

Elektronická archivace – výzva pro odborníky více oborů

PhDr. Tomáš Kalina, Ing. Miroslav Kunt, Státní ústřední archiv v Praze

Otištěno: ISSS 2005 - Internet ve státní správě a samosprávě Hradec Králové. Sborník příspěvků. Triada, spol. s.r.o., Praha, 2005, s. 85-87.

Současný bouřlivý rozvoj informačních systémů, zejména snaha o zavedení elektronického podpisu, eGovernmentu apod. akceleruje využívání elektronické formy informací. Nutno konstatovat, že elektronická archivace, tedy **trvalé uchování** elektronických dokumentů, není v rámci tohoto rozvoje cílevědomě řešena vůbec. Příčinou je pravděpodobně to, že ani tvůrce systémů ani jejich uživatelé trvalé ukládání, tedy uchování dokumentů řádově na stovky let z důvodů především historických, v zásadě nezajímá. Jejich zájem je pochopitelně orientován krátkodobě, provozně – přístup k informaci v reálném čase, rychlá obnova systému při výpadku. Veřejná správa, ale i ostatní subjekty (původci dokumentů) by však problematiku archivace elektronických dokumentů měly vzít konečně na vědomí. Bohužel informatici a IT manažeři nevědí většinou o elektronické archivaci nic. Že jde o nadnesené tvrzení? **Za archivaci se obvykle vydává proces, který je ve skutečnosti zálohováním.**

Několik pojmů na úvod.

Pojmem “**elektronický dokument**” (dále jen ED) rozumíme veškeré soubory digitálních záznamů, které zprostředkovávají textové (alfanumerické), obrazové a zvukové informace a které vznikají, uchovávají se a využívají se **pouze prostřednictvím výpočetní techniky**. Digitální záznam je charakterizován nespojitostí. Nezabýváme se zde dokumenty zaznamenanými pomocí spojitého (analogového) signálu - např. nahrávka z magnetofonu. Hlavní nevýhodou analogového signálu je náchylnost na rušení (šum), s každým dalším kopírováním se jeho parametry zhoršují. U digitálního signálu sice dochází ke zkreslením, ale ta lze technicky snadno odstranit - při “kopírování” nedojde ke změnám. Praktickým důsledkem, kromě toho, že **neexistuje originál a kopie**, je možnost přenosu digitálního signálu beze změny informace, tedy i kvality. Důsledkem je **odmaterializování ED, tzn. nezávislost na nosiči (médiu)**.

Samotný ED - a nejen on - se skládá **ze tří složek, tvořících úplnou informaci**, na které musíme brát ohled:

- 1) **obsah**, tj. vlastní informace zaznamenané v ED (příklad: text tohoto příspěvku),
- 2) **struktura**, tj. organizace obsahu (příklad: zvýraznění nadpisů, styl a velikost písma, odstavce, způsob stránkování)
- 3) **kontext**, tj. vazby na jiné ED (příklad v papírové podobě: tento příspěvek je někde otištěn, na určitých stránkách; karta je založena na určitém místě kartotéky, v elektronické formě např. hypertextové odkazy v internetových stránkách).

U ED popis struktury a kontextu zabezpečují tzv. **metainformace (metadata)**, které nemají přímý vliv na obsah. Většina odborníků jim přikládá velký význam: mimo jiné také identifikují proces tvorby, úprav a původu vlastních dat.

Ty ED, které vzhledem k dokumentárnímu významu mají trvalou hodnotu, se stávají po posouzení obsahu v rámci **jejich odborného výběru** dle zákona o archivnictví a jeho prováděcích vyhlášek specifickým druhem **archiválie**.

Z nutnosti využívat elektronické dokumenty pouze prostřednictvím výpočetní techniky plynou jejich základní úskalí. **Výpočetní techniku (hardware, HW) nelze trvale uchovávat.**

Stárnutí elektronických součástek není možné zastavit a “postavit” nově starou mikroprocesorovou techniku nebude možné. K HW lze zařadit i média. K čemu je garantovaná životnost média na desítky let, když ho nebude možné přechít? Daleko obtížněji řešitelným problémem je závislost operačního systému a tedy i většiny aplikačního SW na konkrétní HW platformě (nebo několika platformách používaných v současnosti).

Principy elektronické archivace

a) média (nosiče)

Otázkou vhodných médií se u nás zabývala hlavně Národní knihovna ČR. Z její studie plyne jednoznačný závěr: stárnutí média nelze zastavit, je potřeba ho sledovat a data včas (než bude poškození nevratné) migrovat na novější médium. Dodejme, že migrace je řešením i v případě morálního zastarání nosiče, čímž je odstraněn těžko řešitelný problém přečtení dat nepodporovaných médií bez čtecího zařízení. Z hlediska archivace se odborníci v ČR i v zahraničí kloní k CD-R, resp. CD-ROM. Na rozdíl od např. DVD jsou u CD-R měřitelné parametry stárnutí. Jakmile budou zvládnuta podobná měření u DVD, je pravděpodobné, že teprve potom tato technologie plně nahradí CD-R. Většina pracovišť však používá ještě další médium, pokud možno odlišného typu – např. ve francouzském Národním archivu jsou to DLT pásky.

b) problematika datových formátů

Jak bylo naznačeno výše, je téměř jisté, že ve vzdálené budoucnosti nebude možné používat dnešní programové vybavení (SW). Existuje směr, který požaduje výrobu **emulátorů**, které by toto umožňovaly na nových HW/SW platformách. Tento směr má však velmi malou podporu a v dlouhodobém výhledu není považován za bezpečný.

Druhou, ve světě dnes již téměř výhradně používanou metodou, je **migrace**. Jejím průvodním jevem je archivace ED bez původního SW. Aby však bylo možné takto archivované ED v budoucnu využívat, musí být v tzv. softwarově nezávislém formátu. Takový formát je veřejně popsán a není problém ho snadno dešifrovat. Příkladem jsou formáty TXT, XML (SGML), JPEG, TIFF apod. Opakem jsou interní formáty jednotlivých systémů či výrobců: DOC (Microsoft), databáze (Informix, Oracle aj.) a bohužel i PDF, který jako licencovaný a uzavřený nemá mezi SW nezávislými co pohledávat.

Otázka formátu dat je klíčová jak pro jejich výměnu, tak pro jejich archivaci. Touto problematikou se již dlouho zabývají různé mezinárodní instituce, při Evropské komisi např. DLM forum, jejichž cílem je formulovat mezinárodní standardy a postupy. Asi největší zkušenosti jsou s databázemi. Jejich konverze na text s oddělovači nebo XML formát není problematická. Asi největší problém působí různé produkty kancelářských systémů. Zajímavé řešení navrhlo ve své studii v roce 1996 Německo pro spolkové ministerstvo vnitra. Studie předpokládala převod produktů textových editorů na komprimovaný formát TIFF, čímž by byla zachována kromě obsahu také struktura, nejproblematictější složka ED. Právě u souborů z textových editorů a elektronické pošty je otázkou, zda nejde o perspektivní řešení (spolu s konverzí obsahu do formátu TXT, který umožní fulltextové vyhledávání apod.). Využití XML se nám již nejeví úplně ideální, protože pro zachování všech nuancí struktury ED bude nutné definovat velmi složité DTD a je nebezpečí, že objem metadat převáží nad vlastním obsahem.

Včera bylo pozdě

Archivaci elektronických dokumentů se již zabývá řada zahraničních archivů. Žádný český (slovenský) veřejný archiv se však této problematice dodnes systematicky nevěnuje. Pouze na okraji své oficiální činnosti přistoupilo oddělení informatiky SÚA od roku 1996 k řešení archivace elektronických archiválií (EA). Naše práce byla napřena těmito směry:

- průzkum výskytu ED s důrazem na ED na zastaralých nosičích.
- přejímky EA od původců v předarchivní péči (včetně SÚA samotného) a archivace těchto EA; též získávání a archivace Internetových stránek (tou se systematicky zabývá i Národní knihovna ČR)
- výzkum a technické řešení archivace elektronických archiválií

Z průzkumů u ústředních úřadů docházíme k předběžnému závěru, že všechny ED z období před rokem 1990, které byly pořízeny na starých zařízeních (počítače EC 1035, Robotron aj.) jsou zničeny nebo na původním médiu v původním operačním systému – např. podací protokol úřadu předsednictva vlády, který jsme převzali na 8” disketách. S ohledem na to, že pro magnetická média uplynula doba jejich použitelnosti (např. u magnetických pásek) a na to, že všechny skutečně významné databáze byly zřejmě zničeny konstatujeme, že se tímto nejstarším obdobím již nemá cenu zabývat. Také se nám nepodařilo najít firmu, která by se migrací těchto dat na dnešní systém ještě zabývala.

O to alarmující je skutečnost, že stejně ohroženy jsou i novější významné ED, zejména databáze. To platí např. o datech kupónové privatizace, pro budoucnost prvořadým historickým pramenem. Stávající legislativa je zcela nedostatečná, stále je uplatňován pokyn Archivní správy MV z roku 1995, že vše se musí k trvalému uložení vytisknout. Je velmi smutné, že na trvalé uložení není brán ohled při budování ISVS. Proto jsme se pokusili, spíše však jako soukromou iniciativu spolu s atestační firmou a několika odborníky z MV a MI, navrhnout standard pro archivaci elektronických dokumentů, který by veřejnou správu zavázal k péči o ně. Nejnověji usnesení vlády č. 11 ze 7. ledna 2004 k dlouhodobému uchovávání a zpřístupňování dokumentů v digitální podobě předpokládá, že se Státní ústřední archiv stane trvalým úložištěm elektronických archiválií s celostátní působností. V těchto souvislostech považujeme za nevyhnutelnou spolupráci archivářů s Ministerstvem informatiky, s ostatními ústředními úřady a veřejnou správou vůbec.