

PROTOKÓŁ z XI posiedzenia Rady do Spraw Cyfryzacji, które odbyło się 25 czerwca 2021 roku, o godzinie 13:00 w formie wideokonferencji.

Archiwizacja zasobów Internetu – Pan Karol Krawczyk

Zagadnienie archiwizacji zasobów Internetu w Polsce pojawia się w przestrzeni publicznej od lat. Pojawia się bo jest to ważne zagadnienie ale jednocześnie systematycznie znika, bo jest pod wieloma względami trudne. Niewątpliwie zadanie to powinno być realizowane od mniej więcej dwudziestu lat ale nie było albo było realizowane szczątkowo. Dlatego teraz należy zastanowić się jak zaprojektować to dobrze, bazując na szeregu doświadczeń, również tych trudnych, państw które przez ten proces przeszły.

Archiwa Państwowe czując odpowiedzialność za funkcjonowanie instytucjonalne państwa polskiego i konieczność podejmowania trudnych tematów i mierzenia się z nimi chcą zainicjować działania dotyczące archiwizacji zasobów polskiego Internetu.

Archiwizacja zasobów Internetu to projekt, niezwykle istotny, ponieważ dotyczy rozwiązań związanych z zabezpieczeniem dokumentacji wytworzonej w Internecie przez administrację publiczną – czyli materiałów mających wartość historyczną dla odtworzenia obrazu funkcjonowania państwa, który powinien być zdefiniowany w trzech wymiarach.

Pierwszy wymiar to obszar legislacji. Nie ma obecnie do końca uregulowanych tych kwestii. Brak jest uregulowań w samej ustawie dotyczącej narodowego zasobu archiwalnego ¹, czyli podstawowej ustawie dla działalności archiwów państwowych. Nie tylko w tej ustawie te aspekty musiałyby być rozstrzygnięte, w związku z tym istnieje olbrzymi wymiar prac legislacyjnych.

Drugi obszar to obszar metodyczny, tzn. opracowanie metodycznych zagadnień na podstawie formalno-prawnych rozstrzygnięć, które będą pozwalały, aby gromadzony zasób archiwalny był uporządkowany, możliwy do opracowania, udostępnienia i bezpiecznego wieczystego przechowywania.

Trzecia płaszczyzna – technologiczna, czyli zagadnienia dotyczące technologii i typu narzędzi do pobierania danych, gromadzenia, zachowywania w sposób trwały oraz udostępniania, a także wszelkie kwestie związane z infrastrukturą niezbędną do obsłużenia tego procesu.

Pan K. Krawczyk wspomniał, że archiwizacja zasobów Internetu jest to przedsięwzięcie, które przekracza możliwości realizacyjne zarówno Naczelnej Dyrekcji Archiwów Państwowych jak i sieci Archiwów Państwowych. Nie jest możliwa jego realizacja bez zaproszenia i zaangażowania do współpracy różnych środowisk – bez zbudowania szerokiej współpracy i płaszczyzny projektowej dla tego przedsięwzięcia. Przeprowadzono już rozmowy zarówno z Biblioteką Narodową oraz z Ministerstwem Kultury, Dziedzictwa Narodowego i Sportu oraz m.in. przedstawicielami projektu *Kronik@* i *Otwarte Dane Plus* – oraz z innymi środowiskami,

¹ Ustawa z dnia 14 lipca 1983 r. o narodowym zasobie archiwalnym i archiwach (Dz.U. 2020, poz. 164)

które ewentualnie byłyby interesariuszami w tym projekcie i mogłyby włączyć się w jego realizację.

Pan K. Krawczyk zaproponował przygotowanie stanowiska Rady w tej sprawie oraz stworzenia zespołu ad hoc do zajęcia się ww. tematem. Pan Przewodniczący rozpoczął dyskusję w tej sprawie.

Zauważono, że temat jest dużo szerszy. Zasoby archiwalne digitalizowane są bardzo wolno, w sposób chaotyczny, fragmentaryczny oraz niejednorodny technologicznie.

Następnie pojawiły się słowa wsparcia tej inicjatywy i projektu. Polska jest jednym z nielicznych krajów, które jeszcze nie uruchomiły takiego przedsięwzięcia systemowego.

W toku dyskusji jeden z członków Rady zapytał o wagę archiwizacji zasobu Internetu oraz o jej metodykę.

Pan K. Krawczyk wskazał, że Naczelna Dyrekcja Archiwów Państwowych zakończyła wdrożenie systemu „Archiwum Dokumentu Elektronicznego”, który służy temu, aby zawartość wytwarzana w systemach EZD, trafiła już w formie elektronicznej do archiwów. Trudnym zagadnieniem stwarzającym wiele problemów jest obsługa licznych formatów plików.

Wspomniano, że pierwsze próby na poziomie rządowym podejścia do tematu archiwizacji zasobów Internetu były prowadzone w 2000 r. przez Ministerstwo Łączności przy okazji prac nad strategią „ePolska - Plan działań na rzecz rozwoju społeczeństwa informacyjnego w Polsce na lata 2001-2006”. Również kiedy w portalu *Poland.pl* publikowane informacje z czasem zniknęły, NASK stworzył prototyp (pilotaż) systemu służącego archiwizacji publikowanych treści. Jednak projekt ten nie został finalnie uwzględniony w strategii przygotowywanej przez Ministerstwo Łączności. Również Narodowe Archiwum Cyfrowe pilotażowo w latach 2009–2011 archiwizowało niektóre serwisy z domeny gov.pl.

Obecnie digitalizowane treści w ramach np. programu Kultura Cyfrowa są przekazywane w zależności od ich rodzaju do odpowiednich centrów kompetencji – np. materiały biblioteczne do Biblioteki Narodowej, a materiały archiwalne do Narodowego Archiwum Cyfrowego.

Jeden z członków Rady zauważył, że decydenci nie rozumieją celu projektu Archiwów Państwowych oraz wskazał na wielką trudność zapanowania nad zakresem tego projektu. Podkreślono, że najważniejszą kwestią jest decyzja, czy archiwizacja odnosi się do treści internetowych w rozumieniu portali, czy do dokumentów elektronicznych. Przy wszelkiego rodzaju kwestiach archiwizacji pojawia się pojęcie polityki zarządzania danymi zarchiwizowanymi. Oprócz tego pojawia się temat poufności danych oraz polityki dostępu do danych. Trzecia kwestia warta uwagi to doświadczenia na podstawie projektu *Kronik@* w odniesieniu do metadanych. Ze wzorca założeń przy projekcie *Kronik@* warto byłoby czerpać, ponieważ największym wyzwaniem jest duża ilość gestorów danych i każdy z nich będzie chciał swoje dane opisywać według swojego uznania.

W toku dyskusji zauważano, że najważniejszy jest zakres przedmiotu archiwizacji i jej częstotliwość. Wskazano, że problem archiwizacji jest złożony. Należy go podzielić na dwa zasadnicze problemy, czyli problem przechowywania i gromadzenia, dotyczący już posiadanych zarchiwizowanych informacji. Osobnym zagadnieniem jest gromadzenie i przechowywanie tych informacji w formie produktu finalnego jakim są strony internetowe.

Pan Przewodniczący zauważył że ważny jest cel projektu, z którego wyniknie jego zakres. Po rozmowach z zainteresowanymi partnerami może okazać się że celów jest wiele i mogą być one różne. Wtedy musi wkroczyć państwo, które powinno określić główny cel i zakres oraz być gotowe na związane z tym wydatki. Przy powołaniu zespołu ad hoc w tej kwestii pojawia się pytanie, z jakimi instytucjami należałoby prowadzić wstępne rozmowy dla rozeznania celu projektu.

Jeden z członków Rady zauważył, że jakość danych jest bardzo istotna. Standaryzowanie przedmiotu digitalizowanego oraz udostępnianie jest niezmiernie istotne na różnych poziomach.

Pan K. Krawczyk doprecyzował, że zwraca się do Rady z inicjatywą współpracy w projekcie, który Naczelna Dyrekcja Archiwów Państwowych chce zainicjować, tak aby został zrealizowany z sukcesem. Realizację należy zacząć małymi krokami od domeny administracji publicznej, a sztuczna Inteligencja wspierająca procesy klasyfikacji oraz opisu materiałów archiwalnych musi być sprawą niedalekiej przyszłości.

Pan Przewodniczący zwrócił się do członków Rady o zgłaszanie się do zespołu ad hoc ds. archiwizacji zasobów Internetu pod wodzą Pana K. Krawczyka i stworzenie wizji programu i problemów z tym związanych tak, aby było można ten temat na posiedzeniu Rady przedstawić. Kilku członków Rady zgłosiło się do prac w zespole.

[Dyskusja nad agendą konferencji \(współ\)organizowanych przez Radę](#)

Wykorzystując połączenie grup roboczych MC z zespołami Rady, Pan Przewodniczący zwrócił się do kierowników Zespołów Rady oraz członków Rady o pomysły, przemyślenia bądź kwestie do rozstrzygnięcia dotyczące wyjścia Rady w swoich działaniach poza formę spotkań zespołów i Rady. Zastanawiano się, czy z rozmów i przeglądu dokonań grup roboczych MC wynikała potrzeba wspólnych warsztatów między zespołami i grupą roboczą a przedstawicielami innych instytucji, nie tylko rządowych, ale także biznesowych, eksperckich, naukowych itd. Pojawiło się pytanie czy Rada powinna współorganizować np. seminaria eksperckie, naukowe, konferencje naukowe oraz prasowe. Pan Przewodniczący wspomniał także o propozycji Pana Ministra M. Zagórskiego, aby Rada współpracowała z innymi resortami.

Kierownik Zespołu Kompetencji cyfrowych zauważył, że warto spotkać się z przedstawicielami Ministerstwa Edukacji i Nauki, aby przedyskutować dokument Program Rozwoju Kompetencji Cyfrowych oraz inne dokumenty przygotowywane przez MEiN.

Pan Dyrektor J. Paziowski wskazał, że dostrzega potrzebę zarówno dyskusji z administracją tzn. powrotu do rozmów z szerszym gronem osób, które decydują o projektowaniu

rozwiązań dla resortów, ale także silną potrzebę rozmowy z rynkiem. Zasygnalizował, że w formule grupy roboczej jest przestrzeń na zaprezentowanie koncepcji, wysłuchanie dostawców i przekonanie ich przede wszystkim do tego, aby myślenie o architekturze rozwiązań oferowanych administracji było zgodne z uzgodnionymi pryncypiami. Formułę funkcjonowania należy wypracować. Pan Dyrektor podkreślił, że architektura informacyjna państwa jest związana z usługami cyfrowymi, tzn. wszystko co w architekturze powstaje, koncentruje się na ergonomicznej, efektywnej, taniej i bezpiecznej usłudze elektronicznej. Jest to kryterium dla architektury. Należy także rozważyć zależność między zespołami Architektury IT i Usług Cyfrowych.

[Wolne wnioski i dyskusja wewnętrzna](#)

Dyskutowano o formule kolejnych spotkań Rady oraz ustalono, że kolejne posiedzenie Rady po uzgodnieniach i ostatecznej decyzji Pana Ministra J. Cieszyńskiego odbędzie się stacjonarnie w KPRM.

Uczestnicy posiedzenia:

Członkowie Rady:

1. Konrad Ciesiołkiewicz
2. Andrzej Dulka
3. Agnieszka Gryszczyńska
4. Janusz Kosiński
5. Karol Krawczyk
6. Anna Beata Kwiatkowska
7. Mirosław Maj
8. Dariusz Milka
9. Aleksandra Musielak
10. Józef Orzeł - Przewodniczący
11. Bolesław Piasecki
12. Paweł Śniatała
13. Robert Trętowski
14. Mateusz Tykierko
15. Małgorzata Zakrzewska
16. Marcin Zarzecki

Zaproszeni goście:

17. Przemysław Sypniewski, ekspert Rady
18. Wiesław Paluszyński, ekspert Rady

Sekretariat Rady i pracownicy Kancelarii Prezesa Rady Ministrów:

19. Michał Przymusiński, Zastępca Dyrektora Departamentu Rozwoju Usług w KPRM,
20. Anna Michałowska, Naczelnik Wydziału, Departament Polityki Cyfrowej w KPRM,
21. Agnieszka Mituniewicz, Starszy Analityk Biznesowy, Departament Rozwoju Usług w KPRM,
22. Piotr Popardowski, Architekt IT, Departament Rozwoju Usług w KPRM,
23. Krzysztof Głomb, Pełnomocnik Ministra Cyfryzacji do spraw współpracy z administracją samorządową Rzeczypospolitej Polskiej,
24. Jacek Paziewski, Dyrektor Departamentu Architektury Informacyjnej Państwa w KPRM,

25. Monika Skrzyńska, Zastępca Dyrektora Departamentu Architektury Informacyjnej Państwa w KPRM,
26. Katarzyna Staromłyńska-Gójska, KRPM
27. Anna Supeł, KPRM