (ukośnik, ang. forward slash) na końcu URI nie jest zalecane.
5. Ścieżki URI są oznaczane z użyciem małych liter.
6. W URI nie stosuje się rozszerzeń plików oraz nazw formatów danych. Informacje te są przekazywane przez nagłówki HTTP lub w metadanych zasobów.

5. Opis dostępu do zasobów powiązanych za pomocą URI

W REST klient dostarcza stan przez ciało zapytania, parametry zawarte w kwerendzie i nagłówki. Usługa API dostarcza klientowi stan przez ciało komunikatu, kody i nagłówki odpowiedzi. Technicznie oznacza to komunikację za pomocą hiperlinków i hipertekstu.

API musi używać techniki HATEOAS, czyli linków zawartych w ciele lub nagłówkach odpowiedzi, które zawierają identyfikatory URI do zasobów powiązanych.

6. Manipulowanie zasobami – filtrowanie, sortowanie, stronicowanie

W założeniu szereg klienckich aplikacji będzie korzystać z zasobów API automatycznie. Takie aplikacje rzadko realizują zadania wymagające pobrania z API zbioru danych w pełnym przekroju informacyjnym oraz pełnej zawartości rekordów. Znacznie częściej potrzebują odpowiednio spreparowanego, niewielkiego wycinka zamiast pełnego zakresu dostępnych danych. Wycinek ten w

¹⁴⁶⁾ RFC 2616 ścieżka i zapytanie w URL mają rozróżnianą wielkość liter <http://w3.org/Protocols/rfc2616/rfc2616-sec3.html#sec3.2.3>.

¹⁴⁷⁾ RFC 1736, <https://www.ietf.org/rfc/rfc1736.txt>.

¹⁴⁸⁾ RFC 1737, <https://www.ietf.org/rfc/rfc1737.txt>.

¹⁴⁹⁾ RFC 3986, <https://www.ietf.org/rfc/rfc3986.txt>.

stosunku do oryginalnego zbioru może być filtrowany względem zakresu pobieranych danych, w kontekście określonych kryteriów, porcji danych pobieranej za jednym razem oraz w odpowiednim uporządkowaniu. Dlatego założono, że API udostępnia funkcjonalność filtrowania, sortowania i stronicowania zasobów, która będzie określana mianem manipulacji danymi.

Zapytania i odpowiedzi systemu są realizowane zgodnie ze standardem JSON API w wersji 1.0 lub nowszej, chyba że standard API stanowi inaczej. W odpowiedziach zaleca się uzupełnienie JSON API o JSON-LD w celu dodania kontekstu koniecznego do automatycznego przetwarzania danych API.

Podstawowe zasady filtrowania, sortowania i stronicowania zasobów wraz z przykładami znajdują się w załączniku nr 4 do „Standardu API”.

Rekomendacje dotyczące manipulowania zasobami

W kontekście manipulowania reprezentacjami zasobów – filtrowaniem, sortowaniem i stronicowaniem – formułuje się następujące rekomendacje standaryzujące:

1. Proste stronicowanie, filtrowanie i sortowanie reprezentacji zasobów odbywają się przez URI.
2. Równoległe API może obsługiwać zapytania przesyłane w ciele wiadomości w zakresie szerszym niż przez URI.
3. Standardem wykorzystywanym przez OpenAPI jest JSON API w wersji 1.0 lub nowszej, chyba że niniejsza specyfikacja wprowadza od tego odstępstwa.
4. W odpowiedziach zaleca się dostarczanie dodatkowego kontekstu za pomocą JSON-LD.
5. Zapytania asynchroniczne są obsługiwane przez kody HTTP o numerach 202 i 303.
6. Obsługa kwerend SQL jest zabroniona.

7. Metadane

Zasoby udostępniane przez API są wyposażane w metadane, stanowiące część odpowiedzi na zapytanie o reprezentację zasobu. Metadane to inaczej dane o danych, które, opisując zasoby, ułatwiają ich zrozumienie oraz zarządzanie nimi. Są szczególnie istotne w procesie maszynowego przetwarzania informacji, dlatego istotne jest wykorzystanie rodzajów metadanych tam, gdzie tylko jest to możliwe. W przypadku API metadane muszą uwzględniać lokalne wytyczne określone przez POD oraz rozporządzenie Rady Ministrów z dnia 12 marca 2014 r. w sprawie Centralnego Repozytorium Informacji Publicznej.

Tabela, w której przedstawiono minimalny zestaw metadanych dla zasobu, znajduje się w załączniku nr 5 do „Standardu API”.

Rekomendacje dotyczące metadanych

W kontekście określania minimalnego zestawu metadanych formułuje się następujące rekomendacje standaryzujące:

1. Zasoby są opisywane za pomocą minimalnego zestawu metadanych określonego w niniejszym dokumencie – chyba że określone metadane nie dotyczą konkretnego rodzaju treści.
2. Gdy metadane nie występują dla określonego rodzaju zasobów, nie są dołączane do metadanych danego zasobu.
3. Metadane określone przez dostawcę zasobu są określane zgodnie ze zdefiniowanymi zasadami nazewnictwa, w miarę możliwości z wykorzystaniem standardowych rozwiązań (schema.org, DCMI itp.).
4. W przypadku wprowadzania niestandardowych metadanych, konieczne jest dostarczenie słownika, do którego będą odnosiły się opisy metadanych.

8. Mechanizm linków do nawigacji

API zapewnia funkcjonalność linków do nawigacji. Polega ona na dołączaniu w ciele zasobu poniższych linków (oznaczanych jako „links”), zgodnie z JSON API 1.0 lub późniejszym.

Objaśnienia i przykłady stosowania znajdują się w załączniku nr 6 do „Standardu API”.

Rekomendacje dotyczące linków do nawigacji

W kontekście linków do nawigacji formułuje się następującą rekomendację standaryzującą: W odpowiedzi na przesłane przez użytkowników zapytania API udostępnia stany wraz ze zdefiniowanymi w JSON API 1.0 linkami ułatwiającymi poruszanie się między zasobami powiązanymi. Linki są oznaczane:

- a) *prev* (poprzedni),
- b) *self* (bieżący),
- c) *next* (następny),
- d) *first* (pierwszy),
- e) *last* (ostatni).

9. Wersjonowanie

Wersjonowanie jest jedną z najistotniejszych kwestii do rozważenia przy tworzeniu API. Od tego jak wersjonowanie zostanie zaplanowane i zaimplementowane zależy nie tylko wygoda, ale i możliwość korzystania z API w przyszłości oraz dostępność jego zasobów.

Zasady wersjonowania, metody oznaczania wersji wraz z przykładami oraz techniki zarządzania wersjami są zawarte w załączniku nr 7 do „Standardu API”.

Rekomendacje dotyczące wersjonowania API

W kontekście wersjonowania API formułuje się następujące rekomendacje standaryzujące:

1. Zmiany API oznacza się za pomocą standardu wersjonowania semantycznego API¹⁵⁰⁾.
2. Do oznaczenia i wywoływania obsługi wersją o konkretnym numerze ma znaczenie tylko pierwszy człon (MAJOR) wersji, a pozostałe (MINOR i PATCH) są ignorowane, przy czym zachowują jedynie wartość informacyjną w kontekście wersji interfejsu.
3. Obowiązkowe jest odwołanie się do wersji przez element URL (/api/vXX/).
4. Dopuszczalne jest odwoływanie się do wersji dodatkowo przez parametr URI „v=XX”, dodatkowy nagłówek „api-version=XX” lub w nagłówku „Accept:”.
5. Jeżeli wersja API nie została określona, używa się wersji bieżącej (najnowszej stabilnej).
6. Informacje o aktualnej wersji można pobrać przez zasób /version.
7. Przed planowaniem zmiany wersji należy oznaczać wersję bieżącą statusem „deprecated”.
8. Poza uzasadnionymi przypadkami nie należy wycofywać obsługi wersji API szybciej niż dwa lata po powiadomieniu użytkowników.
9. Po wyłączeniu obsługi starej wersji API odwołania do niej należy obsłużyć, w zależności od kontekstu, błędem HTTP o kodzie nr 406 lub 410.

10. Obsługa zdarzeń HTTP

API będące tzw. RESTful API opiera się na HyperText Transfer Protocol (HTTP), włączając w to wybrane metody, kody i nagłówki odpowiedzi. API korzysta głównie z metody GET. W API zakłada się niewykorzystywanie metod umożliwiających modyfikację zasobów. HTTP dostarcza pięć kategorii oraz

¹⁵⁰⁾ Zob. <https://semver.org/spec/v2.0.0.html>

ponad 30 kodów odpowiedzi. Jednak nie wszystkie kody będą w API przydatne, dlatego API przewiduje obsługę wybranych kodów.

Tabelę grup kodów zgodnie z RFC 7231 oraz przykłady odpowiedzi i komunikatów znajdują się w załączniku nr 8 do „Standardu API”.

Rekomendacje dotyczące kodów odpowiedzi w ramach obsługi zdarzeń HTTP

W kontekście obsługi zdarzeń HTTP – kody odpowiedzi, formułuje się następujące rekomendacje standaryzujące:

- 1) Określa się minimalny obsługiwany zestaw kodów odpowiedzi:

Tabela 10. Minimalny obsługiwany zestaw kodów odpowiedzi

Grupa kodów	Znaczenie	Informacja na temat błędu
2xx	200	OK
	201	Utworzono nowy zasób
	202	Zaakceptowano zapytanie, odpowiedź jeszcze nie jest gotowa
	204	Brak zawartości
3xx	301	Przeniesiono na stałe
	303	Zobacz inne: odpowiedź na zapytanie dostępna pod innym adresem
	304	Nie modyfikowano od czasu wskazanego w zapytaniu
4xx	400	Błędne żądanie
	401	Brak autoryzacji
	403	Dostęp zabroniony
	404	Nie znaleziono
	405	Nieodpowiednia metoda
	406	Nieakceptowalne – treść nie może zostać przygotowana w sposób zasygnalizowany nagłówkami
	409	Konflikt
	410	Zasób usunięty i już nigdy nie będzie dostępny pod tym adresem
	422	Zapytanie niepoprawne semantycznie
	429	Zbyt wiele zapytań w jednostce czasu
5xx	500	Wewnętrzny błąd serwera
	502	Błąd bramki – serwer otrzymał niepoprawną odpowiedź od serwera nadrzędnego
	503	Usługa czasowo niedostępna
	504	Przekroczony czas oczekiwania na odpowiedź serwera wewnętrznego

Do przesłania informacji o numerze błędu, wykorzystuje się element Status-Line w nagłówku HTTP.

- 2) W przypadku kodu błędów opis uzupełnia się informacją na temat błędów, która powinna być zawarta w ciele odpowiedzi.

Kody z kategorii 4xx i 5xx informują o błędach. Użytkownik nie ma wpływu na błędy z kategorii 5xx, gdyż leżą one po stronie serwera. Przy uzyskaniu błędu z kategorii 4xx istnieje prawdopodobieństwo wystąpienia błędów charakterystycznych w określonych sytuacjach. Ranking tych błędów może być tworzony w oparciu o cykliczną analizę logów API. Ranking ten powinien służyć do aktualizacji informacji objaśniających błędy.

Rekomendacje dotyczące informacji o błędach w ramach obsługi zdarzeń HTTP

W kontekście obsługi zdarzeń HTTP – kody błędów grupy 4xx, formułuje się następujące rekomendacje standaryzujące:

Tabela 11. Informacje objaśniające w ciele odpowiedzi HTTP

Nazwa objaśnienia	Identyfikator objaśnienia
Wiadomość o błędzie	error-result
Wiadomość objaśniająca przyczynę błędu	error-reason
Wiadomość objaśniająca możliwe rozwiązanie (opcjonalnie, tam gdzie jest to możliwe)	error-solution
Kod błędu zdefiniowany przez dostawcę	error-code
Link do dokumentacji z objaśnieniem numeru błędu dla danej wersji interfejsu	error-help

Komunikaty muszą dostarczać konkretnej informacji zwrotnej. Niedopuszczalne są komunikaty, które nie pomagają w rozwiązaniu problemu przez użytkownika. Oznacza to, że komunikat powinien być:

- sformułowany precyzyjnie w sposób konkretny,
- możliwy do sprawdzenia w kontekście przyczynienia się do rozwiązania problemu,
- osiągalny, czyli informacje/zalecenia zawarte w komunikacie powinny być wykonalne przez użytkownika w kontekście informacji, którymi dysponuje,
- wartością dodaną w porównaniu do sytuacji, w której komunikat by nie występował.

11. Uwierzytelnienie i autoryzacja

Systemy implementujące API mogą korzystać z systemu uwierzytelniania użytkowników. Nie jest dopuszczalne zamknięcie API dla użytkowników niewierzytelnionych. Dozwolone jest wprowadzenie dla nich dodatkowych ograniczeń w korzystaniu z API.

Opisy technologii i rozwiązań dotyczących uwierzytelniania i autoryzacji znajdują się w załączniku nr 9 do „Standardu API”.

Rekomendacje dotyczące uwierzytelniania i autoryzacji

W kontekście uwierzytelniania i autoryzacji formułuje się następujące rekomendacje standaryzujące:

- Uwierzytelnianie i autoryzacja użytkowników w API mogą być oparte o framework OpenID Connect.
- Stosowanie OAuth 2.0. jest dopuszczone, ale w połączeniu z innymi rozwiązaniami wypełniającymi lukę autoryzacji.
- Obowiązkowe jest stosowanie szyfrowania SSL.
- Uwierzytelnianie użytkowników API nie jest obowiązkowe.

12. Określenie reguł wprowadzania limitów danych per IP i per API-Key

W przypadku zbyt dużego obciążenia systemów udostępniających dane publiczne dozwolone jest wprowadzenie szczególnych ograniczeń na konkretny adres IP lub dla konkretnego klucza API.

Rekomendacje dotyczące reguł wprowadzania limitów danych

W kontekście reguł określania limitów danych formułuje się następujące rekomendacje standaryzujące:

1. Interfejs API nie powinien dopuszczać do nadmiernego obciążania zasobów API uniemożliwiającego korzystanie z niego przez użytkowników, a w razie konieczności powinien dawać możliwość nałożenia limitów per-IP oraz per-API-Key.
2. Udostępniony webowy kreator zapytań może być osobno, dodatkowo zabezpieczony przed wykorzystaniem go do nadmiernego obciążania zasobów API.
3. Klucze należy konstruować zgodnie z zasadami konstrukcji tokenów Web JSON, z domyślnie ustawionym prawem dostępu tylko do odczytu dla wszystkich zasobów API.
4. W przypadku wprowadzenia ograniczeń dla API muszą zostać opracowane polityka korzystania z API oraz plany korzystania z API przez użytkowników, rozróżniające użytkowników niezarejestrowanych, użytkowników posiadających API-Key oraz użytkowników z białych list – jeżeli takie rozróżnienie zostało wprowadzone.
5. W przypadku wprowadzenia limitów musi zostać utworzona biała lista użytkowników korzystających z API z priorytetem wyższym niż inni użytkownicy – w szczególności dotyczy to systemów publicznych, dla których dane są danymi referencyjnymi. Reguły wpisu na białą listę powinny być jednoznacznie sformułowane.
6. Gdy nie narusza to stabilności systemu, dopuszczone jest obciążanie API w sposób wykraczający poza plany obciążeń, szczególnie przez użytkowników z białych list.
7. Należy używać reguł dławiących destrukcyjnych użytkowników, w przypadku zagrożenia zablokowaniem API (DOS attack). Zaleca się stosowanie polityki długoterminowego uznawania adresów IP za wrogie (czarne listy) i wykorzystywanie zewnętrznych czarnych list¹⁵¹⁾ do analizy zdarzeń i rozszerzania listy adresów blokowanych długoterminowo.
8. Transfer może być ograniczany dla użytkowników generujących transfer i obciążenie zagrażające ciągłości i stabilności API.
9. Transfer może być ograniczany dla użytkowników w oparciu o dynamicznie aktualizowane zasady wynikające z analiz faktycznego sposobu korzystania z API przez użytkowników,
10. Transfer może być ograniczany dla użytkowników z czarnej listy adresów IP.
11. Adresy z czarnej listy adresów IP powinny być weryfikowane co najmniej co 6 miesięcy i usuwane z niej w przypadku braku podejrzanego ruchu z danego adresu.
12. Ograniczenie transferu powinno być w każdym przypadku ostatecznością wynikającą ze stosowania określonych reguł, a nacisk należy położyć na podnoszenie wydajności API.
13. Niektóre funkcjonalności API mogą być ograniczone ze względu na możliwość generowania obciążeń dla rozwiązania IT API.
14. Nałożenie limitów powinno skutkować błędem HTTP o kodzie 429 z informacją o polityce limitów oraz możliwości dodania do białej listy.

13. Opis standardu dokumentacji usług API

Rekomendacje dotyczące standardów dokumentacji usług API

W kontekście standardów dokumentacji usług API formułuje się następujące rekomendacje standaryzujące:

¹⁵¹⁾ Np. <http://apility.io>.

1. Dokumentacja API musi zawierać co najmniej:
 1. opis zasobów – opis informacji dostarczanej przez każdy zasób, znaczenia zasobu, punktów dostępowych, metod do każdego punktu dostępowego,
 2. opis obsługiwanych metod i punktów dostępowych, sposobu dostępu do zasobów, dopuszczonych metod interakcji z zasobami,
 3. parametry obsługiwane przez punkt dostępowy, w szczególności dotyczące:
 - a) metody HTTP (GET, POST, PATCH itp.),
 - b) nagłówków,
 - c) ścieżki dostępu (URI),
 - d) parametrów żądania (query),
 - e) ciała żądania (body), np. JSON,
 4. przykłady kodu żądań obrazujące użycie złożonych, ale nieskomplikowanych zapytań,
 5. przykłady odpowiedzi: struktur i treści przesyłanych klientowi w odpowiedzi na żądania z powyższego punktu – w formie zapisu w dokumentacji lub odwołania do rzeczywistego systemu.
2. Zaleca się, aby dokumentacja API:
 - ✓ stosowała oznaczenia zapewniające możliwość odczytu maszynowego,
 - ✓ była dynamiczna i interaktywna,
 - ✓ automatycznie generowała przykłady kodu dla bibliotek klienta.
3. W zakresie realizacji wytycznych dotyczących dokumentacji oraz uzupełniająco zaleca się stosowanie standardu OpenAPI w obowiązującej obecnie wersji 3.0. lub nowszej.

Na rynku dostępne są narzędzia do generowania dynamicznej i interaktywnej dokumentacji API. Umożliwiają one budowanie kreatorów zapytań do API, a niektóre dedykowane są do współpracy z API typu RESTful. Do tworzenia dokumentacji zaleca się rozwiązania wspomagające implementację standardu OpenAPI z rodziny rozwiązań Swagger, w szczególności Swagger Editor, oraz Swagger UI.

14. Odstępstwa od standardu JSON API

Poniższe odstępstwa są umotywowane koniecznością odciążenia serwerów danych publicznych przez zmniejszenie potrzeby kopiowania całej bazy w celu uzyskania konkretnych odpowiedzi lub ułatwienie korzystania z danych.

Rekomendacje dotyczące odstępstw od JSON API:

1. API musi wspierać wybór pól do dostarczenia klientowi (specyfikacja JSON API mówi, że jest to opcjonalne – sekcja Sparse Fieldsets).
2. API musi wspierać sortowanie pól tam, gdzie jest to logicznie uzasadnione rodzajem danych (specyfikacja JSON API mówi, że jest to opcjonalne – sekcja Sorting).
3. API musi wspierać filtrowanie (specyfikacja JSON API mówi, że jest to wysoce zalecane – sekcja Filtering).
4. API musi wspierać dokładność informacji o błędach (specyfikacja JSON API mówi, że jest to opcjonalne – sekcja Error Objects).

Załącznik nr 1 do „Standardu API”

Kategorie obecnych i potencjalnych odbiorców danych wykorzystujących API – szczegółowa charakterystyka grup użytkowników

Odbiorcami korzystającymi z danych publicznych (obecnie i docelowo) są osoby fizyczne, przedsiębiorstwa, jednostki administracji publicznej (centralnej i samorządu terytorialnego), organizacje pozarządowe, podmioty nauki, kultury i inni interesariusze. Dane publiczne mogą opisywać rzeczywistość w dowolnym aspekcie, a interesariusze mogą przetwarzać te dane na wiele sposobów. Kształtuje to metody korzystania z API oraz wpływa na wymagania przed nim stawiane. Dlatego na tym etapie ważne jest dokonanie podziału odbiorców danych na grupy wyodrębnione ze względu na charakterystykę ich potrzeb i sposobów wykorzystywania API, zamiast grupowania według wyżej wymienionych cech podmiotowych.

Większość użytkowników API będzie korzystała z produktów pochodnych opracowanych na podstawie danych pochodzących z API (usług, aplikacji komputerowych, informacji przetworzonych, know-how) często nieświadomie. Jest to tzw. grupa użytkowników pośrednich, która nie korzysta wprost z API, ale czerpie z niego korzyści i może mieć wpływ na pożądany kierunek jego ewolucji. Pozostałych użytkowników można podzielić na dwie grupy:

- użytkownicy okazjonalni (*ad hoc*),
- użytkownicy profesjonalni.

Grupa użytkowników okazjonalnych (*ad hoc*)

Stanowią ją użytkownicy, którzy będą korzystała z API doraźnie dla rozwiązania lub poznania zagadnień związanych np. z zadaniami w pracy, w szkole, na uczelni, często z ciekawości. Do grupy tej będą należeć zarówno osoby fizyczne (obywatele), jak i przedsiębiorcy prowadzący działalność gospodarczą w różnych jej formach prawnych i organizacyjnych (spółki prywatne i publiczne, organizacje wirtualne, JST, start-upy), organizacje pozarządowe, jednostki naukowe, instytucje obszaru kultury oraz pozostałe jednostki administracji publicznej. Użytkownicy z tej grupy to osoby w znacznej mierze bez zaplecza technicznego, posiadające podstawowe kompetencje cyfrowe, takie jak: obsługa systemu plików, tworzenie dokumentów cyfrowych, obsługa skrzynki e-mail przez stronę internetową, działania w mediach społecznościowych oraz interaktywne przeglądanie stron internetowych.

Do opisywanej grupy należą także innowatorzy lub wcześnie naśladowcy. To oni w pierwszej kolejności zidentyfikują potrzebę korzystania z danych udostępnionych przez API i weryfikacji, na ile te dostępne zasoby będą przydatne do szybkiego rozwiązywania konkretnych zadań. Użytkownicy ci nie będą skłonni do przeznaczania ani znacznych nakładów, ani zasobów (w szczególności czasu), aby zgłębiać i eksperymentować z API. Dlatego po wejściu w interakcję z API powinni szybko uzyskać pozytywne efekty, które będą stanowiły wzmocnienie i zachętę do dalszego eksplorowania zasobów API. Jest to istotna cecha konieczna do uwzględnienia przy tworzeniu zarówno API, jak i jego dokumentacji. Dokumentacja API powinna na prostych przykładach obrazować przydatność danych publicznych w rozwiązywaniu codziennych zadań. Czynności techniczne, w szczególności związane z zachowaniem poprawności składni, które mogą użytkownikom nastręczać trudności, powinny przyjąć formę gotowych elementów – pasujących do siebie bloków funkcjonalnych.

Mimo że użytkownicy *ad hoc* stanowią grupę użytkowników nieprofesjonalnych, to jednak ta grupa jest bardzo ważna, gdyż zakłada się, że to właśnie ona będzie inicjować, a nierzadko decydować o zaangażowaniu API w instytucjach. Mając na uwadze, że świadomość potencjału oraz możliwości korzystania z danych publicznych jest jeszcze stosunkowo na niskim poziomie¹⁵²⁾ (stosownie do wczesnej fazy rozwoju tego rodzaju źródeł informacji), należy położyć odpowiedni nacisk na możliwość uzyskania szybkiej satysfakcji przez użytkowników tej grupy podczas korzystania z API. Ten aspekt może być decydujący dla powodzenia i popularności przygotowanego interfejsu.

Grupa użytkowników zaawansowanych

¹⁵²⁾ Jest to jedna z głównych barier przy korzystaniu z danych publicznych. Zob. M. Możejewski, „Strategia, standaryzacja i pomiar jakości informacji – determinanty efektywnego ponownego wykorzystania informacji sektora publicznego”, [w:] A. Gołębiowska, P.B. Zientarski (red.), „Ponowne wykorzystanie informacji sektora publicznego w administracji”, Senat Rzeczypospolitej Polskiej, Warszawa 2017, s. 65–90.

Druga grupa bezpośrednich użytkowników API. W odróżnieniu od użytkowników okazjonalnych, dysponuje ona wiedzą techniczną i ma już sformułowane bezpośrednie cele i korzyści korzystania z API, gdyż podjęła decyzję o chęci korzystania z niego. To oznacza, że użytkownicy ci są zdeterminowani do integracji usług API z tworzonymi nowymi lub modyfikowanymi systemami teleinformatycznymi. Użytkowników tych cechuje potrzeba korzystania z danych w sposób przewidywalny. Oczekują dostępności, odpowiedniej jakości danych oraz wydajności mechanizmu API, które zaspokoją każde bieżące potrzeby (których często sami nie są w stanie określić z wyprzedzeniem). Zatem w grupie użytkowników zaawansowanych znajdują się także użytkownicy okazjonalni, którzy wypróbowali już API i określili swoje oczekiwania i cele w stosunku do API. Posiadają przy tym wiedzę techniczną, zaplecze techniczne oraz zasoby, które mają być wzbogacane w oparciu o dane zasilane z API. W odróżnieniu od użytkowników okazjonalnych użytkownicy profesjonalni posiadają też wiedzę na temat zaawansowanych narzędzi i form komunikacji z API. Użytkownikami profesjonalnymi w głównej mierze są deweloperzy. Są oni użytkownikami, którzy mają mało czasu, podchodzą do problemów zadaniowo i chcą szybko uzyskiwać satysfakcjonujące efekty.

Załącznik nr 2 do „Standardu API”

Architektura usług sieciowych REST

Standard API jest oparty na standardzie REST. Ponadto referencyjnie wykorzystuje format reprezentacji danych zdefiniowany przez standard JavaScript Object Notation, czyli JSON. Jako standard notacji JSON przyjmuje się JSON Schema, a jako standard komunikatów (zapytania i odpowiedzi między użytkownikiem a systemem) przyjmuje się JSON API. Kontekst ułatwiający korzystanie z danych w sposób automatyczny jest generowany zgodnie ze standardem JSON-LD.

REST – (ang. Representational State Transfer) jest popularnym stylem architektonicznym stosowanym w celu budowy usług wymiany danych w sieci Internet, takich jak API. API zbudowane na zasadach zdefiniowanych przez REST, zwane jest RESTful API. REST wykorzystuje protokół HTTP z uwzględnieniem jego zalet i ograniczeń. Komunikaty REST wykorzystują statusy odpowiedzi HTTP, takie jak 2xx (żądanie obsłużone poprawnie), 4xx (komunikaty o błędzie po stronie klienta) czy 5xx (błędy po stronie serwera).

REST jest stylem zaproponowanym w doktoracie Roya Fieldinga¹⁵³⁾ z 2000 r. Stanowi on odwzorowanie idei efektywnej dystrybucji treści w rozproszonym systemie hipertekstowym. Idea ta uwzględnia jednak istotne ograniczenia, które stanowią jednocześnie kluczowe wytyczne dla REST:

1. Zasoby (ang. resource)

W REST kluczowym elementem jest zasób. Zasób to każda porcja informacji, której postać może być określona. Przykładami zasobów są dokumenty, obrazy, informacje aktualne w określonym czasie, pogrupowane zasoby czy obiekty, pozwalające się w uproszczony sposób scharakteryzować. Zasób jest odwzorowaniem uproszczonych cech obiektu. Cechy te muszą być spełnione (odwzorowane) zgodnie z definicją obiektu. Tylko obiekt, który odpowiada definicji odwzorowania, może być zasobem. Oznacza to, że porcja informacji staje się zasobem, gdy określone są wszystkie jej cechy wskazane w jego definicji. Zależność tę można zdefiniować za pomocą funkcji przynależności na podstawie warunku logicznego, który przyporządkowuje zasobowi w określonym czasie unikalny identyfikator lub określone wartości. Zasób może być określony w postaci nieziennej (gdy przyjmuje wartości statyczne), zmiennej (gdy przyjmuje wartości dynamiczne w czasie), ale może też być pusty (brak przyjętych wartości). Jedyny element, który jest wymagany i stały w określeniu zasobu, to semantyka mapowania, tj. struktura jego elementów znaczeniowych. Nawet w przypadku zmienionych, braku lub tożsamy wartości pozwala ona na odróżnienie jednego zasobu od drugiego.

2. Komunikacja w modelu klient – serwer

¹⁵³⁾ R. Fielding „Architectural Styles and the Design of Network-based Software Architectures”, University of California, Irvine 2010.

W wymianie informacji opartej na REST istnieje wyraźna granica między klientem i serwerem. Domeną klienta jest interfejs użytkownika i generowanie zapytań. Domeną serwera jest obsługa żądań klientów, w szczególności zarządzanie przechowywaniem i dostępem do danych, obciążeniem i bezpieczeństwem. Elementy te mogą być rozwijane po stronie klienta i serwera niezależnie od siebie tak długo, jak długo komunikacja między stronami odbywa się przez wspólnie jednoznacznie rozumiany protokół.

3. Samoopisujące się komunikaty

Każdy komunikat wyposażony jest w odpowiednie informacje, które pozwalają na zarządzanie tym zasobem. Oznacza to, że komunikat oprócz treści zawiera także metadane pozwalające umieścić ją w kontekście i prawidłowo zinterpretować.

4. Bezstanowość (ang. stateless)

Oznacza rozpatrywanie zapytań niezależnie od innych (tego samego lub innego użytkownika). Cecha ta związana z wymianą informacji w modelu klient – serwer powoduje, że nie są akumulowane informacje o poprzednich transakcjach klienta, co pozwala na odciążenie serwera, jednak wymaga dodatkowych mechanizmów, jeżeli podczas wymiany informacji potrzebne są dodatkowe informacje na temat użytkownika lub jego żądań¹⁵⁴). Wszystkie operacje klient – serwer powinny być bezstanowe, a zarządzanie stanem powinno odbywać się po stronie klienta.

5. Buforowanie reprezentacji

W kontekście REST odpowiedzi serwera mogą być po stronie klienta przechowywane w buforze (cache), tak aby mogły być użyte podczas realizacji kolejnych zapytań bez konieczności pobierania ich pełnych wyników. Jednak, aby uniknąć odczytu przez klienta nieodpowiednich lub nieaktualnych danych, pobierane treści powinny być oznaczane jako możliwe do buforowania, ze wskazaniem, na jakich warunkach (głównie czasu ważności treści). Zaleca się, aby wszystkie zasoby mogły być buforowane, chyba że taka możliwość zostanie wprost wyłączona. Prawidłowe określenie tych parametrów pozwala eliminować kolejne interakcje klienta i serwera. Buforowanie wpływa zatem pozytywnie na szybkość dostarczania treści do odbiorcy, pozwala korzystać z treści, nawet gdy serwer jest chwilowo niedostępny, oraz zmniejsza obciążenie serwera. Oznaczenie zasobów informacyjnych¹⁵⁵) powinno być określane przez ich dostawców na podstawie odrębnej polityki określającej wytyczne dla warunków buforowania.

6. Zastosowanie jednolitego interfejsu i unikalna identyfikacja zasobów

Zasoby w systemach REST powinny być jednoznacznie oznaczane i identyfikowane za pomocą adresu URL. W przypadku standardu API działania wykonywane na tych zasobach (np. PUT, GET) powinny być wykonywane za pomocą poleceń protokołu HTTP z włączonym szyfrowaniem (HTTPS).

7. Enkapsulacja systemów warstwami (ang. Layered System)

W uproszeniu przesłonięcie systemów zewnętrznymi powłokami, które mają na celu wystandaryzowaną wymianę informacji. Zaletą takiego podejścia jest to, że każdy system pozostaje czarną skrzynką, a o jego przydatności decyduje tylko oddzielający go interfejs. Szczegóły wewnętrzne systemu serwera pozostają poza percepcją klientów.

8. Kod na żądanie (ang. Code-On-Demand)

Opcjonalnie – serwery są zdolne do rozszerzania lub dostosowania funkcjonalności, z jakiej korzysta klient, przez przesyłanie mu logiki, która może być wykonywana. Przykładami takiego działania może być przesyłanie klientom apletów Java lub kodu JavaScript, który wykonuje się u klienta. Jest to cecha, która w połączeniu z pozostałymi prowadzi do zwiększania skalowalności, adaptowalności, przenośności i modyfikowalności rozwiązań budowanych w oparciu o REST.

¹⁵⁴) Takie informacje mogą być dostarczane przez mechanizm ciasteczek, sesji lub parametrów w nagłówku HTTP.

¹⁵⁵) Zasoby informacyjne – informacje publiczne, są udostępniane w centralnym repozytorium, o szczególnym znaczeniu dla rozwoju innowacyjności w państwie i rozwoju społeczeństwa informacyjnego, które ze względu na sposób przechowywania i udostępniania pozwalają na ich ponowne wykorzystywanie w rozumieniu przepisów ustawy z dnia 25 lutego o ponownym wykorzystywaniu informacji sektora publicznego, w sposób użyteczny i efektywny.

Załącznik nr 3 do „Standardu API”

Opis składni oraz elementów URI – definicje, szczegółowy opis składni oraz przykłady zastosowań

Definicje związane z oznaczeniem zasobów:

Schemat (ang. authority) – komponent URI, który definiuje protokół wymiany informacji (np. HTTP) oraz przestrzeń

Zasób główny (ang. docroot) – zasób, który jest nadrzędny wobec pozostałych w API w modelu REST

Dokument (ang. document) – reprezentacja pojedynczego wystąpienia zasobu

Kolekcja (ang. collection) – grupa zasobów po stronie serwera

Magazyn (ang. store) – repozytorium zarządzane przez klienta

Kontroler (ang. controller) – opis logiki proceduralnej

CRUD – akronim od nazw czterech podstawowych funkcji operowania na danych w bazach danych: create (utwórz), retrieve (odczytaj), update (aktualizuj) i delete (usuń)

„Bezpieczne metody” HTTP – metody GET i HEAD, które umożliwiają uzyskiwanie reprezentacji zasobów, ale nie pozwalają na modyfikację i usuwanie zasobów na serwerze

Zapytanie (ang. query) – fragment URI występujący za ścieżką, ale przed opcjonalnym fragmentem

URI (z ang. Uniform Resource Identifier) – unikalny identyfikator zasobów, stosowany zgodnie z RFC 1736¹⁵⁶⁾, RFC 1737¹⁵⁷⁾ oraz ustandaryzowany zgodnie z RFC 3986¹⁵⁸⁾. Podstawowe oznaczenie URI przyjmuje następującą postać:

```
URI = <schemat> „:” <część hierarchiczna> [ „?” <pytanie> ] [ „#”  
      <fragment> ]
```

Ogólna postać URI zakłada występowanie schematu (np. protokołu), określonej hierarchii (np. domeny i/lub odwołania lokalnego), zapytań oraz wskazanie wybranych fragmentów. Część hierarchiczna może zawierać odwołanie do domeny podmiotu i/lub ścieżki lokalnej. Elementy **<pytanie>** oraz **<fragment>** są składowymi, które nie muszą być podane. Element **<pytanie>** zwany jest też kwerendą.

1) Domena w URI powinna wskazywać jednoznacznie na właściciela API. Domena ta powinna udostępniać funkcjonalność API przez subdomenę „api”. Odstępstwa od tej reguły są dopuszczalne, w sytuacji gdy poddomena API jest już zajęta do innych celów.

Przykład:

```
https://api.dane.gov.pl
```

Powyższe zalecenie wynika z zastosowania tzw. dobrej praktyki przy udostępnianiu danych przez API. Mimo że popularne są także odmienne podejścia (np. udostępnianie zasobów API w ścieżce), to wydzielenie zasobów w zaproponowany sposób jest szczególnie stosowane przez liczne światowe API udostępniające informacje sektora publicznego¹⁵⁹⁾. Takie rozwiązanie poprawia również czytelność i rozpoznawalność zasobów API. Dlatego warto tę konwencję utrzymać. Z drugiej strony API dostępne w wydzielonej subdomenie pozwala na zwiększenie bezpieczeństwa dzięki możliwości izolacji ustawień (per domena) oraz ułatwia zarządzanie przyszłą migracją. Takie

¹⁵⁶⁾ RFC 1736, <https://www.ietf.org/rfc/rfc1736.txt>.

¹⁵⁷⁾ RFC 1737, <https://www.ietf.org/rfc/rfc1737.txt>.

¹⁵⁸⁾ RFC 3986, <https://www.ietf.org/rfc/rfc3986.txt>.

¹⁵⁹⁾ Światowe zasoby danych dostępne poprzez API w subdomenie: api.data.gov, api.nasa.gov, data.norge.no, api.trafikinfo.trafikverket.se.

zalecenie wynika również z literatury przedmiotu, która określa, że stosując wytyczne REST dla API, powinno udostępniać się API w subdomenie¹⁶⁰⁾.

W uzasadnionych przypadkach dopuszczalne jest udostępnianie API w głównej domenie identyfikującej podmiot zarządzający API, pod adresem /api.

2) API może zostać wyposażone w portal dla deweloperów, który zawiera podstawową dokumentację API – lub odnośnik do niej – oraz informację o ewentualnych limitach nałożonych na API i możliwości uzyskania klucza aplikacji (API-Key). Portal powinien być integralną częścią API i być dostępny przez subdomenę „dev”.

W przypadku braku oddzielnego serwisu dla deweloperów strona w domenie API musi zawierać informacje o limitach w wykorzystaniu API oraz o możliwości uzyskania klucza aplikacji.

Wytyczne, podobnie jak w przypadku punktu pierwszego, wynikają z zastosowania przyjętej dobrej praktyki do zapewniania wsparcia deweloperom w jednym wydzielonym miejscu.

3) Znak „/” (ukośnik, ang. forward slash) służy oznaczeniu zależności hierarchicznej między zasobami. Stosowanie znaku „/” na końcu URI nie jest zalecane.

4) Oznaczanie ścieżki URI należy oznaczać z użyciem małych liter. URI zapisane wielkimi i małymi literami w powyższym przykładzie w praktyce są tożsame, gdyż zgodnie z RFC 3986 nie jest rozróżniana wielkość liter w elemencie reprezentowanym przez domenę i subdomenę. Ponieważ wytyczne w stosunku do nazewnictwa zasobów wskazują konieczność używania małych liter, dla zachowania jednoznaczności także dla URI formułuje się taką zasadę.

5) W URI nie stosuje się rozszerzeń plików oraz nazw formatów danych. Informacje te są przekazywane przez nagłówki HTTP lub w parametrach. Używanie w URI nazwy użytkownika z hasłem może stanowić poważną lukę w bezpieczeństwie i z tego powodu nie jest dozwolone. Informacja o wykorzystanym kluczu aplikacji może zostać przekazana jako parametr żądania lub w nagłówkach HTTP.

Załącznik nr 4 do „Standardu API”

Manipulowanie zasobami: filtrowanie, sortowanie, stronicowanie – zasady wraz z przykładami

Z założenia szereg klienckich aplikacji będzie korzystać z zasobów API automatycznie. Takie aplikacje rzadko realizują zadania wymagające pobrania z API zbioru danych w pełnym przekroju informacyjnym oraz pełnej zawartości rekordów. Znacznie częściej potrzebują odpowiednio spreparowanego, niewielkiego wycinka zamiast pełnego zakresu dostępnych danych. Wycinek ten w stosunku do oryginalnego zbioru może być filtrowany względem zakresu pobieranych danych, w kontekście określonych kryteriów, porcji danych pobieranej za jednym razem oraz w odpowiednim uporządkowaniu. Dlatego założono, że API udostępnia funkcjonalność filtrowania, sortowania i stronicowania zasobów, która będzie określana mianem manipulacji danymi.

Zapytania i odpowiedzi systemu powinny być realizowane zgodnie ze standardem JSON API w wersji 1.0 lub nowszej. W odpowiedziach zaleca się uzupełnienie JSON API o JSON-LD w celu dodania kontekstu koniecznego do automatycznego przetwarzania danych API.

Interakcja z API w zakresie filtrowania, sortowania i stronicowania wybranych zasobów oparta jest na mechanizmie URI, który zawiera sekcję **<pytanie>**. Jest to element nieobowiązkowy w URI, lecz niezbędny w przypadku potrzeby manipulacji zakresem lub strukturą danych przed pobraniem

¹⁶⁰⁾ Zob. M. Masse, „REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces”, O’Reilly Media, Sebastopol 2012, str. 14.

Podstawowe zasady manipulacji danymi przedstawiono za pomocą kolejnych przykładów.

Punktem odniesienia do wykonywania operacji manipulowania danymi jest pełna lista zasobów, czyli

```
URI = <schemat> „:” <część hierarchiczna> [ „?” <pytanie> ] [ „#”  
      <fragment> ]
```

tzw. kolekcja, uzyskiwana za pomocą następującego odwołania do zasobu głównego:

```
GET /
```

W wyniku uzyskiwana jest odpowiedź, która jest reprezentacją stanu zawierającą listę wszystkich zasobów.

Natomiast po wykonaniu kodu z przykładu:

```
GET /pojazdy
```

uzyskiwana jest reprezentacja stanu zawierająca listę elementów zasobu „pojazdy”.

Filtrowanie zasobów służy do ograniczania wyników jedynie do tych, które spełniają zadane warunki. W rezultacie kolekcję ogranicza się jedynie do listy zawierającej zasoby odpowiadające kryteriom przekazanym w URI w części **<pytanie>**.

```
GET/filter[<nazwa atrybutu>]=<szukana wartość>
```

```
GET/pojazdy?województwo=14&data-od=20160101&data-  
do=20181201&filter[model]=JUMPER
```

W wyniku uzyskiwana jest reprezentacja stanu zawierająca pojazdy, które dla wymiaru „model” spełniają warunek „JUMPER”, z województwa mazowieckiego, zarejestrowane po raz pierwszy w Polsce między 2016-01-01 a 2018-12-01.

W przypadku filtrowania po więcej niż jednym kryterium do rozdzielania warunków korzysta się z operatora „&”:

```
GET/pojazdy?województwo=02&typ-daty=2&data-od=20190901&filter[  
marka]=OPEL&filter[rodzaj-paliwa]=BENZYNA
```

W wyniku można ograniczyć liczbę zwracanych pól jedynie do wybranych, np. marka, model, rodzaj paliwa:

```
GET/pojazdy?województwo=02&typ-daty=2&data-od=20190901&fields=  
marka,model,rodzaj-paliwa
```

Sortowanie zasobów służy do szeregowania elementów rosnąco lub malejąco według określonego kryterium lub określonych kryteriów. Domyślnie zasoby powinny być sortowane rosnąco według zadanego kryterium. Kolejność podanych kryteriów jest określona przez kolejność sortowania. Aby posortować w odwrotnej kolejności należy poprzedzić nazwę parametru znakiem minus.

```
GET/pojazdy?województwo=02&data-od=20190901&pokaz-wszystkie-  
pola=true&sort=-data-pierwszej-rejestracji-w-kraju,marka
```

Wynik: sortowanie za pomocą więcej niż jednego kryterium.

Jeżeli serwer nie obsługuje danego rodzaju sortowania dla zadanego parametru, serwer powinien zwracać błąd. Jeżeli parametr „sort” nie został wskazany, serwer używa domyślnego sortowania dla zwracanych elementów.

Stronicowanie zasobów służy do ograniczania wyników zapytania do określonej liczby elementów zwanych stroną. Stronicowanie oparte jest na dwóch parametrach:

page – oznacza numer strony,

limit – oznacza liczbę elementów wyświetlanych na jednej stronie.

```
GET/słowniki?limit=100&page=1
```

Oznacza wyświetlenie 100 elementów na pierwszej stronie.

Podsumowując powyższe, kod obrazujący przykład najbardziej złożonego ze wszystkich powyższych warunków będzie miał następującą postać:

```
GET/pojazdy?województwo=02&data-od=20190901&data-  
do=20191117&limit=100&page=478&typ-daty=2&tylko-  
zarejestrowane=true&pokaz- wszystkie-  
pola=true&fields=marka,model,rodzaj-paliwa
```

W przypadku gdy przygotowanie odpowiedzi zajmie więcej czasu, serwer powinien odpowiedzieć komunikatem HTTP o numerze 202, ze wskazaniem zasobu informującego o szacunkowym czasie pozostałym do końca realizacji żądania.

```
HTTP /1.1 202 Accepted  
Location: /queue/[id_żądania]
```

Do momentu przygotowania danych, pod adresem /queue/[id_żądania] znajduje się informacja o przygotowywaniu odpowiedzi, wysyłana ze statusem o numerze 200. W momencie gdy zasób jest gotowy do przekazania odbiorcy, pod tym adresem znajdują się odpowiedź o numerze 303 i lokalizacja odpowiedzi na pierwotne zapytanie. Po upływie okresu, po którym przygotowane w ten sposób dane są usuwane z serwera, zapytanie o lokalizację docelową zwraca kod błędu 410.

Załącznik nr 5 do „Standardu API”

Metadane

Zasoby udostępniane przez API są wyposażane w metadane, stanowiące część odpowiedzi na zapytanie o reprezentację zasobu. Metadane to inaczej dane o danych, które, opisując zasoby, ułatwiają ich zrozumienie oraz zarządzanie nimi. Są szczególnie ważne w procesie maszynowego przetwarzania informacji, dlatego istotne jest wykorzystanie rodzajów metadanych tam, gdzie tylko jest to możliwe. W przypadku standardu API, metadane muszą uwzględniać lokalne wytyczne określone przez POD oraz rozporządzenie Rady Ministrów z dnia 12 marca 2014 r. w sprawie Centralnego Repozytorium Informacji Publicznej (Dz. U. poz. 361, z późn. zm.). Dlatego w poniższej tabeli przedstawiono minimalny zestaw metadanych dla zasobu z założeniem, że może zostać rozbudowany przez twórcę konkretnej implementacji API, gdy będzie to konieczne. Metadane niewykorzystane w konkretnym przypadku (np. schema:expires) są pomijane.

Tabela 12. Minimalny zestaw metadanych, który może być rozszerzony

Nazwa	Oznaczenie	Znaczenie
Nazwa zasobu	dc:title	Nazwa zasobu informacyjnego, która umożliwia jego identyfikację (ew. nazwa zasobu z elementami charakterystycznymi) http://dublincore.org/documents/dcmi-terms/#terms-title
URL zasobu	schema:url	Identyfikator lokalizacji do pobrania zasobu informacyjnego, w przypadku gdy stanowi osobny plik (URL) http://schema.org/url
Identyfikator dostawcy zasobu	schema:provider	Opis dostawcy zasobu informacyjnego http://schema.org/provider
Data publikacji zasobu	schema:datePublished	Data i czas publikacji zasobu, z dokładnością do jednej sekundy Data i czas zgodnie ze standardem ISO-8601 https://schema.org/datePublished
Data modyfikacji zasobu	schema:dateModified	Data ostatniej modyfikacji zasobu z dokładnością do jednej sekundy Data i czas zgodnie ze standardem ISO-8601 https://schema.org/dateModified
Data zakończenia publikacji zasobu	schema:expires	Data określająca, do kiedy zasób jest dostępny w API – jeżeli zasób będzie dostępny przez określony czas https://schema.org/expires
Opis zasobu	dc:description	Opis zasobu informacyjnego http://dublincore.org/documents/dcmi-terms/#terms-description
Słowa kluczowe	schema:keywords	Słowa kluczowe, czyli zestaw słów lub wyrażeń zwięźle opisujących zasób informacyjny https://schema.org/keywords
Zakres czasowy danych	schema:temporalCoverage	Informuje, że zbiór zawiera dane za konkretne okresy https://schema.org/temporalCoverage
Język zasobu	dc:language	Określa język zasobu http://dublincore.org/documents/dcmi-terms/#terms-language

Nazwa	Oznaczenie	Znaczenie
Częstotliwość aktualizacji	sy:updatePeriod	Informuje o tym, jaki jest interwał aktualizacji zasobu, np. co miesiąc, co rok, co godzinę, co 10 sekund
	sy:updateFrequency	Zdefiniowana w purl.org: http://web.resource.org/rss/1.0/modules/syndication/
	sy:updateBase	Nie występuje w przypadku danych stałych lub o nieprzewidywalnym czasie aktualności
Kodowanie	schema:contentType	Określa informację o rodzaju zasobu i jego kodowaniu https://schema.org/contentType
Licencja	dc:rights	Określenie rodzaju licencji w przypadku wprowadzenia ograniczeń w wykorzystaniu zasobu http://dublincore.org/documents/dcmi-terms/#terms-rights
Katalog	schema:isPartOf	Określenie przynależności do grupy np. zasobów https://schema.org/isPartOf
Suma kontrolna	sp:checksum	Wartość, która pozwala na sprawdzenie zawartości pliku https://schema.parkeon.com/checksum

Powyższa tabela zawiera minimalny zestaw metadanych, który może być rozszerzany. Ponadto szczególną rolę odgrywa właściwie dobrany i bezbłędny zestaw metadanych określany przez dostawcę, gdyż są to informacje najczęściej wykorzystywane do precyzyjnego manipulowania danymi w ramach zasobów.

Załącznik nr 6 do „Standard API”

Mechanizm linków do nawigacji – objaśnienia i przykłady stosowania

W odpowiedzi na zapytania przesyłane przez użytkowników API udostępnia stany zawierające odpowiednie zasoby informacyjne. Jednak nie wszyscy użytkownicy potrzebują w odpowiedzi tego, aby otrzymywać całe zasoby. Większość aplikacji użytkowników korzystających z API potrzebuje w danym momencie często jedynie niewielkiej porcji informacji odfiltrowanej z zasobu. Takie informacje są filtrowane za pomocą dodatkowych warunków zapytania zawartych w URI. Jednak przesłanie użytkownikowi samego stanu zawierającego odfiltrowany zasób nie jest wystarczające. Dlatego standard API zapewnia funkcjonalność linków do nawigacji. Polega ona na dołączaniu w ciele zasobu poniższych linków (oznaczanych jako „links”), zgodnie z JSON API 1.0 lub późniejszym:

Tabela 13.

Nazwa linku	Objaśnienie
prev	Link do poprzedniego powiązanego zasobu
self	Link do bieżącego zasobu

next	Link do następnego powiązanego zasobu
last	Link do ostatniego powiązanego zasobu
first	Link do pierwszego powiązanego zasobu

Załącznik nr 7 do „Standardu API”

Wersjonowanie – zasady, metody oznaczania wersji wraz z przykładami oraz techniki zarządzania wersjami

Wersjonowanie jest jedną z najistotniejszych kwestii do rozważenia przy tworzeniu API. Od tego, jak wersjonowanie zostanie zaplanowane i zaimplementowane, zależy w przyszłości nie tylko wygodą, ale i możliwość korzystania z API oraz dostępność jego zasobów. Wiele organizacji udostępniających API nie udoskonala go później. Przyczyną jest często nie tyle brak pomysłów na zmiany, ile niemożność ich dokonania, gdyż w API nie przewidziano i nie uwzględniono jego ewolucyjnego charakteru.

Wśród twórców API przeważają następujące podejścia do wersjonowania API:

- w adresie (np. [https://api.dane.gov.pl/api/v1/\[...\]](https://api.dane.gov.pl/api/v1/[...]))
- w parametrze (np. <https://api.dane.gov.pl/resources/17323,pojazdy?v=2>)
- w nagłówku Accept zapytania HTTP (Accept: application/danepubliczne.v1+json)
- przez dodatkowy nagłówek zapytania HTTP (api-version: 1)

Wersjonowanie w adresie (wariant pierwszy) w standardzie API jest obowiązkowe. Ze względu na przydatne cechy pozostałych rozwiązań i jednocześnie względną prostotę zaimplementowania wszystkich jednocześnie dopuszczone jest umożliwienie korzystania dodatkowo z innych wariantów przy formułowaniu zapytania przez klienta.

Przykład czterech równoważnych zapytań:

- GET /v2/cepik/pojazdy
- GET /cepik/pojazdy?v=2
- GET /cepik/pojazdy
Accept: application/vnd.kowalski.user+json; version=2
- GET cepik/pojazdy api-version: 2

W przypadku braku wskazania wersji API przez klienta, system powinien przyjmować wykorzystanie najnowszej stabilnej wersji API.

Oznaczenie wersji

Oznaczenie wersji składa się z trzech liczb oddzielonych kropkami, np.: 1.2.5, gdzie kolejna pozycja oznacza następujące rodzaje wersji oznaczające:

- wersja MAJOR, gdy dokonuje się zmian niekompatybilnych z poprzednią wersją API,
- wersja MINOR, gdy dodaje się nową funkcjonalność, która jest kompatybilna z poprzednimi wersjami,
- wersja PATCH, gdy naprawia się błąd, nie zrywając kompatybilności z poprzednimi wersjami.

Dodając nowe funkcjonalności i linki, należy dążyć do wstecznej kompatybilności, która nie będzie zakłócała funkcjonalności, z których korzystają dotychczasowi użytkownicy i ich aplikacje. Jeżeli kompatybilność wsteczna zostanie zakłócona, należy zwiększyć wersję MAJOR.

Do oznaczenia i wywoływania obsługi wersją o konkretnym numerze znaczenie będzie mieć tylko pierwszy człon (MAJOR) wersji, a pozostałe (MINOR I PATCH) będą ignorowane. API powinno wyświetlać numer aktualnie używanej wersji co najmniej w dedykowanej metodzie oraz informację o spodziewanym terminie wyłączenia aktualnej wersji, jeżeli taki jest znany. Odwołanie do wersji API

może być wykonane przez parametr URI zawierający element MAJOR wersji. Jeżeli zawiera pozostałe elementy oznaczenia semantycznego wersji, są one ignorowane.

Zarządzanie zmianami wersji

Należy jednak pamiętać, że wersjonowanie nie jest jedynie kwestią techniczną. Istota modyfikacji API tkwi w celu i zakresie zmian. Zmiany, które są pożądane, mogą oddziaływać na API tak dalece, że aplikacje użytkowników po wprowadzeniu zmiany tracą dostęp do dotychczasowych zasobów. Oprócz zarządzania zmianami istotnymi są komunikacja z użytkownikami oraz przewidywalność API. Dlatego rozwiązania organizacyjne towarzyszące API powinny zapewniać możliwość komunikowania użytkownikom API zakresu planowanych i wprowadzanych zmian oraz dostarczania informacji na temat przewidywanego okresu działania danej wersji i wsparcia dla API. Spełnienie tego pozwala budować przewidywalne usługi API. Dlatego przed planowaną zmianą wersji należy oznaczyć wersję bieżącą statusem „deprecated” (przestarzałe), który poinformuje dotychczasowych użytkowników, aby nie podejmowali działań opierających się na tej wersji. Należy zakomunikować ten fakt dotychczasowym użytkownikom na stronie z opisem API oraz w opisie aktualnej wersji dostępnym pod adresem/version, a także określić okres wsparcia, zakończenia wsparcia i zakończenia funkcjonowania bieżącej wersji. Ogłaszany okres zakończenia funkcjonowania wersji nie powinien być krótszy niż dwa lata – chyba że istnieje uzasadniona potrzeba skrócenia tego czasu – i należy go dostosować do potrzeb kluczowych użytkowników. W chwili komunikacji zmian użytkownikom powinna być już dostępna dokumentacja dla nowej wersji API, tak aby mogli oni rozpocząć prace nad nowym interfejsem. Po upływie terminu wyłączenia API zapytanie nawiązujące do jego nieaktualnej wersji będzie zwracać błąd HTTP 410 (Gone) w przypadku wersjonowaniu API w URL lub 406 w przypadku innych metod, a w ciele błędu powinny znaleźć się objaśnienia na temat wyłączenia API oraz adres.

Załącznik nr 8 do „Standardu API”

Obsługa zdarzeń HTTP – tabela grup kodów zgodnie z RFC 7231 oraz przykłady odpowiedzi i komunikatów

API będące tzw. RESTful API opiera się na HyperText Transfer Protocol (HTTP), włączając w to wybrane metody, kody i nagłówki odpowiedzi. Standard API korzysta głównie z metody GET. W standardzie API zakłada się niewykorzystywanie metod umożliwiających modyfikację zasobów. HTTP dostarcza pięć kategorii oraz ponad 30 kodów odpowiedzi. Jednak nie wszystkie kody będą w API przydatne, dlatego standard API przewiduje obsługę wybranych kodów. Poniżej przedstawiono grupy kodów i ich znaczenie¹⁶¹⁾.

Tabela 14. Grupy kodów

Grupa kodów	Znaczenie	Objaśnienie
1xx	Informacyjne	Komunikaty transferu na poziomie protokołu
2xx	Powodzenie operacji	Wskazuje na powodzenie realizacji żądania klienta
3xx	Przekierowanie	Informuje, że do ukończenia żądania klient musi podjąć dodatkową akcję
4xx	Błąd klienta	Wskazanie błędu leżącego po stronie klienta

¹⁶¹⁾ Na podstawie RFC 7231.

5xx	Błąd serwera	Wskazanie błędu, za który odpowiada serwer
-----	--------------	--------------------------------------------

Załącznik nr 9 do „Standardu API”

Uwierzytelnianie i autoryzacja – opisy technologii i rozwiązań

Systemy implementujące API mogą korzystać z systemu uwierzytelniania użytkowników. Nie jest dopuszczalne zamknięcie API dla użytkowników nieuwierzytelnionych. Dozwolone jest wprowadzenie dla nich dodatkowych ograniczeń w korzystaniu z API.

Uwierzytelnianie użytkownika może być przeprowadzane na wiele sposobów i jest związane z identyfikacją, autentykacją¹⁶²⁾ i autoryzacją.

Autentykacja może być realizowana za pomocą metody Basic Auth (RFC 2617). Jest to prosty schemat interakcji, w którym podaje się identyfikator użytkownika i hasło, które wcześniej użytkownik uzyskuje po rejestracji konta na powiązanej stronie.

Podczas wszystkich operacji związanych z rozpoznawaniem użytkownika konieczne jest korzystanie z protokołu HTTPS. Niedopuszczalne jest korzystanie z metod pozbawionych szyfrowania.

W przypadku konieczności autoryzowania poszczególnych użytkowników zaleca się korzystanie z metody OAuth 2.0 (RFC 6749), (RFC 6750). OAuth 2.0 jest wyposażony w protokół, którego działanie ma następujący przebieg:

1. Klient wysyła do właściciela zasobów żądanie autoryzacyjne. Żądanie autoryzacyjne może być wysłane również metodą pośrednią. W tej metodzie żądanie wysyłane jest przez serwer autoryzacji jako pośrednika. Jest to metoda preferowana.
2. Właściciel zasobu wysyła do klienta uprawnienie do autoryzacji. Uprawnienie do autoryzacji zależy od używanej przez klienta metody do żądania autoryzacyjnego oraz metod wspieranych przez serwer autoryzacji.
3. Klient wysyła do serwera autoryzacji uprawnienie do autoryzacji, uwierzytelnia się i żąda tokena.
4. Serwer autoryzujący uwierzytelnia klienta, waliduje uprawnienie do autoryzacji i wysyła do klienta token autoryzujący, gdy uprawnienie i uwierzytelnienie zakończą się pomyślnie.
5. Klient przesyła do właściciela zasobów żądanie dostępu do zasobów oraz token autoryzujący.
6. Właściciel zasobu waliduje token, a jeżeli to się powiedzie, przesyła klientowi chroniony zasób.

Zaletą takiego podejścia jest unikalna identyfikacja czynności wykonywanych przez konkretnego klienta na zasobach API. W OAuth zewnętrzne aplikacje nie mają styczności z danymi uwierzytelniającymi użytkowników. Dodatkowo tokeny mogą ekspirować po określonym czasie, co pozwala na automatyczne wygaszanie uprawnień klienta API, który w przypadku wygaśnięcia może przejść przez procedurę odświeżenia tokena. W OAuth jeden serwer autoryzacyjny chroni wiele zasobów. Prowadzi to do redukcji kont zakładanych przez użytkowników i posługiwania się w nich często tymi samymi hasłami. Powoduje też, że użytkownicy chętniej podejmują integrację, gdyż jej rozpoczęcie nie wymaga realizowania dodatkowych czynności związanych z zakładaniem nowego konta (co realizuje cel zmniejszania bariery rozpoczęcia pracy z API).

Mimo że OAuth 2.0 nie jest jeszcze standardem zatwierdzonym przez IETF (status Proposed Standard), jest on rekomendowany jako rozwinięcie OAuth 1.0 pozbawione części jego wad oraz dostarczające bardziej kompleksowej funkcjonalności niż BasicAuth.

¹⁶²⁾ Autentykacja (uwierzytelnianie) – weryfikuje użytkownika, potwierdzając, że jest tym, za kogo się podaje – innymi słowy nadaje użytkownikowi kontekst.

Protokół OAuth nie jest jednak protokołem autentykacyjnym¹⁶³⁾, gdyż dokonuje jedynie autoryzacji – nie służy do weryfikacji użytkownika, czy jest tym, za kogo się podaje. Zastosowanie czystego OAuth 2.0 do autentykacji będzie rozwiązaniem błędnym generującym m.in. następujące problemy:

- tokeny będą uznawane w systemie za poświadczenie uwierzytelnienia w systemie, podczas gdy w założeniu OAuth realizuje tylko funkcję delegowania autoryzacji,
- dostęp do chronionego fragmentu staje się dowodem na uwierzytelnienie, a wykorzystanie tokenów na okaziciela (Bearer Tokens) skutkuje możliwością podszycia się pod wybraną tożsamość.

Naprzeciw opisanym wyżej problemom wychodzi OpenID Connect 1.0, które jest prostą warstwą nad OAuth 2.0. OpenID Connect umożliwia weryfikację wykonywaną za pomocą serwera autoryzującego, ale dodatkowo również zapewnia pozyskanie profilu klienta w sposób dedykowany stylowi REST i jest oparty na tokenach JSON Web Token (JWT)¹⁶⁴⁾. OpenID Connect obsługuje wszystkie rodzaje klientów, włączając w to klientów opartych na przeglądarce, mobilnych czy funkcjonujących bezpośrednio z Javascript. Specyfikacja OpenID Connect jest adaptowalna i rozszerzalna oraz umożliwia obsługę sesji, kiedy jest to konieczne.

¹⁶³⁾ Zob. <https://oauth.net/articles/authentication>.

¹⁶⁴⁾ Zob. Standard JWT dostarczany przez IETF <https://tools.ietf.org/html/draft-ietf-oauth-json-web-token-32>.