

# Wyjątki od prawa autorskiego na rzecz nauki w kontekście technologii cyfrowych

Teresa Świąćkowska

ORCID 0000-0002-3979-5600

*Katedra Informatologii, Wydział Dziennikarstwa, Informacji i Bibliologii,  
Uniwersytet Warszawski*

---

## Abstrakt

**Cel/Teza:** Celem artykułu jest opis i analiza wyjątku od prawa autorskiego dotyczącego eksploracji tekstów i danych dla celów naukowych w kontekście historycznego rozwoju wyjątków dla celów naukowych.

**Koncepcja/Metody badań:** Badanie obejmuje przegląd literatury przedmiotu i akty prawne. W badaniu zastosowano metodę krytycznej analizy tekstów w podejściu tematycznym.

**Wyniki i wnioski:** Omówiono rozwój wyjątków od prawa autorskiego w ujęciu chronologicznym oraz przedstawiono prawne uwarunkowania eksploracji tekstów i danych. Omówiono także wprowadzone ostatnio wyjątki od prawa autorskiego w tym zakresie i wskazano zalety i ograniczenia tych rozwiązań.

**Zastosowanie praktyczne:** Wskazano możliwości rozwoju nowych usług informacyjnych w związku z wprowadzanymi ostatnio regulacjami prawnymi.

**Oryginalność/Wartość poznawcza:** W artykule przedstawiono nowe zjawisko eksploracji tekstów i danych dla celów naukowych w interdyscyplinarnym ujęciu, łączącym zagadnienia prawa autorskiego z problematyką dostępu do informacji.

## Słowa kluczowe

Badania naukowe. Eksploracja tekstów i danych. Prawo autorskie. Wyjątki.

*Otrzymano: 1 sierpnia 2019. Zrecenzowano: 2 października 2019. Poprawiono: 14 października 2019.  
Zaakceptowano: 21 listopada 2019.*

---

## 1. Wprowadzenie. Przedmiot i treść prawa autorskiego

Prawo autorskie jest zaliczane do szerszej kategorii prawa własności intelektualnej, do której zalicza się też prawo patentowe i prawo znaków towarowych. Terminu „własność intelektualna” używa się na określenie dóbr niematerialnych, do których nie mają zastosowania prawa ochrony własności ogólnej, kształtowane w europejskiej kulturze nowożytnej, m.in. w oparciu o tradycję rzymskiego prawa rzeczowego<sup>1</sup>. Tym, co odróżnia prawo własności intelektualnej od prawa własności rzeczowej, jest m.in. ograniczający charakter definicji przedmiotu prawa i czasowe ograniczenie jego ochrony. Ograniczenia te są uzasadniane ważnymi interesami społecznymi, takimi jak wolność intelektualna, czy potrzeby edukacji i dostępu do informacji i wiedzy.

---

<sup>1</sup> Do tego ostatniego zalicza się też pewne dobra niematerialne, jak np. papiery wartościowe.

Prawo autorskie chroni szeroko rozumiane rezultaty działalności twórczej<sup>2</sup>. Chronione są dzieła, które zawierają komponenty twórcze i indywidualne<sup>3</sup>. Ochronie z definicji nie podlegają idee, prawa nauki, koncepcje matematyczne, chronione są natomiast sposoby indywidualnego ich wyrażania. Taka koncepcja przedmiotu ochrony w prawie autorskim opiera się na odróżnieniu treści od formy jej wyrazu. Pojawiła się ona w XVIII w., w wyniku pierwszych sporów o zasady ochrony twórczości intelektualnej. Odróżnienie idei od sposobu ich ekspresji było odpowiedzią zwolenników koncepcji własności literackiej na krytykę, że idei i myśli nie można posiadać, nie można nimi dysponować, jak innymi rzeczami, gdyż doprowadziłoby to do paraliżu społecznej komunikacji i uniemożliwiło rozwój nauki (zob. Świąćkowska, 2018). Ograniczenie przedmiotu ochrony do form wyrażania treści ma zapewniać przestrzeń twórczą i wolność intelektualną, pozwala np. reinterpretować fakty, czy rozwijać lub krytykować koncepcje i teorie.

W XVIII i XIX w., kiedy toczono spory, jak zdefiniować przedmiot ochrony twórczości intelektualnej, zakres tej ochrony był w praktyce rozumiany dość wąsko, jako prawo do druku i obrotu handlowego (zob. Świąćkowska, 2018). Z czasem, wraz z rozwojem nowych form towarowych, zakres ochrony był rozszerzany, m.in. o to, co dziś nazywa się prawem do tworzenia utworów zależnych, czyli np. tłumaczeń, adaptacji. Obecnie zakres ochrony jest rozumiany bardzo szeroko. Poza osobistymi prawami do autorstwa, twórcom przysługuje wyłączność dysponowania prawem do niemalże wszystkich sposobów korzystania z utworu i wynagrodzenia za takowe<sup>4</sup>. W polskim prawie jest to wyrażone następująco: „Jeżeli ustawa nie stanowi inaczej, twórcy przysługuje wyłączne prawo do korzystania z utworu i rozporządzania nim na wszystkich polach eksploatacji oraz do wynagrodzenia za korzystanie z utworu” (Ustawa, 1994, 2019). „Pole eksploatacji” oznacza sposób wykorzystania utworu i jest to np. kopiowanie, upowszechnianie, ale też wypożyczanie, czy odtwarzanie<sup>5</sup>. Według Janusza Barty i Ryszarda Markiewicza, sposobem korzystania z utworu, którego nie kwalifikuje się jako oddzielne pole eksploatacji, jest samo zapoznawanie się z utworem: czytanie książki czy oglądanie filmu w kinie czy telewizji (Barta & Markiewicz, 2016, 155). To wyłączenie odczytywania z monopolu autorskiego nie jest już jednak oczywiste przy odczycie komputerowym, zwłaszcza, że odczyt komputerowy wymaga często uprzedniego sporządzenia kopii cyfrowej. Wyjaśnienie statusu prawnego odczytu komputerowego jest

<sup>2</sup> W polskim prawie definiowane jako utwory. Art. 1 Ustawy o prawie autorskim mówi: „Przedmiotem prawa autorskiego jest każdy przejaw działalności twórczej o indywidualnym charakterze, ustalony w jakiejkolwiek postaci, niezależnie od wartości, przeznaczenia i sposobu wyrażenia (utwór)” (Ustawa, 1994, 2019).

<sup>3</sup> Interpretacje wymogów względem komponentów twórczych i indywidualnych różnią się w zależności od podejścia i mogą być daleko rozbieżne (zob. Markiewicz, 2018).

<sup>4</sup> Rozszerza się również zakres pojęcia przedmiotu prawa autorskiego o nowe kategorie utworów. Przykładem jest objęcie ochroną prawnoautorską programów komputerowych.

<sup>5</sup> Art. 50 polskiej ustawy o prawie autorskim wymienia typowe przykłady pól eksploatacji: „Odrębne pola eksploatacji stanowią w szczególności: (1) w zakresie utrwalania i zwielokrotniania utworu – wytwarzanie określoną techniką egzemplarzy utworu, w tym techniką drukarską, reprograficzną, zapisu magnetycznego oraz techniką cyfrową; (2) w zakresie obrotu oryginałem albo egzemplarzami, na których utwór utrwalono – wprowadzanie do obrotu, użyczenie lub najem oryginału albo egzemplarzy; (3) w zakresie rozpowszechniania utworu w sposób inny niż określony w pkt 2 – publiczne wykonanie, wystawienie, wyświetlenie, odtworzenie oraz nadawanie i reemitowanie, a także publiczne udostępnianie utworu w taki sposób, aby każdy mógł mieć do niego dostęp w miejscu i w czasie przez siebie wybranym” (Ustawa, 1994, 2019).

niezwykle ważne w sytuacji, gdy komunikacja cyfrowa staje się standardem współczesnego społeczeństwa. I tak np. funkcjonowanie Internetu w wysokim stopniu bazuje na tworzeniu kopii cyfrowych. W tym przypadku czysto technologiczne wykorzystanie utworu zostało wyłączone z ochrony prawnoautorskiej na zasadzie wyjątku. W polskiej ustawie wyjątek ten zapisano w art. 23 w następującym brzmieniu<sup>1</sup>:

Nie wymaga zezwolenia twórcy tymczasowe zwielokrotnienie, o charakterze przejściowym lub incydentalnym, niemające samodzielnego znaczenia gospodarczego, a stanowiące integralną i niezbędną część procesu technologicznego, którego celem jest wyłącznie umożliwienie: 1) przekazu utworu w systemie teleinformatycznym pomiędzy osobami trzecimi przez pośrednika lub 2) zgodnego z prawem korzystania z utworu (Ustawa, 1994, 2019).

Wraz z upowszechnianiem się technik cyfrowych pojawiają się nowe sposoby korzystania z utworów, bazujące na odczycie komputerowym, co do których brak pewności, czy podlegają ochronie, czy też są spod niej wyłączone. Zaliczyć można do nich m.in. eksplorację danych (ang. *data mining*) znajdujące zastosowanie w wielu dziedzinach działalności, w tym w badaniach i komunikacji naukowej. Niniejszy artykuł poświęcony jest wyjątkom od prawa autorskiego w odniesieniu do eksploracji danych i tekstów dla celów naukowych.

## 2. Wyjątki i ograniczenia prawa autorskiego

Przy szeroko zdefiniowanym zakresie ochrony prawa autorskiego rośnie znaczenie wyłączeń i wyjątków. Wyłączenia odnoszą się do pewnych elementów i kategorii wytworów intelektualnych ustawowo wyłączonych z ochrony. Prawo autorskie nie chroni odkryć, idei, procedur, metod, zasad działania oraz koncepcji matematycznych. Z ochrony wyłączone są też akty normatywne; urzędowe dokumenty, materiały, znaki i symbole; opublikowane opisy patentowe; proste informacje prasowe<sup>6</sup>. Wyjątki to natomiast ustawowo dopuszczone sposoby wykorzystania utworu, które nie wymagają zgody autora<sup>7</sup>. Wyłączenia i wyjątki są podstawą wolności intelektualnej, jak też pełnią ważną rolę w zapewnieniu powszechnego dostępu do edukacji, informacji i kultury. Do najstarszych i najbardziej znanych należą wyjątki w celach cytatu, krytyki, przeglądu, czy relacji prasowych. Tradycyjnie uznawane są też wyjątki dla celów edukacji i nauki. Do ważnych dóbr publicznych zaliczają się również wyjątki dla bibliotek, zezwalające na pożyczanie i udostępnianie książek i innych dokumentów. Wyjątki pełnią też ważną rolę we wspieraniu bieżącej twórczości, gdyż do produkcji nowych dzieł niezbędny jest dostęp do wcześniejszego dorobku nauki i kultury (Samuelson, 2017).

Prawo autorskie ma charakter monopolu i jako takie ma wiele potencjalnych niekorzystnych rezultatów na obieg informacji i wiedzy. Wyjątki i ograniczenia łagodzą lub zapobiegają pewnym niekorzystnym zjawiskom związanym z monopolem, ponieważ pozwalają na wykorzystanie dzieł bez zezwolenia. Ograniczenia prawa autorskiego zapobiegają np. sytuacjom, w których właściciele praw odmawialiby zezwolenia na korzystanie z utworów z powodów

<sup>6</sup> Prawnicy nie zawsze zgadzają się co do tego, jak rozumieć te kategorie. Dotyczy to zwłaszcza określić dokumenty urzędowe, oraz proste notatki prasowe (zob. Barta & Markiewicz, 2016, 90).

<sup>7</sup> Dla określenia tego rodzaju ustawowych wyłączeń prawa autorskiego używa się też terminu „ograniczenia prawa autorskiego”, a w języku polskim przyjęło się określenie wyjątków terminem „dozwolony użytek”.

społecznie nie akceptowanych, np. w celu uniknięcia krytyki, czy debaty. Mathew Sag zwraca uwagę, że w odniesieniu do dóbr intelektualnych rynek zawodzi w przypadkach, kiedy koszt uzyskania pozwolenia na wykorzystanie chronionego utworu jest wyższy, niż zysk, który można osiągnąć na jego podstawie (Sag, 2009, 1616). W takiej sytuacji prawo autorskie staje się czynnikiem blokującym dostęp do utworów.

W erze mediów drukowanych i analogowych wyjątki były tworzone głównie na poziomie ustaw krajowych<sup>8</sup>. Ich zakresy i rodzaje nadal różnią się w zależności od państwa, co ujawnia się też na poziomie terminologii (Goldstein & Hugenholtz, 2010, 360). Ogólne ramy stanowienia wyjątków dla państw członkowskich konwencji berneńskiej uchwalono na konferencji w Sztokholmie w 1967 r. Warunki, które zapisano wtedy w art. 9 (2), znane są w literaturze przedmiotu pod nazwą testu trójstopniowego. Są one następujące: wyjątki od prawa autorskiego dotyczyć mogą jedynie określonych przypadków, nie wyrządzają szkody normalnemu korzystaniu z dzieła oraz nie przynoszą nieuzasadnionego uszczerbku prawowitym interesom autora<sup>9</sup>. W rewizji sztokholmskiej konwencji berneńskiej test trójstopniowy odnosił się tylko do wyjątków od prawa do reprodukcji utworów (Ricketson & Ginsburg, 2006, 763). Międzynarodowe traktaty z lat 90. XX w. rozciągnęły go na wszystkie wyjątki. W Porozumieniu w Sprawie Handlowych Aspektów Praw Własności Intelektualnej (TRIPS) z 1994 r. ograniczenia testu trójstopniowego dotyczą już wyjątków dla wszystkich kategorii przedmiotów prawa własności intelektualnej. Artykuł 13 tego porozumienia mówi:

Członkowie zawiążą ograniczenia lub wyjątki od praw wyłącznych do określonych przypadków specjalnych, które nie będą w sprzeczności z normalnym wykorzystaniem dzieła i nie spowodują bezzasadnej szkody prawomocnym interesom posiadacza prawa własności dzieła (Porozumienie TRISP, 1994).

W ślad za Porozumieniem TRIPS, Traktat WIPO o prawie autorskim z 1996 r. nakazuje stosować test trójstronny w odniesieniu do wyjątków od wszystkich praw chronionych przez konwencję berneńską (Traktat WIPO, 1996, art. 10)<sup>10</sup>.

Wyjątki od prawa autorskiego występują współcześnie w dwóch głównych formach: mogą być zdefiniowane przez ogólne zasady, dopuszczające dozwolone korzystanie lub są określone przez listę bardziej szczegółowych, dopuszczalnych sposobów dozwolonego korzystania z utworów. Druga z tych form jest charakterystyczna dla ustawodawstwa europejskiego. Skończoną listę wyjątków zawiera dyrektywa z 2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym, tzw. dyrektywa internetowa. Wyjątki zdefiniowane dyrektywą internetową są dodatkowo ograniczone ogólnymi warunkami określonymi w teście trójstronnym, a w stosunku do niektórych, jak np. dozwolonego użytku prywatnego, wprowadza się też obowiązek wynagradzania autorów (Dyrektywa UE, 2001, art. 5). Wyjątki wymienione w dyrektywie internetowej nie

<sup>8</sup> Kwestię tę w pierwszych dziesięcioleciach funkcjonowania konwencji berneńskiej postanowiono pozostawić do rozstrzygnięcia na poziomie ustawodawstwa krajowego, gdyż rozwiązania w poszczególnych krajach różniły się od siebie znacząco (Ricketson & Ginsburg, 2006).

<sup>9</sup> Co do trzeciego kroku, zapobiegać uszczerbku interesom autora można przez wymóg wynagrodzenia za korzystanie z dzieł w ramach wyjątku. Niektóre wyjątki zawierają więc wymóg wynagrodzenia autora. Z dozwolonym użytkowaniem osobistym zdefiniowanym w polskiej ustawie w artykule 23 związane jest pobieranie dodatkowych opłat od urzędzeń reprograficznych i nośników informacji. W 2015 r. wprowadzono natomiast wymóg wynagrodzenia za wypożyczanie w bibliotekach publicznych. W ustawie polskiej do wymogu testu trójstopniowego odnosi się art. 35.

<sup>10</sup> Więcej na temat problematyki testu trójstopniowego zob. Sikorski, 2010.

są obowiązkowe, państwa członkowskie mogą je dowolnie określać w swoich krajowych ustawach, nie mogą one jednak wychodzić poza wyznaczone ramy.

Zamknięta, ograniczająca lista wyjątków jest problematyczna o tyle, że nie zostawia miejsca na nowe sposoby korzystania z utworów, które pojawiają się wraz z rozwojem technologii cyfrowych. W takich przypadkach zachodzi konieczność ustawowego wprowadzania nowych wyjątków. W przypadku prawa europejskiego nowe wyjątki wprowadza się, przyjmując nowe dyrektywy. I tak np. dyrektywa z dnia 25 października 2012 r. w sprawie niektórych dozwolonych sposobów korzystania z utworów osieroconych definiuje wyjątek korzystania z dzieł osieroconych<sup>11</sup>. Nowe wyjątki są ustanawiane w odpowiedzi na pojawiające się potrzeby i problemy z pewnym opóźnieniem.

Niektórzy autorzy uważają, że bardziej elastycznym rozwiązaniem jest zdefiniowanie ogólnych warunków dla wyjątków, które są otwarte na włączanie nowych sposobów korzystania z utworów, gdyż decyzje o tym są podejmowane na poziomie sądów bez konieczności wprowadzania nowych legislacji (Hugenholtz & Senftleben, 2011, 29; Reichman & Okediji, 2012, 1405). Takie rozwiązanie istnieje w ustawodawstwie Stanów Zjednoczonych. W prawie północnoamerykańskim dopuszcza się korzystanie z utworów chronionych bez zgody autora na zasadzie uczciwego wykorzystania – *fair use*. Art. 107 Copyright Act z 1976 r. mówi, że nie narusza prawa autorskiego wykorzystanie utworów w celu krytyki, komentowania, informowania o bieżących wydarzeniach, nauczania, nauki i badań naukowych. Ustawa podaje też cztery konteksty, które powinny być uwzględnione przy interpretacji poszczególnych przypadków *fair use*. Należy uwzględnić: cel i charakter wykorzystania utworu, w tym czy jest to wykorzystanie komercyjne, non profit, ewentualnie dla celów edukacji; charakter wykorzystanego utworu; ilość i istotność elementów przejętych w stosunku do całości utworu oraz wpływ na potencjalny rynek i wartość danego utworu (Copyright Law 1976, 2016). Wadą otwartego zdefiniowania wyjątków jest to, że wymykają się one jednoznaczному określaniu i pozostawiają duży zakres niepewności, czy konkretne wykorzystania mieszczą się w ramach *fair use*, czy nie (Sag, 2009, 1615)<sup>12</sup>.

### 3. Wyjątki w kontekście technologii cyfrowych

W latach 90 XX w. w prawie autorskim wprowadzono szereg zmian, uwzględniających rozwój technologii cyfrowych. Zmiany te miały głównie na celu wzmocnienie ochrony przed nieautoryzowaną reprodukcją dzieł w postaci cyfrowej. Rozszerzeniu zakresu ochrony nie towarzyszyły jednak odpowiednie zmiany w definicji wyjątków. Wielu badaczy twierdzi, że w odniesieniu do środowiska cyfrowego mamy do czynienia z zawężeniem uprawnień użytkowników. Niva Elkin-Koren jest zdania, że zmiany w cyfrowym ekosystemie niekorzystnie wpływają na dostęp do wiedzy i zagrażają wolnościom obywatelskim. Wiąże się to z jednej strony z rosnącym uzależnieniem od treści cyfrowej oraz z umacnianiem ochrony prawa autorskiego z drugiej (Elkin-Koren, 2017, 133). Zmiana formy dystrybucji ze sprzedaży kopii na usługi dostarczania dostępu do treści pozwala na większą kontrolę

<sup>11</sup> Więcej na temat tego wyjątku zob. Świątkowska, 2016.

<sup>12</sup> Tito Rendas jest zdania, że zastosowane w UE podejście zamkniętej listy też nie eliminuje niepewności (Rendas, 2015).

sposobów korzystania z treści i jej dalszej dystrybucji. Kontrola ta jest wzmocniona przez ustalenia kontraktowe, które są zaprogramowywaną, integralną częścią oferty dostępu. Takie ustalenia kontraktowe mogą uniemożliwiać korzystanie z ustawowych wyjątków. Podobne rezultaty daje zastosowanie automatycznych filtrów blokujących możliwość publikowania treści chronionych.

Kwestią budzącą dużo kontrowersji jest prawna ochrona zabezpieczeń technicznych, wprowadzona w 1996 r. Traktatem WIPO. Artykuł 11 Traktatu mówi:

Umawiające się Strony zapewnią właściwą ochronę prawną oraz skuteczne środki prawne przeciwko obchodzeniu skutecznych środków technicznych stosowanych przez autorów w związku z wykonywaniem praw na podstawie niniejszego Traktatu lub Konwencji berneńskiej i ograniczających podejmowanie wobec ich utworów działań, na które oni nie udzielili zezwolenia albo które nie są prawnie dozwolone (Traktat WIPO, 1996).

W Stanach Zjednoczonych ochronę zabezpieczeń technicznych ustanowiono ustawą Digital Millennium Copyright Act (DMCA) z 1998. Odpowiednie przepisy znalazły się również w europejskiej dyrektywie internetowej. Techniczne zabezpieczenia przed kopiowaniem blokują korzystanie z utworów w formie cyfrowej dozwolone w ramach wyjątków (Amstrong, 2006; Dusollier, 2003; Mulligan, 2003). Większość obowiązujących wyjątków została utworzona z myślą o utworach w formie kodeksu drukowanego lub zapisanych techniką analogową i nie przystają one do ery cyfrowej.

Według Saga nowe technologie ujawniły potrzebę nowego, nieznanego dotychczas wyjątku w odniesieniu do technologii bazujących na kopiach cyfrowych (Sag, 2009). Od technologii bazujących na kopiach cyfrowych w ogromnym stopniu uzależnione jest funkcjonowanie współczesnego Internetu, a najbardziej znane zastosowania takich technologii znajdziemy w wyszukiwarkach internetowych. Proces wyszukiwania wymaga stworzenia kopii stron internetowych, w celu ich analizy komputerowej. Kopie stron i ich zindeksowanych treści są przechowywane w tymczasowym repozytorium, zwanym *cache*. Skopiowane treści nie są wykorzystywane w ich pierwotnej formie (chronionej prawem autorskim), służą jedynie jako dane i metadane, na których operują algorytmy przeszukiwania. Kopie stron są potrzebne do przetwarzania zapytań, jak też do wyświetlania fragmentów stron, odnoszących się do zapytań<sup>13</sup>.

Na podobnej zasadzie działają narzędzia cyfrowe, służące do eksploracji i analizy dużych zbiorów danych. Technologie bazujące na kopiach cyfrowych są wykorzystywane w narzędziach do wykrywania plagiatów, znajdują też coraz więcej zastosowania w nauce, jak też w badaniach biznesowych dla celów marketingowych. Ich zastosowanie może być jednak utrudnione ze względu na naruszenia prawa autorskiego. Przykładem tego rodzaju sytuacji jest projekt Google Books, który wywołał skargi autorów i wydawców, a jego realizacja została opóźniona i ograniczona przez długoletni proces. Projekt zakładał digitalizację zbiorów bibliotek naukowych, w celu ich pełnotekstowego przeszukiwania. Jednakże kiedy pod koniec 2004 r. Google udostępnił usługę Google Books Search online, wydawcy i autorzy którzy posiadali prawa do zdigitalizowanych książek, zarzucali firmie ich naruszenie.

<sup>13</sup> Wyszukiwarka Google w wynikach wyszukiwania wyświetla linki zarówno do aktualnej strony, jak też do kopii tej strony, na podstawie której przeprowadzone było przeszukiwanie. Dzięki temu użytkownicy mogą obserwować różnice wynikające z aktualizacji stron i zrozumieć dlaczego czasem otrzymują wyniki, które nie są relewantne do zapytań.

Twierdzili, że tworzenie kopii cyfrowych i udostępnianie urywków tekstu bez ich zgody jest naruszeniem prawa autorskiego. Firma broniła się, twierdząc, że kopiowanie dla celów wyszukiwania jest dozwolone w ramach *fair use* (Manuel, 2010). W 2005 r. autorzy i wydawcy zaskarżyli firmę Google przed sądem. Proces był skomplikowany i trwał długo, gdyż skarga miała charakter zbiorowy, w prawie Stanów Zjednoczonych określane jako *class action*, co oznacza, że była składana w imieniu szerokiej grupy autorów i wydawców. Pod koniec 2013 r. sąd nowojorski odrzucił skargę stowarzyszenia autorów, potwierdzając wyjątek dla tworzenia kopii dla celów indeksowania i przeszukiwania. Sądy drugiej instancji w 2014 r. i trzeciej w 2016 r. odrzuciły apelację i utrzymały w mocy wyrok gwarantujący *fair use* dla celów wyszukiwania. Wyrok uzasadniono tym, że kopiowane utwory zostały wykorzystane w sposób przekształcający (ang. *transformative*), czyli odmienny od ich normalnego wykorzystania (Sag, 2019). Taki sposób wykorzystania utworu tylko do celów komputerowej analizy jest określanej literaturze anglosaskiej jako *non-expressive use*, które to określenie odnosi się do rozróżnienia w definicji przedmiotu prawa autorskiego treści i formy ich wyrażania<sup>14</sup>.

W Stanach Zjednoczonych sądy potwierdziły również w kilku innych wyrokach, że tego rodzaju wykorzystanie utworów mieści się w ramach *fair use* (Sag, 2019). Na poziomie ustawodawstwa UE odpowiedni wyjątek wprowadzono w marcu 2019 r., ale ma on znacznie węższy zakres niż w USA, gdyż podmiotami uprawnionymi są instytucje badawcze i instytucje dziedzictwa kulturowego, a także obowiązuje tylko dla celów badań naukowych.

## 4. Wyjątki dla nauki

Wyjątki dla nauki, podobnie jak inne ograniczenia prawa autorskiego, były do końca XX w. ustanawiane głównie przez ustawodawstwa krajowe. W konwencji berneńskiej pierwszy przepis dotyczący wyjątku dla celów nauki i edukacji pojawił się w 1948 r. i nie był obowiązujący. Miał on też relatywnie wąski zakres, pozwalał na wykorzystanie urywków dzieł literackich i artystycznych w pracach naukowych. W rewizji paryskiej konwencji berneńskiej z 1971 r. zdefiniowano wyjątki w celu cytatu i nauczania. Artykuł 10 Aktu Paryskiego w punkcie pierwszym pozwala na

(...) przytaczanie cytatów z dzieła już legalnie udostępnionego odbiorcom, pod warunkiem że jest to zgodne z przyjętymi zwyczajami i w stopniu uzasadnionym przez zamierzony cel, a także cytatów z artykułów, zamieszczonych w dziennikach i czasopismach, w formie przeglądów prasy (Akt Paryski, 1990).

Punkt drugi art. 10 pozwala natomiast na możliwość

(...) legalnego korzystania, w stopniu uzasadnionym przez zamierzony cel, z dzieł literackich lub artystycznych dla zilustrowania nauczania za pomocą środków publikacji, audycji radiowo-telewizyjnych, zapisów dźwiękowych lub wizualnych pod warunkiem, że takie korzystanie jest zgodne z przyjętym zwyczajem (Akt Paryski, 1990).

<sup>14</sup> Nie wszyscy jednak zgadzają się z takim uzasadnieniem wyjątku dla celów automatycznego przetwarzania tekstu. Maurizio Borghi i Stavroula Karapapa zwracają uwagę, że ten nowy sposób wykorzystania utworów ma potencjał bardzo lukratywnego rynku w bliskiej przyszłości i nie uwzględnia się w nim zgody autorów i wynagrodzenia dla nich (Borghi & Karapapa, 2011).

Sformułowanie „dla zilustrowania” powtórzono również w wyjątku dla celów naukowych w dyrektywie internetowej z 2001 r. Punkt 3 art. 5 mówi, że państwa członkowskie mogą przewidzieć wyjątki lub ograniczenia w odniesieniu do zwielokrotniania i publicznego udostępniania utworów w przypadku

(...) korzystania wyłącznie w celach zilustrowania w ramach nauczania lub badań naukowych tak długo jak źródło, łącznie z nazwiskiem autora, zostanie podane, poza przypadkami, w których okaże się to niemożliwe i w stopniu uzasadnionym przez cel niehandlowy, który ma być osiągnięty (Dyrektywa UE, 2001).

Zwrot „wyłącznie w celach zilustrowania” został wprowadzony, ponieważ krajowe ustawy niektórych państw członkowskich zawierały już takie sformułowanie przed wprowadzeniem dyrektywy. Jest on jednak krytykowany, jako nieprecyzyjny i wprowadzający element niepewności co do dozwolonych sposobów korzystania z utworów (Reichman & Okediji, 2012, 1381).

Jerome H. Reichman i Ruth L. Okediji są zdania, że brak wyraźnie zdefiniowanych wyjątków w konwencji berneńskiej był rekompensowany w okresie dominacji technologii druku przez wyjątek użytku prywatnego, obejmujący również indywidualne badania (Reichman & Okediji, 2017). Prawo użytku prywatnego jest wymienione również na liście wyjątków w dyrektywie internetowej z 2001 r., jego wykorzystanie w środowisku cyfrowym jest jednak ograniczone przez przepisy wprowadzone w latach 90. XX w. i później, dostosowujące prawo autorskie do kontekstu technologii cyfrowych. Poza tym dyrektywa nakłada obowiązek wynagradzania autorów za korzystanie w ramach tego wyjątku. Prawo użytku prywatnego nie może być też stosowane do niektórych ważnych współcześnie sposobów wykorzystania utworów dla celów nauki, które wymagają pracy zespołowej i wsparcia instytucjonalnego. Należy do nich m.in. eksploracja tekstów i danych (Text and Data Mining – TDM).

## 5. Text and Data Mining. Wyjątki dla celów naukowych

Text and Data Mining (TDM) jest techniką badawczą, służącą do pozyskiwania informacji z dużych ilości danych za pomocą programów komputerowych (Geiger et al., 2018). TDM zyskuje rosnące zastosowanie w nauce, m.in. w naukach o życiu, medycynie, lingwistyce czy krytyce literackiej (Ignatow & Mihalcea, 2018). Celem TDM jest metodyczne pozyskiwanie wiedzy, wzorów, czy trendów przez cyfrowe przetwarzanie ogromnych ilości nieusystematyzowanych danych. Źródłem danych są zarówno gotowe bazy czasopism naukowych, informacji gospodarczych, czy artykułów prasowych (np. EBSCO, Factiva, LexisNexis), jak też rozproszone dane, np. te tworzone przez użytkowników Internetu. Eksploracja cyfrowa danych w bazach komercyjnych nie zawsze jest jednak możliwa, nawet dla posiadaczy licencji. Niektóre komercyjne bazy danych oferują własne narzędzia do eksploracji danych, ale są one dodatkowo płatne, a jako narzędzia standardowe nie zawsze odpowiadają zamierzonym celom badaczy. Dla celów TDM wykorzystuje się też ogólnodostępne bazy danych, np. zawierające informacje publiczne, dokumenty urzędowe czy korpusy tekstów do badań językoznawczych<sup>15</sup>.

<sup>15</sup> Przykłady wielu takich źródeł zob. Ignatow & Mihalcea, 2018.

Jak zauważa Eleonora Rosati, istnieje wiele metod, praktyk i celów stosowania TDM, co utrudnia przedstawienie i zilustrowanie ogólnych zasad wykorzystanie tej metody badawczej. Można jednak wyróżnić analitycznie trzy główne etapy TDM, które w praktyce nie muszą stanowić oddzielnych kroków postępowania: dostęp do treści; wydobycie i/lub kopiowanie treści; pozyskiwanie tekstu i danych oraz tworzenie wiedzy (Rosati, 2019). Dwa pierwsze etapy mogą wchodzić w kolizję z prawem autorskim lub prawem ochrony baz danych. Jak zaznaczono wyżej, wyjątek kopiowania dla celów komputerowej analizy potwierdziły jako pierwsze sądy amerykańskie. W marcu 2019 r. Parlament Europejski przyjął dyrektywę *Prawa autorskie na jednolitym rynku cyfrowym*, która przewiduje m.in. wyjątek dla celów TDM na potrzeby badań naukowych. Art. 3 dyrektywy mówi, że państwa członkowskie wprowadzają wyjątek zezwalający na zwielokrotnienia i pobrania dokonywane (...), przez organizacje badawcze i instytucje dziedzictwa kulturowego w celu przeprowadzenia, dla potrzeb badań naukowych, eksploracji tekstów i danych na utworach lub innych przedmiotach objętych ochroną, do których mają zgodny z prawem dostęp (Dyrektywa UE, 2019).

Zdefiniowany w ten sposób wyjątek ma dość wąski zakres. Użytkownikom treści z licencjonowanym dostępem ma gwarantować możliwość TDM, gdyż obowiązuje niezależnie od ustaleń kontraktowych. W odniesieniu do tekstów ogólnie dostępnych w Internecie, uprawnione do nich podmioty mogą jednak zastrzec prawa do ich wykorzystania dla celów TDM

(...) za pomocą środków nadających się do odczytu maszynowego, łącznie z metadanymi i warunkami korzystania ze strony internetowej lub usługi (Dyrektywa UE, 2019).

Możliwość takiego zastrzeżenia może poważnie wpłynąć na dostęp do tekstów i danych dla celów badań naukowych. Hilty i Richter są zdania, że wymóg zgodnego z prawem dostępu będzie niekorzystny dla mniej zamożnych instytucji badawczych, których nie stać na zakup licencji na dostęp do odpowiednich baz danych (Hilty & Richter, 2017).

## 6. Wnioski i perspektywy w odniesieniu do Polski

Państwa członkowskie są zobowiązane do implementacji dyrektywy w ciągu dwóch lat. Problem polega na tym, że zawiera ona również inne przepisy budzące dużo kontrowersji i sprzeciwu ze strony niektórych państw członkowskich, co może spowodować opóźnienia w implementacji i dodatkowo wydłużyć proces wprowadzania omawianych wyjątków<sup>16</sup>. Niektóre kraje europejskie (Wielka Brytania, Francja i Niemcy) wprowadziły już wyjątki dotyczące TDM w ustawach krajowych, ale są one wężziej zdefiniowane niż wyjątek określony dyrektywą.

Polska nie ma oddzielnego wyjątku dla TDM, ale ustawa o prawie autorskim z 1994 r. (tekst ujednolicony z 2019 r.) zapewnia relatywnie szeroki wyjątek dla celów nauki. W art. 27 pozwala się instytucjom naukowym i oświatowym

<sup>16</sup> Projekt dyrektywy europejskiej złożono w 2016 r., kiedy w Stanach Zjednoczonych istniało już kilka orzeczeń sądowych potwierdzających *fair use* dla TDM. W USA proces w sprawie Authors Guild vs. Google trwał w sumie ponad dekadę. Ustalanie nowych wyjątków trwa więc długo w obydwu systemach prawnych, europejskim i północnoamerykańskim, z tym, że rozwiązania europejskie są jeszcze opóźnione w stosunku do północnoamerykańskich.

(...) na potrzeby zilustrowania treści przekazywanych w celach dydaktycznych lub w celu prowadzenia działalności naukowej korzystać z rozpowszechnionych utworów w oryginale i w tłumaczeniu oraz wielokrotnie w tym celu rozpowszechnione drobne utwory lub fragmenty większych utworów (Ustawa, 1994, 2019).

O ile można mieć wątpliwości, czy art. 27 pozwala na korzystanie w celach TDM i czy obejmuje również te przypadki, kiedy eksploracja tekstów bazuje na kopiach cyfrowych całych utworów, to dyrektywa wątpliwości te rozwiewa. Z drugiej strony dyrektywa wprowadza ograniczenie, którego nie ma w polskiej ustawie, dotyczące zgodnego z prawem dostępu.

Inną zmianą wprowadzoną przez dyrektywę jest to, że otwiera ona możliwość korzystania z utworów w celach eksploracji tekstów bibliotekom publicznym i innym instytucjom dziedzictwa kulturowego, które nie są podmiotem uprawnionym w art. 27<sup>17</sup>. Otwiera to potencjalnie przed tego rodzaju instytucjami nowe pole działalności, tj. zapewnianie dostępu do tekstów i danych dla celów TDM<sup>18</sup>. Polskie biblioteki i archiwa, które dysponują zbiorami cyfrowymi, zawierającymi dokumenty chronione prawem autorskim, będą mogły legalnie przetwarzać je komputerowo dla celów naukowych.

W Polsce zastosowanie TDM jest jeszcze niewielkie, ale wykorzystanie tej metody w innych krajach wskazuje na to, że ma ona duży potencjał rozwoju (Rosati, 2019). Niewątpliwie potrzebne są dalsze badania w tym kierunku.

## Bibliografia

- Akt Paryski (1990). Akt Paryski Konwencji Berneńskiej o Ochronie Dzieł Literackich i Artystycznych sporządzony w Paryżu dnia 24 lipca 1971 r. *DzU*, 82.
- Armstrong, T. K. (2006). Digital Rights Management and the Process of Fair Use. Faculty Articles and Other Publications [online], Paper 146. Cincinnati College of Law. Scholarship and Publications, [25.07.2019], [http://scholarship.law.uc.edu/fac\\_pubs/146](http://scholarship.law.uc.edu/fac_pubs/146)
- Barta, J., Markiewicz, R. (2016). *Prawo autorskie*. Warszawa: Wolters Kluwer.
- Borghi, M., Karapapa, S. (2011). Non-display Uses of Copyright Works: Google Books and Beyond. *Queen Mary Journal of Intellectual Property*, 1(1), 21–52.
- Copyright Law (1976, 2016). Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code, December 16 [online]. Gopyright.gov [25.07.2019], <https://www.copyright.gov/title17/>
- Dusollier, S. (2003). Fair Use by Design in the European Copyright Directive of 2001, *Communications of the ACH*, 47 (4), 51–55.
- Dyrektywa UE (2001). Dyrektywa 2001/29/WE Parlamentu Europejskiego i Rady (UW) z dnia 22 maja 2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym. *Dziennik Urzędowy Unii Europejskiej*, 17(1), 230–239.
- Dyrektywa UE (2019). Dyrektywa 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

<sup>17</sup> Art. 27 definiuje podmioty uprawnione następująco: „Instytucje oświatowe oraz podmioty, o których mowa w art. 7, ust. 1, pkt 1, 2 i 4–8 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce” (Ustawa, 1994, 2019).

<sup>18</sup> Przykładu tego rodzaju usług dostarcza inicjatywa HathiTrust, fundacji która oferuje badaczom dostęp do ogromnych kolekcji cyfrowych w celu ich komputerowej eksploatacji dla celów badań naukowych (Plale et al., 2019). Zbiory Hathi Trust to cyfrowe kopie zbiorów bibliotek naukowych, których znaczna część to utwory nadal chronione prawem autorskim. Są one udostępniane wyłącznie w celach badawczych i tylko do analizy komputerowej (Plale et al., 2019, 53).

- zmiany dyrektyw 96/9/WE i 2001/29/WE. *Dziennik Urzędowy Unii Europejskiej* [online] 130/92, 17 maja. EUR-lex [10.10.2019], <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:32019L0790&from=EN>
- Elkin-Koren, N. (2017). Copyright in a Digital Ecosystem. A User Rights Approach. In: R. L. Okediji (ed.), *Copyright Law in an Age of Limitations and Exceptions* (132–168). Cambridge: Cambridge University Press.
- Geiger, Ch., Frosio, G., Bulayenko, O. (2018). Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? Legal Analysis and Policy Recommendations. *IIC – International Review of Intellectual Property and Competition Law*, 49(7), 814–844.
- Goldstein, P., Hugenholtz, P. B. (2010). *International Copyright: Principles, Law, and Practice*. Oxford: New York: Oxford University Press.
- Ghosh, Y. (2017). Jurisprudential Analysis of the Rights of the Users. In: M. K. Sinha, V. Mahalwar (eds), *Copyright Law in the Digital World. Challenges and Opportunities* (61–96). Singapore: Springer.
- Ignatow, G., Mihalcea, R. (2018). *An Introduction to Text Mining Research Design, Data Collection, and Analysis*. Thousand Oaks: Sage.
- Hilty, R., Richter, H. (2017). Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B Exceptions and Limitations (Art. 3 – Text and Data Mining). *Max Planck Institute for Innovation & Competition Research Paper* [online] 17(2), January 14. SSRN, [25.07.2019], <https://ssrn.com/abstract=2900110> or <http://dx.doi.org/10.2139/ssrn.2900110>
- Hugenholtz, P., Senftleben, M. (2012). Fair Use in Europe: In Search of Flexibilities, [online], *Institute for Information Law Research Paper*, 2012–33, February 29. SSRN [10.10.2018], <https://ssrn.com/abstract=2013239> or <http://dx.doi.org/10.2139/ssrn.2013239>
- Iavindrasana, J., Cohen, G. Depeursinge, A., Müller, H., Meyer, R., Geissbuhler, A. (2009). Clinical Data Mining: A Review. *Methods of Information in Medicine* [online], 121–133, <https://doi.org/10.1055/s-0038-1638651>
- Kamocki, P., Ketzan, E., Wildgans, J., Witt, A. (2019). New Exceptions for Text and Data Mining and Their Possible Impact on the CLARIN Infrastructure. Selected Papers from the CLARIN Annual Conference 2018. Linköping Electronic Conference Proceedings [online], 159: 66–71. Linköping University Electronic Press [30.08.2019], <http://www.ep.liu.se/ecp/159/007/ecp18159007.pdf>
- Manuel, K. M. (2010). The Google Library Project: Is Digitization for Purposes of Online Indexing Fair Use under Copyright Law? In: B. D. Rhodes (red), *Copyright Law And A Brief Look At The Google Library Project* (131–143). New York: Nova Science Publishers.
- Mulligan, D. K. (2003). Digital Rights Management and Fair Use by Design. *Communications of the ACH*, 47(4), 31–33.
- Plale, B. A., Dickson, E., Kouper, I., Liyanage, S. H., Ma, Y., McDonald, R. H., Walsh, J. A., Withana, S. (2019). Safe Open Science for Restricted Data. *Data and Information Management* [online], 3(1), 50–60. DeGruyter [25.07.2019], <https://www.degruyter.com/downloadpdf/j/dim.2019.3.issue-1/dim-2019-0005/dim-2019-0005.pdf>
- Porozumienie TRISP (1994). Porozumienie w Sprawie Handlowych Aspektów Praw Własności Intelektualnej [online]. Biblioteka Sejmowa [31.08.2019], <http://libr.sejm.gov.pl/tek01/txt/inne/1994-r0.html>
- Reichman, J.H., Okediji, R. L. (2012). When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale. *Minnesota Law Review*, 96(4), 1362–1480.
- Rendas, T. (2015) Destereotyping the Copyright Wars: The ‘Fair Use vs. Closed List’ Debate in the EU, [online], September 8. SSRN [10.10.2019], <https://ssrn.com/abstract=2657482> or <http://dx.doi.org/10.2139/ssrn.2657482>
- Ricketson, S., Ginsburg, J.C. (2006). *International Copyright and Neighbouring Rights: The Berne Convention and Beyond*. Oxford: New York: Oxford University Press.

- Rosati, E. (2019) Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI. *Asia Pacific Law Review* [online], September 12. SSRN [12.10.2019], SSRN: <https://ssrn.com/abstract=3452376>
- Sag, M. (2009). Copyright and Copy-Reliant Technology. *Northwestern University Law Review*, 103, 1607–1682.
- Sag, M. (2019). The New Legal Landscape for Text Mining and Machine Learning. *Journal of the Copyright Society of the USA* [online], 66, February 9. SSRN [25.07.2019], <https://ssrn.com/abstract=3331606> or <http://dx.doi.org/10.2139/ssrn.3331606>
- Samuelson, P. (2017). Justifications for Copyright Limitations and Exceptions. In: R. L. Okediji (ed.) *Copyright Law in an Age of Limitations and Exceptions* (12–59). Cambridge: Cambridge University Press.
- Sikorski, R. (2010) Ocena dozwolonego użytku w prawie autorskim w świetle kryteriów testu trójstronnego. W: J. Kępiński, K. Klafkowska-Waśniowska, R. Sikorski (red.). *Granice prawa autorskiego* (25–52). Warszawa: Wydawnictwo C.H. Beck.
- SPARC (n.d.) SPARC Europe Guidance to Members Publishers' Content-Mining Policies. SPARC Europe [25.07.2019], [https://sparceurope.org/test2017/wp-content/uploads/dlm\\_uploads/2017/12/TDMlicenceguidance\\_SPARCEurope\\_170214-1-2.pdf](https://sparceurope.org/test2017/wp-content/uploads/dlm_uploads/2017/12/TDMlicenceguidance_SPARCEurope_170214-1-2.pdf)
- Świąćkowska, T. (2016). Choroba sieroca humanistyki cyfrowej, czyli o problemach z niewyjaśnionymi prawami autorskimi. W: B. Sosińska-Kalata; M. Przystek-Samokowa; Z. Wiorogórska (red.) *Nauka o informacji w okresie zmian. Informatologia i humanistyka cyfrowa* (109–120). Warszawa: SBP.
- Świąćkowska, T. (2018). *Kochani krwiopijce. Własność literacka i prawo autorskie w dziewiętnastowiecznej Polsce*. Kraków: Universitas.
- Traktat WIPO (1996). Traktat WIPO o prawie autorskim. *Dziennik Urzędowy Unii Europejskiej*, [online] 33(11), 210–216, [22.07.2019], [https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:22000A0411\(01\)&from=PL](https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:22000A0411(01)&from=PL)
- Ustawa (1994, 2019). Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych. *DzU*, 24, poz. 83. Tekst ujednolicony: 2019 [online]. Biblioteka Sejmowa [30.08.2019], <http://prawo.sejm.gov.pl/isap.nsf/download.xsp/WDU19940240083/U/D19940083Lj.pdf>

---

## Exceptions to the Copyright Law for the Purpose of Scientific Research in the Context of Digital Technologies

**Purpose/Thesis:** The article describes and analyzes exceptions to the copyright law for the purpose of text and data mining from a historical perspective.

**Approach/Methods:** The research presented in the article consisted of thematically oriented critical analysis of specialist literature and of legal acts.

**Results and conclusions:** The article presents the legal conditions of text and data mining; it discusses the recently established exceptions to the copyright law regarding this process [this process =text and data mining] and explores the affordances and limitations of this solution.

**Practical implications:** The article suggests further directions for the development of new information services in connection with the latest legal regulations.

**Originality/Value:** The article investigates the phenomenon of text and data mining for scientific purposes from an interdisciplinary perspective accounting for the copyright issues and the question of access to information.

### Keywords

Copyright. Limitations and exceptions. Scientific research. Text and data mining.

*Dr hab. TERESA ŚWIĘĆKOWSKA jest adiunktem w Katedrze Informatologii Wydziału Dziennikarstwa, Informacji i Bibliologii, Uniwersytetu Warszawskiego. Specjalizuje się w problematyce prawa autorskiego w działalności informacyjnej i bibliotecznej. Jest autorką wielu artykułów z tego zakresu opublikowanych w czasopismach i pracach zbiorowych. Ostatnio wydała książkę Kochani krwio pijce. Własność literacka i prawo autorskie w dziewiętnastowiecznej Polsce (Kraków 2018).*

*Kontakt z autorką*

*t.swieckowska@uw.edu.pl*

*Katedra Informatologii*

*Wydział Dziennikarstwa, Informacji i Bibliologii*

*Uniwersytet Warszawski*

*ul. Nowy Świat 69*

*00-046 Warszawa*