

RLG International Archival Gateways Meeting

Kew, 23-24 maja 2005

Spotkanie odbyło się w The National Archives. Wzięli w nim udział przedstawiciele RLG i NARA ze Stanów Zjednoczonych oraz archiwów Wielkiej Brytanii, Francji, Holandii, Niemiec, Szwecji i Polski. Wprowadzenie do tematyki przedstawiła Anne Van Camp, podkreślając, że Research Libraries Group, istniejąca od 30 lat, ma doświadczenia w udostępnianiu danych zgromadzonych w różnych systemach. Określiła cele spotkania, tzn. wymiana doświadczeń, a przede wszystkim ustalenie działań, jakie należy podjąć, aby zbudować jeden portal (bramę dostępową do wielu portali) przedstawiający dane z sektora kultury, oraz zapoczątkowanie stałej współpracy międzynarodowej. Jednym z celów ma być rozszerzony dostęp do informacji, integrującej dane bibliograficzne i pomoce archiwalne. Aby to osiągnąć, konieczna będzie standaryzacja danych. Z drugiej strony jednak model portalu wydaje się bardziej zbliżony do rozwiązań stosowanych w Internecie (hiperłącza) niż do dotychczas stosowanego modelu ściśle określonych punktów dostępu (haseł indeksowych). W modelu tym najważniejszy jest użytkownik, który ma prawo do samodzielnego konstruowania zapytań, nawet jeśli wydają się one nie mieć sensu, nie zaś archiwista i bibliotekarz „reglamentujący” zestawy informacji. Jednocześnie w portalu tym informacje muszą być podane w zrozumiałym dla przeciętnego odbiorcy sposób, gdyż w Internecie nie będzie stojącego obok fachowca, który udzieliłby niezbędnych wyjaśnień. A. Van Camp krótko przedstawiła też charakter i działania RLG.

Następnie przedstawiciele różnych krajów i projektów prezentowali własne doświadczenia i zamierzenia. Pracownicy The National Archives: James Strachan, Bill Stocking, Dick Sargent, omówili kolejno:

- A2A – projekt rozpoczęty w 2000 r. dzięki grantowi Skarbu Państwa (1 mln funtów), a kontynuowany dzięki wielu grantom z funduszu loteryjnego (ok. 3 mln funtów); objął on 400 z 700 instytucji gromadzących archiwalia dotyczące historii Anglii; w jego ramach opisano 92 000 zespołów, stosując standard EAD – opisy liczą w sumie 8,5 mln rekordów oraz mają połączenie z 1 mln cyfrowych kopii dokumentów; system ten ma również ok. 1000 linków do różnych stron internetowych;
- Catalog (d. PROCAT) – inwentarze dokumentów z XI-XXI w. przechowywanych w TNA; w sumie 9,5 mln opisanych jednostek archiwalnych; opisy przygotowywane są na 7 poziomach – od departamentu do dokumentu; aktualizacja systemu online dokonywana jest codziennie; statystyka wykorzystania systemu mówi o 150 tys. użytkownikach miesięcznie poprzez Internet oraz 9,5 tys. na miejscu; koszt systemu wyniósł 1 mln funtów, a roczne podtrzymanie – 10% tej sumy;
- National Register of Archives – rozpoczęty w 1945 r., uczestniczy w nim ok. 3 tys instytucji; w obecnym systemie informatycznym stosuje się EAD do opisów archiwalnych (ponad 277 tys. zespołów), EAC do tworzenia haseł wzorcowych (ponad 48 tys. osób, 116 tys. ciał zbiorowych i 9 tys. rodzin), a do opisu archiwum planowane jest stosowanie

przygotowywanego EAG; system zawiera też linki do serwisów instytucji przechowujących archiwalia – co roku sprawdzana jest ich aktualność.

Na tym tle zupełnie inaczej przedstawia się Walia, co omówił Gwyn Jenkins. Do tej pory opisano w niej – na poziomie zespołu – 90% zespołów. Opisy gromadzone są w bazie danych MySQL, a w Internecie prezentowane w HTML. Pracami tymi kieruje Biblioteka Narodowa Walii.

Istniejący w Anglii oddzielny Archives Hub, obejmujący archiwa 19 szkół wyższych, przedstawiła Amanda Hill. Program rozpoczęto w 1999 r. przy wykorzystaniu funduszy (1,2 mln funtów) zebranych przez uczestniczące uczelnie. Jest to baza inwentarzy utworzona w oparciu o Cheshire3, utrzymywana na uniwersytecie w Manchester. Dane i ich jakość zależą i zależą od twórcy – AH jedynie je prezentuje. Automatyczne zbieranie danych (przez e-mail i ftp) umożliwia od 2002 r. stosowanie EAD.

Claire Sibille poinformowała, że francuskie Ministerstwo Kultury jest odpowiedzialne za prezentację kultury. Na lata 2005-2008 planowane jest opracowanie portalu internetowego, do którego Dyrekcja Archiwów Francji dołączy bazy danych informujące o zasobie archiwalnym. Następnie przedstawiła kilka projektów archiwalnych realizowanych we Francji: Aquitaine patrimoine; Pleade (aplikacja do sporządzania pomocy archiwalnych i prezentacji w Internecie przy użyciu EAD) i Navimages (ewidencja obrazów cyfrowych) – docelowo mają być połączone; France-Genealogique (wspólnie ze Federacją Genealogiczną); IREL (Instruments de Recherche En Ligne – umożliwiające wyszukiwanie pełnotekstowe); NOMINA – aplikacja prowadząca do baz danych informujących o ok. 13 mln osób.

Przygotowany został także portal Archives Canada-France, który zawiera opisy zespołów dotyczących Nowej Francji, przygotowywane są opisy serii i podserii. Postać osobnego portalu ma też przyjąć obecna BORA (Base d'Orientation et de Recherche dans les Archives), która zawiera opisy ok. 5 000 zespołów zgodne z ISAD i sporządzone w EAD. C.Sibille podkreślała kilkakrotnie, że do przygotowywania tych systemów wykorzystywane jest oprogramowanie open source, dostosowywane do konkretnych potrzeb.

W archiwach holenderskich, jak przedstawili to Jacques Bogaarts i Ma Oeh Pe, nie ma centralnego systemu informacji – tworzone są osobne zbiory danych, zwykle na bazie programów opracowywanych indywidualnie przez firmy komputerowe. Centralny system ABS/Archeion jest dopiero planowany. Archiwa zwykle przygotowują opis zbliżone do ISAD, który jest bliski holenderskiej metodyce archiwalnej. EAD jest stosowany natomiast z powodzeniem do opisu zasobu archiwalnego Holenderskiego Instytutu Nauk Społecznych. Holendrzy zastanawiają się, jak i czy wszystkie elementy EAD należy wypełniać.

Przedstawiciel NARA Jim Hastings zwrócił uwagę, że do lat 90. XX w. w USA nie stosowano żadnych standardów opisu. Dopiero utworzenie centralnej bazy danych Archival Research Catalog stało się podstawą do ujednolicania opisów w NARA, 14 archiwach regionalnych i 12 bibliotekach prezydenckich. Podstawowym poziomem opisu w tym systemie jest seria (opisano ich ok. 14 000, wśród nich ok. 40% serii przechowywanych w NARA). Zawiera on także linki do cyfrowych kopii opisanych dokumentów. System ma obecnie ok. 600 tys. opisów i 1 mln haseł wzorcowych, a jego aktualizacja w Internecie

odbywa się co dwa tygodnie. Obecnie rozważane jest zharmonizowanie zawartości ARC ze standardami ISAD i ISAAR.

Angelika Menne-Haritz stwierdziła, że niemiecka metodyka wymaga tworzenia pomocy archiwalnych zawierających informacje o materiałach archiwalnych, ich twórcach i miejscach przechowywania. W celu ich gromadzenia i prezentacji zostało wykonane specjalne narzędzie informatyczne Midosa XML – baza danych o strukturze zgodnej z ISAD, umożliwiająca automatyczny eksport i prezentację danych w EAD w portalu archiwalnym. Midosa umożliwi prezentację w postaci e-książki, na którą składają się 4 warstwy: dane, połączenia (linki), wydruki i instrukcja posługiwania się pomocą archiwalną. Warstwy te składają się na: opis archiwum, przewodnik po zasobie, szczegółowe pomoce archiwalne i zdigitalizowaną wersję dokumentu. Poruszanie się pomiędzy elementami systemu (na ekranie) umożliwia 4 strategie: nawigacja, pełnotekstowe wyszukiwanie, wyszukiwanie według haseł indeksowych i przeglądanie. Uzupełnieniem tej prezentacji mogą być kopie cyfrowe – na razie nieliczne dokumenty. Kopie te wykonuje się z mikrofilmów, które stanowią podstawową formę zabezpieczającą.

Nową wersję systemu zintegrowanego ARKIS, działającą w szwedzkim Rikarkivet od 1999 r., przedstawił Martin Bjersby. System zawiera, obok opisu zasobu archiwalnego, także moduły funkcjonalne – od przejęć począwszy, poprzez opracowanie, magazynowanie na udostępnianiu skończywszy. Opisano w nim już 70 tys. zespołów, 600 tys. serii, 5,1 mln j.a., 400 tys. mikrofilmów, 18 mln obrazów cyfrowych (w formacie dejavu). Osobny system informacyjny stanowi NAD, który przedstawił Per-Gunnar Ottoson. Podkreślił on odmienną od przyjętych w innych krajach filozofię tworzenia systemu. Głównymi jego elementami są słowniki haseł geograficznych i kartoteki haseł wzorcowych ciał zbiorowych, osób i rodzin (EAC). Do nich podłączone są informacje o związanych jednostkach archiwalnych. Opisy archiwaliów (EAD) sporządza się w lokalnych systemach, a przed transferem danych do NAD odbywa się ich weryfikacja, m.in. przy zastosowaniu słowników kontrolnych, i ewentualne poprawianie (centralne). Opisy archiwaliów dołącza się do haseł wzorcowych, sporządzanych jedynie centralnie (EAC). Kilkuletnie korzystanie z systemu pozwoliło na swoisty bilans kosztów (przygotowanie narzędzi informatycznych i ich utrzymanie, rozwój zasad opracowania i słowników kontrolowanych, szkolenia operatorów i kontrola poprawności danych) i zysków (znacząca poprawa jakości danych) tych działań.

Ostatnia prezentacja dotyczyła archiwów polskich. Anna Laszuk i Hubert Wajs omówili zasady tworzenia i udostępniania w Internecie pomocy archiwalnych, m.in. bazy SEZAM, IZA, SCRINIUM, KITA, PRADZIAD, ELA, RMP, baza płacowa i inwentarz Metryki Koronnej przygotowany w EAD.

Drugi dzień spotkania poświęcony był dyskusji, głównie nad możliwościami interoperacyjności funkcjonujących systemów. Punktem wyjścia do dyskusji była prezentacja (Louise Craven i Bill Stocking z TNA) angielskiej propozycji National Council on Archives Interoperability Protocol. Opracowany on został przez grupę ekspertów w celu zharmonizowania istniejących angielskich systemów archiwalnych: TNA, A2A, AIM25, NAS, NLW, Archives Hub, BL, NRA. W trakcie przeglądu systemów zespół ten zwrócił

uwagę na znikome wykorzystanie w praktyce zaleceń eGIF przez archiwa lokalne. Stwierdzono także dość powszechną zbieżność opisów z ISAD, ale różne źródła haseł wzorcowych (nie wszyscy stosują ISAAR). Interoperability Protocol ma umożliwić TNA opracowanie Global Search Engine – oprogramowania pośredniego między bazami danych (MySQL, EAD) a użytkownikiem internetowym (middleware). Wymaga to jednak znacznych nakładów finansowych, które trudno uzyskać, jako że oprogramowanie to będzie - w sumie ukryte – tworzyło jedynie wirtualne zbiory danych, a nie widoczne konkretne byty. Drugą prezentacją wprowadzającą do dyskusji (Arnold Arcolio z RLG) było wyliczenie 6 możliwych typów wyszukiwarki internetowej. Rozpoczęło to wymianę opinii na temat zalet i zagrożeń, jakie wiążą się z wyszukiwarką Google. Z jednej strony zwróconą uwagę, że jeśli Google uznawane są za dobre i powszechnie używane, to należałoby stworzyć dla archiwów coś lepszego. Powstaje tu pytanie, dlaczego użytkownicy archiwów mieliby przestać korzystać z Google. Aby być konkurencyjny, nowy Gateway powinien mieć chwytliwą nazwę, np. Arhcial Explorer, Navigator lub Discovery. A. Menne-Haritz poinformowała, że istnieje wspólna inicjatywa francusko-niemiecka opracowania europejskiej wyszukiwarki konkurencyjnej dla Google, ale jeszcze nie ma rezultatów. Z drugiej strony rozważano możliwość takiego opracowania zawartości serwisów archiwalnych, aby były one w pełni przeszukiwane przez Google. Przed tym jednak należy przeprowadzić analizę wykorzystania archiwalnych serwisów internetowych. Przykłady takich analiz pokazali Anglicy w odniesieniu do A2A (<http://www.a2a.org.uk/stats/statistics.htm>). Z przedstawionych raportów wynika, że na razie interoperacyjność systemów dla archiwów lub sektora kultury próbuje się uzyskać na poziomie narodowym, choć bez konkretnych spektakularnych sukcesów. Jakość danych uzyskuje się przez stosowanie słowników kontrolnych, tezaurusów i kartotek wzorcowych, czasem wspólnych z bibliotekami. Przykładem tego jest elektroniczna wersja brytyjskiego słownika biograficznego, opracowana w Oxfordzie, do której podłączono opisy archiwaliów dotyczących konkretnych osób oraz dane bibliograficzne dotyczących ich publikacji.

Jak wobec tego zapewnić sukces projektom międzynarodowym? Podkreślono, że archiwa chcą wykazać swoje znaczenie i aktywność, a tym samym uzyskać fundusze państwowe właśnie poprzez budowę portali narodowych. Idea międzynarodowej interoperacyjności zakłada zatarcie tego indywidualnego charakteru. Użytkownik portalu globalnego nie będzie powiem zwracał uwagi na twórcę uzyskanej informacji. Zgodzono się, że podstawą tworzenia zbiorów danych wspólnie przeszukiwanych powinny być EAD i EAC, a OAI może być formatem zbierania i wymiany metadanych. EAD może być także formatem do transportu i przejmowania danych, jak stosuje się to w Szwecji w przypadku przejmowania do archiwów dokumentów elektronicznych. Zalecono także analizę dokumentu angielskiego dotyczącego NCA Interoperability Protocol i zapowiedziano dalsze spotkania robocze.

Anna Laszuk, Hubert Wajs