

Josh Cowls<sup>1</sup>  
Comparative Media Studies  
Massachusetts Institute of Technology

## Kultura brytyjskiej sieci web

**Streszczenie:** Autor przedstawia brytyjski projekt BUDDAH, w którym naukowcy prowadzili badania z zakresu nauk humanistycznych korzystając ze zgromadzonych zasobów archiwalnych pobranych z sieci. Celem projektu była m.in. odpowiedź na pytanie, czy archiwizowanie stron internetowych dla celów badawczych ma sens. W artykule opisano wiele różnych badań, metodologii, studiów przypadków oraz narzędzi technicznych stworzonych, aby zrealizować te badania.

**Słowa kluczowe:** archiwizacja stron internetowych, projekt British Library, BUDDAH, badania humanistyczne, humanistyka cyfrowa

### Wstęp

W artykule przedstawiono wnioski i spostrzeżenia dotyczące projektu *Big UK Domain Data for the Arts and Humanities* (BUDDAH) prowadzonego przez British Library, Instytut Badań Historycznych na Uniwersytecie w Londynie oraz Oxford Internet Institute na Uniwersytecie w Oksfordzie. Projekt trwał od stycznia 2014 r. do marca 2015 r. Głównym jego celem było ułatwienie wykorzystania pakietów zawierających dane z domeny .uk, o rozmiarze 65 terabajtów, indeksowanej w latach 1996–2013. Indeksowania danych (*crawls*) były przeprowadzone przez Internet Archive, który na masową skalę przechwytyuje i archiwizuje strony internetowe (Kahle, 1997). Zestaw tych danych (zwany dalej archiwum internetowym) został zakupiony dla British Library przez JISC, organizację pozarządową w Wielkiej Brytanii, która wspomaga wykorzystanie technologii cyfrowych w nauce i edukacji. Jak wynika z rozważań naukowców, zbiór danych poddany był podobnym ograniczeniom i wyzwaniom, jakie były udziałem innych archiwów internetowych. Archiwum, choć ogromne, nie zawiera wszystkich stron z domeny .uk, a zgromadzone zasoby mają daty przechwycenia zamiast dat ich utworzenia. Mnożą się zatem pytania o reprezentatywność archiwum.

W ramach projektu zaproszono dziesięciu badaczy z zakresu humanistyki do wykorzystania internetowego zestawu danych w celu przeprowadzenia innowacyjnych badań. Naukowcy, którzy otrzymali stypendium za udział w projekcie, w czasie jego trwania byli doktorantami lub studiowali na wyższym poziomie – byli zatem ekspertami w swoich dziedzinach.

W tekście opisano sposoby wykorzystywania zestawu danych przez naukowców oraz wnioski z ich badań. Inicjatywa ta była oryginalna z kilku powodów. Stanowiła rzadką okazję do wykorzystania tego ekscytującego, lecz wymagającego źródła danych dla naukowców z małym doświadczeniem w korzystaniu z archiwów internetowych lub bez żadnych

---

<sup>1</sup> Josh Cowls jest absolwentem i badaczem w Comparative Media Studies w Massachusetts Institute of Technology. Przed wstąpieniem do MIT, Josh uzyskał tytuł magistra na studiach Social Science of the Internet i był asystentem w Oxford Internet Institute. Jego prace dotyczą wpływu nowych technologii i danych m.in. na kampanie polityczne, środowisko akademickie i media. Kontakt: <https://joshcowls.com/contact/>.

praktyk w tym zakresie. Ponadto, proces ten tak skonstruowano, by archiwum mogło być rozwijane w sposób powtarzalny (iteratywny), z informacją zwrotną od badaczy, co ułatwiały regularne spotkania naukowców i programistów. To wspólne działanie było szczególnie ważne dla rozwoju Shine, interfejsu używanego do prowadzenia pełnotekstowych przeszukiwań w archiwum. Ze względu na skalę i różnorodność danych w nim zawartych, interfejs wyszukiwania stał się niezbędnym narzędziem do nawigacji w archiwum. Jednak interfejs wyszukiwawczy służył również jako platforma do koncepcyjnej nawigacji w celu badań archiwów internetowych samych w sobie.

Artykuł rozpoczyna się opisem dziesięciu projektów, krótko przedstawiających istotę badań, przyjętych metod oraz wyłonionych wniosków. Doświadczenia zostały następnie podane syntezie w części dyskusyjnej, uzupełnionej serią szerszych refleksji na temat wyzwań związanych z prowadzeniem badań z wykorzystaniem archiwów internetowych oraz ich implikacji dla nauk humanistycznych, wynikających podczas korzystania z tego cennego, ale jak dotąd nieodkrytego zasobu.

## Streszczenia projektów

### Reakcje online na kryzysy instytucjonalne : BBC Online i pokłosie po Jimmym Savile

Rowan Aust z Royal Holloway, University of London, w swoim studium przypadku skoncentrowała się na następstwach poważnego skandalu, który miał miejsce w samym sercu British Broadcasting Corporation (BBC). W roku 2011, po śmierci Jimmy Savile'a – kultowego redaktora BBC, który od dziesięcioleci słynął z wiodących programów radiowych i telewizyjnych oraz z pracy charytatywnej – pojawiło się wobec niego szereg zarzutów o nadużycia seksualne. Aust opisała rewelacje na temat Savile'a jako „przerwaną stabilność”, która podtrzymuje pamięć kulturową, a jej badania koncentrowały się na zrozumieniu, jak BBC – jako wybitna instytucja – zareagowała na zarzuty. W tym celu analizowała zmiany w treściach odnoszących się do Savile'a, na przestrzeni czasu, na stronach internetowych BBC (Aust, 2015: 3).

Aust rozpoczęła badania przez przeprowadzenie powtarzającego się cyklu poszukiwań związanych z Savilem i BBC w całym archiwum [powstałym w ramach projektu BUDDAH – przyp. red.] oraz na samej stronie internetowej BBC. Porównując je z bieżącym serwisem zauważyła szereg przypadków, w których dokonano zmian w reakcji na skandal. Aust znalazła tylko jeden przypadek – wywiad z Savilem w wieloletniej serii programów radiowych BBC *Desert Island Discs* – w którym wyraźne usunięto odnośnik do materiałów internetowych z nim związanych. W innym miejscu na stronie internetowej, gdzie starano się usunąć lub zmodyfikować treść, zrobiono to niekonsekwentnie i chaotycznie. W niektórych przypadkach odnośniki usunięto lub zniszczono, a w innych, zmodyfikowano zawartość lub nadano stronie szary kolor.

Aust kontynuowała analizę porównawczą źródeł pierwotnych kontaktując się z kontrolerem polityki redakcyjnej (Controller of Editorial Policy). Dzięki korespondencji dowiedziała się, że BBC miała wytyczne zawierające procedury usuwania treści online, wdrożone dopiero w 2014 r., zaraz po pierwszych zarzutach. Badania Aust sugerują, że odwrotnie do zasad – *domniemania, że materiały opublikowane w Internecie stają się częścią trwałego, dostępnego archiwum* – opisanych w tych wytycznych, niektóre wrażliwe treści zostały po cichu usunięte (2015: 8). Mimo to próbując kontrolować narrację wokół sprawy Savile'a,

działania skutecznego wdrażania polityki „przykrywania”, usuwania lub modyfikacji treści przez BBC wydawały się być ograniczane, ze względu na skalę jej obecności online. Badania Aust są znaczące w pewnych warunkach – dają nowe światło na poważne i istotne sprawy – ale są one także lekcją na rzecz szerszego utrzymania i modyfikacji dużych archiwów instytucjonalnych i pokazują konsekwencje tych działań dla pamięci kulturowej.

### **Archiwum stron internetowych a literatura bitników<sup>2</sup>**

Rona Cran z University College London starała się zbadać w swoim projekcie akademicką i publiczną recepcję literatury bitników w Wielkiej Brytanii i ustalić, co odzwierciedlały archiwalne strony internetowe oraz czy archiwa internetowe są w związku z tym użytecznym narzędziem badawczym dla szerszych badań literackich. Ogólnie rzecz biorąc, Cran stwierdziła, że archiwum *ma wielką wagę i ogromny potencjał jako narzędzie badawcze dla badaczy literatury*, a pracę z archiwum opisała jako *wyzwolenie sensu badania jednocześnie przeszłości, jak i przyszłości, wejścia na nieznane terytorium, a także odkrywania zapomnianych artefaktów* (Cran, 2015: 3).

Jakkolwiek jest to obiecujący obszar studiów, Cran napotkała wiele wyzwań badawczych przy wykorzystaniu archiwum. Jednym z nich było „geograficzne” ograniczenie zbioru danych archiwalnych – wiele istotnych materiałów prawdopodobnie opublikowano w domach innych niż .uk. Ponadto dane dostępne w archiwum brytyjskim były *znacznie bardziej fragmentaryczne i nieporównywalne* niż dane z kolekcji literackich z innych witryn internetowych, bardziej ujednoczonych i kompletnych (2015: 2). Potrzebny był więc proces *szukania i nadawania sensu*. Autorka uważa, że do terytorium reprezentowanego przez strony archiwalne *powinny być sporządzone mapy* przez uczonych i inne zainteresowane osoby (2015: 3).

Cran stwierdziła, że archiwum w jego obecnym „bałaganie”, w niezmapowanym i arbitralnym stanie, idealnie współgra z emocjonalnym rozumieniem literatury bitników. Cran narysowała podobieństwa między stylami pisarskimi bitników – takimi jak William Burroughs, którego powieści *czyta się [...] jak niekontrolowane wymioty niedomagającej maszyny* czy nieuporządkowane zbiory danych (2015: 5). Co więcej, bitnicy cenili sobie pojęcia rozdrobnienia, elips i właściwej dla nich nieświadomości, które są *pozytywnymi aspektami archiwum internetowego w jego obecnej formie* (2015: 6). Z tego punktu widzenia badania Cran pokazują, z archiwum internetowego możemy nie tylko dowiedzieć się dużo o bitnikach, ale też nauczyć się wiele o archiwum sieciowym ich literatury.

### **Ujawnienie brytyjskiego eurosceptycyzmu w archiwum sieciowym**

Richard Deswarte z University of East Anglia próbował dowiedzieć się w swoim projekcie, czy i jak brytyjski eurosceptycyzm został odzwierciedlony w archiwum internetowym. Deswarte rozpoczął od stworzenia listy słów kluczowych odpowiednich dla miejsca Wielkiej Brytanii w Unii Europejskiej, zawierających terminy „referendum”, „eurosceptycyzm” oraz „UKIP” (akronim Eurosceptic United Kingdom Independence Party), których następnie szukano w archiwum (Deswarte, 2015: 3). Niestety, liczba wyników dla tych zapytań była ogromna, co wykluczało bliższą analizę wszystkich, czy nawet znaczących przykładów z dostępnych zasobów. Nawet gdy przeprowadzono filtrowanie wyników wyszukiwania –

---

<sup>2</sup> Beat Generation Literature – pojęcie z literaturoznawstwa określające pokolenie bitników, takich pisarzy jak Jacka Kerouac, Allen Ginsberg i Williama S. Burroughs [przyp. tłum.].

np. przez ograniczenie zakresu dat i usunięcie najliczniejszych pod względem witryn subdomen, takich jak serwisy wiadomości – liczba wyników była duża.

Niemniej jednak Deswarte potrafił znaleźć wiele elementów istotnych z punktu widzenia jego badań, w tym pełen tekst starej wypowiedzi lidera UKIP Nigela Farage'a oraz serię filmów dokumentalnych. Pomimo naukowej wartości tych badań, Deswarte odnotowuje, że *ich odkrycie było przypadkowe, a nie oparte na solidnym metodologicznym podejściu do analizy przyrastającej góry materiałów* (2015: 4). Różne wyzwania, w tym obfitość dostępnych danych oraz kwestie dotyczące jakości i spójności przechwytywania stron internetowych, wykluczają analizę ilościową na dużą skalę w celu opisanie bardziej ogólnych wzorców dotyczących eurosceptycyzmu lub powiązania ich z trendami istniejącymi offline. W jednym przypadku wyszukiwanie akronimu UKIP dało setki wyników znajdujących się na stronach sportowych regionalnej gazety. Znacznego wysiłku ze strony badacza wymagało stwierdzenie, że był to rezultat pojawienia się w aktualnościach banneru tej partii. Ten przykład ilustruje, w jaki sposób pozornie drobne elementy na stronie internetowej mogą powodować poważne problemy, których wyeliminowanie jest niezwykle czasochłonne. Projekt Deswarte'a pokazuje, że chociaż archiwa internetowe mogą zawierać cenne materiały, to znajdowanie ich i powiązanie z szerszymi trendami społecznymi w różnych momentach, stanowi obecnie duże wyzwanie dla historyków.

### **Poszukiwanie domu w historycznej sieci: etnosemiotyczne studium londyńsko-francuskich zwyczajów, jakie pojawiają się w blogach**

Saskia Huc-Hepher z University of Westminster wykorzystała liczne archiwa internetowe do opisanie etnosemiotycznego – podejścia łączącego badanie etnograficzne z analizą semiotyczną – studium przypadku francuskiej społeczności w Londynie. Huc-Hepher starała się *myśleć mniej o zarządzaniu dużą ilością danych pozwalających dotrzeć do nowych, głębszych znaczeń*, skupiając się na niewielkim zbiorze podstawowych źródeł napisanych po francusku i wykorzystując *opowiadania z prywatnego życia będące światłem w tunelu*, by uczynić zrozumiałym ogromne archiwum internetowe (Huc-Hepher, 2015).

Huc-Hepher przeszukiwała archiwum, aby stworzyć korpus odpowiednich witryn: blogów napisanych przez francuskich emigrantów mieszkających w Londynie. Istniały zalety, jak i ograniczenia tego podejścia. Użycie przy wyszukiwaniu terminów w języku francuskim okazało się być skutecznym filtrem dla stron z archiwum .uk. Jednakże wizualne elementy blogów w korpusie – w tym czcionki, zdjęcia i filmy – były często niedoskonałe lub zniszczone, grożąc *w końcu narażeniem na szwank bardzo ważnego multimodalnego podejścia semiotycznego*.

Pomimo tych niedociągnięć, Huc-Hepher potrafiła zlokalizować w wielu różnych archiwach sporo blogów autorstwa Francuzów mieszkających w Londynie. Ze względu na ograniczenia archiwum internetowego [powstałego w ramach projektu BUDDAH – przyp. red.] – np. fakt, że zawiera ono tylko witryny z domeny .uk – odnaleziono w nim tylko jeden blog, a pozostałe przykłady zlokalizowano w innych archiwach. Poprzez analizę multimodalną Huc-Hepher mogła ocenić całą serię elementów poszczególnych blogów, w tym palety kolorów, zawartość i układ bannerów, typografię czy tekst. Co ważniejsze, dzięki analizie tych blogów w pewnym okresie czasu – wykorzystując przechwytywanie tego samego bloga w różnych momentach, w różnych archiwach – Huc-Hepher potrafiła wysledzić subtelne, ale znaczące zmiany w emocjonalnym podejściu blogera w odniesieniu do Londynu.

W wielu przypadkach obserwowano stopniową integrację blogerów z ich nowym środowiskiem, odkrywając *połowiczność dyspozycji*<sup>3</sup> lub *mieszanie przyzwyczajęń*. Warto zauważyć, że Huc-Hepher zdołała przeprowadzić bogatą, refleksyjną analizę na niewielkiej próbie źródeł. Pokazuje to, że wkład do badań może mieć nawet jedna strona lub obiekt, co przypomina nam o znaczeniu archiwizacji, która w większym stopniu może wspomóc przyszłe badania.

### **Przechwytywanie, upamiętnienie oraz historycy obywatelscy: cyfrowe archiwa dotyczące opowieści więźniów wojennych z II wojny światowej**

Dr Alison Kay z Northumbria University skupiła się w swoim projekcie badawczym na wykorzystaniu archiwum internetowego do zlokalizowania i zbadania niewielkich archiwów cyfrowych – mikrokolekcji i opowieści – związanych z przeżyciami więźniów wojennych w czasie II wojny światowej (Kay, 2015). Kluczowym podejściem Kay było iteracyjne rozwijanie ciągów wyszukiwawczych, które pozwoliłyby na otrzymanie rezultatów odnoszących się przede wszystkim do więźniów wojennych z II wojny światowej, zwłaszcza w odniesieniu do ich osobistych narracji utrwalonych w pamięci. Wiązało się to z różnymi technikami filtrowania, w tym z wykluczaniem pewnych domen, np. mediów (jak [bbc.co.uk](http://bbc.co.uk)) czy witryn komercyjnych (jak [amazon.co.uk](http://amazon.co.uk)) oraz wyszukiwaniem w celu wyeliminowania nieistotnych wyników. Strategia okazała się skuteczna, bez tych filtrów liczba rezultatów w przypadku jednego z prostych wyszukiwań Kay była nie do przejścia – 53 638; z ograniczeniami było 206 – bardziej wykonalne. Ogólnie rzecz biorąc, w przypadku 11 istotnych terminów wyszukiwawczych liczba ta wyniosła 2894. Z jednej strony oznaczało to znaczący spadek z 24 727 wstępnie przefiltrowanych wyników, ale z drugiej strony, nadal pozostało zbyt dużo, aby naukowiec mógł samodzielnie rozwiązać problem w trakcie realizacji projektu.

Mimo dużej liczby wyników i ograniczonego czasu na ich ocenę, projekt Kay dał pouczający przegląd cennego materiału historycznego, który został przechwycony w ramach archiwum internetowego. Identyfikując kilka projektów gromadzących w sieci wspomnienia wojenne i filtrując tylko rezultaty z tych domen, Kay była w stanie zbadać, jaka część wspomnień z bieżącego internetu została udostępniona w archiwum internetowym. Rozstrzygnięcia różniły się w zależności od projektu. W przypadku *Wartime Memories* wyszukiwanie zwróciło prawie 10 000 wyników. Może to stanowić do dwóch trzecich z 15 000 wojennych opowieści i świadectw, choć jak zauważa Kay, duplikaty mogą to sztucznie zwiększać. Jednak w przypadku projektu *BBC's People's War* otrzymano tylko 346 wyników – niewielki odsetek z 47 000 wątków, które według projektu istnieją w internecie. Dlatego w badaniach Kay sugeruje, że choć niedoskonałe i niekompletne (lub w przypadku duplikatów „nadmiarowe”), to jednak kolekcje wspomnień zawartych w archiwum mogą okazać się cennym źródłem materiałów dla historyków i szerzej – dla społeczeństwa.

### **Bariery cyfrowe i dostępna sieć: osoby niepełnosprawne, informacja i internet**

Gareth Millward z London School of Hygiene and Tropical Medicine starał się zbadać w swoim projekcie, w jaki sposób informacje w formacie dostępnym dla niepełnosprawnych użytkowników były prezentowane w internecie. Szczególnie koncentrował się na dostępności informacji tworzonej przez organizacje osób niepełnosprawnych i przez nich samych. Wstępne badanie rozpoczęło od szeregu

<sup>3</sup> W oryg. *habitus* – nabyte umiejętności i kompetencje jednostki [przyp. red.].

przeszukiwać całego zbioru danych, starając się odkryć, w jakim stopniu organizacje osób niepełnosprawnych są reprezentowane na przestrzeni czasu. Ogólnie rzecz biorąc analiza wykazała, że Royal National Institute of the Blind (RNIB) wypadł lepiej niż jego odpowiednik pod względem liczby odnośników w zbiorze danych (Millward, 2015). Szerzej ujmując, publiczne organizacje charytatywne cieszyły się lepszym zasięgiem w porównaniu do bardziej ukierunkowanych organizacji lobbingowych. Millward zbadał także, w jakim stopniu strony internetowe organizacji osób niepełnosprawnych są spójne z zasadami dostępności opisanymi w wytycznych World Wide Web Consortium's Web Accessibility Initiative, wykorzystujących narzędzia do sprawdzania poprawności kodu.

Raport Millwarda wskazał szereg wyzwań w przeprowadzaniu tego rodzaju analiz. Po pierwsze, nazwy organizacji osób niepełnosprawnych miały znaczący wpływ na to, czy można dokładnie przeanalizować ich zasięg. Organizacje o ogólnych nazwach, takie jak Scope i Mind, dały znaczną część nieistotnych wyników, nawet w przypadku dodatkowych określeń zawartych w wyrażeniu wyszukiwawczym, takich jak „niepełnosprawność”. Po drugie, walidacja kodu nie pozwalała na podobne porównanie między witrynami internetowymi. Im dłuższe stawały się strony, tym bardziej rosła liczba błędów dostępności w kodzie. Ostatnim, bardziej powszechnym wyzwaniem, na które wskazywał Millward, był ogromny rozmiar zbioru danych i ilość potencjalnie odpowiedniego materiału. Wymagało to podejmowania szeregu ogólnych decyzji, które miały na celu zmniejszenie rozmiarów korpusu do tego stopnia, że pogłębiona analiza byłaby niemożliwa. Ostatecznie, *korpus miał niewielkie znaczenie naukowe i trudno było obronić wyniki jako reprezentatywne czy obiektywne* (2015: 8), choć wyłoniły się z tego badania interesujące wnioski.

Millward zarysował wiele obszarów, dla których można w przyszłości rozszerzyć taką analizę. Zamiast przeprowadzania masowego przeszukiwania całego archiwum, bardziej dokładnym sposobem szacowania względnego zasięgu i wpływu różnych organizacji mogłaby być analiza linków. Analiza jakościowa pojedynczych stron internetowych, podobnie jak ustne historie zawierające dane dotyczące organizacyjnej strategii online, mogłyby zasilić ustalenia ilościowe. Co więcej, znaczenie tego kierunku badań nie jest kwestionowane. W przypadku RNIB można np. zobaczyć *ciągłość związaną z dostępem do braillea przez sieć, jako podstawową część misji charytatywnej* (2010: 8). W związku z tym używanie archiwów internetowych jako głównego źródła odkrywania tego zjawiska jest nieuniknione.

## **Historia brytyjskich firm w sieci web**

Marta Musso z University of Cambridge podjęła próbę zbadania rozprzestrzeniania się korzystania z internetu wśród brytyjskich firm, ze szczególnym uwzględnieniem okresu 1996–1999, kiedy strony firm były na ogół „w powijakach”. Musso wykorzystywała wiele źródeł danych i metod, porównując przechwycone witryny firm z archiwum ze współczesnymi ich wersjami w bieżącym internecie. Prowadziła też badania ankietowe i przeglądała listy rejestracyjne stron internetowych oraz artykuły ze współczesnych gazet (Musso, 2015).

Musso zaczęła od nakreślenia stopniowego pobierania rejestracji witryn przez duże firmy internetowe na wczesnym etapie rozwoju sieci. Łącząc informacje z różnych archiwalnych i innych źródeł stwierdziła, że pomimo doboru przykładowych firm spośród działających w Wielkiej Brytanii, wiele z nich używało rdzennych domen .com, zamiast .uk. Musso zauważyła ponadto w tym okresie *ogólną tendencję [w firmach] do przejścia na domenę*

.com i całkowitego uproszczenia adresu. Odkryła także, że starsze i większe firmy skłonne były rejestrować witryny wcześniej niż inne, zwykle przed rokiem 1996. Dzięki odpowiedziom na kwestionariusz ankiety dowiedziała się także, że firma BP zarejestrowała domenę adresową bp.com już w 1989 r., chociaż służyła jedynie jako zastępcza przez wiele lat.

Po tych odkryciach Musso przeprowadziła cykl przeszukiwań archiwum według słów związanych z biznesem i handlem, które dały w rezultacie serię katalogów firmowych z tego okresu. Katalogi istniały krótko, od około roku 2000, a zaawansowane wyszukiwanie sprawiło, że były dużo mniej użyteczne. Badając niektóre z tych witryn bardziej szczegółowo, Musso zauważyła, że wiele z nich *naśladuje fizyczne przestrzenie handlowe*, zapożyczając elementy z otoczenia handlu offline – np. przez pokazanie *drzwi wejściowych* i przyjaznego dla użytkownika menu – czyniąc przeglądanie online bardziej przyjaznym dla użytkownika. Podsumowując wnioski Musso sugeruje, że wczesne doświadczenia firm w internecie odzwierciedlają bardziej ogólne wzorce korzystania z internetu, co prowadzi do *przechodzenia z osobistych technik komunikacji [...] w przestrzeń publiczną, wirtualną rzeczywistość, w której wszyscy mają dostęp do każdej przestrzeni*.

### **Ministerstwo Obrony – rozwój online**

Harry Raffal z University of Hull zbadął w swoim projekcie archiwum internetowe [powstałe w ramach projektu BUDDAH – przyp. red.], by prześledzić rozwój stron internetowych Brytyjskich Sił Zbrojnych (Britain's Armed Forces) – the Army, the Navy and the Air Force – oraz Ministerstwa Obrony (Ministry of Defence, MoD). Raffal rozpoczął od stworzenia korpusu pięciu iteracji każdej ze stron w okresie 1996–2013 (Raffal, 2015). Korpus ten został poddany analizie tematycznej, za pomocą kodowania różnych elementów na stronie, głównie tekstu, filmów wideo, obrazów i elementów nawigacyjnych. Raffal przeprowadził również analizę linków na witrynie MoD, w odniesieniu do kilku stron pomocniczych organizacji rekrutacyjnych i edukacyjnych.

Analizy te ujawniły różne cele obecności Sił Zbrojnych w sieci. Dzięki analizie tematycznej Raffal stwierdził, że choć treści i projekty strony internetowej MoD ulegały zmianie na przestrzeni czasu, zgodnie z rozwojem standardów i trendów internetowych, początkowe intencje zaprezentowania się ministerstwa online – promowanie *wizerunku firmy* i służba *potrzebom biznesowym oraz prezentacyjnym* – pozostają w dużej mierze aktualne do dziś. W przypadku Sił Zbrojnych główną ich troską pozostała rekrutacja, zwłaszcza w przypadku Armii, która zintegrowała reklamy telewizyjne z interaktywną treścią internetową i przestała się w używanej terminologii rekrutacyjnej ze słowa „kariera” na bardziej potoczne „przyłącz się”. Raffal przeciwstawia to kontynuacji wykorzystywania słowa „kariera” w Marynarce Wojennej i Siłach Powietrznych, dla których wieloletnie nazewnictwo czy bardziej techniczne stanowisko pozostało normą.

Analiza linków dokonana przez Raffala przyniosła również interesujące wyniki. Nieoczekiwanie uwypukliła prominentną rolę ministerialnej agencji Supporting Britain's Reservists and Employers Agency (SaBRE) wśród witryn internetowych Sił Zbrojnych. Wysoka koncentracja powiązanych linków ze stron władz lokalnych i stowarzyszeń rezerwistów sugerowała, że SaBRE *jako organizacja realizowała co najmniej część spraw, które miały na celu budowanie wsparcia dla członków Sił Zbrojnych*. Ogólnie rzecz biorąc, badania Raffala odniosły korzyść z tytułu utworzenia systematycznego, samodzielnego korpusu witryn

i wykorzystania różnych metod wykrywania istotnych wzorców na stronach Brytyjskich Sił Zbrojnych.

### **Szukając publicznej archeologii<sup>4</sup> w archiwach internetowych**

Lorna Richardson, obecnie z Umeå University, starała się zbadać w swoim projekcie obraz archeologii publicznej w archiwum internetowym [powstałym w ramach projektu BUDDAH – przyp. red.] (Richardson, 2015). Podobnie jak w przypadku innych projektów, ogrom dostępnych źródeł zmusił Richardson do zawężenia i doprecyzowania kryteriów wyszukiwania, od dobrze znanego stanowiska archeologicznego Stonehenge do trzech „skromniejszych” obiektów, co dawało znacznie mniej wyników. Richardson zastosowała ngram – funkcję archiwów służącą do wyświetlania wyników na przestrzeni czasu. Stwierdziła, że dla stanowiska Ness of Brodgar, znaleziska archeologicznego odkrytego w Orkney pod koniec lat 90., *Ngram pięknie wizualizuje informacje, w jaki sposób witryny archeologiczne pokazywały postępy w odkryciach*. Richardson mogła pokazać, że informacje pochodzące ze stanowiska archeologicznego stopniowo trafiają na witryny archeologiczne.

Ogólnie ujmując Richardson zauważa, że w archiwum znajdują się setki tysięcy stron zawierających materiały potencjalnie istotne dla archeologii publicznej. Podejście Richardson do pracy z masowymi danymi przekształciło się w *łatwe do wykonanie ćwiczenia bazujące na kilku kluczowych stanowiskach i terminach archeologicznych*. W istocie podejście to można rozszerzyć na bardziej ogólnie badania archiwów internetowych. Jak Richardson sugeruje, *stosowanie podejścia archeologicznego do zbadania, zrekonstruowania i ponownego wykorzystania technologii dawnych iteracji sieci World Wide Web* może poprawić naukowe pojmowanie tego, w jaki sposób efektywnie korzystać z archiwum.

### **Czy istnieją sieci internetowe dla społeczności poetów?**

Helen Taylor z Royal Holloway, University of London przeprowadziła badania nad obecnością sieci poezji w archiwum internetowym [powstałym w ramach projektu BUDDAH – przyp. red.] (Taylor, 2015). Sieci poezji funkcjonowały przed powstaniem internetu, umożliwiając nieformalne rozpowszechnianie treści między różnymi odbiorcami. Taylor próbowała zbadać, czy sieci poezji online *istnieją same w sobie czy też obecność online jako grupy jest jedynie rodzajem „miejsca zastępczego”<sup>5</sup> lub „katalogu”<sup>6</sup>* (2015: 2). Taylor ograniczyła swoją analizę do dwóch witryn poetyckich, które reprezentują obie strony tego spektrum: Forum Poezji (The Poetry Forum), miejsce do dzielenia się i komentowania własnej poezji oraz strony internetowej Towarzystwa Poetyckiego Uniwersytetu w Oksfordzie (Oxford University Poetry Society's).

Analizując The Poetry Forum, Taylor zaobserwowała wiele cech, które wskazują na to, że jest to prawdziwa społeczność. Opcjonalny proces rejestracji zachęca autorów do tworzenia profilu promującego dzielenie się oryginalnymi utworami i komentarzami, jakkolwiek nie musi on wiernie odzwierciedlać ich osobowości. Taylor przypuszcza, że strona *jest przykładem tego, jak sieć poetycka faktycznie funkcjonuje online [...]; tego rodzaju interakcje i wymiana nie byłyby możliwe przed powstaniem internetu* (2015: 3). Taylor odkrywa, że jest to zdecydowanym przeciwieństwem Towarzystwa Poetyckiego z Uniwersytetu

<sup>4</sup> Archeologia publiczna jest w Wielkiej Brytanii praktyką prezentowania opinii publicznej danych archeologicznych i ich interpretacji [przyp. red.].

<sup>5</sup> W oryg. placeholder [przyp. red.].

<sup>6</sup> W oryg. directory [przyp. red.].



w Oksfordzie, gdzie *obecność online służy tylko do* [zachęty – przyp. red.] *zaangażowania się offline* (2015: 4). Biorąc pod uwagę czasowo reprezentatywną serię przechwytywań stron, Taylor stwierdziła, że wiersze członków [towarzystwa – przyp. red.] bardzo rzadko pojawiają się na witrynie, nawet jeśli nie ma tam narzędzi do komentowania czy dyskusji. Taylor przypuszcza, że rozważano decyzję, żeby nie przenosić debat do sieci, zważywszy na wrywkowe użycie różnych adresów URL na przestrzeni czasu oraz częste pojawianie się komunikatów „zmieniono adres”.

Wyróżniając te dwa przykłady znajdujące się w archiwum internetowym, badania Taylor wykazują różnorodność podejść do tworzenia i utrzymywania sieci poetyckich. Podsumowuje ona, że chociaż sieć była kreatywna w ułatwianiu łączenia geograficznie rozproszonych uczestników, w warunkach gdy znajomości wirtualne nie są konieczne, witryna może odgrywać zupełnie inną rolę.

## Dyskusja

Zapewnienie dziesięciu naukowcom bezpośredniego dostępu do danych za pośrednictwem interfejsu wyszukiwania i ściśle zaangażowanie ich w proces rozwojowy narzędzi przyniosło szereg nowych spojrzeń na użyteczność archiwów internetowych dla badań humanistycznych. Każdy z badaczy poświęcił znaczne części w swoich raportach z projektów rozważaniom nad potencjałem i zagrożeniami prowadzenia badań przy użyciu zbiorów danych z archiwum internetowego. W dalszej części artykułu refleksje te zostaną poddane syntezie, a szerzej opisane zostaną trzy tematy: jak badacze zdefiniowali archiwa internetowe [konceptualizacja archiwów – przyp. red.]; podejmowane strategie badawcze; oraz znaczenie narzędzi wyszukiwawczych do nawigacji w archiwum z ogromną ilością danych.

## Konceptualizacja archiwów

Dziesięciu badaczy realizujących projekt – choć wcześniej miało niewielkie lub żadne doświadczenia w tym zakresie – znalazło wiele charakterystycznych zjawisk dotyczących archiwów internetowych, które były już podkreślane przez innych uczonych (Brügger, 2012; Schneider and Foot, 2004). Wielu z nich zauważyło zarówno podobieństwa, jak i różnice między danymi z archiwów internetowych a tradycyjnymi archiwami. Richard Deswarte zasugerował, że *pod wieloma względami termin „archiwum” jest błędną nazwą, ponieważ to, co naukowiec naprawdę napotyka nie jest stroną internetową w swojej pierwotnej formie, ale rekonstrukcją poprzedniej wersji strony internetowej – często niekompletną* (Deswarte, 2015: 6). Alison Kay wypowiedziała się zauważyła w podobnym duchu: *jako historyk muszę przypominać sobie, że strony online znikają. Mamy [tylko] ich reprezentacje*. Dlatego archiwa internetowe mogą mieć mniej wspólnego z archiwami historycznymi – które zazwyczaj mają charakter tekstowy – niż z archeologicznymi artefaktami, metaforą proponowaną przez Lornę Richardson. W związku z tym istnieje techniczna rozbieżność między tym, w jaki sposób strona internetowa pojawiła się w bieżącym internecie, a jak jest odtwarzana w archiwum internetowym, wraz z czasową różnicą między pierwszym użyciem artefaktu archeologicznego a jego znalezieniem. Rona Cran zauważyła również podobieństwa koncepcyjne między archiwum internetowym a *niekontrolowanymi wymiotami niedomagającej maszyny*, które charakteryzują literaturę bitników (Cran, 2015: 5).

Nie należy jednak wyolbrzymiać koncepcyjnego odejścia od tradycyjnych archiwów, których odpowiednikami są archiwa internetowe. Jak zauważa Alison Kay, *masowe drukowa-*

*nie było w ciągu ostatnich kilku latach dokuczliwe pod względem objętości, tak jak dziś wyzwaniem jest archiwizowanie sieci. W istocie, podobnie jak w przypadku archiwów internetowych, większość archiwów historycznych podlega w pewnym stopniu arbitralności w odniesieniu do tego, co jest lub nie jest zachowywane. W przypadku archiwizacji internetowej zarówno czynniki techniczne, jak i kuratorskie mogą wpływać na to, co jest przechowywane, a co odrzucane. Archiwum internetowe wykorzystane w niniejszych badaniach zasadniczo reprezentuje tylko część archiwum z domeny .uk ze znacznie większego zbioru Internet Archive, który przechowuje ogromne obszary bieżącej sieci i powiązań między witrynami. Jednak nawet przy tej wielkiej skali działają zasady zarządzania i polityki Internet Archive, polegające na poszanowaniu działania protokołów robots.txt podczas indeksowania. Niezależnie od tego, czy strona pojawia się w Internet Archive czy nie, jest to rezultat zarówno warunków technicznych, jak i wynikających z nadzoru nad danymi. Pod pewnym względem fakt, że czynniki techniczne czy czynniki subiektywne wpływają na to, co pojawia się w archiwum internetowym można je uznać za odświeżające zobiektywizowane narzędzie [...], co sprawia, że ostateczna kolekcja nie jest tylko refleksją [archiwisty], a bardziej materiałem samym w sobie, jak sugerowała Saskia Huc-Hepher.*

Naukowcy uznali, że archiwa internetowe, przynajmniej w ich obecnym stanie, reprezentują ciekawą pozycję w stosunku do poprzednich źródeł danych. Są zarówno podobne, jak i odróżniające się od tradycyjnych archiwów historycznych, a także łączą je związki pojęciowe z tradycjami archeologicznymi i literackimi. W dalszym rozdziale wyjaśniono, w jaki sposób te różne perspektywy wyobrażeń o archiwach internetowych stanowią alternatywne sposoby wykorzystywania ich w celach badawczych.

## **Podjęmowane strategie**

Szeroki zakres tematyki badań realizowanych przez dziesięciu naukowców w ramach omawianego projektu był odzwierciedlony w dużym stopniu przez różnorodność podejść metodologicznych, pomimo że wszyscy badacze mieli dostęp do tych samych danych i tych samych narzędzi. Te zróżnicowane podejścia można jednak pogrupować w dwie przeciwstawne strategie: podejście „część całości” lub podejście „całość części”. Czy badacze poradzą sobie z całym zbiorem danych archiwalnych przy jego ogromnym rozmiarze i złożoności, używając wyszukiwarki do wyodrębnienia konkretnych elementów ze zarchiwizowanej sieci, związanej z ich badaniami (przy podejściu „część całości”)? Czy raczej powinni ograniczyć swoje zainteresowanie badawcze do wcześniej zdefiniowanego, merytorycznie istotnego podzbioru zasobów (przy podejściu „całość części”)? Te strategie nie wykluczają się wzajemnie, nawet w indywidualnym projekcie, a wielu naukowców wykorzystało obydwie w swoich badaniach. Nakreślone tu rozróżnienie pomaga wyjaśnić mocne i słabe strony prowadzenia cennych badań.

Przykładem podejścia „część całości” jest badanie Richarda Deswarte związane z euro-sceptycyzmem. Jedną wyraźną zaletą strategii Deswarte'a polegającej na przeszukiwaniu archiwum jako całości jest to, że wyniki są reprezentatywne dla samego archiwum, umożliwiając analizę wzdłużną powszechnego występowania określonego zjawiska (w tym przypadku euro-sceptycyzmu). Nie oznacza to jednak, że archiwum jest samo w sobie reprezentatywne dla społeczeństwa. Liczne wspomniane wyżej problemy, takie jak niekonsekwentne przechwytywanie stron, rodzą wątpliwości co do rzeczywistej reprezentatywności czy odzwierciedlenia społeczeństwa w archiwum. Ponadto inne praktyczne kwestie ograni-

czają użyteczność tej strategii. Jak stwierdził Deswarte, liczba wyników zwróconych podczas całościowego przeszukiwania archiwum może być ogromna, zwłaszcza gdy badacze poszukują dowodów dotyczących zjawisk ogólnych (Deswarte, 2015). Z tą strategią jest związany również problem „i co z tego”? opisany przez Garetha Millwarda, który twierdził, że *jedynie poprzez posegregowanie tych wyników możemy uzyskać jakikolwiek realny sens użyteczny dla badaczy* (Millward, 2015: 5). Przy tak rozległej sieci, „wyławianie” ogromnych zasobów może być po prostu zbyt czasochłonne. Problem ten zauważyli również Alison Kay i Lorna Richardson.

Usiłowanie zgromadzenia i analizy „części całości” archiwum jest zatem trudne. Inni badacze przyjęli z kolei strategię „całość części”. To podejście oznacza dokładne skupianie się na wcześniej zdefiniowanym zbiorze źródeł – zazwyczaj na jednej witrynie lub małej liczbie stron – i analizowaniu ich jako pewnej całości. Projekty wykorzystujące tę metodę to przede wszystkim badania witryn Ministerstwa Obrony prowadzone przez Harry’ego Raffalla, analiza dwóch różnych sieci poetyckich Heleny Taylor oraz dociekania Rowana Austa dotyczące skandalu wokół Jimmy’ego Savilego, odzwierciedlonego na witrynach BBC. I znowu, mamy oczywiste zalety tego podejścia. Naukowcy mogą na początku korzystać z funkcji filtrowania domen przy wyszukiwaniu, co daje znacznie mniejszy zbiór wyników, pozwalający na bardziej dogłębną analizę „line-by-line” wszystkich zwróconych wyników. Mimo że taka metoda może dać zwartą i bardziej spójną wewnątrznie grupę zasobów, którą można analizować jakościowo, należy zwrócić uwagę, że elementy wybrane do analizy nie są reprezentatywne dla archiwum jako całości.

Obydwie strategie – „części całości” i „całość części” – mają mocne i słabe strony przy prowadzeniu wartościowych badań z wykorzystaniem archiwum. Podsumowano je w tabeli 11.1. Generalnie podejmowanie decyzji o wyborze strategii powinno się odbywać w oparciu o charakter pytania badawczego. Jeśli koncentruje się ono na szerokim spektrum społecznym lub historycznym, sensowny będzie wybór metody „części całości”, przy jednoczesnym uwzględnieniu poważnego wyzwania jakim jest skala, które to podejście za sobą pociąga. Jeśli natomiast badania koncentrują się na konkretnej jednostce lub zdarzeniu, szczególnie jeśli jest to związane a priori z określoną subdomeną lub witryną, podejście „całość części” będzie najlepsze.

Omówiono tu dwa przeciwstawne sposoby prowadzenia istotnych badań przy użyciu dużych zbiorów danych z archiwum internetu. Oczywiście, obie przedstawione strategie nie wykluczają się wzajemnie, ponieważ wielu badaczy wykorzystywało jedną i drugą na różnych etapach. Nie są one również ściśle dychotomiczne. „Sfera web” (Schneider i Foot, 2004) – trzecia co do wielkości z pięciu „warstw” analizy sieci wymienionych przez Brüggera (2012) – mogłaby być jednostką analizy w każdej z metod, w zależności od wielkości i charakteru „sfery”, o której mowa [w pytaniu badawczym – przyp. red.]. Ponadto obie strategie obejmują na pewnym etapie procesu badawczego przeszukiwanie archiwum. W dalszej części omówiono bardziej szczegółowo cel i proces wyszukiwania.

Tab. 11.1 Porównanie strategii w badaniach z archiwum internetowym.

Wybrana strategia	Podsumowanie procesu	Kluczowe zalety	Kluczowe ograniczenia
Część całości	Przeszukiwanie całego archiwum pod kątem szerszego zjawiska historycznego/społecznego (filtrowanie przebiega głównie a posteriori)	Możliwość traktowania archiwum jako całości i posiadanie jasnego stanowiska względem archiwum	- Archiwum jest niekoniecznie reprezentatywne dla społeczeństwa - Dla analizy ilościowej istotne są ustrukturyzowane dane, które w archiwach internetowych nie są uporządkowane w sposób przydatny
Całość części	Przeszukiwanie określonej subdomeny lub witryny pod kątem konkretnej jednostki lub wydarzenia (filtrowanie przebiega głównie a priori)	Możliwość przyjęcia wrażliwego, gruntownego pojmowania witryn internetowych i ich elementów	- Trudno ostatecznie stwierdzić, w jaki sposób analizowane zasoby dotyczą lub reprezentują archiwum jako całość - Istotne wyniki wyszukiwania są odnajdywane przypadkowo, a nie w sposób usystematyzowany

## Korzystanie z narzędzi wyszukiwawczych

Niezależnie od strategii wykorzystywanej przez naukowców, ogromny rozmiar bazy danych oznacza konieczność korzystania z interfejsu wyszukiwania umożliwiającego dostęp do zasobów, ich pogrupowanie oraz analizowanie zgodnie z pytaniami badawczymi. Interfejs Shine, z którego korzystali badacze, został opracowany w trakcie realizacji projektu i jest w dużej mierze oparty na doświadczeniach badaczy. Rozwój i stosowanie interfejsu automatycznie spowodowały napływ wielu pytań koncepcyjnych oraz problemów. Badania nie odbywają się w próżni – jest to coś, co naukowcy humaniści doceniają bardziej niż inni. Wielu badaczy zauważyło, że ich wcześniejsze doświadczenia z użyciem wyszukiwarek do poruszania się w bieżącej sieci wpłynęły na założenia dotyczące wyszukiwania w archiwach internetowych. Richard Deswarte sformułował wyrażenie – *Googlowe myślenie* – do opisanego zestawu oczekiwań badaczy względem działania interfejsu wyszukiwawczego (Deswarte, 2015: 9).

Kluczowa różnica – w zasadzie prosta, ale w praktyce dezorientująca – odnosi się do porządkowania wyników wyszukiwania. Algorytmy opracowane przez Google i inne wyszukiwarki są tak dochodowe, jak nieuchwytnie, szeregując w ułamku sekundy miliardy wyników według trafności. W przypadku interfejsu wyszukiwarki Shine rozmiar indeksu jest ogromny – łącznie z setkami tysięcy wyników na podstawowe zapytania – ale porządkowanie wyników jest znacznie mniej zaawansowane. Naukowcy mogą sortować wyniki np. według tytułu danej pozycji lub daty jej przechwycenia, ale nie według trafności. Tak więc, jak stwierdził Richard Deswarte, *wszystkie wyniki* – nie tylko kilka pierwszych – *będą potencjalnie interesujące* (Deswarte, 2015: 7).

Patrząc z innej perspektywy, niektórzy badacze uznali jednak to ograniczenie jako dające swobodę, a nawet więcej możliwości. Rona Cran przyjęła *celowo nieusystematyzowane podejście do archiwum, traktując je jako coś podobnego do dużej grupy nieposortowanych dokumentów, w odróżnieniu od Google*. Dzięki temu *mogła konfrontować to ze swoją per-*

spektywą. Dla Cran proces ten *zwiększył intelektualną wiarygodność* jej studiów, ponieważ *wykorzystywała procesy argumentacji i selekcji, które były unikalne* dla niej jako badacza humanisty (Cran, 2015: 5). Tym samym dzięki ograniczeniom interfejsu wyszukiwawczego jej ekspertyzy naukowe nabrały nowego znaczenia.

To poczucie większej kontroli nad procesem badawczym zostało wzmocnione przez dalsze zmiany interfejsu. Jak zauważono, proces rozwoju odbywał się iteracyjnie, w trakcie odpowiadania na opinie badaczy. Co najważniejsze, zaraz po zgłoszeniu przez badaczy potrzeby tworzenia własnego zestawu rezultatów poprzez wydzielenie wyników wyszukiwania i tworzenie trwałej kolekcji, pojawiła się ta nowa funkcja. Po połączeniu jej z zaawansowanymi narzędziami wyszukiwania, umożliwiającymi przeszukiwanie tylko określonych domen lub zawężanie według czasu, dano naukowcom ulepszony zestaw narzędzi, dzięki któremu można było pracować z dostępnymi danymi.

Jednak nawet przy włączeniu tych narzędzi nadal niezmiernie trudne jest definitywne stwierdzanie zawartości archiwum w ogóle, szczególnie dla tych badaczy, którzy szeroko ujęli swe pytania badawcze. Jednym z podejść byłoby utworzenie małej próbki danych, która następnie mogłaby być rzeczowo analizowana przez badacza. Jednak Richard Deswarte zauważył, że chociaż *ustrukturyzowane dane oznaczają, że można podjąć jasną i naukowo uzasadnioną decyzję dotyczącą rozmiaru i znaczenia reprezentatywnych próbek [...], niefortunne i problematyczne jest to, że dane w archiwach internetowych są prawie całkowicie pozbawione struktury, przynajmniej w zakresie treści* (Deswarte, 2015: 3). Kolejny problem został wyjaśniony przez Garetha Millwarda sugerującego, że w rzeczywistości wystąpił względny niedostatek danych, pozwalający w tradycyjny sposób odpowiedzieć na pytania dotyczące historycznych zagadnień. Tradycyjni historycy *zwyczajowo identyfikują problem i bazę źródłową, wracają do archiwum, a potem wydobywają to, co mogą, dopóki zasób się nie wyczerpie. Tylko to jest możliwe, ponieważ mamy niewiele dowodów, które przetrwały* (Millward, 2015: 6).

Podsumowując, doświadczenia naukowców z funkcją wyszukiwawczą wydają się być po równi frustrujące jak i twórcze. Frustrujące, ze względu na brak jakiegokolwiek istotnego porządku, dzięki któremu naukowcy mogliby się zająć bogatymi zasobami. Twórcze, gdyż przy wcześniej określonym warunku trafności, naukowcy muszą podejmować więcej decyzji w oparciu o wiedzę ze swojej dziedziny. Oczywiście, perspektywy te nie wykluczają się wzajemnie. Najbardziej frustrujące doświadczenia są łagodzone przez np. bardziej wydajne i pasujące funkcje wyszukiwawcze, a jednocześnie umożliwiające badaczom podejmowanie świadomej decyzji, co do tego, co należy do korpusu. Jednak pomimo udoskonaleń technicznych, naukowcy i programiści muszą nadal krytycznie oceniać rezultaty przeszukiwania pełnotekstowego. Richard Deswarte stwierdził, że *jego dominacja jako głównego podejścia do dostępu do archiwów internetowych nie może być kwestionowana*, a Alison Kay nadmieniła, że *historycy muszą wnieść także wkład do dyskusji na temat źródeł jutra* (Deswarte, 2015: 9).

## **Wnioski**

Każdy z omówionych tutaj dziesięciu studiów przypadków przyczynił się do rozwoju badań nad archiwami internetowymi zarówno w określonych dziedzinach, jak i dzięki zastosowanym innowacyjnym podejściom metodologicznym. Odzwierciedlają to wstępne cele projektu BUDDAH, który zakładał nie tylko *podkreślenie wartości archiwów internetowych jako*

źródeł do badań w zakresie nauk humanistycznych i nauk z obszaru sztuki, ale także ustanowienie teoretycznych i metodycznych ram analizy archiwów internetowych (Big UK Domain Data for the Arts and Humanities, n.d.). Poprzednia część artykułu pokazuje, że takie ramy są w fazie rozwoju, aczkolwiek jeszcze na etapie powstawania. W projekcie wykazano też znaczenie łączenia perspektyw naukowców na każdym etapie opracowywania zestawu danych.

W artykule wyjaśniono również, jak wiele pozostaje do zrobienia, aby wykorzystać duży potencjał archiwów internetowych jako źródła do badań w zakresie sztuki i humanistyki. Opisano nowe strategie i techniki stosowane przez naukowców w projekcie, choć wiele kwestii pozostaje nierozwiązanych. Należą do nich np.: jak postępować z niestarannymi i niekompletnymi danymi; jak badania archiwów internetowych są wcielane do głównego nurtu wielu różnych dyscyplin; w jaki sposób można z powodzeniem zaprezentować wyniki zapytań do zbioru danych.

Od czasu zakończenia poszczególnych projektów badawczych, zarówno podstawowy zestaw danych archiwalnych, jak i używany do nich interfejs Shine, były stale rozwijane, w dużej mierze w odpowiedzi na opisane tu doświadczenia badaczy (Jackson, 2016). Interfejs jest teraz dostępny dla wszystkich zainteresowanych prowadzeniem własnych badań nad obrazem społeczeństwa brytyjskiego w internecie w latach 1996–2013. Narzędzie wyszukiwawcze oferuje obecnie dwie główne funkcje: przeglądanie wyników oraz analizę trendów, pokazującą podobne występowanie słowa lub frazy z archiwum w czasie 18 lat. Narzędzia te oraz więcej wskazówek można znaleźć na stronie <https://www.webarchive.org.uk/shine>.

Należy podkreślić, że osoby, które zetkną się w przyszłości z archiwami internetowymi, nie będą miały podobnej przyjemności, jakiej doświadczyli naukowcy w tym projekcie, np. kontaktu z osobami tworzącymi zbiory danych, a także umiejętności dzielenia się w grupie ideami i rozwiązaniami. Ponieważ archiwa internetowe nadal się rozwijają, należy zawsze pamiętać o badaczach jako grupie użytkowników, nawet jeśli nie zawsze będą dostępni. W artykule zaprezentowano też wiele osiągnięć naukowców, napotykanych przez nich wyzwań i, co najważniejsze, sposobów definiowania i traktowania archiwów internetowych jako źródła naukowego. Miejmy zatem nadzieję, że niniejszy artykuł – i badania, które zostały tu przedstawione – posłużą jako źródło nie tylko uczonym zaangażowanym w ten powstający obszar badań, ale także wszystkim zaangażowanym na każdym etapie rozwoju archiwów internetowych i wykorzystania ich do celów badawczych.

#### **Bibliografia<sup>7</sup>:**

1. Aust, R. (2015). Online reactions to institutional crises: BBC Online and the aftermath of Jimmy Savile. Retrieved 19 April 2016, from <http://sas-space.sas.ac.uk/6100/>.
2. Big UK Domain Data for the Arts and Humanities (n.d.). Retrieved 19 April 2016, from <http://buddah.projects.history.ac.uk/about/aims-and-objectives/>.
3. Brügger, N. (2012). Web history and the web as a historical source. *Zeithistorische Forschungen* 9(2): 316–25.
4. Cran, R. (2015). 'all writing is in fact cut ups': The UK Web Archive and Beat literature. Retrieved 19 April 2016, from <http://sas-space.sas.ac.uk/6101/>.

---

<sup>7</sup> Styl bibliografii zachowano zgodnie z oryginałem. Wszystkie odesłania do stron internetowych przedstawiają wersję aktualną w dn. 5.05. 2017 r. [przyp. red.].

5. Deswarte, R. (2015). Revealing British Euroscepticism in the UK Web Domain and Archive Case Study. Retrieved 19 April 2016, from <http://sas-space.sas.ac.uk/6103/>.
6. Huc- Hopher, S. (2015). Searching for Home in the Historic Web: An Ethnosemiotic Study of London-French Habitus as Displayed in Blogs. Retrieved 19 April 2016, from <http://sas-space.sas.ac.uk/6252/>.
7. Jackson, A. (2016). Introducing SHINE 2.0 – A Historical Search Engine. UK Web Archive Blog. Retrieved 7 March 2016, from <http://britishlibrary.typepad.co.uk/webarchive/2016/02/updating-our-historical-search-service.html>.
8. Kahle, B. (1997). Preserving the internet. *Scientific American* 276(3) : 82–3.
9. Kay, A. (2015). Capture, commemoration and the citizen historian: digital shoebox archives relating to PoWs in the Second World War. Retrieved 19 April 2016 from <http://sas-space.sas.ac.uk/6248/>.
10. Millward, G. (2015). Digital barriers and the accessible web: disabled people, information and the internet. Retrieved 19 April 2016, from <http://sas-space.sas.ac.uk/6104/>.
11. Musso, M. (2015). A history of UK companies on the web. Retrieved 19 April 2016, from <http://sas-space.sas.ac.uk/6251/>.
12. Raffal, H. (2015). The Online Development of the Ministry of Defence (MoD) and Armed Forces. Retrieved 19 April 2016 from <http://sas-space.sas.ac.uk/6250/>.
13. Richardson, L. (2015). Looking for public archaeology in the web archives. Retrieved 19 April 2016 from <http://sas-space.sas.ac.uk/6249/>.
14. Schneider, S. M. and Foot, K. A. (2004). The web as an object of study. *New Media and Society* 6(1): 114–22.
15. Taylor, H. (2015). Do online networks exist for the poetry community? Retrieved 19 April 2016, from <http://sas-space.sas.ac.uk/6105/>.

*Przekład Bożena Bednarek-Michalska i Lidia Derfert-Wolf, Toruń/Bydgoszcz, maj 2017.*

*Tłumaczenia dokonano za zgodą autora. Oryginał jest opublikowany na licencji Creative Common 4.0: COWLS, J. Cultures of the UK web. W: The Web as History. Using Web Archives to Understand the Past and the Present. Ed. Niels Brügger, Ralph Schroeder [online]. UCL Press University College London, 2017, s. 220–237. [Dostęp 21.04.2017]. ISBN: 978-1-911307-59-4. Dostępny w: <https://www.ucl.ac.uk/ucl-press/browse-books/the-web-as-history>.*