6. Zabezpieczenie wiarygodności zasobów cyfrowych deponowanych w repozytoriach instytucjonalnych

Aneta Januszko-Szakiel

Wprowadzenie

Instytucje sektora biznesu, administracji, a także nauki i kultury, w bieżącej działalności wytwarzają i operują rozmaitymi dokumentami w cyfrowej postaci. Dokumenty te tworzą cyfrowe kolekcje i są deponowane w systemach repozytoryjnych, gdzie podlegają krótko- bądź długoterminowej archiwizacji.

System repozytoryjny, określany również jako depozytowy bądź archiwalny, jest definiowany jako organizacja ludzi, narzędzi oraz przyjętych rozwiązań organizacyjnych, technicznych i prawnych, powołany w celu zgromadzenia, bezpiecznego przechowania oraz zapewnienia dostępu i użyteczności zgromadzonych zasobów w określonym czasie[1]. Repozytoria biznesowe, w zależności od rodzaju przechowywanych dokumentów, realizowanych zadań oraz procedur prawnych mają zapewnić dostępność i użyteczność zdeponowanych materiałów w okresie od trzech do pięćdziesięciu lat[2], natomiast systemy repozytoryjne instytucji pamięci powinny gwarantować utrzymanie użyteczności zasobów przez okres stu i więcej lat[3]; w niektórych przypadkach wieczyście[4]. Docelowo system repozytoryj-

---

[1] Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data Systems. CCSDS 650.0-B-1. Blue Book, Iss. 1. January 2002. Consultative Committee for Space Data System, Washington D.C., [online:] http://public.ccsds.org/publications/archive/650x0b1.pdf, [dostęp: 10.11.2012]; A. Januszko-Szakiel, Open Archival Information System – standard w zakresie archiwizacji publikacji elektronicznych, „Przegląd Biblioteczny" 2005, nr (73)3, s. 342.

[2] W. Sasin, *Przechowywanie i archiwizowanie dokumentacji przedsiębiorstwa według nowych zasad normatywnych. Poradnik dla wszystkich firm z instrukcją wzorcową. Stan prawny na dzień 1 lutego 2004 r.*, Wydawnictwo „Sigma", Skierniewice 2004; M. Konstankiewicz, *Wykaz ważniejszych resortowych aktów prawnych regulujących zasady postępowania z dokumentacją. Uzupełnienie i kontynuacja (część XVIII)*, „Archiwista Polski" 2005, nr 3(39), s. 49–62; *idem*, *Wykaz ważniejszych aktów prawnych regulujących zasady postępowania z dokumentacją. Uzupełnienie i kontynuacja (część XVIII)*, „Archiwista Polski" 2006, nr 2(42), s. 53–60.

[3] U. M. Borghoff, *Vergleich bestehender Archivierungssysteme*, „Nestor Materialien" 2005, nr 3, [online:] http://files.d-nb.de/nestor/materialien/nestor_mat_03.pdf, [dostęp: 21.11.2012].

[4] Ustawa z dnia 27 czerwca 1997 r. o bibliotekach, Dz.U. z 1997 r., Nr 85, poz. 539.

ny ma zapewnić obecnym i przyszłym użytkownikom możliwość dostępu do kompletnej kolekcji użytecznych dokumentów, w ramach posiadanych przez nich praw dostępu[5]. Użyteczność dokumentów cyfrowych wiąże się z efektywnym korzystaniem z zapisanych w nich treści (informacji). Jest to możliwe, wówczas gdy użytkownik ma pewność, że treści, które czyta, słucha, ogląda, a także, na które powołuje się, są autentyczne i niezafałszowane, to znaczy, że pochodzą od ich autorów i od dnia opublikowania nie uległy zmianie; przedstawiają dokładnie to, co było zamierzeniem ich twórców[6]. Systemy repozytoryjne mają zatem za zadanie zabezpieczać i zagwarantować wiarygodność zdeponowanych zasobów cyfrowych.

Celem bezpiecznego deponowania zasobów cyfrowych jest zapewnienie, że dokumenty i zapisane w nich treści zostały zachowane w niezmienionej postaci. Proces ten polega na zachowaniu niezmienionej substancji dokumentu cyfrowego w postaci kodu zerojedynkowego oraz na zapewnieniu platformy programowo-sprzętowej, która będzie w stanie zdekodować dane cyfrowe i przedstawić je w postaci czytelnej dla użytkownika.

Organizatorzy bezpiecznych i trwałych systemów repozytoryjnych odwołują się do standardu archiwizacji zasobów cyfrowych OAIS[7].

6.1. OAIS – Open Archival Information System

OAIS to referencyjny model organizacji i przebiegu procesu trwałej ochrony obiektów cyfrowych, stworzony przez *Consultative Committee for Space Data Systems* (CCSDS)[8] na potrzeby archiwizacji i wymiany danych cyfrowych, zawierających informacje z badań przestrzeni kosmicznej agencji NASA. W lutym 2003 roku model OAIS został zaakceptowany przez International Organization for Standardization ISO jako norma postępo-

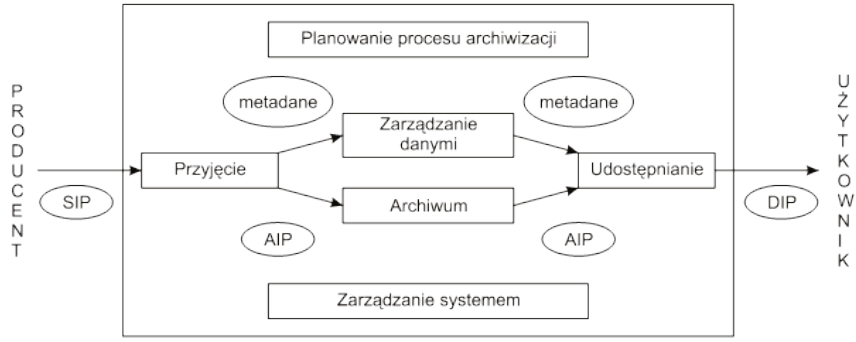
[5] T. Bilski, *Pamięć. Nośniki i systemy przechowywania danych*, Wydawnictwa Naukowo-Techniczne, Warszawa 2008, s. 423–425; *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. Version 1: Entwurf zur öffentlichen Kommentierung*, „Nestor – Materialien" 2006, nr 8, [online:] http://edoc.hu-berlin.de/series/nestormaterialien/2006-8/PDF/8.pdf [dostęp: 16.11.2012].

[6] W. Coy, *Perspektiven der Langzeitarchivierung multimedialer Objekte*, „Nestor Materialien" 2006, nr 5, [online:] http://files.dnb.de/nestor/materialien/nestor_mat_05.pdf, [dostęp: 10.11.2012].

[7] *Reference Model*…, *op. cit.*

[8] Komitet CCSDS został powołany w 1982 roku. Jest organizacją składającą się z przedstawicieli wielu światowych agencji badań przestrzeni kosmicznej i podlega bezpośrednio agencji NASA. *The Consultative Committee for Space Data System*, Reston VA, USA. CCSDS/AIAA, 2010, [online:] http://www.ccsds.org/, [dostęp: 16.11.2012].

wania w zakresie długoterminowej archiwizacji danych cyfrowych (ISO: nr 14721:2003). Obecnie jest uznawany za uniwersalny model organizowania i funkcjonowania repozytoriów cyfrowych i stosowany do gromadzenia, trwałego i bezpiecznego archiwizowania oraz udostępniania różnych typów dokumentów cyfrowych[9].



Rysunek 9. Funkcjonowanie archiwum OAIS. Na podstawie modelu referencyjnego OAIS

Źródło: *Reference Model…*, *op. cit*.

Repozytorium zasobów cyfrowych, zorganizowane na podstawie modelu OAIS obejmuje sześć funkcjonalnych jednostek oraz drogę dokumentu cyfrowego od producenta do użytkownika. Jednostka „Przyjęcie" odpowiedzialna jest za przyjęcie zgłoszeniowego pakietu informacyjnego SIP (*Submission Information Package* – SIP) oraz za przygotowanie go do umieszczenia oraz administrowania nim w repozytorium. W zakresie jej zadań znajduje się m.in. kontrola kompletności oraz autentyczności zgłoszeniowego pakietu informacyjnego, przekształcenie pakietu SIP w pakiet gotowy do archiwizacji oraz stworzenie do niego metadanych. Następnie archiwizowany pakiet informacyjny AIP (*Archive Information Package* – AIP) przekazywany jest do jednostki zajmującej się archiwizacją, tj. do „Archiwum", a metadane odsyłane do jednostki „Zarządzanie danymi", odpowiedzialnej za zarządzanie zdeponowanymi zasobami.

Kolejną istotną jednostką funkcjonalną systemu OAIS jest „Archiwum", odpowiedzialne za zapis, właściwe przechowywanie pakietów informacyjnych (AIP) oraz możliwość ich odczytu. „Archiwum" odpowiada za długo-

[9] *Reference Model…*, *op. cit*.; *Trusted Digital Repositories. Attributes and Responsibilities. An RLG-OCLC Report* RLG. The Research Libraries Group, Mountain View, CA 2002, [online:] http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf, [dostęp: 16.11.2012].

terminowe przechowywanie i zapewnienie nienaruszalności pakietów AIP, okresowe przenoszenie danych na media nowszej generacji, migrację do aktualnie stosowanych formatów lub systemów, a w przypadku awarii systemu za ich rekonstrukcję. Na żądanie „Archiwum" przekazuje określony pakiet AIP do jednostki „Udostępnianie".

W archiwum elektronicznym niezbędna jest jednostka „Zarządzanie danymi". Jej zadaniem jest utrzymywanie i udostępnianie szerokiego wachlarza informacji. Przykładami mogą być katalogi i inwentarze, na których podstawie można wyszukać i uzyskać określone zasoby z repozytorium, a także statystyki dotyczące udostępniania zasobów. Do zadań tej jednostki należy także kontrola bezpieczeństwa danych oraz inne procedury związane z ochroną repozytorium narzucane przez OAIS.

Poprawne funkcjonowanie całego repozytorium jest uzależnione od prac jednostki administrującej procesami w nim zachodzącymi. „Zarządzanie systemem" zajmuje się negocjowaniem warunków z deponentami, na których podstawie dokumenty są transferowane do repozytorium, czuwa nad kontrolą zgodności dostarczonych dokumentów ze standardami repozytorium oraz przejmuje odpowiedzialność za utrzymywanie sprawności sprzętu i oprogramowania w repozytorium. Czyni także starania na rzecz rozwoju oraz nadzoru nad standardami, potrzebnymi dla funkcjonowania repozytorium.

Repozytoria zgodne z modelem OAIS starają się zapewnić na przyszłość stabilny dostęp do przechowywanych w nich różnorodnych zasobów cyfrowych. W tym celu w OAIS wyodrębniono jednostkę „Planowanie procesu archiwizacji", która zajmuje się m.in. obserwowaniem rozwoju rynku sprzętu i oprogramowania, testowaniem nowych rozwiązań, kontrolą czy archiwizowane dokumenty dadzą się uruchomić i odczytać. Jednostka ta jest odpowiedzialna za wszelkie decyzje dotyczące strategii postępowania, m.in. częstotliwości odświeżania danych, działań mających na celu dostosowanie rozwiązań do zmieniających się warunków sprzętowo-programowych (emulacja lub migracja) i udostępnienia treści dokumentów w zmienionych warunkach.

Proces udostępniania dokumentów cyfrowych w archiwach OAIS został określony terminem *Access*. W ramach udostępniania zasobów użytkownikom umożliwia się przeglądanie zawartości repozytorium poprzez katalogi online, określenie lokalizacji i dostępności określonych zbiorów. Na zamówienie użytkownika system tworzy i wysyła użytkowe pakiety informacyjne typu DIP (Dissemination Information Package – DIP).

Na podstawie przytoczonego opisu możliwe staje się prześledzenie drogi cyfrowego dokumentu przez repozytorium zgodne z założeniami OAIS.

Deponent, który chce odesłać do repozytorium dokument cyfrowy w celu jego długoterminowego przechowania, powinien nadać dokumentowi właściwą – ustaloną wcześniej z repozytorium – formę oraz dołączyć wszelkie dodatkowe informacje o dokumencie wraz z metadanymi. Dokument wraz z metadanymi przesyłany jest w postaci zgłoszeniowego pakietu informacyjnego SIP do działu przyjęcia, gdzie zostaje „rozpakowany" oraz sprawdzony pod względem kompletności i poprawności wszelkich danych niezbędnych do przyjęcia i długoterminowego zarchiwizowania dokumentu. W systemie repozytoryjnym każdy dokument przyjmuje następnie postać AIP, jest zapisywany na serwerze archiwum i przechowywany w sposób umożliwiający jego długotrwałą, stabilną użyteczność. Wygenerowane metadane deponowanych dokumentów są odsyłane do działu, zajmującego się ich zarządzaniem. Archiwizowany obiekt AIP może być przekształcony w formę DIP (*Dissemination Information Package* – DIP), udostępnianą na żądanie użytkownikowi, w postaci umożliwiającej jego zrozumienie.

Autorzy modelu OAIS zapewniają, że może on być stosowany przy organizacji wszelkich repozytoriów (magazynów, archiwów) danych cyfrowych, ze specjalnym przeznaczeniem dla tych, które mają być odpowiedzialne za długoterminowe przechowywanie zdeponowanych zasobów. Do modelu OAIS odwołują się pracownicy Poznańskiego Centrum Superkomputerowo-Sieciowego, autorzy i producenci Usługi Powszechnej Archiwizacji Platon – U4 oraz dArceo, dedykowanych bezpiecznemu i trwałemu składowaniu cyfrowego zasobu polskiej nauki i kultury[10].

6.2. Atrybuty zasobów repozytoryjnych podlegające trwałej ochronie

Funkcjonujące i planowane repozytoria cyfrowe, zwłaszcza te przechowujące dziedzictwo nauki i kultury, dążą do spełniania wymagań instytucji wiarygodnych, autentycznych, stabilnych i niezawodnych (ang. *trusted digital repository, trustworthy digital repositories,* niem. *vertrauenswürdiges digitales Langzeitarchiv*)[11]. Wiarygodne repozytorium cyfrowe to takie, które

[10] *KMD Krajowy Magazyn Danych*, [online:] http://kmd.pcss.pl/index.html [dostęp: 28.11.2012].

[11] *Trusted Digital Repositories. Attributes and Responsibilities*. An RLG-OCLC Report [online:]http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf[dostęp: 12.02.2010]; *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive*. Version 1. [Entwurf zur öffentlichen Kommentierung], „Nestor Materialien" 2006, nr 8, [online:] http://edoc.hu-

gwarantuje dostępność przechowywanych i zarządzanych w nim dokumentów obecnie i w odległej przyszłości, przyjmuje odpowiedzialność za przeprowadzanie prac konserwatorskich w imieniu swoich deponentów oraz na rzecz potrzeb obecnych i przyszłych użytkowników. Dostępność (ang. *availibility*) to atrybut ściśle związany z użytecznością dokumentów, czyli możliwością odczytu i interpretacji ich treści w założonym czasie, przez osobę lub instytucję do tego uprawnioną.

Kolejne atrybuty zasobów repozytoryjnych podlegające ochronie to autentyczność i integralność[12]. Integralność danych (ang. *data integrity*), określana także jako spójność, to funkcja bezpieczeństwa polegająca na zagwarantowaniu niezmienności danych, poprzez blokowanie możliwości ich dodania lub usunięcia w nieautoryzowany sposób. Autentyczność (ang. *authenticity*) natomiast to właściwość odpowiadająca za to, że tożsamość podmiotu lub zasobu jest taka jak zadeklarowana. Atrybut ten jest odpowiedzialny za potwierdzenie, że dokument zdeponowany i udostępniany użytkownikom repozytorium jest dokładnie tym samym dokumentem, którego autentyczność potwierdził deponent na etapie zgłoszenia i przyjmowania go do kolekcji repozytoryjnej.

W technice informatycznej i telekomunikacyjnej ochrona integralności zapobiega przypadkowemu zniekształceniu danych podczas odczytu, zapisu, transmisji lub magazynowania. W celu ochrony integralności danych wykonuje się sumy kontrolne i stosuje kody korekcyjne. W bezpieczeństwie teleinformatycznym natomiast ochrona integralności zapobiega celowej modyfikacji danych dokonanej z użyciem zaawansowanych technik, mających na celu ukrycie faktu dokonania zmiany. Wykorzystuje się tutaj techniki kryptograficzne, np. kody MAC odporne na celowe manipulacje. Systemy repozytoryjne, chroniąc integralność danych, dbają o kompletność zdeponowanej kolekcji, prawdziwość dokumentów i gwarantują, że nie zostały one poddane modyfikacji lub manipulacji[13].

W literaturze przedmiotu, obok ochrony dostępności oraz użyteczności dokumentów cyfrowych, wśród celów, którym długoterminowa ochrona ma służyć, wymienia się także poufność. Przez pojęcie poufności należy rozu-

berlin.de/series/nestor-materialien/2006-8/PDF/8.pdf [dostęp: 20.11.2012]; *Trustworthy Repositories Audit & Certification. Criteria and Checklist. Version 1.0. 2007* [online:] http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf [dostęp: 14.11.2012].

[12] *Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources*. An RLG-OCLC Report. Draft for Public Comment [online:], http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2172.pdf, [dostęp: 15.11.2012].

[13] *Społeczeństwo informacyjne*, red. J. Papińska-Kacperek, Wydawnictwo Naukowe PWN, Warszawa 2008, s. 256–257.

mieć stan, w którym dokument nie jest i nie może być ujawniony osobom, podmiotom bądź procesom nieupoważnionym. Zapewnienie poufności dokumentu może wynikać z takich przesłanek jak ochrona prywatności i interesów własnych deponenta, ochrona interesów instytucji archiwizującej, czyli depozytariusza, obowiązujące akty prawne[14].

Zgodnie z modelem OAIS w jednostce *Planowanie procesu archiwizacji* jest opracowywana strategia trwałej ochrony systemu repozytoryjnego. Ochronie podlegają przede wszystkim dane cyfrowe w postaci kodu zerojedynkowego, ale procedurami bezpieczeństwa obejmuje się również algorytmy dotyczące postępowania z danymi zakodowane w metadanych, oprogramowanie, sprzęt, budynki. Strategia ochrony powinna uwzględniać rozmaite rodzaje zagrożeń; błędy ludzkie, awarie sprzętu, oprogramowania i innych części infrastruktury, katastrofy naturalne, destrukcyjne oprogramowania, włamania do systemów[15].

6.3. Metadane zasobów repozytoryjnych

System repozytoryjny powinien zapewnić odpowiedni stopień bezpieczeństwa i nienaruszalności nie tylko zdeponowanych dokumentów, ale także ich metadanych. Metadane opisują dokumenty przechowywane w systemach repozytoryjnych, ich strukturę, atrybuty, ewentualne modyfikacje, wskazują na wytwórcę dokumentu, jego autora, datę powstania, daty transmisji wewnątrz systemu lub na zewnątrz[16]. Umożliwiają wyszukanie, wykorzystanie i administrowanie dokumentem[17]. W strategii ochrony zasobów cyfrowych metadane odgrywają zasadniczą rolę, ponieważ są jedynym sposobem uchwycenia kontekstu archiwizowanych dokumentów[18], czyli wszelkich informacji o pochodzeniu, procesie powstawania, o czasie i celu powstania, o dotychczasowej hi-

[14] T. Bilski, *Pamięć. Nośniki i systemy*…, *op. cit*., s. 423; *Kriterienkatalog Vertrauenswürdige*…, *op. cit*.

[15] *Reference Model*…, *op. cit*.; *Strategie i modele gospodarki elektronicznej*, red. C.M. Olszak, E. Ziemby, Wydawnictwo Naukowe PWN, Warszawa 2007, s. 399–404.

[16] A. Freedman, *Encyklopedia komputerów,* Wydawnictwo Helion, Gliwice 2004, s 456; J. Adamus, *Metadane w archiwizacji dokumentów elektronicznych*, „Zagadnienia Informacji Naukowej" 2009, nr 2(94), s. 14–15.

[17] M. Nahotko, *Biblioteki cyfrowych książek w środowisku akademickim i bibliotekarskim*, [w:] *Cyfrowy świat bibliotek – problemy techniczne, prawne, wdrożeniowe*, Materiały konferencyjne z IX edycji seminarium z cyklu „Archiwizowanie i digitalizacja", 17–18 stycznia 2006 r., Centrum Promocji Informacji, Warszawa, s. 184.

[18] *Ibidem*, s. 31.

storii, warunkach dostępu, sposobach użytkowania i wszelkich powiązaniach dokumentu z innymi komponentami, pozostającymi w repozytorium bądź poza nim. Metadane powinny wspomóc procesy migracji danych przez kolejne generacje sprzętu komputerowego i oprogramowania, umożliwiać rekonstrukcję procesu decyzyjnego dotyczącego prac na dokumentach cyfrowych, dostarczyć rejestr kontroli dokumentu przez cały cykl jego życia[19].

Metadane zwykle ujmowane są w następujące grupy: metadane opisowe, techniczne, administracyjne, strukturalne oraz prawne, niekiedy określane również jako użytkowe[20].

Z punktu widzenia strategii ochrony użyteczności zasobów repozytoryjnych istotną rolę odgrywają metadane techniczne, w których zakodowana jest informacja o okresie przechowywania, platformie sprzętowo-programowej potrzebnej do odczytu i prezentacji treści dokumentu, opis zastosowanego formatu zapisu publikacji wraz z informacjami o dokumentacji tego formatu oraz informacja o zastosowanym nośniku danych. Metadane techniczne określane bywają jako konserwacyjne[21], dlatego że zawierają informacje o planowanych bądź przeprowadzonych dotychczas pracach konserwatorskich na dokumentach. Dodatkowo mogą być uzupełnione o dokumentację sporządzaną podczas takich prac bądź do niej odsyłać. Techniczne metadane są też istotnym czynnikiem umożliwiającym automatyzację określonych prac konserwatorskich na zasobach cyfrowych, np. migracji i emulacji. Mogą też dostarczać dokładny opis zastosowanego identyfikatora trwałego[22] do dokumentu repozytoryjnego.

Kolejny rodzaj metadanych wspomagających procesy ochrony użyteczności zasobów repozytoryjnych to metadane administracyjne. Dotyczą one

---

[19] C. Lupovici, J. Masanès, *Metadata for the Long Term Preservation of Electronic Publications*, Nedlib Report Series 2. The Hague: Koninklijke Bibliotheek, 2000, s. 3–4.

[20] *Metadata, The Tech Terms Komputer Dictionary 2005*, [online:] http://www.techterms.com/definition/metadata [dostęp: 12.11.2012]; A. Januszko-Szakiel, *Dysertacje via Internet. Projekt elektronicznej archiwizacji rozpraw naukowych w Niemieckiej Bibliotece Narodowej*, „Przegląd Biblioteczny" 2006, z. 2, s. 141.

[21] *Ibidem*.

[22] Identyfikator trwały (PI) to niezmienna (określana też jako stabilna, unikalna, permanentna) nazwa, którą przyporządkowuje się do dokumentu cyfrowego przechowywanego w repozytorium jeden raz na cały cykl jego „życia". Zadaniem PI jest jednoznaczna i trwała identyfikacja dokumentu oraz przynależnych do niego metadanych, niezależnie od miejsca (instytucji), w którym dokument został zapisany i jest archiwizowany, z uwzględnieniem różnorodnych systemów, zmian, z uwzględnieniem występowania obiektów w różnych wersjach, postaciach, formach reprezentacji. Zob. *Persistent Identifiers for Cultural Heritage*, Digital Preservation Europe, [online:] http://www.digitalpreservationeurope.eu/publications/briefs/persistent_identifiers.pdf [dostęp: 20.11.2012].

zarządzania dokumentami cyfrowymi w repozytorium. Zawierają między innymi informacje o istniejących wersjach określonego dokumentu, o sporządzanych kopiach. Ważnym elementem tych metadanych są informacje o takich parametrach dokumentów, jak integralność, autentyczność, o wynikach sporządzania sum kontrolnych *etc*. Metadane administracyjne informują również o uprawnieniach poszczególnych pracowników systemu repozytoryjnego do wykonywania określonych czynności na dokumentach.

6.5. Techniczne aspekty ochrony zasobów cyfrowych

Techniczne aspekty zapewnienia ochrony wiarygodności, autentyczności, integralności oraz poufności zasobów cyfrowych można rozpatrywać z punktu widzenia zabezpieczenia fizycznej infrastruktury repozytorium cyfrowego, polityki uprawnień oraz uwierzytelniania użytkowników systemu, a także metod zabezpieczania danych przed niepowołanym odczytem za pomocą metod kryptograficznych.

Podstawowym wymaganiem, które powinien spełniać każdy system repozytoryjny, jest zapewnienie fizycznej nienaruszalności infrastruktury repozytorium i zgromadzonych w nim danych. Zagrożenia fizycznego bezpieczeństwa systemu mogą wynikać zarówno z nieuprawnionych działań ludzi, nieprawidłowej pracy infrastruktury technicznej, jak i działania sił natury. Aby zminimalizować możliwy wpływ zagrożeń fizycznych, konieczne jest odpowiednie zaprojektowanie, zlokalizowanie oraz wyposażenie budynku repozytorium.

Lokalizacja budynku repozytorium powinna być wybrana w taki sposób, aby zminimalizować możliwość jego podtopienia zarówno przez wody opadowe, jak i powierzchniowe, a także zminimalizować możliwość wystąpienia zagrożeń związanych z działalnością ludzi (np. katastrof lotniczych w lokalizacjach położonych blisko korytarzy powietrznych lub w pobliżu lotniska). Dodatkowym zabezpieczeniem szczególnie cennych zasobów powinno być także wykonywanie kopii zapasowej zgromadzonych zasobów, która jest przechowywana w lokalizacji geograficznie oddalonej od repozytorium. Pozwala to na zachowanie zasobu w przypadku zniszczenia repozytorium w wyniku nieprzewidzianych katastrof naturalnych lub np. działań wojennych.

Właściwie zlokalizowany budynek powinien być także odpowiednio wyposażony w systemy zabezpieczające umieszczone w nim urządzenia techniczne.

Jednym z podstawowych zagrożeń podczas pracy urządzeń elektronicznych jest możliwość wystąpienia pożaru, w wyniku którego bardzo prawdopodobne jest uszkodzenie wrażliwych na wysoką temperaturę urządzeń, a w szczególności nośników danych, co pociąga za sobą ich nieodwracalną utratę. W celu zapobiegania rozprzestrzenianiu się ewentualnego pożaru, a także minimalizacji zniszczeń, jakie może powodować, zalecany jest podział budynku na sekcje rozdzielane ścianami ognioodpornymi oraz zastosowanie urządzeń tłumiących ogień. Jednym z najbardziej skutecznych systemów stosowanych w przypadku pomieszczeń z urządzeniami elektronicznymi są urządzenia, które w przypadku wykrycia oznak pożaru wypierają z pomieszczenia powietrze poprzez wtłoczenie mieszaniny gazów o działaniu gaśniczym zgromadzonych w butlach wysokociśnieniowych. Wtłoczenie do pomieszczenia gazów gaśniczych powoduje wyparcie z niego powietrza, dzięki czemu spada stężenie tlenu niezbędnego do podtrzymywania procesu spalania oraz powoduje gwałtowne obniżenie temperatury poprzez rozprężenie gazów gaśniczych. Ze względu na możliwość przebywania w pomieszczeniu ludzi, konieczny jest jednak dobór parametrów gazowego systemu gaśniczego w taki sposób, aby osoby, którym nie udało się ewakuować z pomieszczenia, mogły przeżyć w atmosferze zmodyfikowanej gazami gaśniczymi.

Systemy zabezpieczające powinny także obejmować kontrolę fizycznego dostępu do urządzeń poprzez wykorzystanie ochrony perymetrycznej wnętrza oraz otoczenia budynku, realizowane za pomocą np. czujników geofonicznych, systemów elektromagnetycznej lub impedancyjnej detekcji ruchu, czujników ruchu pracujących w zakresie ultradźwięków lub podczerwieni, a także systemów monitorowania wizyjnego. Pozwala to na minimalizację możliwości celowego działania osób trzecich, których celem mogłoby być zniszczenie infrastruktury lub też ingerencja w zgromadzony zasób[23].

Kolejnym aspektem, który powinien zostać szczegółowo rozważony, jest uwierzytelnianie użytkowników systemu, a więc weryfikacja ich tożsamości. Stosowane obecnie metody można podzielić na trzy grupy: pierwszą z nich stanowią stosowane od lat rozwiązania oparte na identyfikacji użytkownika za pomocą nazwy użytkownika (login) oraz odpowiadającego mu hasła, drugą grupę rozwiązań stanowią biometryczne systemy identyfikacji i uwierzytelniania użytkowników, natomiast trzecią grupę stanowią rozwiązania będące połączeniem elementów obu wyżej wymienionych rodzajów[24].

---

[23] Systemy ochrony zewnętrznej, [online:] http://www.alkam-security.pl/_cms/view/104/systemy-ochrony-zewnetrznej.html [dostęp: 28.11.2012].

[24] M. Kaeo, *Tworzenie bezpiecznych sieci*, Mikom, Warszawa 2000, *passim*; W. Stallings, *Kryptografia i bezpieczeństwo sieci komputerowych. Matematyka szyfrów i techniki kryptologii*,

Uwierzytelnianie oparte na znajomości identyfikatora (nazwy użytkownika i loginu) i odpowiadającego mu hasła pozwala dowolnej osobie, która zdobędzie te informacje, na dostęp do systemu na prawach przyznanych użytkownikowi, którego tożsamość wykorzystuje. Login i hasło mogą zostać „przejęte" na wiele sposobów, zaczynając od podejrzenia hasła zapisanego przez użytkownika obawiającego się, że je zapomni, poprzez wykorzystanie oprogramowania pozwalającego na generowanie i sprawdzanie poprawności kolejnych kombinacji znaków, po fizyczne wymuszenie ujawnienia danych niezbędnych do uwierzytelnienia.

Drugą grupą metod uwierzytelniania są systemy rozpoznające charakterystyczne cechy osobnicze ludzi – kształt twarzy, wzór tęczówki lub siatkówki oka, brzmienie głosu, geometrię dłoni lub układ linii papilarnych. Projektowane aktualnie systemy biometryczne charakteryzują się coraz wyższą niezawodnością, jednakże niepozwalającą jeszcze na ich zastosowanie jako jedynego źródła uwierzytelniania w systemach, w których wymagany jest bardzo wysoki poziom bezpieczeństwa[25].

W celu podniesienia poziomu bezpieczeństwa systemów uwierzytelniania opartych na metodach biometrycznych mogą być wykorzystywane kombinacje kilku metod biometrycznych, np. skan siatkówki i tęczówki oka wraz z analizą układu linii papilarnych. Zastosowanie równocześnie kilku metod potwierdzenia identyfikacji użytkownika redukuje ryzyko błędnego uwierzytelnienia. Często spotykanym połączeniem jest także wykorzystanie technik biometrycznych jako uzupełnienia klasycznego systemu opartego na loginie i haśle.

Użytkownik uwierzytelniony w systemie może w nim wykonywać działania zgodne z przyznanymi uprawnieniami, odnotowanymi w metadanych technicznych i administracyjnych. Uprawnienia te są zróżnicowane dla danego użytkownika lub grupy użytkowników. Mówi się wówczas o poziomowaniu dostępu do zasobów i operacji, które można na zasobach wykonać (użytkownicy o najniższym poziomie dostępu otrzymują zezwolenie na odczyt ściśle określonych dokumentów, natomiast użytkownicy posiadający status administratora (np. pracownicy działu długoterminowej archiwizacji) posiadają uprawnienia zarówno do odczytu, jak i prac konserwatorskich na danych cyfrowych. Poziom uprawnień przypisanych danemu użytkownikowi może wynikać zarówno z polityki bezpieczeństwa repozytorium, uwarunkowań

Helion, Gliwice 2012, s. 48–49, 408–409; M. Horton, C. Mugge, *Bezpieczeństwo sieci. Notes antyhakera*, Wydawnictwo Translator, Warszawa 2004, s. 124–125.

[25] *Społeczeństwo informacyjne*, red. J. Papińska-Kacperek, Wydawnictwo Naukowe PWN, Warszawa 2008, s. 295.

prawnych, jak i ochrony poufności zasobów. Niezależnie od poziomu uprawnień, wszystkie operacje na danych, wykonywane przez użytkowników wewnętrznych (pracowników repozytorium) oraz użytkowników zewnętrznych (klientów, odbiorców treści deponowanych dokumentów), które mogą wywołać utratę autentyczności dokumentów, powinny być rejestrowane[26].

Właściwe zaplanowanie oraz wdrożenie wspomnianych metod ochrony zgromadzonego zasobu cyfrowego, zarówno przed fizycznym zniszczeniem, jak i nieautoryzowanym fizycznym dostępem, nie gwarantują pełnej ochrony autentyczności, integralności oraz poufności zasobów cyfrowych. Większość repozytoriów cyfrowych świadczy usługi polegające na udostępnianiu treści przechowywanych zasobów za pomocą łącz teleinformatycznych, co powoduje konieczność zapewnienia bezpieczeństwa przesyłanych danych i dostarczenia ich do ściśle określonego odbiorcy. Wykorzystanie łącz teleinformatycznych do udostępniania zasobów powoduje także możliwość podszywania się osób trzecich pod uwierzytelnionego użytkownika, w celu pozyskania poufnych informacji lub też dokonania w repozytorium działań o charakterze destrukcyjnym.

W celu uzyskania wysokiego poziomu bezpieczeństwa na etapie udostępniania dokumentów poprzez łącza teleinformatyczne stosowane są techniki kryptograficzne. Kryptografia jest to dziedzina zajmująca się metodami utajniania informacji poprzez jej szyfrowanie. Dzięki kryptografii można zamienić normalny, zrozumiały tekst lub innego rodzaju wiadomość w taki sposób, że stanie się niezrozumiała dla nieupoważnionego odbiorcy. Właściwy odbiorca wiadomości może po jej otrzymaniu przekształcić ją ponownie do czytelnej postaci. Do niedawna głównymi odbiorcami rozwiązań kryptograficznych były instytucje rządowe, placówki dyplomatyczne oraz wojsko. Rozwój elektronizacji obiegu informacji spowodował, że obszar zastosowań kryptografii znacznie się powiększył. Ze względu na przedmiot szyfrowania można wyróżnić szyfrowanie pojedynczych plików lub systemów plikowych, a także szyfrowanie transmisji[27].

W ostatnich latach postępuje bardzo szybki rozwój kryptografii, która musi sprostać nowym wymaganiom, wynikającym z wcześniej nieznanych

[26] *Reference Model…*, *op. cit*.; *Trusted Digital Repositories. Attributes and Responsibilities*. An RLG-OCLC Report, [online:] http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf [dostęp: 20.11.2012].

[27] *Zabezpieczenia kryptograficzne*, [online:] http://www.b-skrzypczyk.republika.pl/ ; *Społeczeństwo informacyjne…, op. cit*., s. 300–311 [dostęp: 20.11.2012]; D.E. Denning, *Wojna informacyjna i bezpieczeństwo informacji,* Wydawnictwa Naukowo-Techniczne, Warszawa 2002, s. 326–424.

oczekiwań, dotyczących m.in. bezpieczeństwa transmisji danych w sieciach teleinformatycznych. Jednym z owoców gwałtownego rozwoju kryptografii jest powstanie algorytmu AES (Advanced Encryption Standard), którego założenia zostały zdefiniowane właśnie pod kątem zastosowania w internecie – miał być szybki, wydajny i przede wszystkim bezpieczny. Równocześnie jednak okazało się, że nie wystarczy samo zastosowanie silnego szyfru, należy go jeszcze stosować w bezpieczny sposób. Dopiero prawidłowe zaprojektowanie kompletnego systemu poprawnie wykorzystującego zarówno metody uwierzytelniania użytkowników, jak i protokoły kryptograficzne pozwala na niezawodne i bezpieczne działanie w oparciu o sprawdzone algorytmy szyfrujące. W stosunkowo niedługim okresie, jaki upłynął od cywilnego upowszechnienia kryptografii, ustalono pewien ogólny schemat protokołu służącego do bezpiecznej komunikacji. Dotyczy to zarówno doboru odpowiednich algorytmów zapewniających poufność oraz integralność przekazu, jak i metod projektowania i późniejszego testowania danego protokołu. W przypadku protokołów stosowanych obecnie w internecie standardem stał się otwarty model rozwoju, co oznacza, że stworzony przez grupę ekspertów projekt jest następnie przedstawiany do publicznego wglądu oraz oceny. Na tym etapie wskazywane są najsłabsze punkty projektu, które następnie podlegają korekcie. W ten sposób powstały dwa najszerzej obecnie stosowane protokoły kryptograficzne: SSL v.3 oraz IPSec, używany w sieciach VPN. Praktyka dowiodła, że otwartość procesu projektowania takich protokołów i publiczna ocena dodatnio wpływają na ich bezpieczeństwo, podczas gdy utrzymywanie protokołu w tajemnicy skutkuje prawie zawsze jego złamaniem później, kiedy jest on już szeroko wdrożony i kiedy naprawienie błędów oznacza poważne koszty.

Kryptograficzne zabezpieczenie transmisji danych może być realizowane programowo lub sprzętowo. Programowa realizacja procesu kryptograficznego wymaga stosunkowo dużych mocy obliczeniowych, w związku z czym coraz częściej zastosowanie znajdują specjalistyczne urządzenia zabezpieczające transmisję. Przykładem takiego rozwiązania jest rodzina szyfratorów CompCrypt. Są to urządzenia przeznaczone do zapewnienia wysokiego poziomu bezpieczeństwa informacji niejawnych przesyłanych w sieciach teleinformatycznych. CompCrypt wykorzystują Narodowe Algorytmy Szyfrujące – dzięki czemu mogą być wykorzystywane do ochrony informacji niejawnych – od klauzuli „Zastrzeżone" do „Ściśle tajne". Zastosowany w urządzeniach Narodowy Algorytm Szyfrujący jest faktycznie oparty na zmodyfikowanym algorytmie 3DES (rodzaj i zakres modyfikacji jest tajny).

Powodem zastosowania dodatkowych przekształceń jest prawdopodobnie udaremnienie prób wykorzystania sprzętowych implementacji algorytmu 3DES do szukania kluczy deszyfrujących[28].

Strategia ochrony zasobów cyfrowych składowanych i archiwizowanych w repozytoriach instytucjonalnych, nawet jeśli została opracowana z uwzględnieniem wszelkich możliwych zagrożeń oraz sprawdzonych i skutecznych metod, technik i narzędzi zabezpieczenia wszystkich elementów składowych systemu repozytoryjnego, powinna podlegać okresowej rewizji. Bardzo ważne jest obserwowanie zmian technologicznych (tzw. *technology watch*) oraz zmieniających się oczekiwań klientów repozytorium, czyli deponentów i odbiorców deponowanych treści. W systemach repozytoryjnych obowiązuje zarządzanie zmianami, cykliczne weryfikowanie obowiązujących założeń i w razie potrzeby ich aktualizowanie.

Organizatorzy repozytoriów powinni zastanowić się również nad zawarciem umowy z firmą ubezpieczeniową. Odpowiednio dobrana polisa ubezpieczeniowa może stanowić przydatny element polityki zabezpieczenia systemu repozytoryjnego[29].

Podsumowanie

Zadaniem systemów repozytoryjnych jest zagwarantowanie bezpiecznego przechowania i dostępu do kompletnej kolekcji autentycznych zasobów cyfrowych. W celu zabezpieczenia kompletności i prawdziwości zdeponowanych dokumentów konieczne jest opracowanie strategii ochrony oryginalnego kodu zerojedynkowego oraz pozostałych elementów repozytorium, umożliwiających jego udostępnienie, odczyt i przedstawienie w postaci zrozumiałej dla użytkownika.

Zasoby deponowane w repozytoriach instytucjonalnych zwykle są grupowane w kolekcje z określeniem praw dostępu do nich, zakresu ich użytkowania, przyporządkowania metod ochrony, nadania uprawnień i odpowiedzialności za ochronę bezpieczeństwa[30]. Zaleca się powielanie kolekcji depozytowych. Oprócz podstawowej kolekcji repozytoryjnej tworzone są ich tzw. kopie lustrzane i przechowywane w oddalonym miejscu. Możliwe

[28] *Narodowy Algorytm Szyfrujący*, [online:] http://ipsec.pl/kryptografia/narodowy-algorytm-szyfrujacy-nasz.html, [dostęp: 26.11.2012].

[29] T. Kifner, *Polityka bezpieczeństwa i ochrony informacji*, Helion, Gliwice 1999, s. 119–120.

[30] *Strategie i modele gospodarki elektronicznej*, red. C.M. Olszak, E. Ziemba, Wydawnictwo Naukowe PWN, Warszawa 2007, s. 402–404.

jest przechowywanie kopii kolekcji w kilku miejscach; powstaje wówczas rozbudowany system replikacji geograficznych.

W zależności od przyjętych celów i założeń systemu repozytoryjnego, strategia ochrony dokumentów to połączenie rozmaitych metod, narzędzi i technik. Niektóre z nich to: metadane techniczne i administracyjne, identyfikatory trwałe, sumy kontrolne, systemy kryptograficzne, biometryka, uwierzytelnianie.