

ЦИФРОВАТА ПАМЕТ. ПРОБЛЕМИ НА ДИГИТАЛИЗАЦИЯТА ЧАСТ ПЪРВА

Юри Тодоров

Софийски университет „Св. Климент Охридски“

Резюме. Статията разглежда проблема за възможностите за съхранение и трайно запазване на интелектуалния ресурс на нашето съвремие чрез дигитализация, мобилност и миграция и обръща внимание на нов ресурс – информацията, която се пази в тайна. Централизираната информация в огромни хранилища на историческата памет никога няма да може да догони информационната лавина по отношение на наличния капацитет. Модерният архивист е мениджър на корпоративното поведение за архивните информационни ресурси. Налага се преходът към т.нар. виртуален архив, който да си сътрудничи с останалите глобално достъпни информационни ресурси. Много хора възприемат интернет като автоматичен израз на обществен и технологичен прогрес. Впоследствие става ясно, че информацията има и тъмна страна, а това е тайната. Междувременно „Уикилийкс“ се превръща в събирателно пространство за повече от един милион документи, успява да ни даде представа за това как би изглеждал светът без граници.

Keywords: information resources, digitalization, mobility, centralized information

Научната общност в съвременното информационно общество се занимава все повече с проблема как да се съхрани огромната информационна лавина по сигурен и надежден начин за бъдещите поколения. Динамичното развитие в областта на информационните технологии невинаги е съпроводено с благоприятни условия за реализирането на тази задача. Надеждите на архивистите и библиотекарите, до известна степен и на музейните работници, са ориентирани към спасяването на ценните фондове със средствата на дигитализацията. Дигитализираните обекти могат да се копират, обработват и изпращат на големи разстояния. Една част от историческото наследство на човечеството може да бъде обработено по този начин и да се увековечи в цифрово съответствие. Проблемите, които възникват тук са предизвикани от непрекъснатата промяна на файловете формати и технологиите за съхранение на информацията. Така дългосрочното архивиране се превръща в задача, която никога няма да приключи. Данните трябва да се движат, да се преобразуват, ако те останат в покой, ще бъдат загубени. Усилията в тази

област затова са насочени към разработване на методи, които да направят цифровото наследство на нашето поколение достъпно за изследователите и след столетия. Цифровите обекти, снимки, текстове, филми и програми би трябвало да се съпровождат от своеобразна инструкция за употреба – това ще даде възможност след време да се разкрие съдържанието на обектите дори когато технологиите и средствата за съхранение са напълно различни. Веднага възниква въпросът къде да се съхранява този ценен информационен масив. Да се създаде информационен център, който да изпълнява всички условия от гледна точка на сигурността и надеждността? От самото начало е ясно, че това е един погрешен подход. Централизираната информация в подобни средища на историческата памет никога няма да може да догони информационната лавина по отношение на наличния капацитет. Затова перспективното решение би трябвало да залага на многократното копиране на данните и пълната децентрализация на информационните масиви. Колкото повече копия имаме, толкова по-голям е шансът след време те да бъдат реконструирани. Човешкият мозък все още е най-добре функциониращият орган за съхранение на информация въобще. Информацията в мозъка не се натрупва на едно място, а се разпределя из целия орган. Това очевидно е добра стратегия. Това, което се използва непрекъснато, стои на повърхността, докато отдавна случилите се събития са отложени в дългосрочната памет. Преценката за ценността на информацията прави възможно отсяването на маловажното от същественото, което да се съхранява завинаги.

Модерният архивист

Хилари Дженкинсон (Jenkinson, 1922) и Теодор Шеленберг (Schellenberg, 1956) са известни мислители в областта на архивната теория и практика на XX век. И двамата дават отговор на кризата в областта на модерната архивистика с редица влиятелни публикации. Те често са поставяни на противоположни позиции с оглед на теоретичните им постановки и идеи. Дженкинсон нарича идеите за селекция на документите опасни, докато Шеленберг определя Дженкинсон като стара вкаменелост.

Дженкинсон вижда себе си в ролята на пасивен пазител, който иска да съхрани всичко. Шеленберг се определя като прагматичен интервенционист, баща на процедурата за определяне на ценността на документа. Задачата, която Дженкинсон си поставя, е да изследва естеството и характеристиките на архивните документи, за да достигне до фундаменталното разбиране за архивните принципи, които от своя страна да доведат до създаване на архивите на съвременното и на бъдещето. Дженкинсон вижда архива като съставен от взаимно свързани документи и именно контекстуалната цялост му дава смисъл и е ценност, достойна за съхранение.

Теорията на Шеленберг възниква от необходимостта да намали огромния обем от документи чрез селекция сред масата от текстове и да предостави този интелигентен избор на изследователите. Шеленберг поставя процеса на селекция като централен за ролята на архивиста. Според него, докато документите се създават, за да задоволяват нуждите на техните създатели, то причината при избора им за трайно съхранение е

съвсем различна. Документите имат две стойности: първична стойност, като свидетелство за автора, и вторична стойност, която има отношение към историческите и културните му функции за останалите (но не и за самия създател). Вторичната стойност от своя страна има две свойства: качеството на свидетелство за начина, по който институциите се организират и функционират и информационна стойност, която се отнася до определен предмет, дейност, проблеми и условия.

Във връзка с дебата за това как най-добре да управляваме и съхраняваме електронните документи имаме очевидни паралели към проблемите, свързани със селекцията на документи и определянето на ценността. От едната страна са неодаженинсонистите, които настояват за това да не се прави разлика между традиционни и електронни документи. Тази постановка връща селекцията при източника на създаване, съобразно правните и административните му потребности. Тя трябва да се извършва от фондообразувателя с помощта и съвета на архивиста.

От другата страна са постшеленбергианците, които определят електронните документи за коренно различни. Тази разлика изисква пълно преосмисляне на архивната теория и практика даже по отношение на пазителския контрол над документа. Електронните документи подчертават обстоятелството, че документът съществува във фазите създаване, използване и съхранение. Тази интерпретация измества фокуса за отговорността на архивиста от съхранението на документи за исторически цели в полза на ролята на архивиста като отговорник за самия документ. Постановката определя и правно-административната роля, която архивите имат в съвременното общество. Архивистите трябва да се откажат от пазителската си роля не само заради цената на съхранението на електронните документи с всичките хардуерни и софтуерни проблеми и непрекъснатата необходимост от миграция на документите.

Модерният архивист е мениджър на корпоративното поведение за архивните информационни ресурси. Той регулира съхранението на информация като надзорник, управляващ собствеността на институцията. Необходими са входни точки или виртуални архиви без стени, които позволяват на потребителите да получават достъп до документите по всяко време и отвсякъде. Документите се поддържат и управляват от техните създатели, които осигуряват и необходимите електронни устройства. Ролята на архивиста е да работи заедно с авторите на документите, като определя тези документи, които са от значение за институционалната или корпоративната памет. Гледището на Дженкинсон относно ценността на документа не може да бъде валидно за модерните условия на живот и не отговаря на очакванията на обществото за ролята на модерния архив.

Стратегии за дигитализация

Днес навсякъде се прави опит части от историческото наследство да се дигитализират и да се съхранят за следващите поколения. Особено критични в това отношение са записите, които са правени с аналогови технологии (на звук, изображения, филми). Ако днес не успеем да преобразуваме тези записи в цифров

вид, след около петнадесет години те няма да могат да бъдат възпроизвеждани и ще останат безвъзвратно загубени за човечеството.

Съвременните системи за работа с информация, т.нар. системи за управление на знания и съдържание (knowledge, content management), бележат развитие в насока на изоставяне от първоначалната документална ориентираност към интелигентните редакторски системи, които гарантират мултимедийност и поддържане на информационни потоци в смисъла на съвременния workflow management.

Характерната особеност на съвременните системи за управление на съдържанието е разделянето между структурата, начина на представяне и съдържанието на сведенията, които влизат в състава на системата. Структурата се отнася до елементите на текстовия документ (дата, адресат, адресант, описание, текст и т.н.). Към начина на представяне принадлежат стиловите схеми и форматирането, а съдържанието е самият текст и аудиовизуалните модули, които са включени в документа. Основен момент в архитектурата на документа, в смисъла на управлението на съдържанието, се състои в боравенето с трите му съставни части – съдържание, структура и начин на представяне.

Какво е отражението на тази концепция в рамките на информатиката върху класическите институции, които боравят с информация, т.е. архиви, музеи и библиотеки? Да разгледаме например т.нар. специални архиви, които отдавна се занимават с документиране и архивизиране на информация от съвременни технически медии. Днес те се преориентират. Сложната обработка на публикациите се замества от мултимедийни бази данни, интелектуалната работа по разкриване на съдържанието се прави от средствата за търсене в свободен текст, традиционните потребители на аудиовизуални архиви вече не са зависими от работата на архива, глобално наличната информация в Интернет много често превъзхожда възможностите на собствения, ведомствен конвенционален архив. Дейността на аудиовизуалните архиви, в смисъла на информационния мениджмънт, се състои в планиране, контрол и управление, както на вътрешните, така и на външните информационни ресурси. Новият аспект на външните ресурси е типичен при работата с мултимедийна информация. Потребителите на историческите архиви допълват сведенията си от Интернет и предлаганите там информационни ресурси. Затова те очакват разнопосочност на предлаганата информация както по отношение на съдържанието, така и по формата и вида на представяне. В този смисъл се налага преходът към т.нар. „виртуален архив“, който да си сътрудничи с останалите глобално достъпни информационни ресурси. Осъществяването на тази идея е особено полезна за университетското образование и може да си сътрудничи с концепцията за „виртуалната библиотека“. В това отношение особено подходящи са музейните и литературните архиви, както и ведомствените университетски архиви.

Тъй като днес информацията е „горивото“ на иновацията, класическите производители на информация (библиотеки и архиви) придобиват нова роля и функции, извършвайки прехода към електронни и мултимедийни информационни продукти и услуги. Този процес се демонстрира от края на 90-те години на миналия век, когато класическите библиографски университетски специалности започват да се преиме-

нуват и да си създават нов етикет като информационен мениджмънт, информационни фондове, технологии и т.н. Именуването дава израз на едно далеч по-комплексно измерение по отношение на класическите методи и постулати на библиографията. Това е разкриване на съдържанието и търсене на информация независимо от институционалните компетентности чрез разпределяне на информационния ресурс в мрежи. Методите на нормиране и контролиране от централно ведомство и централизация на информационните ресурси, които бяха интересни в средата на 70-те години от миналия век, днес се заместват от процесите на децентрализация. В съвременния информационно доминиран свят възникват навсякъде информационни продукти и ресурси, паралелно и неконтролирано. По такъв начин постулатът за целостта на данните се оказва една илюзия. Ще продължават ли библиотеки и архиви да се отказват от функциите си, за да останат само пазители на физическите документи, докато други информационни институции започват да предлагат електронно съдържание? Ето и някои предложения за преодоляване на този проблем:

- архиви и библиотеки трябва да предлагат усилено информационни продукти в глобалната мрежа;

- софтуерът за управление на съдържанието в архиви и библиотеки задължително трябва да има веб-интерфейс, който да се използва най-малко на ниво каталог;

- професията на архивиста трябва да стане мултимедийно ориентирана;

- традиционните компетентности в областта на информационното търсене трябва да намерят ново развитие в нехартиените цифрови масиви на разпределените децентрализирани информационни системи;

- обучението и квалификацията трябва да се ориентират към практически компетентности в областта на информационните технологии.

Сигурно има критици на очертаващите се днес тенденции, които биха твърдели, че до средата на столетието читалнята няма да има друга алтернатива. Не бива да забравяме, че днес вече са актуални тенденциите на децентрализация в библиотеките, които се изразяват в изоставяне на централните каталози за сметка на съчетаването на няколко разпределени бази данни – т.нар. OPAC (Online Public Access Catalog). Така се създава един виртуален общ каталог. Отказът от създаване на централизирани ресурси и ориентацията към Интернет е характерна за съвременните виртуални библиотеки. Основни белези на тази стратегия са: Интернет вместо централен компютър, разпределение на информационния ресурс вместо централна база данни и виртуален каталог (търсеща машина) вместо централен каталог. Концепцията се допълва от поддържане на отворени стандарти, техническа независимост на локалните системи от мрежовия ресурс, хетерогенност на интегрираните системи. По такъв начин се постига отворена организация и самостоятелност, независимост на участниците, която е характерна за съвременната интернет-технология. Все още сме далеч от типичното за тази среда мултимедийно съчетаване на информационните обекти, но все по-често информационните услуги започват да се

допълват от предлагане на цифрови копия на реални документи в онлайн режим – услуга, която е типична за съвременните виртуални библиотеки. Архивите в САЩ и Европа залагат на известния от 1996 г. стандарт за дефиниране на документния тип с помощта на метаданни EAD (Encoded Archival Description).

Перспективният информационен мениджмънт в бъдеще няма да се различава съществено в приложенията си в архиви, библиотеки и музеи – нито в техническо, нито в организационно отношение. Разпределените информационни ресурси ще бележат тенденция към интеграция и унификация по пътя на самостоятелността в полза на обогатяването на знанията на човечеството.

Проблеми на автентичността на електронното съдържание

Автентичността, в смисъла на електронното създаване и поддържане на файлове, няма нищо общо с понятието оригинал, действителност или истина. Тук автентичността трябва да се разбира в смисъла на надеждността и гарантираността на възпроизвеждане на документа. Затова автентичен електронен документ е този, при който се доверяваме на начина на представяне пред неговия читател и когато това представяне отговаря на намеренията на неговия автор или изпълнител. Тук автентичността първоначално е въпрос на текстовата надеждност при описанието на минали събития и не е свързана с оригиналността или истинността им. Ако приемем, че автентичността е въпрос на надеждно възпроизвеждане, тогава възниква още един проблем – за степента на тази надеждност. Защото тази категория не може да се определи в абсолютен смисъл, тя не може да се измери в абсолютни стойности. Трябва да приемем, че понятието автентичност е нещо относително. Автентичността на един документ може да се определи само когато е известна връзката му със заобикалящите го обстоятелства: кога е създаден той и къде е съхраняван от времето на създаването му до днес. От гледна точка на контекста автентичността на документа може да се определи на базата на неговото съдържание. При тези условия за определяне на автентичността на електронния документ не е от решаващо значение каква е била първичната форма на създаване на документа. Съществено е запазването на неговото първоначално съдържание при пренасянето на текста от един носител на информация на друг, т.е. автентичността не е свързана с изискването за оригиналност във физически смисъл.

По-нататък при определяне на автентичността може да се постави изискването съдържанието да има по-важно значение от формата. В такъв случай би трябвало да се отдели повече внимание на самото послание, отколкото на начина, по който е направено то. Но, ако се придържаме към принципа, че при оценката на автентичността формата и съдържанието са равнопоставени, тогава трябва да приравним автентичността с понятието оригиналност. Но все пак би трябвало да има разлика между автентичност и оригиналност, доколкото оригиналната форма е нещо абсолютно, а автентичността е относителна.

След казаното дотук възниква следващият проблем: възможно ли е да се направи отделяне на формата от съдържанието, без да се накърнява автентичността

и как може да бъде определена тя. Съществуват обстоятелства, при които формата не е от решаващо значение за съдържанието. Когато документът се възстановява с технически средства на хартия или в електронен вид, възпроизвежда се в нова материална среда, той продължава да бъде автентичен. Целта на неговото възпроизвеждане е запазване на съдържанието и при наличие на четим текст съответният документ може да се разглежда като автентичен. Автентичността е функция на обвързаността на документа с обстоятелството дали формата е отделима от съдържанието. Тази зависимост е по-силно изразена и от по-голямо значение, когато формата не е обвързана със съдържанието. Надеждното определяне на автентичността тогава до голяма степен зависи от контекста и обстоятелствата.

Да предположим, че даден документ е изготвен с текстообработваща програма и е изпратен с електронна поща като прикачен файл. Когато получателят чете писмото с неговата текстообработваща програма, съществува известна опасност той да вижда писмото по начин, който е различен от вида на документа при неговото изготвяне на мястото на подателя. Основната причина за това е, че не всички текстообработващи програми могат да интерпретират правилно форматите на други програми. Когато писмото се отлага под формата на файл при неговия подател, проблеми с автентичността няма да има, тъй като тук съществува електронно копие на изпратения документ в условията на неговото създаване. По-нататък обаче този документ всъщност зависи от данни, които не съществуват на компютъра на получателя на документа. Това може да бъде специфичен шрифт или връзка с обект, които не присъстват при получателя. От друга страна, когато съхраненият в архива на подателя документ е престоял там няколко години или дори само месеци, налице е опасността, че неговата текстообработваща програма ще се е променила, с което нараства вероятността за нарушаване на първоначалното състояние на документа. В действителност няма гаранция, че документът в стария формат изобщо ще може да бъде прочетен от новата програма. Ако новата програма има възможност да разчита по-стари документи, възникват други проблеми с автентичността. За да може да работи със своите документи, авторът разрешава запазването на писмото в новия формат на програмата, с което замества оригиналния формат с нов. Когато документът се предава в друг архив, например обществен или институционален архив, също възникват проблеми с автентичността, ако там документите се получават само във формати, които основно се различават от тези на авторите.

Проблемите с автентичността и отделянето на формата от съдържанието са сходни както при хартиените, така и при електронните документи. При последните обаче затрудненията са повече поради зависимостта им от непрекъснато променящите се средства за работа с тях. Оттук възниква въпросът дали тази зависимост може да бъде сведена до степеня на зависимост от средата на възпроизвеждане на конвенционалните документи, а, ако не, то по какъв начин тя да бъде сведена до минимум. Трите основни подхода могат да бъдат описани, както следва: според музейната стратегия трябва да се откажем от отделяне на форма

и съдържание и да запазваме старите устройства, носители и програми, с които документите са били създадени. При емуляционната стратегия се извършва частично отделяне на формата от съдържанието. Документите се прехвърлят на нови носители и се изготвят програми, които работят на нови устройства, за да имитират работата на старите програми, с които са били направени документите. В миграционната стратегия се извършва отделяне на формата от съдържанието, като се преминава на нови носители, нови устройства и нови формати на данните.

Ако отделим внимание единствено на автентичността, то най-подходящи са музейната и емуляционната стратегия, но те биха довели до бърза загуба на електронни документи, защото нито една от тези стратегии не е практически приложима, те са технически и икономически неосъществими. Да се надяваме, че един ден ще имаме такова ниво на стандартизация в областта на информационните технологии, че да може да се реализира миграционната стратегия. Казано по друг начин, за да запазим документите, ще трябва да ги променяме. В същото време трябва да намалим честотата на промените до необходимия минимум, като прилагаме стандарти, и то предимно широко приети официални стандарти или такива, които са запазена търговска марка. Трябва да се предвидят мерки за преодоляване на проблемите или поне за тяхното намаляване, защото, след като са възникнали, отстраняването им би било много скъпо или дори неосъществимо. Трябва да сме подготвени за ситуации, при които отделянето на формата от съдържанието е трудно осъществимо или разликите между формата, която носи съдържанието и формата, в която трябва да се преобразува съдържанието, са размити.

Би трябвало да се избягват функции, които са динамични или които могат да се развиват по различни насоки или да се разгръщат на няколко нива. По-лесно е да се архивира звукозапис или видеофилм, отколкото електронна таблица или текстов документ, който съдържа скрити бележки. Звуко- и видеозаписите са непрекъснати и се възпроизвеждат последователно, докато електронните таблици включват формули, числа, диаграми, а текстовите документи се показват с или без бележки или включват връзки към части от същия или към други документи. След създаването си един електронен документ трябва да се архивира възможно най-бързо, защото закъснението повишава риска от нарушаване на автентичността. Непосредственото архивиране гарантира целостта на документа, съдържанието му се влага и интегрира, така че да не се изгражда от свързани към него части, които да са извън обсега на архива. Непосредственото архивиране намалява риска от разместване на съдържанието, премахване на части от него или промяната им на по-късен етап. Трябва да се търси добър компромис в избора на форма на представяне на документа, която от една страна да улеснява неговата повторна употреба и същевременно да запазва автентичността му. Когато не е възможно да се изпълнят двете изисквания, би трябвало да се отдаде предимство на автентичността. Ако целта на съхранението е използваемостта на документа на по-късен етап, тогава той трябва да се съхранява в редактируем формат. В същото време трябва да е налице възможност за сравнение заедно с автентичния начин на представяне на документа. Това е възможно, когато документът се запазва в

архив в оригиналния формат на текстообработващата програма заедно с придружаващото го графично представяне (например като TIFF графика).

Един многообещаващ подход е едно и също съдържание да бъде представено в няколко форми – на хартия, HTML, DjVu или PDF. Това силно отделяне на формата от съдържанието всъщност означава, че се дава по-голямо значение на съдържанието, отколкото на начина на представяне. В същото време е възможно определени представяния да бъдат зададени като единствено отговарящи на истината, т.е. автентични, в случай, че автентичността и външният вид са от особено значение.

Необходимостта от отделяне на формата от съдържанието при електронните документи и архиви налага да се обърне особено внимание на проблемите около тях. Представянето на едно и също съдържание в няколко различни форми подсказва, че съществуват неясноти около доказателствената сила на електронните документи и тяхната автентичност, а въпросът за автентичността на електронни документи и архиви изисква обширен дебат и интердисциплинно взаимодействие.

ЛИТЕРАТУРА

Jenkinson, H. (1922). A manual of archive administration including the problems of war archives and archive making. The Clarendon Press, Oxford. (http://openlibrary.org/books/OL7224051M/A_manual_of_archive_administration_including_the_problems_of_war_archives_and_archive_making).

Schellenberg, T. R. (1956). The Appraisal of modern public records. Washington: U.S. Gov. Print. Off. (<http://www.worldcat.org/title/modern-archives-principles-and-techniques/oclc/7126776>)

DIGITAL MEMORY. PROBLEMS OF DIGITALIZATION

Abstract. The paper deals with the problem of storage options and the capabilities of permanently preservation of the intellectual resources of our times through digitalization, mobility and migration and draws attention to a new resource – information that is kept in secret. Centralized information in large repositories of historical memory will never be able to catch the avalanche of information regarding the available capacity. The modern archivist is a manager of the corporate conduct for the archival information resources. A transition is needed to the so called “virtual archive“ to cooperate with other global information resources available. Many people perceive the Internet as an automatic expression of social and technological progress. Now it becomes clear that information has a dark side and this is the secret.

Assoc. Prof. Yuri Todorov, PhD

✉ Sofia University „St. Kliment Ohridski“
15, Tsar Osvoboditel Blvd.
1504, Sofia, Bulgaria
E-mail: todorov@clio.uni-sofia.bg