

ILONA PAWEŁOSZEK

Wydział Zarządzania
Politechnika Częstochowska

Wybrane problemy wdrożenia koncepcji otwartych danych w e-administracji

1. Wstęp

Powszechne prawo dostępu do informacji stało się przedmiotem dyskusji publicznych od połowy XVIII w. Pionierem w tym zakresie były państwa skandynawskie. W Szwecji w 1776 r. uchwalono Akt wolności prasy, który stał się częścią szwedzkiej konstytucji i gwarantował dostęp publiczny do oficjalnych dokumentów. Pomimo tytułu odnoszącego się do prasy, dokument zapewniał dostępność informacji każdemu obywatelowi¹. Kolejnymi państwami, które adaptowały prawo wolnego dostępu do informacji, były w drugiej połowie XX w. Stany Zjednoczone (1996 r.), Dania i Norwegia (1970 r.), Francja i Finlandia (1978 r.), Kanada, Nowa Zelandia i Australia (1982 r.) oraz Wielka Brytania (2000 r.). O ile w erze przemysłowej prawo to dawało stosunkowo niewielkie przywileje obywatelom, o tyle nie ulega wątpliwości, iż w ostatnim dziesięcioleciu technologia cyfrowa wywarła wielki wpływ na funkcjonowanie administracji publicznej i możliwość egzekwowania przez społeczeństwo swojego prawa do informacji.

Zmiany, jakie zapoczątkował rozwój technologii informacyjnej, nie tylko wywarły silny wpływ na organizację pracy urzędów i komunikację wewnętrzną, ale przede wszystkim umożliwiły państwu interakcję z obywatelami. Nastąpiło wdrożenie wielu inicjatyw mających na celu obniżenie kosztów dzięki wprowadzeniu samoobsługi petentów w różnych obszarach administracji. W Polsce platforma ePUAP umożliwia załatwienie przez Internet ponad 600 rodzajów spraw urzędowych².

¹ T. Mendel, *Freedom of Information: A comparative legal survey*, UNESCO, New Delhi 2003, s. 79.

² <http://epuap.gov.pl>.

Jednakże cyfrowa rewolucja w administracji nie ogranicza się jedynie do usprawnienia przesyłania danych, ograniczenia liczby błędów i redukcji kosztów. Coraz większą rolę w komunikacji z obywatelami odgrywają media społecznościowe – platformy takie jak Facebook czy Twitter. W ten sposób jednostki administracji informują o swoich działaniach lub odpowiadają na pytania.

Oprócz samoobsługi petentów i komunikacji za pomocą mediów społecznościowych główny trend, który pojawił się w ostatnich latach i jest postrzegany jako ogromne źródło innowacji, stanowi inicjatywa otwartych danych (ang. *Open Data*). Koncepcja ta opisuje zbiory danych, które są bez ograniczeń dostępne i mogą być używane, jak również przetwarzane i publikowane przez każdego bez restrykcji dotyczących praw autorskich lub patentów, z zastrzeżeniem co najwyżej do wymogu podania źródła czy umożliwienia dalszej dystrybucji przetworzonych treści na takich samych zasadach³. Celem inicjatywy *Open Data* jest w szczególności udostępnienie wszystkich danych nieosobowych (nienaruszających praw obywateli do ochrony danych osobowych) i niekomercyjnych (nienaruszających tajemnicy handlowej) posiadanych przez organizacje rządowe.

W tym nurcie instytucje administracji publicznej w wielu krajach rozpoczęły udostępnianie danych rządowych na swoich portalach w sieci Internet w postaci usług webowych (ang. *Web Services*) lub przez interfejsy typu REST. W założeniu ma to prowadzić do lepszego zrozumienia polityki instytucji administracji publicznej przez obywateli, uczynienia jej w ten sposób bardziej wiarygodną i skuteczną. Otwartość danych rządowych dla społeczeństwa ma również zwiększyć uczestnictwo obywateli we współpracy z organami władzy oraz pobudzić innowacyjność. Publikowane dane odnoszą się do wielu dziedzin, najbardziej popularne to: ochrona środowiska, dane i statystyki dotyczące zatrudnienia, budżety jednostek samorządowych, mapy, rozkłady jazdy komunikacji publicznej itp. Publikować je mogą indywidualne jednostki administracji lub dostawcy większych repozytoriów, które gromadzą publiczne zbiory danych i udostępniają je w scentralizowanej i standaryzowanej formie. Pomimo tego samego celu leżącego u podstaw ich udostępniania, repozytoria mogą bardzo się od siebie różnić. Różnice te mogą dotyczyć dziedzin, zakresu, rozmiaru, a także stosowanych standardów zapisu danych. Niewątpliwie najważniejszą cechą zbiorów otwartych danych jest ich użyteczność w świetle konkretnych zadań. Aby przyczynić się do rozwoju otwartego społeczeństwa, platformy publikacji

³ P. Koziarski, R. Kabaciński, M. Lis, P. Kaczmarek, *Open Access. Analiza zjawiska z punktu widzenia polskiego naukowca*, s. 8, <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>.

otwartych danych powinny spełniać określone wymogi prawne, administracyjne, jak również techniczne.

Celem niniejszego artykułu jest prezentacja korzyści oraz problemów i zagrożeń wynikających z wykorzystania otwartych danych.

2. Istota otwartego rządu i otwartych danych

Otwarty rząd to pojęcie obejmujące wiele różnorodnych terminów i koncepcji. Często otwarty rząd traktowany jest jako synonim słowa e-government i wykorzystania technologii informacyjno-komunikacyjnej w administracji. Jednakże takie porównanie nie jest do końca poprawne, gdyż otwarty rząd jest paradygmatem teoretycznym, natomiast technologia jest narzędziem, którym rząd się posługuje. Należy też nadmienić, iż administracja najczęściej opiera się na technologiach już sprawdzonych w biznesie (np. systemach zarządzania treścią, obiegu dokumentów, systemach klasy ERP, rozwiązaniach mobilnych czy rozwiązaniach funkcjonujących w chmurze).

W wielu źródłach literaturowych istnieje zgodność co do tego, że otwartość jest podstawowym przymiotem e-rządu funkcjonującego w duchu demokracji XXI w.⁴ OECD w następujący sposób definiuje cechy otwartego rządu: jawność (przejrzystość) działań rządu, dostępność usług rządowych i informacji oraz reagowanie na nowe pomysły, żądania i potrzeby⁵. Te elementy są postrzegane jako podstawa do realizacji innych korzyści dla rządu i społeczeństwa, wśród których można wyliczyć: tworzenie bazy wiedzy do ustalania polityki, umacnianie integralności, zniechęcanie do korupcji i budowanie społecznego zaufania⁶. Wszystkie wymienione zadania otwartego rządu leżą głównie w sferze informacyjnej i mogą być w dużej mierze zrealizowane przez implementację mechanizmów udostępniania otwartych danych rządowych.

Wdrażanie mechanizmów otwartego rządu ma na celu zwiększenie zaangażowania społecznego, zainteresowanie obywateli sferą publiczną, umożliwienie im uczestniczenia w procesach decyzyjnych rządu oraz generalnie budowanie społeczeństwa obywatelskiego przez pobudzanie interakcji pomiędzy administracją

⁴ D. Banisar, *Effective Open Government: Improving public access to government information*, OECD Publishing, Paris 2005.

⁵ *Modernising Government: The way forward*, red. T. Curristine, B. Abbott, OECD Publishing, Paris 2005.

⁶ Ibidem.

a obywatelami. Kluczowym zadaniem otwartego rządu jest zatem otwieranie dostępu do generowanych przez jednostki administracji zasobów informacyjnych i wskazywanie możliwości ich dalszego wykorzystania. Dostęp do publicznych zasobów umożliwia obywatelom użycie ich w twórczy sposób w rozmaitych aplikacjach⁷. Publikowane dane pozwalają zorientować się obywatelom w zakresie działań i celów instytucji, a wiedza ta daje im możliwość większego zaangażowania w kwestie publiczne. Otwartość rządu w tym zakresie można rozumieć jako wyjście naprzeciw obywatelom i uczynienie z nich partnerów w dialogu przez zwiększanie ich wiedzy na temat sposobów i efektów funkcjonowania organizacji rządowych.

W praktycznej realizacji idei otwartego rządu ważne jest zintensyfikowanie komunikacji państwo–obywatel. W Polsce w odniesieniu do kluczowej kwestii, jaką jest dostęp do informacji publicznej (i możliwość jej wykorzystania), zdecydowanie widać, że proces wdrażania koncepcji otwartego rządu, mimo wyraźnych sygnałów o jego rozpoczęciu, znajduje się w fazie embrionalnej⁸. Rysunek 1 pokazuje etapy ewolucji państwa w stronę realizacji koncepcji otwartego rządu.



Rysunek 1. Droga do otwartego rządu

Źródło: *Otwarty rząd w Polsce. Kulisy programu Opengov*, red. G. Młynarski, Fundacja Projekt: Polska, Warszawa 2013, s. 33.

⁷ <http://centrumcyfrowe.pl/projekty/otwarty-rzad-w-polsce>.

⁸ *Otwarty rząd w Polsce. Kulisy programu Opengov*, red. G. Młynarski, Fundacja Projekt: Polska, Warszawa 2013, s. 32.

Wprowadzenie opartych na technologiach informacyjnych i komunikacyjnych mechanizmów może w Polsce pomóc rozwiązać wiele problemów, takich jak⁹:

- niski poziom dialogu obywatelskiego,
- brak konsultacji społecznych dotyczących decyzji administracyjnych i ustawodawstwa,
- deficyt komunikacji pomiędzy instytucjami publicznymi a obywatelami,
- niewystarczająca transparentność działań rządu i samorządów,
- niski poziom zaufania do instytucji publicznych.

Pogłębienie zaangażowania obywateli przez wykorzystanie technologii wymaga zintegrowanych działań wielu podmiotów różnych szczebli administracji, organizacji pozarządowych oraz biznesu, wreszcie przychylnego nastawienia samych obywateli.

3. Techniczne aspekty udostępniania i wykorzystywania otwartych danych

Otwarte dane publikowane przez rząd lub samorzady lokalne mogą być cennym źródłem informacji, pod warunkiem, że będą opublikowane w odpowiedniej, użytecznej formie. Podstawą skutecznego udostępniania danych jest możliwość ich odczytania przez maszyny oraz otwarta licencja pozwalająca na użytkowanie danych.

Otwarta licencja na format danych jest niezbędnym czynnikiem wykorzystania zasobów *Open Data*. Ilustruje to sytuacja¹⁰, która miała miejsce w Anglii w 2008 r. Programista A. Nicolson był zaangażowany w kampanię przeciw budowie drogi na terenach zielonych w miejscowości Westbury w hrabstwie Wiltshire. Interesował go dostęp i wykorzystanie danych na temat ruchu drogowego, które były podstawą oceny i decyzji o realizacji projektu budowy drogi. Zgodnie z ustawą o swobodnym dostępie do informacji¹¹, otrzymał dane, o które prosił, jednakże samorząd lokalny dostarczył dane w formacie, który mógł być odczytany jedynie przy użyciu specjalistycznego oprogramowania firmy Saturn, specjalizującej się w tworzeniu aplikacji służących modelowaniu i przewidywaniu

⁹ <http://centrumcyfrowe.pl/projekty/otwarty-rzad-w-polsce>.

¹⁰ <http://opendatahandbook.org/en/appendices/file-formats.html>.

¹¹ The Freedom of Information Act 2000 (c.36) – jest aktem Parlamentu Zjednoczonego Królestwa uchwalonym w 2000 r., który ustanawia publiczne prawo dostępu do informacji posiadanej przez władze publiczne, <http://www.foia.gov>.

ruchu drogowego. Dostawca oprogramowania nie oferuje wersji przeznaczonej tylko do przeglądania plików, dlatego też Nicholson wraz z grupą działającą przeciw budowie drogi musiał nabyć licencję na oprogramowanie, której cena (wersji edukacyjnej) wynosiła 600 EUR.

Obecnie w wielu państwach ustawy dotyczące swobodnego dostępu do danych zawierają klauzulę dotyczącą formatu danych. Dla przykładu, jako złotą zasadę uznaje się postanowienie prezydenta USA B. Obamy zawarte w Dyrektywie otwartego rządu (ang. *Open Government Directive*) z grudnia 2009 r., stwierdzające¹²: „W możliwie najszerszym zakresie oraz z zastrzeżeniem obowiązujących ograniczeń agencje powinny publikować informacje on-line w otwartym formacie, który może być otworzony, pobrany, indeksowany i przeszukiwany przez powszechnie używane aplikacje wyszukiwawcze sieci Web. Otwarty format to taki, który jest niezależny od platformy, czytelny maszynowo i udostępniony publicznie bez restrykcji, utrudniających użytkowanie informacji”.

Obecnie zbiory danych najczęściej są publikowane w formie wymagającej od użytkowników w celu odnalezienia interesującego ich zbioru danych wejścia na stronę internetową odpowiedniej instytucji, która je publikuje. Często zachodzi potrzeba integracji danych z kilku źródeł, co jest skomplikowane i czasochłonne. Problem integracji pojawia się, gdyż każda organizacja posiada inną politykę dotyczącą publikacji danych, co sprawia, że są one heterogeniczne pod względem formatu.

Prawdziwa korzyść płynąca z udostępniania otwartych danych wynika z możliwości integracji różnych źródeł informacji. Integracja taka polega na możliwości łatwego zlokalizowania i zestawienia danych odnoszących się do określonego obiektu lub zjawiska, zwłaszcza gdy dane te pochodzą z różnych źródeł i przedstawiają badany fragment rzeczywistości w odmiennych kontekstach.

Łatwo wyobrazić sobie możliwość kompleksowego raportu wygenerowanego dla określonego obszaru geograficznego. Raport taki mógłby pokazywać np.: wzrost ekonomiczny, wykorzystanie odnawialnych źródeł energii, śmiertelność i publiczne wydatki na edukację czy służbę zdrowia. Dane tego typu można zwizualizować w różnych aspektach przy pomocy wachlarza narzędzi, np.: publicznie dostępnych serwisów map, tabeli przestawnych lub wykresów. Do niedawna tego typu analizy i wizualizacje możliwe były jedynie w drogich

¹² http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf.

systemach *Business Intelligence*, dostępnych dla wybranych grup zawodowych, jak informatycy i kadra menedżerska dużych firm.

Integracja otwartych danych nie może być dokonana wprost, ponieważ mają one różnorodny charakter, nawet jeśli dotyczą tego samego zjawiska, niekoniecznie można to stwierdzić w prosty sposób, patrząc na zestawienia w tabelach. Kolumny tabel mogą być różnie nazwane, dodatkowo w przypadku danych przekrojowych pozostaje kwestia ustalenia punktu lub przedziału czasowego, jaki one opisują. Aby ustalić rzeczywiste znaczenie danych, należałoby zatem analizować nie tylko ich treść, ale także metadane im towarzyszące, opisujące kontekst, w jakim dane były gromadzone. Dodatkowo trzeba wziąć pod uwagę różnorodność jednostek miary oraz skali, jakimi mogli się posługiwać twórcy danych w różnych jednostkach administracyjnych. W krajach, które niedawno rozpoczęły upublicznianie danych rządowych, pochodzą one z systemów, które były nastawione wyłącznie na zaspakajanie potrzeb informacyjnych ich bezpośrednich użytkowników. Udostępnianie danych społeczeństwu wymaga planowania już na etapie zbierania danych, aby gromadzić je w sposób, który umożliwia potem ich adaptację do otwartego formatu i udostępnienie w zrozumiałej (dla maszyn lub ludzi) formie.

Odpowiednie oprogramowanie może wspomagać obecności użytkowników w zadaniach integracji danych. Po dokonaniu integracji użytkownicy mogą eksplorować nowe dane, ekstrahować potrzebne informacje i – jeśli potrzeba – wizualizować rezultaty. Sposób, w jaki użytkownicy używają otwartych danych, jest w znacznym stopniu uwarunkowany formą ich publikacji. Duże ilości danych generowane przez urzędy administracji publicznej i organizacje mogą być publikowane małym nakładem pracy w „surowej” formie, w jakiej są gromadzone, jednakże nie są one wtedy zbyt użyteczne dla użytkownika zewnętrznego. Wiele zbiorów danych podlega zatem procesom oczyszczania i kustomizacji przed ich opublikowaniem. Można zaobserwować dwa style publikowania danych – pierwszy to forma czytelna dla człowieka, drugi natomiast to format czytelny dla maszyn¹³.

Termin „połączone dane” (ang. *Linked Data*) odnosi się do zbioru najlepszych praktyk związanych z publikacją i łączeniem strukturalnych danych w sieci. Praktyki, o których mowa, są coraz częściej wykorzystywane przez dostawców danych, co pozwala stworzyć globalną przestrzeń danych zawierającą miliony hiperłączy, wiążących powiązane ze sobą znaczeniowo dane. Zbiór taki określany

¹³ Przykładem mogą być dane publikowane przez „New York Times” w formacie RDF do pobrania ze strony: <http://data.nytimes.com>.

jest często mianem *Web of Data*, czyli sieć danych¹⁴. Techniczne rozumienie pojęcia *Linked Data* odnosi się do danych publikowanych w sieci w sposób zrozumiały dla maszyn, o jasno zdefiniowanym znaczeniu, połączonych z innymi zewnętrznymi zbiorami danych i możliwymi do wykorzystania przez zewnętrzne źródła danych.

T. Berners-Lee nakreślił zbiór reguł publikowania danych w sieci w sposób zapewniający, że wszystkie dane staną się częścią globalnej przestrzeni danych¹⁵:

1. Wykorzystanie adresów URI jako nazw – aby jednoznacznie identyfikować encje, dane są semantycznie wzbogacone i używają zestandaryzowanego słownictwa.
2. Używanie standardu HTTP do tworzenia identyfikatorów URI – na podstawie paradygmatu REST każdy zbiór danych otrzymuje unikalny identyfikator, za pomocą którego można do niego dotrzeć z każdego miejsca w sieci przy użyciu protokołu HTTP¹⁶.
3. Wykorzystanie standardów RDFi SPARQL w celu wyszukiwania informacji.
4. Połączenie danych przez URI, aby wskazać ich relację z innymi zbiorami – w ten sposób możliwe jest wyszukiwanie nowych danych, eksplorując hiperłącza między zasobami, połączenia tego typu zwiększają wartość informacji.

Powyższe punkty stanowią techniczne postawy publikowania otwartych połączonych danych (ang. *Linked Open Data* – LOD), można zauważyć, iż kluczową kwestią jest tutaj wykorzystanie standardów.

Rozpatrując strategie różnych inicjatyw w zakresie *Open Data* (jak np. *Data.gov*), można zauważyć, iż nie ma ustalonego standardu ani zgodności co do tego, w jakiej postaci dane powinny być udostępniane. Widać tu bowiem dwie tendencje udostępniania danych:

- 1) w postaci „surowej”,
- 2) po obróbce statystycznej.

Jedna i druga forma mają swoje zalety i wady. Udostępnianie surowych danych (bezpośrednio w postaci tabel z baz danych) jest mniej kosztowne dla

¹⁴ *Reasoning Web. Semantic Technologies for the Web of Data – 7th International Summer School 2011*, red. A. Polleres, C. d'Amato, M. Arenas, S. Handschuh, P. Kroner, S. Ossowski, P.F. Patel-Schneider, Galway, Ireland, August 23–27, Springer, Berlin–Heidelberg 2011, s. 29.

¹⁵ T. Berners-Lee, C. Bizer, T. Heath, *Linked Data – The Story So Far*, <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>.

¹⁶ A. Logean, O. Lavrovsky, R.K. Straumann, P. Gassner, *Sustainability through open data: Examples from Switzerland*, http://www.researchgate.net/publication/233793818_Sustainability_through_Open_Data_Examples_from_Switzerland.

instytucji publikującej. Taka forma daje ich potencjalnym użytkownikom duże możliwości analityczne, ale jednocześnie wymaga od nich umiejętności informatycznych i wiedzy w zakresie statystycznej obróbki danych. Zawęży ona zatem krąg odbiorców i generalnie zniechęca do zaglądania na strony rządowe osoby bez odpowiedniego wykształcenia. Ponadto – według A. Meijera¹⁷ – publikowanie surowych danych może przynieść niepożądane efekty: zamiast działać zgodnie z zasadą przejrzystości działań rządu, może podważyć zaufanie społeczne, gdyż dane są wyjęte z kontekstu, przyjmują głównie ustrukturalizowaną formę numeryczną oraz kierują uwagę społeczeństwa na wąskie i niekoniecznie ważne (aczkolwiek wymierne) zjawiska.

Natomiast udostępnianie danych po obróbce statystycznej (w postaci raportów, wykresów, dobrze opisanych i zinterpretowanych wskaźników) sprawia, iż użytkownicy dostają gotową, łatwą do przyswojenia informację. Jednakże istnieje niebezpieczeństwo, iż raporty i statystyki mogą być tworzone w sposób tendencyjny, ukazują lub wyolbrzymiają trendy pokazujące aktualną władzę w korzystnym świetle i tuszują zjawiska, które mogłyby wywołać głosy dezaprobaty ze strony społeczeństwa.

4. Ryzyko udostępniania otwartych danych

Jak wykazują doświadczenia państw, które od kilku lat udostępniają otwarte dane (min. Stany Zjednoczone, Wielka Brytania), samo umieszczenie w Internecie danych, które pochodzą z różnych systemów informatycznych danej instytucji, nie realizuje szczytnego celu takiego przedsięwzięcia, nawet jeśli te dane są w odpowiednim otwartym formacie. Dane potrzebują bowiem właściwej interpretacji i muszą być publikowane w sposób przemyślany.

Poniższy przykład prezentuje skutki niewłaściwego wykorzystania otwartych danych. Sytuacja dotyczyła zastosowania otwartych danych na temat przestępczości do publikowania interaktywnych map. Serwis internetowy www.police.uk pozwalał na wyszukiwanie przestępstw i antyspołecznych zachowań odnotowanych przez policję w ciągu ostatniego miesiąca w określonej lokalizacji, którą można było podać w postaci kodu pocztowego lub nazwy ulicy.

¹⁷ A. Meijer, *Understanding modern transparency*, „International Review of Administrative Sciences” 2009, vol. 75(2), s. 255–269.

Stworzenie map przestępczości stanowiło oczywistą korzyść publiczną, wskazując policji, gdzie należy skierować uwagę i środki, aby poprawić taktykę działania. Korzyścią dla społeczności była możliwość łatwego zidentyfikowania ryzykownych obszarów miasta, aby unikać niebezpieczeństw, zachować wzmożoną czujność i wymagać ochrony ze strony policji, jeśli zajdzie taka potrzeba. Jednakże realizacja projektu natrafiła na pewne problemy. Na przykład w grudniu 2011 r. niewielka, spokojna uliczka Surrey w mieście Portsmouth, na której mieścił się tylko jeden pub (przeważnie odwiedzany przez pracowników pobliskiego urzędu pocztowego), kilka domów mieszkalnych oraz parking, okazała się bardzo niebezpiecznym miejscem, w którym – według mapy – odnotowano aż 136 przestępstw w ciągu ostatniego miesiąca, podczas gdy w rzeczywistości miały tam miejsce tylko dwa. Dla porównania, w najbardziej niebezpiecznym mieście w Wielkiej Brytanii odnotowano wtedy 152 przestępstwa¹⁸. Paradoksalna sytuacja została wyjaśniona po dokładniejszej analizie specyfiki źródeł danych i działania aplikacji. Okazało się, że przestępstwa rejestrowane były według kodów pocztowych, spokojna Surrey Street posiadała ten sam kod pocztowy co pobliskie centrum handlowe, gdzie zdarzały się kradzieże w sklepach. Kod pocztowy był również ten sam dla pobliskiej ulicy barów i klubów, w których często odnotowywano interwencje policji w związku z przypadkami przemycy wywołanymi przez osoby pod wpływem alkoholu.

Badanie wykonane w tym samym roku przez firmę ubezpieczeniową wykazało, że 11% respondentów deklaruje, że było świadkiem przestępstwa w swojej dzielnicy, ale nie zgłaszało go w obawie, że podanie takiej informacji do ogólnej wiadomości zmniejszy ich szanse na sprzedaż lub wynajęcie ich domu.

Wadą koncepcji otwartych danych jest ignorowanie informacyjnego sprzężenia zwrotnego. Należy mieć na uwadze to, iż sam akt otwarcia źródła danych może mieć wpływ na jego późniejszą wiarygodność¹⁹. Nasuwa się również inny wniosek z przedstawionego wyżej studium przypadku – sama możliwość dostępu do dowolnych danych nie jest wiele warta, gdy brakuje umiejętności i wiedzy pozwalającej te dane właściwie interpretować.

Pomimo wszystkich pozytywnych efektów, udostępnianie otwartych danych może potencjalnie wywierać negatywne skutki społeczne w postaci naruszenia prywatności lub niewłaściwego (niezgodnego z prawem) ich wykorzystania. Dla przykładu, w Stanach Zjednoczonych rejestr osób posiadających broń przez długi

¹⁸ <http://www.bbc.co.uk/news/uk-england-hampshire-12332347>.

¹⁹ D. Hand, *Open data is a force for good, but not without risks*, <http://www.theguardian.com/society/2012/jul/10/open-data-force-for-good-risks>.

czas był publicznym dokumentem, jednakże informacje tam zawarte udzielane były na życzenie petenta, nie zaś jako otwarty zbiór danych, do którego dowolna osoba ma dostęp w dowolnym czasie. Mimo iż jest to dokument publiczny, udostępnienie go w formie otwartych danych, które mogą być w łatwy sposób zestawione z danymi geograficznymi, stwarza niebezpieczeństwo zarówno dla posiadaczy, jak i dla osób nieposiadających broni. Informacja ta może być anonimowo, bez większych trudności pobrana i wykorzystana przez przestępców do planowania włamań do określonych posesji. Należy zatem zauważyć, iż nie tylko udostępnienie poprzednio tajnych danych, ale także ułatwienie dostępu i przetwarzania tych, które już są dostępne publicznie, może stwarzać zagrożenie dla prywatności²⁰. Pojawia się więc pytanie, czy wszystkie publiczne dokumenty powinny być udostępniane w Internecie na tych samych zasadach. Pewnym, aczkolwiek niedoskonałym rozwiązaniem zdaje się kontrolowane udostępnianie niektórych danych przez konieczność założenia konta na portalu rządowym z podaniem podstawowych danych osobowych.

Kwestią dyskusyjną pozostaje także odpowiedzialność za niezgodne z prawem użycie danych, z pewnością w najbliższym czasie kwestie te będą wymagały uaktualnień obowiązującego prawa w wielu krajach. Na przykład w Wielkiej Brytanii pojawiła się propozycja przerwania odpowiedzialności prawnej za wykorzystanie danych z osoby, która publikuje dane, na osobę, która dokonuje ich niezgodnego z prawem wykorzystania. Propozycja dotyczyła zmian w prawie w stronę modelu, w którym prawo kontroluje użycie danych, a nie dostęp do nich. Taki model jest jednak skomplikowany w realizacji, zwłaszcza w kontekście globalizacji, gdy informacja może łatwo przekraczać granice państw²¹.

Obok licznych korzyści, otwarte dane mogą także budzić pewne obawy i wątpliwości. W założeniu udostępnianie dotyczy danych odpersonalizowanych, które nie naruszają niczyich praw do prywatności ani tajemnic przedsiębiorstw. O ile pojedyncze źródła danych mogą spełniać takie założenie, o tyle w odniesieniu do połączonych danych z wielu źródeł nie można z całą pewnością przewidzieć, czy w efekcie analizy danych nie zaistnieje możliwość naruszenia prywatności²².

²⁰ <http://www.opendataresearch.org/content/2013/501/open-data-privacy-discussion-notes>.

²¹ Ibidem.

²² Przypadek taki jest opisany w: K. O'Hara, E. Whitley, P. Whittall, *Avoiding the Jigsaw Effect: Experiences With Ministry of Justice Reoffending Data*, 2011, <http://eprints.soton.ac.uk/273072>.

5. Stadia rozwoju koncepcji otwartych danych

Aby w pełni wykorzystać możliwości, jakie dają otwarte dane, należy je osadzić w odpowiednim kontekście, który tworzy z nich nową wiedzę i umożliwia korzystanie z niej przez usługi i aplikacje. Otwarte połączone dane (LOD) można rozpatrywać zarówno w kontekście ich wykorzystania, jak i publikacji. T. Berners-Lee wyznaczył ścieżkę rozwoju otwartych danych w stronę otwartych połączonych danych w postaci modelu pięciu gwiazdek (tabela 1). Model ten przedstawia charakterystykę rozwiązań praktycznych w zakresie otwartych danych z uwzględnieniem stopnia realizacji koncepcji LOD.

Tabela 1. Plan realizacji koncepcji otwartych danych

Generacja	Znaczenie	Przykład
★	opublikowane w sieci Web, otwarta licencja	dokument PDF zawierający tabele
★★	jw., lecz format strukturalny	arkusz kalkulacyjny Excel, PC-Axis
★★★	jw., lecz w nieopatentowanym formacie	CSV, SDMX
★★★★	jw. + metadane, posiadają hiperłącza	RDF/XML
★★★★★	jw. + hiperłącza do innych zbiorów danych tworzące kontekst	RDFa z hiperłączami zewnętrznymi

Źródło: *Linking Government Data*, red. D. Wood, Springer, New York–Dordrecht–Heidelberg–London 2011, s. 138, cyt za: T. Berners-Lee, C. Bizer, T. Heath, *Linked Data – The Story So Far*, <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> (data odczytu 11.11.2013).

Każda z pokazanych powyżej generacji rozwoju koncepcji otwartych danych przynosi korzyści i wiąże się z pewnymi kosztami zarówno po stronie instytucji publikującej, jak i po stronie konsumenta danych. Poniższe zestawienie (tabela 2) wyszczególnia niektóre z nich.

Tabela 2. Koszty i korzyści na poszczególnych etapach rozwoju koncepcji *Open Data*

★ Koszty i korzyści	
Konsument	Instytucja publikująca
<ul style="list-style-type: none"> • może zobaczyć dane, • może drukować dane, • może przechowywać lokalnie (na dysku twardym USB itp.), • może wpisać manualnie dane do innego systemu 	<ul style="list-style-type: none"> • łatwość publikacji
★★ Koszty i korzyści	
Konsument	Instytucja publikująca
<p>Wszystko co oferuje generacja ★ oraz:</p> <ul style="list-style-type: none"> • możliwość bezpośredniego przetwarzania i agregacji za pomocą odpowiedniego oprogramowania, • możliwość eksportu do innego strukturalnego formatu 	<ul style="list-style-type: none"> • łatwość publikacji
★★★ Koszty i korzyści	
Konsument	Instytucja publikująca
<p>Wszystko co oferuje generacja ★★ oraz:</p> <ul style="list-style-type: none"> • brak konieczności opłat licencyjnych za oprogramowanie otwierające pliki z danymi 	<ul style="list-style-type: none"> • łatwość publikacji
★★★★ Koszty i korzyści	
Konsument	Instytucja publikująca
<p>Wszystko co oferuje generacja ★★★ oraz:</p> <ul style="list-style-type: none"> • tworzenie hiperłączy do zbioru z każdego innego miejsca w sieci lub lokalnie, • dodawanie zakładek, • ponowne użycie wybranych danych 	<ul style="list-style-type: none"> • stworzenie przekrojów danych w celu umożliwienia zaawansowanej analizy, • przypisanie adresów URI do danych, • przemyślenie kwestii formatu reprezentacji danych, • większa kontrola danych i możliwość optymalizacji dostępu do nich (np. równoważenie obciążenia, <i>caching</i> itp.)
★★★★★ Koszty i korzyści	
Konsument	Instytucja publikująca
<p>Wszystko co oferuje generacja ★★★★ oraz:</p> <ul style="list-style-type: none"> • odkrywanie nowych danych przez łączenie wielu zbiorów, • dostęp do schematów danych 	<ul style="list-style-type: none"> • inwestycja w łączenia danych z innymi zbiorami w sieci, • ułatwienie odnajdywania danych, • zwiększenie wartości danych

Źródło: <http://www.semantic-web.at/LOD-TheEssentials.pdf>.

Zaprezentowane powyżej koszty i korzyści w każdym stadium rozwoju są ograniczone do kwestii technicznych i – ze względu na znajomość możliwości technologii dostępnej już dziś – stosunkowo łatwe do przewidzenia. Dużo trudniejsza i budząca więcej wątpliwości jest kwestia społecznych kosztów i korzyści płynących z wdrażania inicjatyw otwartego rządu.

6. Podsumowanie

Inicjatywa otwartych danych w administracji publicznej może przynieść wiele korzyści natury społecznej i ekonomicznej. Możliwość ujawnienia faktów dotyczących funkcjonowania rządu i samorządów stwarza większe poczucie odpowiedzialności wśród decydentów. Otwarte dane mogą stanowić jawny dowód na skuteczność (bądź jej brak) polityki rządu, której efekty mają odzwierciedlenie w publikowanych wskaźnikach.

Aplikacje stworzone na bazie otwartych danych udostępnione publicznie zwiększają wiedzę społeczeństwa oraz świadomość dotyczącą takich wskaźników, jak: zagrożenie naturalnego środowiska, przestępczość, osiągnięcia naukowe, poziom edukacji, dostępność usług publicznych. Otwarte dane mogą także pobudzać wzrost gospodarczy i sprzyjać przedsiębiorczości również w wymiarze międzynarodowym²³.

Obecnie temat otwartych danych jest jeszcze stosunkowo nowy, wiele państw dopiero przystosowuje swoje systemy informacyjne do tego, aby w zadowalający sposób móc sprostać wyzwaniu upubliczniania danych. Otwarte połączone dane są zatem ciekawym i obiecującym kierunkiem badań w zakresie stawianych im wymagań technicznych i jakościowych oraz możliwości tworzenia nowych, coraz doskonalszych aplikacji. Jednakże jeszcze ważniejszym aspektem wdrażania koncepcji otwartych danych są efekty społeczne i polityczne, jakie mogą wywołać w dalszej perspektywie. W tym obszarze niezbędne jest szukanie sposobów na znalezienie równowagi pomiędzy szybkością, jakością i zakresem udostępniania a ochroną danych osobowych i prywatności²⁴.

²³ K. Węcel, *Linked Data in Enterprisesm*, w: *Systemy wspomaganie organizacji*, red. T. Porębska-Miąc, H. Sroka H, SWO, Wydawnictwo Uniwersytetu Ekonomicznego, Katowice 2013, s. 287–309.

²⁴ http://centrumcyfrowe.pl/wp-content/uploads/2011/01/Mapa-otwartego-rzadu-w-Polsce_draft.pdf.

Bibliografia

1. Banisar D., *Effective Open Government: Improving public access to government information*, OECD Publishing, Paris 2005.
2. Kozierski P., Kabaciński R., Lis M., Kaczmarek P., *Open Access. Analiza zjawiska z punktu widzenia polskiego naukowca*, Wydawnictwo Impuls, Poznań–Kraków 2013.
3. *Linking Government Data*, red. D. Wood, Springer, New York–Dordrecht–Heidelberg–London 2011.
4. Meijer A., *Understanding modern transparency*, „International Review of Administrative Sciences” 2009, vol. 75(2).
5. Mendel T., *Freedom of Information: A comparative legal survey*, UNESCO, New Delhi 2003.
6. *Modernising Government: The way forward*, red. T. Curristine, B. Abbott, OECD Publishing, Paris 2005.
7. *Otwarty rząd w Polsce. Kulisy programu Opengov*, red. G. Młynarski, Fundacja Projekt: Polska, Warszawa 2013.
8. *Reasoning Web. Semantic Technologies for the Web of Data – 7th International Summer School 2011*, red. A. Polleres, C. d’Amato, M. Arenas, S. Handschuh, P. Kroner, S. Ossowski, P.F. Patel-Schneider, Galway, Ireland, August 23–27, Springer, Berlin–Heidenberg 2011.
9. Węcel K., *Linked Data in Enterprises*, w: *Systemy wspomagania organizacji*, red. T. Porębska-Miąc, H. Sroka, SWO, Wydawnictwo Uniwersytetu Ekonomicznego, Katowice 2013.

Źródła sieciowe

1. Berners-Lee T., Bizer C., Heath T., *Linked Data – The Story So Far*, <http://tom-heath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> (data odczytu 11.11.2013).
2. <http://centrumcyfrowe.pl/projekty/otwarty-rzad-w-polsce> (data odczytu 11.11.2013).
3. http://centrumcyfrowe.pl/wp-content/uploads/2011/01/Mapa-otwartego-rzadu-w-Polsce_draft.pdf (data odczytu 11.11.2013).
4. <http://data.nytimes.com> (data odczytu 11.11.2013).
5. <http://eprints.soton.ac.uk/273072> (data odczytu 11.11.2013).
6. <http://epuap.gov.pl> (data odczytu 11.11.2013).
7. <http://opendatahandbook.org/en/appendices/file-formats.html> (data odczytu 11.11.2013).

8. <http://opendatahandbook.org/pdf/OpenDataHandbook.pdf> (data odczytu 11.11.2013).
9. http://www2012.wwwconference.org/proceedings/nocompanion/wwwweb-sci2012_braunschweig.pdf (data odczytu 11.11.2013).
10. <http://www.w3.org/DesignIssues/LinkedData.html> (data odczytu 11.11.2013).
11. <http://www.bbc.co.uk/news/uk-england-hampshire-12332347> (data odczytu 11.11.2013).
12. <http://www.foia.gov> (data odczytu 11.11.2013).
13. <http://www.opendataresearch.org/content/2013/501/open-data-privacy-discussion-notes> (data odczytu 11.11.2013).
14. <http://www.theguardian.com/society/2012/jul/10/open-data-force-for-good-risks> (data odczytu 11.11.2013).
15. http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf (data odczytu: 11.11.2013).
16. Logean A., Lavrovsky O., Straumann R.K., Gassner P., *Sustainability through open data: Examples from Switzerland*, http://www.researchgate.net/publication/233793818_Sustainability_through_Open_Data_Examples_from_Switzerland (data odczytu 11.11.2013).

* * *

Selected issues of implementing the concept of open data in e-government

Summary

In the last few years, open data has grown into an almost mainstream activity for many governments around the world. While opening up datasets provides great opportunities, the idea of open data also raises various kinds of challenges to its stakeholders. The paper discusses the selected rules of publishing open data by governments, as well as some aspects of social and economic impact on society.

Keywords: open data, open linked data, open government