

Aleksandra Tokarska  
Marcin Wilkowski  
Laboratorium Cyfrowe Humanistyki  
Uniwersytet Warszawski  
[m@wilkowski.org](mailto:m@wilkowski.org)

## Seminarium „Archiwizacja Internetu” w LaCH UW

**Słowa kluczowe:** archiwa webu; egzemplarz obowiązkowy; prawo autorskie

2 marca 2017 r. w siedzibie Laboratorium Cyfrowego Humanistyki UW (LaCh UW) odbyło się pierwsze seminarium dotyczące archiwizacji internetu. Spotkanie zainaugurowało nowy kierunek działań LaCH UW, których podstawą ma być historyczna refleksja o nowych mediach. Celem seminarium było także wzmocnienie dyskusji o archiwizacji polskiego webu oraz wypracowanie strategii dalszych działań.

W spotkaniu wzięło udział 20 osób, przedstawicielek i przedstawicieli następujących instytucji: Naczelnej Dyrekcji Archiwów Państwowych (NDAP), Biblioteki Narodowej (BN), Biblioteki Jagiellońskiej (BJ), Uniwersytetu Mikołaja Kopernika w Toruniu (UMK), Ministerstwa Kultury i Dziedzictwa Narodowego (MKiDN), Stowarzyszenia EBIB, Helsińskiej Fundacji Praw Człowieka (HFPC), Akademii Pedagogiki Specjalnej (APS), Politechniki Wrocławskiej (PWr), Archiwum Państwowego (AP) w Toruniu oraz Uniwersytetu Warszawskiego (UW). Dzięki dużej różnorodności grupy zagadnienie archiwizacji webu udało się zaprezentować z bardzo różnych perspektyw, wyznaczając wiele kierunków możliwych działań. Ujawniły się też od razu dziedzinowe podziały w zapotrzebowaniu na wiedzę i kompetencje w tym zakresie.

Pierwszy z prelegentów – Mateusz Brzeziński (UMK) – zaprezentował polskie fora internetowe, które mogą być źródłem informacji historycznych, wskazując na konieczność archiwizacji również tego typu serwisów. Zwrócił uwagę na ich strukturę i wynikające z niej problemy z archiwizowaniem. Następnie Magdalena Grodzińska (UW) przedstawiła format plików WARC oraz język Python jako popularne rozwiązania wykorzystywane w profesjonalnej archiwizacji stron WWW. Podczas dyskusji udało się wskazać także najważniejsze ograniczenia związane ze stosowaniem tych narzędzi. Dorota Głowacka (Wydział Prawa i Administracji Uniwersytetu Łódzkiego, HFPC) podniosła kwestię ochrony prywatności w kontekście internetowych archiwów mediów. Ważnym wątkiem, który powracał później wielokrotnie podczas seminarium okazał się brak narzędzi prawnych, regulujących jakiegokolwiek działania z zakresu *text and data mining*. Kolejne wystąpienie poświęcone było egzemplarzom obowiązkowym i zasobom internetu. Ewa Dąbrowska (UJ) przedstawiła obecnie obowiązujące przepisy oraz wynikające z nich trudności, a także dokonała krótkiego przeglądu rozwiązań stosowanych w innych krajach europejskich. Następnie głos zabrał Bartłomiej Konopa (UMK), który wskazał na konieczność rozwiązania nie tylko praktycznych, ale również teoretycznych zagadnień związanych z archiwizacją webu. Analizie poddany został status witryny internetowej, która mogłaby być uznawana zarówno za dokumentację, jak i za publikację. Autor zwrócił uwagę na wynikające z tego rozróżnienia kon-

sekwencje. Wystąpienie porównywało bibliotekarskie i archiwistyczne podejście do zagadnienia archiwizacji webu. Barbara Szczepańska (EBIB) zaprezentowała studium przypadku ataku hakerskiego na stronę Komisji Nadzoru Finansowego oraz konsekwencje braku systemowych, państwowych rozwiązań w zakresie archiwizacji internetu. Osobny referat poświęcony został wyzwaniom związanym z archiwizacją mediów społecznościowych. Aleksandra Tokarska (UW) zaprezentowała najważniejsze problemy oraz rekomendacje zawarte w raporcie *Preserving Social Media*, opublikowanym przez Digital Preservation Coalition w 2016 r.<sup>1</sup> Jako ostatnia głos zabrała Justyna Kęska (UMK), która zarysowała możliwości archiwizowania stron WWW w odniesieniu do obecnej sytuacji w prawie polskim. Zaprezentowane zostały również rozwiązania stosowane w innych systemach prawnych.

Przebieg seminarium został zarejestrowany, a nagranie wszystkich wystąpień oraz towarzyszących im dyskusji dostępne jest na YouTube (<https://youtu.be/63ieicTef4E2>).

Marcowe seminarium było pierwszym spotkaniem z planowanego cyklu. Kolejne odbędzie się tuż po opublikowaniu przez MKiDN raportu horyzontalnego dotyczącego inicjatyw archiwizacji internetu na świecie oraz statusu prawnego tego typu projektów. Planujemy także warsztaty nt. korzystania z archiwów webu oraz podstaw archiwizacji stron WWW. Kolejne warsztaty i wydarzenia organizowane będą we współpracy z różnymi instytucjami, a ich tematy dopasowane będą do potrzeb uczestników. Zachęcamy do kontaktu wszystkich, którym bliskie są tematy związane z archiwizacją webu.

W roku 2016 obchodziliśmy dwudziestolecie działań archiwizacji webu podejmowanych przez fundację Internet Archive. Nie jest to już zagadnienie nowe, przez 20 lat rozwinął się wokół niego bogaty zasób literatury naukowej, raportów, metod i narzędzi. Z powodzeniem funkcjonują także branżowe organizacje łączące badaczy i instytucje archiwizujące web lub pracujące na jego archiwach. Zależy nam na tym, aby marcowe seminarium w LaCH UW traktować jako wyraźny sygnał pilnego podjęcia tego tematu przez instytucje państwowe i naukowe, do których zadań należy pozyskiwanie, gromadzenie i udostępnianie historycznych zbiorów webu.

---

TOKARSKA, A., WILKOWSKI, M. Seminarium "Archiwizacja Internetu" w LaCH UW. *Biuletyn EBIB* [online]. 2017, nr 2 (172), Archiwizacja internetu. [Dostęp 20.05.2017]. ISSN 1507-7187. Dostępny w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/531>.

<sup>1</sup> THOMSON, S. D. *Preserving Social Media. DPC Technology Watch Report* [online]. Digital Preservation Coalition, 2016. [Dostęp 27.04.2017]. Dostępny w: <http://www.dpconline.org/docman/technology-watch-reports/1486-twr16-01/file>.

<sup>2</sup> Odesłanie do strony internetowej przedstawia wersję aktualną w dn. 10.05.2017 r.