

Marzena Kowalska

Dlaczego za dostęp do pewnych zasobów wiedzy (danych) musimy zapłacić, inne zaś dostępne są bezpłatnie - próba weryfikacji idei otwartego dostępu do zasobów wiedzy

Forum Bibliotek Medycznych 5/1 (9), 521-526

2012

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

**DLACZEGO ZA DOSTĘP DO PEWNYCH ZASOBÓW WIEDZY
(DANYCH) MUSIMY ZAPŁACIĆ, INNE ZAŚ DOSTĘPNE SĄ
BEZPŁATNIE - PRÓBA WERYFIKACJI IDEI OTWARTEGO DOSTĘPU
DO ZASOBÓW WIEDZY**

Abstract

The subject of the speech is an attempt to verify the idea of an open access to the resources of knowledge which also include Polish and worldwide statistics databases. It is commonly known that using global reliable statistics at each level of data aggregation comprises the basis for creating and making a decision pertaining to each citizen. The idea of giving an access to the data collected by Polish statistics or EUROSTAT will be analysed. Methods and rules of giving an access to data by those units will also be compared. An attention will be given to the necessity of making a payment for giving an access to some data despite the fact that the public statistics is financed from the state budget. The other databases not only the statistical ones and the rules of an access and use of their resources will also be analysed.

Streszczenie

Przedmiotem wystąpienia będzie próba weryfikacji idei otwartego dostępu do zasobów wiedzy, do których zaliczyć również można bazy polskiej i światowej statystyki. Powszechnie bowiem wiadomo, że posługiwanie się wiarygodnymi danymi statystycznymi na każdym poziomie agregacji danych stanowi podstawę do tworzenia i podejmowania decyzji dotyczących każdego obywatela. Analizie poddana zostanie między innymi idea udostępniania danych gromadzonych przez polską statystykę oraz EUROSTAT. Porównane zostaną także sposoby i zasady udostępniania danych przez te dwie jednostki. Uwaga poświęcona zostanie również konieczności dokonania opłaty za udostępnienie pewnych danych pomimo faktu, że statystyka publiczna finansowana jest ze środków budżetu państwa. Analizie zostaną poddane również inne bazy, nie tylko statystyczne, oraz zasady dostępu i korzystania z ich zasobów.

Idea Otwartego Dostępu, czyli udostępnianie efektów badań naukowych społeczeństwu za darmo, a w szerszym kontekście, wszelkich danych produkowanych za pieniądze podatników, o ile nie są one oczywiście tajne, staje się coraz bardziej popularna. Niemniej jednak rodzi wiele pytań, czasami kontrowersji. Jej założenia nie są bowiem wolne od sprzeczności i potencjalnych konfliktów związanych z prawem autorskim oraz interesami ekonomicznymi podmiotów działających choćby na polu nauki. Kolejnym dużym problemem jest także niewystarczająca infrastruktura uniemożliwiająca swobodny dostęp do zbiorów.

XXI wiek w Polsce musi zapisać się jako ten, w którym dokonał się m.in. wolny i darmowy dostęp do treści, dostępnych nie tylko w publikacjach naukowych, ale przede wszystkim w Internecie, będącym obecnie jednym z podstawowych kanałów komunikacji. Ponadto należy dążyć do tego, aby coraz bardziej powszechne wyko-

rzystywanie zasobów cyfrowych w badaniach naukowych oraz w rozpowszechnianiu wiedzy stały się głównymi cechami szczególnymi nauki XXI wieku.

W kontekście podjętego tematu warto zaznaczyć, że dane statystyczne, obok publikacji relacjonujących wyniki badań, są zasadniczym elementem całego procesu badawczego, dlatego również ich znaczenie, zwłaszcza w ostatnich latach, zasadniczo wzrosło. Dodatkowo ich rola wzrosła gdy możliwe okazało się ich efektywne udostępnianie w Internecie. Łatwo dostępne i przejrzyste dane statystyczne niewątpliwie często stanowią podstawę dalszej pracy naukowo-badawczej lub są jej istotnym uzupełnieniem. Zatem idea wolnego i pełnego dostępu do treści publikacji i danych uzupełniają się wzajemnie. W takim kontekście dzielenie się danymi oraz ich udostępnianie powinno być normą. Niemniej jednak dostęp do danych statystycznych, również danych statystyki publicznej, jest często ograniczany, a mało przejrzyste standardy publikacji i udostępniania tych danych powodują ograniczone możliwości z ich korzystania. Koniecznym jest nie tylko zapewnienie pełnej dostępności do danych, zwłaszcza danych, których zbieranie finansowane jest środkami płynącymi z budżetu państwa, ale i „wymuszenia” nie tylko na ośrodkach naukowych, lecz także na innych instytucjach publicznych publikowania wyników prac, które powstały dzięki udziałowi publicznych pieniędzy. „Podatnicy już zapłacili za te informacje, musimy przynajmniej zwrócić je tym, którzy chcą z nich korzystać w nowy sposób, pomagając obywatelom w tworzeniu miejsc pracy i stymulowaniu wzrostu gospodarczego”, stwierdziła Neelie Kroes, komisarz UE ds. agendy cyfrowej w kontekście działań Komisji Europejskiej, która w zeszłym roku uruchomiła europejską strategię otwartego dostępu do danych rządowych¹.

Idea Open Access na polu statystyki publicznej skutecznie mogłaby zostać wprowadzana dzięki m.in. udoskonalaniu oraz pełnym udostępnianiu prowadzonych baz danych. Tak pojmowana otwartość zasobów/danych jest bowiem podstawą budowania zaufania społecznego do statystyki, a co za tym idzie i zdobycia poparcia dla starań o rozszerzenie przedsięwzięcia Open Access na inne, poza ściśle naukowe pola. Oczywiście jest bowiem, że nieograniczony i przede wszystkim bezpłatny dostęp do danych wspiera rozwój naukowy czy wzrost gospodarczy, ponadto społeczeństwo ma prawo do obszerniejszej informacji o wynikach prowadzonych badań, szczególnie, że są one finansowane ze środków publicznych.

W ostatnich latach obok terminu Open Access zaczęto stosować termin Open Data (otwarte dane), który odnosi się do różnorodnych projektów mających na celu zniesienie barier stojących na drodze do pełnej dostępności i interoperacyjności danych naukowych. Niestety działania na rzecz otwartości danych nie mają jednak jeszcze

¹ Komunikat prasowy Komisji Europejskiej, *Agenda cyfrowa: Zamienianie danych rządowych w złoto*, Bruksela12. 06. 2011, dostępne: <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/1524&format=HTML&aged=0&language=PL&guiLanguage=en> [27.08.2012]

takiej rangi ani tego stopnia instytucjonalizacji co ruch Open Access². Niemniej jednak z powodzeniem odbyło się w Polsce kilka akcji mających na celu promowanie ruchu Open Data. Przykładem takich działań był zorganizowany w 2010 roku w Centrum Cyfrowym w Warszawie hackaton „otwarte dane publiczne”. Tego dnia kilkanaście osób – w większości informatyków, ale nie tylko – spotkało się by wspólnie pracować nad aplikacjami wykorzystującymi otwarte dane publiczne. Spotkanie, które miało formułę sprintu informatycznego, odbyło się w ramach międzynarodowej inicjatywy –Open Data Day – której celem jest skłaniać lokalne, regionalne i państwowe władze do wdrażania polityki otwartości.

Hackaton pokazał, że w ciągu kilku godzin informatycy oraz osoby nie zajmujące się programowaniem mogą stworzyć rozmaite aplikacje w oparciu o istniejące i dostępne narzędzia, dane i źródła. Uczestnicy, pracując w grupach, podjęli się kilku zadań m.in. przetworzenia i wizualizacji danych Głównego Urzędu Statystycznego. Zasadniczym problemem okazały się trudności, związane głównie z brakiem jednorodnego, czytelnego sposobu prezentowania informacji i danych przez tę instytucję publiczną. Stosunkowo szybko okazało się, że pracę z Bankiem Danych Lokalnych istotnie utrudnia fakt, że GUS nie udostępnia publicznego API pozwalającego na łatwy i wygodny dostęp do danych. Zamiast tego radzić trzeba sobie ze złożonymi formularzami, a prezentowane dane posiadają różnego rodzaju uchybienia i luki. Dla przykładu, brakuje informacji na jakim poziomie podziału administracyjnego zostały zebrane dane. Niektóre zbiera się na poziomie województwa, inne na poziomie gminy lub podregionu – w efekcie porównanie danych różnych typów staje się praktycznie niemożliwe. Sposób publikacji danych GUS sprawia, że trzeba wiele czasu i wysiłku, by przygotować je tak, by stały się czytelne dla komputera³.

W kontekście podjętego tematu uwagę należy zwrócić także na inne wątpliwości, które pojawiają się wśród użytkowników nie będących informatykami, poszukujących danych statystycznych w zasobach polskiej statystyki publicznej. Wątpliwości wzbudzają bowiem niektóre procedury udostępniania danych z zasobów polskiej statystyki publicznej, a mianowicie ilość nieudostępnianych danych, sprzedaż danych przez tę instytucję publiczną oraz nieprzejrzysty sposób ich prezentacji w Internecie. Idealnie wręcz, owe kwestie w swojej analizie⁴ wyłuszczył Marek Tatała, ekspert Forum Obywatelskiego Rozwoju, które są, jak sędzę, głosem wszystkich korzystających, a nie do

² Justyna Hofmokl, Alek Tarkowski, Bożena Bednarek-Michalska, Krzysztof Siewicz, Jakub Szprot: *Przewodnik po otwartej nauce*, Warszawa 2009 s. 59, dostępne: <http://otwartanauka.pl/wp-content/uploads/2010/01/przewodnik-po-otwartej-nauce.pdf> [27.08.2012]

³ Open Data Day 2010, dostępne: <http://centrumcyfrowe.pl/projekty/open-government-data-camp/open-data-day-2010/> [27.08.2012]

⁴ Marek Tatała: *GUS ogranicza dostęp do statystyki publicznej*, „FOR Ostrzeża. Komunikat nr 17 z 11 lipca 2012”, dostępne: <http://www.for.org.pl/pl/a/2239,FOR-ostrzeza-nr-17-GUS-ogranicza-dostep-do-statystyki-publicznej> [dostęp: 28.08.2012]

końca usatysfakcjonowanych, odbiorców danych statystycznych.

Po pierwsze, jak słusznie zauważył Tatała, GUS blokował dostęp do pełnych wersji elektronicznych sześciu z dziesięciu regularnie publikowanych roczników. Z kolei ok. 70% urzędów statystycznych w UE, które publikują na swoich stronach narodowe roczniki statystyczne, udostępnia ich pełne wersje bezpłatnie. GUS powinien zatem nie tylko dołączyć do tej otwartej na obywateli grupy urzędów, ale także umożliwić pełny dostęp do wszystkich publikowanych przez siebie roczników statystycznych na swojej stronie internetowej. Największe zaskoczenie czy wątpliwości wśród zewnętrznych odbiorców danych statystycznych wzbudza zjawisko sprzedaży pewnych, nieopublikowanych danych statystyki publicznej. Jest to o tyle dyskusyjne, ponieważ działalność GUS finansowana jest z pieniędzy podatników. GUS żąda więc często opłaty za coś, za co już raz zapłaciliśmy w podatkach. Częstym tłumaczeniem urzędu jest dodatkowa praca, którą urzędnicy muszą włożyć w udostępnienie zamawianych danych. Na stronie internetowej GUS próżno jednak szukać cennika takich „dodatkowych usług”, a wyceny przedstawiane przez GUS nie posiadają uzasadnienia pokazującego, skąd wynika żądana przez urząd cena. Dostęp do części danych i informacji jest nadal ograniczany przez GUS, co zdaje się być zaprzeczeniem podstawowej idei publicznego charakteru gromadzonych przez GUS statystyk i powszechnej dostępności danych⁵.

Z kolei strona internetowa GUS⁶ w następujący sposób informuje o płatnych zasadach udostępniania danych: Zamówienia na dane niepublikowane oraz na informacje z prowadzonych urzędowych rejestrów realizowane są odpłatnie. Podstawą do opracowania kalkulacji kosztów udostępniania danych jest aktualnie obowiązujący „Cennik za usługi świadczone przez służby statystyki publicznej”. Zakres i formę udostępnianych informacji statystycznych określa Program Badań Statystycznych Statystyki Publicznej zatwierdzany corocznie przez Radę Ministrów. Wszystkie inne agregacje danych statystycznych poza tymi, które wymieniono w Programie - mogą być wykonywane (przy zachowaniu tajemnicy statystycznej) na specjalne zamówienie odbiorcy za odpłatnością. Z kolei Dyrektor Departamentu Informacji GUS w następujący sposób ustosunkował się do zarzutów stawianych przez Marka Tatałę: „(...) odpłatnie realizujemy na podstawie art. 21 ustawy o statystyce publicznej, opracowania i analizy oraz przeprowadzamy badania. Koszt realizacji w przypadku opracowań i analiz jest kalkulowany na podstawie roboczo-godzin potrzebnych pracownikom do wykonania zamówienia (...). Stwierdzenie, że GUS pobiera opłaty za dane, które zostały zebrane za pieniądze podatników jest całkowicie bezpodstawne, gdyż to dane są przedmiotem zamówień płatnych, lecz ich opracowanie i analiza zgodnie z potrzebami zamawiającego, a więc wkład intelektualny i czas pracowników poświęcony na realizację zamówienia (...). Warto także nadmienić, że pod koniec 2013 roku zakończymy

⁵ Tamże.

⁶ http://www.stat.gov.pl/gus/5548_PLK_HTML_HTML.htm [dostęp: 28.08.2012]

realizowany przez nas projekt pod nazwą System Informacyjny Statystyki Publicznej (...). Jednym z celów projektu jest udostępnianie użytkownikom Publicznych Baz Danych, zawierających maksymalną ilość danych możliwych do udostępnienia, poprzez usprawnienie wyszukiwarki, prezentację danych w atrakcyjnej formie (zobrazowanie danych na mapach, dynamiczne wykresy, itp.), a także udostępnianie użytkownikom prostych narzędzi analitycznych”⁷.

Ponadto zaniedbaniem polskiej statystyki jest poświęcanie zbyt małej uwagi potrzebom wszystkich odbiorców danych statystycznych. W porównaniu z tym, co dzieje się w świecie, polskie publikacje statystyczne mogą wydawać się nudne, a informacje statystyczne przedstawiane w przestarzały i nieatrakcyjny sposób. Ponadto źródłem wielu informacji statystycznych o Polsce jest Eurostat, a nie GUS.

Niemniej jednak należy zwrócić uwagę na znaczne różnice w istocie funkcjonowania organów statystyki polskiej oraz instytucji statystyki Unii Europejskiej. Główną rolą Eurostatu jest bowiem wyłącznie przetwarzanie i publikowanie porównywalnych informacji statystycznych na poziomie europejskim. Prace skupiają się również wokół osiągnięcia wspólnego statystycznego „języka”, który musi obejmować koncepcje, metody, struktury i normy techniczne. Ponadto, co jest najważniejsze, Eurostat nie zbiera danych. Odbywa się to w państwach członkowskich przez ich instytucje statystyczne, które sprawdzają i analizują krajowe dane i wysyłają je do Eurostatu. Rolą Eurostatu jest konsolidacja danych i upewnienie się, że przy zharmonizowanej metodologii są one porównywalne⁸.

Eurostat udostępnia szczegółowe dane statystyczne w 9 zakresach tematycznych: ogólne statystyki oraz statystyki regionalne, gospodarka i finanse, ludność i warunki socjalne, przemysł, handel i usługi, rolnictwo i rybołówstwo, handel zagraniczny, transport, środowisko i energia, nauka i technologia. Ponadto zaproponowana przez Eurostat baza danych zdaje się pełniej niż BDL zaspokajać potrzeby użytkowników, m.in. poprzez szerokie możliwości ekstrakcji danych, poprzez możliwość wyboru informacji (wskaźnik, jednostka, czas, itd.) oraz dostępne formaty (XLS, TXT, HTML, SPSS, itp.). Dodatkowe narzędzie Easy Comext- oraz baza COMEXT zawierająca dane nt. handlu zagranicznego krajów członkowskich UE, dane nt. handlu między krajami członkowskimi jak i wymiany handlowej z krajami spoza UE charakteryzują się wieloma zaletami, których brak w polskich bazach danych: dostęp do danych pochodzących z różnych źródeł (UE, ONZ, MFW itp.) oraz do danych o różnych strukturach (wg różnych klasyfikacji, takich jak CN, SITC, REV2, SITC REV3, itp.).

Dodatkowo, poprzez rejestrację otrzymuje się powiadomienia przez e-mail, a także możliwość dostosowania drzewka nawigacji danych do własnych potrzeb oraz możli-

⁷ Odpowiedź z Departamentu Informacji GUS z 20.07.2012, dostępne: <http://www.for.org.pl/pl/a-/2252,Komunikat-na-temat-dzialan-GUS> [dostęp: 27.08. 2012]

⁸ http://epp.eurostat.ec.europa.eu/portal/page/portal/about_eurostat/introduction

wość ekstrakcji większej ilości danych. Ponadto po wysłaniu zapytania dotyczącego danych statystycznych Eurostat dostarcza odpowiedzi w przeciągu jednego dnia dla zapytań standardowych, a w przeciągu 5 dni użytkownik otrzymuje odpowiedź na pytanie, które wymaga pomocy wydziału produkcyjnego Eurostatu.

Promocja modeli Open Access oraz Open Data jest koniecznym warunkiem dla zapewnienia polskiej nauce, polskiej statystyce konkurencyjności oraz potencjału innowacyjnego. Ponadto otwarty i pełny dostęp do danych instytucji publicznych przyczyni się do zwiększenia ich przydatności, efektywności oraz jakości prac, które powstają w ich oparciu. Pomimo faktu, że jako kraj jesteśmy dopiero u narodzin ruchu „otwartych danych” to w miarę udostępniania kolejnych ich zasobów będzie można je łączyć i krzyżować, co stworzy pole do powstania aplikacji jeszcze skuteczniej ułatwiających pracę ludziom nauki, obywatelom czy przedsiębiorcom. Wszystko zaś czynione jest po to, aby dostęp do nauki, wyników prowadzonych prac naukowo-badawczych był równy dla każdego, niezależnie od wykształcenia, stanu materialnego czy miejsca zamieszkania. Sukcesem będzie, jeżeli plany te wejdą w życie. I temu należy kibicować oraz mieć nadzieję, że również u nas Open Access zyska większe poparcie.

Bibliografia

Hofmokl Justyna, Tarkowski Alek, Bednarek-Michalska Bożena, Siewicz Krzysztof, Szprot Jakub, *Przewodnik po otwartej nauce*, Warszawa 2009, s. 59, dostępne: <http://otwartanauka.pl/wp-content/uploads/2010/01/przewodnik-po-otwartej-nauce.pdf> [27.08.2012]

Komunikat prasowy Komisji Europejskiej, *Agenda cyfrowa: Zamienianie danych rządowych w złoto*, Bruksela 12. 06. 2011, dostępne: <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/1524&format=HTML&aged=0&language=PL&guiLanguage=en> [27.08.2012]

Open Data Day 2010, dostępne: <http://centrumcyfrowe.pl/projekty/open-government-data-camp/open-data-day-2010/> [27.08.2012]

Odpowiedź z Departamentu Informacji GUS z 20.07.2012, dostępne: <http://www.for.org.pl/pl/a-12252,Komunikat-na-temat-dzialan-GUS> [dostęp: 27.08. 2012]

Tatała Marek, *GUS ogranicza dostęp do statystyki publicznej*, „FOR Ostrzeża. Komunikat nr 17 z 11 lipca 2012”, dostępne: <http://www.for.org.pl/pl/a/2239,FOR-ostrzeza-nr-17-GUS-ogranicza-dostep-do-statystyki-publicznej> [dostęp: 28.08.2012]

http://www.stat.gov.pl/gus/5548_PLK_HTML.htm [dostęp: 28.08.2012]