

InterPARES Summer School San Benedetto (Włochy), 7–11 lipca 2023 r.

InterPARES (International Research on Permanent Authentic Records in Electronic Systems)¹ to międzynarodowy program badawczy, zainicjowany w 1998 r. przez prof. Lucianę Duranti – specjalistkę w dziedzinie archiwistyki, dyplomatyki i dokumentu elektronicznego z Uniwersytetu Kolumbii Brytyjskiej w Vancouver². Program prowadzony jest dzięki zaangażowaniu wielu instytucji badawczych, włączając w to ponad 50 uniwersytetów, oraz organizacji publicznych (w tym archiwów) i prywatnych w Ameryce Północnej, Łacińskiej, Europie Zachodniej, Afryce, Australii i Azji. Jego główny cel stanowi wypracowanie teoretycznych i metodologicznych ram dla procedur zapewniających wiarygodność i dostępność archiwów elektronicznych oraz wysoką jakość zarządzania informacją, a tym samym trwałą pamięć cyfrową. Cztery pierwsze edycje poświęcono interdyscyplinarnym badaniom, prowadzonym na styku archiwistyki, zarządzania dokumentacją, prawa, technologii informatycznych, inżynierii komputerowej i nauki o informacji. Dotyczyły one trwałej archiwizacji dokumentów elektronicznych, a więc zachowania przez długi czas ich wiarygodności, autentyczności i integralności oraz dostępności (w kontekście technologicznym i fizycznym) w środowiskach cyfrowych.

Piąta edycja InterPARES, realizowana w latach 2021–2026 pod nazwą InterPARES Trust AI, skupiała się na zagadnieniach zastosowania sztucznej inteligencji

¹ Strona internetowa projektu, zob. The InterPARES Project, <http://www.interpares.org> [dostęp: 1.10.2023].

² Prof. Luciana Duranti była gościem archiwów państwowych w 2022 r. Podczas pierwszego z cyklu międzynarodowych spotkań archiwalnych pn. „Konarski Lectures” wygłosiła wykład „Why a world gone digital needs archival theory more than ever before”. Zob. L. Duranti, *Why a world gone digital needs archival theory more than ever before*, „Archeion” 2022, t. 123, s. 10–30, DOI.10.4467/26581264ARC.22.015.16668.

cji (*artificial intelligence*, AI) w celu zapewnienia cyfrowej dostępności dokumentacji elektronicznej³.

Wyniki dotychczasowych studiów teoretycznych InterPARES Trust AI zaprezentowano uczestnikom Szkoły Letniej odbywającej się w San Benedetto del Tronto we Włoszech od 7 do 11 lipca 2023 r. Wzięło w niej udział 35 młodych archiwistów, zarządców dokumentacji i badaczy, a także przedstawiciele innych dziedzin związanych z rozwojem nowoczesnych technologii cyfrowych, takich jak informatyka, prawo i biznes⁴.

InterPARES Summer School otworzyło sympozjum pn. „Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records – an overview of InterPARES Trust AI”. Podzielono je na 10 sesji wykładowych, prezentujących najważniejsze problemy badawcze podejmowane w ramach InterPARES Trust AI, a także metodologie badań i oczekiwane rezultaty, które miałyby w sposób systemowy objąć wszystkie aspekty archiwizacji cyfrowej. W wykładzie inauguracyjnym prof. L. Duranti podkreśliła, że obecna edycja InterPARES różni się od poprzednich podejściem do zagadnienia procesów archiwizacji. Jej tematyka nie dotyczy – jak poprzednio – poszczególnych typów dokumentów wytworzonych w określonych technologiach, lecz zastosowania sztucznej inteligencji w celu długoterminowego zachowania wszelkich dokumentów cyfrowych, utrwalonych na dowolnych nośnikach i pochodzących z dowolnego okresu, w taki sposób, aby zagwarantować ich wiarygodność i autentyczność. Zadanie to wymaga zapewnienia odpowiednich narzędzi. Powinny być one przejrzyste, w otwartym dostępie, zintegrowane, interoperacyjne i trwałe.

W InterPARES Trust AI zidentyfikowano i wyodrębniono siedem obszarów tematycznych, obejmujących wszystkie aspekty dotyczące zastosowania technologii AI w archiwizacji. Istotnym elementem metody badawczej jest odwołanie do procesów aktotwórczych, wpływających na wytwarzanie, zarządzanie i wykorzystanie (także ponowne) dokumentacji elektronicznej. W założeniu te trzy wymiary stanowią punkt wyjścia do określenia miejsca praktycznego zastosowania sztucznej inteligencji w archiwach. Przyjęty model koncepcyjny zakłada zatem kształtowanie przestrzeni za pomocą nowoczesnej technologii i wypełnienie jej

³ Strona internetowa projektu, zob. InterPARES Trust AI (2021–2026), <https://interparestrustai.org/> [dostęp: 1.10.2023].

⁴ Informacje o Szkole Letniej, zob. InterPARES Trust. InterPARES Summer School – July 2023 – San Benedetto del Tronto (AP), <https://interparestrustai.org/trust/article/inter pares-summer-school-july-2023> [dostęp: 1.10.2023].

bogatą i różnorodną treścią cyfrowego dziedzictwa kulturowego. Należy podkreślić, że to ostatnie obejmuje zarówno odzwierciedlone cyfrowo wersje analogowych archiwaliów, jak i dokumentację wytwarzaną w formacie binarym.

Nakreślone powyżej cele badawcze są realizowane przez 33 projekty naukowe, z których większość wiąże się bezpośrednio z rozwojem archiwistyki cyfrowej. Ich praktyczne wykorzystanie ma doprowadzić w przyszłości do przemodelowania koncepcji archiwum rozumianego jako zbiór akt i dokumentów w archiwum pojmowane jako zbiór danych. To nowe podejście oznacza, między innymi, że tradycyjne metody archiwalnej oceny dokumentów staną się nieadekwatne. Potrzebne będą zarówno nowe narzędzia, oparte na relacjach człowiek-maszyna, jak i nowe metody wykorzystania i eksploracji archiwów – nowoczesnych składnic (miejsc przechowywania) informacji. Zakłada się, że efektem końcowym projektu InterPARES AI będzie raport prezentujący najważniejsze osiągnięcia w zakresie zastosowania AI w archiwach, problemy wymagające rozwiązania i bariery. Taka diagnoza przyczyni się do rozwoju badań nad opracowaniem praktycznych rozwiązań dotyczących przeniesienia najnowszych technologii AI na grunt archiwalny.

Podczas sesji plenarnej omówiono szczegółowo założenia kilku projektów badawczych, zwracając szczególną uwagę na fazę koncepcyjną badań. W latach 2022–2023 prace studyjne grup roboczych InterPARES skupiły się na określeniu zarówno korzyści, jak i ryzyk związanych z wykorzystaniem technologii AI w zarządzaniu dokumentami i archiwami, opracowaniu rekomendacji w zakresie zmiany prawa, identyfikacji ograniczeń zastosowania poszczególnych rozwiązań AI, analizie kierunków rozwoju tej technologii oraz badaniu rynku i potrzeb użytkowników w zakresie wdrożenia nowych rozwiązań.

Sztuczna inteligencja znajduje obecnie zastosowanie przede wszystkim w rozpoznawaniu tekstu dokumentów historycznych; coraz częściej badacze wykorzystują ją także jako złożone narzędzie wnioskowania statystycznego. Celem studium pn. „AI in the Middle ages – Arrangement of ancient documents via appearance-based recognition”⁵ jest analiza funkcjonalności głębokich sieci neuronowych, mogących mieć zastosowanie w pracy archiwistów. W założeniach badania przyjęto, iż analiza komputerowa źródeł historycznych zapewnia obiektywną metodę oceny treści dokumentów, niezależną od interpretacji subiektywnej, a jednocześnie charakteryzującą się większą dokładnością, czułością i spoisto-

⁵ Opis badań prowadzonych w ramach projektu, zob. InterPARES Trust AI. Research Studies, https://interparestrustai.org/trust/about_research/studies [dostęp 1.09.2023].

ścią. W badaniu testowano narzędzia AI do rozpoznawania pergaminów i dokumentów średniowiecznych. W pierwszym etapie dokonywano ręcznego tagowania dużych zbiorów danych, służących następnie do uczenia się maszynowego (Machine Learning, ML). W drugim etapie głęboka sieć neuronowa pozwalała na automatyczne wykonywanie czynności rozpoznawania tekstów.

Koncepcja kolejnego badania – „Intensional logic-based AI for the Records in Contexts Conceptual Model (RiC-CM) and Ontology (RiC-O)” – jest przykładem próby nawiązania porozumienia pomiędzy archiwistyką a sztuczną inteligencją w obszarze metodyki archiwalnej. Polega na identyfikacji konkretnych technologii AI, które mogą mieć zastosowanie do analizy i rozszerzenia ontologii Records in Context (RiC-O), tworzącej terminologię i zasady uporządkowanego opisu dowolnego typu zasobu archiwalnego w oparciu o szeroki kontekst jego powstawania (twórca-agent i jego działalność). Należy przy tym podkreślić, że ontologie są nowym elementem sztucznej inteligencji, który wykazał swoją przydatność jako narzędzie konceptualne służące modelowaniu wiedzy. RiC i AI tworzą spójną podstawę budowy opisów archiwalnych i relacji pomiędzy nimi w formie ujednoczonego słownika pojęciowego. Ontologia RiC ma więc pełnić rolę wspólnego, powszechnie rozumianego języka światowej archiwistyki.

W projekcie InterPARES skupiono się także na relacji pomiędzy sztuczną inteligencją i językoznawstwem w kontekście zastosowań AI do wyodrębnienia i analizy metadanych dotyczących zasobów cyfrowych. Wskazuje się, że w sytuacji rosnącej lawinowo ilości informacji w zbiorach danych oraz wynikających stąd coraz większych wymagań wobec technologii ich przechowywania, automatyczna klasyfikacja dokumentów elektronicznych jest niezbędna. Z tego względu wiele uwagi poświęcono technikom przetwarzania języka naturalnego (Natural Language Processing, NPL). Model funkcjonalny, zapewniający autentyczność, wiarygodność, integralność i dostępność danych cyfrowych, pozwala na tworzenie schematów określających funkcje poszczególnych elementów tekstu oraz systematyzację jednostek nazewniczych według zdefiniowanych klas i typów. Umożliwia także analizę wątków tematycznych opisujących dokumenty tekstowe. W ramach badania „Records Classification using Natural Language Processing Techniques to Support Trustworthy Public Digital Record” porównano wydajność tradycyjnych technik klasyfikacji tekstów dokumentów z najnowszymi technologiami przetwarzania języka naturalnego.

Dostępność zaawansowanych technologicznie narzędzi sztucznej inteligencji i uczenia maszynowego stanowi wyzwanie dla tych profesji, które tradycyjnie zaj-

mują się dokumentowaniem ludzkiej działalności, nie zaś obcych im procesów zachodzących w świecie cyfrowym. Wstępne prace projektu „Preserving AI Techniques as Paradata” oparto na koncepcji paradanych, jako danych opisujących kontekst (w tym środowisko technologiczne) wykorzystania narzędzi AI w działalności archiwalnej. Zachowanie takich paradanych jest niezbędne do zrozumienia procesów archiwizacyjnych, w których użyto technologii AI, podobnie jak zachowanie metadanych jest niezbędne do zrozumienia procesów powstawania dokumentacji elektronicznej. Rozwój AI otwiera przed archiwistyką nowe możliwości. Sprawia, że granice dotychczasowej praktyki archiwalnej ulegają przesunięciu i wymagają opracowania nowych pojęć i słowników.

Analiza interakcji zachodzących między sztuczną inteligencją a dokumentacją we wszystkich aspektach jej cyklu życia (od wytworzenia, poprzez zarządzanie, wykorzystanie i archiwizację) otwiera dyskusję dotyczącą etycznych granic cyfrowej dostępności. Dotyka ona kwestii zaufania, odpowiedzialności, kontroli i obiektywizmu. Celem projektu pn. „Comparative assessment of ethical codes of the archival/records management and artificial intelligence communities” jest opracowanie katalogu dobrych praktyk, mających zastosowanie do etycznej oceny działalności archiwalnej w świecie nowych technologii. Wstępne założenie, zgodnie z którym dokumentacja powinna być autentycznym (wiarygodnym) dowodem działalności człowieka, rozszerzono o podejście multidyscyplinarne, uwzględniające komponent AI. Badanie zostanie uzupełnione analizą porównawczą kodeksów etycznych zawodu archiwisty/zarządcy dokumentacji i etyki dotyczącej godnej zaufania sztucznej inteligencji.

W ramach badań pn. „Users’ approaches and behaviors in accessing records and archives in the perspective of AI: a global user study” opublikowano pierwszy, wstępny raport krajowy obejmujący badania użytkowników archiwów państwowych w Mediolanie, Turynie, Wenecji i Ascoli Piceno⁶. Dokument, oparty o analizę danych badania ankietowego przeprowadzonego online w 2022 r., dostarcza interesujących informacji na temat doświadczeń i oczekiwań użytkowników włoskich w zakresie dostępu do cyfrowych usług archiwalnych oraz ich stosunku do rozwiązań opartych na sztucznej inteligencji. Jak wynika z ankiety, respondenci oczekują od nowoczesnych technologii zwiększenia dostępności zdigitalizowa-

⁶ P. Feliciati, G. Di Marcantonio, *Users’ approaches and behaviors in accessing records and archives in the perspective of AI: a global user study Phase 1 – Italian study. April 2023*, https://interparestrustai.org/assets/public/dissemination/RA05Users_AI_firstreport-april2023.pdf [dostęp: 1.09.2023].

nych zasobów i ogólnej poprawy jakości wyszukiwania informacji. Pozytywnie oceniono zastosowanie sztucznej inteligencji do mechanizmów wyszukiwania, choć jednocześnie wskazano na potrzebę uzyskania dodatkowych informacji na temat zalet i wad użytkowania tego narzędzia w systemach informacji archiwalnej (ok. 65% badanych). W podsumowaniu raportu autorzy porównali uzyskane dane z wynikami sondażu IPSOS z 2022 r. dotyczącego wad i zalet produktów i usług wykorzystujących sztuczną inteligencję⁷.

Tematyka projektu pn. „Increasing Access to Photos, Videos and Social Media records through AI-generated Descriptive Metadata” dotyczy możliwości zastosowania sztucznej inteligencji i uczenia maszynowego do rozpoznania i klasyfikowania dokumentacji nietekstowej, głównie fotografii cyfrowych, filmów i nagrań. Automatyczne rozpoznawanie mowy, maszynowe tłumaczenie dialogów filmowych i kodowanie czasowe scen wskazują na możliwość poprawy dostępu do cyfrowego dziedzictwa kulturowego. Takie narzędzia mogą znacznie przyspieszyć tworzenie metadanych na poziomie jednostki archiwalnej oraz pomóc archiwistom porządkować i opracowywać duże zbiory fotografii, filmów i nagrań dźwiękowych.

Główną częścią programu InterPARES Summer School było podsumowanie wyników projektów realizowanych we wcześniejszych edycjach. Nawiązanie do problematyki badań nad sztuczną inteligencją stanowiło uzupełnienie treści wykładów i pozostawiło pewien niedosyt wśród uczestników spotkania. Zaprezentowany dorobek studiów zrealizowanych w poprzednich edycjach skłania do refleksji, iż nowoczesne technologie wymagają opracowania nowej, posługującej się narzędziami cyfrowymi, metodyki archiwalnej. Istotna przy tym jest rola archiwów jako podmiotu wytwarzającego i skutecznie wdrażającego nowe rozwiązania, a nie jedynie wykorzystującego powszechnie dostępne narzędzia i technologie. Koncepty zaprezentowane w poszczególnych projektach mogą stanowić dobry punkt wyjściowy do kontynuacji i rozwijania dalszych badań.

Na dokładniejsze omówienie zasługują następujące, zaprezentowane podczas Szkoły Letniej, sesje i wykłady:

Corinne Rogers, „Authenticity and Authentication”. Zgodnie z teorią archiwistyki autentyczność określana jest na podstawie pochodzenia i kontekstu dokumentu oraz nośnika, na którym został zapisany. W środowisku cyfrowym treść, struktura

⁷ Zob. N. Maslej, L. Fattorini, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, H. Ngo, J.C. Niebles, V. Parli, Y. Shoham, R. Wald, J. Clark, R. Perrault, *The AI Index 2023 Annual Report, Stanford, April 2023*, https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf [dostęp: 1.09.2023].

i forma nie są już ze sobą nierozzerwalnie związane, a domniemanie autentyczności opiera się na badaniu elementów tożsamości i integralności zapisu cyfrowego. Profesor Uniwersytetu Kolumbii Brytyjskiej w Vancouver zaprezentowała ustalenia InterPARES dotyczące autentyczności i uwierzytelniania dokumentu w środowisku cyfrowym oraz ontologii wiarygodności, do której należy autentyczność.

Hrvoje Stancic, „Technological Authentication”. Era cyfrowa postawiła przed archiwami nowe wyzwania. Choć podstawowe teorie i koncepcje archiwalne powinny być takie same zarówno dla dokumentów analogowych, jak i cyfrowych, to jednak techniczna manifestacja dokumentów cyfrowych, ich różnorodność, szybkość tworzenia, objętość i zmienność wymagają nowoczesnych podejść obliczeniowych. Wykładowca Uniwersytetu w Zagrzebiu przedstawił model blockchain i technologii rozproszonych rejestrów w kontekście wsparcia integralności zapisów świata cyfrowego i możliwości ich wykorzystania w archiwistyce.

Tracey P. Lauriault, „Data as artifacts and as records”. Wykład poświęcony był związkowi między metrologią a danymi oraz rodzajami danych i ich cechami ontologicznymi. Wykładowczyni, profesor z Uniwersytetu Carleton w Ottawie, przybliżyła koncepcje współczesnych systemów, takich jak inteligentne miasta i cyfrowe bliźniaki. Przede wszystkim jednak słuchacze starali się znaleźć odpowiedzi na pytania o kwestie etyczne i zagadnienia sprawiedliwości społecznej, związane bezpośrednio z wykorzystaniem sztucznej inteligencji, a także dokonać oceny społecznych i kulturowych rezultatów działalności człowieka w środowisku cyfrowym.

Erik Borglund, „Records in the Cloud”. Nowoczesna usługa w chmurze została uruchomiona przez Amazon w 2002 r. w ramach infrastruktury wewnętrznej. Cztery lata później, w 2006 r. „chmurę” wprowadzono na rynek zewnętrzny, tworząc nowoczesną technologię przetwarzania oraz przechowywania danych, które nie są magazynowane na dyskach komputerowych, ale w wirtualnej przestrzeni serwerowej. Pojawienie się dostawców tego typu usług stanowiło wyzwanie dla archiwistów i zarządców dokumentacji. Przedstawiciel Uniwersytetu Środkowej Szwecji (Mittuniversitetet) w Sundsvall skupił się na założeniach projektu InterPARES „Rekordy w chmurze”, ale także na wyzwaniach związanych z tą technologią.

Kenneth Thibodeau, „Proteus Bound: Trustworthy Digital Preservation”. Zachowanie wiarygodności zasobów cyfrowych w długim czasie wydaje się być niemożliwe. Jak można zachować coś w nienaruszonym stanie, gdy zależy to całkowicie od zbioru luźno połączonych i różnorodnych czynników, które zmieniają się w różnym tempie i w sposób nieprzewidywalny? Jest to pytanie, którym InterPARES zajmuje się od samego początku. Pierwsze dwa projekty sformułowały

i opracowały koncepcję łańcucha ochrony, który nie jest sposobem na zachowanie dokumentów cyfrowych, ale sposobem dokumentowania tego, w jaki sposób zostały one zachowane. Podczas wykładu dokonano przeglądu badań InterPARES w zakresie ochrony zasobów cyfrowych, ich wyników, a także zaprezentowano problem badań dokumentacji cyfrowej powstającej współcześnie w dynamicznych systemach nowoczesnej technologii.

Pierluigi Feliciati, „Trusted and easy access to records and archives”. Profesor Uniwersytetu w Maceracie skupił się na zależności zachodzącej między dokumentami cyfrowymi a użytkownikami, służącej zaprojektowaniu, wdrożeniu i podtrzymaniu cyfrowych usług archiwalnych. Podczas wykładu zaprezentowano wyniki badania InterPARES Trust AI „Users’ approaches and behaviors in accessing records and archives in the perspective of AI: a global user study”. Częścią zajęć szkoleniowych było ponadto omówienie i ocena współczesnych możliwości dostępu do archiwów cyfrowych oraz podstawowych mierników jakościowej oceny ich dostępności.

Jessica Bushey, „Managing and Preserving Digital Images Collections”. Podczas tej sesji profesor Uniwersytetu Stanowego w San José w Kalifornii przedstawiła wyniki badań dotyczących zarządzania fotografiami cyfrowymi. Ukazała rolę metadanych w odnajdywaniu informacji przyczyniających się do potwierdzenia autentyczności i wiarygodności obrazu oraz zapoznała uczestników ze standardami i najlepszymi praktykami dotyczącymi metadanych i formatów w zakresie dostępu i przechowywania. Omówiła ponadto archiwizację fotografii publikowanych w mediach społecznościowych. Sesja zakończyła się dyskusją na temat możliwości i potencjalnych przeszkód w wykorzystaniu sztucznej inteligencji do zarządzania i ochrony obrazów cyfrowych.

InterPARES Summer School było wydarzeniem rozwijającym, szczególnie w odniesieniu do poznania stanu wiedzy i badań prowadzonych w obszarze nowoczesnej archiwistyki cyfrowej. Sztuczna inteligencja daje szerokie perspektywy zastosowania i w przyszłości stanowić może ważne narzędzie pracy archiwalnej. Otwarte pozostają pytania o granice technologiczne, etyczne i społeczne jej wykorzystania oraz konsekwencje powierzenia dziedzictwa kulturowego systemom-maszynom opartym na analizie obliczeniowej.

Kamila Pawełczyk-Dura