

Przyjaciele bibliotek cyfrowych 2017

Raport z analizy źródeł ruchu w wybranych polskich bibliotekach cyfrowych

Opracowanie:

Marcin Werla (mwerla@man.poznan.pl), luty 2018 r.

Dział Bibliotek Cyfrowych i Platform Wiedzy, Pion Usług Sieciowych PCSS



[POZNAŃSKIE CENTRUM SUPERKOMPUTEROWO-SIECIOWE](http://www.pcss.edu.pl)

Spis treści

1. Wprowadzenie	2
2. Analiza udziału poszczególnych kanałów ruchu	4
3. Serwisy kierujące ruch do bibliotek cyfrowych	7
Agregatory	9
Serwisy edukacyjne/encyklopedie	10
Serwisy genealogiczne	10
Katalogi/biblioteki	11
Serwisy hobbystyczne	12
Serwisy naukowe / branżowe	13
Wyszukiwarki naukowe	14
4. Podsumowanie	15
Załącznik nr 1. Źródła ruchu w portalu FBC	16

1. Wprowadzenie

Celem niniejszego raportu jest przedstawienie wyników analizy źródeł ruchu w wybranych polskich bibliotekach cyfrowych. Analizie poddano dane pochodzące z sześciu serwisów:

- Biblioteki Cyfrowej Uniwersytetu Wrocławskiego BC UWr (<http://bibliotekacyfrowa.pl>)
- e-Biblioteki Uniwersytetu Warszawskiego e-bUW (<http://ebuw.uw.edu.pl>)
- Małopolskiej Biblioteki Cyfrowej MBC (<http://mbc.malopolska.pl>)
- Repozytorium Cyfrowego Instytutów Naukowych RCIN (<http://rcin.org.pl>)
- Śląskiej Biblioteki Cyfrowej ŚBC (<https://sbc.org.pl>)
- Wielkopolskiej Biblioteki Cyfrowej WBC (<http://www.wbc.poznan.pl>)

Wszystkie te serwisy są, jak na polskie warunki, dużymi bibliotekami cyfrowymi i istnieją co najmniej od kilku lat. Trzy z nich (BC UWr, e-bUW, RCIN) mają charakter instytucjonalny, trzy pozostałe (MBC, ŚBC, WBC) są regionalnymi bibliotekami cyfrowymi. Wszystkie działają w oparciu o oprogramowanie dLibra rozwijane w PCSS.

W tabeli poniżej przedstawiono podstawowe porównanie numeryczne analizowanych serwisów, wzbogacone dodatkowo o dane dotyczące największej polskiej biblioteki cyfrowej: Polona.pl.

Biblioteka cyfrowa	Liczba obiektów on-line na koniec 2017 r.	Liczba odwiedzin wg Google Analytics w 2017 r.	Liczba odwiedzin wg oprogramowania biblioteki cyfrowej w 2017 r.	Liczba wyświetleń treści publikacji wg oprogramowania biblioteki cyfrowej w 2017 r.
(1)	(2)	(3)	(4)	(5)
BC UWr	ok. 75 000	214 228	brak danych	brak danych
e-bUW	331 476	304 124	2 286 534	1 570 129
MBC	93 851	277 860	2 740 893	1 624 011
RCIN	63 640	253 529	8 538 223	2 338 119
ŚBC	293 031	458 940	16 115 771	3 254 270
WBC	318 682	744 535	7 631 571	4 685 031
Polona.pl	ok. 2 100 000	571 980	brak danych	brak danych

Tabela 1. Dane dotyczące ruchu użytkowników w badanych bibliotekach cyfrowych w 2017 r.

Dane z kolumny (2) pochodzą ze stron poszczególnych bibliotek cyfrowych. Dane z kolumny (3) pochodzą z narzędzia Google Analytics¹. Dane z kolumn (4) i (5) pochodzą ze stron poszczególnych

¹ W przypadku serwisu Polona.pl dane te uzyskano poprzez rządowy serwis widok.gov.pl: <https://widok.gov.pl/statistics/polona-pl/>. W przypadku pozostałych serwisów dane uzyskano bezpośrednio z Google Analytics, za zgodą administratorów poszczególnych serwisów, za co autor niniejszego raportu serdecznie dziękuje.

bibliotek cyfrowych (o ile były dostępne). W przypadku bibliotek cyfrowych opartych o system dLibra statystyki te można znaleźć pod adresem <główny adres serwisu>/stats np.: <http://www.wbc.poznan.pl/stats>.

Duża rozbieżność pomiędzy danymi w kolumnie (3) i (4) wynika z kilku czynników bezpośrednio związanych ze sposobem gromadzenia danych²:

- Google Analytics w standardowej konfiguracji gromadzi dane poprzez kod śledzący Java Script, który osadzany jest w stronach WWW interfejsu biblioteki cyfrowej. W związku z tym:
 - Nie ma możliwości śledzenia w ten sposób bezpośredniego dostępu do plików publikacji, np. poprzez przejście z wyników wyszukiwania Google do pliku PDF w bibliotece cyfrowej, z pominięciem interfejsu WWW tej biblioteki (np. strony z metadanymi). Analizy przeprowadzone z wykorzystaniem innego narzędzia Google – Search Console – pokazują, że zwłaszcza dla serwisów publikujących pliki PDF jest to znacząca część skierowań z wyszukiwarki Google.
 - W przypadku użytkowników korzystających ze stron WWW biblioteki cyfrowej istnieje możliwość blokowania śledzenia przez Google Analytics np. poprzez wykorzystanie wtyczek do przeglądarek WWW (np. Ghostery) lub blokowanie wywołań Java Script.
- Oprogramowanie biblioteki cyfrowej zapisuje do logów każde żądanie użytkowników i nie ma możliwości pominięcia tego kroku.
 - W kontekście ograniczeń śledzenia opartego o Java Script (patrzy wyżej), to przede wszystkim tłumaczy wyraźnie większe liczby odnotowywane po stronie oprogramowania biblioteki cyfrowej.
 - Czynnikiem zaburzającym dane gromadzone po stronie serwera biblioteki cyfrowej jest z kolei ruch generowany przez nietypowe roboty wyszukiwarek lub narzędzia generujące spam. Oprogramowanie biblioteki cyfrowej posiada funkcje eliminowania tego typu ruchu ze statystyk, o ile generujące ruch roboty nie maskują się i nie udają zwykłych użytkowników. Natomiast w przypadku zbierania danych o ruchu przez Google Analytics usunięcie ruchu robotów jest realizowane już po stronie Google, choć również może być niepełne³.

Podsumowując, ruch na stronach internetowych można podzielić w sposób widoczny na Rysunku 1 (poniżej). Przedmiotem analizy omawianej w niniejszym raporcie jest ruch generowany przez ludzi. Ze względu na dostępność odpowiedniej jakości danych, omawiany tu ruch ograniczony jest do zakresu otoczonego na Rysunku 1 czerwoną przerywaną linią. A więc jest to ten ruch generowany przez ludzi, który mógł zostać zmierzony poprzez Google Analytics metodami opartymi o Java Script.

Tak jak wspomniano powyżej, pozostały ruch generowany przez ludzi jest to albo ruch w przypadku którego śledzenie Java Script zostało świadomie zablokowane, albo taki który z zewnątrz prowadził bezpośrednio do udostępnianych plików z pominięciem serwisu WWW biblioteki cyfrowej. Pełna analiza tego ruchu wymagałaby dostępu do logów serwerów poszczególnych bibliotek cyfrowych oraz przygotowania narzędzie przetwarzających te logi do postaci danych analogicznych jak te udostępniane przez Google Analytics.

² Por. <https://piwik.pro/blog/javascript-tracking-web-log-analytics/>

³ Por. <https://datadome.co/exclude-bot-traffic-from-google-analytics/>

		Ruch generowany przez roboty		Ruch generowany przez ludzi	
		Wejścia bezpośrednie	Wejścia z wyszukiwarek	Wejścia z social media	Wejścia z innych stron
Ruch widoczny w logach serwera					
	Ruch widoczny w narzędziach opartych o Java Script				

Rysunek 1. Podział ruchu internetowego ze względu na źródła i sposób śledzenia.

2. Analiza udziału poszczególnych kanałów ruchu

Pierwszym krokiem w analizie ruchu użytkowników korzystających z bibliotek cyfrowych może być porównaniu udziału poszczególnych kanałów sprowadzających ten ruch. Standardowo, jak pokazano na Rysunku 1, ruch ten dzieli się na cztery kanały:

- Wejścia z wyszukiwarek – najczęściej do tego kanału ruchu zaliczane są wejścia z popularnych wyszukiwarek ogólnego przeznaczenia, takich jak Google czy Bing. Kanał ten zazwyczaj dzielony jest na ruch bezpłatny (tzw. organiczny), wynikający z naturalnego indeksowania i pozycjonowania stron w wyszukiwarkach, oraz na ruch płatny, wynikający z opłacania czołowych pozycji w wynikach wyszukiwania (np. Google Adwords). W przypadku analizowanych serwisów ruch płatny nie został odnotowany.
- Wejścia bezpośrednie – są to wejścia, co do których brak jest informacji o stronie poprzedzającej. Może to oznaczać, że użytkownik po prostu znał adres i wpisał go w przeglądarkę WWW, ale też mógł użyć zakładki w przeglądarce czy odnośnika w programie pocztowym. Mógł też zainstalować oprogramowanie utrudniające śledzenie w sieci i zablokować przesyłanie informacji o stronie poprzedzającej.
- Wejścia z innych stron – są to wejście, które nastąpiły poprzez kliknięcie odnośnika do serwisu znajdującego się na jakiejś stronie internetowej, która nie została zakwalifikowana jako wyszukiwarka czy medium społecznościowe.
- Wejścia z social media – są to wejście skierowane z serwisów społecznościowych, takich jak Facebook czy Twitter.

W Tabeli 2 poniżej przedstawiono dane liczbowe obrazujące udział poszczególnych kanałów ruchu w kierowaniu użytkowników do badanych bibliotek cyfrowych w 2017 roku. Dla każdej z bibliotek przedstawiono bezwzględne wartości liczbowe oraz procentowy udział danego kanału w liczbie sesji. Kolorem zielonym oznaczono komórki o największych wartościach procentowych, kolorami od żółtego po czerwony komórki o wartościach najmniejszych.

Biblioteka cyfrowa	Wejścia z wyszukiwarek		Wejścia bezpośrednie		Wejścia z innych stron		Wejścia z social media		Łączna liczba sesji
	Liczba sesji	% łącznej liczby sesji	Liczba sesji	% łącznej liczby sesji	Liczba sesji	% łącznej liczby sesji	Liczba sesji	% łącznej liczby sesji	
BC UWr	61 379	28,7%	121 881	56,9%	27 471	12,8%	3 497	1,6%	214 228
e-bUW	162 833	53,5%	77 000	25,3%	61 173	20,1%	3 118	1,0%	304 124
MBC	153 918	55,4%	64 227	23,1%	51 597	18,6%	8 118	2,9%	277 860
RCIN	144 194	56,9%	61 237	24,2%	43 395	17,1%	4 703	1,9%	253 529
ŚBC	215 666	47,0%	95 002	20,7%	135 077	29,4%	13 195	2,9%	458 940
WBC	361 494	48,6%	208 474	28,0%	165 470	22,2%	9 097	1,2%	744 535
Suma	1 099 484	48,8%	627 821	27,9%	484 183	21,5%	41 728	1,9%	2 253 216

Tabela 2. Udział kanałów ruchu w skierowaniach użytkowników do badanych bibliotek cyfrowych w 2017 roku.

Jak widać, dominującym kanałem kierowania ruchu do bibliotek cyfrowych są wyszukiwarki internetowe ogólnego przeznaczenia. Sprowadzają one średnio niecałą połowę ruchu użytkowników. Pokazuje to, jak ważna dla popularności biblioteki cyfrowej jest dobra widoczność udostępnianych materiałów w wyszukiwarkach. Dokładniejsza analiza skierowań z wyszukiwarek pokazuje, że około 97% ruchu w tym kanale (zależnie od biblioteki cyfrowej) pochodzi z wyszukiwarki Google, na drugim miejscu jest Bing (około 2%). Poszczególne pozostałe wyszukiwarki mają zazwyczaj ułamek procenta całego ruchu w kanale. Jedynie w przypadku WBC proporcje są trochę inne (88% Google, 5.5% Bing, 2.5% Yandex).

Na drugim miejscu wśród kanałów są najczęściej wejścia bezpośrednie, odpowiadające za nieco ponad jedną czwartą ruchu. W Tabeli 3 poniżej przedstawiono analizę tego, gdzie zaczynali swoją wizytę użytkownicy, których wejście zostało zaklasyfikowane jako bezpośrednie.

Biblioteka cyfrowa	Wejścia bezpośrednie na stronę główną		Wejścia bezpośrednie na stronę konkretnej publikacji		Wejścia bezpośrednie na inne strony		Łączna liczba sesji z wejściami bezpośrednimi
	Liczba sesji	% łącznej liczby sesji	Liczba sesji	% łącznej liczby sesji	Liczba sesji	% łącznej liczby sesji	
BC UWr	73 657	60,4%	44 747	36,7%	3 477	2,9%	121 881
e-bUW	12 307	17,4%	52 848	74,6%	5 723	8,1%	70 878
MBC	11 245	19,5%	44 812	77,6%	1 711	3,0%	57 768
RCIN	9 868	18,1%	43 163	79,0%	1 583	2,9%	54 614
ŚBC	18 739	19,7%	60 069	63,2%	16 194	17,0%	95 002
WBC	48 117	20,8%	126 273	54,5%	57 334	24,7%	231 724
Suma	173 933	27,5%	371 912	58,9%	86 022	13,6%	631 867

Tabela 3. Strony początkowe wejść bezpośrednich użytkowników badanych bibliotek cyfrowych w 2017 roku.

Analiza ta rozjaśnia w pewnym stopniu duży udział kanału wejść bezpośrednich w ruchu BC UW. Wygląda na to, że spora grupa użytkowników tej biblioteki cyfrowej korzysta z niej zaczynając od bezpośredniego wejścia na stronę główną. Jest to ciekawy wyróżnik BC UW, być może związany z efektywnymi metodami promocji biblioteki w środowisku lokalnym, gdyż około 75% sesji, które zaczęły się od bezpośredniego wejścia na stronę główną było wykonanych przez użytkowników z Wrocławia.

Podwyższona liczba wejść bezpośrednich na strony konkretnych publikacji w e-bUW i MBC jest z kolei częściowo powiązana z dwiema wyjątkowo popularnymi publikacjami:

- W e-bUW w 2017 roku było 2 521 sesji rozpoczynających się od bezpośredniego wejścia na stronę <http://ebuw.uw.edu.pl/dlibra/docmetadata?id=210344>, miały one miejsce w przybliżeniu od początku maja do połowy sierpnia.
- W MBC w 2017 roku było 5 404 sesji rozpoczynających się od bezpośredniego wejścia na stronę <http://mbc.malopolska.pl/dlibra/docmetadata?id=7794>, miały one miejsce głównie pod koniec czerwca oraz we wrześniu. Na stronie 10. tego numeru Życia Literackiego, wydanego w 1972 roku, znajduje się opowiadanie zatytułowane... Harry Potter.

Również w przypadku RCIN dużą rolę odegrał z pewnością charakter udostępnianych zasobów (w dużej mierze współczesne publikacje naukowe w formacie PDF), co widać było już w Tabeli 1, gdzie RCIN wyróżniał się pod względem liczby odwiedzin i wyświetleń treści publikacji, będąc równocześnie najmniejszą biblioteką cyfrową w zestawieniu.

W przypadku Śląskiej i Wielkopolskiej Biblioteki Cyfrowej sporym zainteresowaniem wśród użytkowników wchodzących bezpośrednio na strony tych bibliotek cieszyły się kolekcje tematyczne, zaklasyfikowane w Tabeli 3 jako inne strony. Poza tym dużo wejść bezpośrednich dotyczyło też odnośników do zapamiętanych wcześniej wyników wyszukiwania.

Trzeci pod względem liczności kanał ruchu to wejścia z innych stron internetowych – nieco ponad jedna piąta odwiedzin. Kanał ten zostanie szczegółowo omówiony w dalszej części niniejszego raportu.

Ostatnim kanałem ruchu są media społecznościowe, które przynoszą około dwóch procent odwiedzin. W tym kanale dominuje Facebook, generując od ponad osiemdziesięciu do ponad dziewięćdziesięciu procent odesłań w ramach kanału. Na drugim miejscu jest w większości przypadków serwis Blogger z kilkuprocentowym udziałem, a po nim następuje seria drobniejszych sieci społecznościowych. W tej części analizy wyróżniają się dwie biblioteki cyfrowe:

- W przypadku RCINu, po Facebooku który ma 81% udziału w kanale społecznościowym, właściwie na równi z udziałem po 7% znajdują się Blogger i Twitter. Wyższa pozycja Twittera jest zapewne związana z popularnością tego narzędzia w środowiskach naukowych, choć jeżeli chodzi o wartości bezwzględne, to i tak mowa jest o 224 wejściach z Twittera do RCIN w ciągu całego 2017 r., tak więc trudno tutaj mówić o dużym znaczeniu tego źródła ruchu.
- W przypadku WBC na drugim miejscu wśród odsyłających sieci społecznościowych, z udziałem na poziomie niecałych 10% w tym kanale, znajduje się serwis Wikia, a konkretnie trzy hostowane tam społecznościowe wiki: de.szlachta.wikia.com, poznaw.wikia.com oraz pisarki.wikia.com.

3. Serwisy kierujące ruch do bibliotek cyfrowych

Jak wynika z Tabeli 2, średnio 21,5% sesji jakie miały miejsce w 2017 roku w badanych bibliotekach cyfrowych było spowodowanych przekierowaniem użytkowników z zewnętrznych serwisów WWW innych niż ogólne wyszukiwarki internetowe czy media społecznościowe. Tabela 4 poniżej zawiera podstawowe dane liczbowe odnośnie liczby unikalnych domen internetowych, z których kierowani byli użytkownicy do badanych bibliotek cyfrowych w 2017 roku wraz ze średnią liczbą skierowań przypadających na taką zewnętrzną domenę. Jak widać, ruch ten generowany był w sumie przez serwisy z 4 180 różnych domen internetowych, z których wiele linkowało do więcej niż jednej biblioteki cyfrowej (dlatego też liczba w ostatnim wierszu przedostatniej kolumny nie jest sumą wartości znajdujących się wyżej w tej kolumnie).

Biblioteka cyfrowa	Liczba sesji skierowanych z innych stron	Liczba unikalnych domen kierujących użytkowników	Średnia liczba skierowań na domenę
BC Uwr	27 471	576	47,7
e-bUW	61 173	687	89,0
MBC	51 597	743	69,4
RCIN	43 395	964	45,0
ŚBC	135 077	1 711	78,9
WBC	165 470	1 826	90,6
łącznie	484 183	4 180	115,8

Tabela 4. Liczba serwisów zewnętrznych kierujących ruch do badanych bibliotek cyfrowych w 2017 roku.

Następnym krokiem w analizie danych o ruchu kierowanego z serwisów zewnętrznych było podzielenie poszczególnych źródeł ze względu na liczbę badanych bibliotek cyfrowych, do których źródła te linkują. Inaczej mówiąc, przeprowadzona została analiza porównawcza list domen kierujących ruch do poszczególnych bibliotek cyfrowych, pod kątem wyszukania domen pojawiających się na więcej niż jednej liście. Wyniki tej analizy zawiera Tabela 5.

Liczba bibliotek cyfrowych, do których kieruje źródło	Liczba źródeł	% łącznej liczby źródeł	Procentowy udział w ruchu generowanym przez skierowania zewnętrzne
1	3 052	73,01%	22,90%
2	547	13,09%	11,25%
3	258	6,17%	9,35%
4	126	3,01%	10,08%
5	99	2,37%	5,26%
6	98	2,34%	41,16%

Tabela 5. Rozkład liczby serwisów zewnętrznych kierujących ruch do badanych bibliotek cyfrowych w 2017 roku ze względu na liczbę bibliotek cyfrowych, do których ruch ten jest kierowany.

Dane w Tabeli 5 należy odczytywać tak, że są 3 052 źródła, które kierują ruch tylko do jednej biblioteki cyfrowej i dają one 22,90% ruchu pochodzącego ze skierowań w badanych bibliotekach cyfrowych łącznie. I dalej, są 1 094 źródła kierujące do dwóch różnych bibliotek cyfrowych, i dają one łącznie 11,25% ruchu itd. Jak widać, około 73% zewnętrznych źródeł kieruje tylko do jednej biblioteki cyfrowej, dając 22,90% łącznego ruchu. Dalsze około 19% źródeł kieruje do dwóch lub trzech bibliotek cyfrowych, dając 20,60% ruchu. Pozostałe 8% źródeł (323 różne domeny internetowe) kieruje do ponad połowy bibliotek cyfrowych, generując 56,50% łącznego ruchu pochodzącego ze skierowań z serwisów zewnętrznych.

Aby lepiej poznać charakter serwisów, które kierują najwięcej użytkowników do bibliotek cyfrowych dalszej analizie poddane zostały te serwisy, które kierują ruch do czterech lub więcej spośród badanych bibliotek cyfrowych. Serwisy te zostały podzielone na następujące kategorie:

- Agregator – bezpłatny lub płatny serwis, którego podstawą działania jest agregowanie i/lub przeszukiwanie danych z różnych źródeł, i który nie jest uruchomiony wyłącznie na potrzeby jednej instytucji.
- Serwis edukacyjny / encyklopedia – serwis o charakterze edukacyjnym lub encyklopedycznym, bez wyraźnego nacechowania naukowego.
- Serwis genealogiczny – serwis związany z danymi czy źródłami genealogicznymi lub społecznościami zajmującymi się tą tematyką.
- Katalog / biblioteka – serwis WWW biblioteki, w tym również katalogi i multiwyszukiwarki biblioteczne uruchomione na potrzeby konkretnych instytucji.
- Serwis hobbystyczny – serwis dotyczący określonej tematyki (najczęściej historycznej), o charakterze hobbystycznym, a nie edukacyjnym czy naukowym.
- Serwis naukowy/branżowy – serwis o charakterze naukowym lub branżowym (profesjonalnym, nie hobbystycznym).
- Wyszukiwarka naukowa – wyszukiwarka internetowa skupiająca się na zasobach naukowych.
- Poczta/chat – serwisy do komunikacji (np. poczta WWW), z których użytkownicy przechodzili do bibliotek cyfrowych otrzymując wiadomość z odnośnikiem.
- Biblioteka cyfrowa – serwis biblioteki cyfrowej lub repozytorium.
- Wyszukiwarka ogólna – serwis o charakterze ogólnej wyszukiwarki internetowej, najczęściej niszowy, który nie został przez Google Analytics automatycznie zakwalifikowany do grupy skierowań z wyników wyszukiwania.
- Media – serwisy związane ze współcześnie funkcjonującymi mediami (radio, TV, prasa itp.).
- Inne – serwisy nie pasujące do żadnej z kategorii powyżej.

Kategorie te oraz ich definicje nie wywodzą się z żadnej formalnej metodyki podziału serwisów internetowych, są natomiast wynikiem analizy charakteru serwisów internetowych, które kierowały w 2017 roku największy ruch do badanych bibliotek cyfrowych.

W Tabeli 6 poniżej zebrano dane na temat procentowego udziału źródeł z poszczególnych kategorii w ruchu pochodzącym z zewnętrznych serwisów w poszczególnych bibliotekach cyfrowych. Najbardziej znaczącymi kategoriami źródeł są: agregatory danych, serwisy edukacyjne/encyklopedie, serwisy genealogiczne, serwisy biblioteki (w tym katalogi), serwisy hobbystyczne oraz serwisy i wyszukiwarki naukowe. Taki zestaw kategorii zdaje się potwierdzać prowadzone wcześniej badania

na temat użytkowników bibliotek cyfrowych, gdzie do najważniejszych grupy odbiorców treści bibliotek cyfrowych zaliczeni zostali hobbyści (w tym genealodzy), studenci oraz pracownicy naukowci⁴. W dalszej części raportu powyższe, najważniejsze kategorie źródeł zostały poddane dodatkowej analizie.

Kategoria źródła skierowań	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	Dla całego ruchu ze skierowań łącznie	
Źródła kierujące do 3 lub mniej BC	13,36%	55,68%	32,78%	46,59%	69,86%	46,67%	43,50%	
Źródła kierujące do 4 lub więcej BC	Agregator	12,91%	7,83%	24,24%	23,44%	10,97%	11,88%	15,86%
	Serwis edukacyjny/encyklopedia	11,02%	18,28%	8,19%	12,91%	7,20%	15,57%	13,39%
	Serwis genealogiczny	11,59%	4,64%	3,29%	7,32%	2,75%	17,66%	10,46%
	Katalog/biblioteka	43,66%	3,30%	8,06%	3,60%	3,26%	2,19%	8,53%
	Serwis hobbystyczny	2,45%	2,37%	6,85%	1,89%	1,26%	3,40%	2,94%
	Serwis naukowy/branżowy	2,96%	4,01%	1,91%	0,97%	2,10%	0,54%	1,55%
	Wyszukiwarka naukowa	0,00%	0,00%	11,13%	0,10%	0,49%	0,07%	1,09%
	Poczta/chat	0,41%	0,50%	2,10%	0,66%	0,82%	0,76%	0,78%
	Biblioteka cyfrowa	0,63%	1,32%	0,18%	1,71%	0,14%	0,15%	0,77%
	Wyszukiwarka ogólna	0,49%	0,51%	0,71%	0,45%	0,82%	0,76%	0,61%
	Inne	0,50%	0,47%	0,57%	0,35%	0,31%	0,35%	0,40%
Media	0,03%	1,08%	0,00%	0,00%	0,02%	0,00%	0,12%	

Tabela 6. Procentowy udział poszczególnych kategorii źródeł w ruchu kierowanym z zewnętrznych źródeł do badanych bibliotek cyfrowych w 2017 roku.

Agregatory

W przypadku serwisów agregujących metadane z bibliotek cyfrowych lub katalogów bibliotecznych, kluczową rolę w generowaniu ruchu odgrywają przede wszystkim Federacja Bibliotek Cyfrowych i Europeana (patrz Tabela 7 poniżej).

Nazwa agregatora	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	łącznie
FBC	6 346	3 030	8 112	14 190	2 079	14 049	47 806
Europeana	977	952	2 397	16 411	958	5 114	26 809
EBSCO	562	83	55	1 224	38	481	2 443
KaRo	47	22	13	26	7	55	170
WorldCat	12	0	106	17	25	9	169
OpenDOAR	16	4	21	24	13	13	91
slavistik-portal.de	2	4	7	9	1	10	33
łącznie na BC	7 962	4 095	10 711	31 901	3 121	19 731	77 521

⁴ Mazurek, J. Potrzeby informacyjne kluczowych grup użytkowników polskich bibliotek cyfrowych. Raport z badania. Biuletyn EBIB [on-line]. 2011, nr 7. ISSN 1507-7187. Dostępny w: http://www.ebib.pl/images/stories/numery/125/125_mazurek.pdf.

Tabela 7. Liczba sesji kierowanych do badanych bibliotek cyfrowych w 2017 roku przez serwisy zakwalifikowane do kategorii „Agregatory”.

Co ciekawe, w przypadku Śląskiej Biblioteki Cyfrowej Europeana przekierowała w 2017 roku więcej użytkowników niż FBC, i na podstawie danych z Google Analytics trudno jest określić przyczyny tej sytuacji. Ruch z Europeany do ŚBC jest rozłożony dość równomiernie w ciągu roku i nie jest skupiony na żadnej konkretnej grupie wyjątkowo ciekawych obiektów. Dla innych badanych bibliotek cyfrowych FBC ma jednak pozycję zdecydowanie dominującą, dostarczając odwiedzających również równomiernie na przestrzeni roku. Krótkie podsumowanie informacji o źródłach ruchu w portalu FBC zawarto w Załączniku nr 1 do niniejszego opracowania.

Pod hasłem EBSCO w Tabeli 7 zgrupowanych jest kilka domen związanych z komercyjnymi usługami tej firmy.

Serwisy edukacyjne/encyklopedie

Kategorię serwisów edukacyjnych i encyklopedii dominującą rolę odgrywa Wikipedia i powiązane z nią serwisy. Pozycja „Serwisy na bazie Wikipedii” w Tabeli 8 oznacza takie strony internetowe, które pod własną domeną, w innym układzie interfejsu prezentują skopiowane z Wikipedii strony poszczególnych haseł encyklopedycznych. Widoczny jest też ruch z serwisu Wikimedia Commons, gdzie gromadzone są materiały multimedialne wykorzystywane m.in. w hasłach Wikipedii. Wynika z tego, że część z tych materiałów musi pochodzić z badanych bibliotek cyfrowych. Dodatkowo w przypadku Śląskiej Biblioteki Cyfrowej wyróżnia się serwis ibrbs.pl, pod którym znaleźć można Encyklopedię Województwa Śląskiego tworzoną w Instytucie Badań Regionalnych Biblioteki Śląskiej, której hasła intensywnie wykorzystują zasoby ŚBC (oraz jak widać czasem też innych bibliotek cyfrowych).

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	Łącznie na serwis
Wikipedia	6 085	9 005	3 408	15 610	1 747	23 971	59 826
Wikisource	23	11	73	794	115	392	1 408
Wikiquote	7	413	8	26	0	554	1 008
Wikitionary	452	0	6	1	0	476	935
ibrbs.pl	1	1	18	823	13	10	866
Wikimedia Commons	57	30	102	208	102	169	668
olijp.pl	121	36	0	40	57	140	394
Serwisy na bazie Wikipedii	42	47	4	57	13	109	272
szkolnictwo.pl	4	15	0	2	0	14	35
Łącznie na BC	6 792	9 558	3 619	17 561	2 047	25 835	65 412

Tabela 8. Liczba sesji kierowanych do badanych bibliotek cyfrowych w 2017 roku przez serwisy zakwalifikowane do kategorii „Serwisy edukacyjne / encyklopedie”.

Serwisy genealogiczne

Serwisy genealogiczne linkujące do bibliotek cyfrowych to zarówno hobbystycznie prowadzone indeksy osób, jak i po prostu listy materiałów źródłowych przydatnych w badaniach genealogicznych. Część z tych serwisów to także fora dyskusyjne, na których genealodzy wymieniają się odnośnikami do materiałów z bibliotek cyfrowych. Część z tych serwisów ma charakter regionalny (np. wtg-gniazdo.org

czy mtg-malopolska.org.pl), co przekłada się na większą liczbę przekierowań z danej strony do biblioteki cyfrowej właściwej pod względem regionu (por. Tabela 9 poniżej). Są też takie serwisy, które linkują w miarę równomiernie materiały z różnych bibliotek cyfrowych.

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	łącznie na serwis
genealodzy.pl	4 125	704	808	3 705	175	9 687	19 204
wtg-gniazdo.org	98	11	89	251	36	10 588	11 073
genealogia.okiem.pl	489	243	102	250	0	4 062	5 146
genealogyindexer.org	1 348	421	32	1 735	0	1 558	5 094
forum.ahnenforschung.net	24	4	159	1 982	340	592	3 101
mtg-malopolska.org.pl	7	846	0	366	0	32	1 251
genepedia.pl	13	54	2	591	17	567	1 244
kresy.genealodzy.pl	475	13	0	176	0	457	1 121
kimonibyli.pl	159	0	57	271	0	404	891
Pozostałe serwisy (20)	408	131	204	633	215	1 391	2 982
łącznie na BC	7 146	2 427	1 453	9 960	783	29 338	51 107

Tabela 9. Liczba sesji kierowanych do badanych bibliotek cyfrowych w 2017 roku przez serwisy zakwalifikowane do kategorii „Serwisy genealogiczne”.

Katalogi/biblioteki

Praktycznie każda biblioteka posiada w dzisiejszych czasach elektroniczny katalog zbiorów. Wiele z tych katalogów dostępnych jest on-line, a zawarte w nich rekordy bibliograficzne zawierają odnośniki do bibliotek cyfrowych udostępniających opisane pozycje on-line. Część z bibliotek dodatkowo posiada również serwisy indeksujące różne lokalne źródła danych (np. lokalny katalog, biblioteka cyfrowa, repozytorium, zasoby licencjonowane). Takie serwisy również generują ruch w bibliotekach cyfrowych, o ile tylko indeksują ich zbiory. Katalogi i inne serwisy biblioteczne są z pewnością najliczniejszą kategorią serwisów przekierowujących do bibliotek cyfrowych i jak widać przekierowują nie tylko do zasobów/serwisów tej samej instytucji, ale również do serwisów zewnętrznych.

Z danych zebranych w Tabeli 10 poniżej widać, że katalogiem bibliotecznym generującym największy ruch w badanych bibliotekach cyfrowych jest katalog NUKAT. Na pierwszym miejscu w tabeli jest co prawda strona główna Biblioteki UW, jednak skierowania ujęte w tym akuracie wierszu nie wynikają z korzystania z katalogu bibliotecznego, a z kliknięcia dobrze widocznego na stronie biblioteki odnośnika do serwisu e-bUW (katalog i multiwyszukiwarka BUW funkcjonują pod innymi domenami). Warto również podkreślić, że rola NUKATu w tym kontekście nie ogranicza się jedynie do kierowania użytkowników katalogu NUKAT. Dzięki wykorzystywaniu zawierających link rekordów NUKAT przez inne biblioteki, kolejne katalogi biblioteczne są w stanie kierować użytkowników do zasobów polskich bibliotek cyfrowych. Obraz ten wzmocniają jeszcze dane z Załącznika nr 1, z których wynika, że katalog NUKAT odpowiada za około 40% wizyt kierowanych do portalu Federacji Bibliotek Cyfrowych z zewnętrznych serwisów (innych niż ogólne wyszukiwarki i media społecznościowe). Ruch ten, w wielu przypadkach przechodzi dalej do serwisów poszczególnych bibliotek cyfrowych, widziany na ich poziomie jako ruch skierowany przez FBC. Jest to z pewnością jeden z czynników wysokiej pozycji FBC w Tabeli 7.

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	łącznie na serwis
buw.uw.edu.pl	23 599	0	17	4	0	4	23 624
nukat.edu.pl	1 899	316	614	441	104	976	4 350
www1.bg.us.edu.pl	42	321	196	1217	222	714	2 712
opacwww.bs.katowice.pl	2	25	4	1223	0	0	1 254
katalog.pan.pl	19	20	858	24	6	23	950
opac.bu.umk.pl	359	70	109	70	22	160	790
iaepan.vot.pl	0	1	766	1	3	0	771
chamo.bg.ug.edu.pl :8080	79	110	128	49	25	168	559
lbw.lublin.eu	86	110	65	44	50	138	493
opac.ciniba.edu.pl	20	69	27	284	25	34	459
bu.kul.pl	147	25	0	188	2	84	446
ezb.uni-regensburg.de		16	110	156	4	160	446
uam-hip.pfsl.poznan.pl	31	94	40	22	7	214	408
primo.koszykowa.pl	82	94	94	47	9	24	350
webpac.uci.umk.pl	127	23	56	35	34	45	320
primo-48amu.hosted.exlibrisgroup.com	16	15	48	24	18	185	306
Pozostałe serwisy (60)	413	418	431	1 069	395	711	3 437
łącznie na BC	26 921	1 727	3 563	4 898	926	3 640	41 675

Tabela 10. Liczba sesji kierowanych do badanych bibliotek cyfrowych w 2017 roku przez serwisy zakwalifikowane do kategorii „Katalogi/biblioteki”.

W tym kontekście bardzo zaskakuje wyjątkowo niska pozycja serwisów Biblioteki Narodowej, które w 2017 roku wygenerowały łącznie 46 skierowań użytkowników do badanych bibliotek cyfrowych (katalogi.bn.org.pl: 20; mak.bn.org.pl: 11; bn.org.pl: 9; alpha.bn.org.pl: 4; zin.bn.org.pl: 2) i został przez to w Tabeli 10 ujęte w wierszu „Pozostałe serwisy (60)”. Jeżeli chodzi o skierowania z serwisów BN do portalu Federacji Bibliotek Cyfrowych, to w 2017 roku nastąpiły 3 takie skierowania, wszystkie z domeny szkolenia.bn.org.pl.

Serwisy hobbystyczne

Zebrane w Tabeli 11 serwisy skategoryzowane jako hobbystyczne kierują do zbiorów bibliotek cyfrowych najczęściej w kontekstach historycznych, często powiązanych z turystyką i szukaniem „skarbów” z przeszłości. Są to też serwisy powiązane z literaturą, sztuką, architekturą, sportem, wojskowością czy szeroko rozumianym kolekcjonerstwem (np. <http://buttonarium.eu/>). Wyróżnia się tutaj być może pierwsza pozycja na liście, która generuje stosunkowo dużo odwiedzin przede wszystkim w serwisie RCIN. W większości są to po prostu przejścia z katalogu zasobów WWW w serwisie konik.pl do historycznych map ze zbiorów CBGiOŚ IGiPZ PAN dostępnych w portalu RCIN.

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	łącznie na serwis
konik.pl	0	70	1887	0	21	28	2 006
dws.org.pl	410	81	11	323	12	952	1 789
historycy.org	168	24	181	112	0	982	1 467
forum.vgd.ru	149	12	1	8	0	551	721

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	łącznie na serwis
oshispn.pl	33	165	0	108	64	259	629
eksploratorzy.com.pl	0	11	18	503	5	16	553
bibliepolskie.pl	24	12	0	41	0	404	481
odkrywca.pl	48	0	194	25	1	199	467
stareplanymiast.pl	95	0	161	50	75	31	412
lapidaria.wikidot.com	3	0	1	1	0	393	398
many-roads.com	0	8	0	6	1	371	386
wikiliga.pl	40	72	0	190	7	53	362
witkacologia.eu	36	156	31	99	0	29	351
wykop.pl	7	137	72	32	15	79	342
starbel.narod.ru	14	4	0	30	1	249	298
poszukiwanieskarbow.com	1	12	6	70	1	191	281
coryllus.pl	30	2	121	1	28	81	263
ekartkazwarszawy.pl	110	12	6	101	1	15	245
austro-wegry.info	28	42	18	110	0	42	240
zdziennikaodkrywcy.pl	28	109	0	49	34	0	220
skyscrapercity.com	1	21	68	79	5	43	217
Pozostałe serwisy (37)	285	288	252	640	88	674	2 227
łącznie na BC	1 510	1 238	3 028	2 578	359	5 642	14 355

Tabela 11. Liczba sesji kierowanych do badanych bibliotek cyfrowych w 2017 roku przez serwisy zakwalifikowane do kategorii „Serwisy hobbystyczne”.

Serwisy naukowe / branżowe

Serwisu umieszczone w tej kategorii są zazwyczaj serwisami bądź bazami danych prowadzonymi przez instytucje lub zespoły badawcze, najczęściej związane z humanistyką, kulturą czy sztuką (por. Tabela 12). Dla takich serwisów biblioteki cyfrowe są po prostu doskonałymi bazami danych źródłowych.

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	łącznie na serwis
leksykografia.uw.edu.pl	1 118	227	416	0	0	283	2 044
historiaimedia.org	20	1 304	0	31	59	74	1 488
ebib.pl	273	309	178	251	77	16	1 104
polishhistory.pl	128	120	1	77	196	119	641
archivauskunft.de	0	1	0	592	11	33	637
scriptores.pl	59	8	4	59	0	96	226
manuscripta.pl	6	5	2	20	151	27	211
citeweb.info	0	4	115	37	0	1	157
iaml.pl	35	0	0	50	17	51	153
omnia.ie	2	0	7	120	0	8	137
elzenberg.pl		71	4	3	0	45	123
legalnakultura.pl	27	0	2	22	56	15	122

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	łącznie na serwis
topincs.man.poznan.pl	66	8	1	4	0	42	121
cejsh.icm.edu.pl	0	2	62	24	11	3	102
Pozostałe serwisy (11)	90	40	53	26	20	84	313
łącznie na BC	1 824	2099	845	1 316	598	897	7 579

Tabela 12. Liczba sesji kierowanych do badanych bibliotek cyfrowych w 2017 roku przez serwisy zakwalifikowane do kategorii „Serwisy naukowe / branżowe”.

Jeżeli któryś z serwisów generuje wyraźnie więcej ruchu w jednej bibliotece cyfrowej, zazwyczaj jest to łatwe do wyjaśnienia bezpośrednim powiązaniem instytucjonalnym lub kontekstem geograficzno-historycznym. W przypadku serwisu historiaimedia.org duża liczba skierowań do MBC wynika z wiadomości na temat udostępniania w MBC pełnego archiwum „Przekroju”.

Wyszukiwarki naukowe

Ostatnią z analizowanych kategorii serwisów generujących ruch w bibliotekach cyfrowych są wyszukiwarki naukowe. Google Analytics skierowania z takich specjalistycznych wyszukiwarek nie zalicza do ruchu skierowanego z wyszukiwarek, tylko traktuje jako ruch przekierowany z serwisów zewnętrznych. W praktyce ruch ten ogranicza się do odesłań z serwisu Google Scholar, w wariantach nazw domenowych specyficznych dla różnych krajów (nie ma wejść z serwisów konkurencyjnych, takich jak np. Microsoft Academic). W MBC i e-bUW ruch taki praktycznie nie występuje lub nie jest widziany przez Google Analytics, natomiast w pozostałych badanych bibliotekach cyfrowych ruch jest przede wszystkim napędzany przez polską wersję serwisu (por. Tabela 13). Zdecydowanie wyróżnia się tutaj serwis RCIN, ale też jest on jedynym repozytorium naukowym w grupie badanych serwisów.

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	łącznie na serwis
scholar.google.de	0	0	280	2	1	1	284
scholar.google.fr	1	0	207	2	0	2	212
scholar.google.pl	0	0	4 431	135	138	116	4 820
łącznie na BC	1	0	4 918	139	139	119	5 316

Tabela 13. Liczba sesji kierowanych do badanych bibliotek cyfrowych w 2017 roku przez serwisy zakwalifikowane do kategorii „Wyszukiwarki naukowe”.

Dodatkowa analiza skierowań ruchu z różnych wersji domenowych Google Scholar, również tych które w ramach badanej grupy bibliotek cyfrowych kierują ruch tylko do RCIN, jeszcze bardziej podkreśla widoczność w sieci RCIN jako repozytorium naukowego. Zebrane w Tabeli 14 dane pokazują, że w roku 2017 RCIN miał w umię około 9 500 wejść ze 115 różnych domen Google Scholar (tych, które wyłapało Google Analytics, por. Tabela 1, kolumny 3 i 4). Były to domeny narodowe (np. scholar.google.com.au) lub domeny-proxy poszczególnych uczelni (np. scholar-google-com.proxy.lib.umich.edu). Trudno te liczby porównywać z jakkolwiek spośród badanych bibliotek cyfrowych.

Nazwa serwisu	e-bUW	MBC	RCIN	ŚBC	BC UW	WBC	Łącznie dla wszystkich domen
Łączna liczba domen Google Scholar kierujących do danej BC	1	0	115	13	2	25	116
Łączna liczba skierowań użytkowników z domen Google Scholar	1	0	9 490	162	139	209	10 001

Tabela 14. Liczba sesji kierowanych do badanych bibliotek cyfrowych w 2017 roku przez serwis Google Scholar.

4. Podsumowanie

Opisana w niniejszym raporcie analiza danych o źródłach ruchu w polskich bibliotekach cyfrowych powinna stanowić ciekawe źródło informacji dla wszystkich osób zajmujących się zawodowo bibliotekami cyfrowymi, a w szczególności dla twórców i administratorów takich bibliotek. Omówione powyżej dane pokazują, że tytułowych „przyjaciół bibliotek cyfrowych” należy szukać w kilku miejscach.

Po pierwsze, są to oczywiście wyszukiwarki internetowe, a właściwie przede wszystkim wyszukiwarka Google. Uruchamianie w obecnych czasach biblioteki cyfrowej, która nie jest odpowiednio zoptymalizowana pod kątem indeksowania przez wyszukiwarki internetowe, jest olbrzymim marnotrawstwem potencjalnego bezpłatnego ruchu jaki może przynieść bibliotece cyfrowej zaindeksowanie i dobre spozycjonowanie się w Google. Innym bardzo popularnym serwisem ogólnego przeznaczenia, w którym warto się znaleźć jest Wikipedia i powiązane z nią portale.

Kolejnym kluczowym elementem ekosystemu informacyjnego bibliotek cyfrowych są serwisy agregujące i katalogi biblioteczne, które kierują swoich użytkowników do zbiorów bibliotek cyfrowych. W tej grupie w szczególności wyróżniają się agregatory Federacja Bibliotek Cyfrowych i Europeana oraz katalog NUKAT.

Przyjaciółmi bibliotek cyfrowych są też oczywiście ich użytkownicy. Wielu z nich nie ogranicza się do pasywnego pobierania treści z internetu, ale jest też zaangażowanych w budowanie własnych serwisów internetowych lub po prostu dzieli się swoją wiedzą on-line. Serwisy te stają się z kolei istotnym źródłem skierowań użytkowników do bibliotek cyfrowych i ich zasobów.

Wiele wniosków przedstawionych w niniejszym raporcie mogło zostać sformułowanych dzięki porównaniu danych pochodzących z kilku różnych bibliotek cyfrowych. Otwartość na dzielenie się danymi o ruchu użytkowników i korzystanie tej samej platformy do gromadzenia danych, pozwoliły na znalezienie i głębsze przeanalizowanie nietypowych charakterystyk ruchu w poszczególnych serwisach. Bez możliwości porównania danych pomiędzy bibliotekami, wiele z tych ciekawych obserwacji z pewnością by umknęło.

Dlatego też raport ten ma nie tylko stanowić potencjalny punkt odniesienia dla analizy danych bibliotek cyfrowych nieujętych w opisywanym badaniu, ale ma też być zachętą do rozpoczęcia gromadzenia danych o ruchu użytkowników i do dzielenia się tymi danymi. Możliwość prowadzenia takich analiz ma szansę pozwolić twórcom bibliotek cyfrowych na wychwytywanie i adaptowanie dobrych praktyk wypracowanych, często nawet nieświadomie, przez inne podobne serwisy.

Załącznik nr 1. Źródła ruchu w portalu FBC

Wszystkie prezentowane poniżej dane zostały zebrane za pomocą narzędzia Google Analytics.

Kanał ruchu użytkowników	Liczba sesji w 2017 roku	% łącznej liczby sesji
Wejścia z wyszukiwarek	575 284	70,7%
Wejścia bezpośrednie	87 460	10,7%
Wejścia z innych stron	145 029	17,8%
Wejścia z social media	6 172	0,8%
łącznie	813 945	

Tabela 14. Udział kanałów ruchu w skierowaniach użytkowników do portalu Federacji Bibliotek Cyfrowych w 2017 roku (por. Tabela 2).

Źródło skierowań	Liczba sesji	% łącznej liczby sesji pochodzących ze skierowań
nukat.edu.pl	57 200	39,44%
sbc.org.pl	8 084	5,57%
wbc.poznan.pl	6 885	4,75%
pl.wikipedia.org	4 179	2,88%
bg.us.edu.pl	3 719	2,56%
jbc.bj.uj.edu.pl	3 195	2,20%
delibra.bg.polsl.pl	2 480	1,71%
eupeana.eu	2 351	1,62%
kpbc.umk.pl	2 177	1,50%
ebib.pl	1 968	1,36%
dbc.wroc.pl	1 818	1,25%
czasopisma.fbc.net.pl	1 738	1,20%
com.google.android.googlequicksearchbox	1 702	1,17%
buw.uw.edu.pl	1 676	1,16%
ciniba.edu.pl	1 566	1,08%
pbc.biaman.pl	1 538	1,06%
bibliotekacyfrowa.pl	1 340	0,92%
mbc.malopolska.pl	1 294	0,89%
mbc.cyfrowemazowsze.pl	1 250	0,86%
ebuw.uw.edu.pl	1 188	0,82%
rcin.org.pl	1 073	0,74%

Tabela 15. Domeny serwisów kierujących ruch do portalu FBC w 2017 r. – wszystkie, które wygenerowały w tym okresie powyżej 1 000 wizyt. (por. Tabele 7-14).