

Data przesłania artykułu: 3 VIII 2020 r.

Data przyjęcia artykułu do druku: 12 X 2020 r.

DOI: <http://dx.doi.org/10.12775/AKZ.2020.004>



BARTŁOMIEJ KONOPA

(Archiwum Państwowe w Bydgoszczy)

STRATEGIA SELEKTYWNA JAKO NARZĘDZIE W ARCHIWIZACJI WEBU. ANALIZA WYBRANYCH PRZYKŁADÓW

Słowa kluczowe

archiwizacja Webu, archiwa Webu, archiwa cyfrowe, strategia selektywna

Keywords

Web archiving, Web archives, digital archives, selective strategy

Streszczenie

Poszczególne archiwa i projekty zabezpieczające zasoby WWW, aby zrealizować postawiony przed nimi cel, określają odpowiadające im rozwiązania i metody działania. Wśród nich znajduje się strategia, która wpływa na to, jak wybierane i gromadzone są witryny internetowe i inne materiały online. W artykule omówione zostało jedno z najczęściej stosowanych rozwiązań – strategia gromadzenia selektywnego. W dalszej



Bartłomiej Konopa, tytuł zawodowy magistra uzyskał na kierunku: archiwistyka i zarządzanie dokumentacją na Uniwersytecie Mikołaja Kopernika w Toruniu. Obecnie pracownik Archiwum Państwowego w Bydgoszczy i uczestnik studiów doktoranckich w zakresie historii na Wydziale Nauk Historycznych UMK w Toruniu. Jego zainteresowania badawcze obejmują przede wszystkim archiwizację Internetu, a także historię Torunia, źródłoznawstwo oraz teorię archiwistyki.

E-mail: bartlomiejkonopa@gmail.com

ORCID ID: 0000-0001-9843-5552

jego części przedstawiono 11 przykładów archiwów i inicjatyw, które w wykorzystują tę metodę działania. Do przygotowania analizy wykorzystano trzy rodzaje źródeł, które dostarczały informacji na temat wykorzystywania strategii selektywnej przez poszczególne projekty: ich witryny internetowe, wytyczne oraz opracowania naukowe na ich temat. Podczas ich badania zwrócono szczególną uwagę na to jakie materiały starają się one archiwizować, w jaki sposób dokonują ich identyfikacji, wyboru i jak je gromadzą, a także jakie stosują kryteria. W trakcie gromadzenia danych dało zauważyć brak dostępu dokładnych informacji na temat funkcjonowania niektórych z wybranych przykładów. Z tego powodu, a także dużej ich różnorodności, niemożliwe było zaprezentowanie ich według jednego, konkretnego schematu. Na tej podstawie, wspartej literaturą przedmiotu, możliwe było jednak ustalenie jakiego rodzaju inicjatywy korzystają z strategii selektywnej i jak ją stosują oraz jakie są tego efekty. Przeprowadzony przegląd inicjatyw pozwolił też scharakteryzować to rozwiązanie oraz wysunąć wnioski dotyczące zalet i wad.

Summary

Selective strategy as a tool in web archiving. Analysis of selected examples

Particular archives and projects securing web resources define corresponding solutions and methods of operation in order to achieve the goal set for them. Among them is a strategy that influences how websites and other online material are selected and collected. The article discusses one of the most commonly used solutions – the selective collection strategy. It follows on from 11 examples of archives and initiatives that use this approach. Three types of sources were used to prepare the analysis, which provided information on the use of selective strategy by individual projects: their websites, guidelines and scientific studies on them. During their research, particular attention was paid to what materials they try to archive, how they identify and select them, and how they collect them, and what criteria they apply. In the course of data collection, it was noticed that there was no access to precise information on the functioning of some of the selected examples. For this reason, as well as their great diversity, it was impossible to present them according to one specific scheme. On this basis, supported by the literature on the subject, it was possible, however, to determine what kind of initiatives use the selective strategy and how they apply it, and what are its effects. The review of initiatives also allowed for the characterization of this solution and for drawing conclusions about its advantages and disadvantages.

Wstęp

Potrzeba archiwizacji zasobów pochodzących z sieci World Wide Web jest coraz szerzej dostrzegana i istnieją liczne inicjatywy podejmujące się tego zadania. Projekty te gromadzą witryny internetowe oraz inne materiały dostępne online w bardzo zróżnicowanym zakresie oraz stosują w tym celu różnorodne podejścia i metody, wśród których można jednak dostrzec pewne podobieństwa.

W literaturze poświęconej archiwizacji Webu często mówi się o strategiach, które odpowiadają za określenie sposobu doboru i gromadzenia zasobów. Rozwiązania te mają umożliwić dotarcie i zachowanie takiego wycinka zasobów Sieci, które odpowiadają profilowi danego projektu. Jedną z najczęściej przywoływanych, a także wykorzystywanych, strategii jest gromadzenie selektywne, które stanowić będzie główny przedmiot niniejszego referatu.

Analizę tego zagadnienia należy zacząć od przedstawienia tego, czym generalnie są strategie stosowane podczas archiwizowania WWW oraz krótkiego scharakteryzowania ich dwóch podstawowych rodzajów. W dalszej części artykułu zostanie przedstawione 11 przykładów archiwów i inicjatyw, które w swojej działalności wykorzystują rozwiązanie selektywne. Podczas ich analizy zwrócono szczególną uwagę na to jakie materiały starają się one archiwizować, w jaki sposób dokonują ich identyfikacji, wyboru i jak je gromadzą, a także jakie stosują kryteria, o ile takie informacje były udostępniane. Na tej podstawie, wspartej literaturą przedmiotu, możliwe było ustalenie jakiego rodzaju inicjatywy korzystają z strategii selektywnej i jak ją wykorzystują oraz jakie są tego efekty. Przeprowadzony przegląd inicjatyw pozwolił też scharakteryzować to rozwiązanie oraz wysunąć wnioski dotyczące zalet i wad.

Literatura i źródła

Kwestia strategii stosowanych w archiwizacji Webu jest często prezentowana w opracowaniach podejmujących tą tematykę, zwłaszcza takich opisujących ten proces. Informacje na ten temat znajdują się chociażby w rozdziale *Web Archives* z książki *Remembering and Forgetting in the Digital Era*¹ czy w opracowaniu przygotowanym przez Janne Nielsen². Studium porównawcze podejść i metod, w tym także kwestii selekcji zasobów, wśród projektów archiwizujących przeprowadził Julien Masanès w 2005 r³. Badacz ten, także w innym opracowaniu, szeroko omówił kwestie selekcji, zwracając uwagę na rolę tego procesu

¹ F. Thouvenin, P. Hettich, H. Burkert, U. Gasser, *Web Archives*, [w:] *Remembering and Forgetting in the Digital Age*, red. F. Thouvenin i in., Springer, Cham 2018, s. 84–101 DOI:10.1007/978-3-319-90230-2_6.

² J. Nielsen, *Using web archives in research: an introduction*, Aarhus 2016. http://www.netlab.dk/wp-content/uploads/2016/10/Nielsen_Using_Web_Archives_in_Research.pdf (dostęp 18.09.2020).

³ J. Masanès, *Web Archiving Methods and Approaches: A Comparative Study*, „Library trends” 54 nr 1 (2005): 72–90, <https://doi.org/10.1353/lib.2006.0005>.

i identyfikując problemy pojawiające się na różnych jej etapach⁴. Spostrzeżenia na temat dobru kryteriów selekcji w przypadku archiwum o charakterze narodowym przedstawili Daniel Gomes, Sérgio Freitas i Mário J. Silva⁵. W przypadku takich inicjatyw zdarzały się także przypadki ich ewaluacji, co miało miejsce chociażby w Chorwacji⁶.

Problematyka związana z archiwizacją WWW jest również podejmowana w polskiej literaturze naukowej. Przywołać można tu numery 128⁷ i 172⁸ Biuletynu EBIB w całości poświęcone tej tematyce, pojawiły się w nich m.in. omówienia najważniejszych archiwów WWW oraz problemy związane z egzemplarzem obowiązkowym. Ciekawym omówieniem oddolnych inicjatyw związanych z zachowywaniem zasobów Webu jest artykuł Marcina Wilkowskiego zamieszczony w 6 tomie „Archiwów – Kancelarii – Zbiorów”, gdzie przedstawił różne aspekty takiej działalności⁹. Kwestię strategii selektywnej autor podejmował już wcześniej, m.in. w kontekście jej przydatności do zachowania różnorodności Sieci¹⁰.

Do przygotowania analizy prezentowanej w niniejszym artykule wykorzystano trzy rodzaje źródeł, które dostarczały informacji na temat wykorzystywania strategii selektywnej przez poszczególne projekty. W pierwszej kolejności były to normatywy (o ile były one dostępne) określane jako wskazówki, polityki lub wytyczne, które opisywały m.in. stosowane kryteria i rozwiązania. Zaliczyć można do nich chociażby wskazówki opracowane przez instytucje współpracujące w ramach australijskiego archiwum WWW PANDORA¹¹ oraz Bibliotekę

⁴ Tenże, *Selection for a Web archives*, [w:] *Web Archiving*, red. J. Masanès, Berlin – Nowy Jork 2006, s. 71–91.

⁵ D. Gomes, S. Freitas, M.J. Silva, *Design and Selection Criteria for a National Web Archive*, [w:] *Research and Advanced Technology for Digital Libraries*, red. J. Gonzalo i in., Berlin–Heidelberg 2006, s. 196–207. DOI:10.1007/11863878_17.

⁶ M. Willer, T. Buzina, K. Holub, J. Zajec, M. Milinović, N. Topolščak, *Selective archiving of web resources: a study of processing costs*, „Program” 2008, t. 42, nr 4, s. 341–364. DOI:10.1108/00330330810912043.

⁷ Biuletyn EBIB, *Numer 128 – Archiwizacja Internetu*, http://www.ebib.pl/?page_id=380 (dostęp 18.09.2020).

⁸ Tamże, *Nr 172 (2017): Archiwizacja internetu*, <http://ebibojs.pl/index.php/ebib/issue/view/17> (dostęp 18.09.2020).

⁹ M. Wilkowski, *Oddolne archiwizacje Internetu jako działania społeczne*, „Archiwa – Kancelarie – Zbiory” 2015, nr 6 (8) s. 207–220. DOI: 10.12775/AKZ.2015.007.

¹⁰ B. Konopa, *Zachowanie różnorodności Sieci w archiwach Webu*, „Toruńskie Studia Bibliograficzne” 2019, t. 12, nr 1, s. 45–65. DOI: 10.12775/TSB.2019.003.

¹¹ Wszystkie analizowane wskazówki selekcji dostępne są na stronie internetowej: Pandora Archives, *Selection Guidelines*, <http://pandora.nla.gov.au/guidelines.html> (dostęp 29.07.2020).

Kongresu Stanów Zjednoczonych¹², w których wskazano jakie zasoby powinny być gromadzone. Podobne informacje pozyskiwano także z witryn internetowych poszczególnych inicjatyw, gdzie także umieszczane były informacje na temat ich działalności. Uzupełnienie tego stanowiły opracowania naukowe poświęcone wybranym archiwom. Takowe studia przypadku zostały przygotowane m.in. dla archiwów w Danii¹³ czy Chorwacji¹⁴, a także projektu End of Term Web Archive¹⁵. Analizy te często przybliżają historię danych inicjatyw, tło ich funkcjonowania oraz stosowane rozwiązania, a także problemy na jakie napotkały. W trakcie gromadzenia danych dało zauważyć brak dostępu dokładnych informacji na temat funkcjonowania niektórych z wybranych przykładów. Z tego powodu, a także dużej ich różnorodności, niemożliwe było zaprezentowanie ich według jednego, konkretnego schematu.

Strategie w archiwizacji Webu

Rozważania nad strategią selektywną wykorzystywaną w archiwizacji Webu należy rozpocząć od odpowiedzi na pytanie czym są strategie w kontekście tego zagadnienia. Pod tym terminem kryje się sposób w jaki archiwa i projekty archiwizujące WWW podchodzą do selekcji i gromadzenia zasobów, tak aby zrealizować postanowiony cel. Poszczególne inicjatywy, niezależnie od tego czy zainteresowane są zabezpieczeniem narodowego wycinka Sieci lub dokumentowaniem jakiegoś zagadnienia, powinny określić jak będą przebiegać procesy związane z wyborem materiałów oraz jakimi kryteriami będzie się posługiwać w ich trakcie¹⁶.

¹² Library of Congress, Washington D.C., *Library of Congress Collections Policy Statements Supplementary Guidelines*. *Web Archiving*, <https://www.loc.gov/acq/devpol/webarchive.pdf> (dostęp 31.07.2020).

¹³ S. Schostag, E. Fønss-Jørgensen, *Webarchiving: Legal Deposit of Internet in Denmark. A Curatorial Perspective*, „Microform & Digitization Review” 2012, t. 41, nr 3–4, s. 110–120. DOI: 10.1515/mir-2012-0018.

¹⁴ K. Holub, I. Rudomino, *A decade of web archiving in the National and University Library in Zagreb*, materiały z konferencji IFLA WLIC 2015, Kapsztad, 11–20 sierpnia 2015, s. 1–12. <http://library.ifla.org/1092/1/090-holub-en.pdf> (dostęp 30.07.2020).

¹⁵ T. Seneca, A. Grotke, C. N. Hartman, K. Carpenter, *It takes a village to save the web: The End of Term Web Archive*, „Documents to the People” 2012, t. 40, nr 1, s. 16–23. <https://www.journals.ala.org/index.php/dttp/issue/viewFile/590/346> (dostęp 31.07.2020).

¹⁶ D. Gomes, S. Freitas, M.J. Silva, dz. cyt., s. 197–198; J. Masanès, *Web Archiving Methods*, s. 75–76.

Wybór strategii, jak udało się zauważyć, ma też wpływ na późniejszą strukturę zarchiwizowanych zasobów.

Liczba mnoga i konieczność wyboru przywołana w powyższym akapicie jasno wskazuje, że istnieje więcej niż jedna strategia stosowana w archiwizacji Webu. W opracowaniu *Using Web Archives in Research – An introduction* Janne Nielsen przywołuje następujące ich rodzaje: *broad crawl* (alternatywnie *bulk/snapshot harvesting*), *national/regional domain crawl*, *selective harvesting*, *event harvesting* oraz *thematic harvesting*¹⁷. Wiele inicjatyw korzysta z tych lub zbliżonych określeń opisując rozwiązania stosowane przez nie w archiwizacji WWW, jednak tak rozbudowaną systematykę można znacząco uprościć. Za najbardziej przejrzysty i sensowny należałoby uznać, wielokrotnie przywoływany w literaturze, podział na strategie gromadzenia masowego (ang. *bulk harvesting*) oraz selektywnego (ang. *selective harvesting*).

Pierwsza z wymienionych strategii, a więc gromadzenie masowe, opiera się najczęściej na archiwizowaniu konkretnej domeny lub grupy domen, jak ma to miejsce w przypadku części narodowych projektów archiwizacji Webu. Dzięki wykorzystaniu robotów indeksujących możliwa jest daleko posunięta automatyzacja całego procesu, a udział człowieka w nim sprowadza się do kontrolowania właściwego działania programu. W teorii gromadzenie masowe oferuje perspektywę budowania dużych zbiorów przy niskich nakładach pracy. W praktyce ujawniają się jednak jego pewne mankamenty, przez które zebrane w ten sposób materiały mogą sprawiać wrażenie niekompletnych. Jednym z problemów jest fakt, iż pobieranie znacznej ilości danych musi odbywać się w relatywnie dużych interwałach czasowych, co w przypadku dynamiki, którą cechuje się Internet, może prowadzić do utracenia części informacji. Dostęp do części treści mogą ograniczać robotom dodatkowe przeszkody, np. w postaci odpowiedniej konfiguracji pliku robots.txt lub wymogu logowania na witrynie. Duża liczba pozyskiwanych zrzutów utrudnia także sprawdzanie ich jakości i kompletności. Najbardziej znanym przykładem wykorzystującym gromadzenie masowe jest fundacja Internet Archive, oprócz niej wymienić można również UK Web Archive oraz duński Netarkivet¹⁸.

Druga ze strategii to podejście selektywne, które stanowi przedmiot zainteresowania niniejszego artykułu. Idea tego rozwiązania sprowadza się do manualnego typowania poszczególnych witryn i innych obiektów online, które mają zostać poddane archiwizacji. Wyboru dokonuje się na podstawie wcześniej

¹⁷ J. Nielsen, dz. cyt., s. 21–22.

¹⁸ F. Thouvenin, P. Hettich, H. Burkert, U. Gasser, dz. cyt., s. 94–95.

obronych kryteriów, które są skonstruowane w różnorodny sposób i uwzględniają różne czynniki, tak aby możliwe było zrealizowanie celów postawionych przed danym projektem. Zaletą tego rozwiązania jest to, że pozwala ona na większą precyzję w doborze obiektów, a także na dokładniejszą kontrolę jakości wykonywanych kopii. Ponadto ułatwia to opatrywanie ich metadanymi opisowymi wraz z znacznym pogłębieniem tego rodzaju informacji, a także ich katalogowanie lub grupowanie. Selektywna archiwizacja Webu jest jednak zadaniem wymagającym zdecydowanie większego zaangażowania pracowników. Do ich zadań należy analiza i wybór zasobów, które mają zostać zachowane, porównanie wykonanych zrzutów z oryginałami i rozważania czy i za jaki czas będzie potrzebna kolejna archiwizacja. W ramach tej strategii wymienić można jej bardziej szczegółowe rodzaje: ogólną oraz wymieniane wcześniej dokumentowanie wydarzeń (ang. *event harvesting*) oraz archiwizację tematyczną (ang. *thematic harvesting*)¹⁹.

Omawiana w niniejszym artykule strategia selektywna, zdefiniowana

Przed przejściem do prezentacji konkretnych przykładów należy jeszcze odnotować kilka kwestii. Wiele projektów, które posiada taką możliwość (dotyczy to prawie wyłącznie archiwów o charakterze narodowym), łączy obie wymienione strategie, natomiast pozostała część inicjatyw opiera swoje działanie wyłącznie na archiwizacji selektywnej. Ciekawe jest spostrzeżenie Juliana Masanèsa, który stwierdził, że oba wskazane rozwiązania są selektywne, ponieważ na jakimś etapie działań staje się przed wyborem co archiwizować, a co nie²⁰. Badacz używa względem nich także innych określeń: automatyczna dla masowej oraz manualna dla selektywnej, które oddają istotne cechy tych dwóch podejść do archiwizacji. Dodatkowo wprowadził on podział na archiwizację ekstensywną i intensywną, które odnoszą się do kompletności i jakości gromadzenia zasobów. Pierwsza odnosi się do preferowania archiwizacji o jak najszerszym zakresie kosztem jej głębi, natomiast druga stanowi jej przeciwieństwo. Taki intensywny charakter ma bardzo często strategia selektywna²¹.

Przykłady stosowania strategii selektywnej

Omawianie przypadków stosowania strategii selektywnej w archiwizacji Webu należy rozpocząć od Preserving and Accessing Networked Documentary Re-

¹⁹ Tamże, s. 95–96.

²⁰ J. Masanès, *Selection for a Web archives*, s. 76.

²¹ Tenże, *Web Archiving: Issues and Methods*, [w:] *Web Archiving*, s. 38–40, 44–45.

sources of Australia (PANDORA), czyli archiwum australijskiego Webu. Jest to jedna z pierwszych tego rodzaju inicjatyw na świecie, bo uruchomiona w 1996 r. przez Bibliotekę Narodową Australii, i pierwsza, która wykorzystwała w swoim działaniu strategię selektywną²². Obecnie archiwum PANDORA stanowi część składową większego Australian Web Archive, w skład którego wchodzi także Australian Government Web Archive (działające od 2011 r.) i coroczne zbiory całej domeny krajowej Australii prowadzone od 2005 r.²³. Prezentowane przedsięwzięciu opiera się na współpracy i bierze w nim udział jeszcze 9 innych instytucji: 5 bibliotek stanowych oraz Northern Territory Library and Information Service, National Film and Sound Archive, Australian War Memorial i Australian Institute of Aboriginal and Torres Strait Islander Studies. Dzięki takiej organizacji i podziałowi kompetencji możliwe jest realizowanie głównego celu stawianego przed PANDORA czyli zabezpieczanie witryn internetowych oraz innych publikacji on-line dotyczących Australii i Australijczyków²⁴.

Zastosowany model kooperacji wprowadza podział obszarów działania uczestników projektu według ich właściwości terytorialnej lub specjalizacji – za materiały o charakterze ogólnokrajowym odpowiedzialna jest Biblioteka Narodowa Australii, za regionalne biblioteki stanowe, za wojskowe Australian War Memorial itd. Poszczególne instytucje dokonują selekcji na podstawie opracowanych przez siebie wskazówek, które odnaleźć można na witrynie internetowej archiwum australijskiego Webu. Pomimo faktu, że dokumenty te posiadają odrębny charakter, wskazać można na wiele wspólnych rozwiązań. Podstawowe kryterium w nich zawarte dotyczy treści materiałów typowanych do archiwizacji. Zasoby te powinny informować o Australii lub którymś z jej regionów albo być australijskiego autorstwa i odnosić się do zagadnień społecznym, kulturowym, politycznym etc. ważnych dla mieszkańców Australii lub stanowić wkład w wiedzę międzynarodową. Fakt, czy dana publikacja znajduje się na serwerach wewnątrz kraju lub poza nim oraz samo pochodzenie jej autora nie mają znaczenia w tym procesie, jednak wytyczne zwracają uwagę na wiarygodność oraz autorytet autora. Twórcy wskazówek rozumieją przez to, że powinno być możliwe dokładne ustalenie kto jest odpowiedzialny za powstanie danego

²² Pandora Australia's Web Archive, *Pandora Overview*, <http://pandora.nla.gov.au/overview.html> (dostęp 29.07.2020).

²³ National Library of Australia, *Archived websites*, <https://www.nla.gov.au/what-we-collect/archived-websites> (dostęp 29.07.2020).

²⁴ Pandora Australia's Web Archive, *Pandora Partners*, <http://pandora.nla.gov.au/partners.html> (dostęp 29.07.2020).

źródła i czy jest on w tym zakresie kompetentny. Podkreślają oni również ich znaczenie oraz przyszłą, najlepiej długoterminową, użyteczność dla badań naukowych, powinny one zatem zawierać w sobie „istotną kompilację informacji”. W ramach PANDORA mogą zostać zgromadzone także pozostałości po „dawnym” Internecie oraz przykłady zastosowania innowacyjnych rozwiązań technologicznych. Do wytycznych wprowadzane są również pewne wykluczenia, m.in. rezygnuje się z archiwizacji publikacji, które posiadają formę drukowaną (jest ona preferowana), o ile wersja online nie dostarcza szerszego kontekstu.

Analizowane wytyczne wydają się traktować zasób australijskiego Webu jako publikacje dotyczące tego kraju oraz jego obywateli. Poszczególne obiekty są rozdzielane pomiędzy 18 tematów, które dodatkowo mogą zawierać w sobie podkategorie. Zbiory PANDORA mogą być też przypisane do jednej lub więcej kolekcji tematycznych, tworzonych według zagadnień lub wydarzeń – znajdują się tam przykłady zastosowania podejścia *event harvesting*²⁵. Wyselekcjonowane witryny nie muszą zostać zarchiwizowane w całości, jeżeli jakaś jej część nie odpowiada kryteriom. Dopuszczone jest również dzielenie tych bardziej rozbudowanych i tworzenie z nich osobnych tytułów, o ile wpłynie to na ich wartość informacyjną.

Innym projektem archiwizacji Webu, wykorzystującym strategię selektywną jest funkcjonujący od 2005 r. duński Netarchive.dk prowadzony przez Bibliotekę Królewską w Kopenhadze i Państwową i Uniwersytecką Bibliotekę w Aarhus. Projekt stosuje 3 różne warianty omawianego podejścia, które w pewnym stopniu odbiegają od tych stosowanych w PANDORA i są bardziej sformalizowane. Zmiany w prawie dotyczącym egzemplarza obowiązkowego wprowadzone obowiązujące od początku 2005 r. umożliwiły tej inicjatywie na wprowadzenie dwóch rozwiązań określonych jako *selective crawl* oraz *event crawl*. Pierwsza z nich ma stanowić uzupełnienie gromadzenia masowego (określanego tutaj jako *broad crawl*) i pozwala na archiwizowanie 80–100 witryn, które ze względu na swój bardziej dynamiczny charakter wymagają częściej wykonywanych zrzutów. Większość gromadzonych w ten sposób materiałów stanowią witryny z wiadomościami (około 60%), następne w kolejności są witryny dotyczących społeczeństwa, sektora komercyjnego i państwowego (30%), najmniejsza część to natomiast witryny eksperymentalne, wykorzystujące nowe rozwiązania technologiczne lub sztuka internetowa (10%)²⁶.

²⁵ Wykaz głównych kategorii dostępny jest na stronie głównej projektu: Pandora Australia's Web Archive, zob. <http://pandora.nla.gov.au/> (dostęp: 29.07.2020).

²⁶ S. Schostag, E. Fønss-Jørgensen, dz. cyt., s. 110–112.

Lista tytułów archiwizowanych w ten sposób nie jest stała i ulega ciągłym aktualizacjom. Każdy wytypowany obiekt musi przejść przez poszczególne etapy procesu selekcji, które powtarza się co sześć miesięcy. Na *selective harvesting* składa się 7 etapów: identyfikacja witryny, wstępna ocena (podjęcie decyzji o archiwizacji), analiza (określenie parametrów archiwizacji), wprowadzenie danych do systemu, ocena jakości pierwszej wykonanej kopii, monitoring, ewaluacja (podjęcie decyzji o dalszej archiwizacji). Nowe strony mogą być zgłaszane przez Radę Doradczą oraz przez osoby z zewnątrz przez specjalny formularz²⁷. Zasoby wytypowane do archiwizacji mogą być gromadzone w różnej, indywidualnie ustalonej częstotliwości od 6 razy w ciągu jednego dnia do 1 razu na kwartał, jednak konkretne grupy witryn mają określoną częstotliwość, np. witryny partii politycznych archiwizowane są raz w miesiącu²⁸.

Drugim rozwiązaniem stosowanym w archiwum duńskiego Webu są *event crawls*, w ramach którego powstają kolekcje poświęcone wydarzeniom istotnym dla Danii i jej obywateli. Zbiory mogą dotyczyć zdarzeń przewidywalnych, takich jak wybory, lub przypadkowych, wśród których wymienić można zarówno katastrofy naturalne, jak i wydarzenia o charakterze politycznym lub społecznym. Według założeń takich kolekcji powstawać około 2-3 rocznie. Witryny do archiwizacji wybierane są przez pracowników Netarkivet i gromadzone aż do końca wydarzenia²⁹. Jako trzeci wariant strategii selektywnej wskazać można *special crawls*, które obejmują dokumentowanie małych wydarzeń, archiwizowanie ważnych witryn, które mają zostać zamknięte, gromadzenie materiałów na życzenie badaczy, a także testowanie nowych rozwiązań³⁰. Ponadto można stwierdzić, że w sposób selektywny uzupełniana jest lista stron do archiwizacji masowej, ponieważ zasoby spoza duńskiej domeny krajowej są m.in. identyfikowane przez pracowników z pomocą wyszukiwarki Google lub zgłaszane za pomocą formularza na witrynie projektu³¹.

Podobne rozwiązania jak w PANDORA są stosuje Hrvatski arhiv weba (HAW, Chorwackie Archiwum Webu), projekt uruchomiony przez Narodowa i Uniwersytecką Bibliotekę w Zagrzebiu w 2004 r. Ze względu na możliwości techniczne

²⁷ Tamże, s. 112–114.

²⁸ Netarkivet, *Selective høstninger*, http://netarkivet.dk/om-netarkivet/selective-hostninger_2016/ (dostęp 29.07.2020).

²⁹ S. Schostag, E. Fønss-Jørgensen, dz. cyt., s. 114–115.

³⁰ Netarkivet, *Specialhøstninger*, <http://netarkivet.dk/om-netarkivet/specialhostninger/> (dostęp 29.07.2020).

³¹ S. Schostag, E. Fønss-Jørgensen, dz. cyt., s. 111.

w pierwszym okresie swojej działalności HAW wykorzystywał wyłącznie strategię selektywną, a masowe gromadzenie chorwackiej domeny krajowej rozpoczęło się dopiero w 2011 r. W ramach omawianego w artykule podejścia archiwizowane zasoby przypisywane są do jednej z 30 ogólnych grup tematycznych, które mogą dzielić się następnie na mniejsze podkategorie³². Proces wyboru materiałów do zachowania rozpoczyna się od ich identyfikacji, są one wyszukiwane przez bibliotekarzy, zgłaszane z ISSN lub z zewnątrz, następnie analizowane pod kątem spełniania kryteriów. Jeśli zostanie podjęta pozytywna decyzja o archiwizacji sporządzany jest opis danej witryny w standardzie MARC21 oraz ustalane są parametry archiwizacji. Kryteria wykorzystywane w trakcie selekcji podzielone są na dwie grupy – ogólne, obejmujące także inne materiały gromadzone przez Bibliotekę, zarówno drukowane, jak i cyfrowe, oraz szczegółowe, przeznaczone wyłącznie dla zasobów WWW. Pierwsza z nich dotyczy ogólnych powiązań z narodowością chorwacką, a więc chorwackiego autorstwa, opublikowania na terenie Chorwacji bądź w języku chorwackim³³. Druga grupa związana jest już z właściwą oceną zasobów, w trakcie której pod uwagę bierze się ich zawartość wraz z jej strukturą, format danych lub autora³⁴. W ramach HAW od 2011 r. prowadzony jest również *event harvesting*, którego celem jest udokumentowanie ważnych wydarzeń dotyczących Chorwacji. Jego efektem jest powstanie kilkunastu kolekcji, które przede wszystkim dotyczą wyborów oraz referendum, ale wśród nich znajdują się też zbiory związane z piłkarskimi Mistrzostwami Świata w 2018 czy trzęsieniem ziemi, które nawiedziło Zagrzeb w marcu 2020 r.³⁵

Mniej rozbudowana strategia selektywna funkcjonuje w irlandzkim archiwum Webu działającym od 2011 r. w strukturach tamtejszej Biblioteki Narodowej. Historia jego działalności może zostać podzielona na dwa etapy. Pierwszy, obejmujący lata 2011-2018, związany jest z współpracą z nieistniejącą już Internet Memory Foundation. W tym okresie powstało kilkanaście kolekcji tematycznych, głównie dokumentujących ważne wydarzenia, podzielonych w następujące grupy: referenda, wybory, społeczeństwo oraz kultura. Kolekcje te

³² Wykaz głównych grup tematycznych jest dostępny na stronie: HAW Croatian Web Archive, *Browse by subject*, <https://haw.nsk.hr/en/browse-by-subject/> (dostęp 30.07.2020).

³³ K. Holub, I. Rudomino, dz. cyt., s. 3–7.

³⁴ Wykaz kryteriów selekcji stosowanych w HAW można znaleźć na stronie: HAW Croatian Web Archive, *Selection criteria*, <https://haw.nsk.hr/en/selection-criteria/> (dostęp 30.07.2020).

³⁵ HAW Croatian Web Archive, *Thematic collections*, <https://haw.nsk.hr/en/thematic-collections/> (dostęp 30.07.2020).

są dobrze opisane wraz z ich tłem oraz zakresem gromadzenia. Najciekawszym przykładem jest projekt związany z obchodami Powstania Wielkanocnego – *Remembering 1916, Recording 2016*, w ramach którego zaproszono Irlandczyków do typowania witryn do archiwizacji³⁶. Drugi etap działalności irlandzkiego archiwum WWW rozpoczął się z przejściem we wrześniu 2018 r. do współpracy z Archive-It. Nowa narzędzie umożliwiło tworzenie liczniejszych kolekcji tematycznych, w tym takich o bardziej ogólnym charakterze, poświęconych np. historii Irlandii, rolnictwu lub turystyce. Ich mankamentem jest brak szerszego opisu, który objaśniałby kontekst ich powstawania³⁷.

Wśród projektów archiwizacji Webu posługujących się strategią selektywną znajdują się również inicjatywy, które skupiają się na witrynach należących do różnych instytucji państwowych. Takim przykładem może być UK Government Web Archive (UKGWA) uruchomione w 2003 r. przez Brytyjskie Archiwa Narodowe. Działalność UKGWA związana jest z zabezpieczaniem państwowego zasobu archiwalnego i obejmuje nie tylko witryny internetowe, ale także blogi i media społecznościowe (Twitter, Youtube oraz Flickr). Zgodnie z założeniami archiwizowane powinny być wszystkie witryny należące do centralnych instytucji państwowych Zjednoczonego Królestwa, dlatego też na gromadzony zasób składają się głównie zasoby wytworzone przez ministerstwa oraz pozostałe organy rządowe, agencje, komisje królewskie oraz Narodową Służbę Zdrowia na poziomie krajowym. Ze względu na częste zmiany witryn i powstawanie nowych ich właściciele powinni zgłaszać je do UKGWA, ale mogą być one także identyfikowane przez pracowników projektu. Wybrane witryny są archiwizowane przynajmniej raz i dodatkowo w przypadku ważnych wydarzeń, np. przed i po wyborach. W przypadku witryn należących do tymczasowych organów, np. komisji śledczych, jeżeli większość ich materiałów była dostępna online i została zarchiwizowana, to nie muszą one być przekazane w inny sposób. UKGWA prowadziło też pilotażowe projekty tworzenia kolekcji tematycznych oraz gromadzenia zasobów WWW wytwarzanych przez władze lokalne³⁸.

³⁶ Wykaz kolekcji oraz ich opisy są dostępne na stronie: National Library of Ireland, Dublin, *Selective Web Archive Collections*, <https://www.nli.ie/en/udlist/web-archive-collections.aspx> (dostęp 30.07.2020).

³⁷ Profil Biblioteki Narodowej Irlandii w „Archive-It” wraz z wykazem kolekcji: Archive-It, *National Library of Ireland*, <https://archive-it.org/home/nli> (dostęp 30.07.2020).

³⁸ National Archives, London, *Operational Selection Policy OSP27: UK Central Government Web Estate*, 2014, s. 4–7, <https://www.nationalarchives.gov.uk/documents/information-management/osp27.pdf> (dostęp 30.07.2020).

Innym przykładem związanym z archiwizacją zasobów WWW powiązanych z władzami państwowymi jest End of Term Archive. Specyfiką tego projektu jest zabezpieczanie witryn internetowych należących do różnorodnych instytucji państwowych Stanów Zjednoczonych na poziomie federalnym w celu udokumentowania zmian jakie mogłyby zajść w momencie zmiany na stanowisku prezydenta, a także w przypadku jego reelekcji. Inicjatywę uruchomiono w 2008, a w jej ramach współpracują Internet Archive, Biblioteka Kongresu, Uniwersytet Północnego Teksasu, Cyfrowa Biblioteka Kalifornii oraz U.S. Government Publishing Office, a od 2016 r. także Uniwersytety Stanforda i Georga Washingtona. Wykaz adresów URL do pierwszej archiwizacji powstała w ramach współpracy członków projektów na podstawie posiadanych przez nich wcześniejszych list oraz była uzupełniana o nowe portale. Zbudowane w ten sposób zestawienie zostało poddane ewaluacji, usunięto z niego witryny nieaktywne lub nienależące do organów federalnych, a następnie pozostałe poddano ocenie. O tym czy dana witryna znajduje się w zakresie gromadzenia projektu zależało liczby składowych list, na których się ona znajdowała, oraz decyzji kuratorów. Do współpracy w tym przedsięwzięciu zaproszono m.in. badaczy i akademików związanych z politologią i socjologią, specjalistów od informacji rządowych oraz bibliotekarzy. Późniejsze działania w ramach End of Term Archive wyglądały podobnie, jednak można zauważyć większe zainteresowanie projektem w jego ostatniej edycji z lat 2016–2017³⁹.

Zbliżony charakter posiada inna amerykańska inicjatywa – CyberCemetery. Jest to projekt rozpoczęty przez Bibliotekę Uniwersytetu Północnego Teksasu w 1999 r., wspierany przez U.S. Government Publishing Office oraz National Archives and Records Administration. Przedsięwzięcie to ma jasno określone ramy działania – gromadzi witryny i inne zasoby WWW wytworzone przez federalne agencje i komisje. Aby dana witryna została zarchiwizowana musi być oficjalną stroną rządową, a agencja lub komisja, do której ona należy musi kończyć swoją działalność (został wydany finalny raport lub została wygaszona). Przyczyną archiwizacji może być również zagrożenie strony usunięciem z innych przyczyn. Zasoby interesujące CyberCemetery są identyfikowane i monitorowane przez odpowiedzialne za to osoby, jednakże mogą też być zgłaszane z zewnątrz⁴⁰.

³⁹ T. Seneca, A. Grotke, C. N. Hartman, K. Carpenter, dz. cyt., s. 17–19; *Project background*, „End of Term Archive”, <http://eotarchive.cdlib.org/background.html> (dostęp: 31.07.2020).

⁴⁰ S. Hoffman, *Preserving access to government websites: Development and practice in the Cybercemetery*, [w:] *World Library and Information Congress: Seventy-Fourth IFLA*

Poza projektami o charakterze narodowym i rządowym funkcjonuje wiele innych inicjatyw. Warta omówienia jest działalność w tym zakresie Biblioteki Kongresu Stanów Zjednoczonych, prowadzona od 2000 r., początkowo jako projekt Minerva, obecnie jako część zbiorów cyfrowych tej instytucji. Strony archiwizowane przez tą placówkę nie są gromadzone oddzielnie, a głównie jako część kolekcji, która może dotyczyć konkretnego wydarzenia, przedmiotu lub tematu. Propozycje nowych kolekcji są wysuwane przez ekspertów dziedzinowych i powinny zawierać tytuł proponowanego zbioru, jego ogólne tło oraz uzasadnienie. Dodatkowo może być ono uzupełnione o informacje jakiego rodzaju strony mają być gromadzone, przybliżoną ilość adresów URL, przewidywaną częstotliwość ich archiwizacji oraz określenie czy kolekcja będzie stała lub ciągła. Sugerowana kolekcja powinna być użyteczna dla Kongresu Stanów Zjednoczonych, badaczy oraz szerszej publiczności, jednak może ona dotyczyć innych rejonów świata, o ile zostanie to właściwie uzasadnione. Zasoby do archiwizacji powinny też spełniać któreś z następujących kryteriów: spełnianie obecnych lub przyszłych potrzeb informacyjnych Kongresu i badaczy, dostarczanie unikatowych informacji, zawieranie treści naukowych, ryzyko utraty oraz aktualność informacji. Po otrzymaniu zgody Web Archiving Collection Development Group oraz House Office of the General Counsel specjalista odpowiedzialny za daną kolekcję dokonuje wyboru witryn i parametrów ich archiwizacji (głębokość oraz częstotliwość), powinien też regularnie je sprawdzać czy znajdują w zakresie gromadzenia⁴¹. Biblioteka Kongresu udostępnia opis każdej kolekcji, który dostarcza podstawowych informacji na temat ich powstawania⁴².

Z wytycznych Biblioteki Kongresu wynika, że interesuje się ona także zasobami WWW wytworzone przez poszczególne władze Stanów Zjednoczonych. Kompleksowo gromadzone są materiały związane z władzą sądowniczą oraz ustawodawczą. Wybiórczo traktowane są organy związane z egzekutywą (ze względu na kompetencje innych instytucji), archiwizowane są przede wszystkim agencje na poziomie gabinetu prezydenta Stanów Zjednoczonych, wybiórczo na poziomie federalnym i wyjątkowo na poziomie stanowym⁴³. Biblioteka Kongresu

General Conference and Council, 2008, Quebec, Canada, s. 3–5. <https://digital.library.unt.edu/ark:/67531/metadc67623/> (dostęp 31.07.2020).

⁴¹ Library of Congress, Washington D.C., *Library of Congress Collections Policy*, s. 1–3. <https://www.loc.gov/resource/scd2015.00412018874> (dostęp 31.07.2020).

⁴² Przykładowa kolekcja z zasobu Biblioteki Kongresu: tamże, *Papal Transition 2005 Web Archive*, <https://www.loc.gov/collections/papal-transition-2005-web-archive/about-this-collection/> (dostęp 31.07.2020).

⁴³ *Library of Congress Collections Policy*, s. 3.

współpracuje także innymi amerykańskimi oraz zagranicznymi instytucjami na polu archiwizacji Webu, także przy budowaniu kolekcji tematycznych⁴⁴.

Podejście selektywne w archiwizacji Webu może być wykorzystywane także do gromadzenia źródeł na potrzeby badań naukowych. Jako przykład przywołać można tu Digital Archive for Chinese Studies (DACHS) działające od 2001 r. przy Uniwersytecie w Heidelbergu. Projekt ten zbiera materiały na potrzeby konkretnych badań, a mianowicie studiów sinologicznych. Zagadnienie to jest bardzo rozległe i potrzebne materiały są rozproszone, dlatego też zrezygnowano z archiwizacji masowej na rzecz selektywnej. Informacje o zasobach WWW do zgromadzenia są pozyskiwane są dwiema drogami. Pierwsza jest wciąż rozbudowywana sieć informacyjna składająca się z badaczy i użytkowników Sieci zainteresowanych Chinami, co pozwala łatwiej zidentyfikować i wybrać najważniejsze zasoby dotyczące tej tematyki, a następnie lepiej je zrozumieć i opisać. Materiały pozyskane w ten sposób są przypisywane do jednego z trzech rodzajów kolekcji: ciągłych, dokumentujących wydarzenia oraz tematycznych. Drugim sposobem jest tworzenie kolekcji specjalnych, które powstają w efekcie prac badawczych prowadzonych przez akademików oraz magistrantów⁴⁵.

Podobnie jak inne projekty DACHS posiada kryteria selekcji, zaliczają się do nich: istotność dla obecnych i przyszłych badań sinologicznych, niestabilność (np. blokowanie przez chińską cenzurę), powiązanie z wydarzeniami lub problemami istotnymi dla badaczy oraz użytkowników WWW, autorstwo lub powiązanie z postaciami o dużym wpływie na społeczeństwo. Priorytetowo są traktowane zagadnienia polityczne i społeczne, ale także inne ważne zagadnienia, np.: cenzura, prawa człowieka czy ochrona środowiska. Warty odnotowania jest fakt, iż autorzy projektu są świadomi tego, że selekcja jest obciążona subiektywizmem i polega na osądzie człowieka. W ramach DACHS powstają następujące rodzaje kolekcji: ciągłe (składają się na nie ważne witryny, które często ulegają zmianom, dlatego też są gromadzone w regularnych odstępach czasu), dokumentacja ważnych wydarzeń, tematyczne, specjalne oraz repozytorium cytowań (archiwizowane są linki przywoływane przy przypisach prac naukowych)⁴⁶.

⁴⁴ Library of Congress, Washington, D.C., *Web Archiving Collaborations*, <https://www.loc.gov/programs/web-archiving/about-this-program/web-archiving-collaborations/> (dostęp 31.07.2020).

⁴⁵ H. E. Lecher, *Small Scale Academic Web Archiving: DACHS*, [w:] *Web Archiving*, s. 221–222; Bereichsbibliothek Ostasien, *About DACHS*, https://www.zo.uni-heidelberg.de/boa/digital_resources/dachs/about_en.html (dostęp 31.07.2020).

⁴⁶ Tamże.

Archiwizacja w sposób selektywny nie musi być przeprowadzana przez odpowiednio przygotowanych do tego zadania specjalistów. Ciekawie ilustruje to prowadzony w latach 2008-2018 roku program K-12 Web Archiving, za którego organizację odpowiada Archive-It, początkowo we współpracy z Biblioteką Kongresu. W ramach tego projektu uczniowie amerykańskich szkół z programu K-12 (klasy 1-12 obejmujące etapy *primary*, *middle* i *secondary school*) mieli możliwość przeprowadzenia archiwizacji wybranych przez siebie witryn internetowych. W programie do tej pory wzięło udział 47 różnych szkół, które łącznie przygotowały 377 kolekcji poświęconych różnym aspektom korzystania przez nich z Internetu. Duża część poświęcona jest ich małym ojczyznom, edukacji oraz rozrywce, ale znaleźć wśród nich można także zbiory poświęcone historii kolei czy epidemii wirusa Ebola. Praca uczniów polegała najpierw na grupowym wytypowaniu tematów na dany rok, a następnie selekcji witryn. Do ich zadań należało przygotowanie metadanych opisowych, które stanowią odpowiedź na pytanie dlaczego dany obiekt powinien zostać zarchiwizowany⁴⁷.

Program K-12 jest bardzo interesującą propozycją. Poza zachowaniem niedużych wrywków dziedzictwa kulturowego zawartego w Internecie, pozwolił on młodym osobom zapoznać się bliżej z tym zagadnieniem. Zaangażowani uczniowie dostali możliwość poszerzenia swojej wiedzy o Sieci, a także na temat prowadzenia historycznych badań i znaczenia tego typu źródeł historycznych. Z drugiej strony kolekcje budowane przez uczniów pozwolą innym badaczom poznać sposób w jaki postrzegają otaczającą ich rzeczywistość, zwłaszcza rolę jaką pełnił dla nich Internet w okresie dorastania, a także ich zainteresowania, materiały wykorzystywane do nauki oraz codzienne problemy⁴⁸.

Uwagę warto zwrócić także na inicjatywę ArchiveTeam – luźny i otwarty kolektyw zrzeszający osoby zainteresowane zachowaniem cyfrowego dziedzictwa, do którego każdy może dołączyć i wesprzeć jego działania. Swoją działalność prowadzi on od 2009 r., kiedy Jason Scott (pracownik Internet Archive) zauważył potrzebę zarchiwizowania serwisu GeoCities, którego właściciel, firma Yahoo!, postanowił zamknąć. ArchiveTeam nie gromadzi zasobów powiązanych z określoną narodowością lub państwem, a interesuje się zasobami ważnymi dla ogół-

⁴⁷ Archive-It, *K-12 Web Archiving*, <https://www.archive-it.org/k12/> (dostęp 31.07.2020); wykaz szkół i budowanych przez nie kolekcji dostępny portalu Archive-It: <https://archive-it.org/explore?fc=organizationType%3Ak12ProjectSchools> (dostęp 31.07.2020).

⁴⁸ J. Freeman, *Seen but Not Heard: A Case Study of K-12 Web Archiving and the Importance of Student Participation in the Archives*, „Archival Issues” 2016, t. 37 nr 2, s. 37-38. DOI:archivalissues.11015.

noświatowego środowiska Sieci. Działalność grupy skupia się przede wszystkim na zabezpieczaniu serwisów i portali o charakterze społecznościowym. Zgodnie z jej założeniami w pierwszej kolejności archiwizowane są portale, które prosiły swoich użytkowników o generowanie treści i ją od nich otrzymały, następnie serwisy z treścią popularną lub unikatową, a na końcu wszystkie pozostałe⁴⁹. Każdy może zgłosić potrzebę archiwizacji jakiś zasobów za pomocą otwartego kanału IRC. Prezentowana inicjatywa prowadzi listę witryn, o których wiadomo, że zostaną niebawem usunięte z Sieci (tzw. Deathwatch, znajduje się na niej popularny polski serwis komixy.pl⁵⁰) oraz wykaz portali, które pomimo dobrej kondycji mogą być zagrożone (tzw. Firedrill⁵¹). Zabezpieczanie każdego portalu ma charakter odrębnego projektu⁵².

Podsumowanie

Strategia gromadzenia selektywnego to jedno z dwóch podstawowych rozwiązań, obok gromadzenia masowego, stosowanych w archiwizacji Webu. Podejście to opiera się na manualnym identyfikowaniu i wybieraniu zasobów pochodzących z WWW, które mają zostać zarchiwizowane. W przypadku narodowych archiwów Sieci, które działają również w sposób masowych, stanowią one niejako działania uzupełniające, pozwalające zachować to, co mogłoby być pominięte przez gromadzenie całej domeny wykonywane parę razy w roku. Dla pozostałych przypadków stanowi ona jedyne rozwiązanie umożliwiające postawionych przed nimi celów archiwizacji. Jak pokazują to przedstawione przykłady może ona być stosowana w różnych wariantach, które mogą być używane jednocześnie w zależności od potrzeb. Do częstych praktyk należy przeprowadzanie ogólnej selekcji zasobów zgodnie z zakresem gromadzenia danej inicjatywy i przypisywanie jej, do którejś z ogólnych kategorii. Rozszerzenie tego stanowi archiwizacja tematyczna oraz dokumentowanie wydarzeń, które umożliwiają gromadzenia zasobów skupionych wokół bardziej konkretnych

⁴⁹ ArchiveTeam, *Philosophy*, <https://www.archiveteam.org/index.php?title=Philosophy> (dostęp 31.07.2020).

⁵⁰ Tamże, *Deathwatch*, <https://www.archiveteam.org/index.php?title=Deathwatch> (dostęp 31.07.2020).

⁵¹ Tamże, *Alive... OR ARE THEY*, https://www.archiveteam.org/index.php?title=Alive..._OR_ARE_THEY (dostęp 31.07.2020).

⁵² Tamże, *Projects*, <https://www.archiveteam.org/index.php?title=Projects> (dostęp: 31.07.2020).

zagadnień. Za selektywne można uznać także archiwizowanie witryn i serwisów, które są zagrożone zniknięciem z Sieci.

Efektom stosowania gromadzenia selektywnego jest powstawanie rozbudowanych i ustrukturyzowanych katalogów tematycznych oraz szczegółowych kolekcji stron dotyczących konkretnych wydarzeń i zagadnień. Należy jednak zwrócić uwagę na fakt, że są one powiązane rzeczowo, a nie konieczne organicznie (np. za pomocą hiperłączy). Za selekcję zasobów odpowiedzialni są często eksperci dziedzinowi oraz angażowani są zewnątrzni specjaliści i opiera się ona o określone kryteria, dlatego można oczekiwać, że są one najbardziej relewantne dla danego zagrożenia. Analizując wskazówki i wytyczne stosowane przez poszczególne projekty widać, że starają się one gromadzić przede wszystkim źródła o wysokiej jakości, które posiadają dużą wartość informacyjną.

Ciekawą propozycją jest również zaangażowanie użytkowników Internetu do czynnego udziału w jego archiwizacji. Taką szansę dawał amerykański projekt K-12 czy irlandzki *Remembering 1916, Recoding*, a także nadal tworzy ją możliwość zgłaszania zasobów przez użytkowników pozostałych inicjatyw. Pozwoliłoby to lepiej poznać wycinki Sieci szczególnie dla nich istotne, a także samych internautów oraz ich potrzeby i zainteresowania. Wpłynęłoby to również korzystnie na ich świadomość na temat Internetu i jego kulturowego oddziaływania, a także o potrzebie zachowywania dziedzictwa kulturowego dla przyszłych pokoleń i ich badań. Takie działania przysłużyły się również promowaniu samego zjawiska archiwizacji Webu, które słabo rozpowszechnione i wymaga edukowania na jego temat.

Strategia selektywna ma niewątpliwie swoje zalety, do których zaliczyć można większą elastyczność i dokładność w porównaniu z gromadzeniem masowym. Rozwiązanie to daje większą kontrolę i pozwala także archiwizować wybrane materiały z większą częstotliwością oraz pominać, te które można uznać za bezwartościowe. Stwarza to szansę za gromadzenie zasobów dotyczących konkretnych problemów. Zarchiwizowane w ten sposób zasoby posiadają także konkretną strukturę i często są lepiej opisane. Warto także zwrócić uwagę na rozwiązania, które mogą pozwolić na ulepszenie procesu selekcji, jakim jest rozdzielenie kompetencji lub budowanie sieci informacyjnych, jak ma to miejsce w projektach PANDORA oraz DACHS. Słuszne również wydaje się angażowanie specjalistów i ekspertów od danych zagadnień, co także powinno ułatwić dobór właściwych zasobów.

Omawiany rodzaj gromadzenia nie jest jednak rozwiązaniem idealnym i posiada też swoje wady. Przede wszystkim jest ona zdecydowanie bardziej pracochłonna i wymaga większego zaangażowania człowieka, co przekłada się

na mniejszą ilość zgromadzonych zasobów. Problem ten został zauważony w przypadku chorwackiego archiwum Webu⁵³. Taką działalność mogą też mocno ograniczać limity, jakie są widoczne w inicjatywie Netarkivet. Archiwizacja selektywna w znacznie większej mierze oparta jest o subiektywne decyzje odpowiedzialnych za nią osób, ich kompetencji i motywacji, a także narzędzi, których używa do identyfikacji zasobów⁵⁴. Ponadto selektywnie archiwizowane zasoby są wyciągane z szerszego otoczenia, a inne witryny i materiały, z którymi są one powiązane mogą się nie zachować. Istnieje też naturalne ryzyko, związane z tym, że tematy lub wydarzenia wybierane do budowania mogą w przyszłości wzbudzić zainteresowania.

■ Bibliografia

- ArchiveTeam. „Alive... OR ARE THEY”. Dostęp 31.07. 2020. https://www.archiveteam.org/index.php?title=Alive..._OR_ARE_THEY.
- ArchiveTeam. „Deathwatch”. Dostęp 31.07.2020. <https://www.archiveteam.org/index.php?title=Deathwatch>.
- ArchiveTeam. „Philosophy”. Dostęp 30.07.2020. <https://www.archiveteam.org/index.php?title=Philosophy>.
- ArchiveTeam. „Projects”. Udostępniono 31.07.2020. <https://www.archiveteam.org/index.php?title=Projects>.
- Archive-It. „Explore Archived Content”. Dostęp 30.07.2020. <https://archive-it.org/explore?fc=organizationType%3Ak12ProjectSchools>.
- Archive-It. „National Library of Ireland”. Dostęp 30.07.2020. <https://archive-it.org/home/nli>.
- Archive-It. „K-12 Web Archiving”. Dostęp 31.07.2020. <https://www.archive-it.org/k12/>.
- Bereichsbibliothek Ostasien. „About DACHS”. Dostęp 31.07. 2020. https://www.zo.uni-heidelberg.de/boa/digital_resources/dachs/about_en.html.
- Biuletyn EBIB. „Nr 172 (2017): Archiwizacja internetu”. Dostęp 18.09.2020. <http://ebibojs.pl/index.php/ebib/issue/view/17>.
- Biuletyn EBIB. „Numer 128 – Archiwizacja Internetu”. Dostęp 18.09.2020. http://www.ebib.pl/?page_id=380.
- End of Term Web Archive. „Project Background”. Dostęp 31.07.2020. <http://eotarchive.cdlib.org/background.html>.
- Freeman, JoyEllen. „Seen but Not Heard: A Case Study of K-12 Web Archiving and the Importance of Student Participation in the Archives”. *Archival Issues* 37, nr 23-42 (2016). <https://minds.wisconsin.edu/handle/1793/78511>.

⁵³ M. Willer et al., dz. cyt., s. 362-363.

⁵⁴ J. Masanès, *Selection for a Web archives*, s. 76

- Gomes, Daniel, Sérgio Freitas, i Mário J. Silva. „Design and Selection Criteria for a National Web Archive”. W *Research and Advanced Technology for Digital Libraries*, zredagowali Julio Gonzalo, Costantino Thanos, M. Felisa Verdejo, i Rafael C. Carrasco, 196–207. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006. https://doi.org/10.1007/11863878_17.
- HAW Croatian Web Archive. „Browse by subject”. Dostęp 30.07.2020. <https://haw.nsk.hr/en/browse-by-subject/>.
- HAW Croatian Web Archive. „Selection criteria”. Dostęp 30.07.2020. <https://haw.nsk.hr/en/selection-criteria/>.
- HAW Croatian Web Archive. „Thematic collections”. Dostęp 30.07.2020. <https://haw.nsk.hr/en/thematic-collections/>.
- Hoffman, Starr. „Preserving Access to Government Websites: Development and Practice in the CyberCemetery”. W *World Library and Information Congress: Seventy-Fourth IFLA General Conference and Council, 2008, Quebec, Canada*, 2008. <https://digital.library.unt.edu/ark:/67531/metadc67623/>.
- Holub, Karolina, i Ingeborg Rudomino. “A decade of web archiving in the National and University Library in Zagreb.” Dostęp 30.07.2020. <http://library.ifla.org/1092/1/090-holub-en.pdf>.
- Konopa, Bartłomiej. „Zachowanie różnorodności Sieci w archiwach Webu”. *Toruńskie Studia Bibliologiczne* 12, nr 1 (2019): 45–65. <https://doi.org/10.12775/TSB.2019.003>.
- Library of Congress Washington, D.C. „Library of Congress Collections Policy Statements Supplementary Guidelines”, 2017. Dostęp 31.07.2020. <https://www.loc.gov/acq/devpol/webarchive.pdf>.
- Library of Congress, Washington, D.C. „About This Collection | Papal Transition 2005 Web Archive | Digital Collections”. Digital Collection. Dostęp 31.07.2020. <https://www.loc.gov/collections/papal-transition-2005-web-archive/about-this-collection/>.
- Library of Congress, Washington, D.C. „Web Archiving Collaborations | About This Program | Web Archiving | Programs”. Dostęp 31.07.2020. <https://www.loc.gov/programs/web-archiving/about-this-program/web-archiving-collaborations/>.
- Masanès, Julien. „Selection for a Web archives”. W *Web Archiving*, zredagował Julien Masanès, 71–91. Berlin ; New York: Springer, 2006.
- Masanès, Julien. „Web Archiving Methods and Approaches: A Comparative Study”. *Library trends* 54, nr 1 (2005): 72–90. <https://doi.org/10.1353/lib.2006.0005>.
- Masanès, Julien. „Web Archiving: Issues and Methods”. W *Web Archiving*, zredagował Julien Masanès 1–53. Berlin ; New York: Springer, 2006.
- National Archives, London. „Operational Selection Policy OSP27: UK Central Government Web Estate”, 2014. Dostęp 30.07.2020. <https://www.nationalarchives.gov.uk/documents/information-management/osp27.pdf>.
- National Library of Australia, Canberra. „Archived websites”. Dostęp 29.07.2020. <https://www.nla.gov.au/what-we-collect/archived-websites>.
- National Library of Ireland, Dublin. „Web Archive Collections”. Dostęp 30.07.2020. <https://www.nli.ie/en/udlist/web-archive-collections.aspx>.

- Netarkivet. „Selektive høstninger”. Dostęp 29.07.2020. http://netarkivet.dk/om-netarkivet/selektive-hostninger_2016/.
- Netarkivet. „Specialhøstninger”. Dostęp 29.07.2020. <http://netarkivet.dk/om-netarkivet/specialhostninger/>.
- Nielsen, Janne. *Using web archives in research: an introduction*. Aarhus: NetLab, 2016. http://www.netlab.dk/wp-content/uploads/2016/10/Nielsen_Using_Web_Archives_in_Research.pdf.
- Pandora Australia’s Web Archive. „Selection Guidelines”. Dostęp 29.07.2020. <http://pandora.nla.gov.au/guidelines.html>.
- Pandora Australia’s Web Archive. „Overview”. Dostęp 29.07.2020. <http://pandora.nla.gov.au/overview.html>.
- Pandora Australia’s Web Archive. „Partners”. ”. Dostęp 29.07.2020. <http://pandora.nla.gov.au/partners.html>.
- Schostag, Sabine, i Eva Fønss-Jørgensen. „Webarchiving: Legal Deposit of Internet in Denmark. A Curatorial Perspective”. *Microform & Digitization Review* 41, nr 3–4 (styczeń 2012): 110–20. <https://doi.org/10.1515/mir-2012-0018>.
- Seneca, Tracy, Abigail Grotke, Cathy Nelson Hartman, i Kris Carpenter. „It Takes A Village To Save The Web: The End Of Term Web Archive”. *Documents to the People (DttP)* 40, nr 1 (2012): 16–23. <https://www.journals.ala.org/index.php/dttp/issue/viewFile/590/346>.
- Thouvenin, Florent, Peter Hettich, Herbert Burkert, i Urs Gasser. „4 Web Archives”. W *Remembering and Forgetting in the Digital Age*, zredagowali Florent Thouvenin, Peter Hettich, Herbert Burkert, i Urs Gasser, 84–101. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-90230-2_6.
- Wilkowski, Marcin. „Oddolne archiwizacje Internetu jako działania społeczne”. *Archiwa – Kancelarie – Zbiory*, nr 6(8) (2015): 207–20. <https://doi.org/10.12775/AKZ.2015.007>.
- Willer, Mirna, Tanja Buzina, Karolina Holub, Jasenka Zajec, Miroslav Milinović, i Nebojša Topolščak. „Selective Archiving of Web Resources: A Study of Processing Costs”. *Program* 42, nr 4 (2008): 341–64. <https://doi.org/10.1108/00330330810912043>.

