

## Dokument elektroniczny jako obiekt archiwalny

Właśnie mija 10 lat od czasu gdy Ch. M. Dollar opublikował pracę *Authentic Electronic Records: Long-Term Access Strategies*. (Chicago: 1999), jak dotychczas podstawowy tekst z tego zakresu. W IT każda dekada to w istocie epoka, zasadnicza nowość nie dotyczy jednak proponowanych przez Dollar'a dróg rozwiązania problemów przechowywania owych *electronic records* – a może raczej *digital objects* – w długim czasie, ale zdobytych w tym okresie doświadczeń, przede wszystkim zaś wdrożenia rozwiązań typu *Open Archival Information System (OAIS)*<sup>1</sup>. Opisany w tym standardzie archiwalny system informacji to organizacja ludzi i systemów, którzy przyjmują na siebie odpowiedzialność za przechowanie informacji i zapewnienie jej trwania w długim czasie (*preserve information*) oraz jej udostępnianie (*provide access*) dla społeczności, czyli zapewnienie dostępu do tej informacji w sposób zgodny z potrzebami jej pierwotnych użytkowników (twórców).

Już przed 10 laty Dollar uznał ten projekt za obiecujący i interesujący dla świata archiwalnego. Obecnie w prawie każdym kraju tak Unii Europejskiej jak i po drugiej stronie Atlantyku doświadczenia w zakresie przechowywania dokumentów elektronicznych w długim czasie kojarzone są jednoznacznie z tym rozwiązaniem. Także polski prototyp zbudowany we współpracy NDAP z NASK odwołuje się do tego modelu.

Obligatoryjne obowiązki, które archiwum typu-OAIS musi spełniać zaproponowane w tym modelu referencyjnym to:

1. Określenie kryteriów, jakie materiały są właściwe, aby je zaliczyć do archiwalnego przechowania. (Przedmiot, Pochodzenie, Format).
2. Posiąć wystarczające prawa do własności intelektualnej, wraz z obiektami do przechowania, tak aby móc w sposób uprawniony podejmować kroki potrzebne do właściwego przechowania; np.: przeniesienie do nowego systemu technologicznego.
3. Określenie zakresu w jakim wykorzystywał te informacje pierwotny użytkownik (twórca)
4. Zapewnienie, że informacja została przechowana w takiej formie, która (niezależnie od technologii) jest zrozumiała dla pierwotnych użytkowników. Informacja w kontekście – zapewnienie czytelności także tego kontekstu, tak aby informacja była czytelna i zrozumiała dla przyszłych pokoleń. Oraz określenie wynikających z powyższego metadanych.

---

<sup>1</sup> A ściślej zaproponowanego w tym dokumencie *Reference Model for an Open Archival Information System*, który od 2003 r. jest standardem ISO 14721:2003 – *Space data and information transfer systems -- Open archival information system -- Reference model*.

5. Opracowanie dokumentacji i procedur koniecznych do właściwego wypełnienia obowiązku przechowania; czyli mechanizmy przechowania
6. Opracowanie dokumentacji i procedur koniecznych do właściwego wypełnienia obowiązku udostępniania, czyli mechanizmy udostępniania

Model referencyjny zakłada istnienie 3 podstawowych elementów (części, zbiorów .....)

### 1. Środowiska zewnętrzne (*external environment*) SLAJD 9 na które składają się:

- Osoba lub osoby, organizacja lub system przekazujące informacje + metadane do przechowania w OAIS (*Producer (Producers)*)
- *Management* który formułuje ogólne ramy działalności OAIS takie jak : planowanie strategiczne; rozpoznanie źródła finansowania, ale nie codzienne zarządzanie Archiwum.
- I wreszcie odbiorcy, konsumenci (*consumers*), czyli osoba, organizacja lub system korzystająca z informacji przechowywanych w OAIS

I tak weźmy na przykład zbiory danych elektronicznych z centralnych urzędów rządu brytyjskiego dla którego Archiwum typu OAIS to the National Digital Archive of Datasets (NDAD) a ściślej = operates the National Data Repository (NDR) managed by University of London Computer Centre (ULCC), gdzie najwcześniejsze zbiory pochodzą z 1963 r.

1. Producers = UK gov. Departments
2. Management = The National Archives (TNA)
3. Consumers = ALL (via Web)

### 2. Składników funkcjonalnych lub mechanizmów wewnętrznych (*functional components*)

#### SLAJD 15 i 22, gdzie wyróżnia się takie elementy jak:

- *Ingest* – zbiór procedur odpowiedzialnych za przyjęcie złożonej przez Producenta informacji i przygotowanie jej do włączenia do Archiwum  
Przyjęcie → Walidacja → Przekształcenie na formę stosowaną w Archiwum → Wydobycie i/lub stworzenia opisowych metadanych do wyszukiwania i pomocy archiwalnych → Przesłanie (**SIP**) do Archiwum
- *Archival Storage* – system zarządzający przechowywaniem w długim czasie i podtrzymywaniem cyfrowych obiektów powierzonych OAIS; zasadnicza funkcja to zapewnienie dostępności (Online, nearline, off-line) z czym wiążą się takie problemy jak: kompletność i czytelności *bit stream* (*Odświeżanie nośników; migracje, Error-checking procedures, Evaluation, Disaster recovery policy*) i ułatwienia wyszukiwania danych.

- *Data Management* – podtrzymywanie baz danych z metadanymi identyfikującymi i opisującymi zarchiwizowaną informację (co umożliwia jej wyszukiwanie i tworzenie raportów) oraz danych umożliwiającą administrowanie systemem (zmiana wersji)
- *Preservation Planning* - odpowiedzialny za całość strategii przechowywania, w tym: monitorowanie wewnętrznego środowiska; rekomendacje dotyczące zmian (Systemu, Technologii, Strategii wyszukiwawczych)
- *Access* – obsługa zapytań Consumer'a
  - Przekazywanie zapytań do Data Management
  - Udostępnianie wyników uzyskanych z Archival Storage
  - Dokonywanie koniecznych transformacji formatów
  - Mechanizmy bezpieczeństwa
- *Administration* – codzienne działania systemu, koordynacja współpracy pozostałych elementów
  - Zarządzanie wewnętrznymi procedurami
  - Kontakt z zewnętrznymi podmiotami (Producers, Consumers)

### **3. Obiektów, które są ‘zaszyte’, zarządzane i rozpowszechniane (*Information objects*).**

W tym modelu są one określane jako trzech typów *Information packages* mianowicie: Submission Information Package (SIP), Archival Information Package (AIP) i Dissemination Information Package (DIP). Są to oczywiście *digital objects* – obiekty cyfrowe [określenie to wydaje mi się znacznie bardziej ‘nośne’, niż dotychczasowe ‘dokumenty elektroniczne’], których jednym z istotniejszych własności jest posiadanie metadanych, a zwłaszcza tych które pomagają w przechowaniu obiektu w długim czasie (i opisywać historię przechowania i zmian) oraz jego udostępnianiu.

Producent przekazuje do Archiwum SIP - Information ingested into archive, którego forma (information may not be preserved in the exact form in which it is submitted) nie musi być wcale ‘pierwotna’. Ona mogła zostać wynegocjowana z Producentem, albo utworzona *Ad hoc* (object + metadata) na potrzeby transferu. Owe zmiany mogą dotyczyć takich kwestii jak: zmiany: format; układu czy wreszcie możliwe jest późniejsze uzupełnianie metadanych (czy to przez Producenta, czy też przez samo Archiwum).

Archiwum przechowuje AIP SLAJD 26 na który składają się: informacja + kompletny zestaw metadata (dla przechowywania i dla udostępnienia). Co ważne (i ciekawe) tak naprawdę jest to (logical package) ‘wiązka’ logiczna - nie ma wymogu wspólnej formy fizycznej, dane mogą być ‘zaszyte’ w jednym pojemniku (wrapper), lecz równie dobrze mogą być składowane osobno, jako oddzielne bazy danych na informację i na metadane. Na AIP

składają się informacje co do zawartości (Content Information) oraz dotyczące przechowywania (Preservation Description Information PDI). Content Data Object czyli co tam ‘w środku’ jest: jeden plik czy wiele różnych plików, a może strony www; czy jest to: text, image, video, database, computer program?) a także Representation Information, czyli opis sprzętu i oprogramowania, a także opis zawartości (co znaczą pola, cyfry).

Dla Archiwisty istotne wydają się Preservation Description Information PDI – czyli m. in.: Reference Information (unikatowy identyfikator); Context Information (relacje z całością zawartości danego zespołu); Provenance Information (dzieje przechowywania i zmiany (migracje) oraz Fixity Information (Validacja (sumy kontrolne etc.) Nad tym wszystkim są (Packaging Information) informacje tworzące pojedynczy logiczny zestaw (package or unit) wewnątrz Archiwum oraz (Descriptive Information) czyli pomoc wyszukiwawcza (np.: matadane DC).

Z kolei dla konsumenta przeznaczony jest DIP, który może się różnić od AIP zarówno formą (np.: zamiana z TIFF na JPEG), jak i zawartością (np.: wyszukuje się i udostępnia tylko wskazaną w zapytaniu część danych) i wreszcie towarzyszącymi metadanymi (tylko wybór).

Należy jednak pamiętać, że omawiany tu model referencyjny OAIS – to nadal model teoretyczny. Praktyczne przesłanie z niego wynikające to: *interoperability*, i tu właśnie potrzebne są standardy: Metadata Encoding and Transmission Standard - METS i PREservation Metadata: Implementation Strategies - PREMIS **SLAJD 46**

I tak schodzimy po kolei w dół do kolejnych porcji informacji i danych koniecznych do przechowania i zarządzania obiektów cyfrowych w długim czasie. METS to standard dla obsługi metadanych koniecznych do: zarządzania obiektami cyfrowymi w Archiwum; wymiany obiektów cyfrowych między archiwami (Użytkownikami), a także wspólny format obiektów dla Information management tools/services oraz Interoperable exchange of digital materials. Z jego 7 części **SLAJD 32** najistotniejsze wydają się z tego punktu widzenia: (2) Descriptive Metadata; to co można dołączyć z zewnątrz: Dublin Core czy MARCXML MARC 21 Schema oraz (3) Administrative Metadata, gdzie jako Preservation Metadata jest właśnie PREMIS. Punktem wyjścia do opracowania tego standardu była OAIS, a ściślej właśnie Preservation Description Information – PDI. Wzorcowy słownik PREMIS (PREMIS Data Dictionary) gromadzi i dalej rozwija konceptualne typy obiektów informacyjnych na więcej niż 100 strukturalizowanych i logicznie spójnych jednostek semantycznych. Daje też on określenia i zalecenia jak tych jednostek użyć aby połączyć źródła metadanych dla Intellectual Entities. Określa on “preservation metadata” as *the information a repository uses to support the digital preservation process*. **SLAJD 43** Model danych PREMIS to:

*Intellectual Entity* (a set of content that is considered a single intellectual unit for purposes of management and description: for example, a particular book, map, photograph, or database. An Intellectual Entity can include other Intellectual Entities; for example, a Web site can include a Web page; a Web page can include an image. An Intellectual Entity may have one or more digital representations.)

*Object (or Digital Object)* a discrete unit of information in digital form

*Event*: an action that involves or impacts at least one Object or Agent associated with or known by the preservation repository

*Agent*: person, organization, or software program/system associated with Events in the life of an Object, or with Rights attached to an Object

*Rights*: assertions of one or more rights or permissions pertaining to an Object and/or Agent

Metadane odnośnie do przechowywania obejmują takie zagania i mają dać odpowiedź na takie pytania jak: pochodzenia (provenance): kto był przechowawcą/właścicielem ‘digital object’. Autentyczność, czy *do* jest tym, czy wydaje się być (*what it purports to be*)?, Przechowanie – co do tej pory zrobiono; środowisko techniczne – co jest potrzebne aby go (*render and use*) użyć; prawa własności – jakich Intellectual Property Rights należy przestrzegać? To wszystko po to *do* ‘sam się’ dokumentował w czasie.

Budowanie takich systemów wymaga – i tak ‘sugeruje’ MRA – ścisłej współpracy Archiwów z jednostkami naukowo badawczymi z zakresu techniki (uczelnie, JBR, organizacja). W USA to SDCC w Anglii to ULCC etc. Skończyły się czasy ‘pionierów’!

\* \* \*

Rozwiązania typu OAIS to oczywiście nie koniec problemów z przechowywaniem obiektów cyfrowych, czekają nas kolejne wyzwania związane z rozwojem technologii internetowych: przede wszystkim problematyka Web 2.0 (Wiki, blogi, fora etc.), ale także z coraz powszechniejszą mobilnością (*Blackberry*). Udział ‘pracujących gadżetów (*smart-phones*) w rynku mobilnej telefonii podwoił się w ciągu ostatnich trzech lat i wynosi obecnie około 15% (z ponad miliarda tych urządzeń używanych obecnie na naszym Globie).

Na tegorocznych targach Nokia Word 09 w Stuttgarcie fiński producent ‘telefonów’ NOKIA zaprezentowała nowe nazewnictwo – klasyfikację – swoich produktów: *smart-phon* –

urządzenia z serii **E** będą typowo biznesowe (moduł WiFi, klient e-mail/IM [Instant messaging], synchronizacja, szyfrowanie danych i pełna klawiatura), zaś z serii **X** – rozrywkowo multimedialne. Podobną klasyfikację wprowadza też – choć mniej konsekwentnie, inny producent tych urządzeń – Sony Ericsson. Klasyfikacja ma ułatwić wybór urządzenia, a ponadto jest wyraźnym znakiem rozwoju różnych gałęzi tych urządzeń. Prekursorem urządzeń biznesowych od 10 lat jest wspomniany terminal/komputer kieszonkowy *BlackBerry* kanadyjskiej firmy z Ontario RIM – *Research In Motion* (z bardzo mocnym mechanizmem kodowania wiadomości – algorytm Triple-DES z kluczem o długości 168 bitów lub AES). NOKIA przygotowuje swoją ofertę, w której E63 z prawie takimi samymi możliwościami jak *BlackBerry Bold* jest o 50% tańszy od tego ostatniego, na wielki rynek trzeciego świata, zwłaszcza do Indii. Warto sobie uświadomić, że „na Zachodzie” to pojawienie się PC w ostatniej dekadzie XX w. stanowiło początek przemian technologicznych, zaś społeczeństwo Indyjskie ten etap po prostu ‘przeskoczy’ i od razu będzie pracować na urządzeniach typu *smart-phon*. W numerze wiosna/lato 2009 *The American Archivist*<sup>2</sup>, znajduje się krótki tekst Michell Caswell – *Instant Documentation: Cell-Phon-Generated Records in the Archives*. Jest to pierwszy znany mi tekst poświęcony problematyce przechowywania tej klasy ‘dokumentów’ w naszej fachowej archiwalnej prasie. W tekście chodzi o dokumentację tragicznych wydarzeń takich jak: 11 września 2001 w Ameryce, potem Madryt 2004 i Londyn 2005 oraz protestów w Tybecie z marca 2008 r. Warto jednak zwrócić uwagę, że masowe zastosowania urządzeń typu *smart-phon* zmienia też tryb załatwiania spraw i ich dokumentowania, a jednym z pasjonatów wykorzystania *BlackBerry* jest pewien prawnik, mieszkający do niedawna w Chicago, a obecnie, od 20 stycznia 2009 r. w Białym Domu. To z kolei jest już problem gromadzenia dokumentacji do kolejnej biblioteki prezydenckiej.

W tej sytuacji coraz bardziej zasadne wydaje mi się pytanie o to gdzie jest (i ma być) ów ‘dokument’?, czy raczej *digital object*, gdyż wiele wskazuje na to, że z biegiem czasu nawet administracja publiczna stanie się coraz bardziej ‘mobilna’, gdyż jest to warunkiem dostępu 24/7, a co za tym idzie *paperless*.

---

<sup>2</sup> Michell Caswell – *Instant Documentation: Cell-Phon-Generated Records in the Archive*, *The American Archivist*, vol, 72, nr I, spring/summer 2009, s. 133-145.