

## За някои проблеми на дългосрочното архивиране на информация

*Юри Тодоров*

Нарастването на капацитетните възможности на модерните компютърни системи и мрежи позволява на институции и организации да създават, предоставят за ползване и да съхраняват огромни обеми от цифрова информация. Взаимоотношенията между институциите днес се осъществяват все повече в цифров вид и хартиената кореспонденция се допълва и до голяма степен замества с електронна. Съхранението на дигитална информация е свързано с редица проблеми, които са различни от тези на традиционното архивиране на хартиени и специални документи. Цифровите технологии се използват повсеместно. Този факт е свързан с доста предимства, но носи в същото време и недостатъци заради бързото усъвършенстване на технологиите и непрекъснатите промени във стандартите и форматите за съхранение на информацията. Затова архивната работа в тази област винаги е свързана с осъзнаване на риска за загуба от възможността за реконструиране на информацията, на нейното възпроизвеждане според възможностите на съвременните технологии, така както се изисква от Закона за НАФ, съответно и от Закона за електронното управление в България.

Далеч съм от мисълта, да разглеждам единствено техническите аспекти на този проблем. Задължително трябва да бъдат отчетени всички перспективи в научен, правен, икономически аспект, а и от гледната точка на мисията на институциите на паметта за съхраняване на културното и историческо наследство.

Централен въпрос в дискусиата на поставената проблематика е разбирането за понятието информация и информационен ресурс. Информацията това са знания, които могат да се обменят, и при тази комуникация информацията винаги се представя чрез конкретни данни. За възприемане

на информацията са необходими основни познания, чрез които става възможно да се вникне във смисъла на предаваните сведения. Ако получателят не притежава основните знания като предпоставка за осмисляне на информацията, тя трябва да се представи по начин, който да отговаря на неговите възможности и изисквания. Условието за разчитане и разбиране на информацията подлежат на промени във времето, затова данните своевременно трябва да се представят в разбираеми за получателите информационни обекти.

Описаните обстоятелства поставят редица проблеми при архивирането на цифрова информация. Успешното архивиране на конкретен информационен обект зависи от възможностите за идентификация на данните според с технологията за представянето им във цифров вид. Технологичното разчитане на информацията не означава непосредствено, че тя ще бъде адекватно разбрана и изпълнена със смисъл. Прозрачността на данните на ниво бит е съществен елемент за успеха при съхранението на цифрова информация. Въпреки това, съдържанието винаги стои на преден план като приоритет на електронното архивиране. Точната дефиниция на параметрите на цифровата информация се постига с помощта на метаданните за нейното съхранение. В този смисъл са необходими следните основни категории от описателни сведения за информационни обекти:

**Произходът** описва първоизточника на информация в съдържателен аспект, т.е. кой е създател на първата версия, каква е историята на развитие на информационния обект в пълния му жизнен цикъл. **Контекстът** определя какво е отношението на информационния обект към други части на информационни ресурс като съвкупност от тематично свързани сведения. Информационен ресурс в този смисъл са сведения с еднакъв произход, които притежават тясна съдържателна връзка. Контекстната информация описва причините за създаването на конкретно съдържание и заедно с това включва връзките към други елементи. **Препратките** представляват група от идентификатори, с помощта на които съдържанието се иденти-

фицира еднозначно, т.е. това са атрибути за определяне на отделните инстанции в йерархичната структура на информационния ресурс. **Устойчивостта** на информацията представлява своеобразна защитна обвивка, която предпазва съдържанието от нежелателни и нерегламентирани промени. Затова дигиталните информационни обекти се съпровождат от подходящи средства за доказване на автентичност. **Правата на достъп** определят условията за употреба, съхранение и разпространение на информацията – лицензи, спецификации и мерки за контрол на достъпа.

Информационният ресурс, който е предвиден за дълготрайно съхранение, трябва да се поддържа в разчитаем вид в зависимост от актуалните възможности на целевата група. Като се има предвид субективния характер на това твърдение, електронният архив трябва да осигури оптималното съхранение на цифровата информация. За тази цел се изискват подходящи метаданни и контекстни сведения за информационния ресурс, които са адаптирани за групата от потребители. Въпреки че определен масив от сведения е бил вече дефиниран като разчитаем за дадена целева група, напълно е възможно фундаменталните и „знания“ да са се променили в течение на времето до такава степен, че съществени негови части да станат неразбираеми. Ако архивът не разполага с подобни специализирани познания и съхраняваната там информация е разчитаема само за малцина специалисти, то съществува опасността да настъпи пълна загуба на информацията, в случай че специалистите вече ги няма.

За дълготрайното съхранение на електронни документи в историческия архив трябва да намери приложение функционална схема, състояща се от четири основни модула: регистрация, информационен мениджмънт, цифрово хранилище и достъп. Звеното „**регистрация**“ предоставя услуги за приемане на цифрови масиви от фондообразуватели, за подготовката им за дълготрайно съхранение и тяхното управление в рамките на архива. При това трябва да се спазват изискванията относно стандартите и

форматите на данните както и за придружаващата ги документация, поставени от архива. Важен детайл при регистрацията на електронните документи представлява извличането на описателна информация за тях и попълването и в базата данни, т.е. информационната система на архива.

Модулът „**цифрово архивохранилище**“ изпълнява функции около трайното съхранение, поддържането на жизнеспособността на цифровия масив и евентуалното му възстановяване при загуба на интегритет. Особено внимание се обръща на поддържането на цифровите носители, върху които се пазят електронните документи в йерархията и логиката на документната структура на фондообразувателя. За тази цел е необходим презапис, контрол, създаване на рутинни процедури за възстановяване на загубите. Важна и основна дейност е предоставянето на електронни документи за използване по заявки на потребителите.

**Административната част** на системата гарантира цялостната работа на дигиталния архив. Основна дейност в този смисъл е проверката на съответствие на постъпващите и съхраняваните електронни документи с валидните стандарти за съхранение и описание. В този смисъл се изисква гарантирането на синхронност между използваните хардуерни и софтуерни решения в пълната им функционалност. За целта трябва да се водят актуални отчети за съдържанието на цифровото архивохранилище, както и да се извършва навременната му актуализация и миграция в случай на промяна на стандартите и технологиите. Затова информационният мениджмънт на дигиталния архив означава първостепенно да се поддържат актуални политики и стратегии с намерението да се постигне основната цел, а именно дългосрочното съхранение на архивните фондове в дигитален формат.

Модулът „**достъп**“ предоставя услуги, които с в помощ на крайния потребител за определяне на наличност, описанието и местонамирането на информацията. За осигуряване на достъпа е необходима двупосочна комуникация, която се състои в приемането на заявки, прилагането на

контрол на правата на достъп, ограничаване на достъпа до защитена информация и координация на достъпа до получаването на отговор на информационното запитване.

Основна задача на дигиталния архив е съхранението на информацията с една дългосрочна перспектива. Възможното решение на тази задача се състои в създаването и поддържането на общ единен интерфейс в смисъла на оперативната съвместимост. Архивните решения тук са известните емулационни и миграционни стратегии (музейната стратегия е неосъществима поради високата цена за реализирането ѝ). Непостоянният характер на определящи фактори в областта на компютърните технологии стои в непрекъснато противоречие с основната цел на дигиталния архив, а именно, дълготрайното съхранение на информационния масив. Наред с това бързият прогрес на цифровите технологии превръщат електронните системи за поддържане на цифрови архиви само за няколко години в скъпоструващи и ниско ефективно работещи предприятия. Промяната в технологиите води освен това до непрекъснато усъвършенстване на базисните фундаменти от знания в целевата група, които трябва да се поддържат в синхронност. Цифровата миграция на информационните масиви се изпълнява бавно, изисква много средства и излага информационната система на дигиталния архив пред опасността, с висока степен на вероятност да настъпи загуба на информацията. При това става дума за миграция на съдържанието, на описателните метаданни и стандартите за съхранение върху същите или нови информационни носители. Начините по които се изпълняват тези преобразувания зависят до голяма степен от реализацията на автоматизирани процедури. Високата степен на автоматизация в тази област задължително ще доведе до намаляване на вероятността от загуба на информацията. Цифровата миграция се осъществява в две главни технологии – със и без промяна на битовите последователности. Когато не се прави промяна на битовите образи, вероятността от загуба на информация е по-ниска. Методите, които не променят битовите последовател-

ности се състоят в копиране на масивите от един на друг носител от същия тип с пълно битово съответствие. Възможна е миграция със запазване на информационните обекти чрез трансфериране върху подобен или технологично обновен носител. Копирането означава освежаване на информацията, а трансферирането е репликация със запазване на информационната структура. Наред с това трябва да се имат предвид и методите за миграция с промяна на битовите образи чрез преупаковане и трансформация, при които вероятността от загуба е относително по-висока. Прави се разлика между обратима и необратима трансформация. Обратимата трансформация отразява едно към едно оригиналното представяне на информационния обект. Такава се извършва например при смяна на кодовите таблици и стандарти за битово представяне на знаците от естествени азбуки. Необратимата трансформация води до представяния, при които има загуба на смисловото съответствие. При използване на сложни формати за съхранение на информацията, където от особено значение е груповото отношение и значение в структурата от данни, задоволителната консервация на първоначалното съдържание след прилагане на механизмите за необратима трансформация става трудно доказуема.

Използвана литература:

**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.

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

**John, J.L.** Digital Forensics and Preservation. DPC Technology Watch Report 12-03 November 2012.

**Kirschenbaum, M.G., R Ovenden, G. Redvine.** Digital Forensics and Born-Digital Content in Cultural Heritage Collections. Washington D.C., 2010.