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 informaţionale digitale au devenit în multe cazuri sursa principală de informare pentru multe domenii ale cunoaşterii umane, în special în domeniile ştiinţific şi profesional. Digitizarea masivă, crearea de documente direct în format digital fără echivalent tradiţional, dinamica extraordinară a tehnologiilor şi a aplicaţiilor informatice urmată şi de obsolenţa lor la fel de rapidă, multitudinea standardelor şi formatelor pentru documente electronice, absenţa unor norme unitare de lucru pe traseul ciclului de viaţă al unui document electronic fac ca prezervarea, conservarea şi arhivarea documentelor şi colecţiilor digitale să devină o problemă de maximă importanţă pentru toţi furnizorii de conţinut digital şi mai ales pentru pentru structurile deţinătoare de asemenea documente şi responsabile cu administrarea patrimoniului instituţional ca parte a moştenirii culturale şi ştiinţifice naţionale.

1. Prezervarea, conservarea şi arhivarea digitală – aspecte generale

În prezervarea, conservarea şi arhivarea digitală este necesar să se identifice: ce trebuie prezervat şi arhivat?, pentru cât timp?, cum anume se realizează acest lucru?, de către cine?, care sunt resursele implicate?, care este cadrul legal şi instituţional al arhivării digitale?. Principiile prezervării, conservării şi arhivării resurselor informaţionale şi documentare sunt aceleaşi atât pentru resursele pe suport tradiţional, cât şi pentru resursele digitale, diferite fiind doar modalităţile şi strategiile concrete de realizare a acestor operaţii. Prin urmare, trebuie păstrate, prezervate şi arhivate resursele care fac parte din patrimoniul instituţional, ştiinţific sau cultural conform reglementărilor existente la nivel naţional sau internaţional la care statele au aderat.

BDD-A21834 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:57:19 UTC)

Unesco a adoptat *Charte sur la conservation du patrimoine numerique* prin care recunoaşte patrimoniul digital ca parte a patrimoniului comun şi atrage atenţia asupra riscului ca mare parte din moştenirea culturală a umanităţii să se piardă în absenţa asumării unor măsuri legislative, tehnologice, financiare, profesionale specifice. Conform acestei *Charte*, patrimoniul digital se compune din documente unice în domeniul cunoaşterii şi expresiei umane, fie că acestea sunt de ordin cultural, ştiintific, educativ, administrativ, artistic sau conţin informaţii tehnice, legislative, medicale sau de alt tip, create digital sau convertite sub formă digitală pornind de la resurse tradiţionale existente.

Având o tipologie diversă şi, de asemenea, o structură complexă documentul digital impune decizii şi soluţii particularizate în prezervare, conservare, arhivare. Este necesar să fie precizat ce trebuie prezervat; documentul ca întreg (conţinut informaţional + aplicaţie informatică) sau doar conţinutul informaţional. Se disting următoarele caracteristici care asigură integritatea unui document digital: informaţia conţinută; stabilitatea documentului; referinţa documentului; provenienţa documentului; contextul documentului.Toate aceste informaţii condiţionează integritatea documentului şi sunt create şi trebuie să fie prezervate împreună cu obiectul, pe toată durata lui de viaţă.

Durata arhivării este dată de importanţa resursei din punctul de vedere al conţinutului informaţional. Strategiile de arhivare digitală sunt stabilite în funcţie de nevoile de conservare şi reutilizare specifice fiecărei instituţie, în cazul arhivării resurselor proprii sau în funcţie de prevederile legale, în cazul structurilor instituţionale care au obigaţia de a crea şi gestiona arhive digitale (cazul bibliotecilor naţionale care administrează depozitul legal, al arhivelor naţionale şi locale, anumitor producători şi furnizori specializaţi în conţinut digital). Se pot stabili strategii specifice pentru categorii de documente şi în funcţie de suporturile de stocare şi chiar în funcţie 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 şi aplicative din domeniile Tehnologiei Informaţiei şi Ştiinţelor Informării şi Comunicării, ţin cont de cele două tendinţe: asigurarea unei durate de viaţă 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 şi tehnologiile informatice asociate fiecărei resurse electronice. Componentele hardware, programele informatice, metodele de prelucrare a informaţiilor sunt supuse unei uzuri morale rapide (ciclul de viaţă: 2-5 ani). Două soluţii sunt propuse ca răspuns la perimarea tehnologiilor:

- migrarea datelor (incluzând sau nu şi migrarea formatelor);
- emulaţia mediilor informatice.

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

Emulaţia constă în imitaţia parţială a unui sistem de către un altul într-un context informatic, asigurarea compatibilităţilor sistemelor într-un context informatic dat. Tehnica emulaţiilor 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 şi informaţiilor pe suport hârtie (suport deja testat în timp); conservarea pe microformate; realizarea de variante electronice de rezervă

BDD-A21834 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:57:19 UTC)

nesupuse comunicării; asigurarea redundanţei patrimoniului digital supus conservării şi arhivării (soluţiile de back-up).

Responsabili privind arhivarea resurselor digitale sunt:

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

- Bibliotecile naţionale, prin atribuţia specifică de organizare şi administrare a Depozitului Legal naţional, asigură colectarea, prelucrarea, prezervarea, conservarea, arhivarea patrimoniului documentar al unei ţări, indiferent de suport. Prin urmare, şi resursele digitale constituie obiectul Depozitului Legal. Bibliotecile naţionale evaluează problemele tehnice şi juridice privind colectarea şi conservarea publicaţiilor digitale din reţele şi îşi dezvoltă stategii de implementare şi realizare a Depozitului Legal pentru aceste resurse digitale.

- Comunităţile de creatori de conţinut intelectual (de exemplu cercetătorii), precum şi diverse instituţii (producătoare de arhive tehnice, economice etc.) pot să-şi organizeze arhivele (pe termen lung) prin resurse proprii sau prin organizaţii de cooperare independente, organizaţii care au obligaţia de a asigura accesul la arhive în condiţiile reglementărilor legale existente.

Cadrul legal al arhivării digitale lipseşte sau este insuficient dezvoltat pentru majoritatea statelor. Se regăsesc mai curând bune practici, iniţiative private, proiecte şi cercetări ştiintifice pe această problematică, standarde şi norme tehnice. Abordările care reglementează arhivarea conţinutului digital sunt la nivel instituţional, naţional, european sau internaţional şi se întâlnesc cu precădere în domeniile cultural şi ştiinţific, adică în acele domenii care sunt implicate inclusiv în producerea şi administrarea patrimoniului documentar în format tradiţional.

2. Arhivare web

Imensa cantitate de resurse informaţionale disponibile pe web, instabilitatea şi dinamica acestora fac să fie imposibilă o arhivare exhaustivă. Fără politici şi strategii clare de arhivare web, resurse informaţionale importante riscă să fie pierdute. Asemenea politici şi strategii pot fi rezultatul iniţiativelor individuale, instituţionale sau ale unor proiecte colaborative şi se pot aplica la nivelul unei instituţii 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 porţiuni de Internet (indentificată prin numele de domeniu, de exemplu: .en; .fr).

Arhivarea web este condiţionată de o serie de probleme tehnice şi legale. Din categoria problemelor tehnice se pot identifica: instalarea şi întreţinerea unei infrastructuri IT capabile să facă, în orice moment, lizibile documentele digitale; utilizarea programelor informatice de comprimare a conţinutului şi a serviciilor de protecţie a paginilor web; prezervarea contextului de lectură adică prezervarea unui hyperdocument cu toate linkurile sale, astfel încât să nu fie afectată semnificaţia mesajului informaţional; utilizarea standardelor tehnice şi profesionale în crearea documentelor digitale, astfel încât să poată fi definite clar datele de identificare, provenienţă şi autenticitate necesare pentru arhivare. Aspectele legale şi juridice au în vedere drepturile de autor şi drepturile conexe care pot determina restricţii 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 şi conservării conţinutului digital, accesul fiind restricţionat şi controlat;

BDD-A21834 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:57:19 UTC)

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

Arhivarea conţinutului unui site de către producătorul sau administratorul său. Site-ul web este definit ca un sistem de publicare a resurselor informaţionale şi, prin urmare, este nevoie ca în funcţie de dinamica conţinutului digital publicat sau la anumite intervale de timp să fie actualizat conţinutul şi să se decidă ce se adaugă, ce se modifică şi ce se elimină. Resursele digitale care se vor elimina por 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 referenţiale şi care sunt considerate importante pentru consultări şi utilizări viitoare.

Arhivarea conţinutului digital web ar trebui să se facă la anumite intervale de timp (lunar, semestrial, anual), atunci când conţinutul informaţional nu mai este de actualitate, când există resurse sau secvenţe redundante în acelaşi site, când se decide schimbarea formei de prezentare a site-ului şi dezvoltarea acestuia. Arhivarea conţinutului site-ului se poate realiza prin definirea în site a unei secţiuni de arhivare în care se transferă conţinutul care se doreşte a fi arhivat şi care ulterior este gestionat identic şi 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 siguranţă pe servere diferite sau chiar în locaţii diferite, ca o modalitate de gesionarea a riscurilor de eventuale catastrofe, atacuri informatice sau imposibilitate de acces de pe serverul principal. Arhivarea unui site presupune selectarea şi prezervarea, conservarea acelui conţinut digital care nu mai este de actualitate, dar care este important pe termen mediu şi lung şi care, prin urmare, nu poate fi eliminat.

Arhivare conţinutului web de către producătorii, furnizorii sau intermediari de conţinut digital. Arhivarea web este procesul de colectare de porţiuni din World Wide Web cu scopul prezervării şi conservării conţinutului digital, în aceleaşi condiţii şi reprezentare precum arhivarea site-urilor, în vederea asigurării accesului viitor pentru cercetători, istorici sau public larg. Numărul imens de pagini şi site-uri web existente face ca arhivarea web, chiar parţială şi selectivă, să se realizeze automatizat şi doar cu unele excepţii, precum în cazul bibliotecilor şi arhivelor, să implice şi factorul uman şi să se realizeze semiautomatizat.

Există numeroase structuri instituţionale implicate în arhivarea web. Cea mai cunoscută organizaţie este Internet Archive care îşi propune ca, printr-un demers de culegere automatizată a datelor, să arhiveze întreg spaţiul web. Bibliotecile naţionale, arhivele naţionale, diferite consorţii şi organizaţii sunt implicate în arhivarea resurselor web culturale. De asemenea, există numeroase aplicaţii de arhivare web, precum şi firme care furnizează servicii de acest fel proprietarilor de conţinut digital.

Din punct de vedere tehnic, arhivarea web presupune utilizarea a o serie de tehnici, proceduri, protocoale şi standarde specifice. Deosebit de importante sunt asigurarea integrităţii şi stabilităţii conţinutului digital (toate paginile şi linkurile care constituie site-ul web, toate imaginile fixe şi animate, orice alte aplicaţii incluse, precum JavaScript, etc.); arhivarea metadatelor care definesc resursa digitală în scopul stabilirii autenticităţii şi provenienţei acesteia; asigurarea permanentă a interoperabilităţii sistemelor în cazul în care arhivarea şi accesul se fac distribuit, şi nu centralizat.

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

BDD-A21834 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:57:19 UTC)

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

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

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

Problemele legate de stocarea datelor:

- structura instituţională cu atribuţii de arhivare web trebuie să îşi desfăşoare activitatea în conformitate cu o legislaţie specifică (legislaţie de arhivare digitală sau depozit legal digital şi legislaţia drepturilor de autor);

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

- în cazul resurselor ştiinţifice, stocarea şi arhivarea necorespunzătoare pot produce disfuncţionalităţi în citarea referinţelor 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, recunoaşte şi 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 preferinţă pentru arhivarea selectivă sau tematică.

Problemele legate de migrarea datelor:

- formatele de stocare a datelor, sistemele hardwere şi softwere 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 metadatele preluate împreună cu acestea dar pot fi adăugate la arhivare se pot chiar realiza notiţe bibliografice/webgrafice şi documentaţii special pentru arhivare;

- în lanţul producerii, publicării, arhivării unei resurse web sunt implicaţi mai mulţi responsabili, prin urmare, drepturile de autor, drepturile de publicare, drepturile de acces influenţează regăsirea şi utilizarea acestor resurse.

Concluzii

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

BDD-A21834 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:57:19 UTC)

arhivare. În al treile rând, absenţa unui cadru legal coerent şi care să coroboreze legislaţiile naţionale cu cele internaţionale, privind arhivarea digitală. În al patrulea rând, conţinutul digital este extrem de dinamic şi nu este posibilă conservarea şi arhivare exhaustivă a acestuia cu toate variantele sale intermediare. De asemenea, nu se poate vorbi de norme, tehnici şi 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 şi 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. *I*nteroperability 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 informaţionale şi digitale electronice. Aspecte privind construcţia, achiziţia, prelucrarea, utilizarea, conservarea, conservarea, arhivarea. În: *Cercetare şi dezvoltare în Bibliotecile Naţionale ale României şi Republicii Moldova: volum de lucrări ştiinţifice*. Bucureşti, Chişinău: Editura BNR, Editura BNRM, 2008. p. 43-67.

TÎRZIMAN, Elena. *Resurse Internet şi practici de informare*. Bucureşti: Editura Universităţii din Bucureşti, 2011, ISBN : 978-606-16-0038-0, 264 p.

TÎRZIMAN, Elena. *Colecţii şi biblioteci digitale*. Bucureşti: Editura Universităţii din Bucureşti, 2011, ISBN : 978-973-737-979-5, 274 p.

BDD-A21834 © 2014 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:57:19 UTC)