

PRESERVATION, CONSERVATION AND ARCHIVING OF DIGITAL RESOURCES – SPECIFIC PROBLEMS

Elena TÎRZIMAN

University of Bucharest

Abstract: Massive digitization, production of documents directly in digital format without traditional equivalent, The high dynamics of technologies and of information applications, followed by their equally swift obsolescence, the multiplicity of standards and formats of electronic documents, have lead to the paramount importance of preserving, conserving and archiving digital documents and collections for all who are involved in the production and dissemination of digital content. When digitally preserving, conserving and archiving it is necessary to identify: what should be preserved? for how long? how to actually proceed? by whom? what resources should be involved? which is the legal and institutional framework of digital archiving?.

Those responsible for digital archiving are the producers, the suppliers, the lawful owners who are primarily responsible to archive electronic documents and who ensure their preservation; the communities of intellectual content producers (e.g. the researchers), and various institutions that might organize digital archives (for the long-term) by own resources; national libraries that through their specific mission organize and manage the National Legal Repository ensuring the collection, processing, preservation, conservation and archiving the documentary patrimony of a country, regardless of the support.

Keywords: digital document, digital conservation, digital archiving, legal digital repository.

Introducere

Resursele informaționale digitale au devenit în multe cazuri sursa principală de informare pentru multe domenii ale cunoașterii umane, în special în domeniile științific și profesional. Digitizarea masivă, crearea de documente direct în format digital fără echivalent tradițional, dinamica extraordinară a tehnologiilor și a aplicațiilor informatice urmată și de obsolescența lor la fel de rapidă, multitudinea standardelor și formatelor pentru documente electronice, absența unor norme unitare de lucru pe traseul ciclului de viață al unui document electronic fac ca prezervarea, conservarea și arhivarea documentelor și colecțiilor digitale să devină o problemă de maximă importanță pentru toți furnizorii de conținut digital și mai ales pentru structurile deținătoare de asemenea documente și responsabile cu administrarea patrimoniului instituțional ca parte a moștenirii culturale și științifice naționale.

1. Prezervarea, conservarea și arhivarea digitală – aspecte generale

În prezervarea, conservarea și arhivarea digitală este necesar să se identifice: ce trebuie prezervat și arhivat?, pentru cât timp?, cum anume se realizează acest lucru?, de către cine?, care sunt resursele implicate?, care este cadrul legal și instituțional al arhivării digitale?. Principiile prezervării, conservării și arhivării resurselor informaționale și documentare sunt aceleași atât pentru resursele pe suport tradițional, cât și pentru resursele digitale, diferite fiind doar modalitățile și strategiile concrete de realizare a acestor operații. Prin urmare, trebuie păstrate, prezervate și arhivate resursele care fac parte din patrimoniul instituțional, științific sau cultural conform reglementărilor existente la nivel național sau internațional la care statele au aderat.

Unesco a adoptat *Charte sur la conservation du patrimoine numerique* prin care recunoaște patrimoniul digital ca parte a patrimoniului comun și atrage atenția asupra riscului ca mare parte din moștenirea culturală a umanității să se piardă în absența asumării unor măsuri legislative, tehnologice, financiare, profesionale specifice. Conform acestei *Charte*, patrimoniul digital se compune din documente unice în domeniul cunoașterii și expresiei umane, fie că acestea sunt de ordin cultural, științific, educativ, administrativ, artistic sau conțin informații tehnice, legislative, medicale sau de alt tip, create digital sau convertite sub formă digitală pornind de la resurse tradiționale existente.

Având o tipologie diversă și, de asemenea, o structură complexă documentul digital impune decizii și soluții particularizate în prezervare, conservare, arhivare. Este necesar să fie precizat ce trebuie prezervat; documentul ca întreg (conținut informațional + aplicație informatică) sau doar conținutul informațional. Se disting următoarele caracteristici care asigură integritatea unui document digital: informația conținută; stabilitatea documentului; referința documentului; proveniența documentului; contextul documentului. Toate aceste informații condiționează integritatea documentului și sunt create și trebuie să fie prezervate împreună cu obiectul, pe toată durata lui de viață.

Durata arhivării este dată de importanța resursei din punctul de vedere al conținutului informațional. Strategiile de arhivare digitală sunt stabilite în funcție de nevoile de conservare și reutilizare specifice fiecărei instituție, în cazul arhivării resurselor proprii sau în funcție de prevederile legale, în cazul structurilor instituționale care au obligația de a crea și gestiona arhive digitale (cazul bibliotecilor naționale care administrează depozitul legal, al arhivelor naționale și locale, anumitor producători și furnizori specializați în conținut digital). Se pot stabili strategii specifice pentru categorii de documente și în funcție de suporturile de stocare și chiar în funcție de solicitările de reutilizare anticipate.

Cum anume se realizează arhivarea digitală? Modelele de arhivare pe termen lung a resurselor digitale, rezultate în urma cercetării teoretice și aplicative din domeniile Tehnologiei Informației și Științelor Informării și Comunicării, țin cont de cele două tendințe: asigurarea unei durate de viață cât mai lungi a suportului fizic; reînnoirea continuă a suportului (trecerea pe un alt suport). Concomitent cu evaluarea gradului de perimare a suportului trebuie evaluate și tehnologiile informatice asociate fiecărei resurse electronice. Componentele hardware, programele informatice, metodele de prelucrare a informațiilor sunt supuse unei uzuri morale rapide (ciclul de viață: 2-5 ani). Două soluții sunt propuse ca răspuns la perimarea tehnologiilor:

- migrarea datelor (incluzând sau nu și migrarea formatelor);
- emulația mediilor informatice.

Migrarea datelor înseamnă transferul periodic al unei resurse electronice de pe un suport pe altul, dintr-un context informatic (conținut + program informatic) în altul. Migrarea trebuie să prezerve integritatea unui obiect digital, astfel încât să poată fi utilizat (căutarea informației, afișarea și utilizarea să fie identice cu cele din timpul inițial al producerii resursei informaționale). Migrarea datelor implică o reînnoire a suportului, dar nu este în mod obligatoriu o copie identică, ci un conținut informațional identic într-un nou context tehnic. Se realizează la fiecare schimbare a contextului informatic.

Emulația constă în imitația parțială a unui sistem de către un altul într-un context informatic, asigurarea compatibilităților sistemelor într-un context informatic dat. Tehnica emulațiilor aplicată prezervării pe termen lung a resurselor digitale înseamnă că se încearcă definirea unei metode care va permite compatibilizarea sistemelor informatice învechite cu sistemele încă incomplet cunoscute, cele ale viitorului.

Variante alternative: conservarea datelor și informațiilor pe suport hârtie (suport deja testat în timp); conservarea pe microformate; realizarea de variante electronice de rezervă

nesupuse comunicării; asigurarea redundanței patrimoniului digital supus conservării și arhivării (soluțiile de back-up).

Responsabili privind arhivarea resurselor digitale sunt:

- Creatorii, furnizorii, proprietarii de drept care au responsabilitatea inițială a arhivării documentelor electronice și care asigură prezervarea lor. Ei pot transfera această responsabilitate (prin acorduri clare) unor sisteme de arhivare care preiau total sau parțial responsabilitatea arhivării. (Bibliotecile și arhivele pot prelua responsabilități de arhivare; avantajele economice sunt de ambele părți și în plus, structuri de tipul arhive, biblioteci, muzee pot asigura accesul și utilizarea informațiilor).

- Bibliotecile naționale, prin atribuția specifică de organizare și administrare a Depozitului Legal național, asigură colectarea, prelucrarea, prezervarea, conservarea, arhivarea patrimoniului documentar al unei țări, indiferent de suport. Prin urmare, și resursele digitale constituie obiectul Depozitului Legal. Bibliotecile naționale evaluează problemele tehnice și juridice privind colectarea și conservarea publicațiilor digitale din rețele și își dezvoltă strategii de implementare și realizare a Depozitului Legal pentru aceste resurse digitale.

- Comunitățile de creatori de conținut intelectual (de exemplu cercetătorii), precum și diverse instituții (producătoare de arhive tehnice, economice etc.) pot să-și organizeze arhivele (pe termen lung) prin resurse proprii sau prin organizații de cooperare independente, organizații care au obligația de a asigura accesul la arhive în condițiile reglementărilor legale existente.

Cadrul legal al arhivării digitale lipsește sau este insuficient dezvoltat pentru majoritatea statelor. Se regăsesc mai curând bune practici, inițiative private, proiecte și cercetări științifice pe această problematică, standarde și norme tehnice. Abordările care reglementează arhivarea conținutului digital sunt la nivel instituțional, național, european sau internațional și se întâlnesc cu precădere în domeniile cultural și științific, adică în acele domenii care sunt implicate inclusiv în producerea și administrarea patrimoniului documentar în format tradițional.

2. Arhivare web

Imensa cantitate de resurse informaționale disponibile pe web, instabilitatea și dinamica acestora fac să fie imposibilă o arhivare exhaustivă. Fără politici și strategii clare de arhivare web, resurse informaționale importante riscă să fie pierdute. Asemenea politici și strategii pot fi rezultatul inițiativelor individuale, instituționale sau ale unor proiecte colaborative și se pot aplica la nivelul unei instituții sau pentru un site individual, la nivelul unui grup de site-uri, la nivelul unui domeniu tematic dintr-o zonă geografică sau la nivelul unei porțiuni de Internet (identificată prin numele de domeniu, de exemplu: .en; .fr).

Arhivarea web este condiționată de o serie de probleme tehnice și legale. Din categoria problemelor tehnice se pot identifica: instalarea și întreținerea unei infrastructuri IT capabile să facă, în orice moment, lizibile documentele digitale; utilizarea programelor informatice de comprimare a conținutului și a serviciilor de protecție a paginilor web; prezervarea contextului de lectură adică prezervarea unui hyperdocument cu toate linkurile sale, astfel încât să nu fie afectată semnificația mesajului informațional; utilizarea standardelor tehnice și profesionale în crearea documentelor digitale, astfel încât să poată fi definite clar datele de identificare, proveniență și autenticitate necesare pentru arhivare. Aspectele legale și juridice au în vedere drepturile de autor și drepturile conexe care pot determina restricții în accesul la documente incluse în arhive.

Se pot identifica două modele de arhivare:

- model bazat pe un sistem de arhivare închis în care primează obiectivul prezervării și conservării conținutului digital, accesul fiind restricționat și controlat;

- model bazat pe un sistem de arhivare deschis în care sunt urmărite atât obiectivul prezervării și conservării, cât și al accesului liber la resursele arhivate. Sistemul de arhivare deschis se poate aplica oricărei arhive web și în special instituțiilor și organizațiilor cu responsabilități de arhivare pe termen lung și care arhivează resurse informaționale din domeniul public.

Arhivarea conținutului unui site de către producătorul sau administratorul său. Site-ul web este definit ca un sistem de publicare a resurselor informaționale și, prin urmare, este nevoie ca în funcție de dinamica conținutului digital publicat sau la anumite intervale de timp să fie actualizat conținutul și să se decidă ce se adaugă, ce se modifică și ce se elimină. Resursele digitale care se vor elimina pot fi arhivate sau pot fi eliminate definitiv. Arhivarea web are în vedere acele resurse digitale pentru care de ia decizia de a fi păstrate ca documente referențiale și care sunt considerate importante pentru consultări și utilizări viitoare.

Arhivarea conținutului digital web ar trebui să se facă la anumite intervale de timp (lunar, semestrial, anual), atunci când conținutul informațional nu mai este de actualitate, când există resurse sau secvențe redundante în același site, când se decide schimbarea formei de prezentare a site-ului și dezvoltarea acestuia. Arhivarea conținutului site-ului se poate realiza prin definirea în site a unei secțiuni de arhivare în care se transferă conținutul care se dorește a fi arhivat și care ulterior este gestionat identic și unitar cu restul site-ului; se poate crea o arhivă zip de lucru oferită sau nu în acces liber sau se poate arhiva offline pe CD sau DVD.

Backup nu înseamnă arhivarea site-ului chiar dacă arhivarea poate implica backup. Backup-ul presupune realizarea de copii identice de siguranță pe servere diferite sau chiar în locații diferite, ca o modalitate de gestionarea a riscurilor de eventuale catastrofe, atacuri informatice sau imposibilitate de acces de pe serverul principal. Arhivarea unui site presupune selectarea și prezervarea, conservarea acelui conținut digital care nu mai este de actualitate, dar care este important pe termen mediu și lung și care, prin urmare, nu poate fi eliminat.

Arhivarea conținutului web de către producătorii, furnizorii sau intermediarii de conținut digital. Arhivarea web este procesul de colectare de porțiuni din World Wide Web cu scopul prezervării și conservării conținutului digital, în aceleași condiții și reprezentare precum arhivarea site-urilor, în vederea asigurării accesului viitor pentru cercetători, istorici sau public larg. Numărul imens de pagini și site-uri web existente face ca arhivarea web, chiar parțială și selectivă, să se realizeze automatizat și doar cu unele excepții, precum în cazul bibliotecilor și arhivelor, să implice și factorul uman și să se realizeze semiautomatizat.

Există numeroase structuri instituționale implicate în arhivarea web. Cea mai cunoscută organizație este Internet Archive care își propune ca, printr-un demers de culegere automatizată a datelor, să arhiveze întreg spațiul web. Bibliotecile naționale, arhivele naționale, diferite consorții și organizații sunt implicate în arhivarea resurselor web culturale. De asemenea, există numeroase aplicații de arhivare web, precum și firme care furnizează servicii de acest fel proprietarilor de conținut digital.

Din punct de vedere tehnic, arhivarea web presupune utilizarea a o serie de tehnici, proceduri, protocoale și standarde specifice. Deosebit de importante sunt asigurarea integrității și stabilității conținutului digital (toate paginile și linkurile care constituie site-ul web, toate imaginile fixe și animate, orice alte aplicații incluse, precum JavaScript, etc.); arhivarea metadatelor care definesc resursa digitală în scopul stabilirii autenticității și provenienței acesteia; asigurarea permanentă a interoperabilității sistemelor în cazul în care arhivarea și accesul se fac distribuit, și nu centralizat.

Se disting mai multe metode de colectare a metadatelor și resurselor digitale și de constituire a arhivelor digitale, în funcție de strategia de arhivare aleasă. Se poate opta pentru o strategie de arhivare descentralizată care presupune colectarea metadatelor într-o arhivă unitară și arhivarea propriu-zisă a conținutului digital (arhivarea site-ului în locația și formele în care acesta se arhivează). Metoda este utilă în cazul arhivării de site-uri numeroase, cu un

conținut uniform și stabil, dar localizate dispersat. Se poate opta pentru o strategie de arhivare centralizată care presupune colectarea atât a metadatelor de identificare, cât și a conținutului digital și a site-urilor implicite. O asemenea metodă este utilă în cazul în care se consideră că resursele digitale sunt deosebit de importante și trebuie să fie conservate pe termen lung sau chiar că trebuie să fie avută în vedere conservarea perenă (în special site-urile culturale și științifice). Mai poate exista și o strategie mixtă de arhivare care permite ca alături de metadata, pentru unele site-uri să fie colectat și arhivat centralizat și conținutul lor. Este cazul site-urilor volatile sau instabile, dar care au totuși un conținut informațional relevant.

Ținând cont de nivelurile de interoperabilitate existente în rețele, se regăsesc următoarele metode de colectare a resurselor digitale: *Gathering* sau colectare liberă prin utilizare Web Crawler pentru diverse forme de căutare și colectare a resurselor de la diferite organizații; *Harvesting* sau culegerea datelor într-o manieră autonomă, adică definirea unui set limitat de date și servicii și realizarea schimbului pe baza unui protocol (OAI – Open Archive Initiative); *Federation* sau implementarea unor standarde specializate pe care diferite instituții, în special bibliotecile, au convenit să le utilizeze (Z39.50 și MARC etc).

Problemele specifice arhivării web sunt legate de cele trei activități principale: stocarea datelor, migrarea datelor și recuperarea datelor.

Problemele legate de stocarea datelor:

- structura instituțională cu atribuții de arhivare web trebuie să își desfășoare activitatea în conformitate cu o legislație specifică (legislație de arhivare digitală sau depozit legal digital și legislația drepturilor de autor);

- parcurgerea site-urilor web și arhivarea resurselor să se facă la intervale de timp adecvate, astfel încât să fie surprinsă dinamica site-urilor și să poată fi arhivate și resursele digitale care sunt modificate sau înlăturate de pe site;

- în cazul resurselor științifice, stocarea și arhivarea necorespunzătoare pot produce disfuncționalități în citarea referințelor web. De asemenea, este posibil ca în prelucrare să fie descrisă pagina principală, dar resursele să se găsească în paginile secundare;

- sistemele de arhivare nu pot parcurge, recunoaște și stoca toate tipurile de resurse din Internet, din toate tipologiile documentare. Este posibil să nu poată fi arhivate baze de date, sesiuni interactive din web, web-ul profund;

- costurile arhivării web determină o preferință pentru arhivarea selectivă sau tematică.

Problemele legate de migrarea datelor:

- formatele de stocare a datelor, sistemele hardware și software evoluează extrem de rapid; se pune astfel întrebarea: la ce interval de timp trebuie să se realizeze migrarea datelor de pe un suport inferior pe un suport superior?

Problemele legate de recuperarea datelor:

- calitatea prelucrării resurselor digitale care sunt apoi arhivate determină calitatea regăsirii acestora în arhivele web. În arhivele web, resursele digitale sunt în cele mai multe cazuri reprezentate de metadatale preluate împreună cu acestea dar pot fi adăugate la arhivare se pot chiar realiza notițe bibliografice/webgrafice și documentații special pentru arhivare;

- în lanțul producerii, publicării, arhivării unei resurse web sunt implicați mai mulți responsabili, prin urmare, drepturile de autor, drepturile de publicare, drepturile de acces influențează regăsirea și utilizarea acestor resurse.

Concluzii

Dezvoltarea extraordinară a conținutului digital, în diverse tipologii și formate face ca problema arhivării digitale să devină la randul ei o problemă de maximă importanță. Se disting câteva aspecte care evidențiază complexitatea arhivării digitale, În primul rând, diversitatea de tipologii și formate pentru resursele digitale la care se adaugă o diversitate de tehnologii și programe informatice implicate în gestionarea și administrarea acestor resurse. În al doilea rând, multitudinea de furnizori de conținut digital cu responsabilități diverse în

arhivare. În al treilea rând, absența unui cadru legal coerent și care să coroboreze legislațiile naționale cu cele internaționale, privind arhivarea digitală. În al patrulea rând, conținutul digital este extrem de dinamic și nu este posibilă conservarea și arhivare exhaustivă a acestuia cu toate variantele sale intermediare. De asemenea, nu se poate vorbi de norme, tehnici și metodologii unitere de arhivare digitală care să permită uniformizarea unor asemenea practici. Prin urmare, problematica arhivării digitale este în continuă dezvoltare oferind piste de cercetare teoretică și aplicativă multiple în perioada următoare.

Bibliografie

ANGELAKI, Georgia. Digital Legal Deposit in the EU member states: An Overview of Regulatory and Implementation Status Background Report in connection with the Communication on Digital Libraries. Background Report in connection with the Communication on Digital Libraries. [online]. DG INFSO/ Unit E3: Learning and Cultural Heritage February 2006. Disponibil pe Internet: <http://web3.nlib.ee/cenl/docs/Digital%20Legal%20Deposit%20in%20the%20EU%20member%20states.pdf> [accesat 10.11. 2014];

BESSER, Howard. Digital longevity. În: SITTS, Maxine (ed.) *Handbook for Digital Projects: A Management Tool for Preservation and Access*, Andover MA: Northeast Document Conservation Center. [online]. 2000. p. 155-166. Disponibil pe Internet: <http://besser.tsoa.nyu.edu/howard/Papers/sfs-longevity.html> . [accesat 10.11.2014].

CASTILLO-SY, Vivian del. Web-Harvesting: Concept, Issues and Prospects. [online]. p. 1-2. Disponibil pe Internet: paarl.wikispaces.com/file/view/Web-Harvesting.doc. [accesat 10.11.2014].

KUMAR, Shailendra. Interoperability Protocols and Standards in Library and Information Science. [online]. 30 p. Disponibil pe Internet: <http://www.slideshare.net/alibnetweb/interoperability-protocols-and-standards-in-lis> . [accesat 10.11.2014].

UNESCO. Charte de l'Unesco sur la conservation du patrimoine numérique. [online]. 2003, 5 p. Disponibil pe Internet: http://portal.unesco.org/ci/fr/files/13367/10702838853Charter_fr.pdf/Charter_fr.pdf [accesat 10.11. 2014].

UTAS [University of Tasmania]. Web & Learning Services, Archive your website. [online]. Disponibil pe Internet: <http://www.web-services.utas.edu.au/resources/site-management/archive-your-website>. [accesat 27.08.2011].

TÎRZIMAN, Elena. Resurse informaționale și digitale electronice. Aspecte privind construcția, achiziția, prelucrarea, utilizarea, conservarea, conservarea, arhivarea. În: *Cercetare și dezvoltare în Bibliotecile Naționale ale României și Republicii Moldova: volum de lucrări științifice*. București, Chișinău: Editura BNR, Editura BNRM, 2008. p. 43-67.

TÎRZIMAN, Elena. *Resurse Internet și practici de informare*. București: Editura Universității din București, 2011, ISBN : 978-606-16-0038-0, 264 p.

TÎRZIMAN, Elena. *Colecții și biblioteci digitale*. București: Editura Universității din București, 2011, ISBN : 978-973-737-979-5, 274 p.