

Analiza otwartości danych sektora publicznego w Polsce na podstawie Centralnego Repozytorium Informacji Publicznej

1. Wstęp

Rozwój struktur społeczeństwa informacyjnego i upowszechnienie wśród obywateli dostępu do technologii informacyjno-komunikacyjnych (ICT) w znacznym stopniu przyczyniły się do rozwoju idei Otwartego Rządu (*Open Government*) na całym świecie³. Zgodnie z najważniejszymi założeniami tej koncepcji wszelkie działania państwa powinny być jak najbardziej transparentne, a dostępne możliwości technologiczne mają sprzyjać zaangażowaniu obywateli w procesy związane z podejmowaniem istotnych, z obywatelskiego i społecznego punktu widzenia, decyzji⁴.

Celem niniejszego artykułu jest przybliżenie kwestii związanych z realizacją idei *Open Government* (OG) oraz *Open Government Data* (OGD) poprzez analizę otwartości danych sektora publicznego dostępnych w repozytorium informacji publicznej w Polsce. W pierwszej części artykułu omówione zostały pokrótce prawne uwarunkowania OGD, szczególnie dotyczące funkcjonowania Centralnego Repozytorium Informacji Publicznej (CRIP). W dalszej części dokonano analizy porównawczej w zakresie metod udostępniania otwartych danych,

¹ Szkoła Główna Handlowa w Warszawie, Kolegium Analiz Ekonomicznych, Instytut Informatyki i Gospodarki Cyfrowej.

² Uniwersytet Jagielloński, Wydział Zarządzania i Komunikacji Społecznej, Katedra Zarządzania, Ekonomiki Mediów i Reklamy.

³ Zob.: F. Gonzalez-Zapata, R. Heeks, *The multiple meanings of open government data: Understanding different stakeholders and their perspectives*, „Government Information Quarterly” 2015, no. 32, s. 441–452; H.J. Wang, J. Lo, *Adoption of open government data among government agencies*, „Government Information Quarterly” 2016, no. 33, s. 80–88.

⁴ J. Hofmokl, A. Tarkowski, I. Ostrowski, M. Werner, K. Werner, J. Bartosiak, D. Solda-Kutzmann, P. Hess, Raport *Mapa drogowa otwartego rządu w Polsce*, Centrum Cyfrowe, Projekt: Polska, Warszawa 2012, <https://centrumcyfrowe.pl/projekty/mapa-drogowa/> (data odczytu: 5.05.2017).

przedstawiając sytuację Polski na tle innych państw (Stanów Zjednoczonych, Wielkiej Brytanii i Niemiec). Szczegółowa analiza danych sektora publicznego w Polsce koncentruje się na podmiotach dostarczających dane do CRIP i obejmuje ocenę danych pod kątem ich ilości i użyteczności. Szczególna uwaga skupia się na metodach udostępniania danych oraz ich formatach.

Na gruncie polskim próbę analizy metod udostępniania informacji publicznej w różnych państwach przeprowadziły w 2015 r. J. Papińska-Kacperek i K. Polańska⁵. Jednakże od tego czasu idea otwartości danych silnie się rozwija, także w Polsce. Autorzy niniejszego artykułu postanowili w związku z tym przeprowadzić nowe badanie porównujące Polskę z innymi państwami, opierając się na stanie z pierwszego półrocza 2017 r. Szczegółowej analizie poddano również charakter danych dostarczanych za pośrednictwem CRIP przez różne instytucje polskiego sektora publicznego.

2. Inicjatywa *Open Government Data* i jej znaczenie dla rozwoju społeczeństwa informacyjnego i gospodarki

Jak piszą J. Papińska-Kacperek i K. Polańska „Na przecięciu idei *Open Government* oraz otwartości powstała inicjatywa *Open Government Data* (OGD), której celem jest upublicznienie zasobów informacyjnych tworzonych przez administrację publiczną lub na jej zlecenie (z wyłączeniem danych wrażliwych), a także swobodne wykorzystanie i rozpowszechnianie otwartych danych przez każdego obywatela”⁶. Większość inicjatyw w ramach OGD koncentruje się na tym, aby umożliwić obywatelom swobodny dostęp do istotnych informacji dotyczących działalności organów państwowych. Jednym z fundamentalnych aspektów, związanych zarówno z samą ideą *Open Government*, jak i z inicjatywą OGD jest więc upowszechnienie dostępu do informacji publicznej⁷.

Wszystko to ma na celu przede wszystkim zwiększenie przejrzystości zarządzania państwem, zapobieganie różnego rodzaju nadużyciom i niewłaściwym

⁵ J. Papińska-Kacperek, K. Polańska, *Analiza zaawansowania realizacji idei Open Government Data w wybranych krajach*, „Zeszyty Naukowe Uniwersytetu Szczecińskiego” nr 874, „Studia Informatica” 2015, nr 37, s. 103–114.

⁶ Ibidem.

⁷ Zob.: A. Sobczak, T. Kulisiewicz, *Aspekty prawne i ekonomiczne ponownego wykorzystania informacji publicznej dla informatyków – ujęcie praktyczne*, Ośrodek Studiów nad Cyfrowym Państwem, Warszawa 2014, <http://opengovernment.pl> (data odczytu: 8.07.2017).

praktykom. Pozwala to jednak nie tylko na bieżące kontrolowanie działalności rządu, ale przede wszystkim na bardziej aktywne i świadome uczestnictwo obywateli w procesach decyzyjnych. Sprzyjać ma zatem dwukierunkowej aktywności w zakresie przekazywania opinii na temat działań rządowych i stwarzać przestrzeń dla współpracy w kształtowaniu i realizowaniu polityki państwa⁸. Wśród najważniejszych wyróżników koncepcji OGD wymieniane są najczęściej: przejrzystość, partycypacja i współpraca⁹. W otwartym procesie przepływu danych organizacje publiczne udostępniają informacje, które mogą być wykorzystywane przez obywateli, przedsiębiorstwa, naukowców, urzędników służby cywilnej i innych zainteresowanych¹⁰.

Badania praktycznej strony tej idei pokazują dość silną niechęć polskiego sektora publicznego do udzielania informacji na wniosek, mimo istnienia ustawowego obowiązku¹¹. Lecz idea otwartych danych to znacznie więcej niż tylko jawność informacji wyrażana w postaci biernej, tzn. udzielania informacji na wniosek. Co najmniej równie istotny jest czynny sposób informowania, oparty na samodzielnym publikowaniu informacji metodami adekwatnymi do możliwości i potrzeb.

Dostęp do różnorodnych danych sektora publicznego daje także możliwość wykorzystania ich do innych celów, różnych od pierwotnego ich przeznaczenia. Mogą to być cele prywatne, ogólnospołeczne, a w szczególności cele gospodarcze. Istotne jest pytanie w jakim stopniu prawo powinno zezwalać na ponowne wykorzystywanie tego, co powstało za publiczne pieniądze. Dane sektora publicznego mogą być traktowane jako ogólnie dostępne dobro, które można powtórnie wykorzystać także na potrzeby świadczenia komercyjnych usług, pod warunkiem, że nie naruszają one prywatności i bezpieczeństwa.

Znaczna część działań podejmowanych przez państwa na całym świecie w związku z realizacją idei OGD koncentruje się wokół problemów o charakterze prawnym i politycznym. Jednakże, jak zauważają eksperci, zagadnienia te są istotne także np. z gospodarczego punktu widzenia. Otwarte dane, zwłaszcza w kontekście ich ponownego wykorzystywania, mogą być przydatne np. dla rozwoju różnych sektorów gospodarki, w tym np. sektora zaawansowanych

⁸ J. Attard, F. Orlandi, S. Scerri, S. Auer, *A systematic review of open government data initiatives*, „Government Information Quarterly” 2015, no. 32, s. 399–418.

⁹ Zob.: Raport *Mapa drogowa...*

¹⁰ A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, R.S. Alibaks, *Socio-technical Impediments of Open Data*, „Electronic Journal of e-Government” 2012, vol. 10, no. 2, s. 156–172.

¹¹ Zob.: E. Malinowska-Misiąg (red.), *Jawność i przejrzystość finansów publicznych*, Oficyna Wydawnicza SGH, Warszawa 2016.

e-usług, który sytuuje się gdzieś na styku administracji publicznej, przedsiębiorców i obywateli¹². Szczególną wartość otwarte dane zyskują w kontekście możliwości łączenia informacji pochodzących z różnych zbiorów z wykorzystaniem metod określanych jako *Big Data*. Mogą zatem pozytywnie wpływać na rozwój przedsiębiorczości, także w perspektywie międzynarodowej, co w istotny sposób może przekładać się na wzrost gospodarczy¹³.

3. Centralne Repozytorium Informacji Publicznej w polskim systemie prawnym

W Polsce koncepcję otwartości danych publicznych regulują w szczególności dwie ustawy:

- ustawa z dnia 6 września 2001 r. o dostępie do informacji publicznej (tekst jedn.: Dz.U. 2016, poz. 1764);
- ustawa z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego (tekst jedn.: Dz.U. 2018, poz. 1243).

Praktyka i orzecznictwo podkreślają odmienny charakter praw do informacji publicznej i do ponownego wykorzystania informacji publicznej. Pierwsze zaliczane jest do kategorii praw wolnościowych i daje uprawnienie do wiedzy o działaniach władz publicznych (transparentność władzy). Drugie ma charakter prawa gospodarczego, dającego prawo do tworzenia wartości dodanej w oparciu o informacje gromadzone przez sektor publiczny¹⁴.

W pierwszej z wymienionych ustaw nakazano m.in. ogłaszanie stosownych informacji publicznych w Biuletynie Informacji Publicznej (BIP) oraz udostępnianie ich w centralnym repozytorium. Biuletyn Informacji Publicznej rozumiany jest jako ujednolicony system stron w sieci teleinformatycznej. Natomiast 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 w sposób użyteczny i efektywny, są udostępniane w centralnym repozytorium. Mają w nim być przechowywane zasoby informacyjne oraz metadane opisujące

¹² A. Sobczak, T. Kulisiewicz, op.cit., s. 62.

¹³ I. Pawełszek, *Wybrane problemy wdrożenia koncepcji otwartych danych w e-administracji*, „Roczniki Kolegium Analiz Ekonomicznych” 2014, nr 33, s. 455–470.

¹⁴ A. Gałach i in., *Dostęp do informacji publicznej a prawo do prywatności*, C.H. Beck, Warszawa 2015.

ich strukturę. Druga z wymienionych ustaw daje każdemu prawo do ponownego wykorzystania informacji sektora publicznego udostępnionych w systemie teleinformatycznym, a w szczególności na stronie BIP lub w CRIP.

Wymienione przepisy polskie implementują Dyrektywę 2003/98/WE Parlamentu Europejskiego i Rady z dnia 17 listopada 2003 r. w sprawie ponownego wykorzystywania informacji sektora publicznego¹⁵. Zgodnie z dyrektywą i – w jej ślad – z polską ustawą „ponowne wykorzystywanie” oznacza wykorzystywanie przez osoby fizyczne lub prawne dokumentów będących w posiadaniu organów sektora publicznego do celów komercyjnych lub niekomercyjnych innych niż ich pierwotne przeznaczenie w ramach zadań publicznych, dla których te dokumenty zostały wyprodukowane. Organy sektora publicznego udostępniają swoje dokumenty w ich istniejącym formacie lub języku, przy wykorzystaniu środków elektronicznych tam, gdzie jest to możliwe i właściwe.

W polskich warunkach ideę *Open Government Data* realizować ma przede wszystkim prowadzony przez Ministerstwo Cyfryzacji serwis Dane Publiczne¹⁶. Jego zadaniem jest zgromadzenie w jednym miejscu danych o szczególnym znaczeniu dla rozwoju innowacyjności w państwie i rozwoju społeczeństwa informacyjnego. Adresowany jest on zarówno do urzędników państwowych, jak i firm oraz obywateli. Jego podstawowym zadaniem jest realizacja celu CRIP. Funkcjonuje przede wszystkim w oparciu o stosowne rozporządzenia Rady Ministrów¹⁷ oraz ministra administracji i cyfryzacji¹⁸. Podstawowymi dostawcami danych mają być organy administracji rządowej, państwowe instytuty badawcze, fundusze celowe, a także inne państwowe osoby prawne.

4. Metody udostępniania otwartych danych

Obok rzetelności i kompletności danych niezmiernie istotny jest sposób publikowania danych, a w szczególności format zapisu zapewniający ich użyteczność.

¹⁵ Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, Official Journal of the European Union, L 345, s. 0090–0096.

¹⁶ <https://danepubliczne.gov.pl> (data odczytu: 5.05.2017).

¹⁷ Rozporządzenie Rady Ministrów z dnia 12 marca 2014 r. w sprawie Centralnego Repozytorium Informacji Publicznej (Dz.U. 2016, poz. 361).

¹⁸ Rozporządzenie Ministra Administracji i Cyfryzacji z dnia 26 marca 2014 r. w sprawie zasobu informacyjnego przeznaczonego do udostępniania w Centralnym Repozytorium Informacji Publicznej (tekst jedn.: Dz.U. 2016, poz. 1181).

Zalecane jest stosowanie otwartych formatów, czyli niezależnych od platformy, czytelnych maszynowo i udostępnionych publicznie bez restrykcyjnych trudniających użytkowanie informacji¹⁹. Elementem użyteczności jest także możliwość integracji danych różnego pochodzenia. W konsekwencji, aby ustalić rzeczywiste znaczenie danych, konieczna jest analiza nie tylko ich treści, ale także metadanych im towarzyszących, które opisują kontekst, w jakim dane były gromadzone. Stąd termin *linked data*, oznaczający najlepsze praktyki związane z publikacją i łączeniem strukturalnych danych w sieci²⁰.

W analizie przeprowadzonej przez Ministerstwo Cyfryzacji w Programie Otwierania Danych Publicznych²¹ zauważono, że obecnie w Polsce dominuje udostępnianie danych publicznych w formatach nieprzeszukiwalnych, a stosunkowo mały jest udział formatów ustrukturyzowanych. Zaleca się więc stosowanie formatów o jak największej liczbie gwiazdek w skali 1–5, w której poszczególne stopnie oznaczają²²:

- * udostępnienie danych w sieci Web na warunkach otwartej licencji,
- ** udostępnienie danych w formie ustrukturyzowanej,
- *** używanie formatów otwartych,
- **** używanie URI (ang. *Uniform Resource Identifier*) do oznaczania zasobów, aby możliwe było ich wyszukiwanie,
- ***** łączenie danych, aby zapewnić kontekst.

Niezależnie od klasyfikacji danych w kontekście ich formatu zaleca się umożliwianie dostępu do danych z wykorzystaniem programistycznych interfejsów API (ang. *Application Programming Interface*) wraz z opublikowaniem stosownej dokumentacji.

5. CRIP w Polsce na tle innych państw

Dla porównania CRIP w Polsce z analogicznymi repozytoriami na świecie kluczowe jest porównanie metod udostępniania danych, w tym w szczególności

¹⁹ <https://obamawhitehouse.archives.gov/open/documents/open-government-directive> (data odczytu: 5.05.2017).

²⁰ I. Pawełoszek, op.cit.

²¹ https://mc.gov.pl/files/program_otwierania_danych_publicznych_0.pdf (data odczytu: 5.05.2017).

²² T. Berners-Lee, *Linked Data*, <https://www.w3.org/DesignIssues/LinkedData.html> (data odczytu: 5.05.2017).

formatu tych danych, gdyż jest to jeden z podstawowych czynników umożliwiających powtórne wykorzystanie informacji. Do analizy autorzy wybrali 9 popularnych formatów. Następnie wyliczono orientacyjny procent zbiorów udostępnianych w poszczególnych formatach w stosunku do wszystkich badanych formatów. Do porównania wybrano USA jako jednego z liderów otwartości danych, który na dużą skalę jako pierwszy zajął się problemem *Open Government Data*, Wielką Brytanię jako innego z liderów otwartości danych według niektórych rankingów²³ oraz Niemcy, jako inne duże państwo europejskie dysponujące – podobnie jak Polska – znacznym potencjalnym zasobem informacji publicznej.

Porównanie popularności poszczególnych formatów danych w CRIP dla analizowanych krajów przedstawiono w tabeli nr 1.

Tabela nr 1. Popularność formatów danych w CRIP w wybranych państwach

Państwo	html	pdf	csv	xls xlsx	xml	json	gml	rdf	wms
Polska	9%	18%	18%	35%	9%	6%	4%	0%	0%
USA	22%	24%	12%	3%	17%	10%	2%	7%	5%
Wielka Brytania	41%	5%	23%	7%	3%	6%	0%	1%	16%
Niemcy	27%	11%	26%	14%	6%	1%	1%	0%	16%

Źródło: opracowanie własne

Według stanu na początek 2017 r. w polskim CRIP było 90 dostawców i łącznie 693 zbiorów danych. Należy zauważyć, że liczba zbiorów danych jest dość umowna, gdyż pod tym pojęciem kryją się dane o różnym charakterze, w szczególności czasem duże bazy danych składające się z wielu tabel i dużej liczby rekordów, czasem są to odnośniki do zewnętrznych aplikacji, często pojedyncze tabele w układzie arkusza kalkulacyjnego, czasem zaś tylko pojedyncze dokumenty w formacie pdf lub edytora tekstu. Dokładniejsza analiza zawartości polskiego CRIP została przedstawiona w kolejnym rozdziale.

W USA rolę CRIP pełni serwis data.gov²⁴. W chwili przeprowadzania analizy zawierał on 112 948 zbiorów danych (*datasets*), podzielonych na 14 kategorii

²³ Np. *Open data barometer. ODB Global Report. Third Edition*, Web Foundation 2015, <http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-GlobalReport.pdf> (data odczytu: 5.05.2017); K. Granickas, *Understanding the impact of releasing and re-using open government data, European Public Sector Information Platform*, Topic Report No. 2013/08, https://www.europeandataportal.eu/sites/default/files/2013_understanding_the_impact_of_releasing_and_re_using_open_data.pdf (data odczytu: 5.05.2017).

²⁴ <https://www.data.gov/> (data odczytu: 5.05.2017).

i dostarczonych przez 519 instytucji federalnych. W USA w mniejszym stopniu dane są także dostarczane przez władze stanowe i lokalne. Brytyjski CRIP to serwis data.gov.uk²⁵. Dane, zgromadzone w 43 053 zbiorach danych, udostępniło w nim 1411 instytucji, w tym administracja centralna, władze lokalne, a także w niewielkim stopniu sektor prywatny. W Niemczech rolę CRIP pełni serwis govdata.de²⁶. W czasie prowadzenia analizy zawierał on 19 119 zbiorów danych, podzielonych na 14 kategorii.

Porównując strukturę formatów dostępnych plików można zauważyć, że występują znaczące różnice pomiędzy repozytoriami różnych państw. W Polsce przede wszystkim znacznie częściej wykorzystuje się dane w zamkniętym formacie xls/xlsx (w szczególności w porównaniu z USA). Również w porównaniu z USA w Polsce znacznie rzadziej stosuje się dane wykorzystujące koncepcję *linked data* (format rdf) oraz formaty typowe dla API (np. xml, json), choć nie można tego powiedzieć w porównaniu z Wielką Brytanią i Niemcami. W porównaniu z pozostałymi analizowanymi krajami w Polsce wykorzystuje się wyraźnie mniej odwołań do zewnętrznych stron (html). Można postawić tezę, że taka struktura świadczy o wyższej dojrzałości repozytoriów OGD w USA niż w pozostałych badanych państwach, w tym w szczególności w Polsce.

6. Analiza danych w polskim CRIP

Jak wspomniano, podstawowymi dostawcami danych dla polskiego CRIP mają być organy administracji rządowej, państwowe instytuty badawcze, fundusze celowe, a także inne państwowe osoby prawne, w tym w szczególności Zakład Ubezpieczeń Społecznych, Kasa Rolniczego Ubezpieczenia Społecznego i Narodowy Fundusz Zdrowia. W praktyce istotnym dostawcą stały się również organy samorządu terytorialnego, a także w niewielkim stopniu organizacje pozarządowe.

Największą liczbę zbiorów danych dostarcza GUS (70 zbiorów). Wśród nich są roczniki statystyczne składające się z kilkusetstronicowych dokumentów w formacie pdf wraz z załącznikami w postaci tabel excelowych. Jednocześnie jako zbiory danych od GUS traktowany jest dostęp poprzez usługi sieciowe API do rejestrów publicznych TERYT i REGON. Z drugiej strony np. Urząd Dozoru

²⁵ <https://data.gov.uk/publisher> (data odczytu: 5.05.2017).

²⁶ <https://www.govdata.de> (data odczytu: 5.05.2017).

Technicznego dostarcza tylko dwa zbiory, z czego jeden to kilkustronicowa tabela danych teleadresowych oddziałów terenowych w postaci pliku pdf, drugi to wykaz krajowych zakładów uprawnionych przez UDT w postaci adresu strony internetowej z takimi informacjami.

Dostawców danych autorzy podzielili na pięć kategorii: organy administracji rządowej, w tym ministerstwa (razem 19 dostawców), inne urzędy administracji centralnej (24), instytuty badawcze (7), urzędy samorządu terytorialnego (34), pozostałe instytucje centralne i fundusze celowe (6).

W grupie organów administracji rządowej łącznie udostępnianych jest 273 zbiorów danych, lecz tylko w dwóch przypadkach udostępniono narzędzie API. Najwięcej (37 zbiorów) dostarcza Ministerstwo Sprawiedliwości, w przeważającej części w postaci plików csv (format otwarty), co pozwala na ich ocenę na poziomie 3 gwiazdek. Podobną liczbę (35) zbiorów dostarcza Ministerstwo Nauki i Szkolnictwa Wyższego, w przeważającej części w postaci plików xls (zamknięty format ustrukturyzowany), co pozwala na ich ocenę na poziomie 2 gwiazdek. Trzecie pod względem liczby udostępnionych zbiorów danych (27) jest Ministerstwo Gospodarki Morskiej i Żeglugi Śródlądowej. Są to jednak w większości mało liczne rejestry lub nawet pojedyncze dokumenty. Zastosowano tu inny sposób udostępniania danych, opierając się w przeważającej części na linkach do innych stron zawierających właściwe dokumenty oraz na dokumentach pdf (format nieustrukturyzowany), co pozwala na zakwalifikowanie tych danych na poziomie 1 gwiazdki. W omawianej grupie organów administracji rządowej nie ma w ogóle danych klasyfikowanych powyżej 3 gwiazdek. Narzędzie API udostępniły jedynie: Ministerstwo Cyfryzacji w zakresie Systemu Informacyjnego o Regionalnych Sieciach Szerokopasmowych oraz Ministerstwo Rozwoju w zakresie Centralnej Ewidencji i Informacji o Działalności Gospodarczej.

W kategorii innych urzędów administracji centralnej uwagę należy zwrócić na wspomniany wcześniej GUS ze względu na liczbę udostępnianych zbiorów (70), w tym 3 poprzez API. Drugi pod względem liczby dostarczanych zbiorów jest Urząd Ochrony Konkurencji i Konsumentów (26), który udostępnia dane przede wszystkim poprzez pliki excelowe (xls/xlsx) lub csv. Na szczególną uwagę zasługuje Główny Urząd Geodezji i Kartografii, który dostarcza 8 zbiorów, lecz część z nich w najbardziej zaawansowanej formie. W przypadku 4 zbiorów istnieją API, a w przypadku Państwowego Rejestru Nazw Geograficznych wykorzystywany jest format danych połączonych rdf, co pozwala na sklasyfikowanie zbioru na poziomie 5 gwiazdek. Ponadto dla danych Głównego Urzędu Geodezji i Kartografii należy odnotować dużą liczbę alternatywnych metod dostępu do danych. W serwisie znajduje się m.in. link do specjalistycznego portalu informacji

przestrzennych geoportal.gov.pl, realizującego również ideę otwartych danych. Także ze względu na metody udostępniania należy zwrócić uwagę na dane dostarczane przez Narodowy Instytut Dziedzictwa. Dostarcza on 7 zbiorów, lecz w przypadku 4 udostępnia API. Dla dwóch zbiorów – ewidencji pomników historii oraz obiektów wpisanych na listę światowego dziedzictwa UNESCO – udostępniono format danych połączonych z wykorzystaniem modelu rdf. Należy jednak zauważyć, że oba zbiory są mało liczne i rzadko aktualizowane.

Podmioty z kategorii instytutów badawczych zazwyczaj udostępniają tylko pojedyncze zbiory. Wyjątkiem jest Instytut Ekonomiki Rolnictwa i Gospodarki Żywnościowej, dostarczający teoretycznie 15 zbiorów, lecz w większości są to linki do strony internetowej z raportami o rynku rolnym w plikach pdf lub bezpośrednio pliki pdf.

W kategorii pozostałych instytucji centralnych i funduszy celowych uwagę należy zwrócić na Zakład Ubezpieczeń Społecznych ze względu na dużą liczbę (64) dostarczanych zbiorów. Przeważający format to pliki excelowe (xls/xlsx) oraz pdf. Z kolei Biblioteka Narodowa dostarcza tylko 2 zbiory danych, lecz oba dostępne są poprzez API – to baza danych bibliograficznych i portal Polona.pl prezentujący zdigitalizowane zbiory Biblioteki Narodowej oraz instytucji współpracujących. Także dwa API udostępnia jedyna w serwisie fundacja – Projekt: Polska. Tak udostępniane są zbiory serwisu Otwarte Zabytki.

Pozornie dość liczna jest grupa dostawców danych – urzędów samorządowych. Jednak 26 podmiotów to urzędy gmin lub niewielkich miast, które udostępniają jedynie dane adresowe z wykorzystaniem serwisu www.punktyadresowe.pl. Warto jednak odnotować, że dla tych danych udostępniany jest standard API. Jedyne samorządy dostarczające więcej zbiorów to urzędy pięciu dużych miast: Warszawy, Wrocławia, Poznania, Lublina oraz Gdańska. Cztery pierwsze, przynajmniej częściowo, wykorzystują standard API. Przykładowym udostępnionym w ten sposób zbiorem są rozkłady jazdy komunikacji publicznej.

Biorąc pod uwagę wszystkich dostawców, najpopularniejsze są pliki w formatach excelowych (xls i xlsx – 295), pdf (151) i csv (149). Mniej popularne są formaty typowe dla automatyzacji dostępu do danych z wykorzystaniem API (xml – 75, json – 50, gml – 32). Bardzo rzadko stosowany jest format typowy dla *linked open data* (rdf – 3). Często w repozytorium jako zbiór danych występuje odwołanie do strony zewnętrznej (html – 76). Szczegółowe dane z podziałem na kategorie podmiotów wraz z ich liczbą, liczbą zbiorów, liczbą API oraz popularnością wybranych formatów danych przedstawiono w tabeli nr 2.

Tabela nr 2. Charakterystyka polskiego CRIP

Kategoria	Liczba podmiotów	Liczba zbiorów	Liczba API	Html	Pdf	Csv	Xls, xlsx	Doc, docx	Xml	Json	Gml	Rdf
Organy administracji rządowej	19	273	2	34	59	69	132	8	0	1	0	0
Inne urzędy administracji centralnej	24	229	12	19	61	48	94	7	17	0	3	3
Instytuty badawcze	7	23	0	18	2	2	2	0	0	0	0	0
Urzędy samorządu terytorialnego	34	93	75	4	5	29	4	0	57	45	29	0
Pozostałe instytucje	6	75	4	1	24	1	63	2	1	4	0	0
Razem	90	693	93	76	151	149	295	17	75	50	32	3

Źródło: opracowanie własne

7. Podsumowanie

Zgodnie z założeniami CRIP ma służyć udostępnianiu zasobów informacyjnych mających istotne znaczenie zwłaszcza w kontekście rozwoju innowacyjności i struktur społeczeństwa informacyjnego. Dane te, aby mogły spełniać swoją funkcję, muszą być przygotowane w sposób ułatwiający ich przetwarzanie i ponowne wykorzystanie²⁷. Analiza dokonana przez autorów niniejszego artykułu pozwala na stwierdzenie, że zagadnienie to jest dobrze rozumiane przez polskie władze, czego odzwierciedleniem są stosowne akty prawne oraz raporty opracowywane przez instytucje rządowe. Tym niemniej zauważalne jest bardzo różne podejście poszczególnych urzędów administracji publicznej do metod udostępniania danych, co prawdopodobnie wynika z niskiej świadomości problemu na poziomie poszczególnych instytucji. Lepiej można oceniać samo udostępnianie informacji sektora publicznego niż sposób ich udostępniania umożliwiający powtórne ich użycie.

²⁷ A. Sobczak, T. Kulisiewicz, op.cit., s. 28.

Polska w zakresie metod powtórnego wykorzystania danych sektora publicznego pozostaje wyraźnie za światowymi liderami. W szczególności dotyczy to stosowanych formatów udostępnianych danych – szeroko stosowane są formaty niestrukturalne i opatentowane, niemal brak natomiast danych w formatach umożliwiających ich automatyczne łączenie – a także niewielkich możliwości dostępu do danych poprzez interfejsy programistyczne (API). Można w tej sytuacji uznać, że przepisy dotyczące sposobów udostępniania danych są zbyt niejednoznaczne oraz brakuje narzuconych odgórnie odpowiednich standardów.

Z uwagi na dynamikę zmian omawianego zagadnienia autorzy dostrzegają potrzebę prowadzenia dalszych badań w tym obszarze.

Bibliografia

- Attard J., Orlandi F., Scerri S., Auer S., *A systematic review of open government data initiatives*, „Government Information Quarterly”, 2015, no. 32, s. 399–418.
- Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, Official Journal of the European Union, L 345, s. 0090–0096.
- Gałach A. i in., *Dostęp do informacji publicznej a prawo do prywatności*, C.H. Beck, Warszawa 2015.
- Gonzalez-Zapata F., Heeks R., *The multiple meanings of open government data: Understanding different stakeholders and their perspectives*, „Government Information Quarterly” 2015, no. 32, s. 441–452.
- Malinowska-Misiąg E. (red.), *Jawność i przejrzystość finansów publicznych*, Oficyna Wydawnicza SGH, Warszawa 2016.
- Papińska-Kacperek J., Polańska K., *Analiza zaawansowania realizacji idei Open Government Data w wybranych krajach*, „Zeszyty Naukowe Uniwersytetu Szczecińskiego” nr 874, „Studia Informatica” 2015, nr 37, s. 103–114.
- Paweloszek I., *Wybrane problemy wdrożenia koncepcji otwartych danych w e-administracji*, „Roczniki Kolegium Analiz Ekonomicznych” 2014, nr 33, s. 455–470.
- Rozporządzenie Ministra Administracji i Cyfryzacji z dnia 26 marca 2014 r. w sprawie zasobu informacyjnego przeznaczonego do udostępniania w Centralnym Repozytorium Informacji Publicznej, tekst jedn.: Dz.U. 2016, poz. 1181.
- Rozporządzenie Rady Ministrów z dnia 12 marca 2014 r. w sprawie Centralnego Repozytorium Informacji Publicznej, Dz.U. 2016, poz. 361.
- Ustawa z dnia 6 września 2001 r. o dostępie do informacji publicznej, tekst jedn.: Dz.U. 2016, poz. 1764.

Ustawa z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego, tekst jedn.: Dz.U. 2018, poz. 1243.

Wang H.J., Lo J., *Adoption of open government data among government agencies*, „Government Information Quarterly” 2016, no. 33, s. 80–88.

Zuiderwijk A., Janssen M., *Open data policies, their implementation and impact: A framework for comparison*, „Government Information Quarterly” January 2014, vol. 31, iss. 1, s. 17–29.

Zuiderwijk A., Janssen M., Choenni S., Meijer R., Alibaks R.S., *Socio-technical Impediments of Open Data*, „Electronic Journal of e-Government” 2012, vol. 10, no. 2, s. 156–172.

Źródła sieciowe

Berners-Lee T., *Linked Data*, <https://www.w3.org/DesignIssues/LinkedData.html> (data odczytu: 5.05.2017).

Sobczak A., Kulisiewicz T., *Aspekty prawne i ekonomiczne ponownego wykorzystania informacji publicznej dla informatyków – ujęcie praktyczne*, Ośrodek Studiów nad Cyfrowym Państwem, 2014, <http://opengovernment.pl> (data odczytu: 8.07.2017).

Hofmokl J., Tarkowski A., Ostrowski I., Werner M., Werner K., Bartosiak J., Solda-Kutzmann D., Hess P., *Raport Mapa drogowa otwartego rządu w Polsce*, Centrum Cyfrowe, Projekt: Polska, Warszawa 2012, <https://centrumcyfrowe.pl/projekty/mapa-drogowa/> (data odczytu: 5.05.2017).

<https://danepubliczne.gov.pl> (data odczytu: 5.05.2017).

<https://www.data.gov> (data odczytu: 5.05.2017).

<https://data.gov.uk> (data odczytu: 5.05.2017).

<https://www.govdata.de> (data odczytu: 5.05.2017).

<https://obamawhitehouse.archives.gov/open/documents/open-government-directive> (data odczytu: 5.05.2017).

https://mc.gov.pl/files/program_otwierania_danych_publicznych_0.pdf (data odczytu: 5.05.2017).

<http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-GlobalReport.pdf> (data odczytu: 5.05.2017).

https://www.europeandataportal.eu/sites/default/files/2013_understanding_the_impact_of_releasing_and_re_using_open_data.pdf (data odczytu: 5.05.2017).

* * *

Public Sector Data Openness in Poland: The Analysis on the Example of the Central Public Information Repository

Abstract

The paper describes the most important assumptions of the Open Government idea and selected aspects of its implementation in Poland on the example of the Open Government Data initiative. There is described the legal status of releasing and re-use of public information, particularly in the context of the Central Public Information Repository. There is also an evaluation of open data providing in Poland by comparing selected aspects of the Central Public Information Repository functioning analogously to repositories in the United States, Great Britain, and Germany.

Keywords: Open Government, Open Government Data, public information, Central Public Information Repository (in Poland)