

Wybrane aspekty prawne scrapowania ze szczególnym uwzględnieniem informacji i danych publicznych

Selected legal aspects of web-scraping with special regard to public sector data and information scraping

Streszczenie: Autorka omawia prawne aspekty scrapowania. W opracowaniu przeanalizowano stopniowe zmiany w zakresie ram prawnych w UE i w Polsce dotyczące pozyskiwania danych z internetu. Odrębna część opracowania została poświęcona zagadnieniu ekstrakcji ze źródeł dostępnych w internecie szczególnego rodzaju treści w postaci danych i informacji publicznych. Autorka wskazuje na pewne odrębności w odniesieniu do legalności ich pozyskiwania. Zwraca ponadto uwagę na niedawno przyjęte przepisy regulujące eksplorację tekstów i danych oraz zagadnienie *re-use*.

Słowa kluczowe: scrapowanie, dane, pozyskiwanie danych, ekstrakcja danych, agregacja danych, eksploracja tekstów i danych, bazy danych, ochrona baz danych, prawo autorskie, prawo do informacji publicznej, informacja publiczna, otwarte dane, ponowne wykorzystywanie informacji sektora publicznego, prawo UE

Abstract: The Author presents legal aspects of web-scraping. The paper discusses the gradual development of a legal framework regarding internet data mining in the EU and in Poland. Specific attention is paid to scraping of public sector data and information. The Author indicates why different approach should be adopted for data made available through electronic means by the 'public sector body' and 'body governed by public law'. Furthermore, it is explained how the latest legislation on text and data mining and re-use.

Keywords: Web-scraping, data, data extraction, data collection, data aggregation, data mining, text and data mining, databases, databases protection, copyrights law, re-use of public sector information, open data, public sector data, EU law

¹ Radca prawny, specjalista w zakresie prawa własności intelektualnej, prawa nowych technologii i ochrony danych osobowych (kontakt: tel. 881 790 719, adres email: ewa.wojnarska-krajewska@wp.pl).

Wstęp

Niniejsze opracowanie poświęcone jest zagadnieniu automatycznego pobierania danych z internetu. Szczególna uwaga została w nim poświęcona tym danym i informacjom, które zostały udostępnione na stronach internetowych przez podmioty zobowiązane do udostępniania informacji publicznej i informacji do ponownego wykorzystywania.

Dla jasności wyводу należy odwołać się do pojęć z zakresu terminologii informatycznej, które mają kluczowe znaczenie dla zrozumienia prawnych aspektów zagadnienia będącego przedmiotem niniejszej analizy.

Współczesna gospodarka w coraz większym stopniu oparta jest na przetwarzaniu wielkich zbiorów danych. Dla ich określenia używa się pojęcia *big data*. Odnosi się ono do dużej ilości danych, nieustrukturyzowanych, o dużej różnorodności², których rozmiary i szybkość napływania wymagają posłużenia się szczególnymi narzędziami do ich gromadzenia, przechowywania, agregowania, czyli analizowania i zestawiania, a w efekcie do ich przekształcania w nową wartość³.

Narzędziem, które umożliwia ekstrakcję istotnych dla określonego celu danych z wielkich zbiorów i dokonania ich analizy, jest *data mining*, a szerzej *text and data mining* (czyli „eksploracja tekstów i danych”). W efekcie jego zastosowania z dotychczas dostępnych danych można wygenerować nowe informacje lub dokonać prognoz⁴. Dzięki temu mogą powstawać innowacyjne rozwiązania w nauce i gospodarce. Klasycznym przykładem mogą być dane z badań medycznych. Zestawione z dużej

² A. Heiss, *Big Data Challenges in Big Science, Computing and Software for Big Science*, „Computing and Software for Big Science” 2019, <https://link.springer.com/content/pdf/10.1007%2Fs41781-019-0030-7.pdf> [dostęp: 23.08.2020].

³ Zob. A. De Mauro, A.M. Greco, M. Grimaldi, *A formal definition of Big data based on its essential features*, „Library Review” 2016, Vol. 3, s. 122-135.

⁴ S. Selamat, S. Prakoonwit, W. Khan, *A review of data mining in knowledge management: applications/findings for transportation of small and medium enterprises*, „SN Applied Sciences” 2020, <https://link.springer.com/content/pdf/10.1007%2Fs42452-020-2589-3.pdf> [dostęp: 23.08.2020].

liczby baz i przeanalizowane prowadzą do nowych wniosków na temat chorób, uwidaczniają określone tendencje lub pozwalają przewidzieć ryzyka.

Pozyskiwanie danych z wielkich baz odbywa się automatycznie za pomocą techniki, którą określa się mianem *scraping* (*web scraping*, *screen scraping*). *Scraping*, zwany również „scrapowaniem”, polega na ekstrakcji dużych ilości informacji ze stron internetowych za pomocą automatycznych narzędzi – internetowych botów – programów komputerowych, których zadaniem jest systematyczne przeszukiwanie internetu w celu pobrania z nich wyselekcjonowanych informacji według zadanych kryteriów, może iść w parze z indeksowaniem treści⁵.

Scrapowanie dotyczy wszelkiego rodzaju danych dostępnych w internecie. Aktywność ta jest od dawna przedmiotem obaw wielu przedsiębiorców, którzy upubliczniają swoje bazy danych i jednocześnie dążą do zapewnienia im jak najszerzej ochrony. Z drugiej strony – na podstawie dostępnych danych powstają przedsięwzięcia, które opierają się na kreatywnym ich przekształcaniu w produkty o nowym znaczeniu i nowej wartości dla odbiorców.

Szczególne miejsce wśród dostępnych danych zajmują te wytworzone przez podmioty funkcjonujące w sferze publicznej albo przez takie, które mają prawny obowiązek udostępniania określonych informacji. Mają one z mocy prawa inny status niż dane udostępniane przez prywatne podmioty. Szczegółowe regulacje poświęcone zarówno ogólnym zasadom scrapowania i wykorzystywania tak pozyskanych danych, jak i dokonywania tych czynności w odniesieniu do danych i informacji publicznych, zostały omówione w dalszej części pracy.

⁵ A. Zamora, *Making Room For Big Data: Web Scraping and an Affirmative Right to Access Publicly Available Information Online*, “The Journal of Business, Entrepreneurship & the Law” 2019, Vol. 12, s. 203-228; E. Wojnarska-Krajewska, *Crawlowanie – legalne czy nie*, <https://wpprawoautorskie.wordpress.com/2014/04/17/crawlowanie-crawler-roboty-txt/> [dostęp: 24.08.2021].

1. Scrapowanie w świetle strategicznych założeń Unii Europejskiej w zakresie otwierania dostępu do danych sektora publicznego

Przepisy Unii Europejskiej odnoszą się do scrapowania w sposób bezpośredni, jednak tylko w określonym zakresie. Scrapowanie podlega przede wszystkim ograniczeniom prawnym w odniesieniu do chronionych baz danych. Producentów bazy danych, którzy ponieśli istotne nakłady inwestycyjne na ich stworzenie, chronią przepisy dyrektywy 96/9/WE Parlamentu Europejskiego i Rady z dnia 11 marca 1996 r. w sprawie ochrony prawnej baz danych⁶. Jest to o tyle istotna regulacja, że znacząca część stron internetowych opiera się na bazach danych. Za takie bowiem uznać można na przykład zbiory danych o produktach podlegających sprzedaży, dane podmiotów fizycznych udostępniane w ramach serwisów społecznościowych, informacje o filmach, zdjęcia i recenzje dostępne na stronach serwisów VOD i wiele innych.

Zakres tych ograniczeń określa art. 5 dyrektywy, zgodnie z którym producentowi bazy danych przysługuje wyłączne prawo do czasowego lub trwałego zwielokrotniania, w jakikolwiek sposób lub w jakiegokolwiek formie, w całości lub w części bazy danych (art. 5 lit. a) oraz do dokonywania w odniesieniu do niej wszelkich innych czynności określonych w art. 5 lit. b-e oraz art. 7 ust. 1, zgodnie z którym producentowi przysługuje prawo do ochrony przed pobieraniem danych i/lub wtórnym ich wykorzystaniem w całości lub w istotnej części, co do jakości i/lub ilości. Art. 7 ust. 5 dyrektywy stanowi, iż: „niedozwolone jest powtarzające się i systematyczne pobieranie danych i/lub wtórne ich wykorzystanie w nieistotnej części, jeśli czynności te są sprzeczne z normalnym korzystaniem z tej bazy danych lub w sposób nieuzasadniony naruszają słuszne interesy producenta bazy danych”. W tym zakresie zatem, w odniesieniu do

⁶ Dz. Urz. UE L 77 z 27.03.1996 r., s. 20.

baz danych chronionych prawem *sui generis*, scrapowanie zostało wprost ograniczone.

Istotne przepisy regulujące pobieranie treści z internetu zostały wprowadzone również na mocy dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/790 z dnia 17 kwietnia 2019 r. w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym oraz zmiany dyrektyw 96/9/WE i 2001/29/WE⁷. Przywołana dyrektywa uregulowała kwestię eksploracji tekstów i danych, czyli posługiwania się narzędziami umożliwiającymi przetwarzanie dużych ilości informacji w celu uzyskania nowej wiedzy i odkrycia nowych tendencji. W niektórych przypadkach eksploracja tekstów i danych może obejmować czynności chronione prawem autorskim lub prawem do bazy danych *sui generis*, w szczególności zwielokrotnianie utworów lub innych przedmiotów objętych ochroną lub pobieranie treści z bazy danych. W tym zakresie dyrektywa przewiduje wprowadzenie wyjątków, a poza ich zakresem – konieczność uzyskania zezwolenia od podmiotów uprawnionych.

Ponieważ niniejsze opracowanie poświęcone jest scrapowaniu szczególnego rodzaju danych, a mianowicie tych udostępnianych przez podmioty zaliczane do sektora publicznego, należy zwrócić uwagę na wyjątkowe podejście unijnego prawodawcy do tej kategorii danych. Informacje sektora publicznego odnoszą się do wszystkich informacji, które podmioty publiczne wytwarzają, zbierają lub za które płacą. Przykładami są informacje geograficzne, statystyki, dane pogodowe, dane z projektów badawczych finansowanych ze środków publicznych i zdigitalizowane książki z bibliotek. Polityka Unii Europejskiej zmierza do jak najszerzego otwarcia tych zasobów w celu tworzenia wartości dla gospodarki i społeczeństwa poprzez ponowne wykorzystywanie zawartych w nich danych⁸.

⁷ Dz. Urz. UE L 130 z 17.05.2019 r., s. 92.

⁸ Zob. Komisja Europejska, *Shaping Europe's digital future, Open Data*, <https://ec.europa.eu/digital-single-market/en/open-data> [dostęp: 28.08.2021].

Wyżej wskazane podejście znalazło swój wyraz w wielu europejskich strategicznych dokumentach i inicjatywach, ostatnio w Europejskiej Strategii w zakresie danych przyjętej przez Komisję Europejską w dniu 19 lutego 2020 r., zgodnie z którą: „Obywateli należy wyposażać w odpowiednie narzędzia umożliwiające im podejmowanie lepszych decyzji w oparciu o doświadczenia uzyskane w obszarze danych nieosobowych. Dane te powinny natomiast być dostępne dla wszystkich [...]. Dzięki temu społeczeństwo będzie w stanie czerpać jak największe korzyści z innowacji i konkurencji”.

Realizacja tych założeń opiera się na kilku filarach. Po pierwsze transparentność procesu decyzyjnego oraz dostęp do dokumentów urzędowych stanowią wspólnie istotną gwarancję jasności i przejrzystości funkcjonowania samych struktur UE⁹. W Traktacie o funkcjonowaniu Unii Europejskiej w art. 15 ust. 1 zapisano, iż w celu wspierania dobrych rządów i zapewnienia uczestnictwa społeczeństwa obywatelskiego, instytucje, organy i jednostki organizacyjne UE działają z jak największym poszanowaniem zasady otwartości. W wykonaniu delegacji określonej w Traktacie przyjęto rozporządzenie 1049/2001/WE w sprawie publicznego dostępu do dokumentów Parlamentu Europejskiego, Rady i Komisji.

Przepisy unijne wprowadziły ponadto ramy prawne dla szerokiego dostępu do informacji i danych wytwarzanych przez sektor publiczny w krajach członkowskich oraz do ich ponownego wykorzystywania. Odpowiednie przepisy zostały zawarte w dyrektywie Parlamentu Europejskiego i Rady (UE) 2019/1024 z dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego wykorzystywania informacji sektora publicznego (dalej jako: dyrektywa *re-use*)¹⁰, która zastąpiła wcześniejsze unijne regulacje w tym

⁹ B. Opaliński, *Dostęp do informacji publicznej w prawie międzynarodowym i prawie Unii Europejskiej*, [w:] P. Szustakiewicz (red.), *Dostęp do informacji publicznej*, Warszawa 2014, s. 7.

¹⁰ Dz. Urz. UE L 172 z 26.06.2019 r., s. 56.

zakresie. W ramach realizacji unijnej polityki otwartych danych od 2015 r. działa także Europejski Portal Danych, ogólnoeuropejskie repozytorium informacji sektora publicznego, które można ponownie wykorzystać¹¹. Portal oferuje zbiory danych z Komisji Europejskiej, a także innych instytucji i agencji UE, gromadzi także metadane na temat informacji o sektorze publicznym dostępnych w witrynach z publicznymi danymi krajów członkowskich.

2. Scrapowanie a wykorzystywanie treści pobranych ze stron internetowych

Ocena prawna działań podmiotów scrapujących dane z internetu opiera się na dwóch płaszczyznach: przeszukiwania zasobów internetu i pobierania danych z cudzych stron internetowych oraz wykorzystywania pobranych danych, które to sfery w coraz większym stopniu się przenikają i uzupełniają. Problematyka korzystania z udostępnionych publicznie dóbr i treści od dawna była przedmiotem regulacji prawnych. Scrapowanie natomiast jest zagadnieniem stosunkowo nowym, które jednak już znajduje odzwierciedlenie w rosnącej liczbie regulacji.

W zależności od charakteru udostępnionych w internecie treści zakres ich ochrony i dopuszczalnego legalnego ich wykorzystywania regulują przepisy o prawie autorskim i prawach pokrewnych lub inne szczególne regulacje, np. dotyczące ochrony znaków towarowych.

Stosunkowo nowsze są przepisy dotyczące ograniczeń w korzystaniu z samych danych. Jeśli mają one charakter prostych informacji o faktach, same w sobie nie podlegają ochronie autorskoprawnej, nawet jeśli zostały wyodrębnione z chronionego utworu. Nie oznacza to jednak, że mogą być dowolnie wykorzystywane. Szczególną, rozszerzaną na przestrzeni ostatnich

¹¹ Europejski Portal Danych dostępny pod adresem: <https://www.europeandataportal.eu/pl>.

lat ochroną zostały objęte dane osobowe. Zakres ich wykorzystywania określony został w rozporządzeniu 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/WE¹², które jest aktem obowiązującym we wszystkich krajach unijnych, a zatem i w Polsce – bezpośrednio. Problematykę ochrony danych osobowych uzupełnia w Polsce ustawa z dnia 10 maja 2018 r. o ochronie danych osobowych (dalej jako: u.o.d.o.)¹³. Ponieważ za przetwarzanie danych osobowych uznaje się wszelkie czynności dokonywane na danych włącznie z ich zbieraniem, utrwalaniem i przechowywaniem, należy uznać, że ich scrapowanie musi znajdować legalną podstawę prawną w obowiązujących przepisach o ochronie danych osobowych. Jest to wymóg niezależny od regulacji omawianych w dalszej części opracowania.

Inne dane, które w oderwaniu od siebie nie byłyby chronione, mogą podlegać ochronie zestawione w ramach bazy danych. Strony internetowe zawierają często, oprócz innych elementów, również bazy danych (nieraz bardzo rozbudowane), wśród których można wymienić:

- 1) bazy danych, które nie podlegają ochronie prawnej (np. zawierające zestawienia wyników meczów piłkarskich);
- 2) bazy danych chronione na podstawie przepisów ustawy z dnia 27 lipca 2001 r. o ochronie baz danych (dalej jako: u.o.b.d.)¹⁴, (tzw. ochrona *sui generis*).

Przez bazę danych podlegającą ochronie *sui generis* (samoistnej) należy rozumieć: „zbiór danych lub jakichkolwiek innych materiałów i elementów zgromadzonych według określonej metody lub systematyki, indywidualnie dostępnych w jakikolwiek sposób, w tym środkami elektronicznymi, wymagający istotnego co do jakości lub ilości nakładu inwestycyjnego, w celu sporządzenia, weryfikacji lub prezentacji

¹² Dz. Urz. UE L 119 z 4.05.2016 r., s. 1.

¹³ T.j. Dz. U. z 2019 r. poz. 1781.

¹⁴ T.j. Dz. U. z 2021 r. poz. 386.

zawartości” (art. 2 ust. 1 pkt 1 u.o.b.d.). Przez wykorzystywanie bazy danych należy rozumieć zgodnie z art. 2 ust. 1 pkt 3 u.o.b.d. wszystkie postaci publicznego udostępniania jej zawartości. Dopuszczalność takich działań regulują przepisy art. 7 i 8 u.o.b.d.

- 3) bazy danych chronione przez ustawę z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (dalej jako: u.p.a.p.p.)¹⁵, przy czym aby baza mogła podlegać przepisom prawa autorskiego, musi posiadać cechy oryginalne i twórcze.

Powyżej nakreślono w skrócie złożoność problematyki wykorzystania danych przez inne podmioty niż ich pierwotny dysponent. Temat ten nie zostanie szerzej omówiony, jako że niniejsze opracowanie jest poświęcone zagadnieniu scrapowania, które zasadniczo służy późniejszemu wtórnemu wykorzystywaniu pozyskanych treści. Zauważyć natomiast należy, że sfery pobierania i korzystania z danych w coraz większym stopniu się przenikają. Tak jak już wskazano, na przykład dyrektywa o ochronie danych osobowych obejmuje swym zakresem każdą operację na danych, a zatem i ich ekstrakcję ze stron internetowych. Podobnie przepisy o ochronie praw autorskich i praw pokrewnych, w świetle których już sama eksploracja utworów może być rozumiana jako wkroczenie w monopol twórcy. Jednak ograniczenia na etapie scrapowania danych często mają swój specyficzny charakter i różnią się od tych obowiązujących na dalszych etapach korzystania z pozyskanych danych.

W kolejnych rozdziałach omówione zostały przepisy, które regulują tematykę pozyskiwania upublicznianych w internecie treści w drodze scrapowania. Wywód został uzupełniony odpowiednim orzecznictwem i wypowiedziami doktryny. Ogólne regulacje w omawianym zakresie znajdują zastosowanie również do informacji publicznych i informacji udostępnianych przez podmioty sektora publicznego do ponownego wykorzystywania. Szczególny jednak status powyższych treści prowadzi do

¹⁵ T.j. Dz. U. z 2021 r. poz. 1062.

niewielkiego odmiennego podejścia również w odniesieniu do ich scrapowania, co zostanie opisane w dalszej części opracowania.

3. Scrapowanie w świetle przepisów o ochronie baz danych

3.1. Przesłanki ochrony *sui generis* baz danych

Jak zostało wskazane wcześniej, najdalej idące ograniczenia w zakresie scrapowania zostały określone przez przepisy regulujące ochronę baz danych, dlatego zostaną tu one szerzej omówione. Prawo *sui generis* do bazy danych, wynikające z przepisów ustawy o ochronie baz danych (art. 6 w zw. z art. 2 ust. 1 pkt 1 u.o.b.d.), przysługuje jej producentowi, jeśli poniósł istotny nakład inwestycyjny na jej sporządzenie, weryfikację lub prezentację. Inwestycja taka może polegać na zaangażowaniu zasobów ludzkich, środków technicznych i finansowych. Powinna być też istotna z punktu widzenia ilościowego lub jakościowego. Ocena pod względem ilościowym odnosi się przy tym do środków dających się wyrazić liczbowo, a ocena jakościowa – do nakładów, których wysokości nie da się ustalić, takich jak wkład pracy intelektualnej lub wydatkowanie energii (wynika to z motywów 7, 39 i 40 dyrektywy o ochronie baz danych, która stanowi podstawę polskich przepisów i w świetle której muszą być one interpretowane).

Do pojęcia „inwestycji” ponoszonej przez producenta bazy danych odniósł się Trybunał Sprawiedliwości Unii Europejskiej¹⁶, uznając, że pojęcie to obejmuje środki poświęcone na poszukiwanie już istniejących materiałów i ich gromadzenie w bazie danych, nie obejmuje natomiast środków użytych w celu stworzenia materiałów składających się na zawartość bazy danych. W kontekście rozstrzyganej sprawy, która dotyczyła

¹⁶ Zob. wyrok TSUE z dnia 9 listopada 2004 r. w sprawie C-444/02 *Fixtures Marketing Ltd. przeciwko Organismos Prognostiken Agnon Podosfairou AE*.

terminarza spotkań w celu organizacji mistrzostw w piłce nożnej, Trybunał uznał, że pojęcie „inwestycji” nie obejmuje środków poświęconych na wyznaczenie dat, godzin i ustalenie doboru w pary drużyn odnoszących się do poszczególnych spotkań w ramach tych mistrzostw. W konsekwencji bazę danych z terminarzem spotkań piłkarskich TSUE uznał za niepodlegającą ochronie.

W Polsce oceny inwestycji w kontekście baz danych dokonał Sąd Apelacyjny w Szczecinie w wyroku z dnia 2 maja 2012 r. (I ACa 105/12)¹⁷, które to rozstrzygnięcie zostało podtrzymane w wyroku Sądu Najwyższego z dnia 9 maja 2013 r. (II CSK 466/12)¹⁸. W przywołanym wyroku Sąd Apelacyjny uznał, że właściciel portalu motoryzacyjnego nie poniósł istotnej ilościowo ani jakościowo inwestycji w pozyskiwanie danych do portalu ani w ich weryfikowanie, jako że sami użytkownicy zamieszczali ogłoszenia, które nie były poddawane procedurze weryfikacji. Odmienienie natomiast Sąd ocenił nakład poniesiony przez producenta bazy na stworzenie całego systemu ogłoszeń.

Jeżeli baza danych podlega ochronie na podstawie przepisów o bazach danych, nie tylko jej wykorzystywanie lub wykorzystywanie jej części podlega ograniczeniom, ale również pobieranie z niej danych. U.o.b.d. odnosi się bezpośrednio do czynności, które polegają na przeszukiwaniu baz danych i ekstrakcji jej treści.

3.2. Dopuszczalny prawnie zakres pobierania danych z chronionych baz danych

Pobieranie danych oznacza, zgodnie z art. 2 ust. 1 pkt 2 u.o.b.d., przejęcie lub przeniesienie ich na inny nośnik, przy czym przepis odnosi się zarówno do nieograniczonego w czasie przejęcia i przeniesienia, jak i do tego typu czynności dokonanych tylko na pewien czas. Pojęcie pobierania obejmuje swym zakresem każdy sposób i formę przejęcia lub

¹⁷ Zob. uzasadnienie wyroku opubl. w OSASz 2012/4/56-86.

¹⁸ Opubl. w OSNC-ZD 2014/3/46.

przeniesienia. Przeniesienie zakłada umieszczenie danych na nowym nośniku z jednoczesnym ich usunięciem z dotychczasowego nośnika, a przejście to umieszczenie na nowym nośniku bez usunięcia z poprzedniego¹⁹. Do pojęcia „pobierania” danych odniósł się też TSUE, uznając, że należy je rozumieć szeroko jako odnoszące się do każdej czynności dokonanej bez upoważnienia i polegającej na przywłaszczeniu całości lub części zawartości bazy danych, niezależnie od charakteru czy formy zastosowanej w tym kontekście metody²⁰. Za pobieranie należy uznać też przejście elementów chronionej bazy danych do innej bazy, które wskutek jego powtarzalnego i systematycznego charakteru mogło doprowadzić do przeniesienia istotnej części jej zawartości²¹. Należy uznać, że w ramach definicji „pobierania” mieści się scrapowanie, które w części polega na ekstrakcji danych ze stron internetowych za pomocą automatycznych narzędzi.

Przepisami, które bezpośrednio regulują zakres dozwolonego pobierania treści z chronionych prawnie baz danych, są art. 7 oraz art. 8 ust. 2 u.o.b.d. Zgodnie z art. 7 ust. 1 ustawy u.o.b.d. producent bazy danych udostępnionej publicznie nie może zabronić korzystającemu z niej użytkownikowi nie tylko jej wtórnego wykorzystania (o czym była już mowa), ale również pobierania w jakimkolwiek celu nieistotnej, co do jakości lub ilości, części jej zawartości. Przepis odpowiednio odnosi się do części bazy danych (art. 7 ust. 2 u.o.b.d.). Postanowienia umowne sprzeczne z przywołaną regulacją są nieważne (art. 7 ust. 4 u.o.b.d.). Zgodnie z art. 8 ust. 2 u.o.b.d. z kolei: „nie jest dozwolone powtarzające się i systematyczne pobieranie lub wtórne wykorzystanie sprzeczne z normalnym korzystaniem i powodujące nieusprawiedliwione naruszenie słuszych interesów producenta”.

¹⁹ Zob. uzasadnienie wyroku Sądu Apelacyjnego w Szczecinie z dnia 2 maja 2012 r., I ACa 105/12.

²⁰ Zob. wyrok TSUE z dnia 9 listopada 2004 r. w sprawie C-203/02 *The British Horseracing Board Limited przeciwko William Hill*, EU:C:2004:695.

²¹ Zob. wyrok TSUE z dnia 9 października 2008 r. w sprawie C-304/07 *Directmedia Publishing przeciwko Albert-Ludwigs-Universität Freiburg*, EU:C:2008:552.

Z przywołanych regulacji wynikają następujące wnioski:

- 1) zabronione jest pobieranie istotnej części bazy danych podlegającej ochronie *sui generis*;
- 2) producent takiej bazy danych nie może sprzeciwić się pobieraniu nieistotnej, co do jakości lub ilości, części jej zawartości;
- 3) jeżeli pobieranie nieistotnej części bazy danych ma charakter powtarzający się i systematyczny, a ponadto jest sprzeczne z normalnym korzystaniem z tej bazy danych i narusza w sposób nieuzasadniony słusne interesy producenta – jest sprzeczne z prawem.

Jeżeli baza danych nie jest pobierana w całości, lecz jedynie w części, aby ocenić legalność takiego działania, należy w każdym konkretnym przypadku przeanalizować wszystkie przesłanki określone w art. 8 ust. 2 u.o.b.d. oraz wziąć pod uwagę całość przepisów ustawy w kontekście regulacji dyrektywy o ochronie baz danych, co uczyniono poniżej.

Do określenia zakresu zastosowania przepisu art. 8 ust. 2 u.o.b.d. konieczne jest też wyjaśnienie znaczenia określeń: „powtarzający się” i „systematyczny” odnoszących się do zakazów pobierania objętych analizowaną regulacją. Przede wszystkim wskazać należy, że dla zastosowania regulacji wyżej wskazanego art. 8 ust. 2 niezbędne jest, by pobieranie lub wtórne wykorzystanie nieistotnej części bazy danych miało charakter zarówno powtarzający się, jak i systematyczny, a zatem przesłanki te muszą zostać spełnione łącznie²². Wobec braku definicji tych pojęć w zakresie wykładni literalnej stwierdzić można, zestawiając słownikowe znaczenie powołanych pojęć, że nie wystarczy, by pobieranie czy wtórne wykorzystywanie nastąpiło ponownie więcej niż jeden raz, ale także musi być metodyczne, regularne, ewentualnie również zachodzące od pewnego czasu²³.

²² R. Stötzel, S. Wille, *Der Urheberrechtsschutz sui generis für Datenbanken*, “Europäisches Wirtschafts- und Steuerrecht” 2004, H. 11, s. 493.

²³ S. Dubisz (red.), *Uniwersalny słownik języka polskiego*, Warszawa 2003, t. 3, s. 476 i 1480.

Systematyczne i powtarzające się pobieranie stanowi zasadniczo istotę scrapowania, zatem podmioty podejmujące tego typu czynności zawsze rozważyć muszą ich zgodność z regulacją art. 8 ust. 2 u.o.b.d.²⁴

Kolejną przesłanką, którą należy rozważyć przy ocenie legalności scrapowania chronionych baz danych, jest rozmiar i istotność pobranych elementów. Nie wynika ona bezpośrednio z art. 8 ust. 2 u.o.b.d., który stanowi w tym zakresie nieprawidłową implementację dyrektywy o ochronie baz danych, ale z całościowej analizy regulacji dokonanej w świetle celów dyrektywy. Dyrektywa w art. 7 ust. 5 stanowi, iż: „niedozwolone jest powtarzające się i systematyczne pobieranie danych i/lub wtórne ich wykorzystanie w nieistotnej części, jeśli czynności te są sprzeczne z normalnym korzystaniem z tej bazy danych lub w sposób nieuzasadniony naruszają słuszne interesy producenta bazy danych”.

Pojęcie „istotnej pod względem ilościowym” części zawartości bazy danych odnosi się do pobranej ilości danych i należy je oceniać w stosunku do całkowitej ilości danych zawartych w tej bazie. Pojęcie „istotnej pod względem jakościowym” części zawartości bazy danych z kolei odnosi się do rozmiaru inwestycji związanej z uzyskaniem, weryfikacją i prezentacją tej zawartości, która została pobrana lub wtórnie wykorzystana, niezależnie od tego, czy pobrana lub wtórnie wykorzystana zawartość bazy stanowi istotną pod względem ilościowym część całkowitej zawartości chronionej bazy danych. Pojęcie „nieistotnej” części bazy danych obejmuje każdą część, która ani pod względem ilościowym, ani jakościowym nie wchodzi w zakres pojęcia „istotnej” części²⁵.

²⁴ Za spełniające warunki „systematycznego” i „powtarzającego się” pobierania uznano między innymi działanie metapoziomowej wyszukiwarki umożliwiającej publicznie kręgowi odbiorców na wyszukiwanie w ramach całości lub istotnej części zawartości udostępnionej online bazy danych „w czasie rzeczywistym”, zob.: wyrok TSUE z dnia 19 grudnia 2013 r. w sprawie C-202/12 *Innoweb BV przeciwko Wegener ICT Media BV i Wegener Mediaventions BV*, EU:C:2013:850.

²⁵ Zob. wyrok TSUE z dnia 9 listopada 2004 r. w sprawie C-203/02 *The British Horseracing Board Limited przeciwko William Hill*, EU:C:2004:695.

W sprawie, która była podstawą powyższych wniosków TSUE, ostatecznie sąd w Wielkiej Brytanii uznał, że działanie pozwanego polegające na ściąganiu z internetowej bazy danych i wykorzystywania danych z serwisu powoda w postaci informacji o lekarzach, które pozwany zestawiał w swoim serwisie z informacjami zagregowanych z innych serwisów, było nielegalne. Pozwany korzystał z ok. 4300 rekordów z serwisu powoda, co stanowił 10-11% całości. Sąd jednak uznał, że – mimo że nie była to istotna ilościowa część pobranej bazy danych – była istotna jakościowo, a zatem podlegała ochronie.

W doktrynie podkreśla się, że badając „istotność” części bazy danych, należy brać pod uwagę wartość gospodarczą (handlową) tej części, jej rodzaj oraz zakres. Za istotne należy uważać te części, które same „jako takie” mogłyby stanowić przedmiot ochrony *sui generis*. Podobnie należy oceniać te części bazy danych, które faktycznie mogą zastąpić korzystanie z całości bazy danych z punktu widzenia potrzeb określonych użytkowników. W tym kontekście należy badać, czy przejęcie części zawartości bazy otwiera możliwość wytworzenia konkurencyjnego produktu, który byłby w stanie zastępować produkt pierwotny. Przy ustalaniu „istotnej części” bazy danych istotne jest także, czy eksploatacja bazy danych obejmuje jej najbardziej wartościowe elementy²⁶.

Nawet jeżeli pobieranie danych z cudzej bazy danych zostanie dokonane w „nieistotnej” części, może wciąż stanowić naruszenie przepisów prawa. Stanie się tak w sytuacji, gdy będzie sprzeczne z normalnym korzystaniem z tej bazy danych i w sposób nieuzasadniony naruszy słuszne interesy producenta. Należy zwrócić uwagę, że omawiana polska regulacja jest istotnie niezgodna z dyrektywą o ochronie baz danych w zakresie konstrukcji przesłanek zastosowania przepisu art. 8 ust. 2 u.o.b.d. Art. 7 ust. 5 dyrektywy o ochronie baz danych stanowi bowiem, iż aby uznać, że pobieranie danych jest nielegalne, wystarczy, że spełnia jeden

²⁶ Zob. J. Barta, R. Markiewicz, *Ustawa o ochronie baz danych. Komentarz*, Warszawa 2002, s. 117.

z warunków: jest sprzeczne z normalnym korzystaniem z bazy danych lub powoduje nieusprawiedliwione naruszenie słuszych interesów producenta. Polski ustawodawca zamiast łącznika „lub” posłużył się łącznikiem „i”, w związku z czym dla zaktualizowania się uprawnień ochronnych producenta bazy danych przesłanki sprzeczności z normalnym korzystaniem i naruszenia interesów producenta muszą być spełnione łącznie.

Pojęciem wymagającym wyjaśnienia w kontekście omawianego zagadnienia jest „normalne korzystanie”. W doktrynie przyjmuje się, że obejmuje ono czynności, których wykonywania w normalnych, zwykłych warunkach można się spodziewać, zwłaszcza w powiązaniu z poczynioną przez uprawnionego inwestycją i oczekiwaniem dochodu z niej uzyskiwanego. *A contrario*, naruszające normalne korzystanie będą zatem takie formy eksploatacji, które godzą w inwestycję producenta, ograniczając albo wykluczając realizację praw producenta bazy danych lub wykorzystanie przez niego lub z jego upoważnienia bazy danych w oczekiwanym w normalnych warunkach zakresie²⁷.

W rozstrzyganej w Polsce sprawie z tego zakresu Sąd Apelacyjny w Szczecinie w przywołanym już wcześniej wyroku I ACa 105/12 uznał, że usługi świadczone przez serwis internetowy polegające na „przeglądaniu” w wersji on-line ofert pod kątem zleceń klientów i udostępnianiu im oferty w wersji całościowej z pominięciem strony głównej serwisu, z którego oferta ta pochodziła, nie przekraczało granic normalnego korzystania. Normalne korzystanie z bazy danych powódki bowiem polegało nie na wejściu użytkownika na stronę główną serwisu, ale na dotarciu do konkretnej oferty, co ostatecznie następowało.

Wydaje się, że działaniem zakłócającym „normalne korzystanie” z bazy danych są takie działania automatów do przeszukiwania treści strony, które zakłócają jej normalne funkcjonowanie. Ideą działania

²⁷ Zob. E. Traple, [w:] J. Barta, R. Markiewicz (red.), *Prawo autorskie i prawa pokrewne. Komentarz*, Kraków 2005, s. 372-373; zob. także S. Ciupa, *Dozwolony użytek w ustawie o ochronie baz danych*, „Monitor Prawniczy” 2003, nr 12, s. 550.

botów jest jak najszybsze otwieranie kolejnych znalezionych linków, a to powoduje ryzyko wystąpienia różnego rodzaju problemów technicznych, szczególnie wzrostu obciążenia serwera.

W odniesieniu do pojęcia „słuszne interesy producenta bazy danych” w literaturze prezentowane jest stanowisko, że miarą, którą należy mierzyć ograniczenie dla słuszných interesów producenta bazy danych, są interesy użytkowników tej bazy. Równowaga, do której należy dążyć, ma zapewnić wyważenie z jednej strony ekonomicznych interesów producentów związanych z poczynioną inwestycją, a z drugiej strony – interesów użytkowników w zakresie dostępu do informacji inkorporowanych w bazach danych. Szeroka wykładnia „słuszných interesów producenta bazy danych” nie jest zatem dopuszczalna, prowadzi bowiem *de facto* przez ochronę prawem *sui generis* do monopolizacji informacji, tworząc samoistną podstawę ochrony treści, co wprost wykluczono w motywie 46 dyrektywy o ochronie baz danych²⁸.

W istotnym rozstrzygnięciu TSUE z tego zakresu z dnia 19 grudnia 2013 r. w sprawie C202/12 *Innoweb BV przeciwko Wegener ICT Media BV, Wegener Mediaventions BV* Trybunał uznał, iż serwis internetowy pobierający oferty sprzedaży używanych samochodów z różnych platform zajmujących się pośrednictwem przy ich sprzedaży narusza słuszne interesy producentów tych serwisów. Wynika to z faktu, że klient może dokonać zakupu, posiłkując się jedynie danymi z serwisu agregującego oferty bez konieczności wchodzenia do serwisu, który ofertę oryginalnie wystawił, może znaleźć samochód w konkurencyjnej cenie na innej platformie i skorzystać z tej oferty zamiast z oferty skarżącego producenta, co godzi w jego interesy. Ponadto właściciel oryginalnej witryny ponosi koszty jej utrzymania i aktualizowania ofert, a z uwagi na możliwe

²⁸ Por. E. Traple, [w:] J. Barta, R. Markiewicz (red.), *Prawo autorskie...*, s. 372-373; D. Lubasz, M. Namysłowska (red.), *Ustawa o ochronie baz danych. Komentarz*, Warszawa 2013, LEX.

zmniejszenie liczby użytkowników w wyniku działań pozwanego grozi mu utrata wpływów związanych z reklamami umieszczanymi na stronie.

Jak już wcześniej wspomniano, w związku z błędną implementacją dyrektywy o ochronie baz danych do polskiego porządku prawnego warunki sprzeczności z normalnym korzystaniem i ze słusznym interesem producenta muszą być spełnione łącznie. Takie rozumienie polskich przepisów uznał w przywoływanym już wcześniej wyroku I ACa 105/12 Sąd Apelacyjny w Szczecinie, co wydaje się stać w sprzeczności z prounijną wykładnią polskich przepisów prawa (w wyroku kasacyjnym do tego orzeczenia o sygn. II CSK 466/12 Sąd Najwyższy przywołał zarówno przepisy polskie, jak i unijne).

4. Scrapowanie w świetle prawa autorskiego; eksploracja tekstów i danych

Przepisy prawa autorskiego, które należy wziąć pod uwagę przy omawianiu zagadnienia legalności scrapowania, dotyczą ochrony baz danych i pól eksploatacji utworu. Niezwykle istotne regulacje zagadnienia eksploracji tekstów i danych (*text and data mining*) zostały zawarte w dyrektywie w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym, która nie została jeszcze zaimplementowana do polskiego porządku prawnego mimo upływu terminu na to wyznaczonego. Oznacza to, że obowiązujące obecnie w Polsce przepisy prawa powinny być interpretowane w zgodzie z wprowadzonymi przez dyrektywę zasadami mimo braku nowej krajowej regulacji prawnoautorskiej.

Ochronie prawnoautorskiej podlegają zarówno strony internetowe, jeśli mają oryginalny charakter, twórcze elementy stron, jak i bazy danych o cechach utworu, które są częściami takich stron. Zgodnie z art. 3 u.p.a.p.p.: „bazy danych spełniające cechy utworu są przedmiotem prawa autorskiego, nawet jeżeli zawierają niechronione materiały, o ile przyjęty w nich

dobór, układ lub zestawienie ma twórczy charakter, bez uszczerbku dla praw do wykorzystanych utworów”.

W obowiązującym stanie prawnym ochrona *sui generis* w odniesieniu do baz danych uznawana jest zasadniczo za silniejszą od ochrony wynikającej z prawa autorskiego. Jak wskazuje się w literaturze, ustawa o prawie autorskim zapewnia ochronę tylko w tym zakresie, w jakim jego utwór ma twórczy i indywidualny charakter, pozostałe składniki utworu nie są przedmiotem wyłączności i mogą być dowolnie wykorzystywane w innych wytworach niematerialnych (w tym bazach danych). Natomiast ochrona *sui generis* baz danych rozciąga się także na ich zawartość²⁹.

Podkreślić należy, że przedmiotem regulacji prawa autorskiego jest generalnie eksploatacja utworów. Przedmiotem dyskusji zatem było rozstrzygnięcie, czy same czynności, jakim poddawany jest utwór, a więc na przykład jego przeszukiwanie czy ekstrakcja czystych danych niemających cech twórczych z utworu bądź bazy danych o takich cechach, podlegają regulacjom prawnoautorskim, a jeżeli tak – to jak trzeba czynności takie kwalifikować. Należy tu przywołać stanowisko wskazujące, że o ile informacje i dane nie są objęte monopolem prawnoautorskim, to nie jest to równoznaczne ze stwierdzeniem, że żadna z czynności, którym jest poddawany utwór w celu uzyskania informacji i danych, nie wkracza w monopol uprawnionego. Wskazano przy tym, że niezależnie od przyjętego modelu teoretycznego faz analizy danych co najmniej część takich działań jest nierozdzielnie powiązana z reprodukcją co najmniej kilku fragmentów utworów. Co więcej, czynności takie mogą być powtarzane w kolejnych etapach eksploracji danych³⁰. Prowadzi to do uznania, że w określonych okolicznościach podmiot dokonujący czynności scrapowania treści objętych ochroną prawnoautorską samą tą czynnością może

²⁹ Por. J. Barta, R. Markiewicz, *Prawo autorskie. Komentarz*, Warszawa 2013, s. 315.

³⁰ K. Gienas, *Data mining jako problem prawa autorskiego*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2017, z. 1, s. 26.

naruszyć monopol twórcy. Stanowisko takie legło u podstaw skonstruowania przepisów regulujących eksplorację tekstów i danych.

Regulacja zagadnienia eksploracji tekstów i danych (*text and data mining*) jest nową konstrukcją prawną przyjętą w dyrektywie w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym. Odnosi się do wszystkich efektów twórczej działalności chronionych na podstawie przepisów o prawie autorskim, ale swoim zakresem wykracza poza regulacje *stricte* prawnoautorskie. Ponieważ eksploracja tekstów i danych dotyczy wprost zagadnienia scrapowania, omówienie tego zagadnienia jest istotne dla niniejszej pracy.

Pojęcie „eksploracji tekstów i danych” zostało wyjaśnione w motywie 8 preambuły dyrektywy, zgodnie z którym: „Nowe technologie umożliwiają automatyczną matematyczną analizę informacji w postaci cyfrowej, takich jak teksty, dźwięki, obrazy lub dane, powszechnie określaną jako eksploracja tekstów i danych. Eksploracja tekstów i danych umożliwia przetwarzanie dużych ilości informacji w celu uzyskania nowej wiedzy i odkrycia nowych tendencji [...]. W niektórych przypadkach eksploracja tekstów i danych może obejmować czynności chronione prawem autorskim lub prawem do bazy danych *sui generis*, w szczególności zwielokrotnianie utworów lub innych przedmiotów objętych ochroną lub pobieranie treści z bazy danych, co ma miejsce na przykład w przypadku normalizacji danych podczas eksploracji tekstów i danych. Jeżeli nie znajduje zastosowanie wyjątek lub ograniczenie, konieczne jest uzyskanie zezwolenia na dokonanie takich czynności od podmiotów uprawnionych”.

Jak wynika z powyższego, zagadnienie eksploracji tekstów i danych ma tak szeroki charakter, że wykracza poza ramy regulacji przedmiotu ochrony prawa autorskiego i wkracza nawet w zakres ochrony *sui generis* baz danych. Implementacja dyrektywy oznaczać zatem będzie zmiany w zakresie dotychczasowych zasad scrapowania na różnych poziomach.

W literaturze przyjmuje się, że nowy sposób korzystania stanowi nowe pole eksploatacji³¹ i jako taki wymaga odrębnej zgody podmiotu uprawnionego. W motywie 9 wyjaśniono jednocześnie, iż eksploracja tekstów i danych może być również dokonywana w odniesieniu do samych faktów lub danych, które nie są chronione prawem autorskim, i w takich przypadkach nie jest wymagane zezwolenie na podstawie prawa autorskiego, podobnie jak w odniesieniu do eksploracji tekstów i danych, w których nie dokonuje się czynności zwielokrotniania.

Dyrektywa wprowadza kilka szczegółowych wyjątków w zakresie eksploracji tekstów i danych odnoszących się do eksploracji na potrzeby badań naukowych, dydaktyki czy w celu zachowania dziedzictwa kulturowego. Ponadto na mocy art. 4 nakłada na państwa członkowskie obowiązek wprowadzenia wyjątku lub ograniczenia odnoszącego się generalnie do eksploracji tekstów i danych, bez wskazania kontekstu. W takim przypadku wyjątek lub ograniczenie mają zastosowanie pod warunkiem, że korzystanie z utworów i innych przedmiotów objętych ochroną nie zostało wyraźnie zastrzeżone w odpowiedni sposób przez podmioty uprawnione. Przepis ten oznacza zmianę w stosunku do obecnie obowiązujących regulacji, jako że dotyczy nie tylko eksploracji treści poddanych ochronie prawnoautorskiej, ale również przewiduje, że wyjątek ma być wprowadzony w odniesieniu do art. 5 lit. a) i art. 7 ust. 1 dyrektywy o ochronie baz danych. Jego potencjał został znacznie ograniczony zastrzeżeniem, że podmiot uprawniony będzie mógł zastrzec brak zgody na eksplorację tekstów i danych. W odniesieniu do przedsiębiorców zarządzających serwisami internetowymi może to zatem w praktyce nie mieć większego znaczenia. W odniesieniu jednak do informacji i danych publicznych może oznaczać ułatwienia w dostępie.

³¹ K. Gliściński, *Komentarz do art. 50 ustawy o prawie autorskim i prawach pokrewnych*, [w:] A. Michalak (red.), *Ustawa o prawie autorskim i prawach pokrewnych. Komentarz*, Warszawa 2019, s. 390.

5. Scrapowanie informacji publicznych oraz informacji i danych sektora publicznego

5.1. Prawo do informacji publicznej

W Polsce prawo obywatela do uzyskiwania informacji od organów władzy publicznej i osób pełniących funkcje publiczne jest prawem konstytucyjnym. Art. 61 Konstytucji określa je szeroko, obejmując jego zakresem również uzyskiwanie informacji o działalności organów samorządu gospodarczego i zawodowego, a także innych osób oraz jednostek organizacyjnych w zakresie, w jakim wykonują one zadania władzy publicznej i gospodarują mieniem komunalnym lub mieniem Skarbu Państwa.

Ustawa z dnia 6 września 2001 r. o dostępie do informacji publicznej (dalej jako: u.d.i.p.)³² jako informację publiczną definiuje każdą informację o sprawach publicznych (art. 1 ust. 1 u.d.i.p.), a art. 6 ustawy zawiera otwarty katalog informacji podlegających udostępnieniu. Podmioty zobowiązane do udostępniania informacji publicznych zostały wymienione w art. 4 ustawy, przy czym katalog ten nie jest zamknięty:

„Art. 4. 1. Obowiązane do udostępniania informacji publicznej są władze publiczne oraz inne podmioty wykonujące zadania publiczne, w szczególności:

- 1) organy władzy publicznej;
- 2) organy samorządów gospodarczych i zawodowych;
- 3) podmioty reprezentujące zgodnie z odrębnymi przepisami Skarb Państwa;
- 4) podmioty reprezentujące państwowe osoby prawne albo osoby prawne samorządu terytorialnego oraz podmioty reprezentujące inne państwowe jednostki organizacyjne albo jednostki organizacyjne samorządu terytorialnego;

³² T.j. Dz. U. z 2020 r. poz. 2176 ze zm.

5) podmioty reprezentujące inne osoby lub jednostki organizacyjne, które wykonują zadania publiczne lub dysponują majątkiem publicznym, oraz osoby prawne, w których Skarb Państwa, jednostki samorządu terytorialnego lub samorządu gospodarczego albo zawodowego mają pozycję dominującą w rozumieniu przepisów o ochronie konkurencji i konsumentów.

2. Obowiązane do udostępnienia informacji publicznej są organizacje związkowe i pracodawców, reprezentatywne w rozumieniu ustawy z dnia 24 lipca 2015 r. o Radzie Dialogu Społecznego i innych instytucjach dialogu społecznego (Dz. U. z 2018 r. poz. 2232), oraz partie polityczne.

3. Obowiązane do udostępniania informacji publicznej są podmioty, o których mowa w ust. 1 i 2, będące w posiadaniu takich informacji”.

W judykaturze i doktrynie utrwalił się pogląd, że informacją publiczną będzie każda wiadomość wytworzona lub odnosząca się do władz publicznych, a także odnosząca się do innych podmiotów wykonujących funkcje publiczne w zakresie wykonywania przez nie zadań w zakresie władzy publicznej³³.

O uznaniu, że określona informacja jest informacją publiczną, decydują więc dwa elementy łącznie:

- 1) podmiotowy: dany podmiot zaliczony jest do zbioru władz publicznych lub w imieniu państwa wykonuje zadania publiczne;
- 2) przedmiotowy: informacja dotyczy spraw związanych z wykonywaniem przez ten podmiot zadań publicznych.

5.2. Scrapowanie stron internetowych podmiotów zobowiązanych do udostępniania informacji publicznych

Informacje publiczne mogą być udostępniane bądź przez internet (w ramach Biuletynu Informacji Publicznej prowadzonego przez podmioty

³³ Zob. wyrok NSA z dnia 27 czerwca 2013 r., I OSK 513/13, CBOSA; podobnie wyrok NSA z dnia 10 stycznia 2014 r., I OSK 1966/13, CBOSA.

zobowiązane do jej udostępniania, ale również poza nim), bądź na wniosek. W niniejszym opracowaniu ze względu na charakter scrapowania skoncentrowano się jedynie na trybie bezwnioskowym.

Na stronach internetowych podmiotów zobowiązanych do udostępniania informacji publicznej możemy znaleźć treści, które stanowią taką informację, i takie, które nie mają charakteru informacji publicznej. Jest to szczególnie istotne w odniesieniu do podmiotów, które nie są instytucjami publicznymi, takich jak spółki z udziałem Skarbu Państwa, jednostki organizacyjne wykonujące zadania publiczne czy organy samorządu zawodowego lub gospodarczego. Zasady scrapowania treści niemających charakteru informacji publicznej w niczym nie odbiegają od omówionych we wcześniejszej części opracowania. Typ podmiotu, który je udostępnia w internecie, nie ma przy tym znaczenia. Co do zasady również w odniesieniu do treści, które mają charakter informacji publicznych, znajdują zastosowanie regulacje ogólne. Ustawodawca nie wprowadził wyjątku w odniesieniu do pobierania danych udostępnianych przez podmioty zobowiązane do tego na mocy u.d.i.p. Niemniej szczególny charakter takich treści prowadzi do nieco odmiennej interpretacji ogólnych przepisów dotyczących ich automatycznego pobierania z internetu.

Jak wskazano, omawiając przepisy dotyczące baz danych chronionych prawem *sui generis*, producentowi bazy przysługuje wyłączne prawo do wtórnego pobierania lub wykorzystywania jej istotnej części. Producent może jednak udzielić zgody na takie działania innym podmiotom. Niedozwolone jest ponadto powtarzające się i systematyczne pobieranie nieistotnej części bazy danych producenta przez podmiot trzeci. Aby podejmowane przez użytkowników czynności w odniesieniu do baz danych mogły być uznane za niezgodne z przepisami, muszą być sprzeczne z normalnym korzystaniem i powodować nieusprawiedliwione naruszenie interesów producenta.

Jeżeli użytkownik dokonuje scrapowania strony internetowej podmiotu zobowiązanego do udostępniania informacji publicznej, jego

działania podlegają takiej samej ocenie jak w odniesieniu do każdego innego rodzaju witryny internetowej. Jeżeli zatem jego aktywność dotyczy korzystania z chronionej bazy danych w dozwolonym zakresie, nie może naruszać normalnego z niej korzystania. Za naruszające „normalne korzystanie” uznane będą takie formy eksploatacji, które godzą w inwestycję producenta, ograniczając albo wykluczając realizację jego praw. Powinno się do nich zaliczać również te, które zakłócają korzystanie ze strony internetowej. Podmiot dokonujący scrapowania może też naruszyć „słuszne interesy producenta” w różny sposób, ale w szczególności jeżeli jego działania naruszają ekonomiczne interesy producenta związane z poczynioną inwestycją.

Gdy scrapowanie dotyczy informacji publicznych udostępnionych przez podmioty, które są organami państwowymi albo samorządowymi, ale również państwowych czy samorządowych osób prawnych lub jednostek organizacyjnych czy też podmiotów reprezentujących Skarb Państwa, ryzyko naruszenia normalnego korzystania z udostępnionej przez nie bazy danych czy słusznych interesów takich podmiotów jest ograniczone. Wynika to z faktu, że w pełni państwowe czy samorządowe podmioty z założenia nie udostępniają informacji publicznej w ramach komercyjnych serwisów, a ich celem nie jest osiąganie zysków ani pozyskiwanie jak największej liczby odbiorców treści o charakterze informacji publicznych. Działalność tych podmiotów nie ma na celu osiągnięcia zysku, a zatem nie można zagrozić ich ekonomicznym interesom. Natomiast już aktywność programów komputerowych przeszukujących zasoby stron takich podmiotów i zakłócających ich działanie może zostać uznana za sprzeczne z normalnym z niej korzystaniem.

W odniesieniu do omawianej kategorii podmiotów należy również zwrócić uwagę na przepisy dyrektywy w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym. Nakazuje ona państwom członkowskim, o czym była mowa wcześniej, wprowadzenie wyjątku w zakresie eksploracji tekstów i danych. Określa jednocześnie prawo

dysponentów praw autorskich i praw pokrewnych oraz producentów baz danych do zastrzeżenia niedopuszczalności jej dokonywania. Należy jednak uznać, że podmioty, o których mowa w niniejszej części opracowania, są tymi, które zastrzeżeń w zakresie eksploracji tekstów i danych nie powinny zgłaszać w odniesieniu do informacji, co do których prawami dysponują. Wynika to zarówno z charakteru tych podmiotów, jak i ze specyfiki udostępnianych przez nie treści.

Odmiennej charakter mają inne niż wymienione wyżej podmioty zobowiązane do udostępniania informacji publicznej, np. spółki z udziałem Skarbu Państwa. Ich podstawowa działalność może mieć komercyjny charakter, mogą również tworzyć bazy danych, które wymagają znacznych inwestycji niepublicznych środków, ich celem może być osiągnięcie jak największej liczby odwiedzin na swojej stronie. Wszystkie te interesy mogą być chronione przed nielegalnym scrapowaniem na zasadach ogólnych.

5.3. Ponowne wykorzystywanie informacji sektora publicznego i otwarte dane publiczne

Od prawa dostępu do informacji publicznej należy odróżnić prawo do ponownego wykorzystywania informacji sektora publicznego (*re-use*), które znalazło wyraz w odrębnej regulacji, a mianowicie w przepisach ustawy z dnia 11 sierpnia 2021 r. o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego (dalej jako: u.o.d.p.w.i.s.p)³⁴. Ustawa stanowi implementację dyrektywy 2019/1024/UE i zastępuje obowiązującą uprzednio ustawę z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego³⁵.

Prawo do informacji publicznej i prawo do ponownego wykorzystywania informacji sektora publicznego mają odmienny charakter i służą realizacji innych celów. Pierwsze z nich, określone w ustawie o dostępie

³⁴ Dz. U. z 2021 r. poz. 1641, dalej jako u.o.d.p.w.i.s.p

³⁵ T.j. Dz. U. z 2019 r. poz. 1446.

do informacji publicznej, stanowi implementację zasady jawności wyrażonej w art. 61 Konstytucji i realizuje prawo do informacji, które uważane jest za jedno z podstawowych praw człowieka. Prawo do ponownego wykorzystywania informacji sektora publicznego natomiast ma przede wszystkim wymiar ekonomiczny i pozwala podmiotom zainteresowanym osiągać korzyści związane z możliwością dalszego wykorzystywania (przetwarzania) udostępnionej informacji. Celem zasadniczym przyjęcia regulacji w zakresie udostępniania informacji sektora publicznego do ponownego wykorzystywania było zwiększenie podaży otwartych danych, w tym zasobów informacji o istotnej wartości i potencjale do tworzenia nowych usług i produktów, a także stworzenie optymalnego otoczenia regulacyjnego dla efektywnego wykorzystywania informacji sektora publicznego w Polsce³⁶.

Zdefiniowane w art. 2 pkt 8 u.o.d.p.w.i.s.p pojęcie informacji sektora publicznego jest szersze niż pojęcie informacji publicznej. Przez informację sektora publicznego należy rozumieć bowiem „każdą treść lub jej część, niezależnie od sposobu utrwalenia, w szczególności w postaci papierowej, elektronicznej, dźwiękowej, wizualnej lub audiowizualnej, będącą w posiadaniu podmiotu zobowiązanego”.

Za podmioty zobowiązane do udostępniania lub przekazywania informacji sektora publicznego w celu ponownego wykorzystywania zostały uznane, zgodnie z art. 3 u.o.d.p.w.i.s.p, analogiczne jednostki jak w ustawie z 2016 r., a mianowicie:

- 1) jednostki sektora finansów publicznych w rozumieniu przepisów ustawy z dnia 27 sierpnia 2009 r. o finansach publicznych;
- 2) inne państwowe jednostki organizacyjne nieposiadające osobowości prawnej niebędące jednostkami sektora finansów publicznych;

³⁶ Zob. *Uzasadnienie do projektu ustawy o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego*, Druk nr 1338, Internetowy System Aktów Prawnych [dostęp: 24.09.2021].

3) osoby prawne niebędące jednostkami sektora finansów publicznych, utworzone w szczególnym celu zaspokajania potrzeb o charakterze powszechnym, niemających charakteru przemysłowego ani handlowego, jeżeli jednostki o których mowa w tym przepisie, jednostki sektora finansów publicznych oraz inne państwowe jednostki organizacyjne nieposiadające osobowości prawnej, pojedynczo lub wspólnie, bezpośrednio albo pośrednio przez inny podmiot: finansują je w ponad 50% lub posiadają ponad połowę udziałów albo akcji, lub sprawują nadzór nad organem zarządzającym, lub mają prawo do powoływania ponad połowy składu organu nadzorczego lub zarządzającego;

4) związki podmiotów, o których mowa wyżej;

oraz dotatkowo podmioty mieszczące się zakresie definicji „przedsiębiorstwa publicznego” w rozumieniu art. 2 pkt 3 dyrektywy 2019/1024/UE.

Podmioty zobowiązane w celu ponownego wykorzystania udostępniają informacje sektora publicznego:

- 1) w Biuletynie Informacji Publicznej (dotyczy wyłącznie informacji publicznej);
- 2) w Portalu Danych: danepubliczne.gov.pl;
- 3) w inny sposób (np. w innym systemie teleinformatycznym);
- 4) przekazują je również na wniosek (przy czym ten sposób nie ma znaczenia w kontekście zagadnienia będącego przedmiotem niniejszego opracowania).

Ustawa przewiduje podstawę prawną do dobrowolnego udostępniania informacji sektora publicznego za pośrednictwem prowadzonego przez ministra właściwego do spraw informatyzacji portalu dane.gov.pl. Co więcej, polska ustawa w tym zakresie wykracza poza ramy wyznaczone przez dyrektywę 2019/1024/UE, wprowadzając możliwość także dla podmiotów sektora prywatnego dobrowolnego udostępniania w portalu dane.gov.pl zasobów informacji będących w ich posiadaniu. Portal ten w założeniu zatem powinien więc być źródłem bogatego zasobu danych,

z których wszystkie zainteresowane osoby czy podmioty będą mogły wtórnie korzystać.

U.o.d.p.w.i.s.p kompleksowo realizuje zamysł jak najszerszego otwarcia publicznych zasobów. Jednym z elementów jego realizacji jest wprowadzenie nowych, istotnych rozwiązań wynikających z wymogów dyrektywy 2019/1024/UE. W tym celu do ustawy wprowadzono kategorie danych o wysokiej wartości i danych dynamicznych oraz objęto jej zakresem dane badawcze.

Dane o wysokiej wartości zdefiniowane zostały w art. 2 pkt 4 u.o.d.p.w.i.s.p jako „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 wykorzystywaniu tych danych”. Podmioty zobowiązane mają obowiązek udostępniania takich danych do ponownego wykorzystywania co do zasady bezpłatnie, w formie nadającym się do odczytu maszynowego, za pośrednictwem interfejsu programistycznego aplikacji (API), oraz, w odpowiednich przypadkach, do zbiorczego pobrania.

Dane dynamiczne zostały zdefiniowane w art. 4 pkt 3 u.o.d.p.w.i.s.p jako „informacje sektora publicznego utrwalone w postaci elektronicznej podlegające częstym aktualizacjom lub aktualizacjom w czasie rzeczywistym, w tym ze względu na ich zmienność lub szybką dezaktualizację, w szczególności dane wygenerowane przez czujniki”. Podmioty zobowiązane są zobligowane do ich udostępniania do ponownego wykorzystywania natychmiast po ich zgromadzeniu, za pośrednictwem interfejsu programistycznego aplikacji (API), oraz, w odpowiednich przypadkach, do zbiorczego pobrania.

Ustawa objęła swym zakresem również dane badawcze, które zostały zdefiniowane w art. 2 pkt 2 u.o.d.p.w.i.s.p jako „informacje sektora publicznego utrwalone w postaci elektronicznej, inne niż publikacje naukowe, które zostały wytworzone lub zgromadzone w ramach działalności

naukowej i są wykorzystywane jako dowody w procesie badawczym lub służą do weryfikacji poprawności ustaleń i wyników badań”.

W ramach kompleksowego podejścia do otwierania danych w przepisie art. 10 ust. 1 u.o.d.p.w.i.s.p sformułowano zasadę, zgodnie z którą podmioty zobowiązane, o ile to możliwe, powinny udostępniać lub przekazywać informacje sektora publicznego w celu ponownego wykorzystywania, jako otwarte dane.

Samego pojęcia danych ustawa nie definiuje, jednak, w jaki sposób je rozumieć, wskazuje rządowy dokument wprowadzający Program Otwierania Danych. Zgodnie z jego brzmieniem pod pojęciem danych publicznych należy rozumieć „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”³⁷.

U.o.d.p.w.i.s.p definiuje natomiast pojęcie otwartych danych, które zgodnie z jej art. 2 ust. 11 są informacjami sektora publicznego udostępnianymi lub przekazywanymi w postaci elektronicznej, bezwarunkowo lub z uwzględnieniem warunków i które to dane są: kompletne, aktualne, w wersji źródłowej, w otwartym i niezastrzeżonym formacie przeznaczonym do odczytu maszynowego oraz są przeznaczone do bezpłatnego ponownego wykorzystywania na tych samych zasadach dla każdego użytkownika, bez konieczności potwierdzania przez niego tożsamości.

Zgodnie z wytycznymi UE Polska prowadzi program otwierania danych publicznych. Jego założenia zostały określone w Programie Otwierania Danych Publicznych na lata 2021-2027. Otwarte dane w rozumieniu ustawowym powinny spełniać w filary otwartości zdefiniowane w tym Programie. Otwarte dane powinny być również zgodne ze Standardami Otwartości Danych. Zostały one przygotowane przez Ministerstwo

³⁷ Zob. uchwała Rady Ministrów nr 28 z dnia 18 lutego 2021 r. w sprawie Programu Otwierania Danych Publicznych na lata 2021-2027, M. P. z 2021 r. poz. 290, s. 6.

Cyfryzacji i obejmują wytyczne dotyczące: regulacji prawnych, zapewnienia bezpieczeństwa i ochrony prywatności danych, technicznych aspektów ich udostępniania oraz dostępu do baz danych przez API (interfejs programistyczny aplikacji). W 2020 r. opublikowana została nowa wersja dokumentów, zmienionych w trakcie wcześniejszych konsultacji³⁸.

W u.o.d.p.w.i.s.p wprowadzono ograniczenia dotyczące zakresu jej obowiązywania, jak również zasady, tryb oraz warunki udostępniania i ponownego wykorzystywania informacji sektora publicznego. Ograniczenia te, jako mające istotny wpływ na dopuszczalność dokonywania scrapowania danych publicznych, zostaną omówione w dalszej części opracowania.

5.4. Scrapowanie informacji i danych sektora publicznego podlegających ponownemu wykorzystywaniu

Omówione wyżej regulacje opierają się na założeniach jak najszerzego otwarcia danych sektora publicznego i umożliwienia wykorzystywania ich przez wszystkie zainteresowane osoby i podmioty. Służyć to ma wspieraniu innowacyjności europejskich gospodarek. W tym zatem kontekście należy interpretować regulacje polskiej u.o.d.p.w.i.s.p. Zasady te znalazły też swój wyraz w Planie Otwierania Danych na lata 2021-2027 i w Standardach Otwartości Danych, które wśród filarów otwartości wymieniają „dostępność” danych, rozumiejąc przez to wprost możliwość ich pobierania, indeksowania i przeszukiwania zasobów. Jak najszerzej dostępności danych służą też regulacje nakładające na podmioty udostępniające niektóre rodzaje informacji do ponownego wykorzystywania stosowanie formatów nadających się do maszynowego odczytu czy udostępniania danych do zbiorczego pobrania. Działania takie stanowią wprost przygotowanie do tego, aby aplikacje komputerowe mogły łatwo zidentyfikować, rozpoznać i uzyskać określone dane, w tym poszczególne

³⁸ Standardy Otwartości Danych dostępne pod adresem: <https://dane.gov.pl/pl/knowledgebase/useful-materials/standardy-otwartosci-danych> [dostęp: 23.09.2021].

stwierdzenia faktów, i ich wewnętrzną strukturę³⁹. U.o.d.p.w.i.s.p nie wprowadza jednak wyraźnego wyjątku w zakresie prawa do scrapowania.

Prawo do ponownego wykorzystywania podlega ograniczeniom, które określone zostały w art. 6 u.o.d.p.w.i.s.p. Dla niniejszego opracowania szczególne znaczenie ma przepis art. 6 ust. 4 pkt 4 tej ustawy, zgodnie z którym prawo do ponownego wykorzystywania podlega ograniczeniu w zakresie informacji sektora publicznego, do których prawa autorskie i prawa pokrewne oraz prawa do baz danych przysługują podmiotom innym niż podmioty zobowiązane. Przywołany art. 6 wprowadza jednak i inne znaczące ograniczenia w zakresie możliwości ponownego wykorzystywania informacji sektora publicznego, np. w odniesieniu do informacji, których wytwarzanie przez podmioty zobowiązane nie należy do zakresu ich zadań publicznych określonych prawem czy w odniesieniu do części materiałów archiwalnych. Tym samym do dopuszczalności scrapowania takich informacji zastosowanie znajdą zasady szczegółowo omówione w poprzednich rozdziałach.

W odniesieniu natomiast do informacji sektora publicznego nie podlegających ograniczeniom, o których mowa w poprzednim akapicie, zastosowanie znajdzie zasada określona w art. 14 ust. u.o.d.p.w.i.s.p, zgodnie z którą informacje sektora publicznego udostępnia się lub przekazuje w celu ponownego wykorzystywania bezwarunkowo, z wyjątkiem przypadków określonych w ustawie.

Podmiot zobowiązany określa, na podstawie art. 14 ust. 2 u.o.d.p.w.i.s.p, warunki ponownego wykorzystywania informacji sektora publicznego mających cechy utworu lub przedmiotu praw pokrewnych lub stanowiących bazę danych w rozumieniu przepisów ustawy o ochronie baz danych, do których przysługują mu prawa. W art. 15 u.o.d.p.w.i.s.p określono z kolei katalog fakultatywnych warunków ponownego wykorzystywania informacji sektora publicznego, jakie podmiot zobowiązany może ustanowić,

³⁹ Zob. Pogram Otwierania Danych Publicznych..., s. 6.

udostępniając lub przekazując informacje sektora publicznego do ponownego wykorzystywania, przy czym podmiot zobowiązany może jedynie zdecydować o tym, czy zastosuje dodatkowe warunki, sama ich treść natomiast ustalona jest przez ustawodawcę⁴⁰. W odniesieniu do udostępniania szczególnego rodzaju danych, jakimi są dane badawcze, podstawę prawną dla określenia warunków ich wykorzystywania stanowi art. 22 u.o.d.p.w.i.s.p.

Nie ma podstaw prawnych do wprowadzania innych niż opisane powyżej ograniczeń w odniesieniu do informacji sektora publicznego udostępnianej w celu ponownego jej wykorzystywania.

Czynności, które składają się na scrapowanie danych w postaci eksploracji danych, ich pobierania czy agregowania, należy uznać za mieszczące się w ramach pojęcia ponownego wykorzystywania informacji sektora publicznego, które zostało zdefiniowane w art. 2 pkt 12 u.o.d.p.w.i.s.p. jako „wykorzystywanie przez użytkowników informacji sektora publicznego w jakimkolwiek celu”. Tym samym w określonym zakresie podmiot zobowiązany może wprowadzić pewne ograniczenia również na tym etapie korzystania przez użytkownika z informacji.

Dla pełnego obrazu sytuacji prawnej podmiotu zobowiązanego do udostępniania informacji sektora publicznego do ponownego wykorzystywania należy jeszcze przywołać przepisy omawianej już wcześniej dyrektywy 2019/790/UE w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym, która nakłada na państwa członkowskie obowiązek wprowadzenia ograniczeń uprawnień dysponentów praw autorskich i praw pokrewnych oraz producentów chronionych baz danych w zakresie eksploracji tekstów i danych (jeśli nie zastrzegą inaczej).

⁴⁰ W omawianym zakresie art. 15 ustawy jest co do zasady tożsamy z art. 14 ustawy z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego, a zatem odpowiednie zastosowanie znajdują uwagi zawarte w: B. Fischer, A. Piskorz-Ryń, M. Sakowska-Baryła, J. Wyporska-Frankiewicz, *Ustawa o ponownym wykorzystywaniu informacji sektora publicznego. Komentarz*, wyd. 2, Warszawa 2019, s. 307.

Z powyższego wynika wniosek, że podmiot zobowiązany w rozumieniu u.o.d.p.w.i.s.p powinien co do zasady udostępniać dane będące w jego posiadaniu w sposób otwarty i bezwarunkowy. Nie oznacza to jednak możliwości w pełni swobodnego scrapowania takich danych w każdym przypadku. Podmiot zobowiązany może bowiem określić pewne ograniczenia w tym zakresie, jeśli są one uzasadnione na przykład z uwagi na ochronę baz danych lub gdy udostępniane informacje albo bazy danych są objęte prawem autorskim, a on jest ich dysponentem. Należy uznać jednak, że ograniczenia w zakresie scrapowania danych, jeśli w ogóle zostaną przez podmiot zobowiązany określone, powinny mieć wyjątkowy i wąski charakter (np. dotyczyć tylko zapewnienia sprawnego działania strony internetowej). Wynika to zarówno z celu, jakiemu mają służyć przepisy o otwieraniu danych i umożliwieniu ich ponownego wykorzystywania, jak i z zasady jak najszerzego udostępniania danych (również na etapie przeszukiwania zasobów podmiotów sektora publicznego, pobierania czy indeksowania danych) oraz z przepisów, które wprost regulują kwestię udostępniania części danych do maszynowego odczytu czy zbiorczego pobierania. Jeżeli warunki w odniesieniu do scrapowania nie zostały przez podmiot zobowiązany określone, należy uznać, że jest ono prawnie dopuszczalne.

Zakończenie

Narzędziem, które służy do pozyskiwania danych z internetu, jest scrapowanie. Regulacje dotyczące tego zagadnienia są stosunkowo nowe, wywodzone z różnych aktów prawnych i niejednoznaczne. Sytuacja ta stopniowo ulega zmianie. Przepisy o eksploracji tekstów i danych, przyjęte w niewdrożonej jeszcze do polskiego systemu prawnego dyrektywie w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym, wspólnie z obowiązującymi

przepisami o ochronie baz danych stanowią ramy dla scrapowania danych z internetu. Dla omawianego zagadnienia mają również znaczenie regulacje dotyczące poszczególnych kategorii scrapowanych danych, takie jak przepisy prawnoautorskie czy przepisy o ochronie danych osobowych.

W odniesieniu do części podmiotów obowiązanych do udostępniania informacji publicznych, w szczególności takich jak organy państwowe i samorządowe, państwowe czy samorządowe osoby prawne lub jednostki organizacyjne czy podmioty reprezentujące Skarb Państwa, zasadą powinno być niewprowadzanie przez nie ograniczeń w scrapowaniu informacji publicznych, chyba że podmioty te nie dysponują pełnią praw do udostępnianej informacji bądź gdy scrapowanie powoduje ryzyko naruszenia normalnego korzystania z udostępnianych przez nie baz danych.

Podobnie oceniać również należy scrapowanie informacji sektora publicznego, o ile prawami do nich dysponują podmioty zobowiązane do ich udostępnienia w celu ponownego wykorzystywania. W tym przypadku jednak wprost określono zasadę, że przy braku określenia warunków ponownego wykorzystywania informacji sektora publicznego podlega ona wykorzystaniu bezwarunkowo. Jednostki sektora publicznego działające w interesie publicznym i dzięki funduszom pochodzącym od podatników powinny jak najszerszej udostępniać dane i informacje, które wytwarzają lub nabywają w ramach swojej publicznej działalności. Służą to nie tylko wykonywaniu funkcji kontrolnej przez obywateli, ale i pozytywnej stymulacji gospodarki, która dzięki tak dostępnym danym może wytwarzać nowe produkty i usługi. W związku z powyższym należy uznać, iż pomimo możliwości wprowadzenia ograniczeń w korzystaniu z danych sektora publicznego podmioty udostępniające informacje do ponownego wykorzystywania powinny zapewnić jak najszerszą możliwość ich scrapowania. Za uzasadnione można uznać ewentualne określenie warunków, które prowadzą do ograniczenia scrapowania baz danych w sposób zakłócający normalne z nich korzystanie.

Bibliografia

- Barta J., Markiewicz R., *Prawo autorskie, Komentarz*, Warszawa 2013.
- Barta J., Markiewicz R., *Ustawa o ochronie baz danych. Komentarz*, Warszawa 2002.
- Ciupa S., *Dozwolony użytek w ustawie o ochronie baz danych*, „Monitor Prawniczy” 2003, nr 12.
- Fischer B., Piskorz-Ryń A., Sakowska-Baryła M., Wyporska-Frankiewicz J., *Ustawa o ponownym wykorzystywaniu informacji sektora publicznego. Komentarz*, wyd. 2, Warszawa 2019.
- De Mauro A., Greco A.M., Grimaldi M., *A Formal definition of Big data based on its essential features*, “Library Review” 2016, Vol. 3.
- Dubisz S. (red.), *Uniwersalny słownik języka polskiego*, t. 3, Warszawa 2003.
- Europejski Portal Danych, <https://www.europeandataportal.eu/pl>.
- Gienas K., *Data mining jako problem prawa autorskiego*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2017, z. 1.
- Gliściński K., *Komentarz do art. 50 ustawy o prawie autorskim i prawach pokrewnych*, [w:] A. Michalak (red.), *Ustawa o prawie autorskim i prawach pokrewnych. Komentarz*, Warszawa 2019.
- Heiss A., *Big Data Challenges in Big Science, Computing and Software for Big Science*, “Computing and Software for Big Science” 2019, <https://link.springer.com/content/pdf/10.1007%2Fs41781-019-0030-7.pdf>.
- Komisja Europejska, *Shaping Europe's digital future. Open Data*, <https://ec.europa.eu/digital-single-market/en/open-data>.
- Lubasz D., Namysłowska M. (red.), *Ustawa o ochronie baz danych. Komentarz*, Warszawa 2013, LEX.
- Opaliński B., *Dostęp do informacji publicznej w prawie międzynarodowym i prawie Unii Europejskiej*, [w:] P. Szustakiewicz (red.), *Dostęp do informacji publicznej*, Warszawa 2014.
- Selamat S., Prakoonwit S., Khan W., *A review of data mining in knowledge management: applications/findings for transportation of small and medium enterprises*, “SN Applied Sciences” 2020, <https://link.springer.com/content/pdf/10.1007%2Fs42452-020-2589-3.pdf>.

Standardy Otwartości Danych, <https://dane.gov.pl/pl/knowledgebase/useful-materials/standardy-otwartosci-danych>.

Stötzel R., Wille S., *Der Urheberrecht sui generis für Datenbanken*, "Europäisches Wirtschafts- und Steuerrecht" 2004, H. 11.

Traple E., [w:] J. Barta, R. Markiewicz (red.), *Prawo autorskie i prawa pokrewne. Komentarz*, Kraków 2005.

Wojnarska-Krajewska E., *Crawlowanie – legalne czy nie*, <https://wpprawoautorskie.wordpress.com/2014/04/17/crawlowanie-crawler-roboty-txt/>.

Zamora A., *Making Room for Big Data: Web Scraping and an Affirmative Right to Access Publicly Available Information Online*, "The Journal of Business, Entrepreneurship & the Law" 2019, Nr 12.

Uzasadnienie do projektu ustawy o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego.

