

# Dlouhodobá archivace digitálních dat

Walter Schorge

Brno  
18.10.2011

Praha  
25.10.2011

# O čem budeme hovořit

- Zdroje dat v muzejní práci
- Živá data, zálohování a archivace
- Média pro uchovávání dat – typy, výhody a nevýhody
- Formáty dat a morální životnost
- Technická rizika
- Organizační rizika
- Finanční rizika

# Zdroje dat v muzejní práci I

- Evidence sbírek II. stupně
  - Specializované aplikace, databáze
- Fotografická dokumentace stavu sbírkových předmětů
  - Součást specializovaných aplikací
  - Separátní uchování pro jiné využití
- Fotografická a audiovizuální dokumentace aktivit muzea
  - Archiv instituce
  - Pro budoucnost zdroj obrazového materiálu – např. osobnosti

# Zdroje dat v muzejní práci II

- Fotografická, zvuková a audiovizuální dokumentace – terénní práce, výzkum, orální historie
  - Etnologie, archeologie, zoologie, botanika, zanikající techniky a technologie...
  - Práce s pamětníky
- Digitalizace sbírek a dokumentačního materiálu
  - Zpřístupnění prostřednictvím ICT
  - Ochrana originálů při badatelské činnosti
  - Záchranná digitalizace (bezprostřední ohrožení originálů)

# Zdroje dat v muzejní práci III

- „Digitally born“ dokumenty
  - Aktuální problém archivů – digitalizace státní správy
  - Doprovodná dokumentace – web, pdf dokumentace k technickým výrobkům, ...
- Administrativní dokumenty
  - Dokumenty vznikající chodem instituce
  - Korespondence (e-mail)
- **Digitální věk je bezprostředním ohrožením klasické muzejní práce!**

# Živá data, zálohování, archivace

- Živá data – data se kterými se pracuje, průběžně se mění
- Záloha – kopie dat sloužící pro obnovu v případě ztráty primárních dat – havárie, trestný čin, lidská chyba
- Záloha se uchovává zpravidla nejvýše několik týdnů
- Archivace dat – **proces**, který slouží k dlouhodobému uchování dat

# Média pro uchování dat I

- Optická média – CD, DVD, Blu-ray, UDO
- Pásy
- Pevné paměti – flash disky, paměťové karty
- Pevné disky v PC, serverech, speciálních zařízeních

# Média pro uchování dat II

- Paměťová média lze dělit z hledisek:
  - energetická závislost
    - paměti vyžadující pro zachování dat přísun energie (paměti typu RAM)
    - paměti nevyžadující pro zachování dat přísun energie (např HDD, FDD, optické disky, flash, EEPROM, ...)
  - režim práce s daty
    - paměti umožňující pouze čtení dat (paměti CD-ROM, CD, DVD, Blu-ray, ...)
    - paměti umožňující opakované čtení dat ale pouze jedno zapsání informace (paměti CD-R, DVD±R, WORM)
    - paměti umožňující opakované čtení i záznam dat (paměti EEPROM, HDD, FDD, CD-RW, DVD-RW, ...)



# Média pro uchování dat III

- fyzikálního principu
  - retence náboje (polovodičové paměti založené na tranzistorech)
  - změna orientace remanentní (trvající) magnetické indukce (HDD, FDD, pásy)
  - vychýlení laserového svazku (CD, DVD, Blu-ray)
  - rozklad materiálu (optické disky umožňující jeden zápis - CD±R, DVD±R)
  - fázová transformace materiálu (přepisovatelné optické disky (CD-RW, DVD±RW, DVD-RAM, BD-RE)
  - spintronika (spin = kvantová vlastnost elementárních částic)

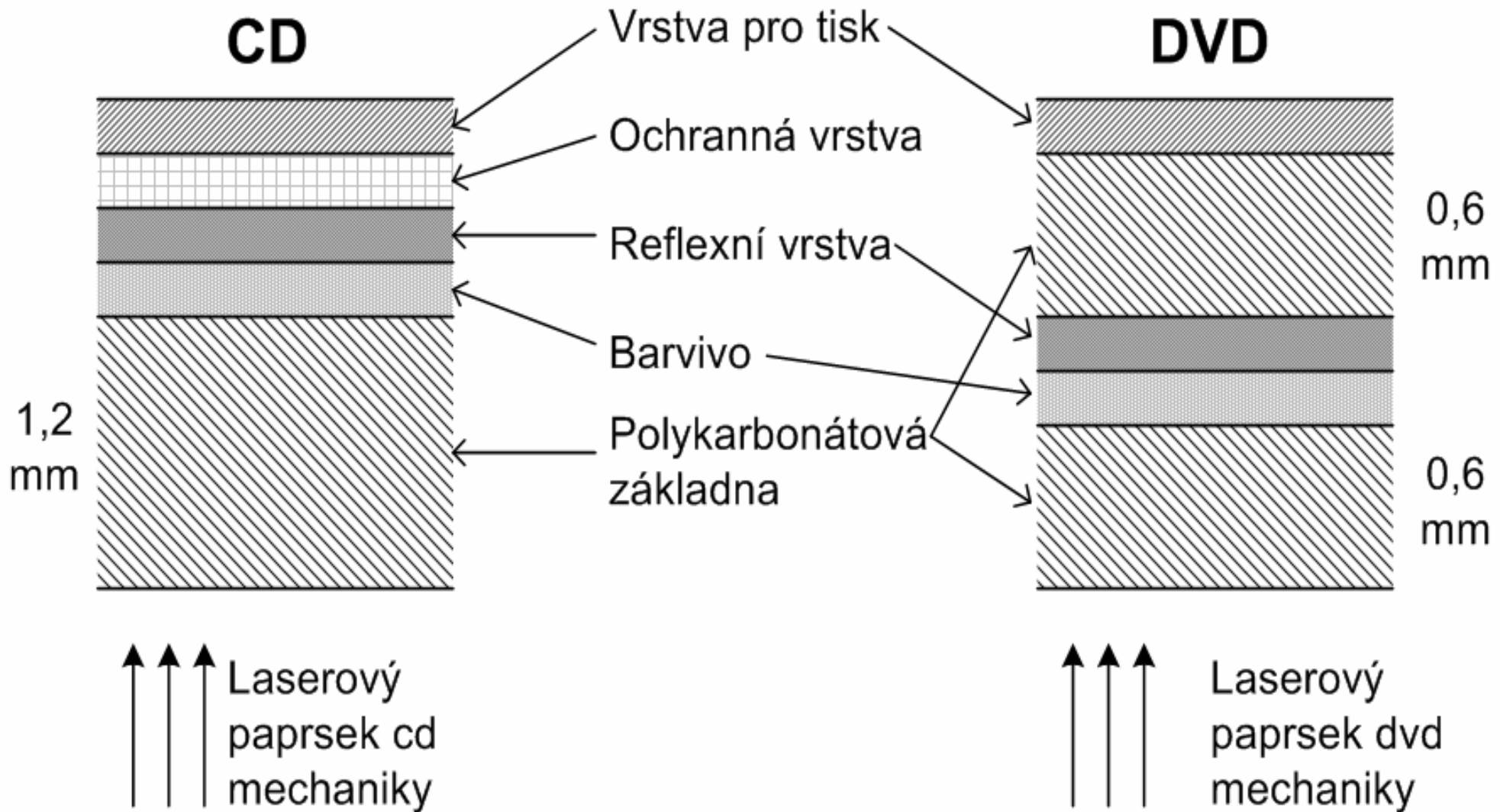
# Média pro uchování dat IV

- přítomnosti pohyblivých segmentů
  - statická média (polovodičové paměti)
  - rotující média (klasické pevné disky, optické disky)
- provedení
  - autonomní paměť (polovodičové paměti, pevné disky)
  - mechanika + médium (optické disky, pásy)
- odolnosti vůči
  - mechanickým šokům (vibrace)
  - teplotě
  - elektromagnetickým polím
  - energetickému záření
- praktických parametrů
  - rychlost
  - kapacita
  - cena za jednotku kapacity

# Parametry CD a DVD I

- Stejné rozměry, standardně průměr 12 cm, tloušťka 1,2 mm
- Datový záznam je zapsán v jedné spirále od středu k okraji disku
- Kapacita 700 MB (CD), 4,7 GB (jednostranné jednovrstvé DVD), 8,5 GB (jednostranné dvouvrstvé DVD)
- U CD je kompaktní polykarbonátová vrstva 1,2 mm z jedné strany, u DVD jsou dvě vrstvy 0,6 mm silné

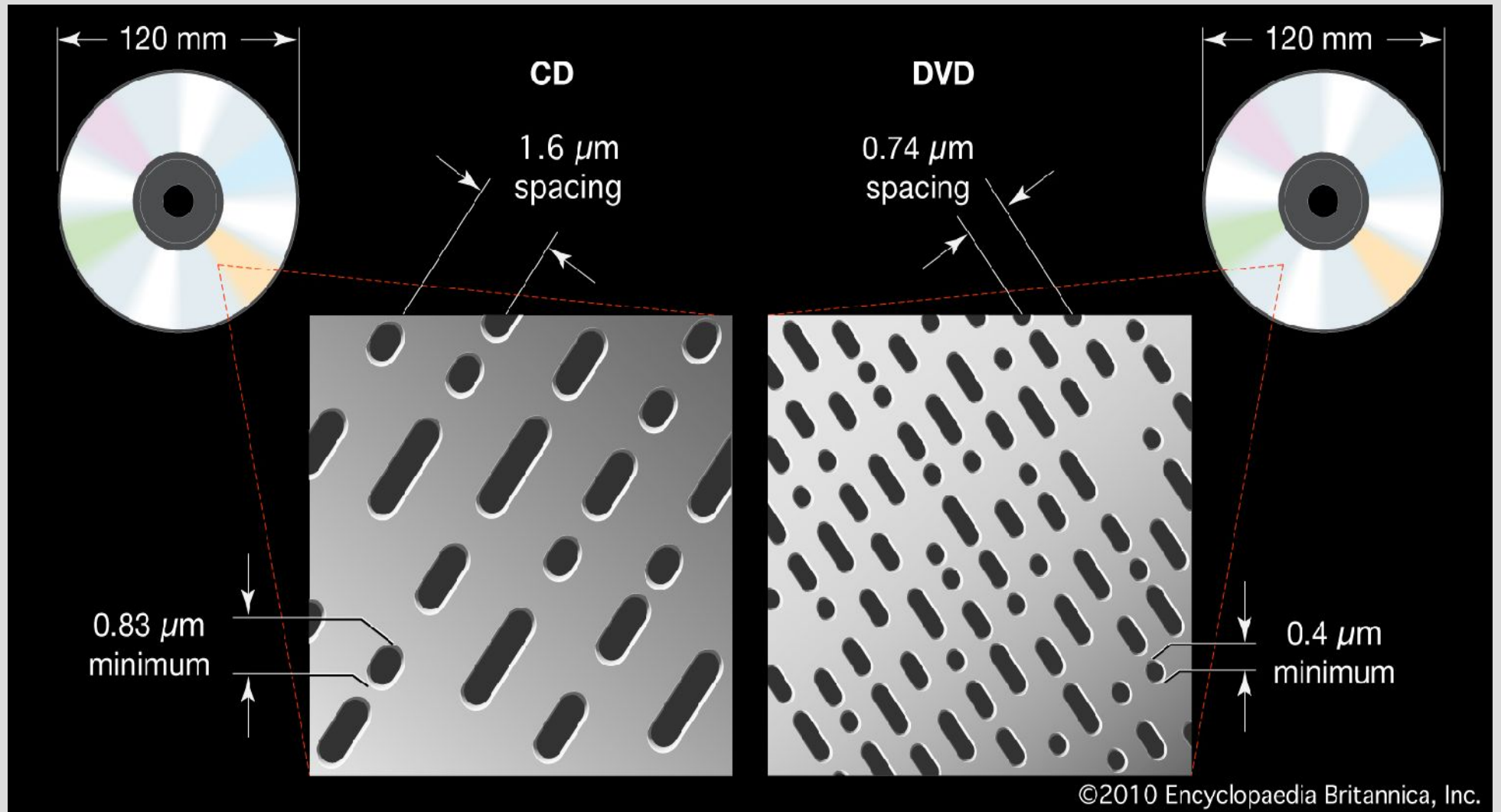
# Parametry CD a DVD II



# Parametry CD a DVD III

- **DVD používá větší hustotu záznamu**
  - U DVD jsou píty kratší (minimální délka píty je u DVD 0,4  $\mu\text{m}$  oproti 0,83  $\mu\text{m}$  u CD)
  - Vzdálenost datových stop je menší (0,74  $\mu\text{m}$  u DVD oproti 1,6  $\mu\text{m}$  u CD)
  - Změnila se také vlnová délka laseru čtecí mechaniky (ze 780 nm – infra červená u CD na 640 nm – červená u DVD)
- **DVD používá jiný souborový systém – Universal Disk Format (UDF) namísto ISO 9660 u CD**

# Parametry CD a DVD IV

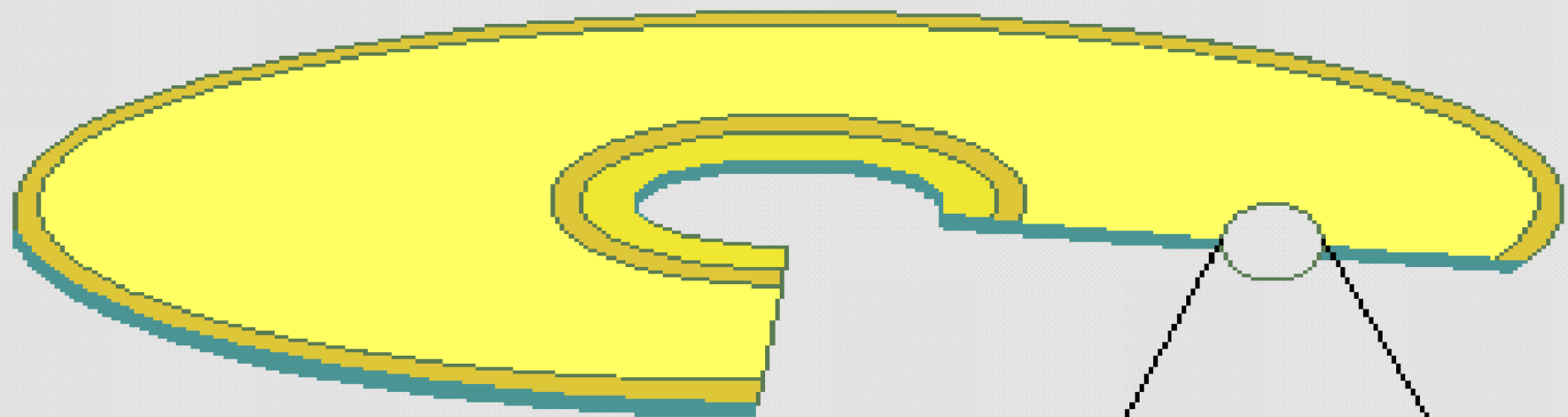


# Parametry CD a DVD V

- Lisovaná média (tovární výroba)
- „Vypalovaná“ média (zapisovatelná na PC)
- CD-R (recortable), CD-RW (rewriteble)
- DVD-R & DVD-RW – vysoké licenční poplatky >> nahrazeny DVD+R & DVD+RW
- Existují kombinované mechaniky DVD $\pm$ RW
- RW média – fázový přechod mezi krystalickou a amorfní (beztvarou) fází kovové slitiny

# Parametry CD a DVD VI

- Struktura zapisovatelného média CD-R



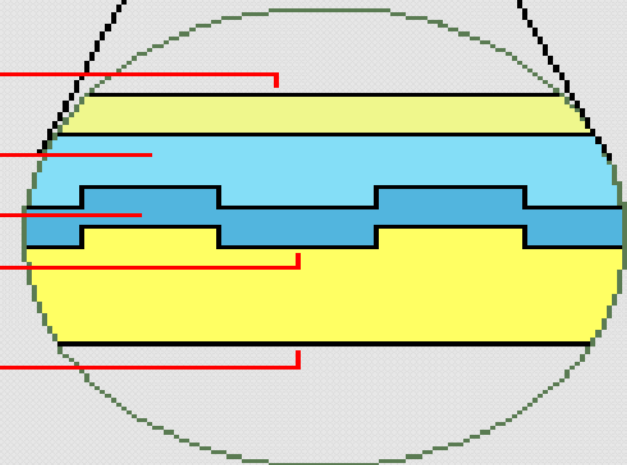
**Ochranná lakovaná  
vrstva**

**Reflexní fólie**

**Záznamové barvivo**

**Předlisovaná drážka**

**Polykarbonátový základ disku**





# Parametry Blu-ray I

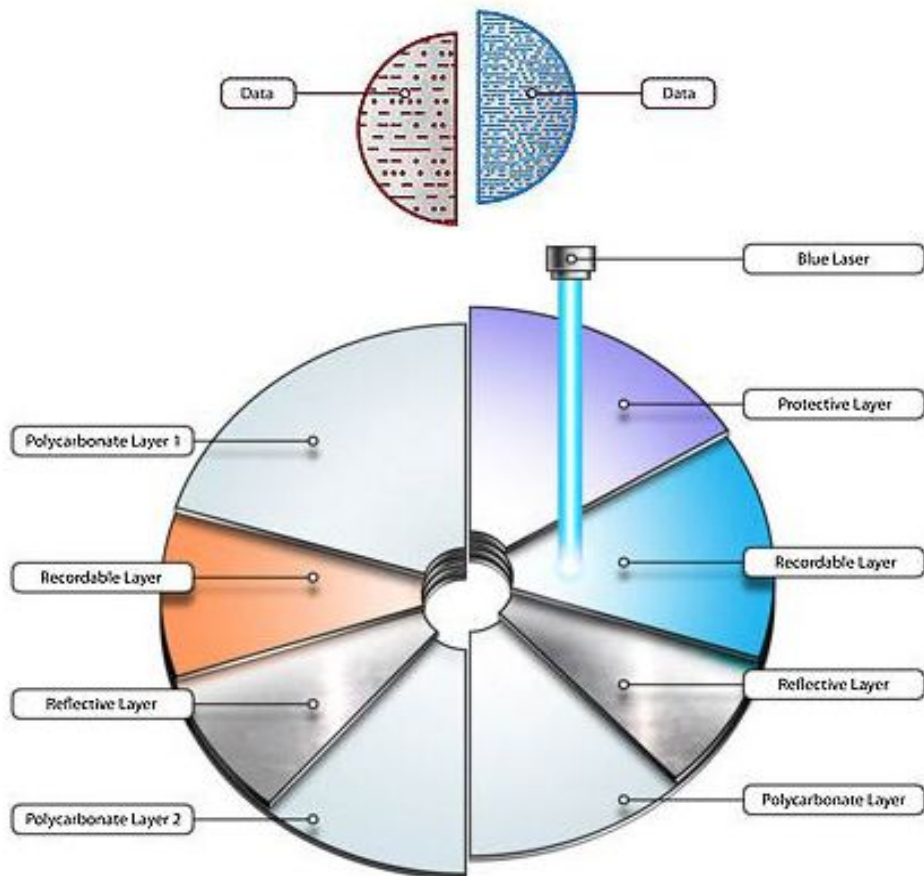
- Třetí generace optických disků
- Data se ukládají do spirály 0,1 mm pod povrch disku, možnost hybridu DVD/BD
- Příčný odstup stop je 0,35  $\mu\text{m}$
- Vlnová délka laseru 405 nm (modrý laser)
- Stejný rozměr jako CD a DVD
- Kapacita až 25 GB (jednostranné, jednovrstvé médium)
- Označení: BD-ROM, BD-R, BD-RE
- Souborový systém UDF (jako u DVD)
- ???

# Parametry Blu-ray II

## How a Blu-ray Disc Works

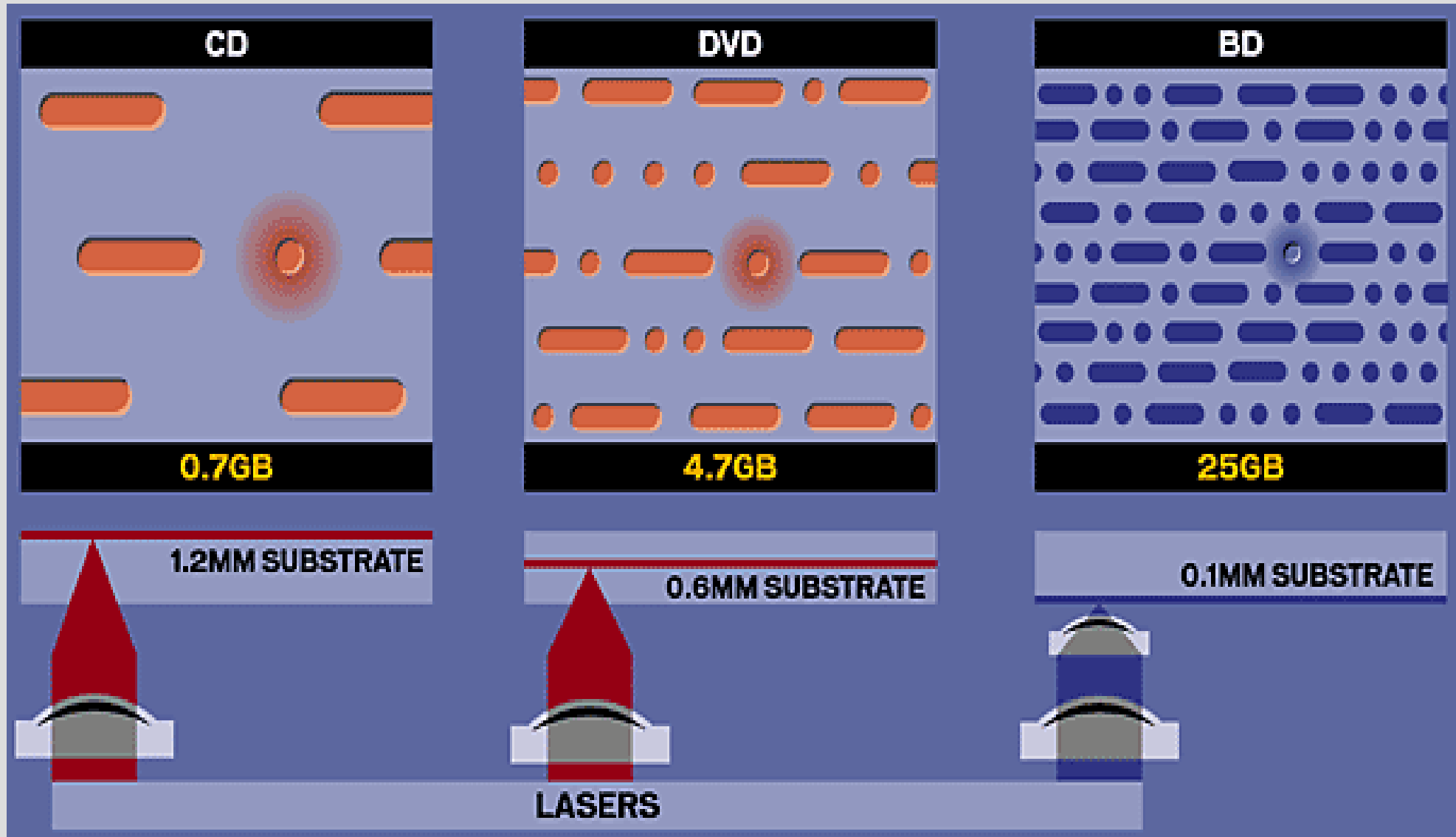
Current standard definition  
DVD using red laser

Blu-ray High Definition  
disc using a blue laser



Porovnání DVD a BD

# Parametry Blu-ray III



# Parametry UDO

- Ultra Density Optical (UDO)
- Technologie podobná Blu-ray
- Větší disk (130 mm), zapouzdřený (5,25“)
- Udávaná životnost > 50 let
- WORM technologie
- Kapacita 60 GB, plán 120 GB (2013)
- Cena disku 60 GB cca 1 500,- Kč
- Používá např. Archiv hl. m. Prahy



# Parametry datových pásek I

- Záznam dat na magnetické pásky (magnetická páska je pevné médium sestávající z magnetické vrstvy nanesené na plastické pásce)
- Obdobná technologie jako u běžných audio a videokazet
- Citlivost na magnetické pole (vč. elektromagnetického)
- Ovlivňování záznamu sousedními závity při dlouhodobém skladování bez přetáčení

# Parametry datových pásek II

- Pásky se používají na zálohování a archivaci velkého objemu dat, běžně uložených na diskových polích serverů, příp. hromadné zálohování pracovních stanic apod.
- Nabízejí funkcionalitu WORM (nelze přepsat, modifikovat data)
- Pásky nejsou vhodné jako primární médium pro přístup k opakovaně využívaným datům

# Parametry datových pásek III

- Nejrozšířenější typy v posledních letech
  - DAT – malá kapacita, malé rozměry, nejzranitelnější
  - Digital Linear Tape (DLT, SDLT) – technologie na ústupu
  - Linear Tape Open (LTO) – aktuální formát, v současnosti verze LTO-5 z roku 2009 (kapacita 1,6/3,2 TB), v nejbližší době bude LTO-6
- Nižší cena média, vyšší pořizovací náklady na mechaniku
- Uvádí se životnost záznamů 15-30 let
- Manipulace s daty je zdlouhavá

# Parametry datových pásek IV



Interní mechanika a cartridge LTO-5





# Parametry datových pásek V

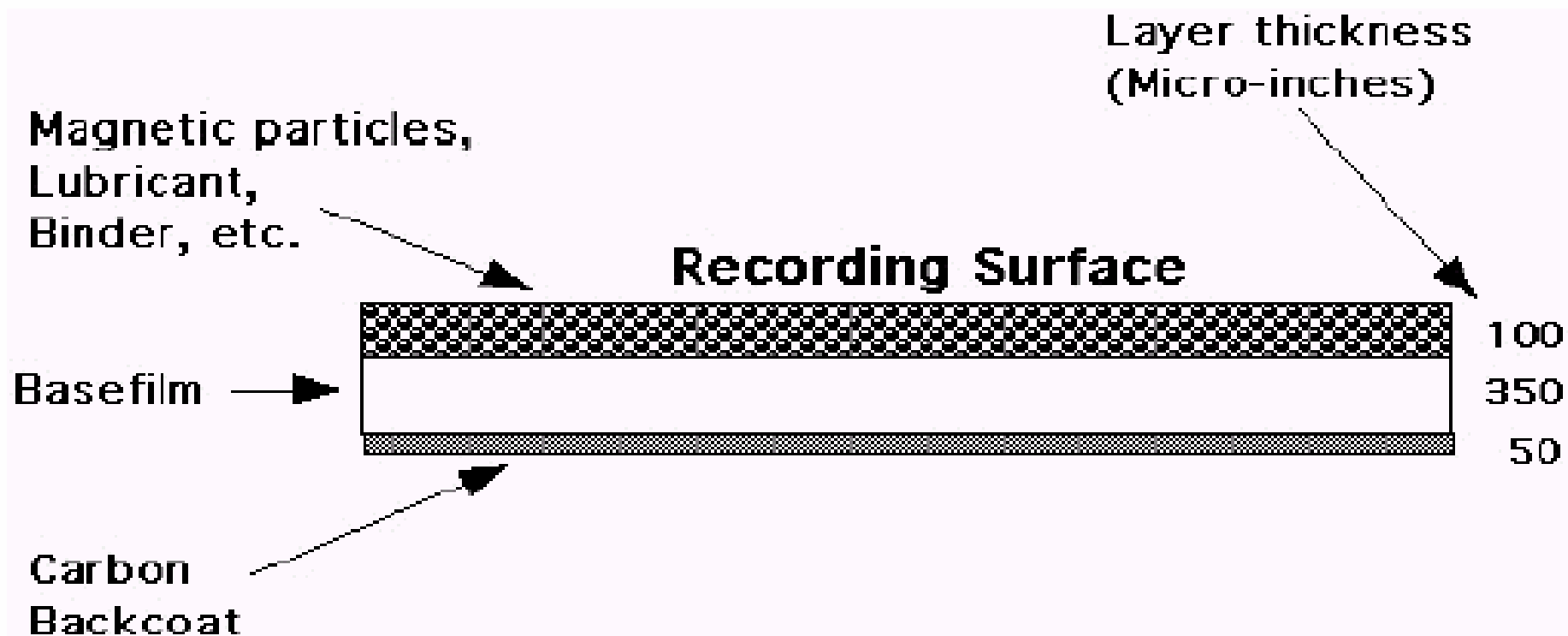
- Kromě standardních mechanik se používají páskové robotické knihovny



- Pásy mají v současné době klesající využití s ohledem na vzestup kapacit pevných disků a pokles jejich cen

# Parametry datových pásek VI

- Řez páskou



# Parametry pevných pamětí I

- USB flash disky



- Paměťové karty



# Parametry pevných pamětí II

- Polovodičové paměti typu flash
- Paměťové karty různých typů:
  - CompactFlash (CF)
  - Memory Stick (MS)
  - Multimedia card (MMC)
  - Secure Digital (SD)
  - Smart Media (SM)
  - xD-Picture card (xD)
- Vhodné pro mobilní zařízení, přenos dat
- Odolné proti vibracím a rázům
- Životnost se udává 10 let
- Kapacita max. desítky GB

# Parametry pevných disků I

- Médium pro „živá“ data
- Nejrychlejší zápis a čtení dat ze všech uvedených médií
- Magnetický záznam dat (klasické rotační disky), paměti flash (SSD disky)
- Počítače (pracovní stanice, notebooky)
- Servery
- Specializovaná disková pole pro ukládání dat
  - zařízení NAS, síť SAN
- Externí HDD

# Parametry pevných disků II

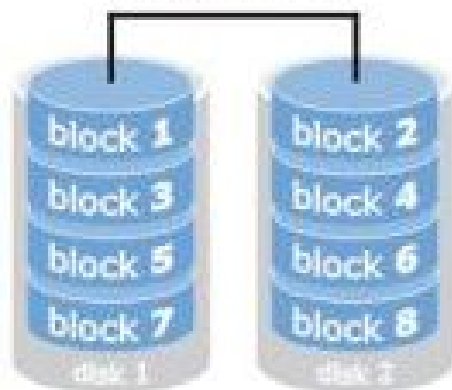
- Počítače používané pro běžnou práci jsou pro archivaci zcela nevhodné (neexistence redundance při ukládání dat, riziko chyby uživatele, u NB riziko ztráty, uložení v méně chráněných prostorech)
- Vhodnější jsou servery a speciální disková pole vybavená RAID (Redundant Array of Inexpensive/Independent Disks – vícenásobné diskové pole laciných/nezávislých disků)

# Parametry pevných disků III

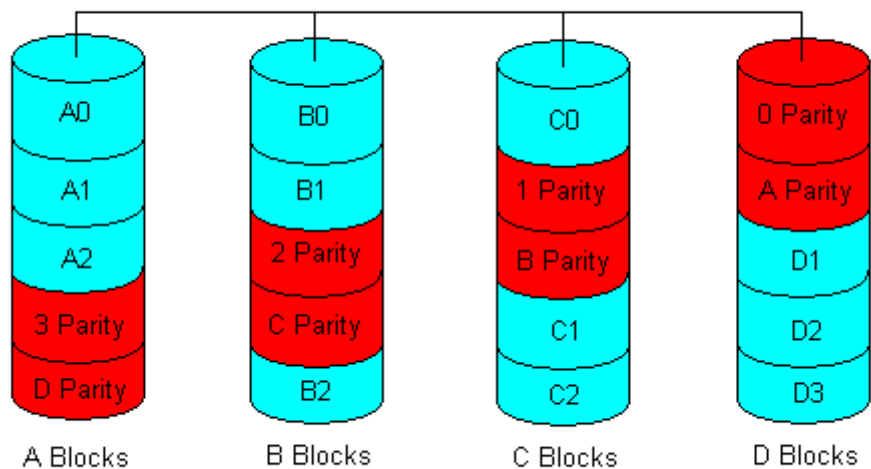
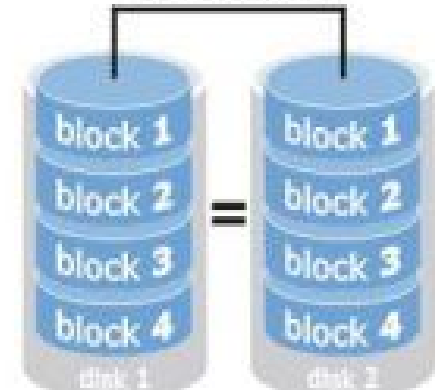
- Servery a disková pole by měla být profesionálně spravována (monitoring, zálohování, prevence)
- Redundantní ukládání dat – RAID
  - RAID 0 – zcela nevhodný
  - RAID 1 – vhodný pro malé objemy dat
  - RAID 5 – nejčastěji používaná varianta
  - RAID 6 – bezpečnější než RAID 5 (duální parita)
  - Existují i další varianty v praxi méně používané

# Parametry pevných disků IV

## RAID 0

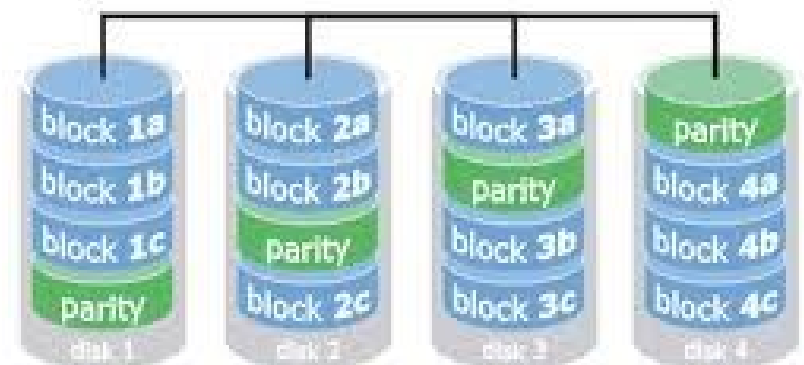


## RAID 1



## RAID 5

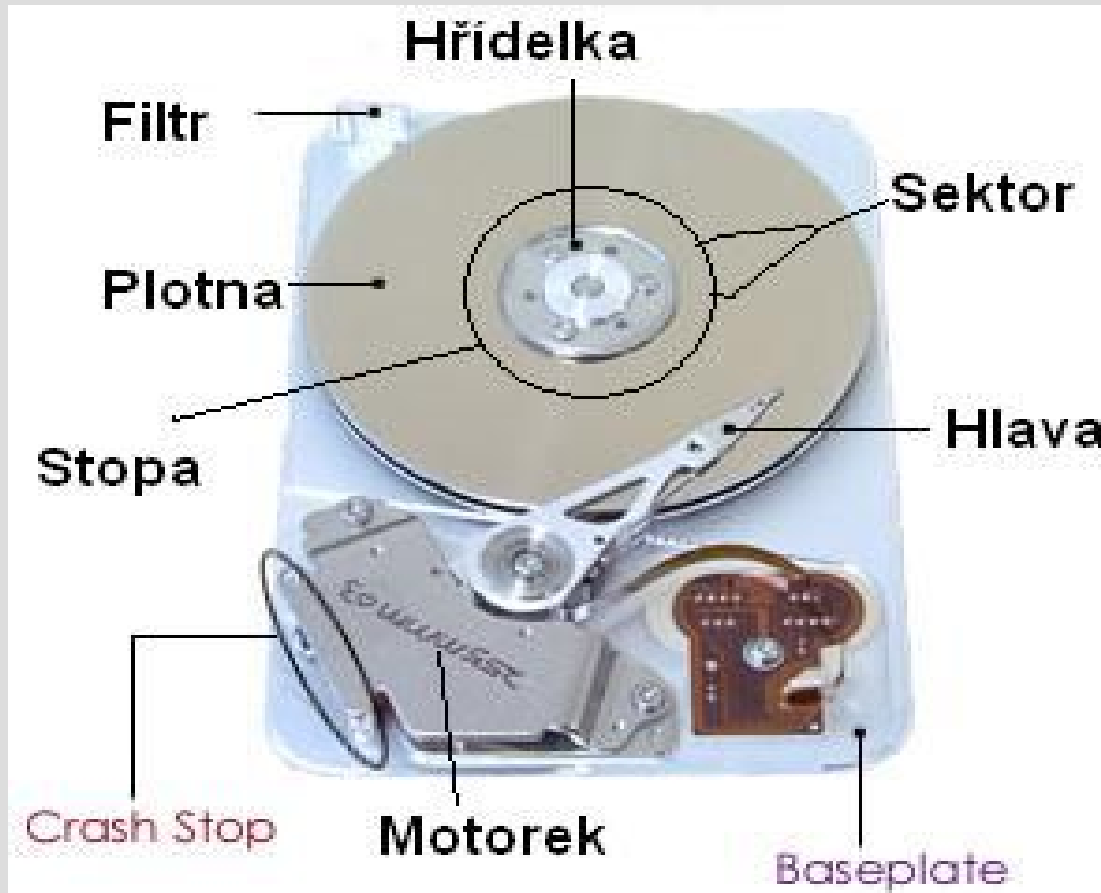
parity across disks





# Parametry pevných disků V

- Konstrukce pevného disku
- U velkokapacitních disků je více ploten a hlav



# Parametry pevných disků VI

- Zařízení NAS lze pořídit v různých velikostech a cenách



# Formáty dat a morální životnost

- Data pro dlouhodobou archivaci je nutné ukládat ve standardizovaných formátech
- V průběhu archivace bude nutně docházet k migraci na nové nosiče, pravděpodobně však i do nových formátů
- Preferované formáty dat – viz např. projekt NDA ČR - [http://www.nacr.cz/Z-Files/nda\\_proj1.pdf](http://www.nacr.cz/Z-Files/nda_proj1.pdf)
  - Výběr pro potřeby českých muzeí:
    - prostý text UTF-8, XML, PDF/A-1 - textové dokumenty
    - TIFF nekomprimovaný, PNG - rastrové obrazy
    - WAV (PCM) – zvuk
    - MPEG-2, AVI nekomprimovaný, QuickTime Movie nekomprimovaný - video

# Technická rizika

- Technická rizika jsou prvním ze tří faktorů, který ohrožuje archivovaná data
  - Porucha technického zařízení (pevné disky)
  - Degradace média vlivem vnějších vlivů (UV záření, vlhkost, teplo, elektromagnetické pole) nebo samovolná degradace materiálu média vedoucí k nečitelnosti dat (např. CD, DVD, pásky)
  - Morální zastarání média a absence zařízení pro čtení dat
  - Nesrozumitelnost proprietárních dat pro budoucí software

# Ohrožení CD a DVD I

- Větší hustota záznamu u DVD může vést k větší chybovosti ve čtení, DVD však mají efektivnější korekci chyb
- Může docházet ke znehodnocování polykarbonátu a k fotochemické degradaci materiálů média
  - Stárnutí materiálů se projevuje změnami v jejich struktuře a vlastnostech. Pro dobrou čitelnost DVD médií jsou z pohledu stárnutí důležité změny v optických vlastnostech, ke kterým by mohlo v průběhu času docházet. Jedná se o průhlednost polykarbonátu, odrazivost reflexní vrstvy a změny v záznamové vrstvě

# Ohrožení CD a DVD II

- Byly provedeny testy DVD+R médií
  - Působení teploty 55, 85 a 100 °C po dobu 48 hodin – na čitelnosti se neprojevovalo
  - Působení teploty 125 °C po dobu 24 hodin způsobilo deformaci média, přestože nedošlo ke změně optických vlastností (kritická teplota u polykarbonátu je 130 °C), média byla nečitelná
  - Vlhkost 100% při teplotě 50 °C po dobu 48 hodin – navlhla místa na záznamové vrstvě, po vysušení bez chyb
  - Dlouhodobější působení vlhkosti způsobuje růst hub na záznamové vrstvě

# Ohrožení CD a DVD III



Disk napadený houbou

# Ohrožení CD a DVD IV

- Test DVD médií – působení slunečního záření – 200 hodin, teplota okolo 40 °C
  - Fotochemická degradace je proces, při kterém dochází vlivem absorbovaného záření ke strukturálním přeměnám v materiálu, což se může projevit jako mechanická nebo optická změna vlastností polykarbonátové složky DVD média.
- Média byla po testu čitelná, ale s velkým množstvím chyb



# Ohrožení pásek

- Magnetické pole – smazání záznamu
- Dlouhodobé uložení bez přetáčení (prokopírování závitů) – zhoršení čitelnosti
- Nevhodné klimatické podmínky – optimální teplota 8 °C (ne méně!) a RV 25%
- Degradace lepidla nebo podložky – lepkavost bránící přehrání (např. problém v Ústavu pro studium totalitních režimů)

# Ohrožení pevných disků

- Porucha disku vlivem vnitřní vady
- Porucha řadiče diskového pole
- Výpadek napájení
- Otřesy, rázy, vibrace
- Nedostatečné chlazení (mj. nezajištění redundance chlazení v serverovně)
- Živelné události
- Přepětí, výkyvy v elektrické síti
- On-line kriminální jednání – viry, hacking

# Organizační rizika

- Druhý faktor, neméně důležitý
- Nastavení pravidel a kontrola jejich dodržování, zajištění kontinuity bez ohledu na personální obsazení
- Ochrana proti protiprávnímu jednání
  - Zabezpečení serverů, diskových polí a off-line médií – cizí osoby, vlastní zaměstnanci a dodavatelé
- Kontrolní a bezpečnostní mechanismy proti lidské chybě

# Finační rizika

- Pořízením dat a požadavkem na jejich uchování vytváříme trvalý závazek financování archivace dat
- Při výběru archivačních médií/technologie je nutné uvažovat nejen pořizovací cenu, ale i TCO (Total Costs of Ownership) – celkové náklady na vlastnictví
  - Udržování, periodická kontrola a migrace dat
  - Investice do obměny technického vybavení
  - Náklady na udržování vhodných skladovacích prostor
  - Elektrická energie
  - Kvalifikované lidské zdroje

# Shrnutí a doporučení I

- S ohledem na místní podmínky a množství dat vybrat kombinaci médií
- Data ukládat na různé typy médií a na různá místa (různé budovy, regiony)
- Optimálně předat data do správy různým osobám
- Nastavit organizační pravidla a jejich dodržování
- Zajistit průběžné financování (podle průzkumů v současné době archivace dat stojí 50% nákladů na IT, očekává se 75%)

# Shrnutí a doporučení II

- Náklady na uložení 1 GB dat (pořizovací náklady média)

Médium	Cena za 1 GB	Cena čtecího zařízení
Pevný disk	1,40 Kč	500,- Kč
DVD+R v krabičce	3,40 Kč	500,- Kč
Páska LTO	0,70 Kč	80 000,- Kč
Paměťová karta SD	60,00 Kč	400,- Kč

# Shrnutí a doporučení III

- Moje doporučení
  - Ukládejte data triplicitně v doporučených formátech včetně metadat
  - První kopii dat ukládejte na pevné disky on-line s řádným zabezpečením
  - Druhou kopii ukládejte na pevné disky on-line či off-line v jiné lokalitě
  - Třetí kopii ukládejte na optická média nebo na pásky (máte-li velký objem dat), média ukládejte na bezpečné místo jako zálohu, tuto pravidelně obnovujte
  - Nezapomeňte na organizační pravidla a zajištění průběžného financování

# Děkuji za pozornost!

Mgr. Walter Schorge  
Národní technické muzeum v Praze  
oddělení informačních služeb a digitalizace  
walter.schorge (zavináč) ntm.cz