

Sebastian GACA* (UP w Krakowie, Polska)

Propozycja modelu archiwizacji polskiego Internetu na bazie doświadczeń brytyjskich (Proposal archiving model of the Polish Internet based on British experiences)

Abstract: *The aim of this paper is to attempt to propose a model of archiving of the Polish Internet, based in this respect on the solutions adopted in Great Britain. It is known that although efforts have been made in Poland in the field of Internet archiving, it does not yet have an Internet archive. The obstacles preventing this action are the low budget of Polish institutions responsible for the protection of national heritage, as well as the unregulated legal situation of the issue of archiving the Polish Internet resource. This means that the valuable Internet resource related to the Polish cultural heritage is constantly and irretrievably lost, which leads to the inability to fully show the values of this heritage to future generations. In this situation, it becomes necessary to archive the resources of the Polish Internet. For the purposes of this paper, the current state of protection of the Polish Internet will be presented, as well as the British model of Internet archiving, which, if implemented in Poland, would enable the implementation of effective protection of its Internet heritage.*

Key words: Internet; polish Internet heritage; UK Government Web Archive; Internet archiving models

Streszczenie: *Celem niniejszego artykułu jest próba zaproponowania modelu archiwizacji polskiego Internetu, opartego w tym zakresie na przyjętych rozwiązaniach w Wielkiej Brytanii. W Polsce były czynione starania w zakresie archiwizacji Internetu, nie dysponuje ona dotychczas archiwum Internetu. Przeshkodami uniemożliwiającymi to działanie są niski budżet instytucji polskich odpowiedzialnych za ochronę narodowego dziedzictwa, a także nieuregulowana sytuacja prawna kwestii archiwizacji polskiego zasobu Internetu. Oznacza to, że cenny zasób Internetu związany z polskim dziedzictwem kulturowym, ulega ciężkiej i bezpowrotnej utracie, która prowadzi do niemożności ukazania przyszłym pokoleniom walorów tego dziedzictwa w sposób pełny. Koniecznym w tej sytuacji staje się archiwizacja zasobu polskiego Internetu. Na potrzeby tego referatu zostanie przedstawiony aktualny stan ochrony polskiego Internetu, a także model brytyjski archiwizacji Internetu, który zaimplementowany w Polsce umożliwiłyby wprowadzenie skutecznej ochrony jej dziedzictwa Internetu.*

Słowa kluczowe: Internet; polskie dziedzictwo Internetu; UK Government Web Archive; modele archiwizacji Internetu

* sebastiangaca97@gmail.com; Uniwersytet Pedagogiczny im. Komisji Edukacji Narodowej w Krakowie; Tekst został częściowo oparty, na napisanej w 2019 r. pod kierunkiem dra Barto-sza Drzewieckiego pracy licencjackiej pt.: „Archiwizacja Internetu – Polska na tle świata”.

Archiwizację Internetu można określić jako proces, w którym dochodzi do gromadzenia, opracowania oraz udostępnienia treści i zawartości stron internetowych¹. Efektem tego procesu są archiwalia internetowe, które można określić jako materiały odrodzone cyfrowo (ang. reborn digital)². Kolekcje zarchiwizowanych materiałów sieciowych będą składać się na Archiwum Internetu³. Archiwizacja Internetu może odbywać się w różnej skali, co będzie determinowane potrzebą instytucji, która ją prowadzi⁴. Na potrzeby niniejszej publikacji zostaną przedstawione państwowe projekty, które są skoncentrowane na archiwizacji stron internetowych domeny krajowej. Zanim jednak to nastąpi, warto uzasadnić potrzebę prowadzenia takich projektów.

Przede wszystkim materiały internetowe mogą stanowić cenne archiwalne źródło informacji. Internet stał się niewątpliwie jednym z obszarów życia człowieka, a obecne czasy pandemiczne dodatkowo to unaocznily. Trudno zatem wyobrazić sobie historię końca XX w. oraz XXI w. opowiedzianą przyszłym pokoleniom, bez zachowanego dziedzictwa Internetu. Wielkość, forma, różnorodność materiałów internetowych, a także liczba osób korzystających z Internetu zapewnia bardzo dużą możliwość prowadzenia zróżnicowanych badań naukowych⁵. Przykładem może być tutaj chociażby utworzone przez amerykańską Narodową Bibliotekę Medyczną w 2017 r. archiwum Internetu Epidemii Opioidowej, które stanowiło jedno z działań edukacyjnych i monitorujących w celu zwalczania kryzysu opioidowego w Stanach Zjednoczonych⁶. Należy zaznaczyć, że problem gromadzenia oraz przechowywania archiwaliów internetowych w porównaniu do tradycyjnych materiałów archiwalnych, zostaje znacząco zre-

¹ CHORAŻYCZEWSKI W., *Archiwistyka dla początkujących*, Toruń 2014, s. 23, <<https://repozytorium.umk.pl/handle/item/2191>> [dostęp 15.06.2021].

² KONOPA B., *Reborn Digital i Black Box – wpływ procesu archiwizacji na zasób archiwów Webu*, „Archiwa – Kancelarie – Zbiory”, 2019, nr. 10, s. 155, <<https://apcz.umk.pl/czasopisma/index.php/AKZ/article/view/AKZ.2019.008/24589>> [dostęp 17.07.2021].

³ *Archiwum sieciowe*, w: *Słowniczek archiwalny Programu Archiwizacji Webu Biblioteki Kongresu*, tłum. WILKOWSKI M., s. 2–3, <http://webarch.uw.edu.pl/wp-content/uploads/2018/12/slowniczek_archiwalny_LOC.odt> [dostęp 15.06.2021].

⁴ NIELSEN J., *Using Web archives in research*, tłum. WILKOWSKI M., s. 10-11, <http://webarch.uw.edu.pl/wp-content/uploads/2018/12/janne_nielsen_wykorzystanie_archiwow_webu_w_badaniach.odt> [dostęp 15.06.2021].

⁵ W maju 2019 roku istniało w Internecie ponad miliard stron internetowych, z czego 14 % pozostawało aktywnych. *Netcraft – May 2019 Web Server Survey*, <<https://news.netcraft.com/archives/2019/05/10/may-2019-web-server-survey.html>> [dostęp 15.06.2021]; W styczniu 2019 roku z Internetu korzystało 57% ludzkości. KEMP S., *Global Digital Report 2019*, <<https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>> [dostęp 15.06.2021].

⁶ SPEAKER S.L. i MOFFATT C., *The Opioid Epidemic: Collecting Now For Future Research*, <<https://circulatingnow.nlm.nih.gov/2018/09/27/the-opioid-epidemic-collecting-now-for-future-research/>> [dostęp 15.06.2021], cit. per: B. Konopa, *Badanie kryzysu opioidowego z wykorzystaniem archiwum Webu*, <<https://webarch.uw.edu.pl/blog/archiwum-kryzysu-opioidowego/>> [dostęp 15.06.2021]; *Opioid Epidemic web archive*, <<https://archive-it.org/collections/9280>>, [dostęp 15.06.2021].

dukowany⁷. Ponadto Internet jest efemeryczny, jednak archiwizacja sieci WWW pozwala na częściowe rozwiązanie problemu efemeryczności⁸.

Mając na uwadze wspomniane korzyści, nieprzypadkowo zaczęły powstawać na świecie archiwa, chroniące dziedzictwo Internetu. Początek archiwizacji Internetu należy datować na 1996 r., kiedy w Stanach Zjednoczonych z inicjatywy Brewster'a Kahle'a zostało założone Internet Archive. Jest to pierwsza oraz największa do dziś organizacja pozarządowa, która przyjęła za główny cel globalną ochronę materiałów online⁹. W 2003 r. Internet Archive oraz 11 instytucji o charakterze bibliotek narodowych powołały do życia International Internet Preservation Consortium (IIPC). Jest to organizacja działająca w skali międzynarodowej, zrzeszająca wszelkiego typu jednostki organizacyjne z całego świata, które prowadzą działalność w zakresie archiwizacji Internetu¹⁰. Do IIPC należy 54 różnych instytucji z 37 krajów, w tym m.in. co najmniej 26 bibliotek narodowych prowadzi krajowe archiwum Internetu i pochodzi w większości z Europy¹¹. Szacunki ilościowe dotyczące projektów, koncentrujących się na archiwizacji określonego wycinka Internetu mówią, że w 2011 roku na świecie było ich 42, a w 2018 roku co najmniej 87¹².

Jak można zauważyć, w większości biblioteki narodowe, a nie archiwa państwowe są odpowiedzialne za ochronę narodowego dziedzictwa Internetu. Wiąże się to z faktem łatwiejszej adaptacji prawnej tych instytucji do prowadzenia archiwizacji krajowej sieci. Bowiem przepisy prawa autorskiego nie pozwalałyby swobodnie archiwizować Internetu żadnej instytucji, gdyby nie to, że w wielu krajach zasoby Internetu objęto znanym w Polsce prawem egzemplarza obowiązkowego (EO) (ang. legal deposit), przysługującym bibliotekom¹³. Przepisy EO nakładają na wydawców „obowiązek bezpłatnego przekazywania publikacji do uprawnionych bibliotek”. Początkowo przepisy EO służyły gromadzeniu przez biblioteki tylko drukowanych materiałów, jednak uznano że obok tych form publikacji występują także materiały internetowe, które można uznać za publikacje według prawa EO¹⁴. Dzięki temu biblioteki w wielu

⁷ WILKOWSKI M., *Archiwistyka Webu: status, cele, metody (nagranie)*, <<https://www.youtube.com/watch?v=fPuC2VIK2N8>> [dostęp 17.07.2021].

⁸ NIELSEN J., op. cit., s. 5.

⁹ TOKARSKA-TRZASKOWSKA A., *Krótką historia archiwizowania Webu*, <<https://webarch.uw.edu.pl/zasoby/krotka-historia/>> [dostęp 15.06.2021]; *Internet Archive*, <<https://archive.org/>> [dostęp 15.06.2021].

¹⁰ *About the IIPC*, <<http://netpreserve.org/about-us/>> [dostęp 15.06.2021]; *IIPC members*, <<http://netpreserve.org/about-us/members/>> [dostęp 15.06.2021].

¹¹ *IIPC members...*,

¹² TOKARSKA-TRZASKOWSKA A., op. cit.

¹³ KONOPA B., *Witryna internetowa – Dokumentacja czy publikacja?*, „Biuletyn EBIB”, 2017, nr. 172, s. 4, <<http://ebibojs.pl/index.php/ebib/article/view/147>> [dostęp 15.06.2021].

¹⁴ DĄBROWSKA E., *Problem archiwizacji internetu w kontekście egzemplarza obowiązkowego: sytuacja w Polsce i wybranych krajach europejskich*, „Biuletyn EBIB”, 2017, nr. 172, s. 1-2, <<http://ebibojs.pl/index.php/ebib/article/view/146>> [dostęp 15.06.2021].

krajach mogą samodzielnie gromadzić kopie zawartości Internetu, bez przymusu udzielania każdorazowo zgody od ich właścicieli¹⁵.

Z punktu widzenia prawa instytucjami predestynowanymi do ochrony polskiego dziedzictwa Internetu są Narodowe Archiwum Cyfrowe (NAC) oraz Biblioteka Narodowa (BN). Takie stanowisko zostało podkreślone w Programie digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020, opracowanym na zlecenie Ministerstwa Kultury i Dziedzictwa Narodowego (MKiDN)¹⁶. Uzasadniono to tym, że NAC jako jedno z archiwów objęte ustawą z dnia 14 lipca 1983 r. o narodowym zasobie archiwalnym i archiwach, ustawą z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne oraz rozporządzeniem MKiDN z dnia 6 lutego 2008 r. w sprawie zmiany nazwy i zakresu działania Archiwum Dokumentacji Mechanicznej w Warszawie, jest właściwą instytucją do prowadzenia archiwizacji dokumentów elektronicznych oraz stron internetowych, „wytworzonych przez podmioty objęte zakresem wyżej wymienionych ustaw m.in. administracji rządowej i samorządowej (instytucje uprawnione do korzystania z domeny gov.pl)”¹⁷. BN we wspomnianym programie została także wskazana jako właściwa do prowadzenia krajowej archiwizacji stron internetowych. Zostało uzasadnione to tym, że BN mogłaby uzupełniać archiwizację pozostałej części Internetu, czyli stron nieurzędowych¹⁸. W tym wypadku BN miałyby gromadzić archiwalia Internetu od ich wydawców, uznane za egzemplarze publikacji, które zostałyby udostępnione publicznie przez nich na terenie RP, bądź za granicą zgodnie z ustawą z dnia 7 listopada 1996 r. o obowiązkowych egzemplarzach bibliotecznych¹⁹.

Za początek starań Polski w zakresie ochrony narodowego dziedzictwa Internetu należy uznać 2008 rok, kiedy BN śladem innych bibliotek narodowych została członkiem IIPC, żeby rok później w ramach tego samego konsorcjum, wziąć udział testowo wraz z siedmioma zagranicznymi bibliotekami w projekcie „’09 European Election Web Harvesting Project”. Projekt ten był realizowany w związku z Wyborami do Parlamentu Europejskiego 2009, a jego celem była archiwizacja wybranych stron internetowych kandydatów. Pod względem technicznym realizacja tego projektu przez BN zakończyła się powodzeniem. Ostatecznie jednak projekt w 2010 r. został

¹⁵ KONOPA B., *Witryna internetowa...*, s. 6.

¹⁶ *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020*, Warszawa 2009, s. 43, <<https://www.nac.gov.pl/wp-content/uploads/2015/05/Program-digitalizacji-2009-2020-1.pdf>> [dostęp 15.06.2021].

¹⁷ Ustawa z dnia 14 lipca 1983 r. o narodowym zasobie archiwalnym i archiwach, Dz.U. 1983, nr 38, poz. 173; Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne, Dz.U. 2005, nr 64, poz. 565; Rozporządzenie Ministra Kultury i Dziedzictwa Narodowego z dnia 6 lutego 2008 r. w sprawie zmiany nazwy i zakresu działania Archiwum Dokumentacji Mechanicznej w Warszawie, Dz.U. 2008, nr 29, poz. 167, cit. per: *Program digitalizacji...*, s. 43.

¹⁸ *Program digitalizacji...*, s. 41, 43.

¹⁹ Ustawa z dnia 7 listopada 1996 r. o obowiązkowych egzemplarzach bibliotecznych, Dz.U. 1996, nr 152, poz. 722, cit. per: *Program digitalizacji...*, s. 41.

zawieszony, na co wpływ miała nieefektywność uzyskiwania odpowiedzi od indagowanych podmiotów²⁰.

Kolejny krok w dążeniu do archiwizacji polskiego Internetu podjęło Narodowe Archiwum Cyfrowe (NAC), które w latach 2009-2011 prowadziło pierwsze istniejące w Polsce krajowe Archiwum Internetu. W dniu 1 kwietnia 2009 roku doszło do pierwszej krajowej archiwizacji przy wykorzystaniu Archiwum Internetu. Pierwszymi zgromadzonymi archiwizacjami polskiego zasobu Internetu stały się strony internetowe archiwów państwowych oraz Naczelnej Dyrekcji Archiwów Państwowych (NDAP). W dniu 26 marca 2010 roku powyższe archiwalia udostępnił publicznie na stronie internetowej pod adresem www.archiwuminternetu.nac.gov.pl²¹. Ostatnie archiwizacje stron internetowych w ramach Archiwum Internetu miały miejsce w kwietniu 2011 r., a w styczniu 2014 r. archiwum przestało być ogólnodostępne²². Wpływ na zamknięcie Archiwum Internetu miały dwa czynniki. Jednym z nich była w tamtym czasie nieuregulowana kwestia prawna. Prowadzono wtedy analizy prawdopodobnie na poziomie MKiDN we współpracy z NDAP i BN, mające rozstrzygnąć problem zakwalifikowania stron internetowych do materiałów archiwalnych gromadzonych przez archiwa państwowe, bądź do publikacji gromadzonych przez BN. Drugi czynnik zamknięcia Archiwum Internetu stanowił fakt, że było ono projektem prototypowym, który miał na celu poznanie możliwości technicznych systemu, zapewniającego archiwizację wybranych stron internetowych. Archiwum Internetu z racji swojej prototypowości wymagało ciągłej pracy nad nim, pociągając za sobą koszty, które NAC nie było w stanie pokryć. Co więcej – jak tłumaczy MKiDN – wobec położenia nacisku na prace związane z modernizacją systemu ZoSIA, a także Szukaj w Archiwach, nie było możliwości pozyskania zasobów ludzkich potrzebnych do kontynuacji prac nad Archiwum Internetu²³.

Los polskiego dziedzictwa Internetu jest w rękach MKiDN, gdyż ono dysponuje i rozporządza budżetem oraz reguluje prawo w zakresie krajowej archiwizacji Internetu. Uzyskana odpowiedź na wniosek o udostępnienie informacji publicznej od MKiDN świadczy o tym, że Ministerstwo planuje powierzyć zadanie archiwizacji Internetu NAC oraz BN²⁴. Archiwum Internetu prowadzone przez NAC zapewne bę-

²⁰ ŚLASKA K. i WASILEWSKA A., *Archiwizacja Internetu – sytuacja w polskim prawie z punktu widzenia bibliotekarzy*, „Biuletyn EBIB”, 2012, nr. 128, s. 2–3, <http://www.ebib.pl/?page_id=380> [dostęp 15.06.2021]; Odpowiedź Biblioteki Narodowej na wniosek o udostępnienie informacji publicznej z dnia 27 czerwca 2019 r.

²¹ Odpowiedź Narodowego Archiwum Cyfrowego na wnioski o udostępnienie informacji publicznej z dnia 31 marca 2019 r., s. 1–2 ; Odpowiedź Ministerstwa Kultury i Dziedzictwa Narodowego na wniosek o udostępnienie informacji publicznej z dnia 3 maja 2019 r., s. 1; *Narodowe Archiwum Cyfrowe – Wizja Projekt Ludzie*, Narodowe Archiwum Cyfrowe, Warszawa 2010, s. 141–142, <https://www.nac.gov.pl/wp-content/uploads/2015/05/NAC_wizja_projekt_ludzie_WWW.pdf> [dostęp 15.06.2021].

²² Odpowiedź Narodowego Archiwum Cyfrowego..., s. 2–4.

²³ Odpowiedź Narodowego Archiwum Cyfrowego..., s. 4–5; Odpowiedź Ministerstwa Kultury i Dziedzictwa Narodowego..., s. 1–2.

²⁴ Odpowiedź Ministerstwa Kultury i Dziedzictwa Narodowego..., s. 2.

dzie gromadziło zasób Internetu tworzony przez polskie instytucje publiczne, a drugie zarządzane przez BN (po uregulowaniu prawa autorskiego i EO) prawdopodobnie będzie archiwizowało pozostałą, nieurzędową część polskiego Internetu. Wydaje się, że rozwiązaniem które pozwoliłoby na ochronę polskiego dziedzictwa Internetu mogłoby być przyjęcie brytyjskiego modelu archiwizacji Internetu.

Wielka Brytania od kilkunastu lat z powodzeniem zabezpiecza swoje dziedzictwo Internetu i jest jednym z pierwszych państw, w których powstało archiwum Internetu²⁵. Wśród inicjatyw podejmowanych na rzecz archiwizacji brytyjskiego wycinka Internetu należy wymienić: UK Government Web Archive (od 2003 r.), UK Web Archive (od 2005 r.), UK Parliament Web Archive (od 2009 r.)²⁶. Prowadzone przez The National Archives, UK Government Web Archive (UKGWA) służy gromadzeniu stron internetowych oraz materiałów z kont mediów społecznościowych brytyjskiego rządu centralnego. Zasób UKGWA tworzą kolekcje: A to Z of archived websites, Video archive, Twitter archive, Flickr archive. Archiwalia Internetu UKGWA są w pełni dostępne online dzięki temu, że archiwizowana treść jest tworem instytucji publicznych²⁷. Założone przez The British Library, UK Web Archive (UKWA) stanowi największe krajowe archiwum Internetu, obejmujące swoim zasobem materiały o wartości naukowej i kulturowej z domeny brytyjskiej („uk”)²⁸. Gromadzenie materiałów internetowych jest możliwe dzięki Legal Deposit Libraries Act 2003 oraz The Legal Deposit Libraries (Non-Print Works) Regulations 2013. UKWA jest współpracą sześciu Bibliotek Depozytowych Wielkiej Brytanii (ang. UK Legal Deposit Libraries), egzekutorów przepisów legal deposit, znanych w Polsce jako przepisy EO. Archiwum złożone jest z trzech kolekcji zbiorów archiwalnych brytyjskiego Internetu²⁹. Najmniejszą z tych trzech kolekcji jest Open UK Web Archive, tworzona przez strony internetowe archiwizowane od 2004 roku, które dotyczą aspektów życia, kultury, czy też krajowych wydarzeń w Wielkiej Brytanii. Jest ona otwartą kolekcją ponieważ każdy może wyznaczyć stronę internetową do archiwizacji za pomocą formularza dostępnego na stronie UKWA (www.webarchive.org.uk)³⁰. Nie jest

²⁵ TOKARSKA-TRZASKOWSKA A., op. cit.

²⁶ *Collection guides – UK Web Archive*, <<https://www.bl.uk/collection-guides/uk-web-archive>> [dostęp 15.06.2021]; *UK Government Web Archive*, <<http://nationalarchives.gov.uk/webarchive/>> [dostęp 15.06.2021]; *UK Parliament Web Archive*, <<http://webarchive.parliament.uk/about>> [dostęp 15.06.2021].

²⁷ *The National Archives – Our role*, <<http://www.nationalarchives.gov.uk/about/our-role/>> [dostęp 15.06.2021]; *UK Government Web Archive – How to use the web archive*, <<http://www.nationalarchives.gov.uk/webarchive/information/>> [dostęp 15.06.2021].

²⁸ *IPC members...; Collection guides – UK Web Archive...*

²⁹ „Według [prawa brytyjskiego] wszystkie drukowane oraz elektroniczne publikacje, włączając w nie strony internetowe z Wielkiej Brytanii, muszą zostać zdeponowane w British Library i na prośbę pięciu innych Bibliotek Depozytowych. Są nimi – Biblioteka Narodowa Szkocji, Biblioteka Narodowa Walii, Biblioteka Bodlejańska, Biblioteki Uniwersytetu Cambridge i Kolegium Trójcy Świętej w Dublinie”. *UKWA – UK Web Archive*, <<https://www.webarchive.org.uk/>> [dostęp 15.06.2021]; *Collection guides – UK Web Archive...*

³⁰ *Collection guides – UK Web Archive...; UK Web Archive – Save a UK website*, <<https://www.webarchive.org.uk/en/ukwa/info/nominate>> [dostęp 15.06.2021].

to równoznaczne z tym, że każda wyznaczona strona internetowa zostanie zarchiwizowana, gdyż zgodę na to musi wydać właściciel tej strony, chyba że zawiera ona licencje Creative Commons³¹. Ponadto Open UK Web Archive ma charakter selektywny, co oznacza że tę kolekcję tworzą wybrane przez pracowników UKWA witryny internetowe wysokiej jakości, których kopie archiwalne zostały dokładnie sprawdzone i opatrzone odpowiednimi uwagami³². Kolejna kolekcja nazwana JISC UK Web Domain Dataset (1996-2013), powstała dzięki współpracy UKWA z Internet Archive oraz brytyjską organizacją non-profit JISC³³. Składa się ze zarchiwizowanej w latach 1996-2013 zawartości domeny brytyjskiej Internetu przez Internet Archive, która została pozyskana przez JISC. Ostatnią kolekcję tworzy najmłodsza Non-Print Legal Deposit UK Web Archive, która stanowi większość gromadzonego zasobu. Zawiera ona miliony stron internetowych z całej domeny brytyjskiej archiwizowanej corocznie od 2013 roku³⁴. Wprowadzone regulacje prawne The Legal Deposit Libraries (Non-Print Works) Regulations 2013 pozwoliły na automatyzację procesu archiwizacji. Gromadzenie materiałów jest prowadzone za pomocą odpowiedniego oprogramowania, bez konieczności otrzymywania każdorazowo zgody od osób dysponujących prawami autorskimi wobec stron internetowych³⁵. Zastosowanym rozwiązaniem, które pozwoliło wyjść naprzeciw prawom autorskim, było ograniczenie dostępu do tej kolekcji. Dostęp do kolekcji Non-Print Legal Deposit UK Web Archive można uzyskać tylko na komputerach w czytelnich sześciu Bibliotek Depozytowych Wielkiej Brytanii³⁶. Instytucją brytyjską, która także archiwizuje część krajowego Internetu jest Parliamentary Archives. Prowadzi ona UK Parliament Web Archive (UKPWA), składające się z pięciu kolekcji archiwaliów internetowych, które zostały podzielone według źródeł ich pochodzenia: Web Archives (strony internetowe), Twitter, YouTube, Instagram, Flickr. Dostęp do archiwalnych materiałów Internetu UKPWA – tak samo jak w przypadku UKGWA – nie jest w żaden sposób ograniczony³⁷.

³¹ *Collection guides – UK Web Archive...*; M. Wilkowski, *Archiwum brytyjskiego WWW: UK Web Archive*, <<http://historiaimedia.org/2011/11/09/archiwum-brytyjskiego-www-uk-web-archive/index.html>> [dostęp 16.07.2021].

³² *Collection guides – UK Web Archive...*; B. Konopa, *Archiwa Internetu jako nowe bazy źródłowe*, „Archiwa – Kancelarie – Zbiory”, 2018, nr. 9, s. 54–55, <https://www.researchgate.net/publication/330196868_Archiwa_Internetu_jako_nowe_bazy_zrodlowe> [dostęp 16.07.2021].

³³ *JISC UK Web Domain Dataset (1996-2013)*, <<http://data.webarchive.org.uk/opendata/ukwa.ds.2/>> [dostęp 15.06.2021].

³⁴ *Collection guides – UK Web Archive...*; *UK Web Archive – Frequently asked questions...*

³⁵ The Legal Deposit Libraries (Non-Print Works) Regulations 2013, no. 777, <<http://www.legislation.gov.uk/uksi/2013/777/contents/made>> [dostęp 15.06.2021]; M. Wilkowski, *Archiwum brytyjskiego WWW...*; B. Konopa, *Archiwa Internetu...*, s. 55.

³⁶ *Collection guides – UK Web Archive...*; *UKWA – UK Web Archive...*

³⁷ *Parliamentary Archives*, <<https://www.parliament.uk/business/publications/parliamentary-archives/>> [dostęp 15.06.2021]; *UK Parliament Web Archive...*

Należy dodać, że przy tworzeniu krajowych archiwów Internetu są stosowane różnorodne strategie archiwizacji zasobów internetowych³⁸. Zdaniem badacza archiwistyki Internetu Bartłomieja Konopy, odpowiednim podziałem tych strategii jest różniczenie ich na gromadzenie masowe (ang. bulk harvesting) oraz selektywne (ang. selective harvesting). W ramach pierwszej strategii archiwizowana jest konkretna domena, bądź grupa domen. Proces archiwizacji jest zautomatyzowany, zapewniony przez odpowiednie oprogramowanie do indeksowania Internetu, którego praca jest jedynie kontrolowana przez człowieka. Pozwala to na gromadzenie dużej ilości zasobów online, przy niewielkich nakładach pracy. Jednakże materiały pozyskane tą drogą mogą tracić na kompletności. Gromadzenie selektywne jak sugeruje nazwa, opiera się na manualnym wyborze za pomocą wcześniej ustalonych kryteriów „poszczególnych witryn i innych obiektów online”, przeznaczonych do archiwizacji. Strategia ta, w porównaniu do poprzedniej, pozwala na większą dokładność nad tym jakie materiały trafią do archiwum oraz większą kontrolę nad jakością uzyskanych archiwalnych kopii materiałów. Dodatkowo gromadzenie selektywne ułatwia opracowanie materiałów pod względem opisu metadanych, czy też ich grupowania, katalogowania. Selektywna archiwizacja wymaga jednak większych nakładów pracy człowieka niż przy archiwizacji masowej³⁹. Zatem klasyfikując wyżej wymienione brytyjskie archiwa Internetu w oparciu o przytoczone strategie archiwizacji można stwierdzić, że w przypadku projektu UKWA obecnie jest stosowane gromadzenie masowe oraz selektywne. Przy pozostałych projektach nie udało się ustalić jak przebiega archiwizacja materiałów internetowych. Można jedynie domniemywać, że archiwizacja podobnie jak w UKWA jest wspomagana odpowiednim oprogramowaniem. Wiadomo jednak, że The National Archives i Parliamentary Archives współpracują z firmą outsourcingową MirrorWeb świadczącą usługi w zakresie archiwizacji Internetu⁴⁰.

Model archiwizacji Internetu przyjęty przez Wielką Brytanie przypomina ten zaproponowany przez MKiDN, gdyż The National Archives i British Library (odpowiedniki NAC i BN, którym planuje się powierzenie archiwizacji polskiego Internetu) archiwizują zasób brytyjskiego Internetu. Zatem mając na uwadze rozwiązania przyjęte w Wielkiej Brytanii, zasób polskiego archiwum Internetu NAC mógłby składać się ze stron internetowych oraz materiałów opublikowanych w mediach społecznościowych, tworzonych przez polski rząd. Ze względu na źródło pochodzenia archiwaliów internetowych mogłyby one zostać podzielone na kolekcje zarchiwizowanych: stron internetowych, filmów na YouTube, tweetów, postów na Facebooku. Przy gromadzeniu tych materiałów można byłoby zastosować strategię gromadzenia masowego. Za-

³⁸ NIELSEN J., op. cit., s. 3, [dostęp 16.07.2021].

³⁹ KONOPA B., *Strategia selektywna jako narzędzie w archiwizacji Webu. Analiza wybranych przykładów*, „Archiwa – Kancelarie – Zbiory”, 2020, nr. 11, s. 102-103, <<https://apcz.umk.pl/czasopisma/index.php/AKZ/article/view/AKZ.2020.004>> [dostęp 16.07.2021].

⁴⁰ *MirrorWeb – Case study: The National Archives*, <<https://www.mirrorweb.com/the-national-archives-case-study/>> [dostęp 16.07.2021]; *UK Parliament Web Archive – About...*

sób polskiego archiwum Internetu mógłby zgodnie z prawem zostać w całości udostępniony na stronie szukajwarchiwach.gov.pl. Do gromadzenia i przechowywania archiwaliów można byłoby wykorzystać serwerownie Centralnego Repozytorium Cyfrowego Archiwów Państwowych (CRC AP) NAC.

BN zajmowałaby się archiwizacją pozostałej, nieurzędowej części Internetu. Dobrym rozwiązaniem – tak jak to miało miejsce w przypadku tworzenia zasobu UKWA – byłoby stworzenie trzech kolekcji w ramach polskiego archiwum Internetu BN, które stanowiłyby:

1. Kolekcja otwarta – składałaby się z archiwaliów internetowych pozyskanych i udostępnionych online za zgodą ich właścicieli, a także tych opartych na licencji Creative Commons. Kolekcję mogliby współtworzyć użytkownicy tego archiwum Internetu, sugerując wybrane materiały internetowe za pomocą formularza dostępnego na stronie internetowej archiwum. Pracownicy stosując strategię gromadzenia selektywnego, wydzielaliby sugerowane materiały. Zbiory całej kolekcji dzięki takiemu sposobowi ich gromadzenia mogłyby być w pełni dostępne online.

2. Kolekcja Internet Archive – byłaby polskim zasobem archiwalnym Internetu nabytym od Internet Archive, udostępnianym również online.

3. Kolekcja o ograniczonym dostępie – tę kolekcję tworzyłyby głównie zbiory archiwalne objęte prawami autorskimi, pozyskane drogą zautomatyzowanego procesu indeksacji polskiego Internetu (strategia gromadzenia masowego), do którego mogłoby dochodzić corocznie. Uzyskane tą drogą zbiory kolekcji mogłyby być dostępne tylko na komputerach w czytelnich bibliotek uprawnionych, dzięki czemu unikniętoby kolizji z prawem autorskim. Zgromadzone kolekcje lub tylko informacje o nich – w zależności od możliwości prawnej – można byłoby udostępnić na stronie Katalogu BN oraz przechowywać na serwerach Repozytorium Dokumentów Elektronicznych BN.

Ponadto do gromadzenia archiwaliów internetowych przez NAC i BN można byłoby wykorzystać oprogramowanie Heritrix udostępniane przez Internet Archive, dzięki czemu udałooby się zredukować koszty, wynikające z opracowywania technologii. Istnieje też możliwość zlecenia zadania stworzenia polskich archiwów Internetu firmie outsourcingowej, dzięki czemu mogłoby dojść do zredukowania kolejnych kosztów oraz przyspieszenia procesu wdrożenia archiwizacji polskiego Internetu. Na takie posunięcie zdecydował się na przykład Rząd Walii, zlecając firmie MirrorWeb archiwizację tworzonego przez ten rząd zasobu Internetu⁴¹. NAC i BN,

⁴¹ Firma MirrorWeb w zakresie świadczonych usług archiwizacji Internetu, poza tym obsługuje Bank Anglii, Izbę Lordów, Skarbiec Jej Królewskiej Mości oraz współpracuje z The National Archives i Parliamentary Archives. R. Gregory, *Wales' Digital History To Be Fully Archived And Searchable*, <<https://wales247.co.uk/wales-digital-history-to-be-fully-archived-and-searchable/>> [dostęp 15.06.2021]; *Mirror Web*, <<https://www.mirrorweb.com>> [dostęp 15.06.2021], cit. per: M. Wiśniewska-Drewniak, *Outsourcing archiwizacji domeny krajowej w Walii*, <<https://webarch.uw.edu.pl/blog/outsourcing-archiwizacji-webu/>> [dostęp 15.06.2021].

którym powierza się archiwizację polskiego Internetu, powinny prowadzić również aktywną współpracę zagraniczną w zakresie archiwizacji Internetu m.in. z IIPC, aby korzystać z najnowszych rozwiązań stosowanych w krajowych archiwach Internetu oraz niwelować napotkane problemy w tej materii. Ponadto należy podkreślić, że wielkość powodzenia archiwizacji polskiego Internetu, jak udowadnia przybliżony wcześniej projekt „’09 European Election Web Harvesting Project”, będzie zależeć także od dobrej woli oraz świadomości instytucji, które zostaną uznane za wytwórców archiwaliów internetowych. Zatem NAC i BN powinny prowadzić działania promocyjne, zmierzające do pogłębiania świadomości społeczeństwa na temat znaczenia dziedzictwa Internetu.

Bibliografia

Źródła

Ustawa z dnia 14 lipca 1983 r. o narodowym zasobie archiwalnym i archiwach, Dz.U. 1983, nr 38, poz. 173

Ustawa z dnia 7 listopada 1996 r. o obowiązkowych egzemplarzach bibliotecznych, Dz.U. 1996, nr 152, poz. 722.

Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne, Dz.U. 2005, nr 64, poz. 565.

Rozporządzenie Ministra Kultury i Dziedzictwa Narodowego z dnia 6 lutego 2008 r. w sprawie zmiany nazwy i zakresu działania Archiwum Dokumentacji Mechanicznej w Warszawie, Dz.U. 2008, nr 29, poz. 167.

The Legal Deposit Libraries (Non-Print Works) Regulations 2013, no. 777, <<http://www.legislation.gov.uk/uksi/2013/777/contents/made>> [dostęp 15.06.2021].

Odpowiedź Ministerstwa Kultury i Dziedzictwa Narodowego na wniosek o udostępnienie informacji publicznej z dnia 3 maja 2019 r.

Odpowiedź Biblioteki Narodowej na wniosek o udostępnienie informacji publicznej z dnia 27 czerwca 2019 r.

Odpowiedź Narodowego Archiwum Cyfrowego na wnioski o udostępnienie informacji publicznej z dnia 31 marca 2019 r.

Opracowania

About the IIPC, <<http://netpreserve.org/about-us/>> [dostęp 15.06.2021]

British Library, <<https://www.bl.uk/>> [dostęp 15.06.2021].

ChORAŹYCZEWSKI W., *Archiwistyka dla początkujących*, Toruń 2014, <<https://repozytorium.umk.pl/handle/item/2191>> [dostęp 15.06.2021].

DĄBROWSKA E., *Problem archiwizacji internetu w kontekście egzemplarza obowiązkowego: sytuacja w Polsce i wybranych krajach europejskich*, „Biuletyn EBIB”, 2017, nr. 172, <<http://ebibojs.pl/index.php/ebib/article/view/146>> [dostęp 15.06.2021].

GACA S., *Archiwizacja Internetu – Polska na tle świata*, Praca licencjacka napisana pod kierunkiem dra Bartosza Drzewieckiego, Kraków 2019.

GREGORY R., *Wales' Digital History To Be Fully Archived And Searchable*, <<https://wales247.co.uk/wales-digital-history-to-be-fully-archived-and-searchable/>> [dostęp 15.06.2021].

IIPC members, <<http://netpreserve.org/about-us/members/>> [dostęp 15.06.2021].

Internet Archive, <<https://archive.org/>> [dostęp 15.06.2021].

JISC UK Web Domain Dataset (1996-2013), <<http://data.webarchive.org.uk/opendata/ukwa.ds.2/>> [dostęp 15.06.2021].

Kemp S., *Global Digital Report 2019*, <<https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates/>> [dostęp 15.06.2021].

KONOPA B., *Archiwa Internetu jako nowe bazy źródłowe*, „Archiwa – Kancelarie – Zbiory”, 2018, nr. 9, s. 49–62, <<https://apcz.umk.pl/czasopisma/index.php/AKZ/article/view/AKZ.2018.003/16523>> [dostęp 15.06.2021].

KONOPA B., *Badanie kryzysu opioidowego z wykorzystaniem archiwum Webu*, <<https://webarch.uw.edu.pl/blog/archiwum-kryzysu-opioidowego/>> [dostęp 15.06.2021].

KONOPA B., *Reborn Digital i Black Box – wpływ procesu archiwizacji na zasób archiwów Webu*, „Archiwa – Kancelarie – Zbiory”, 2019, nr. 10, s. 147–167, <<https://apcz.umk.pl/czasopisma/index.php/AKZ/article/view/AKZ.2019.008/24589>> [dostęp 15.06.2021].

KONOPA B., *Strategia selektywna jako narzędzie w archiwizacji Webu. Analiza wybranych przykładów*, „Archiwa – Kancelarie – Zbiory”, 2020, nr. 11, s. 97–117, <<https://apcz.umk.pl/czasopisma/index.php/AKZ/article/view/AKZ.2020.004>> [dostęp 16.07.2021].

KONOPA B., *Witryna internetowa – Dokumentacja czy publikacja?*, „Biuletyn EBIB”, 2017, nr. 172, <<http://ebibojs.pl/index.php/ebib/article/view/147>> [dostęp 15.06.2021].

MirrorWeb, <<https://www.mirrorweb.com/>> [dostęp 15.06.2021].

Narodowe Archiwum Cyfrowe – *Wizja Projekt Ludzie*, Narodowe Archiwum Cyfrowe, Warszawa 2010, <https://www.nac.gov.pl/wp-content/uploads/2015/05/NAC_wizja_projekt_ludzie_WWW.pdf> [dostęp 15.06.2021].

Netcraft – *May 2019 Web Server Survey*, <<https://news.netcraft.com/archives/2019/05/10/may-2019-web-server-survey.html>> [dostęp 15.06.2021].

NIELSEN J., *Using Web archives in research*, tłum. WILKOWSKI M., <http://webarch.uw.edu.pl/wp-content/uploads/2018/12/janne_nielsen_wykorzystanie_archiwow_webu_w_badaniach.odt> [dostęp 15.06.2021].

Opioid Epidemic web archive, <<https://archive-it.org/collections/9280>>, [dostęp 15.06.2021].

Parliamentary Archives, <<https://www.parliament.uk/business/publications/parliamentary-archives/>> [dostęp 15.06.2021].

Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020, Warszawa 2009, <<https://www.nac.gov.pl/wp-content/uploads/2015/05/Program-digitalizacji-2009-2020-1.pdf>> [dostęp 15.06.2021].

Słowniczek archiwalny Programu Archiwizacji Webu Biblioteki Kongresu, tłum. WILKOWSKI M., <http://webarch.uw.edu.pl/wp-content/uploads/2018/12/slowniczek_archiwalny_LOC.odt> [dostęp 15.06.2021].

SPEAKER S. L. i MOFFATT C., *The Opioid Epidemic: Collecting Now For Future Research*, <<https://circulatingnow.nlm.nih.gov/2018/09/27/the-opioid-epidemic-collecting-now-for-future-research/>> [dostęp 15.06.2021].

ŚLASKA K. i WASILEWSKA A., *Archiwizacja Internetu – sytuacja w polskim prawie z punktu widzenia bibliotekarzy*, „Biuletyn EBIB”, 2012, nr. 128, <http://www.ebib.pl/?page_id=380> [dostęp 15.06.2021].

The National Archives, <<http://www.nationalarchives.gov.uk/>> [dostęp 15.06.2021].

TOKARSKA-TRZASKOWSKA A., *Krótką historia archiwizowania Webu*, <<https://webarch.uw.edu.pl/zasoby/krotka-historia/>> [dostęp 15.06.2021].

UK Parliament Web Archive, <<http://webarchive.parliament.uk/>> [dostęp 15.06.2021].

UKWA – UK Web Archive, <<https://www.webarchive.org.uk/>> [dostęp 15.06.2021].

WILKOWSKI M., *Archiwistyka Webu: status, cele, metody (nagranie)*, <<https://www.youtube.com/watch?v=fPuC2VIK2N8>> [dostęp 17.07.2021].

WILKOWSKI M., *Archiwum brytyjskiego WWW: UK Web Archive*, <<http://historiaimedia.org/2011/11/09/archiwum-brytyjskiego-www-uk-web-archive/index.html>> [dostęp 15.06.2021].

WIŚNIEWSKA-DREWNIAK M., *Outsourcing archiwizacji domeny krajowej w Walii*, <<https://webarch.uw.edu.pl/blog/outsourcing-archiwizacji-webu/>> [dostęp 15.06.2021].

Jak cytować: Gaca, S. (2021). Propozycja modelu archiwizacji polskiego Internetu na bazie doświadczeń brytyjskich. *Officina Historiae* 4, 79-90.



© 2021 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.