

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 opartych na modelu odniesienia *Open Archival Information System* (OAIS)¹. Opisany w tym standardzie archiwalny system informacji to organizacja ludzi i systemów, którzy przyjmują na siebie odpowiedzialność za przechowanie informacji, 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 odniesienia to:

1. Określenie kryteriów, jakie materiały są właściwe, aby je zaliczyć do archiwalnego przechowania (przedmiot, pochodzenie, format).

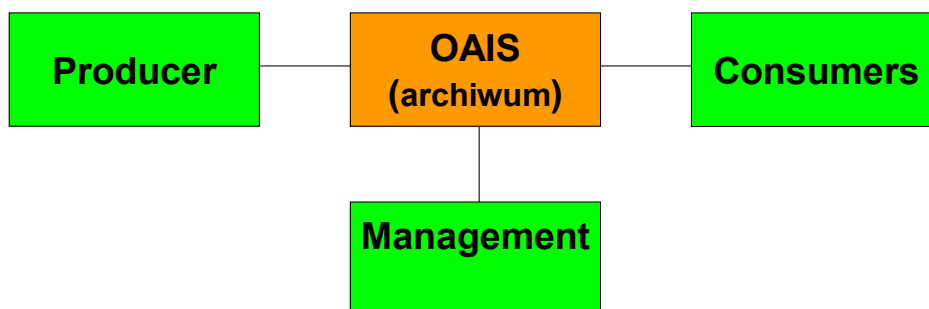
¹ A ściślej zaproponowanego w tym dokumencie *Reference Model for an Open Archival Information System*, który od 2003 r. jest standardem ISO 14721:2003 – *Space data and information transfer systems -- Open archival information system -- Reference model*.

2. Wejście w posiadanie wystarczającego 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 informację 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.
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 odniesienia zakłada istnienie 3 podstawowych elementów części, (zbiorów, czy też grup), są to (rys. 1):

Rys. 1

OAIS - środowisko



1. Środowisko zewnętrzne (*external environment*) na które składają się:

- *Producer; Producers*, czyli osoba/(y), organizacja/(e) lub system/(y) przekazujące informacje + metadane do przechowania w OAIS.
- *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 osoby, organizacje lub systemy korzystająca z informacji przechowywanych w OAIS.

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

1. *Producers*, to odpowiednie ministerstwa, departamenty i urzędy brytyjskiej administracji rządowej (publicznej)
2. *Management*, to *The National Archives* (TNA)
3. *Consumers*, to wszyscy potencjalni zainteresowani (via Web)

2. Składniki funkcjonalne lub mechanizmy wewnętrzne (*functional components*).

3. Obiekty, które są 'zaszyte', zarządzane i rozpowszechniane (*Information objects*).

W tym modelu są one określane jako trzech typów pakietów informacyjnych (*Information packages*) mianowicie: *Submission Information Package* (SIP), *Archival Information Package* (AIP) i *Dissemination Information Package* (DIP). Stanowią je oczywiście *digital objects* – obiekty cyfrowe² posiadające metadane. Zwłaszcza taki ich zestaw, który pomagać ma w przechowaniu obiektu w długim czasie, opisując historię przechowania i zmian; a także w jego udostępnianiu.

Producent przekazuje do Archiwum SIP – (information ingested into archive – informację przyswojoną przez archiwum); cały schemat ilustruje rys. 2. Jej forma (nośnik i format zapis) wcale nie musi być 'pierwotna'. Informacja może być przechowana w innej formie niż ją złożono. Owa forma mogła zostać

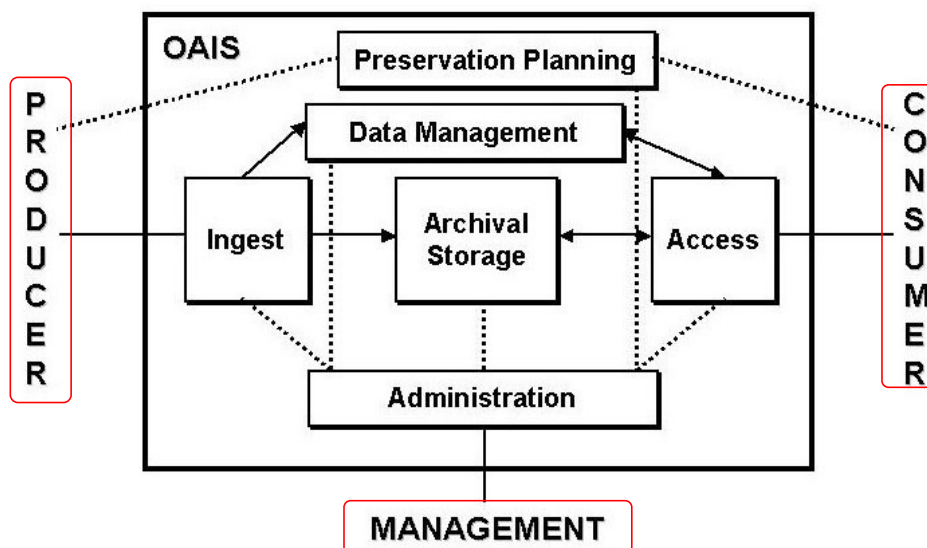
² Określenie to wydaje mi się znacznie bardziej 'nośne', niż dotychczasowe 'dokumenty elektroniczne'.

wynegocjowana z Producentem, albo utworzona *Ad hoc* (object + metadata) na potrzeby transferu. Owe zmiany mogą dotyczyć takich kwestii jak: inne format; odmienny układu; możliwe jest też późniejsze uzupełnianie metadanych (czy to przez Producenta, czy też przez samo Archiwum).

Archiwum przechowuje AIP, na który składają się: informacja + kompletny zestaw metadata potrzebnych tak dla przechowywania jak i dla udostępnienia. Co ważne (i ciekawe) tak naprawdę jest to 'pakiet' logiczna (logical package) - nie ma wymogu wspólnej formy fizycznej, dane mogą być 'objęte' w jednym opakowaniu/obwolucie (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)³ oraz dotyczące przechowywania (Preservation Description Information PDI).

Rys. 2

OAIS – model funkcjonalny



³ Informację o zawartości tworzą: *Content Data Object* czyli co tam 'w środku' jest: jeden plik czy wiele różnych plików, a może strony www; czy jest to: tekst, image, video, database, computer program? oraz *Representation Information*, czyli opis sprzętu i oprogramowania, a także objaśnienie 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 (walidacja; 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.: metadane zapisane zgodnie ze standardem – The Dublin Core Metadata Initiative, DCMI).

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 odniesienia OAIS – to nadal model teoretyczny. Praktyczne przesłanie z niego wynikające to: *interoperability* [współoperatywność], a do tego potrzebne są kolejne standardy: Metadata Encoding and Transmission Standard - METS i PREservation Metadata: Implementation Strategies - PREMIS Oba wprowadzicie wywodzące się ze środowiska bibliotecznego, lecz coraz częściej stosowane też przez elektroniczne archiwa.

I tak schodzimy po kolei w dół do następnych porcji informacji i danych koniecznych do przechowania obiektów cyfrowych i zarządzania nimi w długim czasie. METS to standard dla obsługi metadanych koniecznych do: zarządzania obiektami cyfrowymi w Archiwum oraz wymiany (*interoperable exchange of digital materials*) obiektów cyfrowych między archiwami (Użytkownikami, Konsumentami). Z jego 7 części najistotniejsze wydają się z tego punktu widzenia: (2) *Descriptive Metadata*; to co można dołączyć z zewnątrz: The Dublin Core Metadata Initiative, DCMI czy MARCXML MARC 21 Schema oraz (3) *Administrative Metadata*, gdzie jako rozszerzenie dla *Preservation Metadata*⁴ dodawany jest właśnie PREMIS. Punktem wyjścia do opracowania tego standardu była OAIS, a ściślej właśnie, wspomniane

⁴ *The information a repository uses to support the digital preservation process.*

powyżej *Preservation Description Information* – PDI. Wzorcowy słownik PREMIS (*PREMIS Data Dictionary*) gromadzi i 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*.

Tak opisane metadane mają dać odpowiedź na pytania dotyczące: pochodzenia (*provenance*): kto był przechowawcą/właścicielem '*digital object*'; autentyczność, czy '*digital object*' jest tym, czy wydaje się być (*what it purports to be*); przechowanie – co do tej pory zrobiono i w jakim środowisku technicznym; co jest potrzebne, aby go (*render and use*) użyć; oraz jakich prawa własności (*Intellectual Property Rights*) należy przestrzegać? To wszystko po to '*digital object*' 'sam się' dokumentował w czasie.

Budowanie takich archiwum typu-OAIS wymaga – i tak są zalecenia Międzynarodowej Rady Archiwów – ścisłej współpracy Archiwów z instytucjami naukowymi z zakresu techniki (uczelnie, JBR, organizacja). W USA partneram [National Archives and Records Administration](#) jest San Diego Supercomputer Center (SDSC); w Anglii, jak wspomniałem, jest to *University of London Computer Centre* 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 mobilnych urządzeń używanych obecnie na naszym Globie)⁵.

W ciągu tych 10 lat nastąpiły dwie zmiany: w technologii i w 'kulturze'. Większość nowych technologii jest tak zaprojektowana, że wytwarzane obiekty cyfrowe nie koniecznie mają właściwości oczekiwane przez nas – archiwistów – od dokumentów i łatwo nie poddające się archiwizowaniu. Są to 'twory' znacznie bardziej spersonalizowane lub zindywidualizowane jak: blogi, *instant messaging*, *twitters* (świergotki), czy fora i Wiki; mające różne formaty i składając się z pomieszanych wielu typów plików multimedialnych. Wynika to z upowszechnienia się osiągnięć przemian technologicznych. Już nie tylko urzędy i instytucje administracji publicznej oraz firmy z sektor prywatnego są użytkownikami tej infrastruktury (i twórcami 'archiwów elektronicznych'); z nowoczesnych rozwiązań korzystać może każdy i to często za bardzo małe pieniądze (lub wręcz bez nich z *open sources* – co też często powoduje posługiwanie się niestandardowymi formatami). To co możemy dziś elektronicznie rejestrować (zarchiwizować) to coś, co zawsze nie łatwo poddawało się regułom archiwalnym – codzienność. Tworzące się portale społecznościowe jako główny cel stawiają sobie wymianę informacji, a nie dokumentowanie działalności. By nie wspomnieć o tym, że tu inaczej działają regulacje prawne jak ochrona danych, czy wizerunku – vide fora, *LinkedIn* czy *Facebook*. Społeczeństwo stało się naprawdę „informacyjne”, nie tak jak w modelu sprzed okresu Web 2, gdy miało być odbiorcą jak najszerzego wachlarza usług dostarczanego przez administrację publiczną *on-line*.

Aktotwórcy i archiwiści wiedzą jak istotne są metadane, stąd skłonność tworzenia zawiłych i pełnych szczegółów formularzy do wprowadzania danych; „Użytkownicy 2.0” nie są fanami formularzy i wierzą w automatyczną kategoryzację i wyłapywanie metadanych. Tak samo zresztą nie lubią skomplikowanych wyszukiwań – w *Google* wyszukiwanie jest proste. Aktotwórcy i archiwiści zastanawiają się jakie dokumenty i po jakim czasie można wybrakować;

⁵ TIME, vol 174, no 4, 2009; Technology: Nokia Calling, p. 5-37

„Użytkownicy 2.0” wiedzą, że pamięci masowe są tanie, a będą jeszcze tańsze. O tym zresztą przypomina komunikat otwierający g-mail’a. Przekonanie to wyraża ciekawy projekt biblioteczny LOCKSS (*Lots of Copies Keep Stuff Safe*)⁶, realizowane przez Biblioteki Stanford University, dający bibliotekarzom bezpłatne narzędzia do przechowywania ‘e-zawartości’ (e-content). „Użytkownicy 2.0” sądzą też, że ponieważ wszyscy kopiują wszystko co się da (a na dodatek nie zawsze legalnie) zawsze gdzieś poszukiwany plik się odnajdzie, mniejsza o autentyczność wiadomo w świecie cyfrowym oryginał i kopia są tożsame. Oczywiście są archiwa internetowe, *WaybackMachine*⁷, czy wielki projekt Biblioteki Kongresu Stanów Zjednoczonych „Digital Preservation”⁸ – lecz w większości są to archiwa tematyczne.

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

⁶ <http://www.lockss.org/lockss/Home>

⁷ <http://www.archive.org/web/web.php>

⁸ <http://www.loc.gov/library/digitalpreservation.html>

'przeskoczy' i od razu będzie pracować na urządzeniach typu *smart-phone*. W numerze wiosna/lato 2009 *The American Archivist*⁹, znajduje się krótki tekst Michell Caswell – *Instant Documentation: Cell-Phone-Generated Records in the Archives*. Jest to pierwszy znany mi tekst poświęcony problematyce przechowywania tej klasy 'efemerycznych', jak się potocznie sądzi '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-phone* 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' czy też 'unwired', gdyż jest to warunkiem dostępu 24/7, a co za tym idzie *paperless*.

Hubert Wajs
Legionowo

⁹ Michell Caswell – *Instant Documentation: Cell-Phone-Generated Records in the Archive*, *The American Archivist*, vol, 72, nr I, spring/summer 2009, s. 133-145.