

Úvod do problematiky dlouhodobé ochrany digitálních dokumentů - díl 1.

Jan Hrabal, Zdeněk Hruška

Přístup k bohatému literárnímu dědictví s mnoha typy umělecké a odborné literatury zajišťují od nepaměti sbírky paměťových institucí. Na konci minulého století se pomocí digitalizace začaly převádět klasické (tištěné) dokumenty do digitální podoby. Kvůli fyzické a chemické degradaci starých novin, častému užívání tištěných materiálů a vzácných tiskovin se měla zaručit jejich ochrana před nevratnou ztrátou. Digitalizace však nezaručuje ochranu dokumentů digitálních. Současná doba, která je charakteristická rapidním nárůstem počtu digitálních objektů a rychlými technologickými změnami, přináší řadu obtíží v oblasti ochrany a zpřístupnění digitálních dat pro budoucnost. Jako reakce na tyto problémy se vynořil nový obor Dlouhodobá ochrana digitálních dokumentů, někdy též digitální archivace (v angl. long-term preservation (LTP), digital preservation, resp. digital archiving). Jeho cílem je získávání a šíření poznatků a náležité praktické aktivity v oblasti ochrany a zpřístupnění digitálního obsahu v dlouhodobém horizontu.

Úložná média

Zatímco u tištěných materiálů je intelektuální obsah nerozlučně spjat s nosičem, jenž informaci reprezentuje, v případě digitálních informací obsažených v počítačových souborech je obsah snadno manipulovatelný, kopírovatelný a vymazatelný. Formáty počítačových souborů se spolu se softwarovým vybavením určeným pro jejich zpracování neustále vyvíjí. Stejně tak se vyvíjí i hardwarové prostředky určené pro instalaci softwaru a reprezentaci informací na displejích elektronických přístrojů. Technologické komponenty zastarávají a pod tlakem výrobců jsou nahrazovány novými. Pouhé zálohování dat již nestačí, navíc samotná média určená pro ukládání informací podléhají určité životnosti. Optické nosiče jako jsou CD a DVD nejsou pro dlouhodobou ochranu příliš vhodná. Jejich životnost se liší v závislosti na výrobcu a použitých materiálech (cca od 5 let do maximálně několika desítek) a na dnešní poměry mají malou kapacitu. Navíc se pomalu stávají zastaralou technologií, neboť na trh vstupují počítače, které již nemají hardware na čtení optických disků. Na druhou stranu CD a DVD se v mnoha odvětvích stále používají. Z hlediska poměru cena/výkon se jako nejvhodnější jeví LTO pásky (založené na magnetickém zápisu dat), které mají životnost 15 - 30 let a jejich velikost se pohybuje ve stovkách GB (LTO-5 má např. velikost 1,5 TB, nejnovější generace LTO-6 má 2,5 TB).

Referenční model OAIS - základní dokument LTP

V 90. letech minulého století neexistoval žádný sjednocující rámec pro oblast dlouhodobé ochrany digitálních dokumentů. V roce 2002 vznikl referenční model OAIS (Open Archival

Information System), který se stal ISO standardem a který problematiku konceptuálně a terminologicky ukotvil. V České republice vyšel v roce 2014 jako ČSN ISO 14721.

Celosvětově slouží jako základní pilíř LTP. Poskytuje strategie pro tvorbu praktických kroků dlouhodobé ochrany digitálních dokumentů a určuje řadu požadavků na funkční systém. Definiuje také postavení digitálního repozitáře, což je organizace lidí a technických systémů odpovědných za ochranu a dostupnost digitálních dat v dlouhodobém horizontu. Ideálně funguje jako subsystém digitální knihovny, která umožňuje jen střednědobou ochranu dat, a která má přístup k datům v repozitáři a poskytuje je koncovým uživatelům.

OAIS vznikl původně v kontextu ochrany vesmírných výzkumných dat, nicméně posléze se ukázalo, že koncept ochrany dat je společný řadě dalších institucí, které si uvědomují potřebu chránit data dlouhodobě. Nárůst digitálních objektů je zcela patrný v řadě oblastí jako průmysl, medicína, státní správa, obchodní a akademická sféra, telekomunikace a další. Význam OAIS je pro obor tak velký, že se mu budeme věnovat v jednom z dalších dílů našeho seriálu.

Strategie dlouhodobé ochrany

Existuje několik hlavních strategií pro dlouhodobou ochranu. První z nich je refreshing (obnovení), kdy jsou data zkopírována z jednoho úložného média na jiné (stejněho typu). Tato strategie není dlouhodobou v pravém slova smyslu, mnohem blíží má spíše k pouhému zálohování, protože s daty se nijak nepracuje.

Rovněž spíše krátkodobou strategií je tzv. "technologické muzeum", kdy jsou uchovávána všechna původní hardwarová zařízení a na nich pak běží daný software. I přes veškerou péči však hrozí, že se zařízení porouchá a náhradní díly již nebudou k sehnání.

Další možností je migrace. Nejdříve si popíšeme migraci nosiče - ta se používá ve chvíli, kdy se fyzický nosič stává zastaralým, ale datový formát ještě změnu nevyžaduje. Vhodným příkladem může být zkopírování dat z CD na LTO pásky, nebo z nalezené diskety (pokud je stále funkční) na novější médium. Migrace nosiče se vypořádává s hardwarovým zastaráváním a je tedy spíše krátko- až střednědobou strategií, přihlédneme-li k jednotlivým technologickým změnám probíhajícím zhruba každých 5 - 7 let.

Migrace formátu spočívá ve změně formátu ze zastarávajícího na nový. Může a nemusí být spojena s migrací na nový nosič. Při formátové migraci je nutné určit tzv. "významné vlastnosti" (*significant properties*), což jsou důležité aspekty digitálního dokumentu, které se snažíme převodem ochránit. Je potřeba, aby u obrázku byla převedena i barva? Je nutné u textu převést do nového formátu i rozdělení na odstavce nebo verše? Neztratí se smysl díla, když tyto vlastnosti nebudou převedeny? A umí to nový formát? Takové otázky je nutné si při formátové migraci položit a umět na ně odpovědět. Je také důležité vědět, jaké formáty jsou v našem repozitáři, jaká jsou jejich rizika a jestli nezastarávají. K tomu pomáhá např. registr formátů PRONOM nebo UDFR. Formátová migrace se v dnešní době jeví jako jedna z nejlepších strategií. I když ani tak není ideální, neboť stále je zde nutnost sledovat formáty včetně jejich rizik a umět na ně včas reagovat. Výhledově je tedy potřeba po určité době znovu a znovu migrovat na nové formáty,

což přináší potencionální rizika v podobě ztráty významných vlastností v digitálních dokumentech.

Jiná strategie ochrany, totiž normalizace, má společné znaky s migrací, ale využívá se především při vstupu digitálních dokumentů do repozitáře. Např. při příjmu jsou dokumenty Microsoft Word převedeny do formátu PDF/A, který je vhodným a doporučeným formátem pro dlouhodobou archivaci. Obecně lze tedy říct, že normalizace znamená převod na formáty, které jsou vhodnější pro dlouhodobou ochranu. To jsou zpravidla takové formáty, které jsou dobře popsány a mají širokou podporu - např. formáty MS Office nejsou vhodné, protože jejich popis vlastní jedna firma, ale formáty z Open Office vhodné jsou, protože se jedná o open-source, jehož specifikace je veřejně dostupná.

V poslední době se opět vrací myšlenka emulace, jako další vhodné strategie dlouhodobé ochrany. Emulací je myšleno napodobení původního prostředí, ve kterém byl digitální objekt vytvořen. Používá se pro staré programy i hry, je možné, že někteří z vás znají, nebo mají dokonce nainstalovaný program DOSbox, který se používá pro hraní her původně určených pro MS DOS. Díky emulaci je možné dosáhnout softwarové i hardwarové nezávislosti, na druhou stranu je důležitá dobrá znalost původních prostředí, aby emulace mohla být správně provedena.

Jak můžete vidět, způsobů dlouhodobé ochrany je několik, každý má svoje výhody a nevýhody. Jedna univerzální strategie bohužel neexistuje, i když různé výzkumy se snaží nějakou takovou vytvořit.

Použité zdroje (kam se také můžete podívat pro více informací):

<http://www.digitalpreservation.cz>

http://cs.wikipedia.org/wiki/Linear_Tape_Open

ROSENTHAL, Colin, Asger BLEKINGE-RASMUSSEN a Jan HUTAŘ. *Průvodce plánem důvěryhodného digitálního repozitáře (PLATTER)*. 1. vyd. Praha: Národní knihovna České republiky, 2009, 51 s. ISBN 978-807-0505-694. Dostupné z: <http://www.ndk.cz/platter-cz>

CUBR, Ladislav. *Dlouhodobá ochrana digitálních dokumentů*. 1. vyd. Praha: Národní knihovna České republiky, 2010, 154 s. ISBN 9788070505885.

<http://apps.nationalarchives.gov.uk/pronom>

http://www.dca-project.eu/images/uploads/banners/DCA_D62_Best_practices_for_a_digital_storage_infrastructure_20130506_Version1.pdf

<http://www.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>