

Data przesłania artykułu: 15 IV 2019 r.

Data przyjęcia artykułu do druku: 19 V 2019 r.

DOI: <http://dx.doi.org/10.12775/AKZ.2019.004>



WOJCIECH WOŹNIAK

(Prokuratoria Generalna Rzeczypospolitej Polskiej)

ARCHIWIZACJA INTERNETU – PRÓBA PODSUMOWANIA DOTYCHCZASOWYCH PRAC I USTALEŃ

Słowa kluczowe

archiwizacja Internetu

Keywords

Internet archiving

Streszczenie

Artykuł zawiera omówienie dotychczasowego stanu dyskusji nt. archiwizacji Internetu w Polsce na tle rozważań pojęciowych związanych z rozumieniem terminu „archiwizacja Internetu”. Omówiono kluczowe aspekty archiwizacji stron WWW z punktu widzenia



Wojciech Woźniak, politolog, archiwista, doktor nauk humanistycznych w zakresie historii. Były pracownik AP w Opolu i NAC, którego był dyrektorem. W latach 2016–2018 Naczelnny Dyrektor Archiwów Państwowych. Obecnie pracownik Prokuratury Generalnej RP. Jest współautorem polskiego tłumaczenia standardu EAD oraz opracowania *Narodowe Archiwum Cyfrowe. Wizja, projekt, ludzie* (Warszawa 2010), autorem licznych publikacji w periodykach historycznych i archiwalnych. Współpracuje z Instytutem Historycznym Uniwersytetu Warszawskiego, gdzie prowadzi zajęcia z przedmiotu elektroniczne zarządzanie dokumentacją.

E-mail: wojciech.v@gmail.com

ORCID ID: 0000-0002-6042-6068

archiwów państwowych oraz propozycje realizacji tego zadania sformułowane ze strony środowiska bibliotekarzy i archiwistów.

Summary

Internet archiving – an attempt at summary of previous works and establishments

The article describes the state of discussion about archiving the Internet in Poland, on the background of terminological deliberations on understanding the term “Internet archiving”. The article discusses key aspects of archiving websites from the point of view of state archives, as well as proposals of completing this task formed by librarians and archivists.

Zagadnienie archiwizacji Internetu nie cieszy się dużym zainteresowaniem. Zbadaczy z szeroko rozumianej dziedziny archiwalnej. Pierwsze wzmianki o potrzebie myślenia o archiwizowaniu sieci znaleźć można w artykule Huberta Wajsa, w którym autor pisze o konieczności zwrócenia uwagi na obiekty cyfrowe, niemające formy dokumentów¹: „blogi, instant messaging, twitters (świergotki) czy fora i wiki: mające różne formaty i składające się z pomieszanych wielu typów plików multimedialnych”². Zakres potencjalnego zainteresowania archiwistów zakreślony został przez H. Wajsa bardzo szeroko, co jest zgodne z podobnym podejściem do archiwizacji Internetu prezentowanym przez Marcina Wilkowskiego, który pisząc o archiwizacji Internetu, mieści w tym pojęciu nie tylko klasyczne strony WWW, ale również serwisy społecznościowe i pocztę elektroniczną³. W literaturze archiwistycznej obecne są dwa teksty poświęcone projektowi Archiwum Internetu realizowanemu przez Narodowe Archiwum Cyfrowe⁴ oraz przeglądowy tekst Anny Sobczak⁵. Znacznie bogatsza pod tym względem jest literatura z dziedziny bibliologii, obejmująca przede wszystkim

¹ Zob. też D. Grot, *Ku przeszłości otwartej*, Warszawa 2015, s. 46.

² H. Wajs, *Dokument elektroniczny jako obiekt archiwalny*, „Archiwista Polski” 2011, nr 16, s. 23.

³ M. Wilkowski, *Oddolne archiwizacje Internetu jako działania społeczne*, „Archiwa – Kancelarie – Zbiory” 2015, nr 6 (8), s. 207–220. <http://dx.doi.org/10.12775/AKZ.2015.007>.

⁴ F. Kłębczyk, M. Jędralska, *Serwis „Archiwum Internetu” na tle ogólnych problemów archiwizacji zasobów sieciowych*, „Biuletyn EBIB” 2012, nr 1 (128), s. 1–9. http://www.ebib.pl/images/stories/numery/128/128_jedralska.pdf; F. Kłębczyk, *Archiwizacja zasobów Internetu: kierunki i wyzwania*, „Archiwista Polski” 2012, nr 3, s. 105–112.

⁵ Zob. A. Sobczak, *Internet jako globalne archiwum społeczne – rozważania na temat roli Internetu w dokumentowaniu dziejów ludzkości*, [w:] *Toruńskie Konfrontacje Archiwalne*,

dwa wydania „Biuletynu EBIB” poświęcone w całości zagadnieniu archiwizacji Internetu⁶. Inne publikacje dotyczące tego zagadnienia znaleźć można w licznych opracowaniach zbiorowych z zakresu historii i kultury⁷. Brak jest jednak metodycznego i zwięzłego omówienia stanowiska instytucji odpowiedzialnych za przechowywanie dziedzictwa wobec samego zagadnienia. Wynika to zarówno ze złożoności zagadnienia, jak i trwającej od lat w Polsce dyskusji o tym, który podmiot (lub podmioty) powinien zajmować się archiwizacją treści zamieszczonych w Internecie, a także z niedostatku polskich inicjatyw związanych z archiwizacją Internetu, na co wskazuje Maria Nowina-Konopka⁸. Celem niniejszego tekstu jest prezentacja stanu dyskusji pomiędzy archiwami i bibliotekami oraz spojrzenia na archiwizację Internetu z punktu widzenia archiwistyki.

Archiwista staje wobec treści zamieszczonych w Internecie odrobinę bezradny. Wyrasta przed nim bowiem złożona rzeczywistość informacyjna zakodowana w najróżniejsze formy przekazu. Najpopularniejsze z nich to strony WWW pełniące funkcje informacyjne. Obok nich funkcjonują strony WWW stanowiące interaktywne serwisy i bramy e-usług, za którymi kryją się systemy dziedzinowe lub rejestry publiczne. Spektrum form, w jakich prezentowane są treści zamieszczane w Internecie, dopełniają platformy wypełniane zawartością tworzoną przez użytkowników (user generated content – UGC; lub user created content – UCC), głównie portale społecznościowe⁹. Archiwista, spoglądając na ten obraz, próbuje dostrzec przede wszystkim twórcę dokumentacji i dokumentację przez niego wytworzoną, ewentualnie treści niedokumentacyjne (czy: niedokumentarne, jak proponuje Dariusz Grot¹⁰). Wydaje się to dość proste, jeśli obserwacji poddać pojedynczy serwis WWW określonego podmiotu. Pojawia się jednak coraz więcej serwisów służących wielu podmiotom jako wspólne

t. 4: *Nowa archiwistyka – Archiwa i archiwistyka w ponowoczesnym kontekście kulturowym*, red. W. Chorążyczewski, W. Piasek, A. Rosa, Toruń 2014, s. 237–247.

⁶ „Biuletyn EBIB” 2012, nr 1 i 2017, nr 2.

⁷ Fenomenowi archiwum w kulturze współczesnej poświęcony został np. numer 4 „Kultury Współczesnej”; zob. „Kultura Współczesna” 2011, nr 4, a w nim zob. M. Halawa, *Nowe media i archiwizacja życia codziennego*, s. 27–41. Kilka tekstów o archiwizacji Internetu przygotował też Marcin Wilkowski.

⁸ Zob. M. Nowina-Konopka, *Problematyka zarządzania informacją w procesie archiwizacji zasobów Internetu*. „Media – Kultura – Komunikacja Społeczna” 2014, t. 10, nr 1, s. 30–33.

⁹ Dość oryginalną formą specyficznego archiwum mediów społecznościowych w kontekście aktywności publicznej jest serwis prowadzony przez Fundację ePaństwo <https://mojepanstwo.pl/media> (dostęp 8.04.2019).

¹⁰ Zob. D. Grot, dz. cyt., s. 46.

platformy – ePUAP, obywatel.gov.pl czy wspólny serwis rządowych stron WWW: portal.gov.pl¹¹. Są wreszcie treści budzące wątpliwości, czy w ogóle powinny znaleźć się w zakresie zainteresowania archiwistów. Granice wydają się nieostre, a wątpliwości nabierają charakteru genetycznego, bo jako kwestia wyróżnienia wartości archiwalnej treści w sieci podnoszony jest argument istnienia ich charakteru dokumentacyjnego. Co więcej, postawienie sprawy zabezpieczania treści z Internetu jako „archiwizacji Internetu” każe nagle spojrzeć z perspektywy, do której archiwista nie jest przyzwyczajony. Zamiast patrzeć przez pryzmat twórcy dokumentacji, ma spojrzeć na całość wytworzonej treści, na ogromny ocean informacji i podjąć się archiwizacji tego oceanu, a nie jego poszczególnych elementów o określonej proveniencji. Na zagadnienie nakładają się dodatkowo aspekty techniczne o nietrywialnym charakterze.

Archiwa państwowe, zwłaszcza Narodowe Archiwum Cyfrowe, zobowiązane są do troski o sferę rzeczywistości informacyjnej, jaką jest Internet. Rozporządzenie Ministra Kultury i Dziedzictwa Narodowego w sprawie zmiany nazwy i zakresu działania Archiwum Dokumentacji Mechanicznej z dnia 6 lutego 2008 r. stanowi, że NAC „obejmuje: [...] materiały archiwalne [...] d) stanowiące duże zbiory danych o ustalonej i zdefiniowanej strukturze, w tym rejestry publiczne, e) stanowiące zbiory danych zamieszczone w Internecie”¹². Bardziej szczegółowe regulacje znajdują się w statucie NAC, który zadania archiwum określa następująco: gromadzenie i przechowywanie oraz ewidencjonowanie i opracowywanie zbiorów danych zamieszczonych w Internecie i uznanych za materiały archiwalne, wytwarzanych w szczególności przez podmioty uprawnione do korzystania z domeny gov.pl, oraz zbiorów danych o ustalonej i zdefiniowanej strukturze, w tym rejestrów publicznych¹³. Ramy prawne zostały więc określone (nie miejsce tu na rozważanie, czy nie powinny zostać poddane ewaluacji po nabytych doświadczeniach), choć wydaje się, że nadal brak ram metodycznych. Próbę ustalenia ich podstaw warto rozpocząć od zbadania pojęcia *archiwizacji Internetu* w odniesieniu do pojęcia *archiwizacji*.

Według *Słownika języka polskiego* archiwizacja to „zbieranie i porządkowanie dokumentów, materiałów archiwalnych”¹⁴. Na gruncie archiwistyki pojęcie

¹¹ Zob. <https://www.gov.pl> (dostęp 8.04.2019).

¹² Dz. U., 2008 r., nr 19, poz. 167.

¹³ Decyzja nr 6 Naczelnego Dyrektora Archiwów Państwowych z dnia 27 lutego 2015 r. w sprawie nadania statutu Narodowemu Archiwum Cyfrowemu z późniejszymi zmianami § 6 ust. 1h); 2g), zob. <https://nac.ssdip.bip.gov.pl/statut/statut.html> (dostęp 8.04.2019).

¹⁴ *Słownik języka polskiego*, t. 1, Warszawa 2003, s. 117.

archiwizacji¹⁵ wpisuje się w szersze pojęcie procesu archiwotwórczego. Bohdan Ryszewski pisał, że na archiwizację składa się selekcja i opracowanie archiwalne. Jest to drugi komponent procesu archiwotwórczego następujący po procesie aktotwórczym. „Archiwizacja – pisał B. Ryszewski – to cały kolejny proces, dokonujący się częściowo jeszcze u twórcy registratury, a następnie w archiwum. W trakcie tego procesu registratura zostaje poddana jednorazowej lub przebiegającej etapami selekcji [...] oraz opracowaniu archiwalnemu. Selekcja – w moim rozumieniu – powinna polegać na wyodrębnieniu dokumentacji odbijającej główne i specyficzne funkcje jej twórcy oraz usunięciu z tej dokumentacji tych dokumentów, które zawierają informacje powtarzalne, bez naruszania relacji między dokumentami i jednostkami. Opracowanie archiwalne natomiast polegać winno – jak wiadomo – na rekonstrukcji i ewentualnej przebudowie układu kancelaryjnego w układ archiwalny (bez naruszania ww. relacji) oraz sporządzeniu pomocy archiwalnych. W wyniku selekcji i opracowania archiwalnego nastąpi przekształcenie kancelaryjnego czy też registraturalnego systemu informacyjnego w podstawowy archiwalny system informacyjny, za jaki można uznać zespół archiwalny”¹⁶. Podobnie do sprawy podchodzili autorzy podręcznika *Archiwistyka* – również opisywali archiwizację jako drugi etap procesu archiwotwórczego, w którym następuje przemiana registratury w efekcie selekcji i opracowania archiwalnego¹⁷.

Z kolei Halina Robótka w definicji archiwizacji kładła nacisk na aspekty związane z opracowaniem archiwalnym, którego efektem ma być uporządkowanie zespołu i wytworzenie pomocy archiwalnych¹⁸. Jednocześnie wyróżniła „procesy archiwizacji”, które „[...] obejmują czynności selekcji (oceny wartości połączonej z brakowaniem), przechodzenie do archiwum bieżącego, historycznego, łączenie z opracowaniem archiwalnym”¹⁹.

Elementy tak zdefiniowanego pojęcia archiwizacji są obecne we współczesnym języku potocznym, w tym biznesowym. Istnieje szereg firm świadczących „usługi archiwizacyjne”, w których zakresie najczęściej znajduje się przechowy-

¹⁵ H. Robótka zwraca uwagę, że termin ten nie ma odpowiedników w innych językach, zob. H. Robótka, *Wprowadzenie do archiwistyki*, Toruń 2002, s. 77.

¹⁶ B. Ryszewski, *Problemy i metody badawcze archiwistyki*, Toruń 1985, s. 87–88. Zob. też H. Robótka, *Wprowadzenie do archiwistyki*, Toruń 2003, s. 77–78.

¹⁷ Zob. H. Robótka, B. Ryszewski, A. Tomczak, *Archiwistyka*, Toruń 1989, s. 18. Zob. też H. Robótka, *Wprowadzenie*, s. 71–72.

¹⁸ Zob. H. Robótka, *Opracowanie i opis archiwaliów. Podręcznik akademicki*, Toruń 2010, s. 237.

¹⁹ Tamże, s. 244.

wanie, porządkowanie i udostępnianie dokumentacji. Dodatkowo firmy tego rodzaju zajmują się wspieraniem zarządzania dokumentacją poprzez pomoc przy przygotowywaniu normatywów kancelaryjnych oraz w procesie niszczenia dokumentacji niearchiwalnej. Wielość sposobów postrzegania zakresu pojęcia archiwizacji w tych przypadkach skorelowana jest najczęściej z zakresem działania firmy²⁰. Samo zjawisko percepcji pojęcia archiwizacji w społeczeństwie zdaje się wymagać odrębnej analizy, a dla celów tego tekstu owe znaczenia „komercyjne” zostaną pominięte.

W *Polskim słowniku archiwalnym* brak pojęcia archiwizacji, pojawia się natomiast termin „archiwalizacja” oznaczający: „1. Proces tracenia przez materiały archiwalne twórcy zespołu związku z jego bieżącą działalnością; 2. Przygotowywanie i przekazywanie materiałów archiwalnych do archiwum”²¹. Jest to pojęcie, które jednak nie przyjęło się w języku archiwistów²².

W trakcie dyskusji pomiędzy archiwami i bibliotekami na kwestie terminologiczne zwrócił uwagę Naczelny Dyrektor Archiwów Państwowych – w piśmie z 18 października 2013 r. skierowanym do Departamentu Dziedzictwa Kulturowego MKiDN zauważył, że w rozumieniu potocznym archiwizacja jest też terminem technicznym, oznaczającym wykonywanie kopii bezpieczeństwa (backup). Ponieważ język informatyki jest młodszy od języka archiwistyki, można postawić tezę, że doszło do zawłaszczenia terminu „archiwizacja” przez dziedzinę informatyki²³. Na tym polu archiwum to (1) pliki powstające w wyniku

²⁰ Zob. np. „Głównym celem archiwizacji dokumentów księgowych, jak i innych akt jest przede wszystkim właściwe ich zabezpieczenie przed zniszczeniem, szybki i bezproblemowy dostęp do informacji w nich zawartych oraz opracowanie dokumentów zgodnie z polskim prawem”, <https://www.archiwizacjadokumentow.com.pl/uslugi/archiwizacja-dokumentow/>; „Odpowiednia archiwizacja umożliwia szybkie dotarcie do potrzebnych danych, skuteczniejsze zarządzanie informacją oraz ma istotny wpływ na sukces rynkowy firmy. Konieczność przechowywania dokumentów jest regulowana przepisami, według których należy archiwizować ważne informacje firmowe do czasu ich przedawnienia”, <http://nf.pl/przedsiębiorca/opanuj-chaos-co-musisz-wiedziec-o-archiwizacji-dokumentow,54806,222>; „Archiwizacja dokumentów to nic innego, jak ich porządkowanie według określonych zasad i właściwe przechowywanie”, <http://www.lisaelama.com/3-fakty-o-archiwizacji-dokumentow/>; „Ujmując rzecz w możliwie jak największym skrócie, archiwizacja jest to kompleksowy proces obejmujący tworzenie, zabezpieczenie i klasyfikowanie w odpowiedni sposób całej dokumentacji”, <https://www.projektgamma.pl/strefa-wiedzy/artykuly-ekspertkie/ustawa-o-archiwizacji-dokumentow> (dostęp 8.04.2019).

²¹ *Polski słownik archiwalny*, red. W. Maciejewska, Warszawa 1974, s. 18.

²² Zob. S. Dubisz, „Archiwum”, „archiwizacja”, „archiwizować”, „Poradnik Językowy” 2013, z. 7, s. 113–115.

²³ Zob. W.K. Roman, *Współczesna kultura dokumentacyjna*, Toruń 2013, s. 212–213.

procesu kompresji wielu plików (np. w formatach: ZIP, ARJ, RAR), co prowadzi do znacznej oszczędności miejsca na dysku lub (2) zbiór plików zawierających dane archiwalne²⁴, archiwizacja zaś będzie oznaczać zapewne proces kompresji lub proces tworzenia plików z danymi archiwalnymi. Na zawłaszczenie pojęcia archiwum przez informatyków zwrócił uwagę Paweł Perzyna: „[...] zmiany w rozszerzaniu znaczenia terminu »archiwum« poszły już tak daleko, że adepci informatyki będą go używać dalej głównie do oznaczania skompresowanych plików danych, natomiast archiwiści będą go odnosić do miejsca przechowywania danych (zapisów, dokumentów) niemających już bieżącej wartości użytkowej, bez względu na ich kwalifikację archiwalną i rodzaj nośnika, na jakim się one znajdują”²⁵. Podobne argumenty podnosiła Wanda Roman²⁶.

Pewnym kompromisem pojęciowym okazuje się natomiast wykorzystanie terminu „archiwizacja” do określenia działalności polegającej na „zapewnieniu ciągłości dostępu do materiałów należących do dziedzictwa cyfrowego”²⁷. W tym kontekście archiwizacja to nade wszystko aktywność techniczna polegająca na ciągłym podejmowaniu działań zapewniających trwałość i dostępność treści cyfrowych. Podobnie termin „archiwizacja” definiowano w gronie specjalistów budujących biblioteki cyfrowe jako „[...] ogół metod i praktyk służących zabezpieczeniu i długoterminowemu przechowywaniu wytwarzanych zasobów cyfrowych. Na archiwizację składają się: infrastruktura techniczna, oprogramowanie, sprzęt, interfejsy użytkowników, mechanizmy wymiany danych, mechanizmy zabezpieczania i migracji danych, obsługa techniczna”²⁸. W wymiarze techniki przechowywania pojęcie archiwizacji umieszczał również Damian Kusnik, przedstawiając koncepcję zapisywania danych cyfrowych na mikrofilmach²⁹.

²⁴ P. Adamczewski, *Leksykon informatyki stosowanej*, Poznań 2000, s. 18.

²⁵ P. Perzyna, *Problemy terminologiczne wynikające z komputeryzacji archiwów* [w:] *Komputeryzacja i digitalizacja w archiwach*. (Symposia Archivistica, t. 2), red. R. Leśkiewicz, A. Żeglińska, Warszawa 2016, s. 52–53.

²⁶ Zob. W.K. Roman, *Terminologia archiwalna – ewolucja czy rewolucja* [w:] *Toruńskie Konfrontacje Archiwalne, t. 2: Teoria archiwalna wczoraj–dziś–jutro*, red. W. Chorążyczewski, A. Rosa, Toruń 2011, s. 203–217.

²⁷ Zob. G. Gmiterek, *Długoterminowa archiwizacja zasobów cyfrowych*, „Res Historica” 2013, nr 35, s. 214. Zob. też A. Januszko-Szakiel, *Długoterminowa archiwizacja zasobów cyfrowych – program dla polskich bibliotek*, „Przegląd Biblioteczny” 2011, nr 2.

²⁸ T. Kalota, M. Szala, *Organizacja i logistyka digitalizacji* [w:] *Biblioteki cyfrowe*, red. M. Janiak, M. Krakowska, M. Próchnicka, Warszawa 2012, s. 438.

²⁹ Zob. D. Kusnik, *Długoterminowa archiwizacja danych cyfrowych na mikrofilmach*, „Archiwista Polski” 2014, nr 2, s. 21–30.

Obok pojęcia archiwizacji w ujęciu archiwistycznym i informatycznym pojawiło się trzecie znaczenie: „archiwizacja Internetu”. O konieczności wyodrębnienia tego pojęcia decyduje jego charakter, który zdaje się sytuować je w połowie drogi pomiędzy powyższymi znaczeniami. „Proces archiwizacji informacji z Web-u polega na poszukiwaniu, gromadzeniu i organizacji źródeł informacji w celu zabezpieczenia ich przed zniknięciem z WWW”³⁰. Dość często archiwizowanie Internetu bywa sprowadzane do aspektu technicznego i wówczas jest definiowane jako *harvesting*, czyli „zautomatyzowany proces gromadzenia zbiorów i metadanych, które są następnie indeksowane i składowane w archiwum cyfrowym według ściśle określonych parametrów”³¹. Powyższe definicje abstrahują od wartości informacji – nie odnoszą się do tej kwestii, czyli nie precyzują, czy i jakie kryteria powinny zostać zastosowane do wyboru ani które informacje (a może wszystkie?) powinny zostać zabezpieczone. U podstaw metodologicznych archiwizacji Internetu leży przeświadczenie, że uzasadnieniem podjęcia działań związanych z zabezpieczaniem danych jest samo ich pojawienie się w sieci. W najszerszym rozumieniu podstawą działań związanych z archiwizacją WWW jest chęć zabezpieczenia całości Internetu, czyli tego wszystkiego, co zostało w sieci zamieszczone (bez względu na to, czy dostępne jest w sposób otwarty, czy też wymaga szczególnych uprawnień).

Pewną próbą doprecyzowania pojęcia archiwizacji Internetu była wypracowana na spotkaniu przedstawicieli NDAP i BN w styczniu 2014 r. definicja, zgodnie z którą „pod pojęciem archiwizacji Internetu rozumiane będzie działanie polegające na zbieraniu treści prezentowanych w Internecie na stronach WWW poprzez pozyskiwanie kodu stron internetowych, przechowywanie tego kodu oraz jego udostępnianie na określonych zasadach, mające na celu zabezpieczenie w celach historycznych informacji prezentowanych w Internecie w danym okresie”³². Nie wydaje się jednak, że podejście to wnosi nowe elementy do wcześniej prezentowanych definicji. Skupiono się raczej na doprecyzowaniu technicznym („pozyskiwanie kodu stron internetowych”) oraz przedmiotowym („strony www”). Wskazano też trochę odmiennie cel archiwizacji Internetu – już

³⁰ M.A. Janowska, *Biblioteki akademickie – trendy dotyczące zasobów elektronicznych* [w:] *Informacja dla nauki a świat zasobów cyfrowych*, red. H. Ganińska, Poznań 2008, s. 168. https://library.put.poznan.pl/konf_idn/art/Konferencja%20BGPP%202008.pdf (dostęp 8.04.2019).

³¹ L. Derfert-Wolf, *Archiwizacja Internetu – wprowadzenie i przegląd wybranych inicjatyw*, „Biuletyn EBIB” 2012, nr 1 (128), s. 8. http://www.ebib.pl/images/stories/numery/128/128_derfert.pdf (dostęp 8.04.2019).

³² Pismo NDAP do dyrektora BN z dn. 4.03.2014 r., DKN.400.9.2013.

nie jako zabezpieczenie przed zniknięciem z sieci, ale jako „zabezpieczenie w celach historycznych informacji prezentowanych w Internecie w danym okresie”. Ponieważ jednak te zmiany nie wydają się zasadnicze, a sama definicja przez fakt, że została sformułowana na zamkniętym spotkaniu i nie zyskała szansy rozpowszechnienia, nie będzie brana pod uwagę w dalszych analizach.

Inne rozumienie archiwizacji Internetu zdaje się wynikać z analizy Bartłomieja Konopy, który spogląda na proces holistycznie i zwraca uwagę, że w rozumieniu bibliotekarzy i badaczy informacji ma on stanowić narzędzie badania zjawiska, jakim jest Internet, a nie poszczególnych całości informacyjnych, którymi są treści publikowane przez poszczególne podmioty (czy to indywidualne, czy zbiorowe)³³. Takim podejściem autor zdaje się wybiegać poza zasadniczą dyskusję związaną z gromadzeniem materiału do badań (treści zamieszczonych w Internecie), a sięga już do sfery prowadzenia badań tak pozyskanej informacji.

Żeby zrozumieć, czym jest archiwizacja Internetu i jakie zadania stoją przed archiwami państwowymi w kontekście treści zamieszczanych w Internecie, należy przeanalizować oba pojęcia archiwizacji: wywodzące się z teorii archiwistycznej oraz używane na oznaczenie działań zabezpieczających wobec treści zamieszczonych w Internecie. W tym celu oba pojęcia zostaną rozłożone na składowe w trzech wymiarach: aktywności, przedmiotu i celu.

	Archiwizacja według metodyki archiwalnej	Archiwizacja Internetu
Aktywność	selekcja	poszukiwanie
	akcesja	gromadzenie
	opracowanie	organizacja
Przedmiot	dokumentacja	źródła informacji
Cel	przekształcenie systemu informacyjnego	zabezpieczenie treści przed zniknięciem z WWW

Archiwizacja w ujęciu metodyki archiwalnej oznacza wybór (*selekcję*). Zakłada się, że z całości dokumentacyjnej wytworzonej przez dany podmiot wybrana zostanie część dokumentacji, która będzie przechowywana przez określony czas dla celów dowodowych lub użytecznych podmiotu, oraz część dokumentacji,

³³ Zob. B. Konopa, *Witryna internetowa – dokumentacja czy publikacja?*, „Biuletyn EBIB” 2017, nr 2, s. 1. <http://open.ebib.pl/ojs/index.php/ebib/article/view/528/684> (dostęp 8.04.2019).

która zostanie potraktowana jako materiały archiwalne i poddana procesowi wieczystego zachowania. Archiwizacja Internetu zajmuje się *poszukiwaniem*, czyli lokalizowaniem treści. Sytuacja jest więc zgoła odmienna od archiwizacji w rozumieniu archiwistyki, która rozpoczyna się już w określonym podmiocie wytwarzającym dokumentację, a więc prowadzącym procesy aktotwórcze i posiadającym (być może na różnym poziomie) określoną kulturę dokumentacyjną³⁴, w której mieści się postępowanie z dokumentacją, a więc i jej zachowywanie.

W pojęciu archiwizacji Internetu pojawia się aspekt *gromadzenia*, dość mocno zarysowany i często dominujący wymiarem technicznym nad całością pojęcia. Wynika to zresztą z preferowanego w tym obszarze rozwiązania, jakim jest aktywne zbieranie (harwestowanie) danych zamieszczonych w sieci przez podmiot zajmujący się archiwizacją Internetu. W definicji tradycyjnej gromadzenie (*akcesja*) wydaje się być raczej traktowane jako rzeczywistość czysto organizacyjna, sprowadzająca się do przekazania dokumentacji do archiwum (zakładowego lub państwowego). W tym wypadku zakłada się aktywność obu stron: podmiotu będącego twórcą dokumentacji i podmiotu prowadzącego archiwum.

Obie definicje odnoszą się do aspektu odpowiedniego przygotowania danych. W tradycyjnym pojmowaniu częścią procesu archiwizacji (lub nią samą w rozumieniu H. Robótki) jest *opracowanie*. W pojęciu archiwizacji Internetu pojawia się *organizowanie* zgromadzonych zasobów. W obu przypadkach jednak działanie sprowadza się do odpowiedniego umieszczenia danych w magazynach/repozytoriach przechowawczy oraz przygotowania zestawu metainformacji opisującej zgromadzone dane w taki sposób, aby mogły być sprawnie odnalezione i wykorzystane przez przyszłych użytkowników. Ten element obu pojęć archiwizacji jest też włączeniem informacji o przejętej dokumentacji do systemu informacji będącego już w posiadaniu gromadzącego.

Sporą rozbieżność w definiowaniu „archiwizacji” należy odnotować we wskazaniu jej przedmiotu. Tradycyjne podejście skupia się na *dokumentacji*, natomiast gdy mowa o archiwizacji Internetu, wprowadza się pojęcie źródeł informacji. Czy jest to jednak rzeczywiste rozróżnienie, czy tylko pozorne? W archiwistyce dokument rozumiany jest dwojako: w sposób wąski i szerzej. Węższe rozumienie jest zakorzenione w dyplomatyce, szersze zaś – w naukach informacyjnych³⁵. W pierwszym dokument to „[...] pismo uwierzytelnione, spo-

³⁴ Na temat pojęcia kultury dokumentacyjnej zob. W.K. Roman, *Współczesna kultura*, s. 46–51.

³⁵ Zob. M. Konstankiewicz, A. Niewęglowski, *Narodowy zasób archiwalny i archiwa. Komentarz*, Warszawa 2016, s. 30.

rządzone z zachowaniem przyjętych w danej epoce i miejscu form zewnętrznych i wewnętrznych, stwierdzające lub ustanawiające pewien stan prawny albo też służące do wykonywania uprawnień³⁶. W drugim znaczeniu dokument to „[...] zapis bez względu na jego postać będący świadectwem jakiegoś faktu (faktów) i zjawisk obiektywnej rzeczywistości lub przejawów myśli ludzkiej³⁷. Podobnie jest na gruncie nauk o informacji – tu dokumentem jest „[...] utrwalona informacja wraz z materiałem, na którym została utrwalona, czyli nośnikiem informacji³⁸, dokumentację zaś określa się jako „zbiór dokumentów, rozumianych jako informacje utrwalone na jakimkolwiek nośniku³⁹. Dokument rozumiany szeroko jest więc niczym innym jak utrwaloną informacją, źródłem informacji⁴⁰. Okazuje się więc, że sprzeczność pomiędzy definicją tradycyjną archiwizacji a definiowaniem archiwizacji Internetu nie istnieje w odniesieniu do przedmiotu działań – w obu przypadkach sednem działalności jest gromadzenie informacji. Teza ta wzmocniana jest widocznym i artykułowanym od lat zainteresowaniem archiwistów wspomnianą wcześniej rzeczywistością niedokumentarną.

Jednak sformułowanie celu działań w obu definicjach jest rozbieżne. Podejście archiwistyczne odwołuje się do sfery informacji, traktując pozyskiwaną dokumentację jako obiekt informacyjny⁴¹. W procesie archiwizacji dokumentacja nabiera nowych cech, które z jednej strony wpisują ją w system informacji archiwalnej, z drugiej – tworzą mechanizm informacyjny dla użytkownika zasobu. W pojęciu archiwizacji Internetu jako cel działań uwypuklone są kwestie techniczne związane z zabezpieczeniem danych przed zniknięciem z Internetu. Przekonanie o ulotności treści zawartych w Internecie jest zresztą jedną z głównych przesłanek podejmowania działań archiwizacyjnych⁴². Tak pojmowany cel archiwizacji Internetu oparty jest na założeniu, że podmiot generujący treści i umieszczający je w sieci nie ma mocy ani obowiązku, ani być może też potrzeby,

³⁶ *Polski słownik archiwalny*, s. 27.

³⁷ Tamże, s. 27–28.

³⁸ W.K. Roman, *Współczesna kultura*, s. 68.

³⁹ R. Degen, *Dokumentacja i jej podział* [w:] *Współczesna dokumentacja urzędowa*, red. H. Robótka, Toruń 2011, s. 13.

⁴⁰ Zob. W. K. Roman, *Podstawy zarządzania informacją*, Toruń 2012, s. 18.

⁴¹ O kształtowaniu się definicji dokumentu i dokumentacji zob. W.K. Roman, *Współczesna kultura*, s. 66–80.

⁴² Zob. B. Konopa, *Witryna internetowa*, s. 2. „Archiwizacja webu to metody i działania, których celem jest zgodne ze standardami pozyskiwanie i zachowywanie kopii zamieniających się w czasie wybranych zasobów sieci www”, pisze M. Wilkowski, M. Wilkowski, *Archiwizacja Webu* [w:] *Historia w przestrzeni publicznej*, red. J. Wojdon, Warszawa 2018, s. 239.

by treści te zabezpieczać do wykorzystania w przyszłości. Tradycyjne podejście do archiwizacji jest ściśle związane z przekonaniem o istnieniu i realizowaniu przez wytwórcę dokumentacji obowiązku jej odpowiedniego zabezpieczenia i przekazania do dalszego przechowywania.

Mając powyższe na uwadze, należy teraz przyjrzeć się dotychczasowym pracom w archiwach państwowych oraz toczącym się dyskusjom. Pierwsze doświadczenia archiwów państwowych z archiwizacją Internetu związane są z pilotażem przeprowadzonym przez Narodowe Archiwum Cyfrowe w latach 2009–2011 w formie serwisu Archiwum Internetu⁴³. Kolejne działania zostały wywołane projektem wprowadzenia zmian w ustawie o egzemplarzu obowiązkowym w 2013 r. Zmiany miały doprowadzić m.in. do nowego zdefiniowania egzemplarza obowiązkowego oraz do stworzenia mechanizmu, w ramach którego strona internetowa byłaby traktowana jako egzemplarz obowiązkowy, jej właściciel zaś byłby zobowiązany przekazania go do Biblioteki Narodowej⁴⁴. W październiku 2013 r. NDAP zaproponował powołanie wspólnego zespołu złożonego z przedstawicieli NDAP, archiwów państwowych i Biblioteki Narodowej do spraw analizy kwestii archiwizacji stron WWW z domeny .pl⁴⁵. Archiwa poproszone o przesłanie uwag dotyczących koncepcji archiwizowania stron WWW jako egzemplarzy obowiązkowych zwracały uwagę m.in. na:

- ryzyko gromadzenia danych nieposiadających trwałej wartości (podkreślano, że archiwa gromadzą dokumentację poddaną selekcji, a więc wybraną z całości ze względu na walory informacyjne),
- fakt, że strony internetowe podmiotów pozostających pod nadzorem archiwów państwowych powinny być traktowane jako materiał archiwalny, ponieważ stanowią ich dokumentację,
- konieczność przeprowadzania selekcji zamiast gromadzenia treści z całej domeny .pl,
- aspekty techniczne: technologię stron (np. oparcie ich na bazach danych, co implikuje trudności z prostym harvestowaniem), ich dynamikę (częstotliwość zmian), sposób gromadzenia (harwestowanie),

⁴³ Zob. F. Kłębczyk, M. Jędralska, *Serwis „Archiwum Internetu”*, s. 1–9 (dostęp 8.04.2019); F. Kłębczyk, *Archiwizacja*, s. 105–112.

⁴⁴ Zob. K. Ślaska, A. Wasilewska, *Archiwizacja Internetu – sytuacja w polskim prawie z punktu widzenia bibliotekarzy*, „Biuletyn EBIB” 2012, nr 1 (128), s. 1–8. http://www.ebib.pl/images/stories/numery/128/128_slaska.pdf.

⁴⁵ Pismo NDAP z 18.10.2013 skierowane do J. Milera, Dyrektora Departamentu Dziedzictwa Kulturowego w MKiDN, DKN.400.9.2013.

- potrzebę opracowania standardów archiwizowania stron WWW, w tym okresy i czasy archiwizowania,
- postulat współpracy pomiędzy NAC a BN w procesie archiwizacji,
- konieczność ustalenia, czy strony WWW są publikacjami, czy nimi nie są,
- fakt, że strony WWW zawierają często materiały wtórne, podczas gdy oryginalne materiały wraz z kontekstem są przechowywane jako dokumentacja podmiotu (podawano w wątpliwość, czy strony powinny być rozpatrywane jako materiał archiwalny),
- fakt, że wyjęcie stron z kontekstu organizacyjnego i prawnego twórcy zmniejsza ich wartość informacyjną; jednocześnie jednak argumentowano, że strony WWW tak czy inaczej są powiązane z dokumentacją wytwarzaną przez podmiot,
- występowanie na stronach materiałów objętych ochroną prawnoautorską.

Ponadto jedno z archiwów poinformowało, że w zatwierdzanych wykazach zaleca wprowadzanie klasy poświęconej projektowaniu i eksploataowaniu stron WWW, oznaczonej kategorią A dla jednostek objętych nadzorem i kategorią B25 dla pozostałych jednostek organizacyjnych, zaznaczając, że w ramach tych klas mieści się również „archiwizacja stron”⁴⁶.

W efekcie przeprowadzonych prac w 2014 r. Naczelny Dyrektor Archiwów Państwowych stwierdzał: „Z przeprowadzonej analizy wynika, że strony internetowe nie stanowią *sensu stricto* materiału archiwalnego w rozumieniu ustawy o narodowym zasobie archiwalnym i archiwach i bliżej im do wydawnictw gromadzonych przez Bibliotekę Narodową lub wręcz do muzealnictwa gromadzącego poszczególne artefakty epoki”⁴⁷.

NDAP argumentował, że treści opublikowane w Internecie przez podmioty wytwarzające materiały archiwalne, wchodzące do państwowego zasobu archiwalnego stanowią kopię dokumentacji podmiotu. Takie stanowisko zakłada, że wszystko, co znajduje się na stronach WWW, powstaje w zasadniczych cyklach dokumentowania i jest jedynie kopią dokumentacji podmiotu. Nie bierze więc pod uwagę, że treści zawarte na WWW mogą stanowić autonomiczną wartość informacyjną nieposiadającą odpowiednika w dokumentacji podmiotu. Przy takim podejściu wartość informacyjna (i ewentualnie historyczna) strony WWW jest sprowadzona do wartości dokumentów źródłowych stanowiących część kancelarii podmiotu. Wówczas kwalifikacja archiwalna stronny WWW staje się

⁴⁶ Zob. odpowiedzi archiwów na pismo NDAP, DKN.400.9.2013.

⁴⁷ Pismo NDAP do dyrektora BN z dn. 4.03.2014, DKN.400.9.2013.

beprzedmiotowa, bo jej wartość nie jest samoistna, ale zależna od wartości dokumentów źródłowych. Takie podejście nie wydaje się słuszne. Po pierwsze na stronie WWW pojawiają się nie tylko kopie dokumentacji podmiotu, ale również treści niemające odzwierciedlenia w dokumentacji podmiotu. Po drugie niemożliwa staje się ocena wartości strony WWW samej w sobie jako chociażby narzędzia komunikowania się podmiotu z otoczeniem.

Podsumowując swoje stanowisko, NDAP konkludował, że strony internetowe mogą być uznane za materiały archiwalne, jeśli znaczenie historyczne uzyskają jako prezentacja treści (bo sama treść, zgodnie z tym, co przytoczono powyżej, zostanie przejęta do archiwum wraz z dokumentacją podmiotu). Na tym polu NDAP widział współpracę archiwów i BN, w ramach której archiwiści mieli uczestniczyć w wartościowaniu niektórych stron WWW. Gromadzenie opublikowanych treści (stron WWW) miało stać się zadaniem BN, co doprowadziłoby do sytuacji, w której część informacji zostanie powielona: w archiwum wystąpi w formie dokumentów źródłowych, w zasobach BN zaś jako publikacja w formie WWW. NDAP zastrzegł natomiast, że w przypadku systemów dziedzinowych nie należy próbować gromadzić ich na podobnych zasadach. W tym przypadku bazy danych stanowiące podstawę systemu dziedzinowego powinny być, o ile stanowią materiały archiwalne, przejmowane do archiwów państwowych po zakończeniu eksploatacji systemu⁴⁸.

Argumenty wspierające powyższe stanowisko pojawiły się w artykule B. Konopy *Witryna internetowa – dokumentacja czy publikacja?*⁴⁹. Autor opiera się jednak na odmiennej argumentacji wychodzącej z tezy, że zaliczenie strony WWW do kategorii dokumentacji albo publikacji pozwoli udzielić odpowiedzi na pytanie, który podmiot powinien być odpowiedzialny za archiwizowanie Internetu. Jeśli uzna się, że strony WWW nie są dokumentacją, to za ich gromadzenie odpowiadać powinny biblioteki, gdyby zaś okazało się, że są dokumentacją, wówczas powinny znaleźć się we właściwości archiwów. Opierając się na szerokiej definicji dokumentu, B. Konopa dochodzi do wniosku, że dla charakterystyki strony WWW bardziej odpowiednia jest kategoria publikacji niż dokumentacji. Sięga również do dyskusji na temat rozgraniczenia materiałów archiwalnych i bibliotecznych, wskazując, że jej konkluzją stało się odwołanie do kategorii proveniencji: „Za archiwalne uznawano materiały charakteryzujące się wcześniejszą przynależnością do funkcjonującej registratury, czyli całości dokumentacji jakiegoś urzędu, i udziałem w toku załatwienia sprawy,

⁴⁸ Tamże.

⁴⁹ Zob. B. Konopa, dz. cyt., s. 1–7.

przez co mogły zyskać znaczenie prawne. Natomiast za materiały biblioteczne uznawano te, które posiadają znamię literackości w najszerszym tego słowa znaczeniu⁵⁰. Przywołanie tych argumentów pokazuje, jak bardzo rzeczywistość współczesnej dokumentacji różni się od wcześniejszych warunków i jak bardzo używane wcześniej kategorie mogą okazać się nieprzydatne. Trudno zaprzeczyć proweniencyjnym powiązaniom strony WWW z działalnością podmiotu i z jego dokumentacją. Treść strony WWW kształtowana jest przez aktywność podmiotu i niejednokrotnie odpowiada nawet indywidualnym preferencjom osób pełniących aktualnie funkcje kierownictwa. Jest też związana z dokumentacją powstającą w podmiocie w związku z prowadzonymi sprawami i częściowo stanowi jej powtórzenie, o czym wspomniano we wcześniej przywoływanych dokumentach NDAP i do czego jeszcze wrócę w dalszej części tekstu.

Jednocześnie nie da się zaprzeczyć, że strona WWW nosi cechy publikacji, przez którą „należy rozumieć dzieła zwielokrotnione dowolną techniką w celu rozpowszechnienia, a w szczególności:

- 1) piśmiennicze, jak: książki, broszury, gazety, czasopisma i inne wydawnictwa ciągłe, druki ulotne, afisze;
- 2) graficzne i graficzno-piśmiennicze, jak: mapy, plakaty, plany, wykresy, tabele, rysunki, ilustracje, nuty;
- 3) audiowizualne utrwalające dźwięk, obraz lub obraz i dźwięk, jak: płyty, taśmy, kasety, przeźrocza, mikrofilmy, mikrofiszki;
- 4) zapisane na informatycznych nośnikach danych;
- 5) oprogramowanie komputerowe⁵¹.

Nietrudno też dostrzec, że różne typy obiektów uznane za publikacje również mogą stanowić materiały archiwalne. Nie da się zatem nałożyć w prosty sposób na rzeczywistość siatki definicji dokumentacja vs publikacja i na tej podstawie dokonać rozróżnienia właściwości pomiędzy archiwami i bibliotekami. Hybrydowy charakter strony WWW nie powinien więc stanowić podstawy do rozważenia, czy właściwe do gromadzenia jej treści jest archiwum, czy biblioteka. Oczywiście jest, że w zasobach archiwów znajduje się wiele materiałów archiwalnych mających cechy publikacji, m.in. druki ulotne, afisze czy audiowizualia.

⁵⁰ Tamże, s. 2. Zob. J. Wiśniewski, *Archiwalia w bibliotekach i muzeach*, Poznań 2000; A. Stebelski, *Rękopis archiwalny i biblioteczny*, „Archeion” 1951, t. 19/20, s. 234–235; B. Horodyski, *Z pogranicza bibliotekarstwa i archiwistyki*, „Przegląd Biblioteczny” 1956, r. 24, z. 3.

⁵¹ Ustawa z dnia 7 listopada 1996 r. o obowiązkowych egzemplarzach bibliotecznych, Dz. U., 2018, poz. 545.

B. Konopa analizuje w swoim tekście prawie setkę wykazów akt, poszukując tam klasy przeznaczonej na dokumentację w formie stron WWW. Z pewnym zdumieniem odkrywa klasy związane z procesem powstawania i utrzymywania stron WWW, ale w przebadanej próbce nie odnajduje klas jednoznacznie przeznaczonych dla samych stron internetowych. Wywodzi stąd wniosek, że przebadane podmioty nie traktują stron WWW jako swojej dokumentacji. Trzeba przyznać autorowi, że taki wniosek może być uprawniony przynajmniej w odniesieniu do samej analizy JRWA. Przy czym należy też zauważyć, że nie jest możliwe udowodnienie nieistnienia określonego zjawiska na podstawie badania nawet reprezentatywnej próbki. Praktyczne działania niektórych podmiotów publicznych świadczą o traktowaniu stron WWW jako własnej dokumentacji. Za takie należy uznać działania Ministerstwa Cyfryzacji, o których będzie mowa poniżej.

Dalsze prace nad ustawą o egzemplarzu obowiązkowym zostały zarzucone. Do sprawy wrócono w 2016 r. Została wówczas przeprowadzona ekspertyza zlecona przez Departament Mecenatu Państwa MKiDN, która zawierała propozycje systemowego rozwiązania w zakresie archiwizacji Internetu w Polsce. W ramach ekspertyzy proponowano:

- wprowadzenie rozwiązań legislacyjnych umożliwiających samodzielne gromadzenie przez uprawnione instytucje zawartości stron internetowych z domeny .pl, niezależnie od charakteru tych stron, ich statusu prawnoautorskiego oraz profilu działalności właściciela czy administratora strony internetowej (stosowne uprawnienia i zobowiązania miałyby zostać zawarte w ustawie o egzemplarzu obowiązkowym),
- nadanie uprawnień do gromadzenia treści z Internetu Bibliotece Narodowej i archiwom państwowym (w ekspertyzie mowa o NDAP), przy czym BN ma współpracować z NDAP przy „wyborze i przechowywaniu zasobów, [...] mających wartość dla polskiego zasobu archiwalnego”,
- podjęcie współpracy pomiędzy BN a NDAP polegającej na:
 - udziale przedstawicieli NDAP w zespole pracującym nad przygotowaniem założeń polityki archiwizacji Internetu,
 - włączeniu do listy stron podlegających archiwizacji stron kluczowych dla polskiego zasobu archiwalnego,
 - współdziałaniu w przechowywaniu i udostępnianiu internetowych zasobów archiwalnych oraz wdrażaniu rozwiązań technologicznych dotyczących zarówno samego procesu archiwizacji Internetu, jak i długookresowego przechowywania zasobów internetowych.

W ekspertyzie proponowano, by udostępnianie rezultatów archiwizacji odbywało się w zgodzie z obowiązującym prawem autorskim w ramach dozwolonego

użytku. Sugerowano, że możliwa jest interpretacja rozszerzająca art. 28 ustawy o prawie autorskim i prawach pokrewnych⁵², która pozwoli na udostępnianie zgromadzonych danych na terminalach we wszystkich bibliotekach i archiwach państwowych.

Pierwszym etapem archiwizacji (zaplanowanym na lata 2017–2019) miało być powołanie zespołu do spraw archiwizacji Internetu, który miałby opracować założenia polityki archiwizacji Internetu oraz przygotować listę stron podlegających archiwizacji. Jako pierwsze miałyby być zarchiwizowane strony instytucji publicznych (zakładano sesje archiwizacyjne raz na kwartał). Jednocześnie na tym etapie należało przeprowadzić zmiany legislacyjne. Na kolejnym etapie (lata 2020–2021), po wejściu w życie zmian legislacyjnych, liczba archiwizowanych stron miała wzrosnąć. Również podjęte miały być decyzje o tym, jakie treści gromadzić (np. treści z portali społecznościowych). Miał to być jednak nadal model selektywny, który dopiero na trzecim etapie (po 2022 r.) przeszedłby w model gromadzenia treści z całej domeny .pl. W tym modelu proponowano archiwizację wszystkich treści raz do roku, a treści stron uznanych za szczególnie wartościowe miały być kopiowane z większą częstotliwością. Przewidywano również archiwizację tematyczną – treści istotnych ze względu na aktualne wydarzenia⁵³.

Komentując ekspertyzę, Naczelny Dyrektor Archiwów Państwowych zwrócił uwagę na niedostateczne podkreślenie, że treści tworzące strony internetowe mogą stanowić materiały archiwalne. Zauważał przy tym, że możliwe jest wprowadzenie modelu odmiennego od harvestowania, a polegającego na przechowywaniu stron internetowych tak samo jak pozostałej dokumentacji podmiotu i przekazywaniu jej do archiwów państwowych, tak jak przekazywana powinna być inna dokumentacja elektroniczna⁵⁴.

Latem 2017 r. Ministerstwo Cyfryzacji przystąpiło do integracji serwisów rządowych w ramach jednego portalu gov.pl. Założono wówczas, że dotychczasowe strony WWW zostaną zarchiwizowane⁵⁵. W efekcie prośby skierowanej do NDAP o wytyczne co do sposobu postępowania Archiwum Akt Nowych, we współpracy z NAC i NDAP, przeprowadziło ekspertyzy w Ministerstwie Cyfryzacji (sierpień 2017), Ministerstwie Energii (październik 2017) i Ministerstwie Rozwoju (listopad 2017). W ich wyniku powstała *Analiza możliwości przejmowania serwisów*

⁵² Dz. U., 2018, poz. 1191, 1293, 1669, 2245, 2339.

⁵³ K. Ślaska, *Archiwizacja Internetu. Ekspertyza przygotowana na zlecenie Ministerstwa Kultury i Dziedzictwa Narodowego*, 2016, DKN.040.8.2017.

⁵⁴ Pismo NDAP do dyrektora DMP z 21.03.2017, DKN.040.8.2017.

⁵⁵ Pismo Minister Cyfryzacji z dn. 14.07.2017, DZIP.510.25.2017.

informacyjnych ministerstw. Podkreślono w niej, że przeprowadzone ekspertyzy we wszystkich przypadkach wykazały, że zawartość stron internetowych ministerstw ma wartość materiału archiwalnego. Wskazano również, że strony ministerstw stanowią całości i jako takie powinny być oceniane z punktu widzenia ich wartości informacyjnej. Jako kryteria oceny wartości archiwalnej wskazano:

- „Wartość informacyjną (strona internetowa może być potraktowana jako narzędzie realizacji polityki informacyjnej lub/i strategii promocyjnej podmiotu. Wydaje się to o tyle uzasadnione, że współcześnie Internet jest medium dominującym, a więc kształtowanie wiedzy o działalności podmiotu oraz funkcjonowania jego otoczenia odbywa się głównie z wykorzystaniem strony www. Budowa strony www, zawarte na niej informacje i dokumenty, sposób ich prezentowania oraz czas publikacji stanowią praktyczną realizację zadań informacyjnych podmiotu. Z tego punktu widzenia znaczenie drugorzędne mieć powinna ewentualna powtarzalność dokumentacji i informacji, a pierwszorzędne wartość informacyjna, jaką może nieść strona www jako całość i która jest sposobem udokumentowania narracji podmiotu budowanej w Internecie z wykorzystaniem własnej strony internetowej).
- Posiłkowo wykorzystane może być kryterium unikatowości w trzech aspektach:
 - rozwiązania technicznego (niektóre ze stron internetowych mogą być wykonane w technologii lub w sposób tak dalece różny od standardowych rozwiązań, że ten fakt powinien zostać zachowany),
 - sposobu przedstawienia treści (specyficzne dla Internetu zaprezentowanie informacji, a nie tylko publikowanie kopii dokumentów urzędowych),
 - samej treści (niektóre materiały, dane bądź informacje są udostępniane jedynie na stronach WWW i choć weryfikacja tego stanu rzeczy może być utrudniona, to wydaje się, że takie kryterium również powinno zostać uwzględnione)⁵⁶.

W analizie zwrócono również uwagę na zachowanie zasady proveniencji (poszanowania zespołowości). Strony internetowe jako wytwór dokumentacyjny danego podmiotu powinny zachować łączność z jego pozostałą dokumentacją. W przypadku dokumentacji elektronicznej będzie to realizowane na poziomie

⁵⁶ Analiza możliwości przejmowania serwisów informacyjnych ministerstw, załącznik do pisma NDAP do Dyrektora Departamentu Własności Intelektualnej i Mediów Ministerstwo Kultury i Dziedzictwa Narodowego z dn. 14.05.2018, DKN.040.8.2017, s. 9–10.

informacyjnym poprzez wspólny system zawierający informacje o zasobie archiwów i udostępniający informację w postaci elektronicznej. W analizie nie wspomniano o drugim wymiarze zasady proveniencji, a mianowicie o zachowaniu zgodności układu (porządku) dokumentacji z układem pierwotnym⁵⁷. Wydaje się jednak, że koncepcja przyporządkowania stronie internetowej określonej klasy z JRWA oraz przekazania jej, zaopatrzonej w stosowne metadane, wyczerpuje postulat zachowania oryginalnego porządku. Zwłaszcza że w przypadku dokumentacji w postaci elektronicznej zasada proveniencji nie może być odczytywana jako postulat fizycznego porządkowania dokumentacji, a musi być rozumiana jako porządkowanie informacyjne⁵⁸.

W konkluzji analizy stwierdzono, że:

- podmioty publiczne powinny przekazywać strony WWW po ich zamknięciu lub po wycofania informacji opublikowanej w Internecie do właściwych archiwów państwowych,
- przekazywane strony powinny być opatrzone metadanymi zgodnie z Rozporządzeniem Ministra Spraw Wewnętrznych i Administracji z dnia 30 października 2006 r. w sprawie niezbędnych elementów struktury dokumentów elektronicznych⁵⁹,
- istnieje konieczność opracowania odrębnego systemu teleinformatycznego do przechowywania stron WWW,
- konieczne jest podjęcie wspólnych prac w gronie ekspertów z archiwów państwowych, BN i MKiDN⁶⁰.

Zaproponowane podejście wiąże się z szeregiem działań mających wpływ nie tylko na archiwa, ale przede wszystkim na wytwórców dokumentacji. Uświadomienie im, że w ramach prowadzonej działalności zobowiązani są również do właściwego zabezpieczania tworzonych przez siebie stron WWW, wymaga szerokiej akcji informacyjnej, poprzedzonej przygotowaniem standardów i wskazówek

⁵⁷ Zob. C. Biernat, *Zasada przynależności zespołowej* [w:] *Historia i współczesność*, red. E. Frącki, B. Woszczyński, Warszawa 1987, s. 339–356; zasada proveniencji (zwana zasadą przynależności zespołowej) rozumiana jako „[...] poszanowanie organicznej struktury wytworzonej przez twórcę całości dokumentacji (niepodzielność zespołu archiwalnego) oraz zachowanie (poszanowanie) związków zachodzących między dokumentacją”. Zob. H. Robótka, *Wprowadzenie*, s. 83.

⁵⁸ W ten sposób dylematy archiwisty rozdartego pomiędzy paradygmatem porządkowania fizycznego a pragnieniem tworzenia systemu informacyjnego są rozwiązywane za pomocą zestawu elementów opisu archiwaliów. Zob. C. Biernat, dz. cyt., s. 355.

⁵⁹ Dz. U., 2006, nr 206, poz. 1517.

⁶⁰ Analiza możliwości przejmowania serwisów informacyjnych.

co do sposobu zabezpieczania, przechowywania i przekazywania stron WWW do archiwów państwowych. Wyjaśnienia i ujednoczenia wymagać będzie także praktyka umiejscawiania dokumentacji, którą stanowi strona WWW, w wykazach akt. Z prowadzonych analiz⁶¹ wynika, że praktyka jest w tym zakresie mocno różnicowana, a zasady nie do końca jasne.

Oba omówione powyżej stanowiska NDAP obrazują ewolucję postrzegania kwestii archiwizacji Internetu przez archiwa państwowe. Pierwotnie NDAP stwierdzał, że strony internetowe nie są materiałem archiwalnym *sensu stricto*, choć niektóre za taki materiał mogą być uznane. W zmodyfikowanym podejściu NDAP zwraca uwagę na kryteria selekcji archiwalnej, które mogą wskazywać, że strony internetowe są w przeważającej większości materiałem archiwalnym. W tym podejściu NDAP wskazuje na traktowanie stron WWW jako integralnych części dokumentacji podmiotów i proponuje wprowadzenie odpowiednich klas do JRWA oraz zasad ich uporządkowanego przekazywania do archiwów państwowych. To istotne odmienności od sposobu postrzegania archiwizacji Internetu w świecie. W miejsce harvestowania stawia bowiem sprawdzoną zasadę przejmowania dokumentacji, która zakończyła swój żywot jako aktualna dokumentacja podmiotu. W przypadku stron WWW zaproponowano, by podmiot prowadzący serwis internetowy był zobowiązany do jego zabezpieczenia i przekazania do archiwum państwowego po zakończeniu eksploatacji serwisu lub w chwili wycofywania z niego treści. Takie podejście zdaje się zmierzać do próby uchwycenia stron WWW w pewnych zamkniętych całościach, a nie do próby dokumentowania przyrastania serwisu w miarę pojawiania się nowych treści (rezygnacja z okresowego harvestowania stron). Celem tego podejścia jest więc zapewnienie zabezpieczenia treści serwisu, a nie zobrazowanie zmian, jakim podlegał. Należy przy tym podkreślić, że harvestowanie w okresach cyklicznych pozwala zapewnić tylko wybiórcze dokumentowanie zmian serwisu, co daje jedynie przybliżone pojęcie o zmianach, jakim podlegał. Zaproponowane przez NDAP podejście wydaje się bliższe tradycji archiwistycznej, zgodnie z którą dokumentem jest pewna zamknięta całość znaczeniowa, a nie element o zmiennej treści. Proponowany sposób podziału kompetencji pomiędzy biblioteki i archiwa zbliża stanowisko NDAP do praktyki brytyjskiej, w której gromadzeniem treści zamieszczanych w Internecie przez podmioty rządowe zajmują się National Archives⁶².

Zmodyfikowane podejście NDAP bazuje na założeniu, że powstanie strony WWW wpisuje się w proces archiwotwórczy podmiotu, co oznacza, że powinny

⁶¹ Zob. B. Konopa, dz. cyt., s. 5–6.

⁶² Zob. <http://www.nationalarchives.gov.uk/webarchive/> (dostęp 8.04.2019).

być one traktowane jako element dokumentacji podmiotu rozumianej jako czynność utrwalania informacji⁶³. Stanowisko NDAP skupia się na serwisach WWW tworzonych przez podmioty publiczne, które wytwarzają materiały archiwalne wchodzące do państwowego zasobu archiwalnego. *Analiza...* oraz wcześniejsze prace pomijają więc sferę treści o potencjalnej wartości materiałów archiwalnych, ale zamieszczanych w Internecie przez inne podmioty, w tym osoby fizyczne⁶⁴. Wydaje się więc, że wypracowania wymaga stanowisko wobec stron WWW, które stanowią materiały archiwalne, ale nie są wytwarzane przez podmioty ustalone jako wytwórcy państwowego zasobu archiwalnego. W tym obszarze archiwa powinny podejmować przede wszystkim akcję edukacyjną i uświadamiającą. Poza zakresem zainteresowania archiwów państwowych pozostają te wszystkie przekazy umieszczone w sieci, które nie mają wartości materiałów archiwalnych. Na obecnym etapie rozpoznania sprawy wydaje się, że to znakomita większość zawartości domeny .pl. Zasadne jest, aby był to obszar działań Biblioteki Narodowej.

W ramach analizy zagadnienia archiwizacji stron WWW na gruncie archiwalnym należy – dla jasności wyводу – zastrzec, że gdy mowa o archiwizacji Internetu, rozważa się przekazywanie i przejmowanie do archiwów, a następnie przygotowywanie do udostępnienia treści zamieszczanych w Internecie i uznanych za materiały archiwalne. Podejście archiwistyczne winno się skupiać na gromadzeniu, zabezpieczaniu i udostępnianiu treści, które straciły już bieżącą przydatność dla wytwarzającego je podmiotu, ale zachowują wartość informacyjną w kontekście historycznym. Nie mieści się w tym podejściu natomiast gromadzenie kolejnych wersji stron internetowych wybranych na podstawie kryterium cyklicznego harvestowania. Dokumentowanie zmian w zawartości i wyglądzie strony powinno być zadaniem podmiotu tworzącego treść strony. Archiwum powinno skupić się na przechowywaniu strony WWW jako zamkniętej całości w momencie, gdy zostanie wycofana z eksploatacji albo gdy podjęta zostanie decyzja o usunięciu z niej treści o istotnym znaczeniu. Podejście takie znajduje swoje oparcie w misji archiwów państwowych, którą jest zachowywanie wybranej dokumentacji wybranych podmiotów ze względu na jej wartość historyczną. Nie ma w tym światła znaczenia, czy rozważane obiekty (tu: strony WWW) należy uznać za dokumentację, czy też publikację. Rozróżnienie zaproponowane przez stosowne ustawy (o narodowym zasobie archiwalnym

⁶³ J. Ratajewski, *Wstęp do informacji naukowej*, Katowice 1973, s. 12.

⁶⁴ Zob. J. Łosowski, *Teoria archiwistyki i teoria archiwalna*. [w:] *Toruńskie Konfrontacje Archiwalne*, t. 2, s. 85–86.

i o egzemplarzu obowiązkowym) jest dalece nieprecyzyjne i niewystarczające, o czym świadczą znacznie wcześniejsze dyskusje w odmiennym kontekście. Odwołanie się w nich do kwestii proveniencji może być jednak wykorzystane do wzmocnienia argumentacji za pozostawieniem stron WWW w związku proveniencyjnym z pozostałą dokumentacją wytwórcy. Takie powiązanie zostanie zachowane jedynie wówczas, gdy strony internetowe znajdą się w tym samym systemie informacyjnym co pozostała dokumentacja podmiotu, a więc w systemie informacji archiwalnej.

Zaproponowane rozwiązanie wyznacza linię demarkacyjną pomiędzy kompetencjami archiwów i bibliotek w zakresie zabezpieczania stron WWW. Archiwa powinny zatroszczyć się o treści zamieszczane w Internecie, a wytwarzane przez podmioty pozostające pod ich nadzorem. Podmioty te powinny swoje strony WWW traktować jak dokumentację własną, odpowiednio je przechowywać i po dziesięciu latach od zakończenia eksploatacji lub dokonania istotnego usunięcia treści przekazywać do archiwów państwowych, tak jak pozostałe dokumenty elektroniczne (chyba że w danym przypadku strona zostanie uznana za dokumentację niearchiwalną). Ten postulat dotyczy oczywiście wyłącznie stron podmiotowych, a nie serwisów umożliwiających realizację usług publicznych lub stanowiących serwisy dostępne do systemów dziedziny (np. rejestrów państwowych, systemów do składania wniosków itp.). Te ostatnie powinny być traktowane jako odrębne całości, a postępowanie z nimi również powinno zostać uregulowane odrębnie. Podobnie odmienna jest sytuacja kont/profilu w serwisach społecznościowych, których los także wymaga analizy. W zakresie kompetencji BN powinny znaleźć się pozostałe strony WWW. Niemniej wydaje się ze wszech miar zasadne, aby współpraca pomiędzy archiwami i BN w zakresie praktycznej realizacji całego procesu była kontynuowana.

■ Bibliografia

- Adamczewski, Piotr. *Leksykon informatyki stosowanej*. Materiały Dydaktyczne – Akademia Ekonomiczna w Poznaniu. Poznań: Wydawnictwo Akademii Ekonomicznej, 2000.
- Biernat, Czesław. „Zasada przynależności zespołowej.” W *Historia i współczesność*, zredagowali Edward Frącki i Bolesław Woszczyński, 339–356. Warszawa: Naczelna Dyrekcja Archiwów Państwowych, 1987.
- Degen, Robert. „Dokumentacja i jej podział”. W *Współczesna dokumentacja urzędowa*, zredagowała Halina Robótka, 11–33. Biblioteka Zarządcy Dokumentacji 2. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, 2011.

- Derfert-Wolf, Lidia. „Archiwizacja Internetu – wprowadzenie i przegląd wybranych inicjatyw.” *Biuletyn EBIB*, nr 1 (2012): 1 – 19. http://www.ebib.pl/images/stories/numery/128/128_derfert.pdf.
- Drabik, Lidia, Stanisław Dubisz, Agnieszka Grzegółka-Maciejewska, i Elżbieta Sobol. *Uniwersalny słownik języka polskiego*. T. 1. Warszawa: Wydawnictwo Naukowe PWN, 2003.
- Dubisz, Stanisław. „«Archiwum», «archiwizacja», «archiwizować».” *Poradnik Językowy*, z. 7 (2013): 113–115.
- Gmiterek, Grzegorz. „Długoterminowa archiwizacja zasobów cyfrowych.” *Res Historica* 35 (2013): 213–235.
- Grot, Dariusz. *Ku przeszłości otwartej: dostępność archiwów państwowych w Polsce w latach 1918–2014*. Warszawa: Naczelna Dyrekcja Archiwów Państwowych. Departament Organizacji i Udostępniania. Wydział Popularyzacji i Wydawnictw, 2015.
- Halawa, Mateusz. „Nowe media i archiwizacja życia codziennego.” *Kultura Współczesna*, nr 4 (2011): 27–41.
- Horodyski, Bogdan. „Z pogranicza bibliotekarstwa i archiwistyki.” *Przegląd Biblioteczny* 24, z. 3 (1956): 201–212. <http://bbc.uw.edu.pl/Content/1151>.
- Janowska, Maria A. „Biblioteki akademickie – trendy dotyczące zasobów elektronicznych.” W *Informacja dla nauki a świat zasobów cyfrowych*, zredagowała Halina Ganińska, 166–171. Poznań: Biblioteka Główna Politechniki Poznańskiej, 2008. http://library.put.poznan.pl/konf_idn/art/4_3.pdf.
- Januszko-Szakiel, Aneta. „Długoterminowa archiwizacja zasobów cyfrowych – program dla polskich bibliotek.” *Przegląd Biblioteczny* 79, nr 2 (2011): 211–230.
- Kalota, Tomasz, i Marcin Szala. „Organizacja i logistyka digitalizacji.” W *Biblioteki cyfrowe*, zredagowały Małgorzata Janiak, Monika Krakowska i Maria Próchnicka, 437–446. Nauka – Dydaktyka – Praktyka 135. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 2012.
- Kłębczyk, Filip. „Archiwizacja zasobów Internetu – kierunki i wyzwania.” *Archiwista Polski* 17, nr 3 (2012): 105–112.
- Kłębczyk, Filip, i Monika Jędralska. „Serwis «Archiwum Internetu» na tle ogólnych problemów archiwizacji zasobów sieciowych.” *Biuletyn EBIB*, nr 1 (2012): 1–9. http://www.ebib.pl/images/stories/numery/128/128_jedralska.pdf.
- Konopa, Bartłomiej. „Witryna internetowa – dokumentacja czy publikacja?.” *Biuletyn EBIB*, nr 2 (2017): 1–7. <http://open.ebib.pl/ojs/index.php/ebib/article/download/528/684>.
- Konstankiewicz, Marek, i Artur Niewęglowski. *Narodowy zasób archiwalny i archiwa: komentarz*. Komentarze praktyczne. Warszawa: Wolters Kluwer, 2016.
- Kusnik, Damian. „Długoterminowa archiwizacja danych cyfrowych na mikrofilmach.” *Archiwista Polski* 19, nr 2 (2014): 21–30.
- Łosowski, Janusz. „Teoria archiwistyki i teoria archiwalna.” W *Teoria archiwalna: wczoraj – dziś – jutro*, zredagowali Waldemar Chorążyczewski i Agnieszka Rosa, 69–89. Toruńskie Konfrontacje Archiwalne 2. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, 2011.

- Maciejewska, Wanda, red. *Polski słownik archiwalny*. Warszawa: Naczelna Dyrekcja Archiwów Państwowych, 1974.
- Nowina-Konopka, Maria. „Problematyka zarządzania informacją w procesie archiwizacji zasobów Internetu.” *Media – Kultura – Komunikacja Społeczna* 10, nr 1 (2014): 28–43.
- Perzyna, Paweł. „Problemy terminologiczne wynikające z komputeryzacji archiwów.” W *Komputeryzacja i digitalizacja w archiwach*, zredagowali Rafał Leśkiewicz i Anna Żeglińska, 47–60. Symposia Archivistica 2. Warszawa: Instytut Pamięci Narodowej. Komisja Ścigania Zbrodni przeciwko Narodowi Polskiemu, 2016.
- Ratajewski, Jerzy. *Wstęp do informacji naukowej*. Katowice: Wydawnictwo Uniwersytetu Śląskiego, 1973.
- Robótka, Halina. *Opracowanie i opis archiwaliów: podręcznik akademicki*. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, 2010.
- Robótka, Halina, Bohdan Ryszewski, i Andrzej Tomczak. *Archiwistyka*. Warszawa: Państwowe Wydawnictwo Naukowe, 1989.
- Robótka, Halina. *Wprowadzenie do archiwistyki*. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika, 2002.
- Roman, Wanda K. *Współczesna kultura dokumentacyjna*. Toruń: Wydawnictwo Adam Marszałek, 2013.
- Roman, Wanda K. *Podstawy zarządzania informacją*. Biblioteka Zarządcy Dokumentacji 3. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, 2012.
- Roman, Wanda K. „Terminologia archiwalna – ewolucja czy rewolucja.” W *Teoria archiwalna: wczoraj – dziś – jutro*, zredagowali Waldemar Chorażyczewski i Agnieszka Rosa, 203–217. Toruńskie Konfrontacje Archiwalne 2. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, 2011.
- Ryszewski, Bohdan. *Problemy i metody badawcze archiwistyki*. Rozprawy / Uniwersytet Mikołaja Kopernika. Toruń: Uniwersytet Mikołaja Kopernika, 1985.
- Sobczak, Anna. „Internet jako globalne archiwum społeczne – rozważania na temat roli internetu w dokumentowaniu dziejów ludzkości.” W *Nowa archiwistyka: archiwa i archiwistyka w ponowoczesnym kontekście kulturowym*, zredagowali Waldemar Chorażyczewski, Wojciech Piasek i Agnieszka Rosa, 237–247. Toruńskie Konfrontacje Archiwalne 4. Toruń: Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, 2014.
- Stebelski, Adam. „Rękopis archiwalny i biblioteczny.” *Archeion* 19/20 (1951): 234–235.
- Ślaska, Katarzyna, i Anna Wasilewska. „Archiwizacja Internetu – sytuacja w polskim prawie z punktu widzenia bibliotekarzy.” *Biuletyn EBIB*, nr 1 (2012): 1–8. http://www.ebib.pl/images/stories/numery/128/128_slaska.pdf.
- Wajs, Hubert. „Dokument elektroniczny jako obiekt archiwalny.” *Archiwista Polski* 16, nr 2 (2011): 19–25.
- Wilkowski, Marcin. „Archiwizacja Webu.” W *Historia w przestrzeni publicznej*, zredagowała Joanna Wojdon, 239–244. Warszawa: PWN, 2018.
- Wilkowski, Marcin. „Oddolne archiwizacje Internetu jako działania społeczne”. *Archiwa – Kancelarie – Zbiory* 6(8) (2016): 207–220. <https://doi.org/10.12775/AKZ.2015.007>.
- Wiśniewski, Jarosław. *Archiwalia w bibliotekach i muzeach*. Poznań: Wydawnictwo Naukowe UAM, 2000.