

Joanna Koźmińska

POLSKI INSTYTUT ANTROPOLOGII,
UNIwersytet WARSZAWSKI

Cyfrowa etnografia.
O projekcie antropologicznej biblioteki cyfrowej

Społeczeństwo informacyjne i nieustający rozwój nowych technologii inspiruje wiele przedsięwzięć inicjowanych na różnych polach nauki, w tym etnologicznej i antropologicznej. W ostatnich latach powstało kilka bardzo ciekawych i nowatorskich projektów związanych z upowszechnieniem technik komputerowych oraz treści etnograficznych w Internecie wśród badaczy zajmujących się etnologią i antropologią kulturową. Projekty te, mające wspomagać warsztat naukowy, obejmują różne aspekty działań etnografów, folklorystów, antropologów, poczynając od pionierskiego projektu Bibliografii Etnografii Polskiej (Ośrodek Dokumentacji i Informacji Etnograficznej, Polskie Towarzystwo Ludoznawcze, Łódź), programu komputerowej analizy tekstów etnograficznych (program EdEt – Edytor Etnograficzny, Instytut Etnologii i Antropologii Uniwersytetu Warszawskiego, Warszawa), projektu digitalizacji tekstów folkloru (Cyfrowe Archiwum Folkloru, Uniwersytet Mikołaja Kopernika, Toruń), poprzez digitalizację archiwum kolbergowskiego (Instytut im. Oskara Kolberga w Poznaniu), aż po próbę stworzenia antropologiczno-etnologicznej biblioteki cyfrowej (Polski Instytut Antropologii, Warszawa). Te niezwykle cenne i ciekawe inicjatywy znajdują się w różnym stopniu realizacji – Bibliografia Etnografii Polskiej płynnie przeszła od wersji papierowej do wersji *on-line*, z programu EdEt korzystają studenci i badacze IEiAK UW i innych ośrodków akademickich, Biblioteka Cyfrowa PIA znajduje się w końcowej fazie realizacji – Cyfrowe Archiwum Folkloru (CAF) boryka się z brakiem środków do podjęcia zadania (Grochowski, Mianeki: 2009).

Idea utworzenia antropologicznej biblioteki cyfrowej powstała w 2007 roku w Polskim Instytucie Antropologii (PIA), młodej instytucji pozarządowej, skupiającej środowisko etnologów i antropologów kulturowych ponad oficjalnymi strukturami akademickimi, muzealnymi czy stowarzyszeniowymi. Polski Instytut Antropologii działający na zasadach fundacji jest otwarty na wszelkie inicjatywy popularyzujące i wspierające wyżej wymienione dziedziny nauki.

Gwałtownie rosnąca ilość bibliotek cyfrowych (w 2003 roku istniała jedynie Wielkopolska Biblioteka Cyfrowa, ale już w 2009 roku aż czterdzieści dziewięć bibliotek

zostało zarejestrowanych w Federacji Bibliotek Cyfrowych) i postęp, jaki się dokonał w zakresie digitalizacji/cyfryzacji książek i czasopism, nie mogły pozostać bez udziału naszego środowiska. W trakcie wędrówek po Internecie, stwierdziliśmy jednak bardzo niewielki zasób polskich fachowych treści etnograficznych.

W kręgu etnografów najdotkliwiej odczuwany jest brak szerokiego dostępu do rodzimych czasopism etnograficznych. Czasopisma te, gromadzone w bibliotekach naukowych, uniwersyteckich czy muzealnych, rzadko wypożyczane są do domu. Skazuje to czytelnika – badacza i entuzjastę – na trud wielogodzinnego pobytu w czytelniach. Jest to sytuacja paradoksalna, w kontekście coraz lepszego dostępu do pełnotekstowych baz czasopism zagranicznych *on-line* prenumerowanych za pośrednictwem bibliotek akademickich i naukowych oraz zasobów dostępnych w ramach *Open Acces*. Biblioteka Uniwersytecka w Warszawie oferuje swoim pracownikom i studentom (po zalogowaniu w sieci uniwersyteckiej możliwy jest także dostęp w domu) ponad sto dwadzieścia tytułów czasopism w dziale antropologia. Inne ośrodki uniwersyteckie i naukowe także prenumerują bazy pełnotekstowe i udostępniają je na zasadach analogicznych do BUW.

W związku z tą sytuacją postanowiliśmy powołać do życia cyfrową kolekcję tematyczną, gromadzącą teksty z szeroko pojętego zakresu etnologii i antropologii kulturowej. Nasz projekt wpisuje się w nurt digitalizowanych zbiorów tematycznych, gromadzących jedną kolekcję, będącą efektem współpracy jednej lub kilku instytucji. Są to najczęściej inicjatywy środowiskowe, w przeciwieństwie do dużych projektów pilotowanych przez biblioteki narodowe w porozumieniu z innymi instytucjami (narodowe repozytoria cyfrowe), dążących do kompletności prezentowanych zasobów. Inną formą są zasoby cyfrowe powstające w oparciu o ideę udostępnienia starannie wyselekcjonowanych najważniejszych i najcenniejszych obiektów bibliotecznych kolekcji (Franke: 2007: 122).

Cyfrowa Kolekcja Tekstów Etnograficznych (skrótowa nazwa Cyfrowa Etnografia – CE), czyli Biblioteka Cyfrowa PIA w pierwszym etapie realizacji ma za zadanie digitalizację i publikację w Internecie trzech czasopism etnograficznych. W związku z koniecznością dokonania wyboru, po licznych dyskusjach i konsultacjach zrezygnowaliśmy z digitalizacji starych periodyków – *Wisły*, *Zbioru Wiadomości do Antropologii Krajowej* (ZWAK) oraz *Materiałów Antropologiczno-Archeologicznych i Etnograficznych* (MAAE), na rzecz czasopism ukazujących się współcześnie: *Ludu*, *Etnografii Polskiej* i *Polskiej Sztuki Ludowej-Konteksty*. Podejmując tę decyzję, mieliśmy między innymi na uwadze stale rosnącą popularność nauk antropologicznych i przydatność aktualnych treści w procesie akademickiej edukacji. Podobnie jak dla Randala Strossa „(...) perspektywa, że da się przeszukiwać treść książek z taką łatwością, z jaką się poruszamy po Sieci, była ekscytująca” (Stross 2009: 124), tak i dla nas ekscytująca i fascynująca stała się perspektywa otwarcia dostępu do polskich, aktualnie ukazujących się czasopism wielu użytkownikom rozsianym na całym świecie, w każdym miejscu i o każdej porze. Głównym naszym założeniem stało się wprowadzenie do biblioteki cyfrowej czasopism w taki sposób, aby każdy zawarty w nich artykuł był traktowany jako osobna publikacja, z pełnym opisem bibliograficznym (a nie tylko, jako część ustrukturyzowanego zbioru tomów, jak to miało miejsce dotychczas w innych polskich bibliotekach cyfrowych). Wzorowaliśmy się na zagranicznych konsorcjach typu Jstor czy Proquest. Dużym utrudnieniem jest fakt, że do tej pory brak jest w d'Librze specjalnego modułu odpowiadającego za rejestrację czasopism. W efekcie, użytkownik zmuszony jest do katalogowania całych zeszytów lub tomów, jako osobnych

obiektów sieciowych. W konsekwencji powoduje to nagromadzenie się w indeksie jednorodnych zapisów (Kolasa: 2007). Założyliśmy, że udostępnimy użytkownikom pełną zawartość czasopism z zachowaniem ich pierwotnego wyglądu. Do dyspozycji użytkownika będą narzędzia wyszukiwawcze, dające możliwość przeglądania udostępnionych czasopism *on-line*, łącznie z pobraniem tekstu na dysk własnego komputera.

Drugim przyjętym założeniem był pomysł zintegrowania biblioteki cyfrowej z nakładką umożliwiającą wyszukiwanie informacji i zarządzającą bibliografią, na przykład programem Zotero. Chcieliśmy stworzyć nowoczesny, bezpłatny warsztat wspomagający pracę naukowców, regionalistów, muzealników i pasjonatów. Następne etapy rozwoju i kontynuacja projektu będą poddane szerszej niż dotychczas dyskusji środowiska skupiającego się wokół PIA.

Realizacja tego przedsięwzięcia stała się możliwa po uzyskaniu w sierpniu 2008 roku dotacji z Ministerstwa Nauki i Szkolnictwa Wyższego oraz Fundacji na rzecz Nauki Polskiej. Po negocjacjach z Wydawcami: Instytutem Archeologii i Etnologii PAN, Polskim Towarzystwem Ludoznawczym, Komitetem Nauk Etnologicznych i Instytutem Sztuki PAN uzyskaliśmy zgodę¹ na opublikowanie w Internecie trzech czasopism:

- Etnografia Polska 1958-2004 (dwadzieścia jeden tysięcy stron) – embargo cztery lata,
- Lud (okres powojenny) 1945-2004 (dwadzieścia trzy tysiące pięćset stron) – embargo cztery lata,
- Polska Sztuka Ludowa – Konteksty 1947-2002 (szesnaście tysięcy dwieście trzydzieści pięć stron) – embargo osiem lat.²

Okresy embargo, wprowadzone przez Wydawców, są powszechną praktyką stosowaną przez konsorcja i wydawnictwa, która chroni bieżącą dystrybucję wersji papierowej.

Poważnym – jeśli nie najpoważniejszym wyzwaniem – stało się, zgodnie z ustawą o ochronie praw autorskich, pozyskanie licencji od autorów i ich spadkobierców, co umożliwi publikację dorobku w Internecie. Licencje uzyskane od Wydawców są niewystarczające. W obawie przed naruszeniem praw autorskich zobowiązani jesteśmy do uzyskania zgody autorów, których teksty znalazły się w digitalizowanych czasopismach. Jeśli nie uda się nam uzyskać akceptacji autorów dla naszego projektu, zmuszeni będziemy do udostępniania tekstów w tak zwanym „dozwolonym użytku” (art.28 ustawy o prawie autorskim i prawach pokrewnych dnia 4 lutego 1994, Dziennik Ustaw 1994, nr 24, poz.83)³.

Zgodnie z artykułem 27 tej ustawy:

Instytucje naukowe i oświatowe mogą, w celach dydaktycznych lub prowadzenia

¹ Dziękujemy Dyrekcjom Instytutów IAiE PAN, IS PAN, Komitetowi Badań Etnologicznych, Polskiemu Towarzystwu Ludoznawczemu oraz Redakcjom Etnografii Polskiej, Ludu i PSL-Kontekstów za wyrażenie zgody na digitalizację wydawanych przez nie czasopism – bez tego nasz projekt nie mógłby zostać zrealizowany.

² Ze względu na uczestnictwo Kontekstów-PSL w Central and Eastern European Online Library (CEEOL) okres karencji jest tak długi.

³ Apelujemy do autorów: aby biblioteka cyfrowa mogła stać się przydatnym źródłem informacji, konieczna jest zgoda Autorów i ich Spadkobierców na publikację w Internecie. Prosimy o poparcie naszego projektu przez wyrażenie zgody na publikację tekstów z wyżej wymienionych czasopism w Bibliotece Cyfrowej PIA. Kontakt: www.pia.org.pl.

własnych badań, korzystać z rozpowszechnionych utworów w oryginale i w tłumaczeniu oraz sporządzać w tym celu egzemplarze fragmentów rozpowszechnionego utworu.

Natomiast artykuł 28 cytowanej ustawy stanowi, że:

Biblioteki, archiwa i szkoły mogą:

- 2) sporządzać lub zlecać sporządzenie egzemplarzy rozpowszechnionych utworów w celu uzupełnienia, zachowania lub ochrony własnych zbiorów,
- 3) udostępniać zbiory dla celów badawczych lub poznawczych za pośrednictwem końcówek systemu informatycznego (terminali) znajdujących się na terenie tych jednostek.

Jednak mamy nadzieję, że uda się uzyskać zgodę jak największej liczby autorów. Pierwsze licencje są już wprawdzie podpisane, lecz uzyskanie następnych będzie wymagało więcej czasu. Jest to duża i skomplikowana operacja, która wymaga dotarcia do ponad półtora tysiąca autorów i ich spadkobierców. Autorzy naszych artykułów często zamieszkują odległe kontynenty, ale mamy zamiar pozyskać także ich akceptację. Pierwsze doświadczenia, jakie stały się naszym udziałem przy podpisywaniu licencji, mówią, iż to będzie czasochłonny proces wymagający dużych nakładów pracy i środków finansowych.

Projekt budowy Biblioteki Cyfrowej PIA składa się z różnych etapów działania i – po uzyskaniu niezbędnych licencji wydawniczych oraz źródeł finansowania – obejmuje:

Skanowanie – ponad sześćdziesiąt tysięcy stron – ze względu na wysoki koszt skanowania przez firmy profesjonalnie zajmujące się digitalizacją, postanowiliśmy we własnym zakresie zeskanować, a następnie poddać procesowi OCR wybrane czasopisma (*skany* w rozdzielczości trzysta DPI, w formacie plików PDF). Skanowanie przysporzyło nam wiele trudności i zajęło najwięcej czasu, ze względu na bardzo złą jakość wybranych czasopism. Marnej jakości papier, wyblakły druk, nieczytelne zdjęcia, minimalnej wielkości marginesy spowodowały, że niektóre skany powtarzane były wielokrotnie, aż do uzyskania w miarę zadowalającego efektu. Mimo dołożenia wszelkich starań, nie zawsze udało uzyskać się optymalną jakość zeskanowanego tekstu.

Zakup i instalacja serwera obsługującego bibliotekę cyfrową. Dzięki współpracy z Instytutem Etnologii i Antropologii Kulturowej UW mogliśmy zainstalować serwer na Uniwersytecie Warszawskim.

Wybór platformy cyfrowej umożliwiającej założenie i funkcjonowanie biblioteki cyfrowej w Internecie. Zdecydowaliśmy się na polskie oprogramowanie dla bibliotek cyfrowych *d'Libra Digital Library Framework*, które powstało w Poznańskim Centrum Superkomputerowo-Sieciowym. Oprócz dużej funkcjonalności przyjętych rozwiązań i wysokiej jakości oprogramowania, platforma cyfrowa *d'Libra* zapewnia profesjonalny serwis i jest stale modyfikowana. *D'Libra* daje twórcy biblioteki cyfrowej możliwość kontroli dostępu do zasobów, zarządzanie zasobem oraz metadanymi. Wadą systemu jest brak zadowalających rozwiązań w przypadku wprowadzania czasopism, co w przypadku Biblioteki Cyfrowej PIA, nastawionej głównie na gromadzenie kolekcji czasopism, stanowi poważny mankament. Należy mieć nadzieję, że i ten problem z czasem zostanie pomyślnie rozwiązany.

D'Libra to najpopularniejsza obecnie platforma do budowy bibliotek cyfrowych, wykorzystywana w większości polskich bibliotek cyfrowych. System ten posiada rozbudowaną aplikację opartą o relację klient – serwer, z pięcioma interfejsami (aplikacje: użytkownika, redaktora, administratora, panel administracyjny oraz konsolę administracyjną (Kolasa 2007) Ważnym elementem przemawiającym za wyborem tego programu

jest możliwość integracji z innymi bibliotekami cyfrowymi używającymi tego oprogramowania. Twórca d'Libry, PCSS zainicjował utworzenie Federacji Bibliotek Cyfrowych (FBC) skupiającej obecnie (koniec 2009 roku) czterdzieści dziewięć bibliotek cyfrowych. Jest to platforma mająca służyć do budowy sieci rozproszonych bibliotek cyfrowych i repozytoriów w Polsce, aby wspomóc popularyzację i ułatwić wykorzystanie zasobów opublikowanych *on-line*. Ponadto d'Libra koordynuje prace związane z polskim uczestnictwem w największym europejskim projekcie digitalizacyjnym – portalu Europeana (około sześć milionów obiektów cyfrowych: obrazów, tekstów, dokumentów dźwiękowych, filmów wideo), co świadczy o dużym uznaniu, jakim się cieszy w środowisku bibliotek cyfrowych. Perspektywa ewentualnego uczestnictwa Cyfrowej Etnografii w olbrzymim europejskim projekcie sieciowym jest niewątpliwie bardzo kusząca.

Publikację w Internecie zasobu trzech czasopism, czyli wprowadzenie do zasobu Internetu artykułów w wersji pełnotekstowej wraz z opisem bibliograficznym (metadanymi). W terminologii przyjętej do opisu danych w bibliotekach cyfrowych używa się terminu meta dane (Nahotko 2005). Platforma d'Libra umożliwia umieszczenie opisów bibliograficznych „obiektów sieciowych”, jakimi w tym przypadku są publikowane artykuły oraz reszta zawartości periodyków (spisy treści, recenzje i inne) w formacie DublinCore (Dublin Core Element Set). Jest to konwencja powszechnie przyjęta do opisu bibliograficznego w bibliotekach cyfrowych, analogicznie do formatu Marc21 stosowanego w zautomatyzowanych katalogach polskich bibliotek (Grabowska 2007: 23). Publikacja w Internecie obejmuje tworzenie i wprowadzanie przez redaktorów cyfrowych metadanych wraz z przesyłaniem plików PDF z zawartością tekstową artykułów na serwer obsługujący Bibliotekę Cyfrową PIA.

Szkolenie administratorów i redaktorów cyfrowych – system Libra wymaga wyszkolenia osób zajmujących się wprowadzaniem i przesyłaniem danych do Biblioteki Cyfrowej wraz z metadanymi.

Promocja projektu – poprzez Internet, wyjazdy do miast skupiających ośrodki akademickie i muzealne zajmujące się antropologią, udział w konferencjach środowiskowych, pokazy, prelekcje dla osób zainteresowanych tą tematyką (na przykład w ramach XIII Festiwalu Nauki odbył się wykład *Antropolog w Internecie: o technikach wyszukiwania informacji i projekcie Cyfrowej Etnografii*, a w podczas konferencji „Badacz osobny: śladami etnografii Jacka Ołędzkiego” w Państwowym Muzeum Etnograficznym w Warszawie odbyła się prezentacja projektu).

Instalacja menedżera bibliografii umożliwiającego wyszukiwanie i archiwizowanie danych oraz przystąpienie do Federacji Bibliotek Cyfrowych.

Prawa autorskie – utworzenie bazy autorów wszystkich publikacji z czasopism Etnografia Polska, Lud i Polska Sztuka Ludowa-Konteksty oraz podpisanie umów licencyjnych, ze wszystkimi autorami i spadkobiercami, do których uda nam się dotrzeć. Jak już wcześniej wspomniano, jest to najtrudniejsza i najbardziej czasochłonna operacja, której daty zakończenia nie jesteśmy w stanie przewidzieć.

Przewidywany termin uruchomienia Biblioteki Cyfrowej PIA na przełomie 2009/2010 roku musiał ulec przesunięciu, ze względu na konieczność uregulowania spraw związanych z licencjami autorskimi. Obecnie, na stronie www.cyfrowaetnografia.pl lub www.digitalethnography.pl, użytkownik ma możliwość przeglądania opisów bibliograficznych ponad 2500 artykułów i innych treści z czasopism Lud, Etnografia Polska i Polska Sztuka Ludowa-Konteksty bez dostępu do ich zawartości.

Na zakończenie, podsumowując opis projektu powstania Biblioteki Cyfrowej PIA, wypada zauważyć, że chociaż udział w tym przedsięwzięciu przysporzył nam

– koordynatorom⁴ – wiele niespodziewanych trudności (zwłaszcza podczas skanowania miernej jakości dokumentów, z nieczytelnymi zdjęciami a także przy zawieraniu umów z tak dużą liczbą autorów), to satysfakcja z więcej niż przychylnych reakcji naszego środowiska sprawia, że warto było podjąć się tego zadania, także ze względu na wyzwanie, jakiemu musieliśmy sprostać. Cyfrowa Etnografia jest otwartą inicjatywą i zapraszamy do wspólnego budowania naszej kolekcji i publikowania własnych prac w Bibliotece Cyfrowej PIA.

Literatura:

- Franke J., Googletheca Universalis?, [w:] Biblioteki cyfrowe: projekty, realizacje, technologie, red. J. Kasperek-Woźniak, J. Franke, Wydawnictwo Bibliotekarzy Polskich, Warszawa 2007, s.121-166.
- Grabowska M., Biblioteka cyfrowa w środowisku wirtualnym. Nowe wyzwania dla katalogów w erze dokumentów elektronicznych, [w:] Biblioteki cyfrowe: projekty, realizacje, technologie, red. J. Kasperek-Woźniak, J. Franke, Wydawnictwo Bibliotekarzy Polskich, Warszawa 2007, s.21-34.
- Folklor w dobie Internetu, Wydawnictwo Uniwersytetu Mikołaja Kopernika, red. P. Grochowski, A. Mianecki, Toruń 2009.
- Kolasa W. M., D'Libra Digital Library Framework – platforma do budowy bibliotek cyfrowych, [w:] Biblioteki cyfrowe: projekty, realizacje, technologie, red. J. Kasperek-Woźniak, J. Franke, Wydawnictwo Bibliotekarzy Polskich, Warszawa 2007, s.67-89.
- Nahotko M., Metadane, 2000, Ebib 6/2000(14), www.oss.wroc.pl/biuletyn/ebib14/nahotko.html.
- Stanisławska-Kloc S., Prawo autorskie a biblioteka cyfrowa, [w:] III konferencja: Internet w bibliotekach. Zasoby elektroniczne: podaż i popyt. Wrocław 2006, 12-14 grudnia 2005 roku, <http://www.ebib.info/publikacje/matkonf/iwb3/arttykul.php?e>.
- Stross R., Planeta Google. Cel: Skatalogować Wszystkie Informacje Świata, Wydawnictwo Emka, Warszawa 2009.

⁴ Koordynatorzy projektu Cyfrowa Etnografia: Joanna Koźmińska (PIA, IEiAK UW), Paweł Krzyworzeka (doktorant IEiAK UW, PIA), Joanna Zamorska (PIA).