

## DIGITIZAREA – O METODĂ MODERNĂ DE CONSERVARE A CĂRȚILOR ȘI A DOCUMENTELOR MEDICALE VECHI

Șef lucr.dr. *Alina Mihaela Pascu*, prof.univ.dr.*Angela Repanovici*,  
prof.univ.dr.*Liliana Rogoza*, mat. *Gabriela Mailat*  
Universitatea „Transilvania” Brașov

### Abstract:

Digitization is a means of representing an object, image, document or a signal (usually an analogue signal) by a discrete set of its points or, respectively, by the reduction of a continuous signal to a discrete signal. The result is called "digital representation" or, more specifically, a "digital image", for the object, and "digital form", for the signal. The method of digitization represents a breakthrough for documents preservation and conservation in libraries. Digitization is a method of creating digital surrogates of analog materials such as books, newspapers, microfilms, and videotapes. Digital preservation comprises also a series of activities aimed towards ensuring access to digital materials over time, with major advantages. The facsimile provided by digitization is available for the users without any strain concerning any further deterioration of the old original documents, turned fragile over time. Digitization significantly improves the process of information dissemination by providing full access to documents of all kind, especially to old, rare, and extremely valuable ones. Moreover, the method significantly reduces amounts of physical storage space in a library.

The present paper shows *Transilvania* University Brasov's Library experience in old medical documents digitization, as part of an European project aimed to provide an old medical digitized book database, to develop online information dissemination services, and to form digitization-skilled personnel.

**Key words:** old medical book, digitization, information dissemination.

### Introducere. Stadiul actual al cercetării în domeniu

Pe plan mondial proiectele de digitizare atrag tot mai mulți cercetători, ingineri, specialiști în calculatoare sau în științele informării și comunicării. Interesul mare în această direcție se explică prin fondurile substanțiale care au început să fie investite în acest tip de cercetare de ultimă oră. Prezentăm câteva exemple:

- Colorado Digitization Project [16, 19] a fost inițiat în 1998 și a reunit eforturile bibliotecilor, muzeelor, societăților istorice și arhivelor din Colorado pentru a crește accesul utilizatorilor la colecțiile speciale și resursele unice deținute de aceste instituții. Proiectul presupune un catalog comun de metadata și a dezvoltat instrumente speciale pentru creatorii de baze de metadata. Acest proiect investighează în același timp folosirea clasificării zecimale Dewey prin *WebDewey* pentru a permite accesul pe subiecte cheie la catalogul comun.
- Un alt exemplu este cel al conferințelor IFLA (*International Federation of Library Association*) care au inițiat secțiuni speciale și seminarii dedicate digitizării. Conferința IFLA

din anul 2003, desfășurată la *Isik Universitesi Library* din Turcia [24, 27] a avut ca temă de dezbateră subiectul digitizării: *Introducing Digitization into Turkish Libraries: Current Attitudes and the Way Forward*.

- Un alt proiect important și complex de digitizare a fost lansat în Spania. AGID (*Archivo General de Indias*) a inițiat un proiect deosebit de ambițios pentru digitizarea colecțiilor sale. Mai mult de un milion de pagini de documente privind istoria Spaniei au fost digitizate. Pe lângă complexitatea proiectului de digitizare, a accesului la documente digitale, s-au dezvoltat și soluții de *hardware* și *software*. Pe termen lung, obiectivul proiectului nu este doar accesibilizarea documentelor, ci și prezervarea lor.
- Noile tehnici de digitizare recent dezvoltate combină în mod optim robotica, electronica digitală și tehnica de calcul, prin crearea liniei de digitizare care oferă în premieră o soluție complet integrată. Cele două avantaje majore ale liniei de digitizare sunt productivitatea foarte ridicată și menținerea integrității fizice a cărții supuse procesului de digitizare. Manipularea documentelor, transferul și

digitizarea lor sunt automate, asigurându-se prezervarea fidelă a conținutului informației, precum și a formei originale a documentului indiferent de tipul acestuia (cărți, reviste, colecții de ziare).

- În octombrie 2002 firma elvețiană *4DigitalbooksASSY*, cea care a conceput *DIGITIZING LINE* [13] a instalat un prim echipament de acest tip la *Green Library*, una dintre bibliotecile Universității *Stanford, California*. Rezultat direct al unui prototip pentru formatul maxim A4, prima versiune industrială a *DIGITIZING LINE* permite scanarea complet automată a volumelor sau a ziarelor până la formatul A2. Echipamentul realizează în premieră mondială întoarcerea automată a paginii pentru întreaga gamă de formate de pagină [1, 4, 5].

Bibliotecile Universității *Stanford* au în total aproximativ opt milioane de cărți digitizate, printre care toate titlurile publicate de *Stanford University Press*, inclusiv conservarea lucrărilor începând cu anul 1923.

În prezent, sistemul *DIGITIZING LINE* instalat la *Stanford* funcționează la capacitate maximă, scanând câteva mii de cărți anual. Michael Keller, director al Bibliotecii Universității din *Stanford*, explică: „Un avantaj: cărțile nu părăsesc niciodată mâinile noastre. Nu trebuie să le retragem din circuit șase sau chiar zece săptămâni. Dacă pentru procesarea unei cărți de 300 pagini sunt necesare 20 de minute, atunci un raft pe zi e o nimica toată. Produsul final e sub controlul nostru.”

Proiectul prin care s-a dezvoltat linia de digitizare la Biblioteca Universității din *Southampton* se numește *BOPCRIS* [17, 18, 26]. Una din principalele lui activități este scanarea documentelor parlamentare și istorice rare folosind *DIGITIZING LINE*, un echipament de înaltă precizie, folosit pentru prima dată în iunie 2004 în Marea Britanie într-o bibliotecă de cercetare și a doua oară în acest domeniu. Simon Brackenbury, director de proiecte de digitizare la *Southampton*, afirmă: „Într-un anume fel, acest proiect este un studiu de caz în automatizare. Linia de digitizare are o productivitate ridicată, (scanează 600 de pagini / oră, propunându-ne să scanăm 1.000.000 într-un an), permițându-ne scanarea unui număr mare de documente, ceea ce va face posibil accesul la documentele istorice, inaccesibile utilizatorilor

până în acest moment.”

„Facilitățile de care dispune această linie de digitizare permite ca milioanele de pagini ale documentelor istorice tipărite să devină accesibile ca informații *full-text* pe Internet în decursul a câtorva ani” [2, 7, 8].

- altă unitate a *DIGITIZING LINE* este instalată lângă *Strasbourg*, în *La Walck*, în cadrul uneia din secțiile de producție ale companiei *Infotéchnique*. Instalația, împreună cu alte trei, constituie un ansamblu menit să digitizeze mii de registre similare formatului A2 [9].

În parteneriat cu *IBM*, compania *Infotéchnique* este implicată în digitizarea cărții funciare (aplicația *AMALFI*). Aceasta asigură o logistică sigură pentru colecția de 4.000 de volume, în 46 de birouri, compusă din circa 2.500.000 de foi scrise de mână începând cu anul 1900, toată informația referindu-se la 1.800.000 de districte, 2.000.000 de proprietari și 4.500.000 terenuri și loturi în co-proprietate. Aceste cifre dau o idee asupra dimensiunii sarcinii de preluare a acestor date. Acest proiect mobilizează în jur de 150 de persoane. David Gray, director al companiei *Infotéchnique*, detalia: „Investiția noastră de aproximativ 4 milioane de Euro în *La Walck* prevede instalarea, la începutul verii anului 2004, a patru scanere de acest tip capabile să digitizeze și proceseze 125 de cărți pe zi. Cu un astfel de proiect, ne vor trebui patru ani pentru a computeriza 46 de birouri în *Alsacia Moselle*.”

În România, guvernul a început să acorde interes tot mai mare acțiunilor de digitizare, în special a materialelor vechi, pentru a oferi acces mai larg și, în același timp, pentru a prezerva documentele. Însă pașii sunt deocamdată mărunți și nu putem vorbi de nivelul atins de alte țări în acest domeniu.

Cele câteva exemple prezentate în cele ce urmează arată stadiul actual al digitizării în țara noastră:

- ANBPR (Asociația Națională a Bibliotecilor Publice), Comisia pentru digitizarea documentelor propune un proiect de digitizare, ca o metodă de păstrare și conservare a documentelor originale;
- CIMEC (Centrul Institutului de Memorie Culturală) lansează ideea digitizării de fotografii și crearea unei baze de date;
- Biblioteca Centrală Universitară din București propune soluții moderne de

conservare și acces pentru manuscrise și carte veche; au digitizat colecția de carte veche, la prețul de de 0,4 euro pentru o pagina A4 scanată.

- Biblioteca Universității *Transilvania* din Brașov realizează o bază de date cu carte veche digitizată pe baza unui proiect european de formare a specialiștilor în digitizare.

#### SCOP

Obiectivul cercetării a fost realizarea și furnizarea resurselor informaționale în formă electronică și dezvoltarea serviciilor *online* de difuzare a documentelor digitale în învățământul superior:

*Dezvoltarea și implementarea modelului educațional formativ și de cercetare – linie automatizată de generare bibliografii online* este un sistem modern de informare și depozitare a cunoștințelor, o modalitate de distribuție a informațiilor. Bibliografiile *online* și colecțiile de documente digitizate constituie resurse informaționale electronice produse de bibliotecă, utilizate în învățământul superior.

#### Metode

##### Organizarea liniei de digitizare

Digitizarea documentelor la Biblioteca Universității *Transilvania* din Brașov a urmărit cu precădere structurarea documentelor redactate pe hârtie, înregistrarea, „formatarea” prin capturare, indexarea, arhivarea, refacerea și distribuirea lor pe cale electronică. Ca rezultat al cercetării putem afirma că imaginile documentelor sunt replica exactă, digitizată a documentelor inițiale, originale, sunt superioare calitativ, se stochează ușor, se accesează în mod eficient, se pot copia și transmite fără a se pierde din calitatea informației vizuale [6, 10, 14].

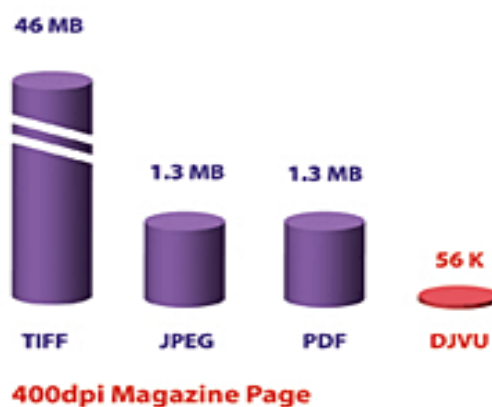


Figura 1.

Operația de digitizare este eficientă folosind tehnologiile actuale.

S-a pus în funcțiune scannerul și s-a instalat *soft-ul* livrat odată cu acesta. Au fost făcute teste de scanare pe diferite formate, la diferite rezoluții și s-au stabilit procedurile de prelucrare a documentelor. S-au determinat parametrii optimi de scanare pentru diferitele tipuri de documente și s-a testat compatibilitatea formatelor cu *soft-ul* de bibliotecă existent. S-au făcut evaluările și s-au stabilit măsuri de adaptare și îmbunătățire a dotării liniei ce urmează să se implementeze în faza a doua a proiectului (Figura 1) (Figura 2) [11, 20].

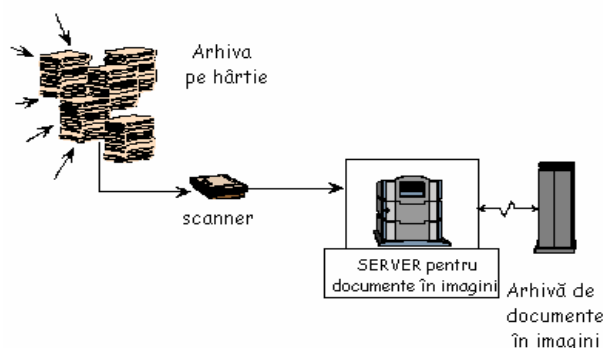


Figura 2. Schema generală a etapelor procesului de digitizare a documentelor

#### Rezultatele procesului de digitizare

##### Calitatea imaginii

Calitatea imaginii la capturare este definită ca rezultatul cumulativ al rezoluției de scanare și de lărgire a proceselor, al compresiei aplicate, al tehnicii sau dispozitivelor de scanare folosite și a abilității operatorului care realizează scanarea [12, 21, 28].

##### Formatele

La Biblioteca Universității *Transilvania* din

Braşov s-a făcut un studiu în ceea ce priveşte mărimea imaginii scanate a documentului în diferitele formate de digitizare.

Informaţia poate fi citită şi înţeleasă numai dacă programul recunoaşte formatul respectiv. Imaginile pot fi salvate în diverse formate. Formatele utilizate au fost:

- JPEG (.jpg) - *Joint Photographic Experts Group*;
- TIFF (.tif) - *Tagged Image File Format*;
- Joint Photographic Experts Group (.pdf) - *Portable Document Format*;
- Djvu

Ca o consecinţă a studiului efectuat putem afirma că în format *Djvu*, oricât de mare ar fi imaginea, aceasta este decodată, în memorie fiind

păstrată doar partea afişată pe ecran, ceea ce necesită mai puţină memorie decât un format *JPG/PNG Viewer* sau un format *PDF Reader*, fapt evident mai ales în cazul hărţilor.

Îmbunătăţirea imaginilor a constat în evidenţierea unor caracteristici ale imaginii (contururi, contrast etc.), pentru a o face mai elocventă pentru accesarea ei în bibliografia *online* pusă la dispoziţia utilizatorilor [15, 22, 29].

*Comparaţii grafice între formatul DjVu şi alte formate de stocare a informaţiei electronice*

Pentru a avea o viziune a compresiunii aplicaţiei dintre formatele *DjVu* şi *LDF* se foloseşte o componentă apropiată a formatului *MRC*:

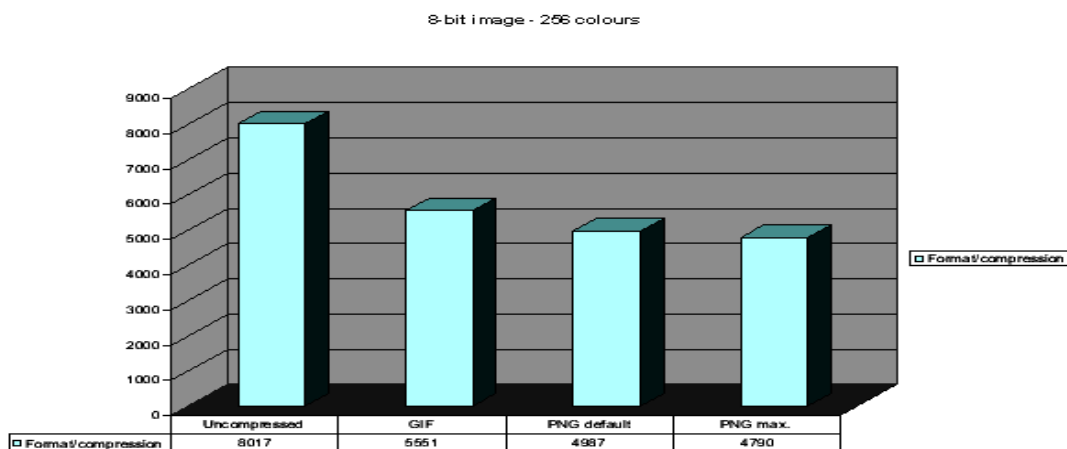
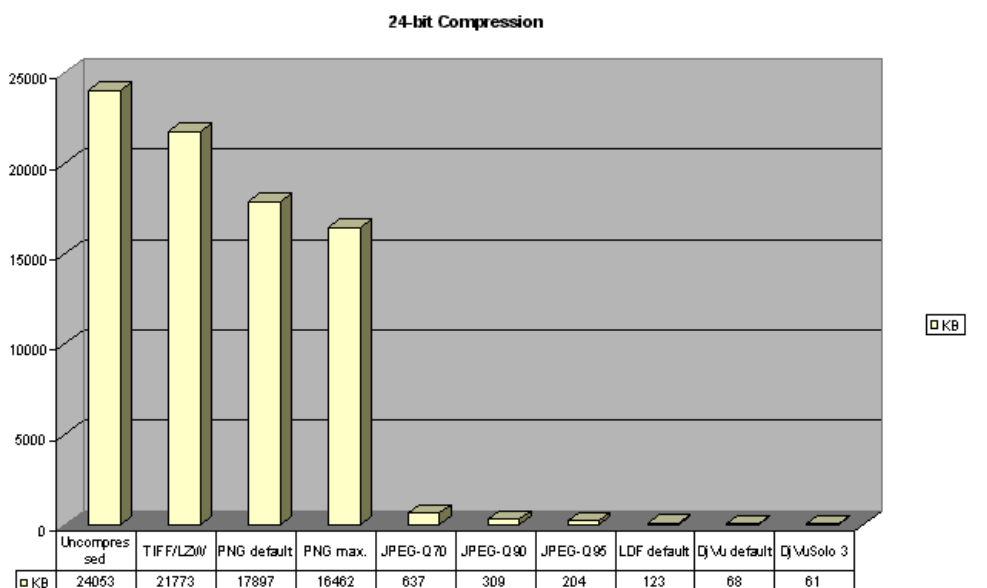


Figura 3. Adâncimea culorii din imaginea textului se reduce la 256 culori în afara celor 16 culori.

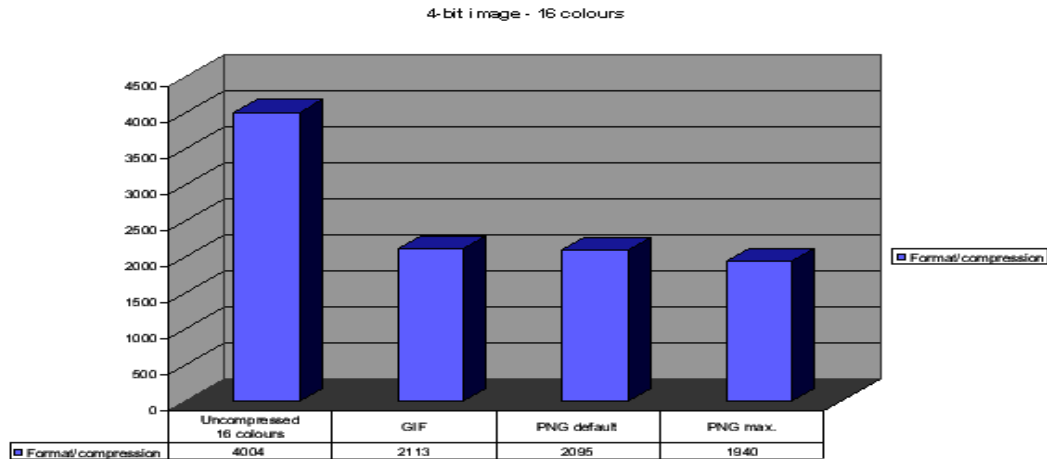


Figura 4. Imagini alb-negru pe 1 bit

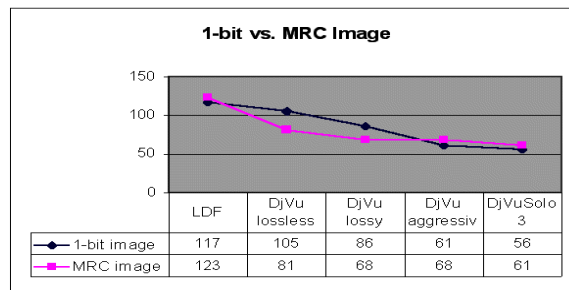
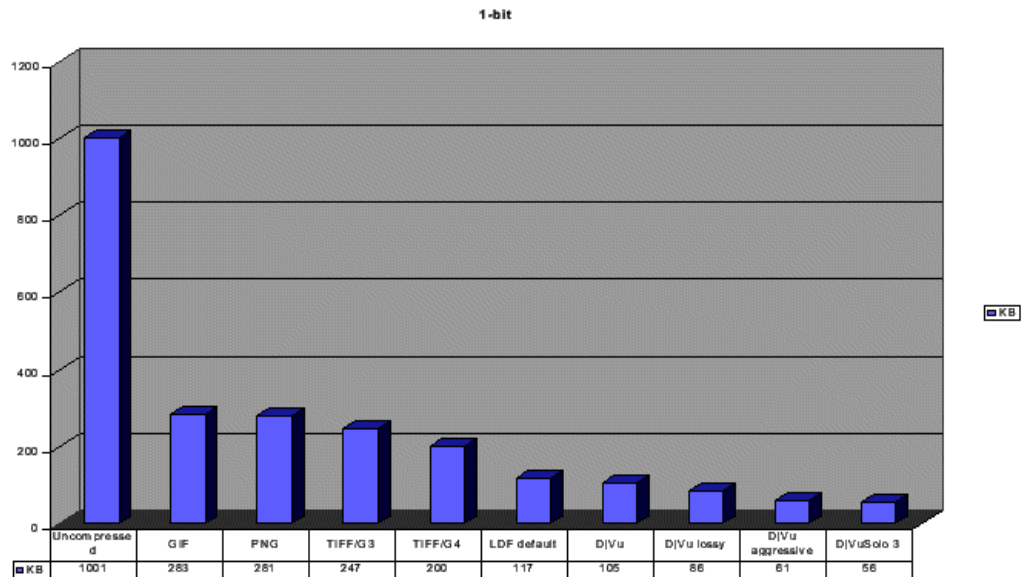


Figura 5. Comparația grafică între imaginea în format 1-bit și în format MRC.

**Nuanța gri a conținutul textului**

În acest caz, se observă că o parte a stratului textului se mișcă de la schema MRC în fundalul segmentului compresat. În formatul DjVu, 1 pagină de un 1-bit are 87 KB, în timp ce un strat de 1-bit în soluția MRC are numai 80 KB (Figura 6) [23, 29].

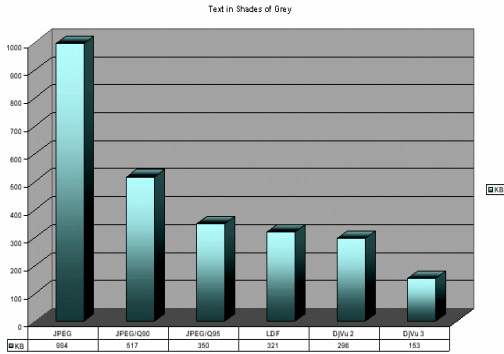
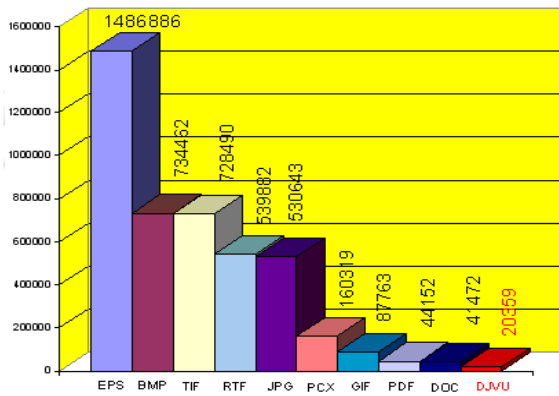


Figura 6.



Comparații între formatul DjVu și celelalte formate pentru stocarea informației (Figura 7)



Comparații între formatul DjVu și alte formate din punct de vedere al dimensiunii memoriei ocupate (Figura 8)

Dintre toate formatele de stocare a imaginilor, formatul DjVu ocupă spațiul cel mai mic, comparativ cu celelalte formate.

Fișierele listate în DjVu au următoarea formă (Figura 9):

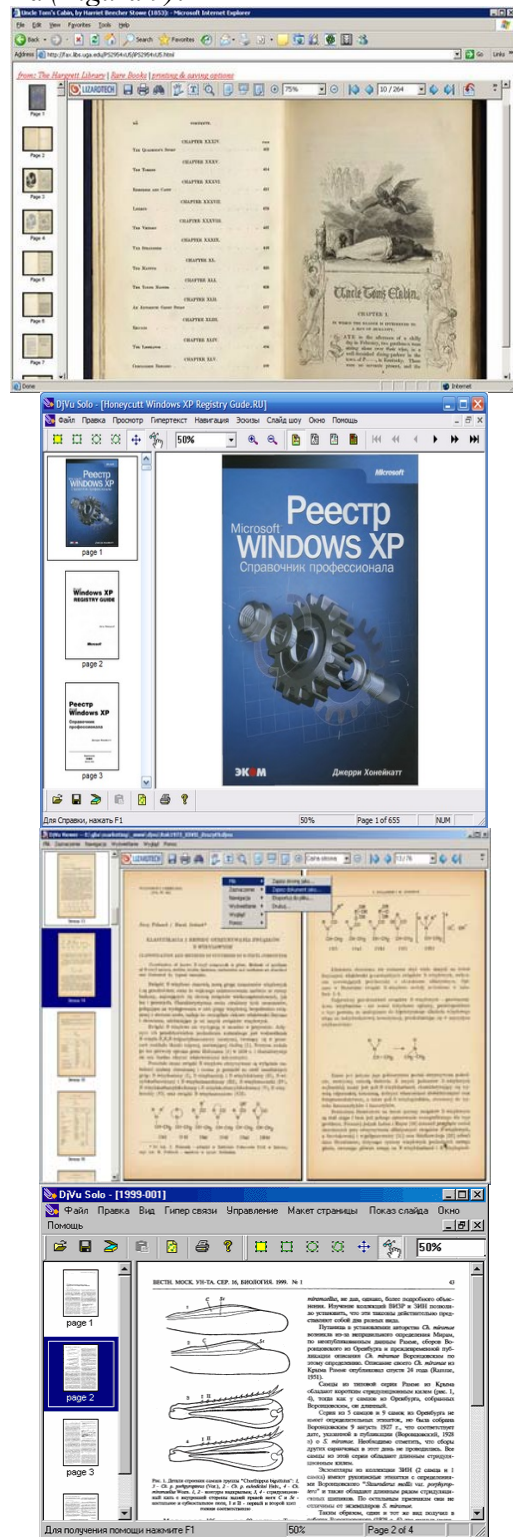


Figura 9.

**Concluzii**

Folosirea acestui procedeu este o modalitate deosebit de eficientă de a prezerva

informațiile, dar și de a crește accesibilitatea utilizatorilor la aceste informații.

Considerăm că această metodă este deosebit de importantă, nu numai pentru biblioteci, ci și pentru alte instituții culturale sau publice, care pot astfel să-și asigure dubla misiune: de păstrătoare ale unor documente (rare, unice) de valoare, dar și de sursă de informare atât pentru cercetătorii din domeniu, cât și pentru publicul interesat.

### BIBLIOGRAFIE

1. Aleger J. - Effective databases for text and document management. *Issues in Science and Technology Librarianship*. 2004; 56: 115-121.
2. Ayre C, Muir A. - The right to preserve: the rights issues of digital preservation. *D-Lib Magazine*. 2004; 34: 231-237.
3. Bakelli Y, Benrahmoun S. - Long-term preservation of electronic theses and dissertations in Algeria. *Libri*. 2003; 53(4): 531-540.
4. Calvert P. - Can a computer be sincere when it says "Have a nice day"? *Library Link Newsletter*. 2003; 46(3): 15-22.
5. Calvert P. - Collection for the information havens. *Library Link Newsletter*. 2003; 46(3): 27-34. Accesat în 26 septembrie, 2007, la <URL:http://www.emeraldinsight.com/librarylink/collection/>.
6. Cunningham SJ. - Report on the 6th International Conference on Asian Digital Libraries (ICADL 2003). December 8 – 11, 2003, Kuala Lumpur, Malaysia. *D-Lib Magazine*. 2004; 34: 267-275.
7. de Stefano P. - Moving image preservation in libraries. *Library Trends*. 2003; 52(1): 133-139.
8. Dean JF. - Digital imaging and conservation: model guidelines. *Library Trends*. 2003; 52(1): 146-154.
9. Digitization, conservation and quality control: XML as a preservation strategy. The long-term preservation of databases. *ERPANET Newsletter*, 2003, October 28.
10. Eason K. - Towards a user-centred approach to digital libraries. *Ariadne*. 2004; 38(30): 234-241.
11. Gurstein M. - Effective use: a community informatics strategy beyond the digital divide. *First Monday*. 2003; 8(12): 515-520.
12. Hain JE. - A brief look at recent development in the preservation and conservation of special collections. *Library Trends*. 2003; 52(1): 245-251.
13. i2S-DIGIBOOK. Accesat în 23 septembrie, 2007, la <URL:http://www.i2s-bookscanner.com/pdf/digibook\_mag\_no6\_en.pdf/>.
14. Joint N. - Staff development and training in the digital library environment. *Library Review*. 2004; 52(8/9): 417-421.
15. \*\*\* Accesat în 29 august, 2007, la <URL:http://www.emeraldinsight.com/librarylink/management/index.htm#article/>.
16. Krowne A. - Building a digital library the commons-based peer production way. *D-Lib Magazine*. 2003; 33: 167-172.
17. *Library Hi Tech Journal*. 2004. Accesat în 10 septembrie, 2007, la <URL:http://www.teriin.org/events/icdl/supporter.htm/>.
18. Lightle KS, Ridgway J. - Generation of XML records across multiple metadata standards. *D-Lib Magazine*. 2003; 33: 425-433.
19. Marieke G, Powell A, Day M. - Improving the quality of metadata in eprint archives. *Ariadne*. 2004; 38(30): 254-261.
20. Maxwell AT. - Mapping information policy frames: the politics of the digital millennium. *Journal of the American Society for Information Science and Technology*. 2004; 55(1): 28-34.
21. May C. - Digital rights management and the breakdown of social norms. *First Monday*. 2004; 8(11): 612-619.
22. Papatheodourou C, Kapidakis S, Sfakakis M, Vassilliou A. - Mining user communities in digital libraries. *Information Technology and Libraries*. 2003; 22(4): 102-108.
23. Rauber A. - Report on the 7th European Conference on Digital Libraries, ECDL 2003, August 17 – 22, 2003, Trondheim, Norway. *D-Lib Magazine*. 2003; 33: 245-249.
24. Reilly J, Bernard F. - Preserving America's printed resources: the roles of repositories, depositories, and collections of record. *Library Link News*. 2004. Accesat în 30 august 2007, la <URL:http://www.emeraldinsight.com/lm.htm/>.
25. Sauperl A. - Cataloger's common ground and shared knowledge. *Journal of the American Society for Information Science and Technology*. 2004; 55(1): 45-52..
26. Shechtman N, Chung M, Roschelle J. - Supporting Member Collaboration in the Math Tools Digital Library: A Formative User Study. *D-Lib Magazine*. 2004; 34: 324-332.
27. Silipigni Connaway L, Lawrence RS. - Comparing library resource allocations for the paper and the digital library: an exploratory study. *D-Lib Magazine*. 2003; 33: 451-458.
28. Titel V. - The digital book: a medial revolution without a new medium. *D-Lib Magazine*. 2003; 33: 514-520.
29. Williams M. - What the resource discovery network is doing for further education. *Ariadne*. 2004; 38(30): 272-277.