

Да помним или да забравяме? (Проблеми на съвременната дигитализация)

Юри Тодоров

Човечеството винаги се стреми да пази постиженията на предишни цивилизации и култури. Известно е, че това е правено с успех за дълги периоди и в продължение на хилядолетия, макар и да има печални изключения, например това с египетската библиотека в Александрия, изгоряла преди 2000 години. Днес се надяваме да запазим съдържанието на ценни документи чрез сканиране и дигитализиране на копия, които да се разпространяват многократно и навсякъде по твърди дискове на компютри в цял свят, без гаранция, че следващите поколения ще могат да ги използват. Оправдано е да си поставим въпроса, какви са рисковете за цифровото съхранение на документалното наследство?

В библиотеките милиони книги чакат своите читатели. Културното наследство, което е натрупано по този начин, трябва да се преобразува в цифров вид, поне частично. Създават се медиатеки под формата на цифрови файлове върху твърди дискове, особено актуални за запазването на исторически записи върху технически носители, например на аудиозаписи от магнитни ленти. Времето, което остава за изпълнението на тази задача, е съвсем кратко. Ако не сме в състояние през следващите 15 до 20 години да пренесем ценните записи от аналогови носители в цифров формат, то те ще бъдат безвъзвратно загубени. Историческото и културното наследство трябва да се дигитализира навреме и да се предостави за ползване на широк кръг потребители. Създаването на глобални енциклопедии в цифров вид е съвременната тенденция на развитие в модерните библиотеки. В близко бъдеще виртуалното ще се възприема за нормално, докато аналогово съхранената информация въобще няма да се използва.

Оправдано ли е, обаче, да поверим вековни цивилизационни ценности на краткотрайното съществуване на компютърните технологии?

Използваните сега електронни памети ще могат ли да съхранят надеждно за бъдещите поколения всички информационни ресурси, които създаваме днес? Информационната лавина води до създаване на все повече и повече битове и байтове. Производството на цифрови данни расте главоломно, затова трябва да се намери решение на проблема за дългосрочното запазване на информацията, която създаваме днес. Тук не става дума за съхранение за период от няколко десетилетия, а по-скоро за столетия и хилядолетия. Този проблем е изключително актуален за архивите и музеите, които пазят за вечни времена и управляват културното и историческото наследство. Така например, няма запазен оригинален запис от първите стъпки на човека на луната през юли 1969 г., останало е едно копие в лошо качество. Оригиналните записи от камерите на НАСА са били изтрети по невнимание преди 35 години ¹, останали са само копия от видео-емисии в лошо качество, които тогава се предаваха в реално време по телевизиите в света. Това е пример, който онагледява ниската надеждност на техническите носители и работата с тях.

Държавното управление не допуска подобни грешки. Известни са методите, които се прилагат за дългосрочно запазване на документи с особено значение и ценност. Принципите на този начин на конвенционално архивиране не са променени и до днес. Държавата непрекъснато произвежда документи на хартия, които след около 15 години се предават на архивите. Това са документи, които се отнасят до националната сигурност, проследяват действието на държавния апарат в течение на годините. Съхраняваното в архива документално наследство е държавно богатство. Проблемите на конвенционалното архивиране недвусмислено се изразяват в недостиг на място за съхранение на хартиените документи. Средно годишно в държавни исторически архиви постъпват хиляди линейни метри от документи. Преходът към електронно управ-

¹ Kaufman, M. The Saga of the Lost Space Tapes. The Washington Post, Jan 31, 2007

ление, който се извършва сега, трябва да подготви архивите за времето след хартията, за електронните документи. Тук е необходим нов начин на архивиране. Електронните документи не се предават физически в архива, а постъпват в неговата компютърна мрежа по защитена линия. Фондообразувателите могат да предават в архив по този начин произволни цифрови обекти, текстови файлове, електронни таблици, изображения, цифрови филми и звукови файлове. Дигиталния фонд може да поеме всичко това на едно място.

Проблемите в тази сфера могат да се онагледят с един пример. Счетоводните документи, изготвени като Ексел таблици, съдържат в клетките си формули. При разпечатване на таблицата виждаме само една част от нейното съдържание, т.е. по този начин не може да се постигне цялостно съхранение на подобни файлове. Необходими са програми, които да подпомогнат архивиста, при преобразуване на постъпващите документи в стандартен формат. По-стари документи, които са редактирани с остарели програми, подлежат също на подобно преобразуване. Текстове, които са изготвени с Уърд 2.0 на Майкрософт, вероятно няма да бъдат разчетени след 15 или 20 години. Затова е наложителна миграцията на електронни документи от остарели и специфични формати в универсален формат, който днес и през следващите години ще се използва като стандартен, напр. pdf. Така електронният документ ще е подготвен за дългия си път в далечното бъдеще.

Съхранението на подобни документи също е сериозен проблем. Необходими са цифрови хранилища, които гарантират висока надеждност и информационна защита. Прилагат се няколко редувантни цифрови масиви за съхранение, които позволяват дублирането на данните, както и произвеждането на електронни застрахователни фондове на определени периоди от време. Само така може да се гарантира необходимата висока степен на защита на електронните документи. Подобна технология би гарантирала електронно архивиране за период от 30 го-

дини. Преобладаващата част от файловете на електронния архив се съхраняват на магнитни ленти, които периодично се проверяват за настъпване на деградация. Данни, които се използват по-често, могат да се съхраняват на магнитни дискове. Това, за сега, е върхът в развитието на цифровата памет за дългосрочно съхранение. От 50-те години на миналия век, когато се изобретява магнитния диск, неговият капацитет непрекъснато расте като се удвоява на всеки 20 месеца. Днес вече е достигнат минималният размер на магнитните площи, при който те запазват своята термична стабилност. Това означава, че след известно време запазената там информация може да бъде загубена. Съществуват материали, които са в състояние да компенсират този недостатък, но проблемът при тях е прекалено високата стабилност, която не позволява презапис на данните. Изглежда, че можем да ползваме твърди дискове, които да бъдат презаписвани, или само такива, които са стабилни в дългосрочен период. Намерено е решение, което позволява съчетаването на двете свойства, т.е. използват се магнитни покрития, които са презаписваеми лесно и под първото покритие имат втори, стабилен слой. Записаната в първия слой информация, се отлага лавинообразно върху долния стабилен слой. Докато информацията в горното покритие е полезна за презаписване, то отложената в долния слой информация се запазва за дълго време непроменена. Така могат да се конструират твърди дискове с няколко магнитни покрития, които гарантират дълготрайно съхранение в комбинация с възможността за презапис върху минимална площ. Това решение задоволява потребността от масови паметни за период от около пет до десет години. След това неминуемо ще се наложи използването на третото измерение, така че дискът да може да съхранява информация в целия си обем, а не само на повърхността.

Несъмнено, дигитализацията на документи от технически аналогови носители (магнитни звукозаписи, филмови ленти) спасява ценни ресурси от безвъзвратното им унищожение. Но дигитализацията не

представлява универсално решение за дългосрочното архивиране, защото жизнения цикъл на цифровите формати е много по-кратък от този на аналоговите медии. Дигитализацията днес е само една първа стъпка в насока на създаването на архиви за дългосрочно съхранение. Файловете трябва да се движат, иначе загиват. Те трябва да се проверяват и презаписват непрекъснато, да са в движение, само така те ще останат живи. Така, както е необходима непрекъснатата грижа за традиционните архивни фондове чрез предаване на знанията от поколение на поколение в смисъла на съвременния архивен мениджмънт. Тук се поставя въпросът, какво да търсим и как се прави търсенето в наличния масив. До сега трансферът на знанията е действал безупречно, затова днес можем да намерим и разчетем документи, които са били създадени преди стотици години. Макар понятията да променят значението си през вековете, чрез разбирането за историческия контекст може да се вникне в съдържанието на понятията и на документите. Без допълнителни знания не е възможно до се разкрие съдържанието на старите текстове. Така, както сега се нуждаем от допълнителни сведения за условията на създаване на документите, в бъдеще ще са необходими подобни изследвания, които разкриват условията на създаване на цифровите файлове. Бъдещите компютърни археолози и архивисти ще трябва да разкриват софтуера на 20-ти век. Без допълнителни сведения историците на бъдещето биха открили само един неразбираем куп от нули и единици. Ние оставяме един неразгадаем лабиринт от формати на файлове, който в бъдеще няма да бъдат разгадани без допълнителни сведения относно начина на създаването им. Всъщност, данните сами по себе си нямат никаква стойност, цифровият файл винаги се разчита само с подходящата програма. Затова е необходимо допълнително помощно средство за разбиране на съдържанието на цифровите файлове. Софтуерът, който ползваме днес, не създава такива помощни средства. Дори ако сме в състояние да предадем дигитализираната информация в пълната и цялост на

следващите поколения, нямаме никаква сигурност, че те ще могат да я разчетат. Необходимо е описание на формата на данните, на тяхната структура, както и описание на софтуера, с който да се разчете съдържанието. Това би бил ключът за цифровото наследство, което ние трябва да оставим на поколенията.

Дигитализацията в библиотеките е ориентирана към разширяване на възможностите за достъп. Гугъл работи върху създаването на цифрови копия на повече от 400 000 книги от периода от 16-ти до 19-ти век, съхранявани в Австрийската библиотека в рамките на европейския проект Europeana ². В резултат на това съдържанието им става достъпно в интернет. Дигитализацията се отразява чувствително върху начина на ползване на библиотеките от техните читатели. Те могат да се възползват от огромното предимство на цифрово съхранената информация, а именно възможността за многократно копиране и презаписване и предоставяне на произволно място за кратко време. Сега се дигитализират особено ценни книги, които така стават достъпни от широк кръг читатели. Наред с единичните много ценни екземпляри, интересни за дигитализация са и периодични публикации, напр. годишни течения от вестници и списания. Към подобни издания интересът расте непрекъснато. Тук изборителното дигитализиране на малки обеми не е интересно. Потребителите искат да търсят в огромни хранилища от данни и да работят с много разполагаеми екземпляри. Непрекъснато расте броят на читателите, които ползват онлайн ресурси от дигитални читални. Музеите също са ориентирани към новите технологии. Ценните експонати трябва да станат достъпни и извън хранилищата и изложбените зали. Изготвят се цифрови изображения за експонати, които са предвидени за различна употреба, за плакати или за научна употреба. Изображенията се комплектуват в съответната база данни.

² <http://www.google.com/hostednews/afp/article/ALeqM5gQRnbbvo2WkH8DzHcf04Fs6y4usg>

Възможно ли е да се справим с огромната лавина от информация, която постъпва от музеи, библиотеки и архиви. Броят на лентовите и дисковите памети, които формират ресурса на един информационен център, никога няма да е достатъчен за постъпващата отвън информация. Решението на този проблем е децентрализацията, т.е. информацията не бива да се натрупва на едно място, а да се разпространява на много места. Колкото повече копия се създават, толкова по-висок е шансът да се възстанови оригинала. Възможно ли е обаче да се пази всичко? Изследванията показват ³, че скоро ще бъде достигнат момент, когато производството на данни ще нараства много по-бързо отколкото производството на памети за съхранението им. Когато се случи това, ще стане ясно, че запазването на информационните ресурси може да става само изборително. Подходът за съхраняване на всичко на всяка цена е обречен на неуспех. Затова е необходим механизъм за оценка на информацията, за преценка, какво може да се изхвърли и какво да се запази. Механизмите за решение на проблема са заложи в природата, подобно на начините на действие на дългосрочната и краткосрочната памет. Така в дигитални архиви загубата ще бъде също толкова важна както съхранението. Само ако забравяме правилно, ще съумеем да запазим тези неща, които ще бъдат полезни в бъдеще.

Библиография:

Brantl, M. et.al. Massendigitalisierung deutscher Drucke des 16. Jahrhunderts. Ein Erfahrungsbericht der Bayerischen Staatsbibliothek. Zeitschrift fuer Bibliothekswesen und Bibliographie 56(2009), 327-338

Conway, P. Preservation in the digital world.

<http://www.clir.org/pubs/reports/conway2/>

Dean, J. F. Digital imaging and conservation: Model guidelines. Library Trends, 52 (2003), pp. 133-137.

Edmondson, R. Audiovisual archiving: Philosophy and principles (document code CI/2004/WS/2). Paris: United Nations Educational, Scientific and Cultural Organization.

<http://unesdoc.unesco.org/images/0013/001364/136477e.pdf>

³ Вж. напр. Lesk, M. How Much Information Is There In the World?
<http://www.lesk.com/mlesk/ksg97/ksg.html>

Foltz, P.W. Models of Human Memory and Computer Information Retrieval: Similar Approaches to Similar Problems. Univ. of Colorado, Boulder CO 80309

Jenkinson, H. Reflections of an archivist. In M. F. Daniels & T. Walch (Eds.), A modern archives reader: Basic readings on archival theory and practice (reprint, pp. 15-21). Washington, 1984

Резюме: Днес се надяваме да запазим съдържанието на ценни документи чрез сканиране и дигитализиране на копия, които да се разпространяват многократно и навсякъде по твърди дискове на компютри в цял свят, без гаранция, че следващите поколения ще могат да ги използват. Оправдано е да си поставим въпроса, какви са рисковете за цифровото съхранение на документалното наследство? Подходът за съхраняване на всичко на всяка цена е обречен на неуспех. Затова е необходим механизъм за оценка на информацията, за преценка, какво може да се изхвърли и какво да се запази. Механизмите за решение на проблема са заложи в природата, подобно на начините на действие на дългосрочната и краткосрочната памет. Така в дигитални архиви загубата ще бъде също толкова важна както съхранението. Само ако забравяме правилно, ще съумеем да запазим тези неща, които ще бъдат полезни в бъдеще.