

Tomasz Parkoła  
Poznańskie Centrum Superkomputerowo-Sieciowe  
[tomasz.parkola@man.poznan.pl](mailto:tomasz.parkola@man.poznan.pl)

## Długoterminowe przechowywanie cyfrowego dziedzictwa kulturowego

**Streszczenie:** Autor podkreśla wagę długoterminowej archiwizacji zasobów cyfrowych, za pomocą praktycznych przykładów i przypadków, z którymi mają do czynienia korzystający z bibliotek cyfrowych. Przedstawia osiągnięcia zagranicznych i polskich ośrodków w tym zakresie, podaje standardy i instytucje, które je opracowują, przytacza szereg narzędzi, które wspierają te procesy.

**Słowa kluczowe:** długoterminowe przechowywanie zasobów; magazyny danych; biblioteki; digitalizacja;

### Wprowadzenie

Digitalizacja i udostępnianie zasobów cyfrowych dziedzictwa kulturowego w internecie są ważnym elementem działań każdej instytucji nauki i kultury. Są one niezbędne, aby zaistnieć w świadomości społeczeństwa informacyjnego na poziomie regionu, kraju czy w kontekście międzynarodowym. To dzięki digitalizacji mamy dostęp do milionów obiektów cyfrowych w europejskim portalu Europeana<sup>1</sup>, amerykańskiej Digital Public Library of America<sup>2</sup> czy w polskiej Federacji Bibliotek Cyfrowych[1]. W Polsce działają zarówno krajowe programy finansowania digitalizacji (np. Wieloletni Program Rządowy KULTURA+ na lata 2011–2015<sup>3</sup>), jak i te współfinansowane międzynarodowo (np. fundusze strukturalne, które w Polsce są zarządzane przez specyficzne Programy Operacyjne<sup>4</sup>). Digitalizacją interesują się również podmioty komercyjne, co widać w projektach partnerstwa publiczno-prywatnego z udziałem takich gigantów technologii ICT jak Google<sup>5</sup>. Tempo cyfrowej dokumentacji obiektów dziedzictwa kulturowego rośnie, a proces ten nabiera charakteru masowego. Stąd też niezwykle istotna staje się odpowiednia organizacja prac związanych z digitalizacją, a także należyte zachowanie rezultatów tych prac.

---

<sup>1</sup> Europeana [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://www.europeana.eu/>.

<sup>2</sup> Digital Public Library of America [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://www.dp.la>.

<sup>3</sup> Program Wieloletni KULTURA+ [on-line]. Ministerstwo Kultury i Dziedzictwa Narodowego [dostęp 9.01.2015]. Dostępny w: <http://www.mkidn.gov.pl/pages/strona-glowna/finanse/program-wieloletni-kultura.php>.

<sup>4</sup> Możliwości finansowania kultury z funduszy europejskich w latach 2007-2013 [on-line]. Ministerstwo Kultury i Dziedzictwa Narodowego [dostęp 9.01.2015]. Dostępny w: [http://www.mkidn.gov.pl/media/docs/Finansowanie\\_kultury-2007-13.pdf](http://www.mkidn.gov.pl/media/docs/Finansowanie_kultury-2007-13.pdf).

<sup>5</sup> Library Partners [on-line]. Google Books [dostęp 9.01.2015]. Dostępny w: <http://www.google.pl/googlebooks/library/partners.html>.

Użytkownicy zasobów cyfrowych oczekują ciągłej i nieprzerwanej dostępności do interesujących ich materiałów, bez względu na urządzenia czy programy, z których korzystają. W krótkiej perspektywie czasu (np. 5–10 lat) zapewnienie dostępu nie jest trudnym zadaniem. Jeśli jednak wydłużymy czas, w ramach którego zasoby powinny być dostępne dla użytkownika, zadanie to staje się trudniejsze i bardziej skomplikowane. Dla przykładu przeanalizujemy sytuację związaną z udostępnianiem obiektów cyfrowych w Polsce, gdzie tylko od 2002 r. udostępnionych zostało w internecie ponad dwa miliony obiektów cyfrowych. Z tych dwóch milionów obiektów około 62% zapisanych jest w formacie DjVu<sup>6</sup>. Format ten ma dobre parametry techniczne, np. kompresję czy zaawansowane możliwości reprezentowania różnych aspektów cyfrowego obiektu (np. tekst i grafika). Dlatego też został uznany za optymalne rozwiązanie w czasie, gdy przepustowość łącz internetowych była niewielka, a umiejętności użytkowników stosunkowo duże (w tym czasie przeciętny użytkownik internetu był z reguły w stanie zainstalować odpowiednią wtyczkę do przeglądarki WWW, która umożliwiłaby mu odczytanie specyficznego formatu pliku, np. PDF czy DjVu).

W ciągu kilkunastu lat zarówno przyzwyczajenia użytkowników, jak i charakter programów komputerowych znacznie się zmieniły — obecnie to programy komputerowe muszą w odpowiedni sposób przedstawić informację użytkownikowi, a on sam nie powinien być angażowany w techniczne szczegóły funkcjonowania internetu czy określonej aplikacji. Jest to również istotne w kontekście bezpieczeństwa programów, danych i komputera użytkownika, np. odpowiednie zabezpieczenie przed instalowaniem zewnętrznego oprogramowania, które potencjalnie może zainfekować komputer.

W rezultacie twórcy najpopularniejszych przeglądarek internetowych zdecydowali się na wbudowane wsparcie najpopularniejszych formatów plików dostępnych w internecie. W szczególności opracowano dedykowane wsparcie dla formatu PDF — obecnie każdy użytkownik może bez problemu wyświetlić plik PDF w przeglądarce Chrome czy Firefox. Inaczej funkcjonuje to w przypadku formatu DjVu — aby wyświetlić ten format należy zainstalować dedykowane oprogramowanie, które niestety traktowane jest przez przeglądarki internetowe jako potencjalnie niebezpieczny program (zawierający np. wirus komputerowy). W efekcie, wielu użytkowników internetu nie otwiera materiałów w formacie DjVu ze względu na strach przed rzekomo niebezpiecznym oprogramowaniem lub ze względu na brak umiejętności instalowania specjalnej wtyczki do przeglądarki stron WWW. Sytuacja robi się jeszcze trudniejsza w przypadku niektórych urządzeń mobilnych, które pojawiły się w ostatnich latach i są w powszechnym użytkowaniu. Przykładowo, urządzenia pracujące z systemem operacyjnym iOS (również MacOS) nie mają możliwości wyświetlenia formatu DjVu bezpośrednio w przeglądarce internetowej.

Z przytoczonego przykładu widać, że choć minęło zaledwie kilkanaście lat od uruchomienia

<sup>6</sup> Federacja Bibliotek Cyfrowych [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://fbc.pionier.net.pl/owoc/attr-stats#format>.

pierwszej biblioteki cyfrowej w Polsce, mamy już do czynienia z trudnościami w dostępie do obiektów zapisanych w formacie DjVu, tzn. użytkownicy internetu z roku na rok mają coraz mniejsze możliwości dotarcia do takich dokumentów cyfrowych. Oczywiście problem dotyczy wersji prezentacyjnych obiektów cyfrowych, zatem jeśli jesteśmy w posiadaniu wersji master, możliwe jest wygenerowanie nowych wersji prezentacyjnych w jednym z bardziej popularnych formatów. Alternatywnie możliwe jest przekonwertowanie formatu DjVu do bardziej popularnego formatu lub dostarczenie użytkownikom kilku formatów do wyboru.

W praktyce problem długoterminowego przechowywania danych cyfrowych jest związany z dynamiką zmian, która ma miejsce w odniesieniu do technologii ICT. W tym kontekście istnieje szereg czynników ryzyka, do których zalicza się między innymi:

- zmiany sprzętowe i technologiczne — urządzenia, które służą do odczytu informacji, np. czytniki CD-ROM, czy całe urządzenia komputerowe (np. komputery PC, urządzenia mobilne) podlegają ciągłym zmianom i mogą wyjść z codziennego użycia. Warto przytoczyć w tym miejscu przykład — zaledwie w ciągu kilkadziesiąt lat nośniki danych, takie jak dyskietki 8", 5.25", a także dyskietki 3.5" w praktyce wyszły z użycia codziennego;
- zmiany oprogramowania — z biegiem lat oprogramowanie wykorzystywane do tworzenia i odczytu informacji cyfrowych jest modyfikowane, aktualizowane, a także wychodzi z użycia na rzecz oprogramowania bardziej nowoczesnego. Zmieniają się zatem również formaty danych, jakie konkretne programy potrafią odczytać czy zapisać. Pojawia się więc kwestia odczytu określonego formatu danych, w jakim zapisano zasoby cyfrowe kilka lub kilkanaście lat wcześniej. Zagrożenie to jest większe w przypadku wykorzystywania oprogramowania komercyjnego, gdyż formaty wykorzystywane przez te narzędzia są najczęściej tajemnicą firmy, niedostępną publicznie. Przykładem może być pakiet biurowy MS Office, w ramach którego do zapisu danych dokumentów tekstowych wykorzystywany był komercyjny format DOC. Dopiero od wersji 2007 pakietu biurowego MS Office domyślnym formatem zapisu danych jest otwarty format Open Office XML (rozszerzenie DOCX). Przytoczony wcześniej przykład formatu DjVu, który posiada otwartą specyfikację, wskazuje również, że sama otwartość i dostępność specyfikacji formatu nie jest warunkiem wystarczającym do odpowiedniego zabezpieczenia materiałów cyfrowych. Ważny jest również poziom wykorzystania i popularności danego formatu;
- brak świadomości istniejących zagrożeń — w instytucjach, które zajmują się digitalizacją i przechowywaniem danych cyfrowych nie ma mocno ugruntowanej świadomości potencjalnych zagrożeń związanych z długoterminowym przechowywaniem. Dane cyfrowe są powszechnie dostępne w internecie, a dzielenie się nimi jest wyjątkowo łatwe i proste. Może to prowadzić do obniżenia czujności u osób odpowiedzialnych za długoterminową archiwizację danych i brak zarządzania ryzykiem, które istnieje w odniesieniu do nich. Istotnym elementem jest tutaj samo

planowanie długoterminowej archiwizacji, włącznie z uwzględnieniem polityki bezpieczeństwa danej instytucji oraz jej możliwości organizacyjno-finansowo-technicznych.

## Inicjatywy międzynarodowe

W skali ogólnoswiatowej zawiązane zostały różnego rodzaju formy współpracy i wspierania działań związanych z przechowywaniem danych. Przykładami inicjatyw wspierających ten proces są m.in.:

1. Open Preservation Foundation — europejska organizacja działająca na rzecz otwartego przechowywania danych cyfrowych;
2. Centrum Kompetencji IMPACT w zakresie digitalizacji (ang. IMPACT Center of Competence in Digitisation) — europejskie centrum kompetencji w zakresie digitalizacji;
3. Digital Preservation Coalition — międzynarodowa organizacja działająca w obszarze długoterminowego przechowywania danych cyfrowych.

## Open Preservation Foundation

Open Preservation Foundation (OPF) to organizacja, która utrzymuje różnego rodzaju narzędzia informatyczne wspomagające proces długoterminowego przechowywania oraz działa na rzecz współdzielenia się wiedzą i doświadczeniami w zakresie przechowywania obiektów cyfrowych. W grupie członków OPF znajdują się instytucje dziedzictwa kulturowego, centra badawczo-rozwojowe oraz firmy komercyjne. Z Polski jedynym członkiem OPF jest Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS). W ramach tej współpracy PCSS rozwija narzędzia służące długoterminowemu przechowywaniu danych, m.in. FITS<sup>7</sup> oraz dArceo<sup>8</sup>.

W zakresie narzędzi utrzymywanych i wspieranych przez OPF jest wiele takich, które wykorzystywane są produkcyjnie w procesach digitalizacji różnego rodzaju instytucji dziedzictwa kulturowego. Narzędzia utrzymywane przez OPF są rozwiązaniami o otwartym kodzie źródłowym i często udostępniane są przy użyciu bardzo liberalnych licencji (np. Apache License v2.0). Pełna lista narzędzi znajduje się na stronach OPF<sup>9</sup>. W skład tych narzędzi wchodzi takie, które są bezpośrednio gotowe do wykorzystania w procesie digitalizacji, jak i te, które są obecnie w fazie rozwojowej. Najciekawszymi narzędziami z punktu widzenia bibliotek cyfrowych i długoterminowego przechowywania plików wzorcowych są:

- Plato[2], który został opracowany przez Uniwersytet Techniczny w Wiedniu. Narzędzie służy do planowania procesu długoterminowego przechowywania danych. Na podstawie informacji wejściowych, takich jak wymagania, co do jakości

<sup>7</sup> *File Information Tool Set (FITS)* [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://fitstool.org/>.

<sup>8</sup> *Darceo* [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://dingo.psnc.pl/darceo>.

<sup>9</sup> *Open Preservation Foundation* [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://openpreservation.org/technology/products/>.

przechowywanych danych, ich rozmiaru czy też potencjalnych form reprezentacji obiektu, oprogramowanie przedstawia alternatywne ścieżki działania. Alternatywy zawierają rekomendacje, co do zastosowania konkretnych mechanizmów przechowywania danych, np. wprowadzenie kompresji obrazów czy też wykorzystanie specyficznego formatu danych. Na tej podstawie możliwe jest przeprowadzenie eksperymentów na testowym zbiorze danych, który jest częścią całej kolekcji podlegającej przechowywaniu. Wyniki eksperymentu wskazują następnie najlepszą metodę przechowywania danych oraz finalny ranking alternatyw. Interesujące przykłady przeprowadzonych analiz z wykorzystaniem narzędzia Plato opisuje artykuł opublikowany w ramach magazynu D-LIB[3];

- C3PO (Clever, Crafty Content Profiling of Objects) — to narzędzie, które przeprowadza analizę metadanych dotyczących przechowywanych zasobów cyfrowych i generuje na tej podstawie różnego rodzaju podsumowania i statystyki. Kluczem do analiz są istniejące już metadane techniczne przechowywanych plików (wspierany jest m.in. format generowany przez otwarte oprogramowanie File Information Tool Set (FITS)). C3PO potrafi przedstawić wyniki analiz na poziomie kilku różnych kolekcji cyfrowych, tworząc równocześnie profil przechowywanych zasobów. Dostarcza bogaty interfejs użytkownika z wieloma wykresami, tabelami i wizualizacjami różnego rodzaju aspektów przechowywanych obiektów. Warto podkreślić, że polskie oprogramowanie dArceo[4] przechowuje metadane techniczne, które mogą być bezpośrednio wykorzystane w ramach C3PO;
- Scout to narzędzie monitorowania kolekcji cyfrowych, które ma na celu identyfikację ryzyka utraty przechowywanych danych. Scout wykorzystuje szereg źródeł informacji, w tym dane pochodzące z bazy formatów PRONOM czy też informacje zbierane przez narzędzia FITS i C3PO. Scout może integrować w swojej bazie wiedzy informacje z różnych instytucji, dzięki czemu możliwe jest współdzielenie wiedzy w zakresie wykorzystywanych formatów i standardów. Na podstawie zebranych informacji przedstawiane są podsumowania i raporty w zakresie bezpieczeństwa zasobów cyfrowych. Oprogramowanie rozwijane jest przez portugalską firmę KEEPS;
- Jpylyzer jest narzędziem, które umożliwia walidację plików w formacie JPEG2000. Jest to jedno z nielicznych narzędzi dostępnych na otwartej licencji, obsługujących format JPEG2000. Jpylyzer umożliwia ekstrahowanie informacji technicznych, dotyczących kluczowych właściwości pliku. Choć nie jest to jeszcze powszechną praktyką, pojawiają się pierwsze zastosowania tego formatu do długoterminowej archiwizacji (np. Wellcome Library [wellcomelibrary.org](http://wellcomelibrary.org)). Obecność oprogramowania Jpylyzer z pewnością ułatwi szerszą adaptację formatu JPEG2000;
- Matchbox jest narzędziem, które umożliwia identyfikację duplikatów w cyfrowych dokumentach graficznych, w szczególności w dokumentach tekstowych, np. manuskryptach czy drukach. Jego zaletą jest to, że identyfikacja przebiega bez względu na rozmiar plików, ich format, obciążenia, przekrzywienia czy pochodzenie. Matchbox może być wykorzystywany na różnych etapach procesu digitalizacji, np.



tuż po skanowaniu/sfotografowaniu zabytku do weryfikacji tego etapu lub do oceny jakości całej kolekcji cyfrowej, np. w przypadku podwykonawstwa dotyczącego skanowania czy fotografowania;

- FIDO (Format Identification for Digital Objects), to oprogramowanie, które umożliwia identyfikację formatów przechowywanych obiektów cyfrowych. Zostało opracowane z myślą o integracji w ramach zautomatyzowanych procesów digitalizacji. Wykorzystuje identyfikatory z bazy formatów PRONOM do opisu analizowanych obiektów.

Wszystkie narzędzia przydatne w procesie przechowywania danych cyfrowych (również te wspomniane powyżej) opisane są w dedykowanym rejestrze COPTR (ang. Community Owned digital Preservation Tools Registry) prowadzonym przez OPF.

Poza szerokim wyborem narzędzi przydatnych w procesie digitalizacji i długoterminowego przechowywania, OPF dostarcza również wielu informacji na temat bieżących działań w zakresie przechowywania obiektów. Informacje te przekazywane są poprzez dedykowany blog OPF oraz kanały społecznościowe, takie jak Twitter czy LinkedIn. OPF organizuje również wirtualne spotkania (tzw. webinaria), podczas których omawiana jest problematyka długoterminowego przechowywania lub prezentowane są narzędzia przydatne w procesie przechowywania danych. Informacje na temat bieżących wydarzeń można śledzić na stronie OPF<sup>10</sup>.

### **Centrum Kompetencji IMPACT w zakresie digitalizacji**

Centrum Kompetencji IMPACT w zakresie digitalizacji (CK IMPACT), to organizacja zrzeszająca instytucje zainteresowane procesem digitalizacji, w szczególności digitalizacji historycznych dokumentów tekstowych. Celem CK IMPACT jest dostarczanie rozwiązań wspomagających proces digitalizacji, włączając w to innowacyjne techniki przetwarzania obrazu, technologie językowe czy przetwarzanie tekstu. Współzałożycielami CK IMPACT są instytucje działające aktywnie w obszarze digitalizacji, włączając w to instytucje dziedzictwa kulturowego, centra badawczo-rozwojowe, jak i firmy komercyjne. Pełna lista dostępna jest na stronie internetowej CK IMPACT<sup>11</sup>. Z Polski członkami CK IMPACT są Poznańskie Centrum Superkomputerowo-Sieciowe w charakterze członka założyciela oraz Biblioteka Uniwersytecka we Wrocławiu w charakterze członka standardowego.

CK IMPACT prowadzi działania w różnych obszarach powiązanych z procesem digitalizacji, w szczególności dotyczy to:

- aktywnego prowadzenia strony internetowej z informacjami na temat narzędzi i zasobów dostępnych poprzez CK IMPACT,
- świadczenia usług konsultacyjnych dotyczących licencjonowania, konfiguracji oraz

<sup>10</sup> *Open Preservation Foundation* [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://openpreservation.org/>.

<sup>11</sup> *The Impact Centre of Competence in Digitisation* [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://www.digitisation.eu/>.

- ewaluacji narzędzi wykorzystywanych w procesie digitalizacji,
- podnoszenia świadomości w obszarze szeroko rozumianej digitalizacji poprzez organizację szkoleń, konkursów oraz grup eksperckich,
  - pomocy uczestnikom CK IMPACT w tworzeniu społeczności badawczych, identyfikacji nadchodzących potrzeb oraz możliwości pozyskiwania funduszy na wspólne projekty,
  - stymulowania udziału uczestników w dedykowanych grupach roboczych dotyczących standardów w digitalizacji.

CK IMPACT utrzymuje m.in. rejestr narzędzi interesujących z punktu widzenia procesu digitalizacji<sup>12</sup>. Obecnie na tej liście dostępnych jest ponad 250 narzędzi. Część z nich jest osiągalna w trybie demonstracji, tzn. istnieje możliwość przetestowania działania tych narzędzi poprzez strony internetowe, bez konieczności ich instalacji na swoim komputerze. W skład narzędzi dostępnych przez platformę demonstracyjną CK IMPACT wchodzi m.in.: narzędzia do obróbki obrazów (np. usuwanie krawędzi), segmentacji dokumentów (np. rozpoznawanie grafik, tekstu) czy wyspecjalizowane silniki rozpoznawania tekstu OCR (ang. Optical Character Recognition).

W kontekście długoterminowego przechowywania danych CK IMPACT udostępnia szczegółowe rekomendacje, opublikowane w 2014 r.<sup>13</sup>. Rekomendacje dotyczą standardów i formatów związanych z procesem digitalizacji, włączając w to rekomendacje dotyczące formatów do długoterminowego przechowywania dokumentów w postaci plików graficznych i tekstowych. Podsumowanie rekomendacji w zakresie długoterminowego przechowywania danych przedstawia Tab.1. Wskazany w tabeli format rekomendowany powinien być rozważony w pierwszej kolejności, natomiast format alternatywny może być wykorzystany tylko wtedy, gdy nie ma możliwości zastosowania formatu rekomendowanego (np. ze względu na koszty wytworzenia). Szczegóły co do właściwości poszczególnych formatów, znajdują się w obszernym opracowaniu udostępnionym na stronie internetowej CK IMPACT<sup>14</sup>.

Tabela 1. Rekomendacje CK IMPACT dot. długoterminowego przechowywania danych

	<b>Rekomendowany format</b>	<b>Alternatywny format</b>
Pliki wzorcowe dla dokumentów graficznych	TIFF	JPEG2000 (JP2)

<sup>12</sup> *Tools for text digitisation* [on-line]. The Impact Centre of Competence in Digitisation [dostęp 9.01.2015]. Dostępny w: <http://www.digitisation.eu/tools-resources/tools-for-text-digitisation/>.

<sup>13</sup> *Recommendations for digitisation projects* [on-line]. The ImpactCentre of Competence in Digitisation [dostęp 9.01.2015]. Dostępny w: <http://www.digitisation.eu/training/recommendations-for-digitisation-projects/>.

<sup>14</sup> *Succeed. Recommendations for metadata and data formats for online availability and long-term preservation* [on-line]. [dostęp 9.01.2015]. Dostępny w: [http://www.digitisation.eu/download/Succeed\\_600555\\_D4.1\\_RecommendationsOnFormatsAndStandards\\_v1.1.pdf](http://www.digitisation.eu/download/Succeed_600555_D4.1_RecommendationsOnFormatsAndStandards_v1.1.pdf).

Pliki wzorcowe dla dokumentów tekstowych	TEI, PDF/A	Plik tekstowy w formacie UTF-8
Format metadanych opisowych	DCMES, MODS	MARC21
Format metadanych strukturalnych	METS	Brak
Format metadanych administracyjnych	PREMIS, MIX, TextMD	Brak
Format rezultatów OCR	ALTO, PAGE	Plik tekstowy w formacie UTF-8

Źródło: [dostęp 9.01.2015].

[http://www.digitisation.eu/download/Succeed\\_600555\\_D4.1\\_RecommendationsOnFormatsAndStandards\\_v1.1.pdf](http://www.digitisation.eu/download/Succeed_600555_D4.1_RecommendationsOnFormatsAndStandards_v1.1.pdf)

## Digital Preservation Coalition

Digital Preservation Coalition (DPC), to międzynarodowa organizacja, która upowszechnia praktyki związane z przechowywaniem danych oraz odgrywa rolę katalizatora i rzecznika dla wszelkich działań w tym obszarze.

Kluczowe działania DPC to m.in.:

- konsultacje i doradztwo — w ramach tego działania dostarczanych jest szereg informacji związanych z problematyką przechowywania danych. Między innymi dostępny jest podręcznik na temat przechowywania danych czy raporty techniczne, które w zaawansowany sposób przedstawiają specyficzne problemy przechowywania danych;
- upowszechnianie informacji — w ramach działań w tym obszarze DPC dostarcza bazę wiedzy na temat aktualnych badań i analiz prowadzonych w kontekście zmieniających się uwarunkowań procesu długoterminowego przechowywania danych. DPC przyznaje również nagrody, wyróżniające narzędzia lub usługi, które w sposób szczególny przysłużyły się środowisku;
- szkolenia oraz organizacja wydarzeń — ten obszar dotyczy organizacji szkoleń dla zainteresowanych instytucji (np. w postaci jednodniowych spotkań w danej instytucji) czy organizacji kilkudniowych szkoleń dla szerszego grona zainteresowanych uczestników (na zasadzie seminarium).

DPC uczestniczy lub uczestniczyło w szeregu projektów związanych z długoterminowym przechowywaniem danych, m.in. APARSEN, 4C, SPRUCE czy TIMBUS. Większość członków DPC pochodzi z Wielkiej Brytanii. Niemniej wśród członków DPC są również tak znaczące instytucje jak ONZ czy NATO. Do DPC dołączają również podmioty komercyjne, np. jednym z nich jest Lloyds Banking Group.



## Inicjatywy krajowe

W Polsce istnieje pięć centrów kompetencji w zakresie digitalizacji obiektów dziedzictwa kulturowego — do spraw materiałów bibliotecznych, muzealnych, audiowizualnych, archiwalnych oraz zabytków. Centra te prowadzą różnego rodzaju działania wspierające procesy digitalizacji i przechowywania danych, również związane z długoterminowym przechowywaniem. W obszarze digitalizacji, długoterminowego przechowywania i udostępniania obiektów cyfrowych w internecie aktywne działania prowadzi również od wielu lat Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS). Wybrane działania opisano w dalszej części niniejszego rozdziału.

### Projekt „E-muzea”

Narodowy Instytut Muzealnictwa i Ochrony Zbiorów (NIMOZ) przygotowuje projekt „E-muzea”, przeznaczony dla sektora muzealnego. Projekt jest przygotowywany do programu Cyfrowa Polska w ramach *II Osi, Priorytet e-Administracja i otwarty rząd, Cel szczegółowy 4: Zwiększenie dostępności i wykorzystania informacji sektora publicznego*. Jednym z trzech głównych celów tego projektu będzie długoterminowe zabezpieczanie zasobów cyfrowych wytworzonych podczas procesu digitalizacji. Zabezpieczanie tych zasobów będzie realizowane przez dedykowane repozytorium cyfrowych zasobów muzealnych, które składać się będzie przede wszystkim z:

- infrastruktury sprzętowej, która umożliwi przechowywanie danych cyfrowych w sposób bezpieczny i niezawodny,
- oprogramowania repozytorium, które będzie kontrolować zgromadzone zasoby, włączając w to proces pozyskiwania danych czy ich weryfikacji.

Poza funkcjami długoterminowego zabezpieczania danych, projekt „E-muzea” będzie prowadził działania w zakresie zwiększania dostępności zasobów muzealnych w formie elektronicznej oraz podnoszenia jakości dokumentacji tychże zasobów. Cele te będą realizowane poprzez digitalizację wybranych kolekcji muzealnych oraz budowę portalu internetowego, który będzie udostępniał cyfrowe wizerunki zdigitalizowanych i zarchiwizowanych obiektów.

NIMOZ udostępnia także szereg zaleceń związanych z procesem digitalizacji. W zakresie długoterminowej archiwizacji ciekawym opracowaniem jest koncepcja systemu trwałych identyfikatorów internetowych dla obiektów dziedzictwa kulturowego[5]. Autor opracowania proponuje trzyczłonowy identyfikator, składający się ze stałego przedrostka *plhoi* (Polish Heritage Identifier), pięciocyfrowego identyfikatora instytucji oraz identyfikatora obiektu. Ponadto w opracowaniu znajduje się propozycja podstawowych funkcji usługi zarządzania identyfikatorami oraz usługi do rozpoznawania tychże identyfikatorów.

## **dArceo i pakiet oprogramowania DInGO**

Poznańskie Centrum Superkomputerowo-Sieciowe, będące operatorem krajowej naukowej sieci *PIONIER — Polski Internet Optyczny* i *Platformy Obsługi Nauki PLATON*, od wielu lat rozwija ogólnopolską infrastrukturę bibliotek cyfrowych[6]. W dużej mierze dzięki tym działaniom obecnie w Polsce istnieje już ponad 100 bibliotek cyfrowych oraz kilka muzeów cyfrowych, które udostępniają użytkownikom internetu ponad dwa miliony obiektów cyfrowych. Tak ogromna skala działań digitalizacyjnych przyniosła w efekcie potrzebę opracowania dedykowanego oprogramowania, które usprawni, zorganizuje i uporządkuje czynności związane z digitalizacją. Działania w tym zakresie zostały podjęte w ramach współpracy PCSS i konsorcjum projektu *Repozytorium Cyfrowe Instytutów Naukowych PAN*<sup>15</sup>. Rezultatem tej współpracy jest oprogramowanie *dLab*, które służy do kompleksowego zarządzania procesem digitalizacji. Ponadto, w ramach prac badawczo-rozwojowych projektu *SYNAT*<sup>16</sup>, PCSS podjęło się opracowania prototypowego zestawu usług dla długoterminowego przechowywania danych źródłowych.

Na bazie tych prac powstał system *dArceo*, zapewniający długoterminowe przechowywanie danych cyfrowych w formatach graficznych, audiowizualnych oraz tekstowych. Warto podkreślić, że oprogramowanie *dArceo* przygotowane zostało w zgodzie z uznanym światowym standardem OAIS[7] oraz, że współpracuje ono z polskim krajowym magazynem danych — usługą powszechnej archiwizacji *PLATON U4*[8]. Oprogramowanie *dLab*, *dArceo* oraz *dLibra* wchodzi w skład tzw. pakietu oprogramowania *DInGO* (ang. „Digitize and Go”), oferującego kompleksowe rozwiązanie dla profesjonalnego i zautomatyzowanego procesu digitalizacji, przechowywania i udostępniania cyfrowych zasobów<sup>17</sup>.

## **Podsumowanie**

Z roku na rok przybywa coraz więcej dokumentów cyfrowych, zarówno tych, które powstają w efekcie digitalizacji, jak i tych powstających od razu w cyfrowej formie. Z tego względu problematyka długoterminowego przechowywania danych cyfrowych staje się z roku na rok coraz ważniejsza i kluczowa dla dostępności materiałów cyfrowych w przyszłości. W kontekście międzynarodowym istnieje szereg inicjatyw zajmujących się tematyką przechowywania i udostępniania cyfrowych zasobów. Przykładami mogą być *Open Preservation Foundation*, *Centrum Kompetencji IMPACT* w zakresie digitalizacji, czy *Digital Preservation Coalition*.

Również na arenie krajowej prowadzone są działania, mające na celu przechowywanie i zabezpieczanie danych. Można tutaj wymienić m.in. Projekt „E-muzea” czy propozycję *NIMOZ*

---

<sup>15</sup> Repozytorium Cyfrowe Instytutów Naukowych PAN [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://rcin.org.pl/>.

<sup>16</sup> Projekt SYNAT finansowany przez Narodowe Centrum Badań i Rozwoju (nr umowy: SP//I/77065/10).

<sup>17</sup> DInGO, [on-line]. [dostęp 9.01.2015]. Dostępny w: <http://dingo.psnc.pl/>.

w zakresie systemu trwałych identyfikatorów obiektów dziedzictwa kulturowego. Poznańskie Centrum Superkomputerowo-Sieciowe, które od lat wspiera instytucje kultury i nauki w zakresie digitalizacji i udostępniania zbiorów, również działa aktywnie na rzecz długoterminowej archiwizacji poprzez rozwój oprogramowania dArceo oraz prace badawczo-rozwojowe prowadzone w tym obszarze.

#### Bibliografia:

1. MAZUREK, C., WERLA, M., Federacja Bibliotek Cyfrowych – studium przypadku. W: Biblioteki cyfrowe, red. M. Janiak, M. Krakowska, M. Próchnicka, Warszawa: SBP, 2012, s. 449. ISBN 978-83-61464-70-9.
2. FARIA, L., BECKER, C., DURETEC, K., FERREIRA, M., RAMALHO, J.C. Supporting the preservation lifecycle in repositories. W: Open Repositories 2013, Charlottetown, 8-12.07.2013.
3. KULOVATIS, H., RAUBER, A., KUGLER, A., BRANTL, M., BEINERT, T., SCHOGER, A. From TIFF to JPEG 2000?. W: D-Lib Magazine. 2009. t. 15, nr 11/12.
4. MAZUREK, C., WERLA, M., PARKOŁA, T. dArceo: Usługi długoterminowego przechowywania danych źródłowych. W: Wokół zagadnień trwałej ochrony zasobów cyfrowych. Red. A. Januszko-Szakiel, Kraków 2013, s. 101.
5. WERLA, M., Koncepcja systemu trwałych identyfikatorów internetowych dla obiektów dziedzictwa kulturowego [online], [dostęp 09.01.2015]. Dostępny w: [http://digitalizacja.nimoz.pl/uploads/zalaczniki/system\\_trwalych\\_identyfikatorow\\_NIMOZ\\_2013.pdf](http://digitalizacja.nimoz.pl/uploads/zalaczniki/system_trwalych_identyfikatorow_NIMOZ_2013.pdf).
6. PARKOŁA, T. MAZUREK, C., WERLA, M. Rozwój bibliotek cyfrowych w Polsce : IX Ogólnopolska Konferencja Bibliotek Szkół Wyższych Niepublicznych "Partnerzy bibliotek", 20 – 21.09. 2007. Firma Księgarska Wiesława Juszczyka, Łódź, 2008. ISBN 978-83-925616-0-6. „Partnerzy bibliotek. Model komunikacji z otoczeniem”, s. 239-246.
7. Reference Model for an Open Archival Information System (OAIS). Magenta Book. Wyd. 2, 2012 [online], [dostęp 17.12.2014]. Dostępny w: <http://dlib.org/dlib/november09/kulovits/11kulovits.html>.
8. BRZEŹNIAK, M., MEYER, N., MIKOŁAJCZAK, R., JANKOWSKI, G., JANKOWSKI, M., Popular Backup/Archival Service and its Application for the Archival of the Network Traffic in the PIONIER Academic Network, W: Computational Methods in Science and Technology, 2010, Special Issue, s. 109.

---

Parkoła, T. Długoterminowe przechowywanie cyfrowego dziedzictwa kulturowego. *Biuletyn EBIB* [on-line] 2014, nr 9 (154), Gromadzenie i zabezpieczanie danych cyfrowych. [Dostęp 20.01.2015]. Dostępny w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/303>. ISSN 1507-7187.