

Data przesłania tekstu: 25 X 2017 r.

Data przyjęcia tekstu do druku: 27 X 2017 r.

DOI: <http://dx.doi.org/10.12775/AKZ.2017.009>



BARTŁOMIEJ KONOPA

(Uniwersytet Mikołaja Kopernika w Toruniu)  
bartlomiejkonopa@gmail.com

**SPOTKANIE WARSZTATOWE „ARCHIWIZACJA INTERNETU”  
ZORGANIZOWANE PRZEZ MINISTERSTWO  
KULTURY I DZIEDZICTWA NARODOWEGO  
Warszawa, 25 VIII 2017 r.**

W dniu 25 VIII 2017 r. w siedzibie Filmoteki Narodowej – Instytutu Audiowizualnego (FINA) odbyło się spotkanie warsztatowe, którego inicjatorem było Ministerstwo Kultury i Dziedzictwa Narodowego. Zjawili się na nim przedstawiciele powyższego urzędu, a także Naczelnego Dyrektora Archiwów Państwowych (NDAP), Narodowego Archiwum Cyfrowego (NAC), Biblioteki Narodowej oraz Uniwersytetu Mikołaja Kopernika w Toruniu (UMK). Organizatorzy obrali za cel rozpowszechnienie podjętego zagadnienia oraz pogłębienie wiedzy na jego temat, co w niedalekiej przyszłości miało przyczynić się do podjęcia w Polsce nowych projektów archiwizacji zasobów Internetu. Aby przybliżyć uczestnikom problemy i wyzwania z tym związane do wygłoszenia referatów zaproszono troje przedstawicieli instytucji, które posiadają już doświadczenia w tym zakresie: Gemmę W. Batson z Internet Archive, Jasona Webbera z British Library oraz Justynę Soliwodę z NAC.

Główne wystąpienia krótkim wstępem poprzedziła Agnieszka Rosa (UMK), która w kilku zdaniach określiła podejmowaną problematykę archiwizacji Internetu. Przypomniała również o trzech związanych z tym projektach przeprowadzonych przez polskie instytucje oraz wspomniała o inicjatywach podjętych przez Marcina Wilkowskiego w ramach działań Laboratorium Cyfrowego Huma-

nistyki Uniwersytetu Warszawskiego. Prowadząca podsumowała także stan badań dotyczący omawianego zagadnienia zawarty w polskiej literaturze naukowej.

Jako pierwsze zaprezentowano wystąpienie Gemmy Batson z Internet Archive. Na początku swojej prezentacji przedstawiła instytucję z której się wywodzi, a więc założoną w 1996 r. w San Francisco organizację non-profit, która pełni funkcje cyfrowej biblioteki i archiwum. Pokrótkie scharakteryzowała także posiadany przez nią zasób w postaci zdigitalizowanych książek, programów telewizyjnych, oprogramowania komputerowego i przede wszystkim zarchiwizowanych witryn internetowych. Dalsza część dotyczyła już stricte archiwizacji Internetu, w trakcie której prelegentka omówiła i porównała dwa projekty prowadzone przez Internet Archive, a mianowicie Wayback Machine oraz Archive-It. Pierwszy z nich, a także starszy, to technologia *opensource* wykorzystywana przy największym archiwum Webu, natomiast drugi stanowi platformę usług umożliwiającą tworzenie dowolnie ukierunkowanych kolekcji. Batson wspomniała również o projektach obejmujących konkretne domeny prowadzone przez biblioteki narodowe i inne instytucje rządowe, aby następnie przejść do pytań, które należy sobie postawić przed podjęciem takiej inicjatywy. Wskazała ona przede wszystkim na konieczność określenia wycinka zasobów sieciowych, które mają zostać objęte archiwizacją, komu i na jakich zasadach będą udostępniane, a także zasobów potrzebnych do ich realizacji.

Jako drugi wypowiedział się przedstawiciel British Library – Jason Webber. W swoim wystąpieniu przedstawił on historię rozwoju UK Web Archive, w którym zawarł cele postawione przed tym projektem oraz wyzwania i problemy, które pojawiły się w trakcie jego realizacji. Uruchomiony w 2004 r. przez Brytyjską Bibliotekę Narodową projekt, w swojej początkowej fazie gromadził witryny internetowe w oparciu o dobrowolne zgody, co pozwalało na pełne ich wykorzystanie, jednak znacznie utrudniało to ich pozyskiwanie. W późniejszym czasie postanowiono pobierać zrzuty witryn organizacji publicznych domniemywając ich pozwolenia na to, a dopiero dokonane w 2013 r. zmiany w prawie pozwoliły na poszerzenie zakresu UK Web Archive na zasoby sieci znajdujące się w brytyjskiej domenie internetowej. Pozyskiwanie zbiorów w ten sposób narzuciło jednak kolejne ograniczenia, które pozwalają na udostępnienie zarchiwizowanych witryn wyłącznie w siedmiu siedzibach bibliotek oraz umożliwiając korzystanie z danej witryny tylko jednej osobie jednocześnie. Webber przedstawił także wykorzystywane w projekcie technologie, zwracając szczególną uwagę na pełnotekstowe indeksowanie zarchiwizowanych witryn. W końcowej części swojej wypowiedzi przedstawił on wyzwania z jakimi musiało się zmierzyć British Library. Wśród nich znalazło się określenie częstotliwości przeprowadzenia, jej

zakresu oraz kwestie dostępu do zgromadzonych zasobów oraz przygotowanie interfejsu przyjaznego dla użytkowników. Podsumowując prelegent zwrócił jeszcze uwagę na rolę Internetu we współczesnym społeczeństwie.

Jako ostatnia głos zabrała Justyna Soliwoda, która szczegółowo opowiedziała o prowadzonym przez NAC projekcie *Archiwum Internetu*. Swoje wystąpienie rozpoczęła od zarysowania kontekstu, w którym przedstawiła funkcje jakie pełnią zasoby internetowe w różnych sferach życia człowieka, a także cechy, które mogą stanowić argument za ich zabezpieczaniem. Zasygnalizowane zostały również rozważania nad przypisaniem witryn internetowych do publikacji bądź materiałów archiwalnych. W następnej części prezentacji jej autorka skupiła się na polskim projekcie archiwizacji. Rozpoczęła ją od przedstawienia normatywów, które zobligowały NAC do podjęcia takiej inicjatywy, następnie zaś poszczególne elementy i etapy tego przedsięwzięcia. Wśród działania podejmowanych od 2009 r. w ramach tego projektu znalazły się wybór i przygotowanie odpowiedniego oprogramowania, przygotowanie serwisu internetowego za pomocą którego udostępniono zgromadzone zbiory oraz przeprowadzenie samej archiwizacji. Swoim zakresem objęła ono witryny internetowe wybranych urzędów centralnych, NDAP oraz archiwów państwowych oraz Muzeum Narodowego w Krakowie i przeprowadzana była co sześć miesięcy oraz w przypadku ważnych wydarzeń. W zakończeniu wystąpienia Soliwoda przedstawiła trudności z jakimi napotkano się w trakcie trwania polskiego *Archiwum Internetu*, które podzielić można na dwie kategorie: pierwszą stanowią te natury technologicznej, natomiast drugą kwestie prawa autorskiego.

Wszystkim wystąpieniom towarzyszyła dyskusja i pytania do prelegentów, w których podejmowano kwestie często towarzyszące rozważaniom poświęconym problematyce archiwizacji Internetu. Poruszano w nich zarówno kwestie techniczne, te związane z komplikacjami prawnymi jak i ze zdefiniowaniem natury samych witryn i innych zasobów Internetowych, a także problemy merytoryczne tego zagadnienia. W czasie debaty wyłoniły się dwa stanowiska. Przedstawiciele pierwszego z nich uznali, iż przed podjęciem inicjatyw archiwizacji Internetu w Polsce należy przygotować pod nie naukowe podstawy, natomiast inni uważali, iż takie projekty należy rozpocząć jak najszybciej, ponieważ część zawartych w Sieci zasobów może bezpowrotnie przepaść. Spotkanie w pełni wypełniło postawione przed nim zadanie, pozwoliło zapoznać się uczestnikom z doświadczeniami przedstawicieli instytucji zajmujących się archiwizacją Webu oraz nawiązanie dyskusji nad rozpoczęciem podobnych projektów w Polsce. Należy mieć nadzieję, że przyczyni się ono do popularyzacji tej świeżej tematyki, a w dalszej perspektywie to podjęcia działań nad zabezpieczeniem zasobów znajdujących się w polskiej „części” Internetu.