

Workshop Digitalisiertes Archivgut in Online-Findbuchern - <daofind>
Berlin, 26-28 maja 2005

Spotkanie poświęcone było zagadnieniom wiążącym się z realizowanym przez Bundesarchiv projektem <daofind> czyli próbą jednoczesnego zastosowania w praktyce trzech międzynarodowych standardów: EAD (Encoded Archival Description), EAC (Encoded Archival Context) i METS (Metadata Encoding and Transmission Standard). Wzięli w nim udział, oprócz pracowników Bundesarchiv zaangażowanych w projekt, przedstawiciele firmy Startext, która opracowuje narzędzia informatyczne, członkowie EAD-Working Group – współautorzy EAD, przedstawiciele Research Libraries Group, współtwórcy standardu METS, prezentujący w Internecie pomoce archiwalne, m.in. z Bundesarchiv, archiwiści z Anglii, Francji i Szwecji, mający dorobek związany z wdrażaniem EAD i EAC.

Spotkanie prowadziła Angelika Menne-Haritz. Na początku krótko przedstawiła Bundesarchiv oraz projekt <daofind>, który jest współfinansowany przez Fundację Mellona i Bundesarchiv. Projekt ma na celu opracowanie narzędzi umożliwiających wspólne zastosowanie 3 standardów do przygotowania prezentacji na specjalnej stronie w Internecie i linkowania różnych elementów informacji o zasobie Bundesarchiv na 3 poziomach: przewodnika (EAD i EAC), pomocy archiwalnej (EAD) i obrazów dokumentów (METS). Dotychczas próby takiej prezentacji czynione były przy zastosowaniu baz danych i HTML, a także programu Midosa eksportującego dane do XML. Realizacja projektu pilotowego ma przynieść także odpowiedzi na szersze pytania: jakie czynniki są istotne dla narodowej i międzynarodowej integracji danych i jakie są perspektywy rozwoju 3 standardów i ewentualnie EAG. Jego celem jest także upowszechnianie w różnych krajach wiedzy o standardach.

Doświadczenia Francji ze stosowania EAD i EAC przedstawiła Claire Sybille. Prace nad standardami rozpoczęto tam w 1999 r., a w latach 2001-2002 otrzymano specjalne fundusze z Ministerstwa Kultury i RLG. Szerzeniu idei EAD w tamtejszych archiwach, a także wśród firm informatycznych sprzyjają regularnie organizowane konferencje (2000, 2002, 2004). Nawiasem mówiąc, zainteresowaniu firm archiwami i EAD/EAC sprzyja obowiązujące we Francji, podobnie jak w Wielkiej Brytanii prawo zalecające stosowanie formatu XML do wymiany danych. C.Sybille omówiła kilka realizowanych obecnie projektów. Bazują one na standardach ISAD i ISAAR oraz głównie programach open source, które dostosowuje się do celów EAD i EAC. W ten sposób powstało narzędzie Pleade, umożliwiające szeroki opis zarówno danych bibliograficznych, jak i pomocy archiwalnych oraz szczegółowych informacji o zasobie. Warto zwrócić uwagę, że eksport danych do EAD udało się uzyskać także ze zintegrowanego systemu Arkheia, stosowanego w niektórych archiwach departamentalnych. Kilka projektów (m.in. Navimages, Minerva) próbuje już łączyć opis zasobu z kopiami cyfrowymi. Biblioteka Narodowa Francji zamierza używać do tych połączeń standardu METS. C.Sybille zwróciła uwagę, że EAD nie daje możliwości podawania niektórych danych mime type i że czasem używany jest tag <physdesc> do łączenia danych.

Podobnie wiele działań podjęto w Wielkiej Brytanii. Przedstawił je Bill Stocking, podkreślając, że nie istniał w metodyce angielskiej określony schemat sporządzania opisów. W ostatnich latach przyjęto jako podstawowy ISAD. Umożliwia to eksport danych do EAD, choć ostatnio niektóre dane wprowadza się od razu w XML. Prowadzone są również prace

nad łączeniem istniejących zbiorów danych oraz wspólnym stosowaniem EAD i EAC. Do standardu METS jeszcze się nie przymierzają, używając czasem Dublin Core do opisu obrazów cyfrowych i publikacji. Rozważana jest możliwość prowadzenia centralnej rejestracji haseł wzorcowych EAC, aby uniknąć powtórzeń i różnic w sformułowaniu. We wcześniej trzeba jednak opracować procedury ich eksportu i importu między różnymi systemami informacyjnymi. Analogiczna baza danych haseł wzorcowych EAC w Szwecji opiera się na razie na ręcznym wprowadzaniu danych.

Generalny przegląd stanu prac nad EAD przedstawił Michael J.Fox. Podkreślił, że standard służy do opisu wielopoziomowego, nie do zarządzania, może jednak być stosowany do wymiany danych opisujących zasób. Nazwał go „XML archiwistyczny” o określonej hierarchii i strukturze. Nie ukrywał też, że stosowanie EAD budziło i budzi pytania natury technologicznej i socjologicznej – czy standard będzie dość elastyczny, aby sprostać potrzebom wielu użytkowników, a z drugiej strony na tyle sztywny, aby nie zatracić możliwości współpracy i wymiany; co ma być podstawowym poziomem opisu, bo w USA istnieją kłopoty z definicją zespołu; czy wszyscy jednoznacznie będą używać tych samych elementów do tych samych celów. I kto zapanuje nad ewentualnymi różnymi wersjami tak, aby system zachował interoperacyjność? Widać obecnie, że system łączący zawartość, strukturę i kontekst nie jest i nie będzie dziełem zakończonym, ale procesem, w którym zachodzą ciągle zmiany. System uwzględniający standard powinien spełniać 4 funkcje: identyfikacja obiektu i jego opis (EAD i EAC), i osobno wyszukiwanie oraz dostarczanie wyników. Proponowana przez niego architektura systemu powinna opierać się na: scentralizowanej bazie danych, którą łatwiej zarządzać i utrzymywać w porządku; zbieraniu danych w jednym miejscu; połączonym wyszukiwaniu przy użyciu różnych protokołów, m.in. php. Dyskusje toczone są np. nad kolejnością prezentowania wyników wyszukiwania, nie podlega przy tym dyskusji jednoznaczna identyfikacja miejsca w strukturze danych aktualnie oglądanych.

Dyskusja, wywiązana po tej prezentacji, doprowadziła do kilku zgodnych wniosków, m.in. tego, że nie należy dążyć do jednego gigantycznego megasystemu, lecz definiować konkretne funkcje (wprowadzanie, prezentowanie, wyszukiwanie danych) spełniane przez systemy i możliwości ich współpracy. Użytkownik powinien w prosty sposób dotrzeć do informacji, otrzymać jasne wyniki wyszukiwania, choć mało istotna jest informacja, gdzie dane są składowane. Zgodzono się też, że wskazane jest zaniechanie stosowania tagu <eadgrp>, który jest mało precyzyjny i mało przydatny w przeciwieństwie do <eacgrp>. Wskazano też na obecną ewolucję EAD, którego II wersja nie jest już DTD, ale strukturą.

Perspektywy i rozwój zainteresowania prezentacją baz danych opartych na EAD przedstawiła Anne Van Camp z RLG. RLG pracę nad opisami archiwaliów rozpoczęło w 1997 r. Do chwili obecnej zgromadziło 50 tys. pomocy archiwalnych w EAD, dołączając je do zbioru opisów bibliograficznych. Zbierając dane od różnych instytucji, zbierają też doświadczenia, dochodząc do wniosku, że podstawą współpracy jest stosowanie standardów i interoperacyjność systemów, a głównym celem jest użytkownik, któremu powinno się ułatwiać dotarcie do danych i ich zrozumienie. Opowiedziała też o spotkaniu w Kew w dniach 23-24 maja i o podjętych tam ustaleniach, m.in. dążeniu do tworzenia prostych narzędzi do zbierania metadanych, przyjrzenia się angielskiej propozycji Interoperability Protocol oraz przygotowania podręczników w różnych językach dla przedstawicieli nowych państw, które zdecydują się przystąpić do współpracy. Prezentowano fragment portalu RLG (<http://eureka.rlg.org/> - niestety, do korzystania potrzebne jest hasło dostępu), w którym do

wyboru jest kilka wersji językowych, m.in. polska i chińska. Przy tej okazji wywiązała się krótka dyskusja na temat wielości języków w systemach, ale bez konkretnych wniosków. Unicode umożliwia już teraz stosowanie różnych alfabetów, a może jeszcze się rozwinie. EAD nie przewiduje miejsca na tłumaczony tytuł, nie widziano jednak przeciwwskazań, aby robić tak jak wynika z polskiej metodyki – podawać tytuł tłumaczony w nawiasach po oryginalnym. Niektórzy obecni byli też zdania, że można opisywać wszystko w języku kraju, który opracowuje opis. Inny wątek dyskusji dotyczył technicznej strony wyszukiwarki przygotowywanej przez RLG – czy jest ona równie sprawna w różnych przeglądarkach internetowych i ich wersjach językowych, i czy spełnia narodowe wymogi bezpieczeństwa. Przedstawiciel firmy Startext pytał o wymogi, jakie ma spełnić provider, aby udostępniane przez niego systemy współpracowały z produktami RLG. A. Van Camp powiedziała, że każdy ich produkt jest testowany w różnych środowiskach i systemach i w ten sposób starają się osiągnąć optymalną jakość. Dyskutowano też nad znaczeniem linków do oryginalnych stron (mogą zapewnić wiarygodność danych), a jednocześnie kłopotami z tym związanymi (zmieniają się czasem adresy pierwotne, a ich właściciele nie są w stanie zawiadomić wszystkich).

Per-Gunnar Ottosson przedstawił standard EAC i stan jego stosowania w archiwach Szwecji. Używany jest on do tworzenia haseł wzorcowych w bazie NAD, która umożliwia już wyszukiwanie wg tytułu lub zawartości dokumentów, a także wg osoby/institucji, a wtedy wynikiem jest wykaz dokumentów – na różnym poziomie opisu - związanych z kryterium wyszukiwania. Ottosson omówił strukturę standardu. Podkreślił także używanie w pracy różnego rodzaju słowników kontrolowanych, także geograficznych, które nie są przedmiotem EAC. Zaznaczył też, że hasła wzorcowe CPF stosowane są w bazie NAD nie tylko w odniesieniu do twórców, ale także do tematyki archiwaliów. Zauważył też dwa wyniki z dotychczasowej praktyki problemy: powtarzające się nazwy ciał zbiorowych lub osób odnoszące się do różnych bytów z różnych okresów, pełniących różne funkcje; powtarzanie się haseł wzorcowych, które dotyczą tej samej osoby opisywanej w różnych językach (królowie, papieże, święci). Ten drugi problem może być rozwiązany w przypadku jednego providera, który umożliwiłaby „podmianę” danych w systemie tak, aby niezależnie od wpisanego hasła wynik wyszukiwania prowadził do właściwej odpowiedzi. Uwagi te wywołały dyskusję i pytania, np. jak, w jakim języku i w jakim kraju formułować hasło wzorcowe osoby, która mieszkała kolejno w kilku krajach, wszędzie zajmując się znaczącymi sprawami. W tym przypadku opowiedziano się raczej za hasłem całościowym, uzgadnianym wielostronnie, niż za hasłami wycinkowymi. Pytania budziło też poprawne określanie jednostki opisu, jej cech organizacyjnych i funkcji, zwłaszcza w sytuacji zmian organizacji ciał zbiorowych i przekazywania między nimi zadań i funkcji. Poprawność jest tu jak najbardziej potrzebna zważywszy na organizację materiałów archiwalnych w archiwach. Jako sposób na koordynację działań przytoczono przykład elektronicznej wersji brytyjskiego słownika biograficznego, opracowanej w Oxfordzie, która jest wykorzystywana jako źródło haseł wzorcowych w archiwalnych i bibliotecznych systemach informacyjnych. Inny przykład praktyczny stanowi program LEAF, w ramach którego wykonano prototyp programu integrującego różne dane, a obecnie biblioteka w Berlinie ma trudności z finansowaniem podtrzymania tego oprogramowania.

Merrilee Proffitt z RLG, współautorka standardu METS, omówiła jego historię i stan obecny. Pierwotnie stosowany do kodowania metadanych zeskanowanych publikacji – obiektów „reborn digital” – a następnie innych zdigitalizowanych form (obrazy, wykresy,

historia mówiona czyli audio i wideo). Standard pozwala na dołączanie innych metadanych – opisowych (MARC21, DC, EAD), administracyjnych i technicznych. Umożliwia też transfer danych między aplikacjami i instytucjami. Pierwszy projekt powstał w latach 1997-2000, jego efektem był XML DTD, składający się z 7 głównych sekcji: nagłówki, opis, dane administracyjne, dołączone pliki, mapa strukturalna, linki strukturalne i behaviour. Ma on strukturę drzewa, którego elementy mogą być dobierane opcjonalnie, rozwijane i łączone z innymi metadanymi. Współdziała także z modelem referencyjnym OAIS. Pierwsza oficjalna prezentacja METS miała miejsce w październiku 2003 r. w Bibliotece Kongresu – uczestniczyli w niej – oprócz amerykańskich twórców – Brytyjczycy i Kanadyjczycy, gdyż Amerykanie czynią starania, aby standard ten upowszechnić i uczynić międzynarodowym. Obecnie używana jest wersja 1.5, a prace prowadzone są nad pozyskiwaniem, przesyłaniem i rozsyłaniem danych. M. Proffitt wspomniała o kłopotach innej natury – np. ze zdefiniowaniem właściciela standardu i pozyskaniem funduszy na dalsze prace. RLG częściowo stosuje METS do kodowania dokumentów elektronicznych prezentowanych w swoim portalu internetowym, m.in. do układania plików w sekwencje umożliwiające czytanie kolejnych stron publikacji. PG Ottosson zaproponował, aby spróbować stosować METS – obok EAD – jako standard przy przejmowaniu dokumentów elektronicznych do archiwów – do eksportu z systemów zarządzania dokumentacją elektroniczną. C.Sibille dodała, że we Francji czynione są pierwsze próby stosowania METS, ale rozwinięto tam własny schemat XML.

Dyskusję rozpoczęła po tym wprowadzeniu najtrafniej określa żartobliwe „it depends”, jak D.Pitti sugerował rozpoczynać odpowiedzi na większość pytań o łączne stosowanie EAD, EAC i METS. Dyskusja toczyła się głównie między przedstawicielami firmy Startext a amerykańskimi współautorami standardów. Poruszano w niej szczegółowe kwestie współpracy standardów (m.in. w ramach California Digital Image Project, który objął kilkaset tysięcy fotografii) także z oprogramowaniem open source, dbania o ograniczoną jednak elastyczność zapewniającą spójność danych, prawidłowego rozmieszczenia i składowania elementów. Przeważał pogląd, że do przechowywania dużych ilości danych dobrze nadają się relacyjne bazy, a XML może być używany do wymiany i prezentacji danych. METS uznano za bardziej odpowiedni do opisu dokumentu niż zespołu (zbyt skomplikowana struktura i dużo elementów składowych), a do łączenia elementów można stosować wspólne tagi. Dyskutanci zgodzili się z PG Ottossonem, że EAC i METS nie są ze sobą bezpośrednio powiązane. W celu dołączenia haseł wzorcowych należy użyć EAD jako pośrednika lub zawrzeć EAD w METS. Podstawą powodzenia kodowania jest prawidłowe stosowanie standardu ISAD.

Ostatnią prezentację przygotowali pracownicy Startext. Omówili pokrótce dostępne na rynku edytory XML – komercyjne (Metal, Oxygen, XMLSPY i VisualStudio) oraz open source (Eclipse, którego powstanie związane jest z IBM). Właśnie edytor Eclipse stał się podstawą budowania przez nich narzędzia Midosa XML Editor dla potrzeb Bundesarchiv, którego funkcje zaprezentowano uczestnikom workshopu. Edytor ten może pobierać dane z baz o dowolnym formacie, jest „wyuczalny”, powtarza raz użyte komponenty, a użytkownikom pozwala nadawać różne uprawnienia. Mimo że wyjściowy program jest bezpłatny, jego dostosowanie do potrzeb EAD, EAC i METS może okazać się kosztowne. Zgromadzeni zgodzili się jednak z tym, że programy open source mają porządniejszą dostępną dokumentację źródłową, są lepiej dopracowane i gwarantują lepsze dostosowanie do konkretnych potrzeb. Niemieccy informatycy pytali też o możliwość tłumaczenia na

niemiecki, i inne języki, fragmentów programu w oryginale angielskich. Amerykańscy twórcy odpowiedzieli, że noszą się z zamiarem przygotowania tagów w językach UE, ale to musi potrwać. Rozważano przy okazji możliwość zachowania programu tylko w wersji angielskiej oraz przygotowania szczegółowych instrukcji użytkownika w innych językach.

Ostatniego dnia, 28 maja, w spotkaniu wziął udział prof. Hartmut Weber prezydent Bundesarchiv. W słowie wstępnym wyraził nadzieję, że wiele archiwów zdecyduje się na stosowanie EAD, EAC i METS, które staną się kluczem do wspólnej wymiany i prezentacji danych. Po nim A. Menne-Haritz przedstawiła zebrane pytania dotyczące koncepcji stosowania 3 standardów w jednym projekcie. W trakcie dyskusji zgodzono się co do pewnych rozwiązań, na wiele jednak kwestii padała odpowiedź „it depends...”, jako że niemiecki projekt <daofind> wydaje się być pierwszym tak dużym polem doświadczalnym, które przyniesie dalsze odpowiedzi, na co zebrani bardzo liczą. Ze spraw natury szerszej można wspomnieć: za celowe uznano zbieranie danych w plikach XML, ich prezentację w Internecie, a długoterminowe przechowywanie danych w bazach; sposób prezentacji danych, ich układ i wygląd jest sprawą mniejszej wagi, istotniejsze jest prawidłowe sformułowanie opisów i przygotowanie sprawnych narzędzi wyszukiwawczych; konieczne jest opracowanie best practice guidelines do EAD; dołączanie obrazu cyfrowego do opisu najpewniej obecnie można wykonywać poprzez linkowanie do miejsca jego umieszczenia; ważne jest zawieranie w nazwach kopii cyfrowych informacji o zdigitalizowanym oryginale; wyjściowy materiał można umieszczać w bazach danych o właściwej strukturze; tłumaczenie tagów jest dopuszczalne, ale musi być konsekwentne i spójne, może łatwiej będzie to robić, kiedy EAD stanie się schematem.

A.Menne-Haritz zapowiedziała organizację w 2006 r. w Berlinie III Europejskiej Konferencji dotyczącej EAD i EAC. Przy tej okazji planuje się też przygotowanie praktycznych workshopów, m.in. dla osób z Europy Centralnej i Wschodniej. W tym ostatnim przypadku A. Menne-Haritz liczy na włączenie się strony polskiej w przygotowanie tematyki do realizacji.

Konferencję zakończyły trzy prezentacje:

C.Sibille pokazała szczegółowo budowę i działanie aplikacji PLEADE, podkreślając znaczenie list wyboru stosowanych przy wypełnianiu wielu pól, oraz IREL (Instruments de Recherche en Ligne), wykonanej przy użyciu XMetal i zawierającej opisy zespołów oraz dołączone opisy inwentarzowe jednostek archiwalnych.

Noel Marx ze Startextu pokazał Intranet Bundesarchiv, sposób wyszukiwania i prezentacji wyników wyszukiwania, a także narzędzie Eclipse – wpisywanie i tagowanie danych.

Sebastian Barteleit przedstawił projekt przebudowy i modernizacji niektórych budynków w Lichterfelde i budowy nowoczesnego głównego gmachu ze szkła i czerwonego granitu, który ma łączyć kilka funkcji.

Anna Laszuk