

*Veslava Osińska**

Małgorzata Kowalska

Instytut Informacji Naukowej i Bibliologii
Uniwersytet Mikołaja Kopernika w Toruniu

Bożena Bednarek-Michalska

Biblioteka Uniwersytecka
Uniwersytet Mikołaja Kopernika w Toruniu

Piotr Malak

Instytut Informacji Naukowej i Bibliotekoznawstwa
Uniwersytet Wrocławski

Czas uwolnienia zasobów a dynamika rozwoju i profil polskich bibliotek cyfrowych¹

Abstrakt. Zasoby polskich bibliotek cyfrowych, udokumentowanych szesnastoletnią historią są zintegrowane na sieciowej platformie FBC (Federacja Bibliotek Cyfrowych). To umożliwia przeszukiwanie wszystkich zdigitalizowanych zbiorów tych instytucji za pomocą jednolitego interfejsu wyszukiwawczego. W FBC obecnie zarejestrowano około 120 bibliotek cyfrowych, a liczba obiektów sięga ponad 2 miliony. System digitalizacji obiektów charakteryzuje się dużym rozproszeniem i brakiem jednolitej polityki planowania. Autorzy wykorzystali metadane FBC do analizy dynamiki zmian zasobów cyfrowej wiedzy. W artykule przedstawiono charakterystyki czasowe rozwoju kolekcji cyfrowych w Polsce oraz zaproponowano nowy wskaźnik dynamiki zmian, oparty na wybranych polach w bazie danych FBC. Różnicę, czyli czas który upływa od powstania dzieła do jego publikacji w Internecie autorzy nazwali czasem uwolnienia zasobu (CUZ), który zbadali dla każdej z ponad stu bibliotek cyfrowych. Za pomocą wizualnej analizy histogramów dowiedziono, że rozkład CUZ nie jest przypadkowy, lecz wykazuje prawidłowości, wskazujące na wspólną politykę digitalizacji dla rozmaitych typów bibliotek cyfrowych. Zaproponowana metoda oparta na charakterystykach czasowych może pomóc w systematyce instytucji dygitalizujących, jak i w monitorowaniu dynamiki rozwoju zasobów cyfrowej wiedzy.

Słowa kluczowe: Polskie Biblioteki Cyfrowe, FBC, analiza dynamiki, ekstrakcja danych, wizualizacja danych

Timeline profile in dynamics analysis of Polish Digital Libraries growth

Abstract. Polish digital libraries having 16-years history integrate their resources on Web platform FBC (Federation of Digital Libraries). It allows users to retrieve relevant objects in digital collections by using uniform interface. The FBC currently provides to the more than 2 million objects from around 120 digital libraries. All

* Adres do korespondencji: Instytut Informacji Naukowej i Bibliologii UMK, ul. Wł. Bojarskiego 1, 87-100 Toruń, e-mail: wiewo@umk.pl

¹ Badania przeprowadzono w ramach projektu badawczego NCN 2014-2017 pt. Badanie struktury i dynamiki cyfrowych zasobów wiedzy za pomocą metod wizualizacji (ang. Information Visualization methods in digital knowledge structure and dynamics study). Projekt dotyczy analizy rozwoju nauk humanistycznych i społecznych (HS) w Polsce przy wykorzystaniu metod wizualizacji. Zob. na stronie projektu: http://www.wizualizacjanauki.umk.pl/pl/#o_projekcie.

system of digitization is distributed with the lack of uniform planning policy. The authors used FBC metadata to analyze the dynamics of changes of digital knowledge. The article presents timeline characteristics of digital collections growth as also proposed a new dynamics index, based on date fields in FBC database. The delay time was introduced as an interval between creation date of document and its later network publishing. Due to visual analysis the authors found that delay time shape on histograms is not random, but depends on the type of digital library. The proposed method can help in these institutions systematics as well as in monitoring the dynamics of digital knowledge growth.

Keywords: Polish Digital Libraries, FBC, dynamics analysis, timeline, data extraction, data visualisation

1. Wstęp

Rozwój bibliotek cyfrowych jest odpowiedzią na stale rosnący przyrost zasobów wiedzy cyfrowej oraz konieczność wypracowania organizacyjnych, technologicznych i prawnych rozwiązań w zakresie zarządzania zgromadzonym w nich dorobkiem ludzkości – tworzonym zarówno współcześnie, jak i w przeszłości. Idea biblioteki cyfrowej, choć nawiązuje do tradycyjnej koncepcji biblioteki jako miejsca, w którym kumulowane są wytwory myśli i twórczości człowieka, związana jest z szerszą funkcjonalnością i co najistotniejsze, dostępnością dla każdego. Uniwersalnego charakteru bibliotekom cyfrowym nadają nie tyle formy udostępnianych materiałów (m.in. dokumenty tekstowe, graficzne, materiał dźwiękowy i wideo, a także interaktywne strony WWW), lecz rodzaje zdigitalizowanych dzieł: rękopisy, stare druki, cenne zbiory ikonograficzne, materiały dydaktyczne, dokumenty życia społecznego – wszystkie stanowiące z jednej strony wartość unikatową w kontekście dziedzictwa dokumentalnego, z drugiej – utylitarną dla potencjalnych użytkowników. To dzięki digitalizacji biblioteki cyfrowe można więc postrzegać jako pewnego rodzaju narzędzie zabezpieczające zbiory – zwłaszcza te utrwalone na kwaśnym papierze, narażone na szybką degradację i zniszczenie czy po prostu zacytane. Jednak udostępnianie zasobów w bibliotekach cyfrowych nie zawsze poprzedzone jest procesem digitalizacji. Kolekcje bibliotek cyfrowych coraz częściej bowiem zawierają obiekty od początku powstałe w formie elektronicznej, tzw. *born digital*. Charakteryzują się one wysoką jakością obrazu cyfrowego, dobrą rozdzielczością oraz większą kompresją, co przekłada się na czytelność dokumentów i zwiększone możliwości operowania ich kopiami cyfrowymi (obróbka, składowanie, przesyłanie, ściąganie).

W Europie najbardziej znaną biblioteką cyfrową, liczącą ponad 45 mln obiektów cyfrowych, jest Europeana² (stan na czerwiec 2016 r.). Ze względu na to, że udostępnia ona rozbudowaną kolekcję cyfrowych dzieł kultury (np. portrety „Mona Lisa”, „Dziewczyna z perłą”) i nauki (np. opracowania Charlesa Darwina, Isaaca Newtona) pełni rolę wirtualnego muzeum i archiwum Europy. Ponieważ w tworzeniu jej zasobów uczestniczy ponad 2 tysiące europejskich instytucji ochrony

² <http://www.europeana.eu/portal/pl>

i dziedzictwa, nie jest ona jednostkową biblioteką cyfrową *sensu stricte*, lecz agregatorem danych.

W Polsce rolę tego rodzaju agregatora pełni – integrująca polskie biblioteki cyfrowe i oferująca jednolity interfejs wyszukiwawczy – Federacja Bibliotek Cyfrowych (FBC)³. Platforma ta jest efektem realizowanego od 2008 r. przez Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS)⁴ projektu, którego celem jest stworzenie sieci rozproszonych bibliotek cyfrowych i repozytoriów w Polsce. Federacja Bibliotek Cyfrowych jest krajowym agregatorem i dostawcą zasobów dla Europeany, co oznacza, że zbiera dane o zbiorach cyfrowych z lokalnych instytucji kultury, a następnie zbiorczo przekazuje je do Europeany. Uczestnictwo w Europeanie sprawia jednak, że zarówno FBC, jak i pojedyncze biblioteki cyfrowe muszą tworzyć swoje zasoby w oparciu o międzynarodowe standardy dotyczące przygotowywania kopii cyfrowych (formaty użytkowe i archiwalne), sporządzania opisu bibliograficznego (Dublin Core) i przesyłania danych (protokół OAI-PMH). Dzięki temu możliwe jest przeglądanie i przeszukiwanie wielu zbiorów bibliotek i repozytoriów cyfrowych z poziomu jednego portalu.

Twierdzenie, że polskie biblioteki cyfrowe dynamicznie się rozwijają jest truizmem. Wystarczy bowiem na bieżąco monitorować liczbę publikacji zdigitalizowanych i udostępnianych w poszczególnych bibliotekach cyfrowych należących do FBC, by się o tym przekonać. Wprost proporcjonalnie do zwiększania zasobów polskich bibliotek cyfrowych rośnie także zainteresowanie nimi w gronie badaczy. Organizacyjnymi, prawnymi, technicznymi i jakościowymi aspektami ich funkcjonowania zajmuje się m.in. wielu współczesnych biblio- i informatologów, (Franke 2007; Janiak i Krakowska 2010; Głowacka 2011; Kolasa 2012; Kowalska 2012; Żernicka 2015). Jednakże większość naukowych opracowań w tym zakresie odwołuje się do danych statystycznych, podawanych na bieżąco przez PCSS. Zarówno przywołani badacze, jak i statystyki FBC dowodzą, że moment powstania pierwszych bibliotek cyfrowych w kraju przypada na lata 2001–2002, m.in. Polska Biblioteka Internetowa (2002) i Wielkopolska Biblioteka Cyfrowa (2002). Od tego czasu tempo przyrostu polskich zasobów cyfrowych znacznie wzrosło, a sam wykaz FBC w momencie kończenia prac nad artykułem (listopad 2016r.) liczy 120 bibliotek i repozytoriów cyfrowych.

Powszechnie wiadomo, że biblioteki cyfrowe mogą się różnić pod względem wielkości i zakresu treściowego zasobów, infrastruktury technicznej, włączając w to interfejs użytkownika oraz audytorium docelowe. Zróżnicowanie takie wynika z faktu, iż mogą one być zakładane i utrzymywane zarówno w ramach porozumienia kilku instytucji (np. konsorcjum, naukowy zespół biblioteczny), jak i pojedynczych podmiotów, np. małej biblioteki publicznej, fundacji, organizacji pozarządowej. Pomimo sporego zainteresowania badaczy problematyką bibliotek

³ <http://fbc.pionier.net.pl/>

⁴ <http://www.man.poznan.pl/online/pl/>

cyfrowych w polskich ośrodkach naukowych, problemy wywodzące się z wyraźnej niejednorodności tych systemów, np. ocena jakości i funkcjonalności, są niewystarczająco zbadane (Żernicka 2015), a tym samym słabo opisane w literaturze przedmiotu. Autorzy opracowania wychodząc naprzeciw tym brakom, dokonali próby systematyki bibliotek cyfrowych, kierując się wskazówkami – wyłonionego specjalnie dla potrzeb badań – eksperta⁵ oraz dwoma zmiennymi – czasem powstawania oraz deponowania poszczególnych publikacji.

2. Biblioteki cyfrowe w Polsce: organizacja i rozwój

2.1. Platforma integracyjna

Federacja Bibliotek Cyfrowych jest siecią zrzeszającą polskie biblioteki cyfrowe, czyli infrastrukturą technologiczno-organizacyjną do agregacji metadanych z rozproszonych systemów informacyjnych. Z punktu widzenia użytkowników jest to dobrze zaprojektowany, umożliwiający użycie operatorów logicznych interfejs wyszukiwawczy zasobów większości kolekcji cyfrowych w kraju. PCSS nadzorujące funkcjonowanie serwisu i jego rozwój realizuje za jego pomocą więcej celów niż tylko filtrowanie informacji. Jest to także „środowisko współpracy i wymiany informacji” pomiędzy twórcami i praktykami w dziedzinie bibliotek cyfrowych, równolegle wspomagające *na poziomie krajowym koordynację procesów digitalizacji* (Mazurek, Werla 2012). Z punktu widzenia badaczy ważną cechą jest zapewnienie aktualizacji danych na temat bibliotek cyfrowych, jak również publikowanie bieżących zestawień i statystyk (od 2015 r. w nowej odsłonie serwisu działa rozbudowany moduł analityczny), bowiem jak już zauważono, dla portale FBC i PCSS są źródłem aktualnej informacji o stanie i rozwoju kolekcji cyfrowych w Polsce.

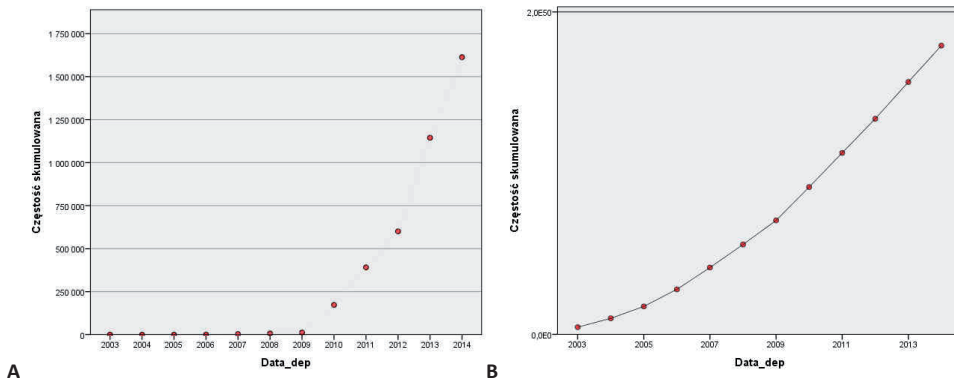
Przy okazji analiz rozwojowych bibliotek cyfrowych w Polsce badacze wymieniają kilka etapów ich powstawania (Głowacka 2011, Kolasa 2012). Pierwszy, wczesny etap rozpoczął się na początku naszego wieku. Złożyły się na to takie procesy, jak: rozwój oprogramowania dLibra i utworzenie Wielkopolskiej Biblioteki Cyfrowej (Mazurek i Werla 2012). Od 2007 r. obserwuje się bardzo szybki wzrost kolekcji cyfrowych w Polsce, co było spowodowane silną determinacją polskich bibliotekarzy oraz powiązane z polityką państwa i finansowaniem digitalizacji w Polsce. W tym właśnie okresie dostrzeżono potrzebę utworzenia agregatora metadanych. W efekcie tego portal FBC uruchomiono 22 czerwca 2007 r. (Mazurek i Werla 2012).

⁵ Do scharakteryzowania stanu rozwoju i rodzajów istniejących bibliotek cyfrowych w Polsce powołał specjalistę w zakresie digitalizacji, biorącego udział w pierwszych projektach i w opracowywaniu polityki ministerialnej, na bieżąco pełniącego rolę doradcy w różnych projektach i który jest autorytetem w środowisku bibliotekarskim.

2.2. Dynamika rozwoju

W roku 2008 w dziewiętnastu polskich bibliotekach cyfrowych zgromadzono około 130 tysięcy obiektów, rok później – już ponad 200 tysięcy. Przyrost obiektów był powolny, lecz systematyczny. Nagły przyrost zasobów i bibliotek nastąpił później po roku 2010, kiedy do Federacji Bibliotek Cyfrowych dołączyły zbiory dużych bibliotek. Do tego grona z pewnością zaliczyć można kolekcję cyfrową Biblioteki Narodowej – POLONA⁶. Przyrost zasobów dobrze ilustruje rysunek 1a. Na rysunku 1b dzięki zastosowaniu logarytmicznej skali zauważalny jest przyrost obiektów w początkowej fazie. Z reguły wykres logarytmiczny stosuje się do wygładzenia dużego wzrostu wartości, w tym przypadku trend wzrostowy dla obu wykresów po roku 2010 niewiele się różni, co świadczy o gwałtownym skoku w gromadzeniu zbiorów cyfrowych.

Rysunek 1. Przyrost zasobów w polskich bibliotekach cyfrowych od początku zaistnienia w skali liniowej (a) i logarytmicznej (b)



Źródło: opracowanie własne na podstawie danych FBC.

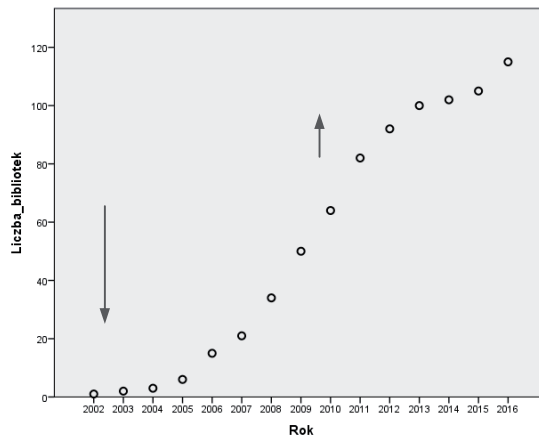
Skala logarytmiczna jest uzasadniona dla dużych rozpiętości analizowanych wartości: od kilkuset obiektów na początku naszego wieku do ponad dwóch milionów w roku 2015. Dla wczesnego okresu rozwoju bibliotek cyfrowych, czyli przed rokiem 2007, na podstawie stopnia pochyłości można wyodrębnić dwie fazy z charakterystycznym „przełomem” w 2004 r. Ta obserwacja zostanie zweryfikowana później, przy analizowaniu specyfiki poszczególnych kolekcji.

Interesującym aspektem w badaniu dynamiki przyrostu cyfrowej wiedzy jest także liczba instytucji gromadzących zasoby, tym bardziej, że różne opracowania podają nieściśle albo rozbieżne dane. Jak z roku na rok rosła liczba bibliotek cyfrowych w Polsce pokazuje rysunek 2. W stałym trendzie wzrostu, który utrzymuje

⁶ <https://polona.pl>

się do dziś widoczne są dwa przełomowe momenty: jeden w latach 2006–2007, oznaczający przyspieszenie dynamiki powstawania ośrodków, drugi w latach 2012–2013 – sygnalizujący spowolnienie. O ile przyspieszenie procesu cyfryzacji łączyć można m.in. z konsekwencjami wejścia Polski do Unii Europejskiej (uruchomienie programów europejskich i krajowych wspierających procesy digitalizacji oraz wyasygnowanie funduszy na ten cel), pojawieniem się firm komercyjnych oferujących usługi digitalizacji dla bibliotek, malejącymi cenami sprzętu komputerowego i skanującego oraz szeroką popularyzacją oprogramowania dLibra (Kowalska 2006), trudno o precyzyjne wyjaśnienie tendencji spadkowych. Można jedynie przypuszczać, że mimo istnienia dużych funduszy unijnych, biblioteki borykały się z takimi problemami, jak: brak dedykowanych digitalizacji grantów, brak fachowców, niekompletność obiektów, kolejne nowelizacje prawa autorskiego (np. kwestia dzieł osieroconych czy utworów *out-of-commerce*), a wreszcie intensyfikacja w zakresie uruchamiania (uznawanych za konkurencyjne) instytucjonalnych repozytoriów cyfrowych⁷, udostępniających nie tylko pozycje natywnie cyfrowe, ale w wielu przypadkach także zdigitalizowane.

Rysunek 2. Kumulacyjna liczba bibliotek cyfrowych w Polsce od momentu powstania pierwszego ośrodka – WBC

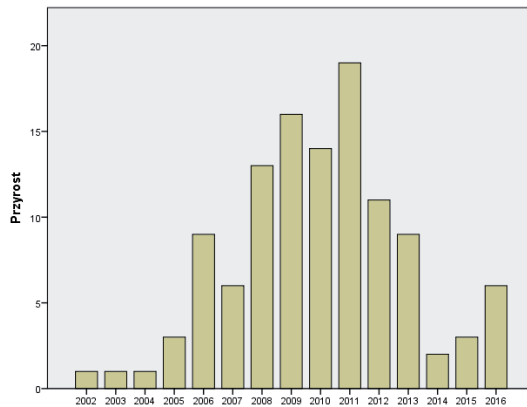


Źródło: opracowanie własne na podstawie danych FBC.

⁷ Tendencje te potwierdzają dane pochodzące z autorytatywnego katalogu repozytoriów cyfrowych OpenDOAR. Według generowanych na bieżąco statystyk po roku 2007 r. nastąpił i nadal trwa intensywny wzrost liczby repozytoriów na świecie. Chociaż w oparciu o dane pochodzące z DOAR trudno formułować jednoznaczne opinie w odniesieniu do polskich repozytoriów, można jednak ostrożnie założyć, że skoro w zasobie tym odnotowuje się również polskie repozytoria, to tendencja wzrostowa także ich dotyczy (dla przypomnienia pierwsze otwarte repozytorium w Polsce – Repozytorium AMUR – uruchomiono w 2010 r.). Zob. [http://www.opendoar.org/one-chart.php?cID=&ctID=&rtID=&cIID=&IID=&potID=&rSoftWareName=&search=&groupby=r.rDateAdded&orderby=&charttype=growth&width=600&height=350&caption=Growth of the OpenDOAR Database – Worldwide](http://www.opendoar.org/one-chart.php?cID=&ctID=&rtID=&cIID=&IID=&potID=&rSoftWareName=&search=&groupby=r.rDateAdded&orderby=&charttype=growth&width=600&height=350&caption=Growth%20of%20the%20OpenDOAR%20Database%20-%20Worldwide)

Rysunek 2 wyraźnie dowodzi, że tendencja wzrostowa utrzymywała się w okresie od 2006 do 2011 r. Uzupełnieniem, wyjaśniającym tę tezę może być rysunek 3 ilustrujący przyrost liczby bibliotek w poszczególnych latach. Wyraźne maksimum przypada na lata 2010–2011, co może potwierdzać prawdziwość sformułowanych wyżej przesłanek przyspieszenia tempa digitalizacji w Polsce i pogląd o doniosłym znaczeniu dużych ośrodków bibliotecznych. Jak ilustruje rysunek 7, dostrzeżona przez Ministerstwo Kultury i Dziedzictwa Narodowego w roku 2007 (początek finansowania projektów digitalizacyjnych ze środków ministerialnych) potrzeba finansowania projektów digitalizacyjnych, znajduje potwierdzenie w punkcie startowym okresu przyspieszenia cyfryzacji w Polsce.

Rysunek 3. Przyrost każdego roku liczby bibliotek cyfrowych w Polsce



Źródło: opracowanie własne na podstawie danych FBC.

Polityka Unii Europejskiej, światowe trendy, nacisk bibliotekarzy na Ministerstwo Kultury i Dziedzictwa Narodowego (MKiDN) oraz Ministerstwo Nauki i Szkolnictwa Wyższego (MNiSW), a także Ministerstwo Archiwizacji i Cyfryzacji (MAiC) spowodowały opracowanie wieloletnich planów digitalizacyjnych dla kraju. I tak, w roku 2009 z inicjatywy MKiDN przygotowany został *Raport o digitalizacji dóbr kultury*, który stanowił pierwsze tak obszerne opracowanie stanu digitalizacji w Polsce, a jednocześnie formułował misję i cele digitalizacji dziedzictwa narodowego oraz zadania w sferze gromadzenia, przechowywania i udostępniania zasobów cyfrowych na lata 2009–2020 (*Raport o digitalizacji* 2009; por. też Kowalska 2010). Z kolei w 2011 r., dzięki staraniom tego samego ministerstwa, uruchomiono Program Wieloletni Kultura+, który w priorytecie „Digitalizacja” wspiera działalność digitalizacyjną m.in. w instytucjach kultury. Celem tego programu jest zarówno rozwój infrastruktury digitalizacyjnej, zabezpieczenie zbiorów dziedzictwa kulturowego poprzez digitalizację, a także ich udostępnianie szerokiemu gronu odbiorców.

Współpraca specjalistów odpowiedzialnych za rozwój poszczególnych bibliotek cyfrowych została usprawniona dzięki FBC. Dobrze działająca sieć nie oznacza jednak, że odległość pomiędzy ośrodkami digitalizacyjnymi jest bez znaczenia. Dlatego warto stworzyć mapę lokalizacji tych instytucji i sprawdzić, które miasta najaktywniej działają na polu digitalizacji, oraz czy można wyodrębnić skupiska metropolitarne wspólnie działające w tym zakresie.

Rysunek 4. Mapa lokalizacji polskich bibliotek cyfrowych w roku 2016



Źródło: opracowanie własne na podstawie danych FBC.

Rysunek 4 wyróżnia Warszawę i Kraków jako miasta, gdzie powstało najwięcej odrębnych instancji bibliotek cyfrowych od początku wieku do 2016 r. Dominacja tych ośrodków wynika z kilku przyczyn: po pierwsze w tych miastach znajduje się wiele placówek bibliotecznych, które realizują proces digitalizacji zbiorów (niektóre z nich, jak np. Akademia Górniczo-Hutnicza, należą do inicjatorów tego procesu w Polsce), po drugie w miastach tych dochodzi do rozdrobnienia procesu digitalizacji, co bierze się z z faktu niechęci wielu placówek do wspólnego tworzenia zasobów w ramach jednej biblioteki cyfrowej (np. brak porozumienia w sprawie tworzenia Małopolskiej Biblioteki Cyfrowej czy uruchamianie odrębnych bibliotek cyfrowych przez poszczególne instytucje funkcjonujące w ramach Uniwersytetu Warszawskiego, np. Bibliologiczna Biblioteka Cyfrowa⁸ i e-biblioteka Uniwersytetu Warszawskiego⁹). Taki stan rzeczy jest także efektem rozbieżnych celów tworzenia zbiorów cyfrowych i ambicji poszczególnych szefów jednostek, którzy koniecznie chcieli mieć odrębną od innych bibliotekę cyfrową. Sytuacja taka znamieną jest jednak nie tylko dla ośrodka krakowskiego i warszawskiego.

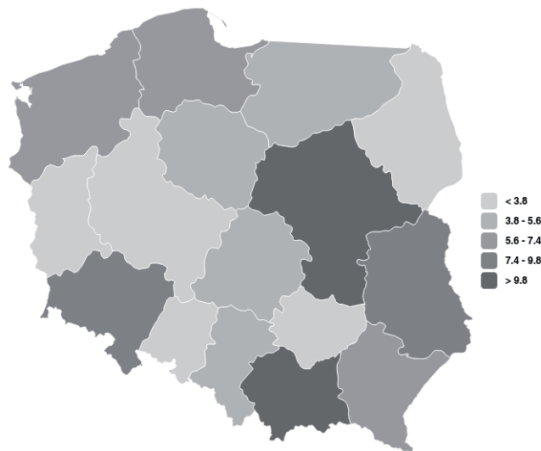
⁸ <http://bbc.uw.edu.pl/dlibra>

⁹ <http://ebuw.uw.edu.pl/dlibra>

W Polsce bowiem powstało wiele bibliotek cyfrowych stworzonych dla potrzeb prezentacji bardzo małych kolekcji. Po ich ucyfrowieniu istnieje niewielkie prawdopodobieństwo, że biblioteki te będą się dalej rozwijały. System jednak będzie trzeba nadal utrzymywać, a obiekty migrować za jakiś czas do innych formatów. Warto więc sobie zadać pytanie, jaki jest efekt ekonomiczny i technologiczny takich decyzji.

Najpopularniejszym w Polsce systemem informatycznym do budowy bibliotek cyfrowych, zarządzania nimi i udostępniania ich zawartości w sieci Internet jest dLibra¹⁰ – oprogramowanie stworzone i rozwijane przez PCSS od 1999 r. Oprogramowanie to zapewnia cały proces upowszechniania zbiorów cyfrowych, włączając w to pracę redaktora biblioteki cyfrowej, korektę metadanych i udostępnianie dzieła w Internecie. Wykorzystanie oprogramowania dLibra w obrębie poszczególnych województw ilustruje rys. 5.

Rysunek 5. Kartogram województw z ilością ośrodków opartych o platformę dLibra



Źródło: opracowanie własne na podstawie danych FBC.

2.3. Istniejące podziały polskich bibliotek cyfrowych

W szesnastoletnim okresie rozwoju polskich bibliotek cyfrowych autorzy wyodrębnili kilka etapów ich powstania. Jak zauważyli, rok powstania jest jedną z prawdopodobnych przyczyn ich zróżnicowania pod względem wielkości, zastosowanej technologii oraz szerokiego spektrum cech indywidualnych. W obrębie kraju różnorodność infrastruktury technicznej bibliotek cyfrowych nie jest duża (Mazurek i in. 2012), natomiast podejście metodologiczne w architekturze kolekcji i strukturze metadanych różni się znacząco. To sprawia, że poszczególne systemy

¹⁰ <http://dlibra.psnc.pl/>

mogą inaczej reagować na szybki wzrost zasobów informacyjnych, ostatecznie uzyskując zróżnicowany poziom funkcjonalności.

Istnieje wiele modeli teoretycznych oceny jakości bibliotek cyfrowych, mających źródło w badaniach zagranicznych, jak i krajowych (Janiak i Krakowska 2012). Wydaje się, iż jednym z najbardziej syntetycznych zestawów kryteriów ewaluacji może być taki, który ocenia kompleksowo najważniejsze elementy. Są nimi: system, interfejs (użyteczność oraz łączność na linii użytkownik – system), jakość usług oraz jakość kolekcji (Janiak i Krakowska 2012). W badaniach jakości usług bibliotek cyfrowych najczęściej używa się kwestionariuszy ankietowych lub sondaży. Nowoczesne metody ewaluacji interfejsu opiera się z kolei o techniki takie, jak *eyetracking*¹¹, *clickstreaming*¹², a nawet *headtracking*¹³ (Osińska 2010b). Spotyka się jednak także podejścia zorientowane na czynnik technologiczny (ocena bibliotek według platformy informatycznej; Żernicka 2015) oraz bazujące na tematycznej zawartości zasobów cyfrowych (próba oceny bibliotek przez pryzmat wybranych zbiorów stanowiących części bibliotek cyfrowych; Kowalska, Jarocki 2010).

Długa lista propozycji kryteriów i atrybutów ewaluacji oraz trudności wykonania samej oceny są prawdopodobnie powodem, dla którego jeszcze nie przeprowadzono jednolitego, wielopoziomowego podziału istniejących w Polsce bibliotek cyfrowych. Jak podkreślają praktycy, w systemie podziału bibliotek cyfrowych mogą być uwzględnione różne aspekty, takie jak np. sposób funkcjonowania, zasady doboru zbiorów, udostępnianie online, schemat opisu obiektów, przygotowanie ich końcowej postaci itp. (Mazurek i in. 2012). Dobrze znanym, zaproponowanym przez specjalistów z PCSS, a wykorzystywanym często przez bibliotekarzy, jest podział bibliotek cyfrowych pod względem organizacyjnym na dwa modele: instytucjonalny i konsorcjalny (Mazurek i in. 2012). W przypadku pierwszego modelu jedna instytucja odpowiada za tworzenie i zarządzanie biblioteką cyfrową oraz za gromadzenie i udostępnianie zasobów. Konsorcjalny model tworzy wiele instytucji, w zakresie kompetencji których leży określenie zasad organizacji i budowania kolekcji. Obok typologii bibliotek cyfrowych zaproponowanej przez PCSS wymienić można jeszcze jedną próbę kategoryzacji tych zasobów – podział stosowany przez redaktorów portalu dla bibliotekarzy EBIB¹⁴. Bazując na podobnych kryteriach podziału, wyłaniają oni trzy typy bibliotek cyfrowych: instytucjonalne (37 pozycji), regionalne (44) oraz wybrane kolekcje cyfrowe skierowane do ściśle określonej grupy odbiorców, np. „Biblia” albo „Biblioteka Wirtualna Matematyki”.

¹¹ Metoda śledzenia ruchu gałek ocznych przy wykonaniu zadanej czynności przy użyciu urządzenia, montowanego na monitorze, zasadą działania przypominającego kamerę internetową.

¹² Metoda rejestracji wszystkich kliknięć myszką przy wykonaniu określonego zadania.

¹³ Metoda rejestracji ruchu głowy, który towarzyszy pracy z dużym monitorem.

¹⁴ Zob. Serwis tematyczny „Biblioteki Cyfrowe” tworzony w ramach portalu dla bibliotekarzy EBIB http://www.ebib.pl/?page_id=666

3. Materiał i metody badawcze

3.1. Zadanie eksperta

Aby dowieść zasadności zastosowania nieco innego podziału bibliotek cyfrowych, autorzy niniejszego opracowania zwrócili się do eksperta z prośbą o pogrupowanie wszystkich bibliotek cyfrowych znajdujących się w wykazie FBC według następujących kryteriów: model organizacyjny biblioteki cyfrowej, zasady jej funkcjonowania, zakres tematyczny udostępnianych materiałów. Założono, że w zależności od tego, przez jaki podmiot tworzona bądź nadzorowana jest biblioteka (uniwersytet, politechnika, biblioteka), inaczej będzie kształtował się nie tylko zakres tematyczny zdeponowanych w niej zasobów, ale również ich zasięg chronologiczny, a konkretnie czas wydania publikacji źródłowych (oryginalnych). Zauważono również, że biblioteki politechnik mają zwykle w swoim zasobie więcej czasopism naukowych i artykułów współczesnych swoich pracowników niż publikacji z domeny publicznej (z tej kategorii są to głównie archiwalne czasopisma techniczne). Wynika to przede wszystkim z funkcji, jakie pełnią te zasoby, tj. m.in. wspierania procesu dydaktycznego uczelni oraz dokumentacji rozwoju nauki i techniki. W tym kontekście przyjęto, że im młodsza politechnika, tym mniej tzw. „starego zasobu”. W efekcie poczynionych założeń i obserwacji analiza ekspercka wyłoniła dwa kryteria podziału bibliotek cyfrowych: kryterium organizacyjne oraz kryterium tematyczne. Według kryterium organizacyjnego biblioteki podzielono na:

- biblioteki regionalne (konsorcyjne),
- biblioteki instytucjonalne (indywidualne),

Według kryterium tematycznego wyróżniono z kolei:

- biblioteki kościelne (teologia, historia),
- biblioteki specjalistyczne (zawierające wąskie kolekcje np. przyrodnicze, czy dot. jednej osoby),
- biblioteki techniczne,
- biblioteki uniwersalne (zawierające kolekcje z różnych dziedzin nauki i obszarów kultury).

Podział zaproponowany przez eksperta jest w zasadzie odzwierciedleniem zakresu tematycznego kolekcji, jakie poszczególne instytucje postanowiły zdigitalizować i umieścić w Internecie. Dowodzi on faktu, że niewiele dały dyskusje bibliotekarzy, jakie odbywały się w Polsce od roku 2004, kiedy to wydawało się, że koncepcja tworzenia jednej biblioteki narodowej i wielu regionalnych ma sens i może prowadzić do spójnego dokumentowania narodowego dziedzictwa w postaci cyfrowej (Bednarek-Michalska 2014a). Zakładano wówczas, iż każda placówka skanuje swój zasób oraz unika dublowania prac innych, w konsekwencji czego – po procesie integracji – powstaje jednolity system obejmujący wszystkie zasoby cyfrowe. Niestety, z braku koordynacji prac w Polsce i myślenia dalekosięż-

nego (każda instytucja przyjęła własną politykę digitalizacji), koncepcja ta się nie przyjęła, co spowodowało powstanie systemu bardzo rozproszonego, składającego się z wielu bibliotek cyfrowych, czasem bardzo małych, czasem funkcjonujących nawet w obrębie jednej instytucji (przywołany przykład Uniwersytetu Warszawskiego). Jedną z przyczyn tego stanu rzeczy był także niski koszt licencji systemu dLibra i brak wyobraźni twórców bibliotek, co do możliwości digitalizacyjnych (osobowych i technicznych) własnych instytucji, jak i dalszego finansowania procesu digitalizacji przez donatorów. Szczęśliwie technologie sieciowe pozwoliły na zbudowanie jednej platformy wyszukiwawczej (FBC) oraz agregację danych ze wszystkich kolekcji. Błędów związanych z jakością zasobów nie udało się jednak uniknąć (Bednarek-Michalska 2014a). Dlatego już dziś wiadomo, że digitalizacja wykonana z niektórych słabych mikrofilmów czy kopii matek wątpliwej jakości w przyszłości będzie bezużyteczna i trzeba będzie ją powtórzyć (Kowalska 2006).

3.2. Czas uwolnienia zasobów – kluczowy parametr w analizach

Jak dowodzą metadane pochodzące z poszczególnych bibliotek cyfrowych, rok uruchomienia danej biblioteki nie jest tożsamy z rokiem zdeponowania pierwszego obiektu. Od oficjalnej premiery mija zazwyczaj rok lub dwa – taką prawidłowość zauważono dla części analizowanego zbioru danych. Opóźnienie w upowszechnianiu zbiorów sięga jednak nawet 6 lat. Przesunięcia czasowe zaobserwowano w 37 na 115 bibliotek (stan na koniec 2015 r.).

W wytypowanym do badań zbiorze danych, liczącym 1,6 mln rekordów uwagę autorów przyciągnęły dwa pola opisu rekordów: data powstania utworu (*dc:Creation_Date*) i data zdeponowania obiektu w danej bibliotece cyfrowej (*dc:Issue_Date*). Różnice wartości dat deponowania były niewielkie: mieściły się w 15 latach (okres funkcjonowania bibliotek cyfrowych). Natomiast w przypadku dat powstania utworów zakres wartości sięgał nawet kilku tysięcy. Sytuacja taka nie dziwi bowiem w bibliotekach cyfrowych udostępnia się różne rodzaje obiektów: starodruki, starożytne kroniki i inne archiwalia, a sam proces ich digitalizacji podyktowany jest wieloma czynnikami: specyfiką kolekcji, kwestią praw autorskich, wyposażeniem pracowni digitalizacyjnej (o ile taka istnieje), funduszami przeznaczonymi na ten cel, polityką zarządzania zasobami instytucjonalnymi, stanem zachowania obiektu analogowego, a wreszcie zapotrzebowaniem naukowym. Cyfryzacja zbiorów często prowadzona jest też *ad hoc*.

W związku z różnicami występującymi pomiędzy datą ukazania się analogowej publikacji a datą jej zdeponowania w bibliotece cyfrowej autorom nasunęły się trzy pytania: 1. jak kształtują się różnice pomiędzy wymienionymi datami dla każdej biblioteki?, 2. czy można znaleźć prawidłowości w rozkładzie czasowym odnoszące się do kilku lub kilkudziesięciu obiektów, 3. czy uzyskane wyniki mogą charakteryzować konkretny typ biblioteki wyróżniony przez eksperta?

Autorzy postanowili przyjrzeć się dokładnie różnicy wartości wcześniej wymienionych pól opisu bibliograficznego (czas powstania i czas zdeponowania utworu). Różnicę tę nazwali **czasem uwolnienia zasobów** (CUZ). Główną hipotezą badawczą było stwierdzenie, że czas uwolnienia zasobów nie jest przypadkowy i może wskazywać na konkretny typ biblioteki cyfrowej. Hipoteza ta zarysowała przed autorami dwa kolejne pytania: w jakim stopniu polityka digitalizacji jest uzależniona od obowiązujących w kraju praw autorskich, a w jakim stopniu od treści dokumentów?

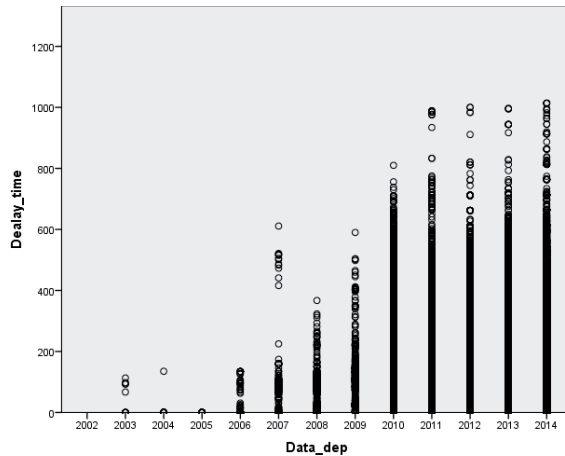
Prawa autorskie do utworu wygasają 70 lat po śmierci autora lub od daty upowszechnienia w przypadkach określonych przez *Ustawę z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych z późniejszymi zmianami* (art. 36). Jeśli autor żyje, to należy podpisać z nim (lub innym właścicielem praw) umowę licencyjną na upowszechnienie utworu w Internecie. Ze współcześnie wydanymi utworami często wiążą wątpliwości prawne wynikające z tego, że autor podpisał umowy z innymi podmiotami przenosząc na nie prawa majątkowe, ale do końca nie wie, w jakim zakresie to zrobił (Bednarek-Michalska 2014b). Umów dawno już nie ma, nie istnieją także wydawcy, którzy je zawierali, nie ma zatem pewności prawnej do kogo utwór należy. Mimo że można przyjąć zasadę, że przed rokiem 2001 nie było pola eksploatacji „Internet”, zatem autor nie mógł oddać praw na tym polu, to bibliotekarze nie są pewni tej interpretacji i wolą nie ryzykować. Takich i innych przypadków jest wiele. Pewność można mieć natomiast np. co do tego, że prawa autorskie do czasopism przedwojennych już wygasły i dzięki temu, że przeszły one do domeny publicznej, chętnie są skanowane i upowszechniane. Dowodem takiego stanu rzeczy jest ich przewaga w FBC. Ponadto czasopisma wykorzystywane są do badań naukowych i zwykle wymagają natychmiastowej decyzji o digitalizacji ze względu na zły stan zachowania.

W związku z takimi przypadkami można się spodziewać, że najwięcej źródeł będzie właśnie pochodzić z domeny publicznej i reprezentować stare czasopisma, książki oraz inne druki. Należy tu zaznaczyć, iż stare gazety i czasopisma są poddawane cyfryzacji w pierwszej kolejności ze względu na szybką degradację „kwaśnego” papieru, na którym je wydrukowano. Digitalizacja starodruków jest kosztowniejsza przez wymóg specjalistycznej konserwacji i zabezpieczenia formy analogowej. To wszystko powoduje, że proces digitalizacji jest wolniejszy i bardziej skomplikowany.

4. Wyniki analiz

Aby wyznaczyć strategię analiz wizualnych autorzy potrzebowali informacji o rozrzucie wartości kluczowej zmiennej – czasu uwolnienia zasobów. Na rysunku 6 zamieszczono wykres punktowy CUZ dla poszczególnych lat powstania bibliotek cyfrowych w Polsce.

Rysunek 6. Rozrzut wartości czasu uwolnienia zasobów względem dat deponowania zasobów

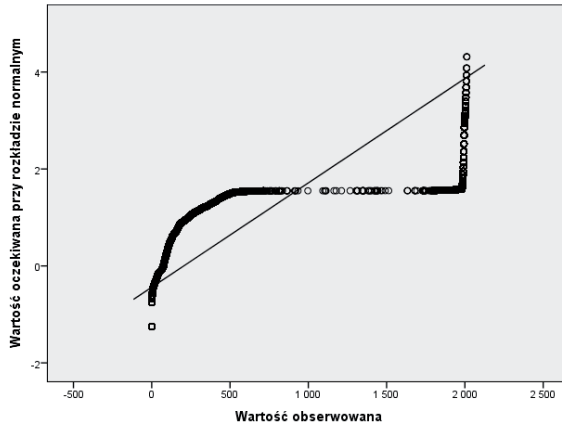


Źródło: opracowanie własne na podstawie danych FBC.

Jak ilustruje wyżej zamieszczony wykres, maksymalne wartości sięgają ponad tysiąc lat. W kolekcji można znaleźć rękopisy pergaminowe z XI–XII wieku, pochodzące z terenów Asyrii, opisy fragmentów architektonicznych z Bliskiego Wschodu, poematy z późnego okresu Imperium Rzymskiego. Na wykresie rozpiętość wartości CUZ kształtuje się od zera do tysiąca i co jest charakterystyczne, wzrasta z roku na rok. Może to świadczyć o przyspieszeniu procesu tworzenia kolekcji cyfrowych w skali całego kraju i rozwoju cyfryzacji. Warto nadmienić, że przy obliczaniu dużych wartości czasu uwolnienia zasobów zdarzały się błędy ze względu na problem z interpretacją zawartości pola „data powstania” (*dc:Creation_date*). Brakowało w nim oczekiwanej wartości albo jego treść nie odnosiła się do daty powstania obiektu. W takim przypadku pomocniczo przeszukano pola „opis” (*dc:Description*) lub „autor” (*dc:Creator*), gdzie występowała data utworzenia obiektu albo wiek twórcy. W wyniku tych działań odfiltrowano blisko 7 tys. rekordów z niejednoznacznie opisaną datą. Dlatego duże wartości CUZ są obciążone błędem rzędu 0,4%.

Dla obalenia głównej hipotezy wykonano test normalności rozkładu CUZ (rys. 7). Wykazał on, iż czas uwolnienia zasobów nie jest przypadkowy. Poza tym, prawa część wykresu – gdzie wartości CUZ sięgają powyżej 1000 – świadczy o działaniu nie podlegającym prawom statystycznym. Ich pochodzenie autorzy wiążą z ludzkim błędem przy wprowadzeniu danych. Ta obserwacja pomogła w wyłonieniu błędnych rekordów w odniesieniu do daty utworzenia utworu. Ten problem ostatecznie rozwiązano poprzez przeszukiwanie wartości daty w sąsiadujących polach.

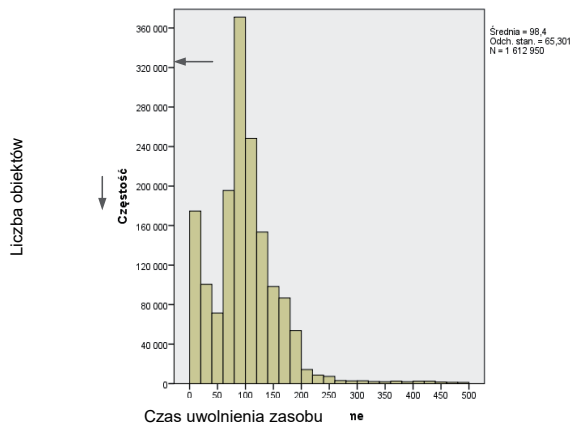
Rysunek 7. Rozkład normalności czasu uwolnienia zasobów



Źródło: opracowanie własne na podstawie danych FBC.

Na kolejnym rysunku (rys. 8) przedstawiono sumaryczny dla wszystkich bibliotek cyfrowych histogram CUZ.

Rysunek 8. Histogram dla zmiennej czasu uwolnienia zasobów w polskich bibliotekach cyfrowych

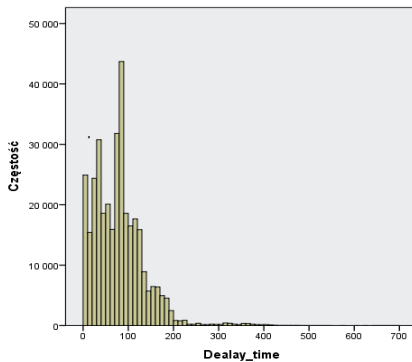


Źródło: opracowanie własne na podstawie danych FBC.

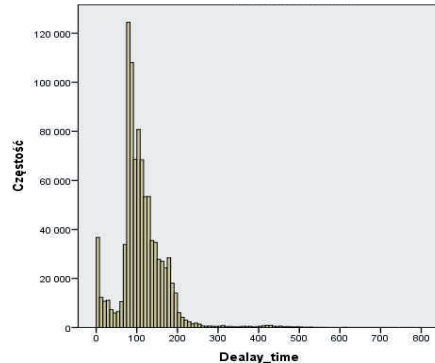
Występujące w rozkładzie dwa charakterystyczne punkty (maksima) przypadające na wartości 0 oraz 80–100 lat, są reprezentowane przez współczesne źródła piśmiennicze, np. artykuły, czasopisma i książki oraz w większości przez XIX-wieczne i przedwojenne materiały. Dla lepszego zrozumienia histogramu, maksima te w dalszym opisie wyników oznaczone zostały odpowiednio jako A i B.

Przyjrzenie się ponad stu wykresom – przygotowanym przez autorów dla poszczególnych bibliotek cyfrowych – wykazało, że charakterystyczne zagłębienie pomiędzy punktami najwyższymi występuje w profilu histogramów wszystkich bibliotek cyfrowych z bazy FBC. Ponieważ zamieszczenie wszystkich wykresów przekracza możliwości edycyjne niniejszego opracowania, naturalną strategią będzie zatem przeanalizowanie czy punkty oznaczone jako A i B są zbalansowane i co wpływa na zmianę w równowadze pomiędzy nimi? W tym celu wykonano skumulowane histogramy dla zdefiniowanych wcześniej grup bibliotek cyfrowych, tym samym znacznie redukując zakres zmiennej nominalnej, która jest reprezentowana przez każdą kolekcję. Zestawienie tych wykresów przedstawiono w układzie tabelarycznym na rysunku 9.

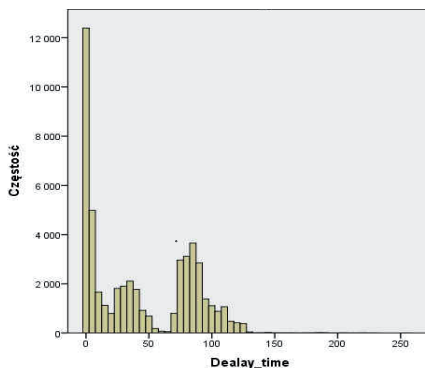
Rysunek 9. Histogramy rozkładu czasu uwolnienia zasobów dla różnych grup polskich bibliotek cyfrowych



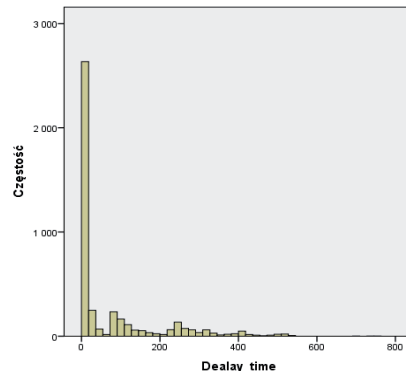
biblioteki regionalne



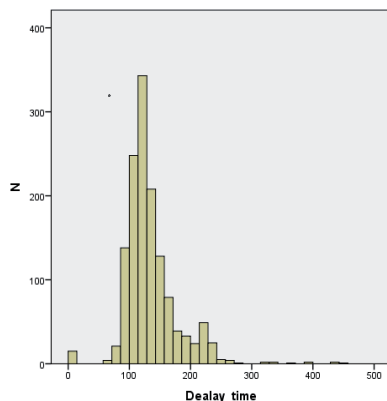
biblioteki uniwersalne



biblioteki techniczne



biblioteki kościelne



biblioteki naukowe

Na osi poziomej odłożona jest nowa zmienna – czas uwolnienia zasobów, na osi pionowej – liczba obiektów dla określonego CUZ.

Źródło: opracowanie własne na podstawie danych FBC.

Dzięki wizualnej analizie zauważono podobieństwa w kolekcjach dla bibliotek technicznych i kościelnych: dominujący punkt najwyższy A w skali całego wykresu dowodzi nadreprezentacji w digitalizowanym zasobie dzieł współczesnych¹⁵ i relatywnie niewielkiej ilości dzieł przedwojennych.

Dla pierwszej grupy bibliotek obok zerowego CUZ pojawia się również sporo obiektów stworzonych przed 20–50 laty. Może to świadczyć o tym, że bibliotekarze pozyskali zgodę od autorów lub innych właścicieli praw na upowszechnienie nowszych prac w Internecie. W przypadku bibliotek kościelnych zasobem dominującym są z kolei czasopisma własne.

Słabo zaznaczony dołek (punkt niższy) pomiędzy punktami A i B dla grupy bibliotek regionalnych może wskazywać na to, że ich polityka digitalizacji w małym stopniu oparta jest na pozyskiwaniu licencji od autorów. Regionalne biblioteki cyfrowe kolekcjonują najczęściej „stare” obiekty opisujące region z wielu perspektyw, a służące mieszkańcom do poszerzania wiedzy o tym regionie. Obiekty współczesne – stanowiące stosunkowo duży odsetek zasobów regionalnych bibliotek cyfrowych – trafiają do nich, dlatego że bibliotekarze mają bezpośredni kontakt z autorami i często zachęcają do uwolnienia dzieł.

Histogramy grupy bibliotek uniwersalnych ukazują najdłuższy zakres czasowy, gdzie można dostrzec charakterystyczne „garby” (300, 420 lat), wskazujące na obecność w kolekcji starodruków. Jak widać, biblioteki te w dużej mierze wykorzystują także dzieła przedwojenne (wyraźny punkt B).

¹⁵ Dla przykładu biblioteka kościelna FIDES posiada w swoim zasobie najwięcej pełnych tekstów artykułów z czasopism współczesnych bądź całych zeszytów tychże czasopism. <http://digital.fides.org.pl/dlibra>

Przeprowadzone analizy dowodzą, że strategia digitalizacyjna jest silnie uzależniona od praw autorskich: masowo udostępnia się obiekty pozbawione ochrony autorskoprawnej, z nowszymi natomiast pojawiają się problemy. Ponieważ w większości placówkami digitalizującymi są ośrodki uniwersyteckie, można przypuszczać, że stosunkowo duży udział zerowego CUZ związany jest z uruchamianymi w tych ośrodkach repozytoriami, gdzie deponowane są prace współczesne. Grupa bibliotek naukowych ma podobny kształt histogramu do uniwersalnych z tym, że punkt A ma tu wymiar znikomy. Biblioteki techniczne – tworzone przez politechniki polskie – zawierają zwykle współczesne prace (często dorobek własnych pracowników) oraz w wyborze publikacje starsze, w tym głównie czasopisma techniczne z XX i XIX wieku. Brak w tych bibliotekach nowszych czasopism technicznych związany jest z coraz częstszym upowszechnianiem tych czasopism w ramach akademickich platform czasopism czy też repozytoriów instytucjonalnych, niekoniecznie zaś w ramach bibliotek cyfrowych.

5. Konkluzje oraz kierunki dalszych działań

Wykonawcy opisanego projektu badawczego skoncentrowali się na udzieleniu odpowiedzi na pytanie, jak się zmieniają się w czasie polskie repozytoria i biblioteki cyfrowe. Wykorzystali do tego nowoczesne metody wizualizacji danych dużego wolumenu, zaś proces wizualizacji podzielili na pięć etapów: gromadzenia, czyszczenia i przetwarzania danych, mapowania i wizualizacji oraz końcowego projektowanie graficznego (Osińska 2016).

Pomimo tego, że FBC stosuje standaryzację Dublin Core, autorzy już na etapie czyszczenia danych spotkali się z licznymi problemami, do których zaliczyli: wprowadzenie błędnych wartości do poszczególnych pól opisu bibliograficznego, stosowane różnych nazw atrybutów podstawowych rekordów oraz brak wypełnienia niektórych pól opisu¹⁶.

W przypadku dat publikacji do konwersji różnych form zapisu daty na jeden wspólny format wykorzystano wyrażenia regularne (Osińska i Malak 2015). Odczytane wartości dat utworzenia i zdeponowania zasobu wymagały głębszego przyjrzenia się ich różnicy, dlatego autorzy swoje działania skupili na analizie wyników w perspektywie czasowej, czyli z wykorzystaniem histogramów.

Różnicę, czyli czas który upływa od powstania dzieła do jego publikacji w Internecie, autorzy nazwali czasem uwolnienia zasobu, który zbadali dla każdej z ponad stu bibliotek cyfrowych. Jak dowiodły badania, polityka cyfryzacji zbiorów bibliotecznych zależy od wielu czynników. Z pewnością są to: stan techniczny pracowni, posiadane prawa autorskie do utworów, fundusz docelowy, stan zachowania obiektów analogowych, jak również zapotrzebowanie naukowe. Autorzy postanowili

¹⁶ Na podobne błędy w interpretacji zawartości poszczególnych pól Dublin Core uwagę zwracali już inni autorzy, m.in. (Potęga, Wróbel 2010; Pulikowski 2009).

zbadać, czy rozkład CUZ jest przypadkowy, czy wręcz odwrotnie – wykazuje prawidłowości, wskazujące na wspólną politykę digitalizacji dla rozmaitych typów bibliotek cyfrowych.

Wykres skumulowany CUZ dla całej kolekcji bibliotek cyfrowych wykazał dwa charakterystyczne maksima, które oznaczono jako punkty A i B (rys. 8). Pierwszy z nich określił obecność współczesnych zbiorów piśmienniczych, drugi – kolekcjonowanie XIX-wiecznych i przedwojennych materiałów.

Histogramy czasowe dla bibliotek pogrupowanych przez eksperta według kryterium organizacyjno-tematycznego prezentowały podobnie dwa maksyma (A i B), względne proporcje których poddane zostały analizie wizualnej. Dominacja punktu A wskazywała na nadreprezentację dzieł współczesnych, wyższy punkt B – na dużą ilość obiektów przedwojennych, nie obciążonych prawami autorskimi.

W kształcie CUZ dla wytypowanych grup zaobserwowano bardzo charakterystyczne różnice, które dało się wyjaśnić dzięki wiedzy eksperckiej. Tym samym główna hipoteza badawcza, że czas uwolnienia zasobów nie jest przypadkowy, lecz związany z polityką digitalizacyjną biblioteki cyfrowej określonego typu, została potwierdzona.

Novum zaprezentowanego podejścia polega na tym, iż z powodzeniem wykorzystano parametr dynamiki rozkładu, oparty na wartościach dat w roboczej bazie danych. Analiza piśmiennictwa fachowego (przeprowadzona na podstawie zawartości bazy *Library, Information Science & Technology Abstracts*) dowodzi, że ani w badaniach polskich, ani zagranicznych badacze pole daty nie było dotąd przedmiotem pogłębionych studiów nad oceną rodowodu i zawartości treściowej bibliotek cyfrowych. Najczęściej atrybut ten poddawano dyskusji w kontekście domniemanej zawartości tego pola (np. *Metadane, zagadnienia słowników kontrolowanych* 2012) i błędów w nim się pojawiających (np. Werla 2010) czy dowodzenia, że nie stanowi on istotnego elementu podczas inicjowania procesu wyszukiwania (np. Costello 2016). Od strony formalnej pole daty zdeponowania zasobu w bibliotece cyfrowej (*date of creation*) jest użyteczne w kontekście przechowywania informacji o tym, ile lat obiekt stworzony przez nieznanego twórcę albo osobę o nieznanym dacie śmierci będzie jeszcze chroniony prawem autorskim (120 lat minus rok zdeponowania) (Xie, Matusiak 2016).

Dane umieszczane w polu *data* wykorzystywano również do zobrazowania efektywności narzędzi wizualizacyjnych (np. Marks, Hussell i in. 2005) oraz mechanizmów działania procesu filtrowania rezultatów wyszukiwania w konkretnych repozytoriach cyfrowych (np. Kostkova, Madle 2013).

Autorzy są świadomi, że z pewnością otrzymaliby większy poziom precyzji omawianego wskaźnika, gdyby rozszerzyli go o bardziej złożony aparat matematyczny, uwzględniający nieliniowość – wszak można poczynić założenie, że wzrost zasobów cyfrowej wiedzy, podobnie jak w każdym układzie dynamicznym, musi być nieliniowy.

Jak wspomniano, polskie biblioteki cyfrowe tworzą niejednolity i rozproszony system. Zaproponowana metoda oparta na charakterystykach czasowych może być krokiem w rozwoju systematyki instytucji digitalizujących (mających krótką, 10-letnią historię), jak również wspomóc monitorowanie dynamiki wzrostu krajowych zasobów cyfrowej wiedzy.

W przyszłości autorzy planują zweryfikowanie zależności kształtu rozkładu czasowego od typu instytucji na danych pochodzących z zagranicznych bibliotek cyfrowych. Dalsze prace kierują na stworzenie automatycznej klasyfikacji dokumentów (wykorzystując metadane) według dziedzin naukowych i odfiltrowanie dokumentów w zakresie nauk humanistyczno-społecznych (Osińska i Małak 2015). Zebrane dane planuje się poddać analizom pod względem współautorstwa, podejmowanej tematyki i zależności czasowych powstawania publikacji. Zakłada się również przygotowanie reprezentacji graficznych wyników i ich porównanie z odpowiednikami wygenerowanymi w oparciu o dane pochodzące z zagranicznych serwisów sieciowych.

Aneks 1

Wypracowana koncepcja podziału polskich bibliotek cyfrowych

Nazwa biblioteki	Typ zasobu
Cyfrowa Biblioteka Druków Ulotnych Polskich i Polski Dotyczących z XVI	specjalistyczne
Armarium. dominikańska biblioteka cyfrowa	kościelne
Biblioteka Cyfrowa Diecezji Legnickiej	kościelne
Cyfrowa Biblioteka Diecezjalna w Sandomierzu	kościelne
Księgozbiór Wirtualny Federacji Bibliotek Kościelnych FIDES	kościelne
Rolnicza Biblioteka Cyfrowa	naukowe
Repozytorium Centrum Otwartej Nauki	naukowe
Computer Science Journal	naukowe
Repozytorium eRIKA	naukowe
Biblioteka Cyfrowa Instytutu Ekologii Terenów Uprzemysłowionych	naukowe
Biblioteka Cyfrowa Katedry Lingwistyki Formalnej Uniwersytetu Warszawskiego	naukowe
Biblioteka Cyfrowa Małopolskiego Towarzystwa Genealogicznego	naukowe
Biblioteka Humanistyczna	naukowe
Otwórz książkę	naukowe
Pedagogiczna Biblioteka Cyfrowa	naukowe
Biblioteka Cyfrowa Polskiego Instytutu Antropologii	naukowe
Repozytorium Cyfrowe Poloników	naukowe
Repozytorium Uniwersytetu Kazimierza Wielkiego w Bydgoszczy	naukowe
Baza Rozpraw Doktorskich Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach	przyrodnicze
Biblioteka Cyfrowych Archiwów Tradycji Lokalnej	regionalne
Cyfrowy Dolny Śląsk	regionalne
Łąwska Biblioteka Cyfrowa	regionalne
Podlaska Biblioteka Cyfrowa	regionalne
Biała Biblioteka Cyfrowa	regionalne
Chełmska Biblioteka Cyfrowa	regionalne
Elbląska Biblioteka Cyfrowa	regionalne
Tarnowsko-Galicyjska Biblioteka Cyfrowa	regionalne
Gdańska Biblioteka Cyfrowa	regionalne
Nowohucka Biblioteka Cyfrowa	regionalne
Inowrocławska Biblioteka Cyfrowa	regionalne
Świętokrzyska Biblioteka Cyfrowa	regionalne
Kolbuszowska Biblioteka Cyfrowa	regionalne
Kędzierzyńsko-Kozielska Biblioteka Cyfrowa	regionalne

Krośnieńska Biblioteka Cyfrowa	regionalne
Biblioteka Cyfrowa –	regionalne
Biblioteka Cyfrowa Wojewódzkiej Biblioteki Publicznej im. Hieronima Łopacińskiego w Lublinie	regionalne
Opolska Biblioteka Cyfrowa	regionalne
Płocka Biblioteka Cyfrowa	regionalne
Radomska Biblioteka Cyfrowa	regionalne
Podkarpacka Biblioteka Cyfrowa	regionalne
Sanocka Biblioteka Cyfrowa	regionalne
Cyfrowa Ziemia Sieradzka	regionalne
Bałtycka Biblioteka Cyfrowa	regionalne
Zachodniopomorska Biblioteka Cyfrowa	regionalne
Tarnobrzeska Biblioteka Cyfrowa	regionalne
Tarnowska Biblioteka Cyfrowa	regionalne
Polonijna Biblioteka Cyfrowa	regionalne
Małopolska Biblioteka Cyfrowa	regionalne
Mazowiecka Biblioteka Cyfrowa	regionalne
Pomorska Biblioteka Cyfrowa	regionalne
Bibliologiczna Biblioteka Cyfrowa	specjalne
Biblioteka Cyfrowa Fundacji Ormiańskiej KZKO	specjalne
Czytelnia kwartalnika „	specjalne
Wydawnictwa EBIB	specjalne
Biblioteka Cyfrowa Ośrodka KARTA	specjalne
Wolne Lektury	specjalne
Muzeum Leszno	specjalne
Cyfrowe Muzeum Narodowe w Warszawie	specjalne
Cyfrowe archiwum KK NSZZ Solidarność	specjalne
Biblioteka Cyfrowa ORE	specjalne
Wirtualne Archiwum Polskich Ormian	specjalne
Publiczne Archiwum Cyfrowe Agnieszki Osieckiej	specjalne
PAUart Katalog zbiorów artystycznych i naukowych Polskiej Akademii Umiejętności	specjalne
Polska Biblioteka Internetowa	specjalne
Zofia Rydet	specjalne
Statystyczna Biblioteka Cyfrowa	specjalne
Biblioteka Multimedialna Teatru NN	specjalne
Dziennik Ustaw RP na Uchodźstwie	specjalne
Wypędzeni 1939	specjalne

Internetowa Biblioteka Wzornictwa	specjalne
Biblioteka Cyfrowa Zachodniopomorskiego Centrum Doskonalenia Nauczycieli	specjalne
Ziemiańska Biblioteka Cyfrowa	specjalne
Akademicka Biblioteka Cyfrowa KRAKÓW	techniczne
Biblioteka Cyfrowa Politechniki Warszawskiej	techniczne
Repozytorium Politechniki Krakowskiej	techniczne
Repozytorium Instytutu Biochemii i Biofizyki Polskiej Akademii Nauk	techniczne
Biblioteka Cyfrowa Instytutu Geodezji i Kartografii	techniczne
Biblioteka Cyfrowa Instytutu Nafty i Gazu	techniczne
Biblioteka Cyfrowa Instytutu Odlewnictwa	techniczne
Biblioteka Cyfrowa Instytutu Łączności	techniczne
Repozytorium Instytucjonalne PCSS	techniczne
Biblioteka Cyfrowa Politechniki Lubelskiej	techniczne
Biblioteka Cyfrowa Politechniki Śląskiej	techniczne
Biblioteka Cyfrowa Politechniki Koszalińskiej	techniczne
Repozytorium Cyfrowe UTP w Bydgoszczy	techniczne
Repozytorium Uniwersytetu im. Adama Mickiewicza	uniwersalne
Biblioteka Cyfrowa Uniwersytetu Łódzkiego	uniwersalne
Biblioteka Cyfrowa Uniwersytetu Śląskiego	uniwersalne
Biblioteka Otwartej Nauki	uniwersalne
Śląska Biblioteka Cyfrowa	uniwersalne
CBN Polona	uniwersalne
Zielonogórska Biblioteka Cyfrowa	uniwersalne
Łódzka Regionalna Biblioteka Cyfrowa	uniwersalne
Dolnośląska Biblioteka Cyfrowa	uniwersalne
e-biblioteka Uniwersytetu Warszawskiego	uniwersalne
ICM	uniwersalne
Jagiellońska Biblioteka Cyfrowa	uniwersalne
Kujawsko-Pomorska Biblioteka Cyfrowa	uniwersalne
Biblioteka Cyfrowa KUL	uniwersalne
Repozytorium Cyfrowe Instytutów Naukowych	uniwersalne
Biblioteka Cyfrowa UMCS	uniwersalne
Repozytorium Uniwersytetu Mikołaja Kopernika	uniwersalne
Repozytorium Uniwersytetu Łódzkiego	uniwersalne
Biblioteka Cyfrowa Uniwersytetu Warmińsko-Mazurskiego	uniwersalne
Biblioteka Cyfrowa Uniwersytetu Wrocławskiego	uniwersalne
Wielkopolska Biblioteka Cyfrowa	uniwersalne

Literatura

- Bednarek-Michalska, B., 2014a, *Otwarte zasoby naukowe i edukacyjne na Uniwersytecie Mikołaja Kopernika*, w: Odlanicka-Poczobutt M., Ziolo K., red., *Biblioteka Akademicka: infrastruktura, uczelnia, otoczenie*, Gliwice: Wydawnictwo Politechniki Śląskiej, 65–75.
- Bednarek-Michalska, B., 2014b, *Prawo autorskie i jego ograniczenia dla polskich bibliotek cyfrowych*, w: Puławski A. *Znaczenie udostępniania kopii cyfrowych regionalnych zbiorów bibliotecznych w sieci*, Stargard Szczeciński: Książnica Stargardzka, 61–72, 51–60, Dokument internetowy: <http://www.stralsund-stargard.eu/panel/images/540070e85f678.pdf> [dostęp: 30.06.2016].
- Börner, K., 2015, *Everyone can map*, Cambridge: The MIT Press.
- CDL GDO, 2011, *CDL Guidelines for Digital Objects*, Dokument internetowy: http://www.cdlib.org/services/access_publishing/dsc/contribute/docs/GDO.pdf [dostęp: 30.09.2016].
- Chen, Ch., 2006, *Information Visualization. Beyond the Horizon*, London: Springer.
- Costello, L., 2016, *Title, Description, and Subject are the Most Important Metadata Fields for Keyword Discoverability*, „Evidence Based Library & Information Practice”, 11.3: 88–90.
- Franke, J., 2007, *Googletheca Universalis?*, w: *Biblioteki cyfrowe. Projekty, realizacje, technologie*, red. J. Woźniak-Kasperek i J. Franke, „Nauka, Dydaktyka, Praktyka” 87: 121–166.
- Garfield, E., 1994, *Scientigraphy: Mapping the tracks of science*, „Current Contents: Social & Behavioural Sciences”, 7: 45.
- Głowacka, E., 2011, *Jakość bibliotek cyfrowych – aspekty i kryteria oceny „e-mentor”*, 2: 39, Dokument Internetowy: <http://www.e-mentor.edu.pl/artikul/index/numer/39/id/828> [dostęp: 30.06.2016].
- Janiak, M., Krakowska, M., 2010, *Ocena bibliotek cyfrowych – kryteria jakości*, w: *Polskie Biblioteki Cyfrowe*, 101–108.
- Janiak, M., Krakowska, M., 2012, *Ocena jakości bibliotek cyfrowych*, w: Janiak M., Krakowska M. i Próchnicka M., red., *Biblioteki cyfrowe*, Warszawa, 135–173.
- Jindal, S.C., Lakshmi, V., 2004, *Digital Libraries: Issues of Digital Environment Vol. 3*, Delhi: Isha Books, 215–266.
- Kolasa, M., 2012, *Biblioteki cyfrowe na świecie. Analiza i rozwój*, w: Janiak M., Krakowska M. i Próchnicka M., red., *Biblioteki cyfrowe*, Warszawa, 66–95.
- Kostkova, P., Madle, G., *What impact do healthcare digital libraries have? An evaluation of national resource of infection control at the point of care using the Impact-ED framework*, 2013, „International Journal on Digital Libraries” 13.2: 77–90.
- Kowalska, M., Jarocki, M., 2010, *Czasopisma polskich bibliotek naukowych w zasobach Federacji Bibliotek Cyfrowych i bibliograficznych bazach danych*, w: *Czasopisma naukowe bibliotek, archiwów, muzeów. Tradycje – role – perspektywy*, red. A. Królczyk, Kórnik: Biblioteka PAN, 27–39.
- Kowalska, M., 2006, *Digitalizacja zbiorów w bibliotekach polskich – próba oceny doświadczeń krajowych*, Biuletyn EBIB 11.81, Dokument internetowy: <http://www.ebib.pl/2006/81/a.php?kowalska> [dostęp: 30.11.2016].
- Kowalska, M., 2007, *Dygitalizacja zbiorów bibliotek polskich*, Warszawa: SBP.
- Kowalska, M., 2010, *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce na lata 2009–2020. Warszawa 2009*, „Przegląd Biblioteczny” 1: 104–111.
- Kowalska, M., 2012, *Koszty i finansowanie digitalizacji*, w: Janiak M., Krakowska M. i Próchnicka M., red., *Biblioteki cyfrowe*, Warszawa, 233–256.
- Marks, L., Hussell, J.A.T., McMahon, T.M., Luce, R.E., 2005, *ActiveGraph: A digital library visualization tool*, „International Journal on Digital Libraries” 5.1: 57–69.
- Malak, P., 2012, *Indeksowanie treści*, Warszawa: SBP.

- Mazurek, C., Parkola, T., Werla, M., 2012, *Tworzenie i rozwój bibliotek cyfrowych*, w: Janiak M., Krakowska M. i Próchnicka M., red., *Biblioteki cyfrowe*, Warszawa, 339–341.
- Mazurek, C., Werla, M., 2012, *Federacja Bibliotek Cyfrowych – studium przypadku*, w: Janiak M., Krakowska M. i Próchnicka M., red., *Biblioteki cyfrowe*, Warszawa, 449–467.
- Gaweł, Ł., red., 2010, *Metadane, zagadnienia słowników kontrolowanych*, Warszawa: Narodowy Instytut Muzealnictwa i Ochrony Zbiorów.
- Osińska, V., 2010a, *Rozwój metod mapowania domen naukowych i potencjał analityczny w nim zawarty*, „Zagadnienia Informacji Naukowej”, 2.96.
- Osińska, V., 2010b, *Wizualizacja i wyszukiwanie dokumentów*, Warszawa: SBP.
- Osińska, V., 2012, *Wizualizacja paradygmatów w nauce*, „Zagadnienia naukoznawstwa”, 48.193.
- Osińska, V., 2016, *Wizualizacja informacji. Studium informatologiczne*, Toruń: UMK, 6.
- Osińska, V., Malak, P., 2015, *Maps and Mapping in Scientometrics*, Materiały konferencyjne *Tools and Methods for Analysing the Scientific Literature and Readers*, WUW (w druku).
- Potęga, J., Wróbel, A., 2010, *The Dublin Core Metadata Element Set, Ver. 1.1 a potrzeby i oczekiwania bibliotekarzy cyfrowych – analiza przypadków*, w: *Polskie biblioteki cyfrowe 2009. Materiały z konferencji*, red. C. Mazurek, M. Stroiński, J. Węglarz, Poznań, 71–78.
- Pulikowski, A., 2009, *Wyszukiwanie pełnotekstowe w zasobach bibliotek cyfrowych* [prezentacja], w: *X Krajowe Forum Informacji Naukowej i Technicznej, Zakopane, 22–24.10.2009*. Dokument internetowy: <http://www.ptin.org.pl/konferencje/10forum/repozytorium/Pulikowski.pdf> [dostęp: 30.09.2016].
- MKiDN, 2009, *Raport o digitalizacji dóbr kultury: Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020*, Warszawa: Ministerstwo Kultury i Dziedzictwa Narodowego, Dokument Internetowy: [http://www.kongreskultury.pl/library/File/RaportDigitalizacja/Program digitalizacji 2009-2020.pdf](http://www.kongreskultury.pl/library/File/RaportDigitalizacja/Program%20digitalizacji%202009-2020.pdf) [dostęp: 16.11.2009].
- Xie, I., Matusiak, K., 2016, *Discover Digital Libraries: Theory and Practice*, Amsterdam: Elsevier, 43–47.
- Werla, M., 2010, *Metadane dokumentów w bibliotekach cyfrowych*, Repozytorium Instytucjonalne PCSS. Dokument internetowy: <http://lib.psnk.pl/Content/284/CPI-Werla.pdf> [dostęp: 30.11.2016].
- Żernicka, K., 2015, *Ocena jakości polskich bibliotek cyfrowych*. Rozprawa doktorska, Repozytorium UMK, Dokument Internetowy: <https://repozytorium.umk.pl/handle/item/2384> [dostęp: 30.11.2016].