

W miarę wzrostu wykorzystania cyfowego zapisu informacji, archiwa zaczynaja przywiązywać coraz większą wagę do mediów w postaci elektronicznej. Dotyczy to zarówno archiwów klasycznych które digitalizują swoje zasoby, jak i instytucji ktore archiwizuja dokumenty i zapisy powstałe w ostatnich dekadach, które sa w coraz większym stopniu cyfrowe od początku ich powstania (‘born digital’).

Zapis cyfrowy wymaga zupełnie innego podejścia do problemu zachowania i zabezpieczenia zasobów archiwalnych. W jednym z poprzednich blogów rozważaliśmy oprogramowanie do inwentaryzacji zasobów, tutaj chciałbym przedyskutować problem pojemności pamięci cyfrowej, potrzebnej do przechowywania zasobów.

Zapis informacji w postaci cyfrowej poprzedza oczywiście powstanie komputerów. Karty dziurkowane były uzywane od poczatku 19 wieku - w krosnach (Joseph Jacquard), w przetwarzaniu informacji (Siemion Korsakow), w opracowywaniu danych spisu powszechnego (Herman Hollerith). Jeszcze niedawno maszyny cyfrowe Odra używaly (za IBM) kart dziurkowanych do zapisu programów i danych. Wkrótce zostały one zastąpione papierową taśmą perforowaną, ale prawdziwe przyspieszenie spowodowało dopiero użycie zapisu magnetycznego. Rewolucja komputerowa to pamięc dyskowa: najpierw mainframe, potem stacje robocze, komputery osobiste, laptopy - to wszystko istniało dzieki możliwości zapisu informacji na dyskach twardych i przenośnych dyskietkach. Dysk twardy ma chyba najdłuższa historię rozwoju technologicznego, i jest do dziś podstawowym medium zapisu danych w Internecie oraz w większości komputerów stacjonarych i laptopów.

Historia dysku twardego jest fascynującą podróżą po kolejnych, rewolucyjnych zmianach technologii, które owocowały urządzeniami coraz lepszymi, o coraz większej gęstości zapisu, coraz większej niezawodności. Napędy z wymienialnymi dyskami zostały zastąpione zintegrowanymi napędami, rozmiary to kolejno standardy 14-calowe, 8-calowe, 5.25 calowe, 3.5 calowe, 2.5 calowe itp. Wielkość dysku mierzy się jednak przede wszystkim liczbą bajtów które

można na nim zapisać, gdzie bajt to liczba 8-bitowa, (0 do 11111111 w formacie dwójkowym, 0 do 255 w dziesiętnym). Kolejne przedrostki pokazują rosnącą pojemność - megabajt (MB) to milion bajtów, gigabajt (GB) to tysiąc milionów, terabajt (TB) to milion milionów bajtów.

Jeden MB to w przybliżeniu obraz o rozmiarze 1024x1024 pixli (bitmap) lub 4 megapixlowe zdjęcie skompresowane w formacie jpg; około 1 minuty muzyki w formacie mp3, albo typowa książka bez ilustracji (500 stron, 2000 znaków na stronę). W Instytucie Piłsudskiego połączony katalog bibliotek polonijnych (IJP, PIASA i PCF) zajmuje 20 MB w formacie MARC.

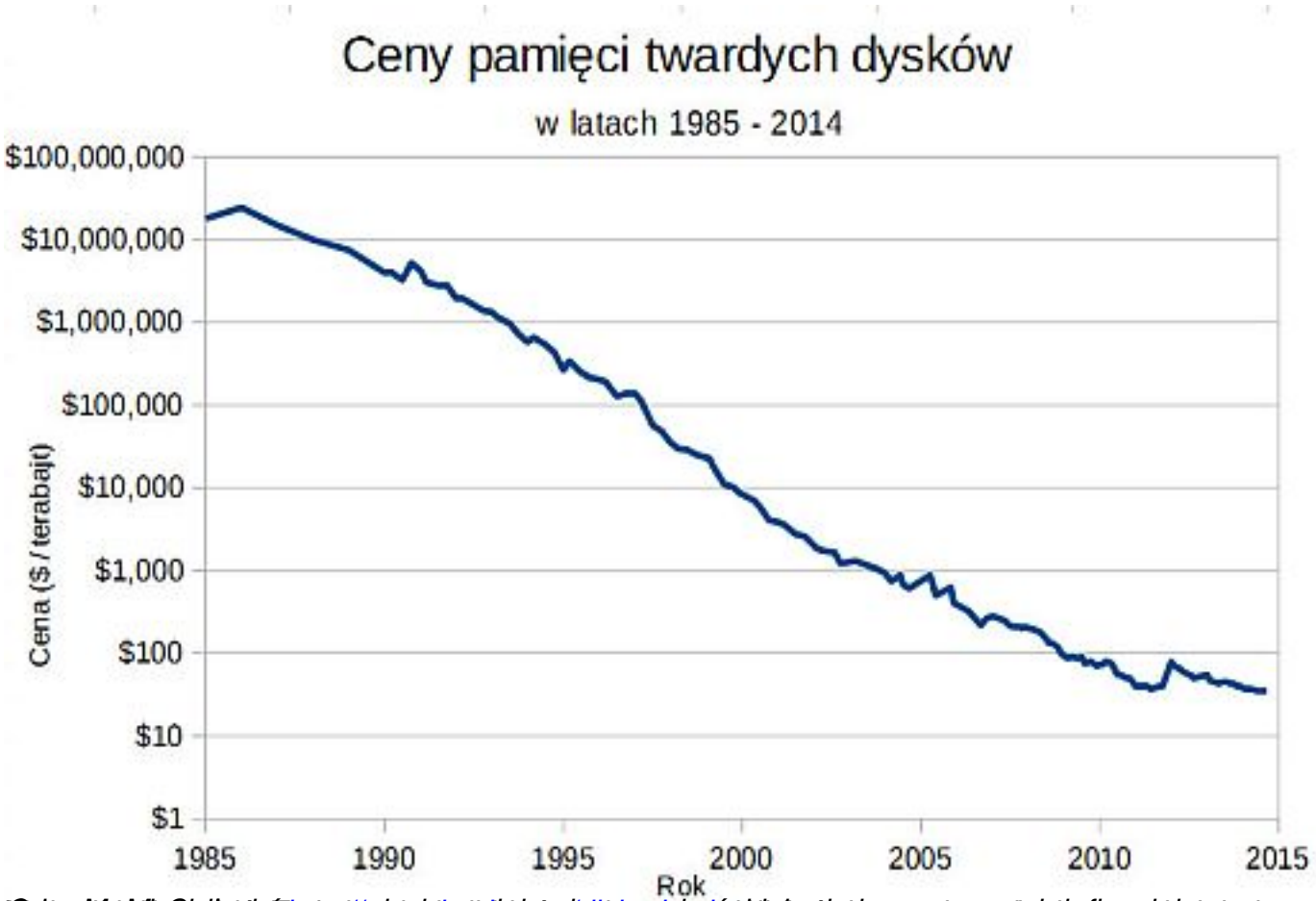
Jeden GB mieści jednogodzinny film o standardowej rozdzielczości, 700 dyskietek 3.5 calowych, tysiąc zdjęć o średnim rozmiarze 3-4 megapixli, 20 tomow encyklopedii, cały ludzki genom. W Instytucie Piłsudskiego zdigitalizowane nagrania wideo wywiadów z osobami wywiezionymi do łagrów w Rosji Sowieckiej zajmują ok. 200 GB.

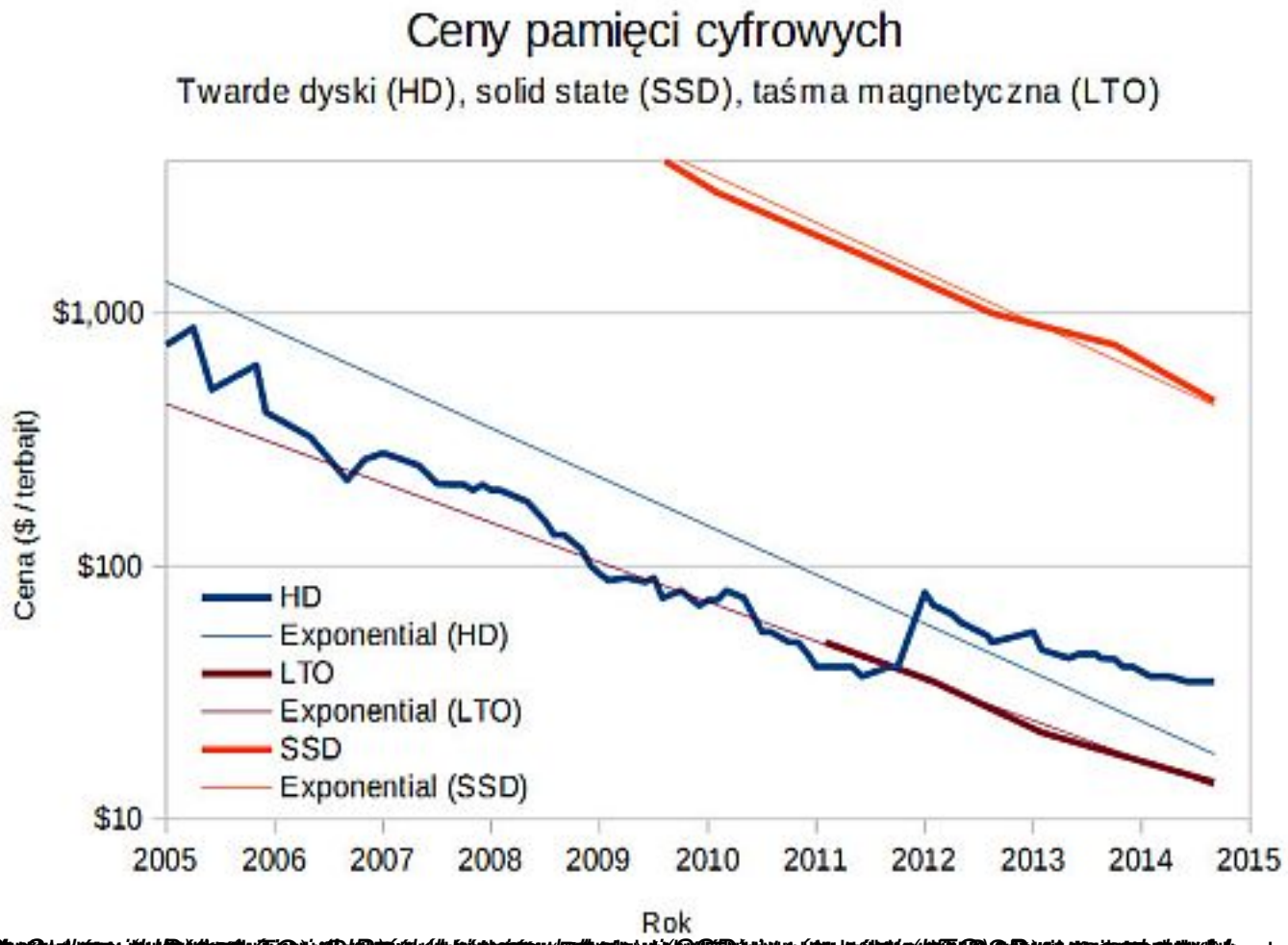
Jeden TB mieści 17 tysięcy godzin (prawie dwa lata) muzyki w postaci mp3, tysiąc godzin (40 dni) filmu, tysiąc milionów stron tekstu, 1500 dysków CD, 40 dysków Blue Ray, 50 tysięcy drzew przerobionych na papier i zadrukowanych. W Instytucie zdigitalizowany zespół Józefa Piłsudskiego, zawierający oryginalne skany kolorowe w rozdzielczości 600 dpi, kopie skompresowane o zmniejszonej rozdzielczości oraz wszystkie metadane zajmuje 2 terabajty objętości.

Gdzie to wszystko pomieścić?

Czy archiwa stać na terabajty pamięci dyskowej? Czy nie lepiej zostac przy mniejszej rozdzielczości i zapisywać wszystko na złotych CD o przedłużonej trwałości? Aby odpowiedzieć sobie na takie pytania, spróbujmy przyjrzeć się kosztom przechowywania informacji na dyskach magnetycznych.

Wykres poniżej pokazuje historyczne ceny pamięci dyskowej w dolarach US na terabajt. Widzimy tu bardzo regularny trend obejmujący 3 dekady produkcji twardych dysków, od 1985 roku. Jeszcze w 2000 roku terabajt kosztował $10,000, dziś kosztuje poniżej $50.


Ceny pamięci twardych dysków
w latach 1985 - 2014
$100,000,000
$10,000,000
$1,000,000
$100,000
$10,000
$1,000
$100
$10
$1
Cena ($ / terabajt)
1985
1990
1995
2000
2005
2010
2015
Rok


Ceny pamięci cyfrowych
Twarde dyski (HD), solid state (SSD), taśma magnetyczna (LTO)
Cena ($ / terbajt)
$1,000
$100
$10
HD
Exponential (HD)
LTO
Exponential (LTO)
SSD
Exponential (SSD)
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Rok

[redacted]