

Praktické zkušenosti s dlouhodobou archivací dat na CD-R discích

Stanislav Psohlavec*

stanislav.psohlavec@aipberoun.cz

1 Úvod

V Národní knihovně ČR se digitalizují již devátý rok vzácné rukopisy a staré tisky. Stejnou dobu se vytvořená data ukládají na CD-R disky. Objem dat překročil 1TB a jejich hodnota je jen těžko vyčíslitelná. Kapacita, která před několika roky budila úctu i obavy z odpovědnosti za jejich zachování, je nyní záležitostí několika málo harddisků. Bezpečné uložení řeší moderní archivační systémy, které uchovávají data spolehlivě i při velmi nepříznivých katastrofických scénářích.

Typickou vlastností archivačních systémů je, že jsou postaveny na neustále běžících kontrolních procesech nad více kopiemi dat a na zajištění samoopravitelnosti běžícího systému. Bezpečnost není primárně postavena na spolehlivosti nosiče, ale na bezpečnosti procesu. Systémy samozřejmě poskytují data on-line.

Naopak při archivaci na CD-R (DVD-R, DVD+R, dále jen CD) se nezávisle na počtu záložních kopií spoléháme na statické vlastnosti záznamu dat na mediu a na jeho odolnost proti působícím destruktivním vlivům. Data nejsou obvykle k dispozici on-line. Přesto je užitečné se zamyslet se nad archivací dat na CD-R, protože je stále ještě běžnou praxí. Zkušenosti s desetiletou archivací cca 10 000 CD-R v NKČR jsou proto užitečné, zvláště co se týká jejich výběru a kontroly. O uložení se zmíním jen pro úplnost disky jsou archivovány při běžných teplotách do 20° C a při nízké vlhkosti do 30 %, ale tyto hodnoty zřejmě nejsou kritické. Disky se používají velmi zřídka a s největší opatrností, obvykle jen pro vytvoření kopie nebo pro měření. Disky se vystavují působení světla jen na nezbytné minimum při manipulaci.

2 Spolehlivost CD-R je v principech

Spolehlivost disku je zajištěna mnoha způsoby, pro naše účely si všimněme následujícího významného principu. Při čtení CD se připouští, že záznam je vždy tak trochu poškozen. K užitečným datům jsou proto na disku zaznamenána pomocná „nadbytečná“ data, která tvoří cca 30 % objemu dat zaznamenaných na disku. S pomocí této „rezervy“ (redundance) lze rekonstruovat původní zaznamenaná data i při významném poškození disku. Disk je stejně dobře použitelný, nezávisle na tom jak je tato rezerva využívána. Disk přestává být čitelný, když se tato rezerva vyčerpá z jakéhokoli důvodu. Rezerva se vyčerpává současně třemi způsoby:

1. počáteční technologické vlastnosti média
2. stárnutí média, degradační procesy
3. poškození při používání.

* AiP Beroun s.r.o., U Stadionu 157, 266 01 Beroun

Nepříjemné je, že o prvních dvou úbytcích obvykle nic nevíme a zatímco poškození či znečištění disku je viditelné, ostatní dva vlivy nejsou patrné. Uživatel je může ovlivnit jen málo a nejvíce je ovlivňuje při nákupu média.

3 Media v desetileté historii používání CD-R v NKČR

Vlastnosti medií nakupovaných a používaných pro archivaci v NK se v průběhu let měnily velmi významně. Lze rozeznat tato období:

1. Drahé, ale kvalitní nosiče se zlatou reflexní vrstvou. Tato CD-R byla určena především pro profesionální použití, měla dobré a velmi stabilní vlastnosti. Redundance byla celá rezervována pro stárnutí a poškození při používání. Pro archivaci dat to byly poklidné první čtyři roky.
2. Po roce 2000 postupně klesaly ceny medií a vypalovaček natolik, že CD-R se stalo populárním médiem především pro kopírování hudebních disků.. Archivace se stala vedlejším způsobem využívání. Přibývalo výrobců a začal mezi nimi tvrdý konkurenční boj o největšího spotřebitele, zákazníka zaměřeného na krátkodobou spotřebu, nikoli na kvalitu (audio disk hraje, když stejně poškozený CD-ROM je již dávno nepoužitelný). Důsledkem byl strmý a příjemný pokles cen, ale také nepříjemný pokles kvality, podmíněný snahou o maximální efektivitu. Výrobci se nesnažili, aby rezerva zůstala zachována pokud možno celá pro uživatele. Trh zaplavila velmi levná CD, bez problémů vyhovují pro audio, produkovaná ve velkém, obvykle na východě. To vedlo k potížím tradičních a solidních výrobců a jejich rezignaci na špičkovou kvalitu.
3. Technologie se dál vyvíjí a zdokonaluje, ale je jen malá motivace zvyšovat kvalitu.
4. Technologie se optimalizuje s ohledem na komerčně dobře využitelnou rychlost záznamu.
5. Technologie se zlepšuje dále, zvyšuje se kvalita, dochází ke stabilizaci trhu.

4 Rozptyl v kvalitě medií

Kromě výše uvedených vlivů, které probíhaly v dlouhých časových obdobích, je významné také to, že výroba medií má z hlediska kvality periodický charakter. Ve výrobě jsou z hlediska kvality tři fáze:

1. Bezprostředně po zahájení výroby jdou media do odpadu. Je pochopitelná snaha tuto domů co nejvíce zkrátit. Obvykle již po dosažení trhem akceptovatelné kvality jdou disky k zákazníkovi. Kvalita se během další uklidnění výroby postupně zvyšuje.
2. Střední část – do výroby se nezasahuje, nejvyšší kvalita.
3. Závěrečná část – kvalita se postupně zhoršuje, po dosažení minimální akceptovatelné kvality je proveden opravný zásah do výroby, výroba začíná znovu.

Rozptyl možných vlastností medií při jejich nákupu značný a náhodný, protože nelze zjistit z které fáze výroby medium pochází. Pro prodejce i výrobce je obchodně problematické nabízet výběr (obchodně: „všechny naše disky jsou nejlepší“). Ideální je přímá dohoda s výrobcem, což bylo využito v našem případě. Jiné řešení je dohodnout se s prodejcem medií na ověření vzorků z právě prodávané šarže a v případě dobrého výsledku z ní koupit jednorázově větší množství.

5 Nezbytnost měření kvality medií

Z výše uvedeného je zřejmé, že pokud chceme zodpovědně archivovat data na CD musíme mít možnost vybrat si media a měřit jejich vlastnosti, zvláště měřit výše uvedenou „rezervu“. Měření je třeba pro:

- výběr media,
- pro kontrolu provedeného záznamu,
- pro kontrolu vzorků z uložených disků, aby bylo zachyceno případný vliv stárnutí.

Měřit je třeba bezprostředně po provedení záznamu, aby v případě nedostatečné kvality mohl být záznam ihned opakován a aby do archivu nebyl uložen disk, který má již od počátku horší „rezervu“.

Měření také upozorní na stárnutí vypalovací mechaniky. Při denním vypalování může být vhodná mechaniku vyměnit již za rok. Měření disků je vhodné po čase opakovat na vzorku archivovaných disků a při zaregistrování významného poklesu „rezervy“ data ze starších disků zkopírovat na nová media. K tomuto doposud nebylo nutno přistoupit.

6 Nové možnosti

Ceny profesionálních přístrojů dovolujících měřit digitální i analogové vlastnosti záznamu se pohybují ve stovkách tisíc korun a jsou pro účely méně rozsáhlé archivace na CD neefektivní a nedosažitelné.

V návaznosti na nové vypalovací mechaniky (máme vyzkoušeno LiteOn) se objevily programy umožňující měřit digitální charakteristiky záznamu. Nové verze známého vypalovacího programu NERO, začínají také tuto možnost nabízet. Měření tímto způsobem sice neposkytuje informace o analogových vlastnostech snímaného signálu, avšak informace o činnosti signálového procesoru (takzvané digitální vlastnosti) jsou v dobré korelaci s profesionálními měřicími přístroji. Tak se tyto velmi levné a dostupné nástroje stávají dobrým pomocníkem a použitelným nástrojem pro kontrolu CD.

7 Kompletní změření archivu a praktické zkušenosti

Kompletní změření archivovaných CD bylo provedeno v roce 2001. Statistický rozbor dat neprokázal ve sledovaném období významný vliv času na hodnotitelné parametry CD-záznamu. Naopak, prokázal se významný vliv počáteční kvality media a vypalovačky.

Vzhledem k prostoru v páskovém datovém úložišti CDH-NKČR bylo dohodnuto, že data dosud archivovaná výhradně na CD-R budou uložena v kopii také do tohoto úložiště. Práce proběhly v tomto roce a nebyl zaregistrován žádný problém s čitelností archivovaných disků.

Stárnutí jakýchkoli medií je jistota, se kterou musíme počítat.

Pokud není zajištěno opravdu spolehlivé uchování dat jinými prostředky, je nutné zajistit sledování „rezervy“, protože z principu nelze vyloučit rychlý nárůst chybovosti.

Výsledky však dovolují závěr, že archivace na CD může mít své opodstatnění, nezbytnou podmínkou je kontrola stavu archivu alespoň na základě vhodně vybraných vzorků.