

WSPÓŁCZESNE TECHNOLOGIE PRZECHOWYWANIA INFORMACJI

Możliwości zastosowania sztucznej inteligencji i blockchain w działalności archiwalnej. Przegląd doświadczeń zagranicznych

Robert Stępień

Uniwersytet Marii Curie-Skłodowskiej w Lublinie / Maria Curie Skłodowska University in Lublin (Poland)
robert.stepien@mail.umcs.pl, ORCID [0000-0003-3453-4185](https://orcid.org/0000-0003-3453-4185)

STRESZCZENIE

W artykule przedstawiono możliwości zastosowania technologii sztucznej inteligencji oraz blockchain w działalności archiwalnej. Przedmiotem analizy były wyniki projektów badawczych realizowanych w ostatnich latach z udziałem zagranicznych archiwów. Tekst artykułu powstał w oparciu o przegląd anglojęzycznej literatury przedmiotu. Szczególną uwagę zwrócono na informacje zawarte w raportach i sprawozdaniach z badań poświęconych problematyce wykorzystania nowych technologii na gruncie archiwalnym. Przedstawione w artykule przykłady dowodzą, że sztuczna inteligencja i blockchain znajdują zastosowanie w różnych obszarach działalności współczesnych archiwów. Technologie na nich oparte mają potencjał, aby zautomatyzować procesy wartościowania i selekcji dokumentacji elektronicznej oraz wspierać identyfikację danych wrażliwych zawartych w e-dokumentach. Innym polem ich zastosowania są systemy wyszukiwania i udostępniania materiałów archiwalnych w postaci cyfrowej, jak też platformy do rozpoznawania tekstu i struktury dokumentów historycznych. Blockchain, czyli technologia rozproszonego rejestru i łańcuchów blokowych, pozwala zachować integralność i autentyczność obiektów cyfrowego dziedzictwa archiwalnego.

SŁOWA KLUCZOWE

archiwistyka,
informatyka,
sztuczna
inteligencja,
blockchain,
dokument
elektroniczny,
archiwa cyfrowe

Possibilities of using artificial intelligence and blockchain in archival activities. Review of international experiences

ABSTRACT

The article presents the possibilities of using artificial intelligence and blockchain technology in archival activity. The analysis covers results of research projects carried out in recent years with participation of foreign archives. The text of the article is based on a review of literature available in English. Particular attention was given to information contained in research reports devoted to application of new technologies in archival work. The examples presented in the article prove that artificial intelligence and blockchain are used in various areas of activity of modern archives. Technologies based on them have the potential to automate the processes of evaluation and selection of electronic documents and to support identification of sensitive data contained in e-documents. Another possible field of application for such technologies are the systems for searching and accessing archival materials

KEYWORDS

archival sciences,
informatics, artificial
intelligence,
blockchain,
electronic
document, digital
archives

in digital form, as well as platforms for recognition of text and structure of historical documents. Blockchain, or distributed ledger and blockchain technology, preserves the integrity and authenticity of digital archival heritage objects.

Cyfryzacja gospodarki i społeczeństwa jest jedną z najbardziej dynamicznych zmian naszych czasów. Za sprawą nowoczesnych technologii radykalnie zmieniają się warunki naszego życia i pracy. W niektórych przypadkach zachodzące przemiany mają wręcz rewolucyjny charakter, przynoszący zupełnie odmienne niż dotychczas wartości, podejścia i sposoby działania¹. Zdaniem ekspertów, w perspektywie 2030 r. takich gwałtownych i nieliniowych zmian otoczenia cywilizacyjnego możemy spodziewać się znacznie więcej². Cyfrowa dysrupcja przeorientuje wiele mechanizmów znanego nam świata, ale uwidoczni też deficyt kompetencji niezbędnych do sprawnego poruszania się w rzeczywistości XXI w.

Postęp technologiczny, choć przynosi wiele nowych wyzwań dla instytucji archiwów i zawodu archiwisty, to otwiera też niespotykane dotąd możliwości wykorzystania innowacyjnych rozwiązań, takich jak sztuczna inteligencja, blockchain czy Internet Rzeczy. Innowacyjność jako imperatyw działania implikuje potrzebę eksperymentowania z nowymi technologiami, dzięki którym możliwa staje się dziś realizacja podstawowej misji archiwów, jaką jest trwałe zachowanie dla potomności świadectw naszych czasów. Wyzwaniem jest nie tylko heterogeniczny i masowy charakter powstających współcześnie dokumentów, ale też łatwość, z jaką można wpływać na integralność ich cyfrowego zapisu oraz autentyczność pochodzenia. Niezwykle ważne w tym kontekście jest utrzymanie zaufania społecznego wobec archiwów jako instytucji chroniących i dostarczających informacji, której wiarygodności nie można zakwestionować. W dobie nowych technologii oznacza to przejście od autorytetu instytucjonalnego archiwum do przejrzystości praktyki archiwalnej, opartej na wiedzy i umiejętnościach sprawnego posługiwania się narzędziami cyfrowymi.

¹ J. Pieriegud, *Cyfryzacja gospodarki i społeczeństwa – wymiar globalny, europejski i krajowy*, [w:] *Cyfryzacja gospodarki i społeczeństwa. Szanse i wyzwania dla sektorów infrastrukturalnych*, red. J. Gajewski, W. Paprocki, J. Pieriegud, Gdańsk 2016, s. 11.

² K. Głomb et al., *Kompetencje cyfrowe w czasach cyfrowej dysrupcji. Studium wyzwań dla Polski w perspektywie roku 2030*, Warszawa 2019, <https://www.mwi.pl/uploads/filemanager/publikacje/Kompetencje--przysz%C5%82o%C5%9Bci--w--czasach--cyfrowej--dysrupcji--studium--2019%2C%20final%2C%207.02.2018.pdf> [dostęp: 30.05.2021].

Podstawowym celem artykułu jest przedstawienie możliwości zastosowania w działalności archiwalnej technologii opartych na sztucznej inteligencji (artificial intelligence – dalej: AI) oraz blockchain. Przedmiotem analizy są wyniki projektów badawczych realizowanych z udziałem zagranicznych instytucji, takich jak Archiwum Narodowe w Londynie (The National Archives), którego przedsięwzięcia od lat dostarczają unikalnej i specjalistycznej wiedzy z zakresu zagadnień wpływu nowych technologii na dziedzinę archiwalną. W przyszłości niektóre z opisanych tu rozwiązań znajdują być może zastosowanie w działalności polskich archiwów. Problemy współczesnej archiwistyki w wielu wymiarach i płaszczyznach wymagają współpracy międzynarodowej i wymiany doświadczeń, co jednoznacznie podkreślają dokumenty wydane m.in. przez Komisję Europejską oraz UNESCO³. Badania z zakresu archiwistyki cyfrowej i nowych technologii w archiwach uwzględniać powinny dorobek ośrodków zagranicznych. Może on bowiem wnieść nową perspektywę i stanowić czynnik rozwoju dla realizacji własnych przedsięwzięć. Warto w tym kontekście przypomnieć myśl Stanisława Nawrockiego, który opisując praktykę zagranicznych archiwów w zakresie użytkowania komputerów, zwracał uwagę, że „[...] nie wystarczy tylko przyglądać się temu, co robią inni, ale trzeba poczynić własne próby, szukać własnych zastosowań i gromadzić własne doświadczenia. Bez odpowiedniego doświadczenia nie jest bowiem możliwe rozwijanie danej dziedziny”⁴.

Artykuł powstał w oparciu o przegląd anglojęzycznej literatury przedmiotu, który wykazał, że omawiana problematyka cieszy się coraz większym zainteresowaniem badawczym, a liczba publikacji w tym zakresie od kilku lat systematycznie wzrasta. O ile większość opublikowanych dawniej prac miała charakter teoretyczny i przeglądowy, to obecnie dostrzec można coraz więcej wydawnictw prezentujących wyniki badań empirycznych. Literatura dotycząca obszaru podjętych badań została pozyskana poprzez analizę bazy abstraktowo-bibliometrycznej Web of Science Core Collection oraz platform oferujących dostęp do elektronicznej wersji czasopism takich wydawnictw jak Emerald, Taylor & Francis

³ *Raport o archiwach Unii po rozszerzeniu: pogłębiona współpraca archiwalna w Europie – plan działania*, oprac. przez Grupę Ekspertów Krajowych ds. Archiwów Państw Członkowskich Unii Europejskiej oraz Instytucji i Organów UE na zamówienie Rady Unii Europejskiej, red. nauk. A. Biernat, W. Stępnia, V. Urbaniak, tł. J. Szymańska, Warszawa 2012; *Zalecenia UNESCO w sprawie zachowania i dostępu do dziedzictwa dokumentacyjnego, w tym dziedzictwa cyfrowego (2015)*, red. T. Komorowski et al., Warszawa 2016.

⁴ S. Nawrocki, *Doświadczenia zagraniczne w zakresie stosowania komputera w archiwach*, „Biuletyn Informatyka i Archiwa” 1978, nr 11, s. 31.

Online oraz Springer. Podstawową przesłanką takiego wyboru źródeł bibliograficznych był zamiar skoncentrowania się na publikacjach wydawanych w renomowanych czasopismach i materiałach konferencyjnych z zakresu nauk ścisłych, społecznych i humanistycznych⁵. Jako główne kryterium doboru literatury przyjęto aspekt praktycznej wiedzy na temat zastosowania sztucznej inteligencji i blockchain w różnych obszarach działalności archiwalnej. Szczególnie interesujące okazały się artykuły opublikowane na łamach czasopism: „Archives and Manuscripts”, „Records Management Journal” oraz „Journal of Documentation”⁶. Podczas pracy nad tekstem analizie poddano też raporty i sprawozdania omawiające koncepcje, metody i narzędzia oraz rezultaty projektów, w których badano potencjał wykorzystania nowych technologii na gruncie archiwalnym. Zwrócono również uwagę na dokumenty określające strategię cyfrowej transformacji, jaką przechodzi The National Archives. Przedstawiono w nich wizję archiwum cyfrowego drugiej generacji, które wprowadza innowacje technologiczne, by na nowo odkryć rolę instytucji archiwum w erze cyfrowej.

Problematyka sztucznej inteligencji i blockchain nie cieszyła się dotąd większym zainteresowaniem środowiska polskich archiwistów, choć w programach konferencji naukowych obecne były referaty poświęcone tym zagadnieniom⁷. W dyskursie naukowym coraz więcej uwagi poświęca się za to kwestii zachowa-

⁵ Dostęp do wskazanych platform internetowych uzyskano w ramach sieci uniwersyteckiej Uniwersytetu Marii Curie-Skłodowskiej. W trakcie przeszukiwania wybranych baz użyto mechanizmów filtrowania treści, sprawdzając występowanie w tytułach, abstraktach i słowach kluczowych kilku zadanych wyrażen, m.in. „artificial intelligence”, „blockchain” oraz „machine learning”. Aby zawęzić wyniki wyszukiwania do artykułów związanych z archiwistyką wykorzystano operatory logiczne oparte na algebrze Boole’a, weryfikując występowanie wskazanych wyrażen w powiązaniu z terminami: „archives”, „archival science”, „archival processing”, „records management”. W kolejnym etapie dokonano wyboru artykułów do dalszej pracy, w oparciu o analizę abstraktów, porządkując je według obszarów tematycznych. Uwagę zwrócono na publikacje o charakterze praktyczno-wdrożeniowym, w których przedstawiono przykłady wykorzystania omawianych technologii w dziedzinie archiwalnej.

⁶ Przykładowo, dwa numery specjalne czasopisma „Records Management Journal” z 2020 r. zostały poświęcone zastosowaniu sztucznej inteligencji oraz blockchain w działalności archiwalnej i zarządzaniu dokumentacją, zob. Emerald Insight. Records Management Journal, <https://www.emerald.com/insight/publication/issn/0956-5698> [dostęp: 30.05.2021].

⁷ Na X Zjeździe Internetowego Forum Archiwalnego w 2016 r. referat pt. „Machine Learning i big data w pracy archiwisty” wygłosił Grzegorz Gałęzowski. Natomiast na konferencji „Dokument elektroniczny w archiwum. Metody postępowania, dobre praktyki i techniki zarządzania” w 2017 r. wystąpienie na temat możliwości zastosowania blockchain w procesach archiwizacji zasobów cyfrowych przedstawił Hrvoje Stančić. Zob.: YouTube. Dokument elektroniczny w archiwum – dzień I, <https://www.youtube.com/watch?v=9R3EPGZvkUY&t=23720s> (od 3:24:12) [dostęp: 30.05.2021].

nia cyfrowego dziedzictwa archiwalnego, w tym koncepcji archiwum cyfrowego. Warto w tym miejscu przywołać pracę Katarzyny Peplowskiej, w której przedstawiono organizację i funkcjonowanie archiwów cyfrowych w wybranych krajach, a jeden z jej rozdziałów omawia doświadczenia brytyjskiego Archiwum Narodowego w zakresie procesów archiwizacji dokumentacji elektronicznej⁸.

Sztuczna inteligencja od lat postrzegana jest jako jedna z technologii mających potencjał, aby zrewolucjonizować funkcjonowanie światowej gospodarki. Niektóre z jej rozwiązań znalazły zastosowanie w takich sektorach jak przemysł, transport, energetyka, handel czy opieka zdrowotna⁹. Technologia AI w coraz szerszym zakresie wspomaga działalność podmiotów sektora administracji publicznej, w tym rozwój systemów i usług e-government¹⁰. Wiele państw i organizacji międzynarodowych dostrzega szanse i wyzwania, jakie niesie ze sobą sztuczna inteligencja, czego wyrazem są działania służące opracowaniu strategii i polityki rozwoju AI oraz wydanie przepisów wspierających innowacje, a przy tym ustalających standardy etyczne korzystania z nowych technologii¹¹. Siłą napędową obecnego rozwoju AI są postępy techniki w zakresie mocy obliczeniowej komputerów i algorytmów oraz dostęp do dużych zbiorów danych.

Termin „sztuczna inteligencja”, choć znany jest od lat 50. XX w. nie posiada jednej, powszechnie uznanej definicji. Zwykle określa się nim zdolność maszyn do wykazywania ludzkich cech i umiejętności w zakresie uczenia się, podejmowania decyzji i rozwiązywania problemów. W informatyce AI oznacza tworzenie

⁸ K. Peplowska, *Archiwa cyfrowe w wybranych krajach europejskich, USA i Australii*, Biblioteka Zarządcy Dokumentacji, t. 8, Toruń 2017, s. 25–54.

⁹ M. Torczyńska, *Sztuczna inteligencja i jej społeczno-kulturowe implikacje w codziennym życiu*, „Kultura i Historia” 2019, nr 36, s. 106–126.

¹⁰ K. Izdebski, *Algorytmy w procesie podejmowania decyzji urzędowych*, „IT w Administracji” 2019, nr 9 (142), s. 24–28.

¹¹ W oficjalnych dokumentach wydanych przez instytucje i organy Unii Europejskiej sztuczna inteligencja jest postrzegana jako centralny element cyfrowej transformacji systemu społeczno-gospodarczego. W lipcu 2020 r. Komisja Europejska opublikowała raport zawierający podsumowanie konsultacji publicznych przeprowadzonych w sprawie Białej Księgi dotyczącej sztucznej inteligencji. 20 października 2020 r. Parlament Europejski przyjął trzy rezolucje dotyczące etyki, odpowiedzialności cywilnej oraz własności intelektualnej w odniesieniu do AI. W grudniu 2020 r. Rada Ministrów polskiego rządu przyjęła dokument w sprawie ustanowienia polityki rozwoju sztucznej inteligencji w Polsce. Zob.: Serwis Rzeczypospolitej Polskiej. Droga do polskiej strategii AI, <https://www.gov.pl/web/cyfryzacja/droga-do-polskiej-strategii-ai> [dostęp: 30.05.2021]; European Commission. A European approach to artificial intelligence, <https://ec.europa.eu/digital-single-market/en/artificial-intelligence> [dostęp: 30.05.2021].

programów i modeli symulujących zachowania wynikające z inteligencji naturalnej. Co ważne, sztuczna inteligencja jest pojęciem nadrzędnym dla szeregu działań, takich jak uczenie maszynowe, sieci neuronowe czy przetwarzanie języka naturalnego, które stanowią obszar praktycznego zastosowania i realizacji koncepcji AI. W dziedzinie archiwalnej coraz częściej stosowane są technologie uczenia maszynowego (machine learning). Pojęcie to oznacza algorytmy i systemy usprawniające swoją wiedzę i wyniki wraz ze zdobywanym doświadczeniem¹². Na podstawie danych dostarczanych do systemu algorytmy uczą się pewnych prawidłowości, aby wykonywać te same lub podobne zadania bardziej efektywnie.

Kilka projektów badawczych realizowanych w ostatnich latach z udziałem zagranicznych archiwów wykazało, że sztuczna inteligencja ma potencjał, aby zautomatyzować procesy wartościowania i selekcji archiwalnej oraz wspierać identyfikację danych wrażliwych zawartych w dokumentacji elektronicznej. W projektach tych wykorzystano oprogramowanie e-Discovery¹³ z obszaru informatyki śledczej, które pozwala na szybkie i skuteczne odnajdywanie informacji w dużych, nieustrukturyzowanych zbiorach danych, gwarantując przy tym ich wartość dowodową. Narzędzia e-Discovery używają funkcji kodowania predykcyjnego, która umożliwia automatyczną klasyfikację dokumentów na podstawie analizy statystycznej i uczenia maszynowego¹⁴. Pierwsze doświadczenia w tym zakresie przyniosły przedsięwzięcia realizowane przez archiwa australijskie i brytyjskie.

Australia jest jednym z państw najbardziej zaawansowanych na świecie pod względem wykorzystania nowych technologii w administracji publicznej¹⁵. Australijskie osiągnięcia w dziedzinie archiwistyki i zarządzania dokumentacją,

¹² P. Flach, *Machine Learning. The Art and Science of Algorithms that Make Sense of Data*, Cambridge 2012, s. 3.

¹³ Pojęcie e-Discovery (electronic discovery) pochodzi z amerykańskiego systemu prawnego i oznacza zbiór procesów i rozwiązań związanych z informatyką śledczą służących pozyskiwaniu, zabezpieczeniu i analizowaniu dowodów elektronicznych (środków dowodowych bazujących na elektronicznie zapisanej informacji) na potrzeby postępowania sądowego. W Wielkiej Brytanii proces ten jest znany jako e-Disclosure. Zob.: D.W. Oard et al., *Evaluation of Information Retrieval for E-discovery*, „Artificial Intelligence and Law” 2010, t. 18, s. 348.

¹⁴ W.M. Hampton, *Predictive Coding: It's Here to Stay*, „E-Discovery Bulletin” 2014, June/July, https://files.skadden.com/sites/default/files/publications/flit_junejuly14_ediscoverybulletin.pdf [dostęp: 30.05.2021].

¹⁵ Świadczy o tym m.in. wysoka pozycja Australii w rankingu „E-Government Development Index”, który ocenia poziom rozwoju e-administracji w państwach członkowskich Organizacji Narodów Zjednoczonych. Zob.: UN E-government Survey 2020, <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2020> [dostęp: 30.05.2021].

choćby te związane z modelem *continuum* dokumentacji oraz rozwojem archiwów cyfrowych, znajdują uznanie na arenie międzynarodowej¹⁶. Archiwa z antypodów były też jednymi z pierwszych, które podjęły inicjatywy w zakresie wykorzystania AI w działalności archiwalnej i dokumentacyjnej. Doświadczenia kilku z tych przedsięwzięć zostały opisane w literaturze przedmiotu¹⁷. Jako przykład posłuży projekt realizowany w latach 2017–2020 przez archiwum państwowe stanu Wiktorii (Public Record Office Victoria), którego celem było zbadanie możliwości wykorzystania narzędzi e-Discovery i uczenia maszynowego do przeprowadzenia analizy dużej serii wiadomości e-mail. Zgodnie z ustawą Public Records Act z 1973 r., powołującą do życia wspomniane archiwum, każdy dokument, niezależnie od formy i postaci, sporządzony lub otrzymany przez funkcjonariusza publicznego w trakcie wykonywania jego obowiązków, ma status akt publicznych i może zostać zakwalifikowany do trwałego zachowania¹⁸. Korespondencja przesyłana drogą e-mailową nie tylko ułatwia komunikację i wymianę informacji, ale dokumentuje też działania i decyzje administracyjne urzędników państwowych. Stanowi dowód w zakresie powierzonych im funkcji i nadanych uprawnień oraz wynikającej z tego odpowiedzialności. Ma zatem znaczenie jako źródło wykorzystywane do poszukiwania informacji publicznej, które może zostać uznane za warte zachowania z historycznego punktu widzenia.

Wielkość i nieustrukturyzowany charakter wiadomości e-mail stanowią duże wyzwanie w procesach zarządzania i archiwizowania zasobów poczty elektronicznej¹⁹. W przypadku administracji rządowej stanu Wiktorii ponad dwadzieścia lat rutynowego tworzenia kopii zapasowych e-maili doprowadziło do powstania zbioru liczącego ponad 67 tys. taśm LTO (Linear Tape Open) i 28 petabaj-

¹⁶ Zob. nową koncepcję „recordkeeping informatics” opracowaną przez australijskich badaczy z Uniwersytetu Monasha: G. Oliver et al., *Recordkeeping Informatics for a Networked Age*, Clayton, Wiktorii 2018.

¹⁷ G. Rolan et al., *More Human than Human? Artificial Intelligence in the Archive*, „Archives and Manuscripts” 2019, t. 47, nr 2, s. 179–203 (artykuł przedstawia sześć projektów w zakresie AI prowadzonych przez archiwa oraz instytucje rządowe w Australii).

¹⁸ Public Records Act 1973, No. 8418/1973, Authorised Version No. 041, S. 2(1) def. of public record.

¹⁹ A. Vellino, I. Alberts, *Assisting the Appraisal of E-mail Records with Automatic Classification*, „Records Management Journal” 2016, t. 26, nr 3, s. 293–294; J. Schneider et al., *Appraising, Processing, and Providing Access to Email in Contemporary Literary Archives*, „Archives and Manuscripts” 2019, t. 47, nr 3, s. 305–306.

tów danych²⁰. Instytucje państwowe zostały narażone w tej sytuacji nie tylko na zwiększone koszty przechowywania danych, ale również problem ograniczonego dostępu do nich, co mogło skutkować podważeniem zasady przejrzystości i jawności działania administracji publicznej. Aby pomóc agencjom rządowym w zmniejszeniu zaległości (backlog) w obsłudze poczty e-mail oraz „uwolnieniu” informacji zawartych w jej zasobach uruchomiono projekt mający na celu przetestowanie rozwiązań e-Discovery w zakresie automatycznej identyfikacji i klasyfikacji wiadomości mailowych. Program Nuix eDiscovery²¹ w ramach procesu deduplikacji danych pozwolił oznaczyć i wygenerować listę powtarzających się wiadomości e-mail. Stanowiły one aż 43% badanego zbioru danych. Program umożliwił też przypisanie informacji kontekstowych (metadanych) do każdej wiadomości, co pozwoliło pogrupować je według obszarów działalności instytucjonalnej oraz nadać priorytety ułatwiające podejmowanie decyzji względem ich długotrwałego zachowania. Aby zidentyfikować e-maile nieposiadające statusu akt publicznych, program dokonał analizy słów kluczowych i domen internetowych zawartych w nagłówkach wiadomości. Projekt dowiódł również, że format NSF (Notes Storage Facility) poczty elektronicznej Lotus Notes IBM, której od końca lat 90. XX w. używa administracja stanu Wiktorja jako głównego środka komunikacji wewnętrznej i zewnętrznej, nie jest odpowiednim formatem w perspektywie celów i założeń długoterminowej archiwizacji cyfrowej.

Projekty badawcze z wykorzystaniem rozwiązań e-Discovery prowadziło też Archiwum Narodowe w Londynie. Ich celem było zbadanie możliwości zastosowania e-Discovery w ramach procesu TAR (Technology Assisted Review), który wykorzystuje algorytmy komputerowe do przeglądu i klasyfikowania dokumentów na podstawie analizy słów kluczowych i metadanych, z użyciem modeli statystycznych i technik głębokiego uczenia. W 2016 r. archiwiści brytyjscy opublikowali raport, który stwierdzał, że narzędzia e-Discovery i TAR mogą wspomóc aktotwórców podczas kwalifikacji, selekcji i oceny wrażliwości²² w ramach pro-

²⁰ Public Office Record Victoria. Email appraisal, disposal and preservation project, <https://prov.vic.gov.au/recordkeeping-government/research-projects/email-appraisal-disposal-preservation-project> [dostęp: 30.05.2021].

²¹ Nuix to australijska firma technologiczna, która oferuje oprogramowanie do analityki śledczej oraz kompleksowego przetwarzania i analizy nieustrukturyzowanych danych. Zob.: Nuix, <https://www.nuix.com/> [dostęp: 30.05.2021].

²² Sensitivity review – przegląd dokumentacji pod kątem występowania w niej danych szczególnie chronionych, takich jak dane osobowe czy klauzule tajności. Większość brytyjskich departamentów rządowych nie wypracowała dotąd skutecznych metod wykrywania danych wrażliwych zawartych w dokumentacji elektronicznej. Jest to proces czasochłonny, ale bez niego archiwa

cedury przekazywania dokumentacji elektronicznej do Archiwum Narodowego²³. Dodatkowo wsparciem tym mogłaby zostać objęta obsługa wniosków o dostęp do informacji na podstawie ustawy Freedom of Information Act oraz zapewnienie tzw. ciągłości cyfrowej (digital continuity) i bezpieczeństwa informacji. W końcowej części raportu omówiono dalsze kierunki działań, jakie archiwum zamierza realizować w obszarze wskazanych technologii, aby wspierać sprawność zarządzania zasobami cyfrowymi przez instytucje państwowe. Pracownicy The National Archives przewidują, że już w najbliższych latach dokumentacja elektroniczna rządu brytyjskiego będzie kwalifikowana do trwałego zachowania w drodze uczenia maszynowego, i że jest to przyszłość archiwów cyfrowych²⁴.

Omówione przykłady zastosowania narzędzi e-Discovery i TAR pokazują, że technologie uczenia maszynowego stanowią realne wsparcie dla procesów zarządzania dokumentacją elektroniczną. Oczywiście rozwiązania te nie gwarantują działania w sposób całkowicie zautomatyzowany. Na każdym etapie pracy wymagany jest udział człowieka. Projekty badawcze dostarczyły jednak obiecujących wyników, które mogą zostać wykorzystane w procesach kwalifikacji i selekcji archiwalnej oraz ocenie „wrażliwości” danych zawartych w e-dokumentach, czy też zarządzaniu nadmiarem wiadomości poczty elektronicznej.

Innym polem zastosowania rozwiązań AI są systemy wyszukiwania i udostępniania materiałów archiwalnych w postaci cyfrowej. Rozwój Internetu i towarzyszących mu technologii sprawia, że zmienia się sposób, w jaki uzyskujemy dostęp do informacji w przestrzeni wirtualnej. Jedną z nowych form prezentacji

historyczne, w świetle obowiązującego prawa, nie będą mogły udostępniać publicznie archiwaliów w postaci cyfrowej. Z tego powodu coraz większym zainteresowaniem cieszy się oprogramowanie, które automatycznie wyszukuje i identyfikuje dane sensytywne zawarte w dokumentacji. Zob.: G. McDonald, C. Macdonald, I. Ounis, *How Sensitivity Classification Effectiveness Impacts Reviewers in Technology-Assisted Sensitivity Review*, [w:] *CHIIR '19: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, Glasgow 2019, s. 337–341, <http://eprints.gla.ac.uk/174446/7/174446.pdf> [dostęp: 30.05.2021]; The National Archives, *The Digital Landscape in Government 2014–15*, Business Intelligence Review, Published: February 2016, s. 24, <https://www.nationalarchives.gov.uk/documents/digital-landscape-in-government-2014-15.pdf> [dostęp: 30.05.2021].

²³ The National Archives, *The Application of Technology-Assisted Review to born-digital records transfer, Inquiries and beyond*, Research Report, Published: February 2016, <https://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf> [dostęp: 30.05.2021].

²⁴ Rozmowa z Johnem Sheridanem, dyrektorem ds. cyfrowych w The National Archives, na temat uczenia maszynowego i jego możliwych zastosowań, zob.: Recordkeeping Roundtable. Recordkeeping Roundcasts Episode 3: Machine learning, <https://rkroundtable.org/2018/08/19/recordkeeping-roundcasts-episode-3-machine-learning/> [dostęp: 30.05.2021].

zawartości baz danych są tzw. bogate interfejsy (generous interfaces), które pozwalają użytkownikowi poznać strukturę zgromadzonych zbiorów przed przystąpieniem do szczegółowych poszukiwań²⁵. Zamiast czasochłonnej eksploracji danych w oknie wyszukiwarki, odbiorca, dzięki dobrze zaprojektowanemu interfejsowi, poprzez odpowiednio zestawione ze sobą miniatury obrazów (graficzną reprezentacją), w intuicyjny sposób może oszacować wielkość i różnorodność przeglądanej kolekcji. Takie połączenie danych pozwala odkrywać nowe obszary badawcze i doświadczać zasobów cyfrowego dziedzictwa w pełniejszy sposób²⁶. Podstawowym elementem tego rozwiązania jest technologia rozpoznawania obrazów (computer vision).

Jednym z przedsięwzięć z tego obszaru jest projekt „Deep Discoveries”, finansowany z grantu brytyjskiej Rady Badawczej do spraw Sztuki i Humanistyki²⁷, jako część programu „Towards a National Collection: Opening UK Heritage to the World”²⁸. W ramach projektu prowadzone są badania nad zastosowaniem funkcji wyszukiwania wizualnego oraz metod głębokiego uczenia do zwiększenia możliwości odkrywania dziedzictwa cyfrowego w przestrzeni wirtualnej. Podstawowym celem projektu jest uruchomienie prototypowej platformy umożliwiającej wyszukiwanie i łączenie obrazów między różnymi cyfrowymi kolekcjami na podstawie właściwości estetycznych obiektów, takich jak wzór, styl i motywy wizualne, a nie lista słów kluczowych. Ten sposób wyszukiwania stwarza możliwość odkrycia nieprzewidzianych powiązań w cyfrowych zbiorach różnych instytucji

²⁵ M. Whitelaw, *Generous Interfaces for Digital Cultural Collections*, „Digital Humanities Quarterly” 2015, t. 9, nr 1, <http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html> [dostęp: 30.05.2021].

²⁶ R. Bomba, *Bazo-danowe interfejsy. Projektowanie interakcji z dużymi zasobami danych kulturowych*, „Kultura Popularna” 2017, nr 1, s. 67–68.

²⁷ A. Knapieńska, *Soft power czy relikw przeszłości. Publiczne finansowanie humanistyki w Stanach Zjednoczonych Ameryki i Wielkiej Brytanii*, [w:] *Systemy publicznego finansowania nauki w ujęciu międzynarodowym*, Warszawa 2016, s. 90–92.

²⁸ Założenia projektu „Deep Discoveries” zostały opisane na stronie: Towards a National Collection, <https://www.nationalcollection.org.uk/> [dostęp: 30.05.2021] oraz w raporcie okresowym: L. Angelova et al., *Deep Discoveries. Interim Report. Foundation Projects*, 2020, <https://www.nationalcollection.org.uk/sites/default/files/2021-02/Deep%20Discoveries.pdf> [dostęp: 30.05.2021]. Głównymi uczestnikami projektu są cztery instytucje: Archiwum Narodowe w Londynie, Uniwersytet Surrey, Muzeum Wiktorii i Alberta w Londynie oraz Królewski Ogród Botaniczny w Edynburgu. W projekcie uczestniczą też trzy instytucje partnerskie – Muzeum Krajowego Wzornictwa i Architektury (Museum of Domestic Design and Architecture), archiwum pracowni projektowej Sanderson Design oraz firma włókiennicza Gainsborough Weaving Company, które reprezentują właścicieli, twórców i odbiorców zbiorów wizualnych.

na skalę krajową. Obecnie projekt koncentruje się na treściach o tematyce botanicznej, aby zweryfikować możliwości, jakie niesie ze sobą ta technologia²⁹.

Sztuczna inteligencja znajduje również zastosowanie przy rozpoznawaniu tekstu dokumentów historycznych. Jednym z projektów tego typu jest READ, finansowany w ramach programu ramowego Unii Europejskiej na rzecz badań naukowych i innowacji – „Horyzont 2020”³⁰. Częścią projektu jest platforma internetowa „Transkribus”, która oferuje narzędzia do automatycznej transkrypcji pisma odręcznego dokumentów historycznych. Platforma pozwala użytkownikom na wyszkolenie komputerowego modelu do rozpoznawania tekstu i struktury dokumentów zapisanych w różnych językach europejskich³¹. Model ten wykorzystuje technologię uczenia maszynowego do porównania istniejących w systemie wzorców językowych i stylu pisma z nowymi dokumentami. Archiwum Narodowe w Londynie w ostatnich latach testowało funkcjonalność tego typu oprogramowania przy rozpoznawaniu tekstu kopii testamentów zapisanych w księgach urzędników sądów kościelnych. Projekt przyniósł obiecujące rezultaty i spotkał się z dużym zainteresowaniem użytkowników wykorzystujących źródła testamentalne w różnych obszarach badawczych. Jednym z wyzwań projektu okazała się być właściwa segmentacja tekstu, czyli podział struktury treści na osobne odcinki i jednostki. Choć proces transkrypcji został w dużej mierze zautomatyzowany, to często wymaga interwencji człowieka. The National Archives w przyszłości planuje zmniejszyć pracochłonność tych działań oraz znaleźć sposoby możliwie najlepszego wykorzystania uzyskiwanych wyników³².

W literaturze przedmiotu odnaleźć można informacje o kolejnych projektach badawczych i narzędziach informatycznych z zakresu AI, które mogą usprawnić pracę archiwistów w różnych obszarach działalności archiwalnej. Dużo uwagi

²⁹ L. Angelova, L. Fulton, *Deep Discoveries: A new way of exploring and connecting digitised image collections*, The National Archives, 13.05.2020, <https://blog.nationalarchives.gov.uk/deep-discoveries-exploring-a-new-way-of-discovering-and-connecting-digitised-collections/> [dostęp: 30.05.2021].

³⁰ Strona internetowa projektu, zob.: READ (Recognition and Enrichment of Archival Documents), <https://readcoop.eu/> [dostęp: 30.05.2021].

³¹ G. Muehlberger et al., *Transforming Scholarship in the Archives Through Handwritten Text Recognition. Transkribus as a Case Study*, „Journal of Documentation” 2019, t. 75, nr 5, s. 954–976.

³² R. Dunley, *Machines reading the archive: handwritten text recognition software*, The National Archives, 19.03.2018, <https://blog.nationalarchives.gov.uk/machines-reading-the-archive-handwritten-text-recognition-software/> [dostęp: 30.05.2021]; F. Mackenzie, *How to teach a computer to read*, The National Archives, 13.05.2019, <https://blog.nationalarchives.gov.uk/how-to-teach-a-computer-to-read/> [dostęp: 30.05.2021].

poświęca się przetwarzaniu języka naturalnego (Natural Language Processing – dalej: NLP) – dziedzinie z pogranicza sztucznej inteligencji i językoznawstwa. Niektóre metody i funkcje NLP pozwalają m.in. automatycznie określać znaczenie poszczególnych fragmentów tekstu i kategoryzować jednostki nazewnicze do zdefiniowanych klas i typów (Named Entity Recognition), a także oznaczać wyrazy częściami mowy (Part-of-speech tagging), wskazywać ich semantyczną rolę w zdaniu (Semantic role labeling) oraz analizować wątki tematyczne, które opisują dokumenty tekstowe (Topic Modeling). W praktyce archiwów takie rozwiązania można wykorzystać zarówno w procesie tworzenia opisu archiwalnego, ekstrakcji i analizy metadanych dla zasobów digitalnych oraz transkrypcji skanowanych dokumentów³³. Znane są projekty zastosowania AI przez polskie archiwa do rozpoznawania i klasyfikowania obiektów w zbiorach audiowizualnych. Trwają bowiem prace nad wykorzystaniem algorytmów sieci neuronowych do opisywania fotografii przechowywanych w Narodowym Archiwum Cyfrowym poprzez nadanie im haseł przedmiotowych na podstawie klasyfikacji zdjęć³⁴. Takie narzędzia mogą znacznie przyspieszyć tworzenie metadanych na poziomie jednostki archiwalnej oraz pomóc archiwistom organizować duże zbiory fotografii, filmów czy nagrań dźwiękowych. W tym zakresie zastosowanie znajduje również technologia rozpoznawania twarzy³⁵.

Jednym z głównych wyzwań archiwizacji cyfrowej jest zachowanie autentyczności, integralności i wiarygodności dokumentów cyfrowych oraz ich metadanych. Spełnienie wymagań technicznych i prawno-organizacyjnych określonych w tym zakresie ma zasadnicze znaczenie dla archiwów, jako instytucji zaufania publicznego, których działalność koncentruje się na zapewnieniu dostępu do kompletnych i autentycznych materiałów archiwalnych. Jak udowodnić jednak, że dokument elektroniczny, który powstał 20 lat temu, jest dokładnie taki sam dzisiaj? Jak zademonstrować, że jedyne wprowadzone w nim zmiany były uzasadnione i nie wpłynęły na jego treść? Jak sprawić, aby w erze cyfrowej instytu-

³³ T. Hutchinson, *Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing*, „Records Management Journal” 2020, t. 30, nr 2, s. 155–174.

³⁴ M. Redzisz, *Piłsudski z czarnej skrzynki*, Sztuczna inteligencja, 23.03.2020, <https://www.sztucznainteligencja.org.pl/pilsudski-z-czarnej-skrzynki/> [dostęp: 30.05.2021].

³⁵ K. Banerjee, M. Anderson, *Batch Metadata Assignment to Archival Photograph Collection Using Facial Recognition Software*, „Code4Lib Journal” 2013, t. 21, <https://journal.code4lib.org/articles/8486> [dostęp: 30.05.2021]; K. Narojczyk, *Technologia rozpoznawania twarzy w warsztacie badawczym historyka*, [w:] *Multimedia a źródła historyczne w nauczaniu i badaniach*, red. M. Ausz, M. Szabaciuk, Lublin 2014, s. 273–294.

cje archiwalne nadal były postrzegane jako „zaufani opiekunowie” akt publicznych? Odpowiedzi na tak sformułowane pytania poszukiwali uczestnicy projektu ARCHANGEL, który w latach 2017–2019 był realizowany z udziałem Archiwum Narodowego w Londynie, Uniwersytetu Surrey oraz Brytyjskiego Instytutu Otwartych Danych³⁶.

Projekt badał funkcjonalność technologii rozproszonej księgi – DLT (Distributed Ledger Technology), powszechnie znanej jako blockchain, oraz sztucznej inteligencji, aby pomóc archiwom chronić integralność, autentyczność i trwałość archiwaliów w postaci cyfrowej. W ramach projektu uruchomiono pierwszy międzynarodowy łańcuch bloków, łączący pięć archiwów narodowych z Wielkiej Brytanii, USA, Australii, Norwegii i Estonii. Blockchain został wykorzystany do generowania za pomocą funkcji haszującej unikalnych skrótów dla dokumentów w postaci binarnej oraz rejestrowania tych danych w tzw. dozwolonym łańcuchu blokowym, do którego dostęp mają tylko uprawnione instytucje³⁷. Innymi słowy, aby zagwarantować niezmiennosc (meta)danych archiwalnych oraz umożliwić ich weryfikację w długim czasie, zostały one udostępnione wspomnianym archiwom poprzez platformę Ethereum³⁸ jako rozproszona baza danych. Każda z instytucji, pełniąc rolę walidatora, zapewnia w ten sposób wzajemną ochronę integralności zasobu innych archiwów, mając możliwość wyszukiwania i przeglądania wyników w całym łańcuchu bloków. W przypadku dokumentu, który został legalnie zmieniony, możliwe jest zarejestrowanie skrótów kodu użytego do wprowadzenia zmiany, a następnie – poprzez ścieżkę audytu – śledzenie, w jaki sposób dokument był edytowany. Dany dokument w każdej chwili można zweryfikować, porównując go z oryginalnym skrótem, w celu potwierdzenia integralności zapi-

³⁶ Strona internetowa projektu, zob.: ARCHANGEL, <https://www.archangel.ac.uk/> [dostęp: 30.05.2021]; Strona internetowa Uniwersytetu Surrey, jednego z wiodących ośrodków badawczych w dziedzinie technologii blockchain, zob.: Surrey Blockchain, <https://blockchain.surrey.ac.uk/index.html> [dostęp: 30.05.2021].

³⁷ Jedną z podstawowych metod dowodzenia autentyczności dokumentów cyfrowych jest tworzenie skrótu kryptograficznego, który reprezentuje zawartość tego dokumentu i staje się jego podpisem cyfrowym. Skróty te wrażliwe są na zmiany w strukturze binarnej pliku, stąd każda modyfikacja dokumentu powoduje zmianę jego skrótu. W ten sposób w dowolnym czasie można sprawdzić, czy zawartość dokumentu uległa zmianie. W projekcie, obok technologii blockchain, wykorzystano sztuczne sieci neuronowe przeszkolone do tworzenia tzw. cyfrowych odcisków (fingerprint), które pozwalają na wyliczenie skrótów dla danych wejściowych, za pomocą których można weryfikować integralność danych.

³⁸ Ethereum to jedna z platform o otwartym oprogramowaniu, wykorzystujących technologię blockchain. Umożliwia ona budowanie zdecentralizowanych aplikacji, takich jak systemy głosowania czy metody płatności.

su³⁹. Takie rozwiązanie w znaczący sposób utrudnia możliwość manipulowania informacją i gwarantuje, że integralność dokumentów historycznych pozostaje niezmieniona, tym samym, jak zauważył Alex Green, zapewnia, że żadna pojedyncza instytucja nie będzie mogła przepisać historii na nowo⁴⁰.

Technologia blockchain została zatem wykorzystana do kryptograficznego zapewnienia integralności dokumentów archiwalnych. Projekt weryfikował możliwość zabezpieczenia informacji cyfrowej przed nieuprawnioną modyfikacją oraz przejścia na tzw. technologiczne zapewnienie zaufania publicznego do archiwów cyfrowych. Badał też potencjał zabezpieczenia integralności i autentyczności plików video z wykorzystaniem funkcji haszowania treści (Temporal Content Hashes) oraz technik głębokiego uczenia. Dowiódł przy tym, że opracowany system pozwala wykrywać wszelkie manipulacje przestrzenne i czasowe w danych wizualnych i dźwiękowych w nagraniach video⁴¹. Wydaje się, że takie rozwiązania mogą w przyszłości znaleźć zastosowanie jako odpowiedź na zagrożenia związane z rozwojem technologii deepfake, wykorzystywanej do tworzenia spreparowanych treści multimedialnych⁴². Z perspektywy archiwów mogą natomiast wspierać procesy długoterminowej ochrony zasobów audiowizualnych, zabezpieczając je np. przed nieuprawnioną ingerencją, choćby w sekwencję klatek filmu czy też szybkość i czas trwania wideo, a także gwarantując prawidłowość i bezpieczeństwo działań w zakresie migracji, konwersji i kompresji danych multimedialnych.

Wciąż trwają prace nad zastosowaniem rozwiązań technologii blockchain w zakresie potrzeb długoterminowej archiwizacji cyfrowej. Współpraca ma zasadnicze znaczenie dla wartości oferowanej przez blockchain. W przyszłości planowane jest zaangażowanie w projekt ARCHANGEL uczestników spoza sektora archiwalnego, takich jak organizacje społeczne zajmujące się monitorowaniem

³⁹ J. Collomosse, *ARCHANGEL: Trusted Archives of Digital Public Documents*, [w:] *DocEng '18: Proceedings of the ACM Symposium on Document Engineering 2018*, Halifax, Nova Scotia, Canada, s. 1–4, <https://arxiv.org/pdf/1804.08342.pdf> [dostęp: 30.05.2021].

⁴⁰ A. Green, *Trustworthy technology: the future of digital archives*, The National Archives, 05.06.2018, <https://blog.nationalarchives.gov.uk/trustworthy-technology-future-digital-archives/> [dostęp: 30.05.2021].

⁴¹ T. Bui et al., *ARCHANGEL: Tamper-proofing Video Archives using Temporal Content Hashes on the Blockchain*, [w:] *Proceedings IEEE CVPR Workshop on Computer Vision AI and Blockchain*, 2019, <https://arxiv.org/pdf/1904.12059.pdf> [dostęp: 30.05.2021].

⁴² O. Wasiuta, S. Wasiuta, *Deepfake jako skomplikowana i fałszywa rzeczywistość*, „*Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate*” 2019, nr 9 (3), s. 19–30.

transparentności działania instytucji publicznych⁴³. Podejście mówiące o wspólnej i rozproszonej odpowiedzialności wydaje się mieć istotne znaczenie w obecnych czasach, kiedy coraz częściej kwestionuje się uprawnienia różnych podmiotów do wykonywania określonych działań. Jako instytucje pamięci i urzędy zaufania publicznego archiwa powinny być w stanie wykazać autentyczność i integralność przechowywanych dokumentów w sposób, którego nie można zakwestionować⁴⁴. ARCHANGEL stanowi przykład tego, jak archiwa wspólnymi siłami mogą chronić integralność dokumentów cyfrowych na dużą skalę oraz budować zaufanie społeczne w odniesieniu do swojej działalności w szybko zmieniającym się środowisku technologicznym.

Na końcu tych rozważań warto przyjrzeć się wizji archiwum cyfrowego drugiej generacji, jaka została przedstawiona w dokumentach strategicznych The National Archives. Należy zauważyć, że samo pojęcie archiwum cyfrowego nie jest jednoznaczne i wciąż ewoluuje⁴⁵. Wydana w 2015 r. encyklopedia archiwistyki przedstawia różne znaczenia terminu „digital archives”, który może oznaczać zarówno zbiory urodzonych cyfrowo dokumentów bądź strony internetowe oferujące dostęp do zdigitalizowanych materiałów archiwalnych, jak też instytucję/repozytorium trwale zabezpieczające dokumenty cyfrowe⁴⁶. W encyklopedii znalazł się również termin „digital repository”, któremu przypisuje się różne znaczenia w zależności od rodzaju zasobów, jakie zawiera repozytorium, czy też usług, jakie są przez nie oferowane⁴⁷. W literaturze przedmiotu występuje też pojęcie zaufanego/wiarygodnego repozytorium cyfrowego (trusted digital repository), którego misją jest zapewnienie niezawodnego i długoterminowego dostępu

⁴³ Open Data Institute, The National Archives, University of Surrey, *ARCHANGEL: guaranteeing the integrity of digital archives. Project Report*, 2019, https://theodi.org/wp-content/uploads/2019/08/2019-08-ARCHANGEL_-digital-artchive-integrity-Open-Data-Institute.pdf, s. 16 [dostęp: 30.05.2021].

⁴⁴ M. Bell et al., *Underscoring Archival Authenticity with Blockchain Technology*, „Insights” 2019, t. 32, s. 1–6, <https://insights.uksg.org/articles/10.1629/uksg.470/> [dostęp: 30.05.2021]; C. Findlay, *Participatory Cultures, Trust Technologies and Decentralisation: Innovation Opportunities for Recordkeeping*, „Archives and Manuscripts” 2017, t. 45, nr 3, s. 176–190.

⁴⁵ P. Perzyna, *Problemy terminologiczne wynikające z komputeryzacji archiwów*, [w:] *Komputeryzacja i digitalizacja w archiwach*, Symposia Archivistica, t. 2, red. R. Leśkiewicz, A. Żeglińska, Warszawa 2016, s. 53–54.

⁴⁶ K. Theimer, *Digital Archives*, [w:] *Encyclopedia of Archival Science*, red. L. Duranti, P.C. Franks, Lanham 2015, s. 157–159. Zob. rec. przygotowaną przez Annę Laszuk: eadem, *Encyclopedia of Archival Science*, edited by Luciana Duranti, Patricia C. Franks, Rowman & Littlefield, 2015, ss. X+454, „Archiwista Polski” 2017, nr 1–2, s. 93–104.

⁴⁷ G. Bak, *Digital Repository*, [w:] *Encyclopedia of Archival...*, s. 170–173.

do obiektów w postaci cyfrowej⁴⁸. Międzynarodowy standard OAIS pod pojęciem archiwum cyfrowego rozumie organizację ludzi i systemów, która przejęła odpowiedzialność za zachowanie informacji oraz jej udostępnianie określonej społeczności⁴⁹. Z kolei standard Trusted Digital Repositories preferuje termin „digital repository”, który używany jest w tym samym znaczeniu, czyli organizacji odpowiedzialnej za długoterminową ochronę zasobów cyfrowych⁵⁰. Należy zauważyć, że termin cyfrowe archiwum/repozytorium w takim ujęciu oznacza poszerzenie jego znaczenia z kwestii technicznego zabezpieczania (infrastruktury systemowej) na zaangażowania różnorodnych środków i zasobów organizacyjnych. Funkcjonowanie archiwum cyfrowego wymaga bowiem nie tylko sprzętu i oprogramowania, ale też ludzi, polityki i procedur działania oraz źródeł finansowania, czyli ogólnie zaplecza instytucjonalnego⁵¹.

W dokumencie „Archives Inspire”, opisującym plany i priorytety działania The National Archives na lata 2015–2019, szeroko rozumiane kwestie cyfrowe uznane zostały za jedno z największych wyzwań, przed jakimi stoi brytyjskie Archiwum Narodowe⁵². Strategia cyfrowa z 2017 r. szerzej omówiła istotę tych wyzwań, przedstawiając wizję archiwum cyfrowego drugiej generacji – takiego które jest „cyfrowe z instynktu i projektu” (digital by instinct and design)⁵³. Przenośnia ta określa jak głęboko w charakter instytucji (instytucjonalne DNA)

⁴⁸ A. Januszko-Szakiel, *Wiarygodność archiwów cyfrowych*, „Przegląd Biblioteczny” 2009, t. 77, z. 3, s. 325–347; eadem, *Archiwistyka cyfrowa. Długoterminowa ochrona dziedzictwa nauki i kultury*, Warszawa 2017, s. 63–82; Strona projektu InterPares Trust, <https://interparestrust.org/terminology/term/trustworthy%20digital%20repository> [dostęp: 30.05.2021].

⁴⁹ Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS). Recommended Practice, CCSDS 650.0-M-2, 2012, <https://public.ccsds.org/pubs/650x0m2.pdf> [dostęp: 30.05.2021].

⁵⁰ Research Libraries Group, Online Computer Library Center, Trusted Digital Repositories: Attributes and Responsibilities, 2002, <https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/repositories-1.pdf>, s. 3, [dostęp: 30.05.2021]; Online Computer Library Center, Center for Research Libraries, Trustworthy Repositories Audit & Certification: Criteria and Checklist, version 1.0, 2007, s. 75, https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf [dostęp: 30.05.2021]; Consultative Committee for Space Data Systems, Audit and Certification of Trustworthy Digital Repositories. Recommended Practice, CCSDS 652.0-M-1, 2011, s. 4, <https://public.ccsds.org/pubs/652x0m1.pdf> [dostęp: 30.05.2021].

⁵¹ R. Allan, *Virtual Research Environments. From Portals to Science Gateways*, Oksford 2009, s. 68.

⁵² The National Archives, Archives Inspire: The National Archives Plans and Priorities 2015–19, <https://www.nationalarchives.gov.uk/documents/archives-inspire-2015-19.pdf>, s. 2, 7, 12 [dostęp: 30.05.2021].

⁵³ The National Archives, Digital Strategy, <https://www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf> [dostęp: 30.05.2021].

wpisane powinno być myślenie, działanie i bycie cyfrowym. Zgodnie z tym podejściem doświadczenia cyfrowe uznać należy za integralny element działalności archiwum i jeden z podstawowych czynników kształtujących jego kulturę organizacyjną⁵⁴.

Obecnie Archiwum Narodowe w Londynie to w pełni funkcjonalne archiwum cyfrowe pierwszej generacji, wyposażone w infrastrukturę technologiczną spełniającą wymagania w zakresie długoterminowej ochrony zasobów cyfrowych o dużym wolumenie danych. Funkcjonalność archiwów cyfrowych w gruncie rzeczy opiera się dziś na zasadach postępowania z dokumentacją w postaci papierowej. Model działania The National Archives jako archiwum cyfrowego przyjmuje praktyki archiwalne stosowane względem akt papierowych i replikuje je przy użyciu technologii informatycznej. Przykładowo wartościowanie i selekcja dokumentacji elektronicznej przebiegają zgodnie z wytycznymi opracowanymi dla dokumentacji tradycyjnej, określonymi przez model cyklu życia zapisów. Przejmowane zaś do zasobu archiwum materiały archiwalne w postaci cyfrowej są opracowywane według ustalonych od dawna standardów opisu archiwalnego⁵⁵.

W strategii cyfrowej z 2017 r. przedstawiono kilka działań, które The National Archives zamierza podjąć w najbliższych latach, aby stać się archiwum cyfrowym drugiej generacji. Dokument mówi o potrzebie dokonania przeglądu i weryfikacji paradygmatów działalności archiwalnej w perspektywie zmieniającej się roli archiwum w erze cyfrowej oraz szans i zagrożeń, jakie niesie archiwizacja dokumentacji elektronicznej. W tym ujęciu archiwum cyfrowe szuka sposobów, aby zakłócić i dostosować ugruntowaną praktykę archiwalną do obecnych i przyszłych wyzwań trwałego zachowania źródeł historycznych w postaci cyfrowej. „Zakłócające” archiwum cyfrowe (disruptive digital archive) rozważa radykalną zmianę zasad i metod dotychczasowego działania. Identyfikuje innowacje i rozwiązania o przełomowym/rewolucyjnym charakterze (disruptive innovations)⁵⁶, takie jak

⁵⁴ J. Müller, *Digital Transformation at Media Archives: Ten Steps to Becoming Digital by Design*, „Journal of Digital Media Management” 2020, t. 8, nr 4, s. 321.

⁵⁵ The National Archives, *Digital Strategy...*, s. 6.

⁵⁶ Konceptję disruptive innovation spopularyzował Clayton M. Christensen w wydanej w 1997 r. książce: *The Innovator's Dilema: When New Technologies Cause Great Firms to Fail*, Harvard Business School Press, Boston 1997 (zob. *Przełomowe innowacje. Możliwości rozwoju czy zagrożenie dla przedsiębiorstwa*, przeł. H. Simbierowicz, Warszawa 2010). Zob. też: J.L. Bower, C.M. Christensen, *Disruptive Technologies: Catching the Wave*, „Harvard Business Review” 1995, t. 73, nr 1, s. 43–53; T. Zaręba, *Self aware networks – cechy techniczne i implikacje technologiczne*, [w:] *Cyfrizacja gospodarki...*, s. 75–77.

uczenie maszynowe czy blockchain, które dają szansę zwiększenia efektywności i sprawności działania instytucji⁵⁷. Celem The National Archives jest przejście z pierwszej do drugiej generacji archiwów cyfrowych, z wykorzystaniem wspomnianej filozofii myślenia i działania. W dokumencie określono zasady, których przyjęcie pozwoli brytyjskiemu archiwum przeprowadzić cyfrową transformację, polegającą na wprowadzeniu i efektywnym wykorzystaniu rozwiązań opartych na nowych technologiach oraz poszukiwaniu innowacji mogących zmienić dotychczasowy model działania. Wymaga to opracowania spójnej i przemyślanej strategii, uwzględniającej potencjał organizacji i posiadane przez nią zasoby.

Jaką funkcjonalność w świetle strategii wykazuje archiwum cyfrowe drugiej generacji? Takie archiwum posiada zdolność zachowania dziedzictwa archiwalnego w różnych typach i formatach zapisu cyfrowego. Obecnie The National Archives w głównej mierze gromadzi formaty plików dokumentów biurowych (np. Word i Excel), pliki audio i wideo, wiadomości poczty elektronicznej oraz media stanowiące połączenie różnych form przekazu informacji (np. strony internetowe). Archiwum dostrzega jednak potrzebę rozwinięcia zdolności zabezpieczenia m.in. strukturalnych baz danych, kodu komputerowego czy też modeli danych poddawanych działaniu sztucznych sieci neuronowych. Należy zauważyć, że wpływ technologii AI na praktykę archiwalną dotyczy nie tylko obszarów i metod ich bezpośredniego zastosowania w pracy archiwalnej. Sztuczna inteligencja coraz częściej staje się jednym z narzędzi wykorzystywanych przez władze państwowe w procesach podejmowania decyzji oraz środków usprawniających funkcjonowanie administracji publicznej. W dyskusjach na temat rządowego wykorzystania rozwiązań AI rozważa się potrzebę wyboru i zachowania w archiwach bieżących i historycznych algorytmów i modeli AI jako nowej formy zapisu, która pozwoli zapewnić długoterminową odpowiedzialność władz publicznych za swoją działalność⁵⁸. Jeżeli rząd wykorzystuje wyniki uczenia maszynowego do podejmowania decyzji, to taki zapis powinien podlegać ocenie jako źródło informacji o wartości dowodowej i historycznej. Dostrzega się przy tym rolę doradczą archi-

⁵⁷ E. Goudarouli, A. Sexton, J. Sheridan, *The Challenge of the Digital and the Future Archive: Through the Lens of The National Archives UK*, „Philosophy & Technology” 2019, t. 32, nr 1, s. 173–183.

⁵⁸ A. Le Sueur, *Robot Government: Automated Decision-Making and its Implications for Parliament*, [w:] *Parliament: Legislation and Accountability*, red. A. Horne, A. Le Sueur, Oksford 2016, s. 183–202; M. Kuziemski, G. Misuraca, *AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings*, „Telecommunications Policy” 2020, t. 44, nr 6, s. 1–13, <https://www.sciencedirect.com/journal/telecommunications-policy/vol/44/issue/6> [dostęp: 30.05.2021].

wistów wobec instytucji państwowych w zakresie korzystania z nowych technologii, nie tylko w celu usprawnienia przepływu pracy, ale też zapewnienia, że zachowane zostaną wszelkiego rodzaju świadectwa procesów decyzyjnych mających wpływ na funkcjonowanie państwa i społeczeństwa⁵⁹.

W świetle strategii archiwum powinno odgrywać aktywną rolę w procesach wdrażania systemów elektronicznego zarządzania dokumentacją w instytucjach publicznych, aby kwestie ochrony zasobów cyfrowych były rozpatrywane na możliwie wczesnym etapie. Archiwizacja według projektu (archiving by design) oznacza bowiem projektowanie systemów informatycznych wspierających proces pracy w taki sposób, aby od samego początku brana była pod uwagę długoterminowa ochrona i dostępność informacji⁶⁰. Rola zaufanego i wiarygodnego doradcy wymaga z kolei ciągłego rozwijania kompetencji w sferze technologicznej. Archiwum powinno wykazywać umiejętność dostosowywania się do środowisk, w jakich informacja cyfrowa jest wytwarzana i użytkowana, aby sprostać współczesnym i przyszłym wyzwaniom jej zachowania. Każda generacja technologii przynosi nowe wymagania w zakresie przechowywania i dostępu, wobec których archiwum cyfrowe jest zobowiązane dotrzymać kroku⁶¹.

*

Przedstawione w artykule przykłady projektów badawczych dowodzą, że sztuczna inteligencja oraz blockchain znajdują zastosowanie w różnych obszarach działalności archiwalnej. Trudno jednakże przewidzieć kierunki zmian i konsekwencje, jakie w dłuższym czasie technologie te przyniosą dla praktyki archiwal-

⁵⁹ A. Seles, A Brave New World: Artificial Intelligence and Archives (referat przedstawiony na konferencji „Archives Today and Tomorrow: Prospering as a Diversity of Records Dramatically Increase”, zorganizowanej w 2019 r. przez Archiwum Narodowe Japonii oraz oddział regionalny Międzynarodowej Rady Archiwów dla Azji Wschodniej – EASTICA), http://www.archives.go.jp/english/news/pdf/20191125_27e_01.pdf oraz http://www.archives.go.jp/english/news/pdf/20191125_27e_02.pdf [dostęp: 30.05.2021]. Anthea Seles od 2018 r. jest Sekretarzem Generalnym Międzynarodowej Rady Archiwów, a wcześniej pracowała w The National Archives w Londynie, gdzie była odpowiedzialna za nadzorowanie procesu przekazywania materiałów archiwalnych w postaci cyfrowej z departamentów rządowych do zasobu archiwum. Podczas pracy w The National Archives uczestniczyła w testowaniu aplikacji uczenia maszynowego mogących zautomatyzować procesy wartościowania i selekcji archiwalnej.

⁶⁰ E. Saaman, *Archiving by design*, National Archives of the Netherlands, <https://www.nationaalarchief.nl/en/archive/knowledge-base/archiving-by-design> [dostęp: 30.05.2021]; Zob. też: Wspólnotowy Serwis Informacyjny Badań i Rozwoju (CORDIS). Nowe podejście do przechowywania treści cyfrowych, <https://cordis.europa.eu/article/id/196579-a-new-approach-to-digital-content-preservation/pl> [dostęp: 30.05.2021].

⁶¹ The National Archives, *Digital Strategy...*, s. 4; S. Renade, *Access Technologies for the Disruptive Digital Archive*, [w:] *Archival Futures*, red. C. Brown, Londyn 2018, s. 79–98.

nej. Zgodnie z tzw. „prawem Amary” mamy tendencję do przeceniania wpływu technologii na krótką metę i niedoszacowywania jej efektów w długim okresie, co sugeruje, że oczekujemy zbyt dużo od innowacji na początku i zbyt mało po pewnym czasie. Stanisław Lem, którego setną rocznicę urodzin obchodzimy w 2021 r., w swojej twórczości przedstawiał prognozy na przyszłość związane z nowymi technologiami, w których przestrzegał też przed skutkami technicznego rozwoju cywilizacji, pisząc: „Każda bez wyjątku nowa technologia ma awers korzyści, i zarazem rewers nowych, nieznanych dotychczas bied”⁶². Technologie cyfrowe otwierają drzwi do świata nowych możliwości, ale przynoszą też niepewność i różnego rodzaju wyzwania. Pozwalają na osiągnięcie większej efektywności pracy, ale wymagają przy tym odpowiednich kompetencji. Współczesne archiwa powinny jednak wykorzystać szanse, jakie przynoszą kolejne fale innowacji technologicznych. Badać potencjał ich wykorzystania w dziedzinie archiwalnej oraz podejmować decyzje o możliwych zastosowaniach. Przyjęcie takiego podejścia sprawi, że będą postrzegane jako instytucje nowoczesne, poszukujące innowacyjności oraz świadome potrzeb technologicznych – własnych i otoczenia.

Źródła

- Consultative Committee for Space Data Systems, Audit and Certification of Trustworthy Digital Repositories. Recommended Practice, CCSDS 652.0-M-1, 2011, <https://public.ccsds.org/pubs/652x0m1.pdf> [dostęp: 30.05.2021].
- Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS). Recommended Practice, CCSDS 650.0-M-2, 2012, <https://public.ccsds.org/pubs/650x0m2.pdf> [dostęp: 30.05.2021].
- Online Computer Library Center, Center for Research Libraries, Trustworthy Repositories Audit & Certification: Criteria and Checklist, version 1.0, 2007, https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf [dostęp: 30.05.2021].
- Public Records Act 1973, No. 8418/1973, Authorised Version No. 041.
- Research Libraries Group, Online Computer Library Center, Trusted Digital Repositories: Attributes and Responsibilities, 2002, <https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/repositories-1.pdf> [dostęp: 30.05.2021].
- The National Archives, Archives Inspire: The National Archives Plans and Priorities 2015–19, <https://www.nationalarchives.gov.uk/documents/archives-inspire-2015-19.pdf> [dostęp: 30.05.2021].

⁶² S. Lem, *Bomba Megabitowa*, posł. J. Jarzębski, Kraków 1999, s. 7.

- The National Archives, Digital Strategy, <https://www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf> [dostęp: 30.05.2021].
- The National Archives, The Application of Technology-Assisted Review to born-digital records transfer, Inquiries and beyond, Research Report, Published: February 2016, <https://www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf> [dostęp: 30.05.2021].
- The National Archives, The Digital Landscape in Government 2014–15, Business Intelligence Review, Published: February 2016, <https://www.nationalarchives.gov.uk/documents/digital-landscape-in-government-2014-15.pdf> [dostęp: 30.05.2021].
- Zalecenia UNESCO w sprawie zachowania i dostępu do dziedzictwa dokumentacyjnego, w tym dziedzictwa cyfrowego (2015)*, red. T. Komorowski et al., Warszawa 2016.

Bibliografia

- Allan R., *Virtual Research Environments. From Portals to Science Gateways*, Oksford 2009.
- Angelova L. et al., *Deep Discoveries. Interim Report. Foundation Projects*, 2020, <https://www.nationalcollection.org.uk/sites/default/files/2021-02/Deep%20Discoveries.pdf> [dostęp: 30.05.2021].
- Angelova L., Fulton L., *Deep Discoveries: A new way of exploring and connecting digitised image collections*, The National Archives, 13.05.2020, <https://blog.nationalarchives.gov.uk/deep-discoveries-exploring-a-new-way-of-discovering-and-connecting-digitised-collections/> [dostęp: 30.05.2021].
- Bak G., *Digital Repository*, [w:] *Encyclopedia of Archival Science*, red. L. Duranti, P.C. Franks, Lanham 2015, s. 170–173.
- Banerjee K., Anderson M., *Batch Metadata Assignment to Archival Photograph Collection Using Facial Recognition Software*, „Code4Lib Journal” 2013, t. 21, <https://journal.code4lib.org/articles/8486> [dostęp: 30.05.2021].
- Bell M. et al., *Underscoring Archival Authenticity with Blockchain Technology*, „Insights” 2019, t. 32, <https://insights.uksg.org/articles/10.1629/uksg.470/> [dostęp: 30.05.2021].
- Bomba R., *Bazo-danowe interfejsy. Projektowanie interakcji z dużymi zasobami danych kulturowych*, „Kultura Popularna” 2017, nr 1, s. 64–73.
- Bower J.L., Christensen C.M., *Disruptive Technologies: Catching the Wave*, „Harvard Business Review” 1995, t. 73, nr 1, s. 43–53.
- Bui T. et al., *ARCHANGEL: Tamper-proofing Video Archives using Temporal Content Hashes on the Blockchain*, [w:] *Proceedings IEEE CVPR Workshop on Computer Vision AI and Blockchain*, 2019, <https://arxiv.org/pdf/1904.12059.pdf> [dostęp: 30.05.2021].

- Christensen C.M., *Przełomowe innowacje. Możliwości rozwoju czy zagrożenie dla przedsiębiorstwa*, przeł. H. Simbierowicz, Warszawa 2010.
- Christensen C.M., *The Innovator's Dilema: When New Technologies Cause Great Firms to Fail*, Harvard Business School Press, Boston 1997.
- Collomosse J., *ARCHANGEL: Trusted Archives of Digital Public Documents*, [w:] *DocEng '18: Proceedings of the ACM Symposium on Document Engineering 2018*, Halifax, Nova Scotia, Canada, <https://arxiv.org/pdf/1804.08342.pdf> [dostęp: 30.05.2021].
- Dunley R., *Machines reading the archive: handwritten text recognition software*, The National Archives, 19.03.2018, <https://blog.nationalarchives.gov.uk/machines-reading-the-archive-handwritten-text-recognition-software/> [dostęp: 30.05.2021].
- Findlay C., *Participatory Cultures, Trust Technologies and Decentralisation: Innovation Opportunities for Recordkeeping*, „Archives and Manuscripts” 2017, t. 45, nr 3, s. 176–190.
- Flach P., *Machine Learning. The Art and Science of Algorithms that Make Sense of Data*, Cambridge 2012.
- Goudarouli E., Sexton A., Sheridan J., *The Challenge of the Digital and the Future Archive: Through the Lens of The National Archives UK*, „Philosophy & Technology” 2019, t. 32, nr 1, s. 173–183.
- Głomb K. et al., *Kompetencje cyfrowe w czasach cyfrowej dysrupcji. Studium wyzwań dla Polski w perspektywie roku 2030*, Warszawa 2019, <https://www.mwi.pl/uploads/filemanager/publikacje/Kompetencje--przysz%C5%82o%C5%9Bci--w--czasach--cyfrowej--dysrupcji--studium--2019%2C%20final%2C%207.02.2018.pdf> [dostęp: 30.05.2021].
- Green A., *Trustworthy technology: the future of digital archives*, The National Archives, 05.06.2018, <https://blog.nationalarchives.gov.uk/trustworthy-technology-future-digital-archives/> [dostęp: 30.05.2021].
- Hampton W.M., *Predictive Coding: It's Here to Stay*, „E-Discovery Bulletin” June/July 2014, https://files.skadden.com/sites/%2Fdefault%2Ffiles%2Fpublications%2Fflit_junejuly14_ediscoverybulletin.pdf [dostęp: 30.05.2021].
- Hutchinson T., *Natural Language Processing and Machine Learning as Practical Toolsets for Archival Processing*, „Records Management Journal” 2020, t. 30, nr 2, s. 155–174.
- Izdebski K., *Algorytmy w procesie podejmowania decyzji urzędowych*, „IT w Administracji” 2019, nr 9 (142), s. 24–28.
- Januszko-Szakiel A., *Archiwistyka cyfrowa. Długoterminowa ochrona dziedzictwa nauki i kultury*, Warszawa 2017.
- Januszko-Szakiel A., *Wiarygodność archiwów cyfrowych*, „Przegląd Biblioteczny” 2009, t. 77, z. 3, s. 325–347.
- Knapieńska A., *Soft power czy relikw przeszłości. Publiczne finansowanie humanistyki w Stanach Zjednoczonych Ameryki i Wielkiej Brytanii*, [w:] *Systemy publicznego finansowania nauki w ujęciu międzynarodowym*, Warszawa 2016, s. 79–96.

- Kuziemski M., Misuraca G., *AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings*, „Telecommunications Policy” 2020, t. 44, nr 6, s. 1–13, <https://www.sciencedirect.com/journal/telecommunications-policy/vol/44/issue/6> [dostęp: 30.05.2021].
- Laszuk A. (rec.), *Encyclopedia of Archival Science*, edited by Luciana Duranti, Patricia C. Franks, Rowman & Littlefield, 2015, ss. X+454, „Archiwista Polski” 2017, nr 1–2, s. 93–104.
- LeSueur A., *Robot Government: Automated Decision-Making and its Implications for Parliament*, [w:] *Parliament: Legislation and Accountability*, red. A. Horne, A. Le Sueur, Oksford 2016, s. 183–202.
- Lem S., *Bomba Megabitowa*, posł. J. Jarzębski, Kraków 1999.
- Mackenzie F., *How to teach a computer to read*, The National Archives, 13.05.2019, <https://blog.nationalarchives.gov.uk/how-to-teach-a-computer-to-read/> [dostęp: 30.05.2021].
- McDonald G., Macdonald C., Ounis I., *How Sensitivity Classification Effectiveness Impacts Reviewers in Technology-Assisted Sensitivity Review*, [w:] *CHIIR '19: Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, Glasgow 2019, s. 337–341, <http://eprints.gla.ac.uk/174446/7/174446.pdf> [dostęp: 30.05.2021].
- Muehlberger G. et al., *Transforming Scholarship in the Archives Through Handwritten Text Recognition. Transkribus as a Case Study*, „Journal of Documentation” 2019, t. 75, nr 5, s. 954–976.
- Müller J., *Digital Transformation at Media Archives: Ten Steps to Becoming Digital by Design*, „Journal of Digital Media Management” 2020, t. 8, nr 4, s. 321–339.
- Naroczny K., *Technologia rozpoznawania twarzy w warsztacie badawczym historyka*, [w:] *Multimedia a źródła historyczne w nauczaniu i badaniach*, red. M. Ausz, M. Szabaciuk, Lublin 2014, s. 273–294.
- Nawrocki S., *Doświadczenia zagraniczne w zakresie stosowania komputera w archiwach*, „Biuletyn Informatyka i Archiwa” 1978, nr 11, s. 6–33 (tekst w nieco zmienionej postaci został opublikowany w czasopiśmie „Archeion” 1982, t. 74, s. 5–26).
- Oard D.W. et al., *Evaluation of Information Retrieval for E-discovery*, „Artificial Intelligence and Law” 2010, t. 18, s. 347–386.
- Oliver G. et al., *Recordkeeping Informatics for a Networked Age*, Clayton, Wiktoria 2018.
- Open Data Institute, The National Archives, University of Surrey, *ARCHANGEL: guaranteeing the integrity of digital archives. Project Report*, 2019, https://theodi.org/wp-content/uploads/2019/08/2019-08-ARCHANGEL_-digital-archhive-integrity-Open-Data-Institute.pdf [dostęp: 30.05.2021].
- Pełowska K., *Archiwa cyfrowe w wybranych krajach europejskich, USA i Australii*, Biblioteka Zarządcy Dokumentacji, t. 8, Toruń 2017.
- Perzyna P., *Problemy terminologiczne wynikające z komputeryzacji archiwów*, [w:] *Komputeryzacja i digitalizacja w archiwach*, Symposia Archivistica, t. 2, red. R. Leśkiewicz, A. Żeglińska, Warszawa 2016, s. 47–69.

- Pieriegud J., *Cyfryzacja gospodarki i społeczeństwa – wymiar globalny, europejski i krajowy*, [w:] *Cyfryzacja gospodarki i społeczeństwa. Szanse i wyzwania dla sektorów infrastrukturalnych*, red. J. Gajewski, W. Paprocki, J. Pieriegud, Gdańsk 2016, s. 11–37.
- Raport o archiwach Unii po rozszerzeniu: pogłębiona współpraca archiwalna w Europie – plan działania*, oprac. przez Grupę Ekspertów Krajowych ds. Archiwów Państw Członkowskich Unii Europejskiej oraz Instytucji i Organów UE na zamówienie Rady Unii Europejskiej, red. nauk. A. Biernat, W. Stępiak, V. Urbaniak, tł. J. Szymańska, Warszawa 2012.
- Redzisz M., *Piłsudski z czarnej skrzynki*, *Sztuczna inteligencja* 23.03.2020, <https://www.sztuczna-inteligencja.org.pl/pilsudski-z-czarnej-skrzynki/> [dostęp: 30.05.2021].
- Renade S., *Access Technologies for the Disruptive Digital Archive*, [w:] *Archival Futures*, red. C. Brown, Londyn 2018, s. 79–98.
- Rolan G. et al., *More Human than Human? Artificial Intelligence in the Archive*, „Archives and Manuscripts” 2019, t. 47, nr 2, s. 179–203.
- Saaman E., *Archiving by design*, National Archives of the Netherlands, <https://www.nationaalarchief.nl/en/archive/knowledge-base/archiving-by-design> [dostęp: 30.05.2021].
- Seles A., *A Brave New World: Artificial Intelligence and Archives* (referat przedstawiony na konferencji „Archives Today and Tomorrow: Prospering as a Diversity of Records Dramatically Increase”, zorganizowanej w 2019 r. przez Archiwum Narodowe Japonii oraz oddział regionalny Międzynarodowej Rady Archiwów dla Azji Wschodniej – EASTICA), http://www.archives.go.jp/english/news/pdf/20191125_27e_01.pdf oraz http://www.archives.go.jp/english/news/pdf/20191125_27e_02.pdf [dostęp: 30.05.2021].
- Schneider J. et al., *Appraising, Processing, and Providing Access to Email in Contemporary Literary Archives*, „Archives and Manuscripts” 2019, t. 47, nr 3, s. 305–326.
- Theimer K., *Digital Archives*, [w:] *Encyclopedia of Archival Science*, red. L. Duranti, P.C. Franks, Lanham 2015, s. 157–159.
- Torczyńska M., *Sztuczna inteligencja i jej społeczno-kulturowe implikacje w codziennym życiu*, „Kultura i Historia” 2019, nr 36, s. 106–126.
- Vellino A., Alberts I., *Assisting the Appraisal of E-mail Records with Automatic Classification*, „Records Management Journal” 2016, t. 26, nr 3, s. 293–313.
- Wasiuta O., Wasiuta S., *Deepfake jako skomplikowana i fałszywa rzeczywistość*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Securitate” 2019, nr 9 (3), s. 19–30.
- Whitelaw M., *Generous Interfaces for Digital Cultural Collections*, „Digital Humanities Quarterly” 2015, t. 9, nr 1, <http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html> [dostęp: 30.05.2021].

Zaręba T., *Self aware networks – cechy techniczne i implikacje technologiczne*, [w:] *Cyfryzacja gospodarki i społeczeństwa. Szanse i wyzwania dla sektorów infrastrukturalnych*, red. J. Gajewski, W. Paprocki, J. Pieriegud, Gdańsk 2016, s. 75–90.

Netografia

ARCHANGEL, <https://www.archangel.ac.uk/> [dostęp: 30.05.2021].

European Commission. A European Approach to Artificial Intelligence, <https://ec.europa.eu/digital-single-market/en/artificial-intelligence> [dostęp: 30.05.2021].

InterPares Trust, <https://interparestrust.org> [dostęp: 30.05.2021].

Nuix, <https://www.nuix.com/> [dostęp: 30.05.2021].

Public Record Office Victoria. Email appraisal, disposal and preservation, <https://prov.vic.gov.au/recordkeeping-government/research-projects/email-appraisal-disposal-preservation-project> [dostęp: 30.05.2021].

READ (Recognition and Enrichment of Archival Documents), <https://readcoop.eu/> [dostęp: 30.05.2021].

Recordkeeping Roundtable. Recordkeeping Roundcasts Episode 3: Machine learning (rozmowa z Johnem Sheridanem, dyrektorem ds. cyfrowych w The National Archives, na temat uczenia maszynowego i jego możliwych zastosowań), <https://rkroundtable.org/2018/08/19/recordkeeping-roundcasts-episode-3-machine-learning/> [dostęp: 30.05.2021].

Records Management Journal, <https://www.emerald.com/insight/publication/issn/0956-5698> [dostęp: 30.05.2021].

Surrey Blockchain, <https://blockchain.surrey.ac.uk/index.html> [dostęp: 30.05.2021].

Serwis Rzeczypospolitej Polskiej. Droga do polskiej strategii AI, <https://www.gov.pl/web/cyfryzacja/droga-do-polskiej-strategii-ai> [dostęp: 30.05.2021].

Towards a National Collection, <https://www.nationalcollection.org.uk/> [dostęp: 30.05.2021].

United Nations. Department of Economic and Social Affairs. Public Institutions, <https://publicadministration.un.org/> [dostęp: 30.05.2021].

YouTube. Dokument elektroniczny w archiwum – dzień I (nagranie z konferencji „Dokument elektroniczny w archiwum. Metody postępowania, dobre praktyki i techniki zarządzania”), <https://www.youtube.com/watch?v=9R3EPGZvkUY&t=23720s> [dostęp: 30.05.2021].