

# Marek Zieliński, Iwona Korga

---

## Zbiory multimedialne Instytutu Józefa Piłsudskiego w Ameryce i ich digitalizacja

---

Przegląd Archiwalny Instytutu Pamięci Narodowej 6, 317-326

---

2013

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach  
dozwolonego użytku.

Marek Zieliński, Iwona Drąg-Korga

## ZBIORY MULTIMEDIALNE INSTYTUTU JÓZEFA PIŁSUDSKIEGO W AMERYCE I ICH DIGITALIZACJA

**I**nstytut Józefa Piłsudskiego w Ameryce powstał w 1943 r. w Nowym Jorku z inicjatywy wybitnych Amerykanów polskiego pochodzenia i emigrantów wojennych z Polski<sup>1</sup>. W tym roku przypada 70. rocznica powołania do życia jednej z obecnie największych polskich instytucji poza granicami Rzeczypospolitej.

Początki Instytutu były trudne: garstka ludzi i niewielkie fundusze. Podstawowe znaczenie miała wizja twórców i program działania Instytutu. Elitę polityczną placówki tworzyli piłsudscy wywodzący się z okresu II Rzeczypospolitej oraz miejscowi – niepodległościowi działacze Polonii. W pierwszej grupie znaleźli się przedwojenni ministrowie: Waław Jędrzejewicz – minister wyznań religijnych i oświecenia publicznego, Ignacy Matuszewski – minister skarbu, Henryk Floyar-Rajchman – minister przemysłu. Grupę polonijną tworzyli: Franciszek Januszewski – wydawca „Dziennika Polskiego” w Detroit, Maksymilian Węgrzynek – wydawca „Nowego Świata” w Nowym Jorku i Lucjan Kupferwasser z Chicago<sup>2</sup>.

W swojej działalności władze Instytutu kierowały się misją określoną przez statut. W 1943 r. wśród głównych zadań ośrodka wymieniono: pozyskiwanie archiwaliów dotyczących okresu po powstaniu styczniowym 1863 r., wydawanie publikacji, popieranie badań nad najnowszą historią Polski oraz współpracę z instytucjami amerykańskimi i polskimi na tym polu. Myślą przewodnią twórców i działaczy Instytutu było początkowo stworzenie organizacji będącej kontynuacją założonego w Warszawie w 1923 r. Instytutu Badań Najnowszej Historii Polski, przemianowanego w 1936 r. na Instytut Józefa Piłsudskiego. Po konferencji w Jalcie (luty 1945 r.), gdy zdano sobie sprawę, że nie istnieje możliwość przeniesienia Instytutu do Polski, kierownictwo podjęło wysiłek prowadzenia niezależnej placówki naukowo-badawczej na terenie Stanów Zjednoczonych<sup>3</sup>.

<sup>1</sup> Sprawozdanie z działalności Instytutu od chwili powstania (4 VII 1943 r.) do 31 XII 1943 r., Nowy Jork 1944.

<sup>2</sup> Szerzej zob. Sprawozdania i biuletyny Instytutu 1943–2012.

<sup>3</sup> *Instytut Józefa Piłsudskiego w Ameryce i jego zbiory*, oprac. J. Cisek, Warszawa 1997, s. 18–19.

## Zasoby Instytutu

Na zasoby archiwalne Instytutu Józefa Piłsudskiego w Ameryce składają się kolekcje dokumentów archiwalnych, fotografie, zbiór kartograficzny, biblioteka, materiały audiowizualne, kolekcja muzealna oraz galeria polskich obrazów. Najbardziej znane są kolekcje dokumentów pisanych i drukowanych. Najcenniejszą część stanowią materiały ze zbiorów Instytutu Józefa Piłsudskiego w Warszawie oraz Wojskowego Biura Historycznego, uratowane z płonącej Warszawy we wrześniu 1939 r. Zbiory te stanowią 20 proc. zasobu, reszta to materiały instytucji polonijnych i wybitnych polskich dyplomatów, wojskowych, działaczy polonijnych<sup>4</sup>.

Najmniej znaną częścią archiwum Instytutu są kolekcje: fotografii i audiowizualna – warto więc omówić je szerzej.

## Fotografie

Kolekcja fotografii (sygn. 701/151) obejmuje zdjęcia z okresu od 1907 r. aż do czasów współczesnych i liczy obecnie ponad 20 100 j.a. Powstała ona dzięki donacjom organizacji polonijnych oraz instytucji Rządu RP na Uchodźstwie. W pierwszej grupie najbardziej zasłużył się Komitet Obrony Narodowej, Komitet Narodowy Amerykanów Polskiego Pochodzenia i Rada Polonii Amerykańskiej. Część przedwojennych fotografii trafiła z Wojskowego Instytutu Naukowo-Wydawniczego i Wydziału Ilustracji Prasowej Polskiej Agencji Telegraficznej. Po 1945 r. do archiwum Instytutu zostały przekazane zbiory fotograficzne zgromadzone przez placówki rządu emigracyjnego, np. Polskie Centrum Informacyjne w Nowym Jorku. Warto odnotować prywatne darowizny, m.in. znanego historyka Polonii – Artura Waldo, dotyczące życia organizacyjnego Polonii, czy kolekcję zdjęć związanych z lotnictwem polskim kpt. Marka Mażyńskiego z Kalifornii. Ciekawe są także zbiory związane z okresem „Solidarności”.

Kolekcja fotografii jest prawie w całości opracowana dzięki wysiłkowi zawodowych archiwistów z Narodowego Archiwum Cyfrowego, Archiwum Akt Nowych i Archiwum Państwowego w Łomży<sup>5</sup>. Została ona podzielona na serie: dział osobowy, zbiór fotografii patrona Instytutu – Józefa Piłsudskiego, okres II Rzeczypospolitej, okres II wojny światowej, demobilizacja, Polska i świat po 1945 r., „Solidarność” oraz architektura i Polonia<sup>6</sup>.

Z kolekcji fotograficznej korzystają badacze poszukujący ilustracji do powstających książek, albumów oraz dokumentaliści filmowi zainteresowani wykorzystaniem zdjęć w filmach. W 2011 r. Instytut Józefa Piłsudskiego w Ameryce i Wojskowe Centrum Edukacji Obywatelskiej wydały publikację poświęconą wojnie polsko-bolszewickiej, gdzie wykorzystano mało znany zbiór zdjęć przechowywany w kolekcji fotograficznej<sup>7</sup>.

<sup>4</sup> *Informator o zasobie archiwalnym Instytutu Józefa Piłsudskiego w Ameryce*, oprac. P. Pietrzyk, Warszawa 2011.

<sup>5</sup> W okresie od 2005 r. do 2012 r. na stażach zawodowych przebywali: Filip Kwiatek, Danuta Bzura, Mariusz Olczak i Bartosz Nowożycki.

<sup>6</sup> *Informator o zasobie archiwalnym Instytutu Józefa Piłsudskiego...*, s. 240–241.

<sup>7</sup> *Wojna polsko-bolszewicka. Komunikaty wojenne Sztabu Generalnego, dokumenty, fotografie*, oprac. K. Paduszek, Warszawa 2011.

Kolekcja fotografii, w większości już opracowana, jest zdigitalizowana tylko w małej części. Proces ten rozpoczęto od fragmentów kolekcji, które mogłyby posłużyć jako ilustracja do zdigitalizowanych dokumentów – jako pierwsze wytypowano materiały z serii „II wojna światowa”. Istnieje jednak pilna potrzeba dalszej digitalizacji i udostępnienia kolekcji fotografii, oczywiście z poszanowaniem praw autorskich. Kolejne prace nad digitalizacją tego zespołu zależeć będą głównie od możliwości finansowych i zdobytych na ten cel funduszy.

### **Kolekcja audiowizualna**

Kolekcja audiowizualna (sygn. 701/152) powstała przede wszystkim dzięki prywatnym przekazom oraz działaniom podejmowanym przez pracowników Instytutu. Na kolekcję składają się kasety VHS, taśmy magnetofonowe, płyty DVD oraz taśmy filmowe 16 mm i 35 mm.

Zbiór kaset VHS zawiera wywiady z osobami, które w okresie 1940–1941 zostały wywiezione ze wschodnich obszarów Polski do łagrów w Rosji Sowieckiej. Nagrań dokonano w latach 1989–1991 dzięki grantowi Fundacji Barbary Piaseckiej-Johnson. W sumie Instytut posiada 56 nagrań, które w 2009 r. zostały przeniesione na pliki cyfrowe, a potem na DVD<sup>8</sup>. W zbiorze kaset znajdują się także nagrania z wykładów w Instytucie z początku lat dziewięćdziesiątych XX w. dokonane przez „Video-Fakt” i Mariana Faryniarza. W 2011 r. do Instytutu trafiło kilkadziesiąt kaset VHS z wywiadami ze świadkami II wojny światowej nagranych przez Bożennę Urbanowicz-Gilbride.

W kolekcji Instytutu znajdują się także nagrania audycji Radia Wolna Europa z lat 1962–1983. Jest to 219 taśm magnetofonowych, które stanowią część kolekcji RWE (sygn. 701/112) liczącej 175 j.a. Kolekcja ta posiada inwentarz, jednak ze względu na brak odpowiedniego sprzętu do odtwarzania nie ma możliwości odsłuchania taśm.

Dopełnieniem zbioru są płyty DVD, które są oryginalnymi wydaniem znanych filmów i seriali. Na przeniesienie do formatu cyfrowego czekają taśmy magnetofonowe, na których zarejestrowano audycje, wywiady oraz wykłady odbywające się w Instytucie Józefa Piłsudskiego, w tym prof. Wacława Jędrzejewicza dla „Grupy Młodych” w latach 1981–1982<sup>9</sup>.

### **Kolekcja filmów**

W archiwum Instytutu Józefa Piłsudskiego przez całe lata przechowywane były taśmy filmowe 16 mm i 35 mm. Spis taśm zawiera 36 sygnatur, razem 82 rolki filmowe w pudełkach metalowych – są to zarówno kopie pozytywowe, jak i negatywy.

Filmy są w większości kopiami, dostępnymi w oryginale w Polsce, jak np. „Śmierć prezydenta”, „Życie Piłsudskiego”, „Pogrzeb marszałka Piłsudskiego”.

W 2011 r. materiały filmowe zostały przekazane do konserwacji i digitalizacji do Instytutu Pamięi Narodowej, gdzie będą przechowywane. Wszystkie taśmy na podkładzie

<sup>8</sup> Konwersji dokonał dr Marek Zieliński, format MPEG2, NTSC, 24 bity 720 × 480.

<sup>9</sup> Skrypty wykładów i nagrania zostały wydane w: W. Jędrzejewicz, *Rola Józefa Piłsudskiego w odbudowie i umacnianiu państwa polskiego*, Warszawa 2013.



niepalnym zostały zdigitalizowane przez IPN i kopie są już w archiwum Instytutu Piłsudskiego. Taśmy filmowe na podkładzie palnym (28 j.a.) – nitrocelulozowym – zostały przekazane do Filмотeki Narodowej, która jako jedyna w Polsce dysponuje specjalnymi magazynami do przechowywania tego rodzaju nośnika. Z czasem filmy te zostaną zdigitalizowane i przekazane Instytutowi w Ameryce.

Archiwum Instytutu nie dysponuje odpowiednimi warunkami do przechowywania taśm filmowych, tj. chłodnią z kontrolowaną temperaturą.

Od wielu lat Instytut Józefa Piłsudskiego prowadzi – jako zadanie priorytetowe – digitalizację zbiorów, które udostępnia w postaci elektronicznej.

### **Digitalizacja**

Od 2007 r. Instytut Józefa Piłsudskiego systematycznie działa w kierunku zabezpieczenia i udostępnienia materiałów archiwalnych w formie cyfrowej. W odróżnieniu od oryginałów zarejestrowanych metodą analogową na nośnikach takich jak papier, taśma filmowa lub magnetyczna, zapis cyfrowy nie ulega starzeniu. To pozornie paradoksalne stwierdzenie wynika z tego, że zapis ten – w odróżnieniu od tradycyjnego – można kopiować nieograniczoną liczbę razy bez strat, a więc informacja tak zapisana nie zależy od czasu życia nośnika.

Należy podkreślić, iż digitalizacja to nie jest jedynie skanowanie dokumentów. W rzeczywistości proces digitalizacji nie zaczyna się ani nie kończy na skanowaniu, które jest tylko jednym z wielu jego etapów. Zespół archiwalny musi być najpierw opracowany i uporządkowany, jeśli zachodzi taka potrzeba. Kolekcja mogła być pozyskana w stanie uporządkowanym lub może wymagać znacznej pracy nad jej organizacją. W tym etapie mieści się także stworzenie hierarchii archiwalnej, tj. podział zasobu na serie i jednostki. To zadanie wykonują w Instytucie zazwyczaj profesjonalni archiwiści. Następnym krokiem jest spaginowanie (w przypadku dokumentów i zdjęć na papierze). Paginacja ma dwa zasadnicze zadania: inwentarzowe (kolejna numeracja zapewnia zachowanie całości zbioru i ułatwia odnalezienie braków) i identyfikacyjne – przy skanowaniu notuje się



lub umieszcza w nazwie pliku dane pozwalające na jednoznaczną identyfikację konkretnej strony – jej adres w systemie archiwalnym.

Dopiero tak przygotowany materiał może zostać zeskanowany. Oczywiście mamy tu na myśli konwersję cyfrową, tj. z obiektu, w którym informacja zapisana jest w postaci analogowej, stworzenie kopii, w której ta sama informacja zapisana jest w formie cyfrowej (albo inaczej dyskretnej). W przypadku dokumentów na papierze czy innym płaskim nośniku, zdjęć i obrazów, proces ten nazywamy skanowaniem, choć można również użyć fotografii cyfrowej, a więc zamiast stopniowej konwersji linia po linii uzyskać cyfrową kopię całego obiektu jednocześnie. Konwersja filmu na taśmie filmowej wygląda podobnie, z tym że trzeba wykonać kopie każdej klatki. Konwersja cyfrowa zapisu na taśmie magnetycznej (audio lub wideo) odbywa się nieco inaczej – zapis w postaci pola magnetycznego na taśmie najpierw jest zamieniany na analogowy sygnał elektryczny, a ten jest przerabiany na cyfrowy za pomocą przetwornika analogowo-cyfrowego. W każdym przypadku otrzymujemy zapis cyfrowy – ciąg zerojedynkowy w żargonie informatycznym – który używając jakiegoś kodowania, zapisuje odzwierciedlenie oryginalnego obiektu.

Jeśli chodzi o kodowanie, dla dokumentów stosujemy skany w formacie TIFF, w kolorze i rozdzielczości skanera 600 dpi. Te same parametry stosujemy do zdjęć dużych rozmiarów, zwiększając rozdzielczość w przypadku odbitek stykowych lub negatywów. Kolejnymi etapami digitalizacji są: zbieranie metadanych i prezentacja albo udostępnienie danych, o czym dalej. W skrócie, na digitalizację składają się następujące etapy:

1. Opracowanie.
2. Paginacja albo oznakowanie części.
3. Skanowanie – konwersja cyfrowa.
4. Opis – zebranie metadanych.
5. Udostępnienie.

## Metadane

Aby zapoznać się z zawartością obiektu analogowego, takiego jak obraz czy pieśń, wystarczą zmysły. Książka i nuty mają dodatkową warstwę kodowania, potrzebna jest znajomość alfabetu i języka książki, języka zapisu muzycznego. Film na taśmie celuloidowej wymaga urządzenia optycznego, a w przypadku taśmy magnetycznej – urządzenia elektronicznego, aby zamienić zapis w formę dostępną naszym zmysłom. Plik w komputerze jest zawsze w formie cyfrowej i wymaga kilku warstw kodowania zanim stanie się dostępny.

Po konwersji cyfrowej otrzymujemy „nagie” odzwierciedlenie oryginału. Aby było ono użyteczne, aby można je było udostępnić, musimy zaopatrzyć je w metadane. Metadane, dosłownie „dane o danych”, są to różnego rodzaju informacje, które opisują konkretny obiekt. Nagi skan jest tylko mapą określającą kolor w każdym punkcie dwuwymiarowej powierzchni. Dopiero metadane pozwalają na znalezienie tego skanu wśród milionów innych, jego sklasyfikowanie, umieszczenie na stronie internetowej itp. Najprostsze metadane mają zwykle strukturę w formie etykiety – zawartości. Na przykład, książka *Bibuła* może być opisana następującymi metadanymi:

```
<author>Piłsudski, Józef</author>  
<title>Bibula</title>  
<publisher>Niezależna Oficyna Wydawnicza</publisher>  
<publication_year>1978</publication_year>
```

Powyższe przykładowe metadane *Bibuly* podane są za pomocą standardowego języka opisu danych zwanego XML, ale nie używają żadnego z publicznie znanych standardów metadanych.

Metadane są klasyfikowane w różne kategorie, tak aby łatwiej zorganizować wiedzę. Typowymi kategoriami metadanych są:

- Metadane opisowe – takie jak autor, tytuł, słowa kluczowe, daty, miejsca, streszczenie itp.
- Metadane techniczne – format pliku, sposób kodowania, parametry obiektu (wymiar, czas trwania) itp.
- Metadane strukturalne wyodrębniające części obiektu, takie jak rozdziały, części, ustępy, tabele itp.
- Metadane dotyczące konserwacji obiektu.
- Metadane dotyczące praw autorskich.

Warto zwrócić uwagę, że zawartość obiektu też może być częścią metadanych, tak jak np. tytuł czy jakikolwiek wyodrębniony i opisany fragment.

Aby móc przechować dane przez dłuższy czas, a także wymieniać się nimi, niesłychanie ważne jest zastosowanie właściwych standardów zapisu metadanych i konsekwentne ich stosowanie. Dlaczego standardy są takie ważne? Kilka lat temu, gdy przenoszono fragmenty archiwum instytucjonalnego Instytutu Piłsudskiego z dyskietek na twarde dyski, okazało się, że wiele dyskietek nie było już czytelnych ze względu na zaniki pola magnetycznego. Największym problemem było jednak odczytanie dokumentów zapisanych w wielu firmowych, a więc zamkniętych, formatach. Dokumenty pisane w WordStar, PCWrite, WordPerfect, Bank Street Writer, XyWrite, First Choice są dziś dostępne z dużym trudem i często konwersja powoduje utratę danych. Zapis metadanych z użyciem alfabetu, takiego jak XML – powszechnie stosowanego na całym świecie – zwiększa szanse na to, że będą one dostępne w przyszłości. Zapis z użyciem publicznie dostępnych i popularnych standardów metadanych zwiększa szanse na to, że dane będą zrozumiałe przez inne systemy i ludzi posługujących się różnymi językami naturalnymi.

Dotykamy tu ważnego, a często zaniedbywanego aspektu danych: przenośności. Bardzo często spotykamy się z opinią, że archiwiści powinni zajmować się archiwami, a dane to jest domena informatyków. W Instytucie uważamy, że w epoce komputerowej to właśnie dane są najcenniejszym zasobem archiwum, na równi z zasobami materialnymi, i należy o nie dbać z taką samą troską. Przenośność danych oznacza, że zasoby archiwalne można skopiować (na pendrive'a, przenośny dysk itp.), schować w kasie pancerniej, a co ważniejsze – otworzyć ponownie w zupełnie innym programie. Oznacza to również możliwość przeniesienia danych w przyszłość, a więc otwarcia tego zasobu za 20 lat. W tym celu używamy standardów metadanych. Po istotnym wysiłku kilku ostatnich lat Instytut Józefa Piłsudskiego dokonał konwersji i zapisał w standardach metadanych: opisy zasobów archiwalnych, katalogi biblioteczne i zasoby zdigitalizowane.

## Standardy

Nie ma jednego standardu metadanych. Istnieje ich wiele, nowych i starych, powiązanych złożonymi zależnościami. Nie jest to zaskakujące, gdyż w pewnym sensie tworzenie tych standardów odzwierciedla próby skodyfikowania i sformalizowania języka naturalnego, aczkolwiek początkowo powoduje zawrót głowy.

Jakimi kryteriami należy się kierować, wybierając standardy do naszej pracy? Wspomniano już o XML, który jest *de facto* głównym alfabetem komunikacji maszynowej i podstawą wyrażania większości nowych standardów metadanych. Wszystkie nowe standardy muszą być wyrażalne w XML (albo lepiej, być w XML wyrażone). Z tym warunkiem wybory są już łatwiejsze, gdyż dane w wielu standardach daje się automatycznie przetłumaczyć na inne (ten proces albo procedura nazywa się *crosswalk*). Oczywiście, jeśli standard wynikowy nie ma danego pojęcia występującego w źródle, nastąpią w tłumaczeniu straty. Przy wyborze standardu zapisu metadanych powinno się zwracać uwagę na to, aby standard był:

- Nowoczesny (tj. używać najnowszej wersji standardu lub najnowocześniejszego standardu).

- Popularny. To kryterium może być w sprzeczności z poprzednim; popularność standardu daje nam zapewnienie tego, że istnieje wiele narzędzi używających takiego standardu, i że wiele źródeł będzie dostępnych. Przykładowo MARC jest bardzo starym standardem opisu bibliograficznego, zastępowanym przez dwie kolejne wersje, ale ciągle powszechnie stosowanym w bibliotekach na całym świecie.

- Międzynarodowy. W wieku komunikacji elektronicznej standard, który obowiązuje tylko w jednym kraju, staje się szybko bezużyteczny.

- Adekwatny do zadania. Oznacza to, że powinien konceptualnie odpowiadać procesowi, jaki stosujemy. Na przykład w katalogu bibliotecznym adekwatny jest standard, którego centralnym konceptem jest dzieło: książka, czasopismo, artykuł. W opisie zasobów archiwalnych standard powinien rozumieć hierarchię archiwum: zespół, seria, jednostka, dokument itp.

- O odpowiedniej szczegółowości (głębokości). Standard powinien w przybliżeniu odpowiadać szczegółowości opisu, jaki stosujemy. Standard nie dość szczegółowy nie pozwoli nam na tak głęboki opis, jak wymagają tego potrzeby. Standard zbyt szczegółowy spowoduje rosnące straty czasu spowodowane koniecznością decydowania czy wybierania szczegółów, które w danym procesie nie są ważne.

- Wyrażalny w XML, o czym już wspomniano, i posiadający narzędzia do opracowania danych. Pierwsze kryterium pozwala na inspekcję i częściowo edycję danych w postaci surowej, z użyciem standardowych narzędzi do obsługi XML. Drugie – wymagane jest do efektywnej i wygodnej pracy.

W pracy Instytutu Józefa Piłsudskiego używane są trzy podstawowe standardy:

**Dublin Core**<sup>10</sup> (**Dublin Core Metadata Initiative**, DC) jest dojrzałym standardem opisu zawartości zasobów przeznaczonym do szerokiego stosowania. Można go wykorzystywać do opisu stron WWW, zdjęć, filmów, dokumentów, książek. Standard ten jest prosty, w swojej podstawowej wersji ma tylko 15 elementów („Tytuł”, „Twórca”, „Temat”, „Opis”, „Data”, „Format” i inne). Przejrzystość jest jego największą zaletą, gdyż

<sup>10</sup> Zob. <http://dublincore.org>.



pozwała na użycie go przez osoby po krótkim przeszkoleniu i nie wymaga długich studiów. Znaczenie i użycie elementów jest dobrze opisane. Standard jest płaski, tj. nie ma żadnej hierarchii elementów. Druga wersja DC wprowadza dodatkowe elementy, które pozwalają na sprecyzowanie/zawężenie pojęć (np. „Obszar pokrycia” można dookreślić jako „Przestrzenny” lub „Czasowy”) oraz na relacje między obiektami. W Instytucie Piłsudskiego używany jest w programie DSpace do opisu digitalizowanych dokumentów i innych obiektów.

**Encoded Archival Description**<sup>11</sup> (EAD) jest również dojrzałym standardem, z założenia przystosowanym do opisu hierarchicznej struktury zasobów archiwalnych. Posiada ok. 150 elementów modyfikowalnych przez ok. 100 atrybutów i jest dużo bardziej złożony od DC. Mimo to również EAD – tak jak i Dublin Core – nadaje się do ręcznej obróbki. Są archiwa, które pracują na „surowym” zapisie EAD, modyfikując go w edytorze XML. Hierarchiczność EAD pozwala na zdefiniowanie dowolnej struktury zespołu archiwalnego (*fonds*), który jest najwyższym elementem w EAD: podzespołu (*subfonds*), serii, foldera, jednostki. Zasoby archiwalne docierają do archiwum często z własną – już określoną przez twórcę archiwum – strukturą. Zasada szacunku dla oryginalnej kolekcji (*respect des fonds*) wymaga, aby tę strukturę w miarę możliwości zachować i EAD daje narzędzia do opisanie takiej struktury. Oprócz tego standard definiuje wszystkie elementy opisu, zarówno wymagane, jak i nieobowiązkowe. EAD jest używany w Instytucie wraz z programem Archivists’ Toolkit do opisu zasobów archiwalnych.

**Machine Readable Cataloging**<sup>12</sup> (MARC) jest najstarszym z trzech omawianych tu standardów, wprowadzonym już w 1971 r. i przeznaczonym do opisu książek i czasopism. Jest standardem binarnym, ale łatwo go przekształcić w czytelną dla człowieka formę. Zamiast nazw elementów występują liczby od 1 do 999 (np. 100 oznacza głównego autora dzieła), a zamiast atrybutów modyfikujących elementy występują pojedyncze litery alfabetu. Ponieważ struktura jest podobna do XML, łatwo przekształcić MARC w postać XML bez żadnych strat. Format ten, MARCXML, nie jest jednak zbyt szeroko stosowany, gdyż wszystkie znaczące programy do katalogowania i obróbki danych bibliograficznych czytają pliki i je zapisują w formacie MARC. Jest to standard o bardzo dużym zasobie danych i stosują go wszystkie (przynajmniej te większe) biblioteki na świecie. W Instytucie Piłsudskiego używany jest MARC wraz z programem bibliotecznym Koha do katalogowania zasobów bibliotecznych<sup>13</sup>.

<sup>11</sup> Standard archiwalny stworzony i utrzymywany przez Towarzystwo Amerykańskich Archiwistów i Bibliotekę Kongresu; zob. <http://www.loc.gov/ead>.

<sup>12</sup> Standardy biblioteczne MARC są utrzymywane przez Bibliotekę Kongresu; zob. <http://www.loc.gov/marc>.

<sup>13</sup> Oprócz wymienionych jest jeszcze wiele innych standardów, niektóre z nich przydatne przy opracowywaniu zasobów multimedialnych. MODS (Metadata Object Description Schema) – unowocześniona pochodna MARC – przeznaczony do opisu różnego rodzaju obiektów. TEI (Text Encoding Initiative) jest standardem szczegółowego opisu zawartości dzieła, pozwalając na oznakowanie wszystkich jego elementów składowych. VRA Core (VRA – Visual Resources Association) to standard opisu dzieł kultury wizualnej. Odróżnia on obiekt (np. budynek, pomnik, malowidło, fotogram) od jego dwuwymiarowego wizerunku i definiuje narzędzia (elementy i atrybuty) do opisu obu rodzajów obiektów, jak również ich kolekcji. NISO MIX (NISO – National Information Standards Organization) jest standardem technicznego opisu obrazów cyfrowych. Wszystkie te standardy wyrażone są w XML. Oprócz tego istnieje wiele standardów opisowych definiujących sposób opisywania i katalogowania obiektów. Ponadto jest też standard zapisu danych konserwacji i wiele innych.

## Blżej kraju

Ostatnim etapem procesu jest udostępnianie zdigitalizowanych zasobów. Zadaniem każdego archiwum jest zachowanie i konserwacja zbiorów i ich udostępnianie. Oprócz klasycznego dostępu do dokumentów udzielanego badaczom na miejscu w archiwum, w dobie Internetu bardzo skuteczną metodą jest pokazywanie ich na stronie WWW. Nie tylko można dotrzeć do większej liczby czytelników, lecz także – dzięki technikom cyfrowym – istotnie zwiększyć możliwości znalezienia poszukiwanych dokumentów.

The screenshot shows the 'Kolekcje Archiwalne Online' interface. At the top, it displays 'Teczka: 701/1/4' and a 'Przejdź' button. Below this is a table with three rows: 'Zbiór' (Archiwum Józefa Piłsudskiego), 'Teczka' (Dokumenty dotyczące Aleksandry Piłsudkiej, korespondencja z H.F. Rajchmanem), and 'Opis teczek' (1940-1953, karta kultury, administracja składowa, notatniki, rękopisy, itp.).

Below the table, it indicates 'Dokument: strona początkowa 1 strona końcowa 74' with navigation icons. The 'Opis dokumentu' section contains a table with the following data:

Tytuł	Korespondencja Aleksandry Piłsudkiej z Henrykiem F. Rajchmanem
Autor	Rajchman, Henryk Thoma; Piłsudka, Aleksandra
Typ	List
Osoby	Piłsudka, Aleksandra; Rajchman, Henryk Thoma
Data	1940-12-27; 1953-01-12
Języki	pl
Data dokumentu	1940-12-27

Below the table, there is a section 'Skany dokumentów' with three thumbnail images of handwritten letters and a postcard.

Instytut Józefa Piłsudskiego w Ameryce podjął się zaplanowanego na wiele lat procesu digitalizacji swoich zbiorów. Od kilku lat czynne są już dwie strony dokumentów *on-line* – jedna poświęcona całkowicie archiwaliom powstań śląskich, druga – zawierająca rosnący zestaw zespołów archiwalnych w całości oddawanych do dyspozycji użytkownikom w Internecie. Zainteresowanie zasobami *on-line* jest fenomenalne i rośnie w miarę dodawania nowych zespołów. Ponad jedną trzecią wszystkich wejść na stronę Instytutu stanowią wizyty w podstronach zdigitalizowanych zasobów archiwalnych. W roku 2012 mieliśmy ponad 10 tys. wizyt na naszej stronie *on-line* i ponad 8 tys. na stronie „Powstania Śląskie”. Ciekawe jest także obejrzenie rozkładu geograficznego wizyt. Najwięcej (75 proc.) pochodzi z Polski, pozostałe zapytania wpływają z całego świata.

Do digitalizacji potrzebne są trzy elementy: chęci, wiedza i umiejętności oraz finansowanie. Instytut w dużym stopniu wykorzystuje umiejętności wolontariuszy, zarówno

na miejscu, jak i zdalnie. Zasób „Powstania Śląskie” jest pierwszym projektem Instytutu, w którym zastosowano *crowdsourcing*<sup>14</sup>, wykorzystując liczną grupę ochotników do opracowywania archiwów powstańców. Instytut Piłsudskiego jest placówką niezależną, finansowaną przez składki członkowskie, hojne zapisy i darowizny Polonii amerykańskiej, jak również przez instytucje polonijne, takie jak Polski Związek Narodowy czy Polsko-Słowiańska Federalna Unia Kredytowa, które nie tylko pomagają go utrzymać, ale łożą na sprzęt i wyposażenie potrzebne do konserwacji i digitalizacji. Instytut był również odbiorcą grantów przeznaczonych na konkretne zadania, jak wspomniany grant Fundacji Barbary Piaseckiej-Johnson czy National Endowment for the Humanities. Przez ostatnie 12 lat współpracuje z polskimi instytucjami kultury, a w szczególności z Naczelną Dyrekcją Archiwów Państwowych, Biblioteką Narodową i Instytutem Pamięci Narodowej. Instytucje te przysyłają archiwistów i bibliotekarzy, którzy uczestniczą w porządkowaniu archiwów, katalogowaniu biblioteki, a także pomagają w digitalizacji. Od 2005 r. Instytut jest odbiorcą grantów Senatu RP i Ministerstwa Spraw Zagranicznych RP, skierowanych m.in. na digitalizację wybranych zasobów.

---

<sup>14</sup> Proces, w którego ramach organizacja (firma, instytucja publiczna, organizacja *non-profit*) przeprowadza *outsourcing* zadań wykonywanych tradycyjnie przez pracowników do niezidentyfikowanej, zwykle bardzo szerokiej grupy ludzi w formie *open call* (ang. *crowd* – tłum, *sourcing* – czerpanie źródła). *Crowdsourcing* umożliwia wszystkim użytkownikom Internetu partycypację w zadaniach.