

**„Jakość danych w systemach informatycznych ZU”
Warszawa, 10 X 2011r.**

**"Problemy jakości w systemach
długotrwałego składowania zasobów
cyfrowych"**

**Bolesław SZAFRAŃSKI, prof. WAT
Wojskowa Akademia Techniczna**

"Jakość i dostępność rejestrów publicznych warunkiem rozwoju usług elektronicznych,, - (przypomnienie ...)

2010.kwiecień

→ **Interoperacyjność i referencyjność rejestrów**

2010.Październik

→ **jakość i dostępność rejestrów**

2011.Kwiecień (?)

→ **jakość dokumentów cyfrowych (NIE)**

2011.Październik

→ **jakość a długotrwałe przechowywanie**

**"Jakość i dostępność rejestrów publicznych warunkiem rozwoju powszechnych usług elektronicznych,, –
(przypomnienie z 2010r ...)**

„ ... powszechne długotrwałe (wieczyste) przechowywanie danych cyfrowych i związany z nim problem jakości takich zasobów” → nowe zagadnienie badawcze, koncepcyjne i projektowe (zgłoszone przez Autora (B. Szafrąńskiego) na seminarium w 2010 roku

Jakość i dostępność zasobów cyfrowych a ich **długotrwałe (wieczyste) składowanie:**

- nowe zagrożenia dla jakości,**
- brak gwarantowanych usług,**
- kwestie efektywności usług a gwałtownie rosnące potrzeby i wymagania**

1. Cele eAdministracji

- Dla każdego zadania publicznego realizowanego przez administrację możliwości pełnego lub częściowego zrealizowania (lub co najmniej zainicjowania) tego zadania drogą elektroniczną,
- „Odmiejscowiony” i realizowany zgodnie z zasadą „braku wykluczenia” dostęp do zadań realizowanych drogą elektroniczną ,
- „Jeden punkt elektronicznego kontaktu” dla usług realizowanych drogą elektroniczną,
- zakaz żądania danych, które już znajdują się w zasobach informacyjnych administracji, bez względu na istniejące podziały administracji.

Tło

2. Governance 2.0+ , reużycie informacji publicznych

- otwarta (informacyjnie) administracja, aktywny, partnerski i współuczestniczący obywatel,
- Możliwość powszechnej partycypacji obywateli i podmiotów w prawnie dozwolonych procesach informacyjnych państwa (zasada „reużywalności” informacji publicznych),
- ochrona obywatelskiego prawa do dostępu do danych

Demokracja → (prawo, dostępność /archiwa/, otwartość),

Architektura → (wydajność, elastyczność, interoperacyjność, reużywalność, koszty),

Jakość → (aktualność, poprawność, kompletność,)

[Źródło: Fragmenty rozmowy P. Waglowskiego ze Zbigniewem 'Gandalfem' Branieckim upublicznionej w całości na portalu VaGla.pl oraz Konferencja „Bezpieczeństwo w Internecie, UKSW, 2011]

3. Europejska Agenda Cyfrowa

- **podjęcie integralne, kompleksowe, infrastruktura informacyjna państwa, to znacznie więcej niż infrastruktura informacyjna dla procesów administracyjnych, ...**
- **wielodziedzinowość, bo państwo to nie tylko administracja:**
 - » **systemy sterowania/zarządzania (Smart Grid, Automated Metering Infrastructure): energetyka, zaopatrzenie w wodę, komunikacja (metro, kolej)**
 - » **System gospodarczy,**
 - » **Telekomunikacja,**
 - » **Zarządzanie kryzysowe,**
 - » **Ochrona zdrowia (telemedycyna),**
 - » **Ochrona środowiska (Inspire),**
 - » **Komunikacja społeczna**

Tło

4. Cyberprzestrzeń i cyberprzestępczość

- Cyberprzestępczość - czyny przestępcze dokonane przy użyciu sieci i systemów teleinformatycznych lub skierowane przeciwko tym sieciom i systemom,
- Cyberprzestrzeń, a wraz z nią nowy typ zagrożenia, czyli cyberprzestępczość, narodziła się w czasie kampanii prezydenckiej Billa Clintona, w 1992 roku,
- Cyberterrorizm (lata 80, „elektroniczny Pearl Harbour”)– przejęcie kontroli nad krytycznymi elementami infrastruktury państwa – energia, woda, telekomunikacja, transport,
- Cyberterrorizm - „ jest to celowy, motywowany politycznie atak przeciw informacji, systemom komputerowym, programom komputerowym i danym.

Wniosek: wszystkie elementy tła wskazują na nieuchronność gwałtowne wzrostu znaczenia jakości długotrwałego przechowywania zasobów cyfrowych dla budowy społeczeństwa informacyjnego (opartego na wiedzy)

„Podstawy dokumentu cyfrowego”

(wybrane tematy seminarium prowadzonego w Polskiej Wytwórni Papierów Wartościowych)

- **„Metody i urządzenia długoterminowego przechowywania danych ,część I. State-of-the-art.”- K.Marasek”, J.P.Walczak;**
- **„Metody i urządzenia długoterminowego przechowywania danych. Sztuka „Archiwizacji Wieczystej” – J.P.Walczak, K.Marasek:**
- **„Cechowanie dokumentów” - J.P. Walczak, K.Marasek;**
- **„Przegląd dostępnych zasobników do długoterminowego przechowywania danych o znacznych rozmiarach” - P.Pisarczyk,**
- **Cyfryzacja archiwalnych zbiorów filmowych w TVP S.A.” - T.Bujak,**
- **„Archiwa audiowizualne w praktyce, części I i II” – A.Kaźmierski:**
- **„Aspekty technologiczne cyfrowego archiwum wieczystego” – W.Wilk;**

„Podstawy dokumentu cyfrowego” **(wybrane tematy seminarium prowadzonego w** **Polskiej Wytwórni Papierów Wartościowych) – cd.**

- **„Aspekty techniczno – organizacyjne długoterminowego przechowywania danych cyfrowych” – W.Skurzak;**
- **"Charakterystyka potrzeb ochrony zdrowia w zakresie długotrwałego przechowywania cyfrowych zasobów danych medycznych. " - dr n. med. Leszek Sikorski, dr inż. Kajetan Wojsyk,**
- **„Długotrwałe przechowywanie zasobów cyfrowych w polskim systemie prawnym” – Kazimierz Schmidt;**
- **„Podstawowe cechy platformy integracyjnej z punktu widzenia systemu długotrwałego przechowywania danych” – M. Zapendowski;**

Polskie definicje ustawowe a europeizacja prawa

- Według Art. 2. 3) dyrektywy 2003/98/WE Parlamentu Europejskiego i Rady z 17 listopada 2003 r. w sprawie ponownego wykorzystywania informacji sektora publicznego,
- **"dokument" oznacza: a) jakąkolwiek treść niezależnie od zastosowanego nośnika (zapisaną na papierze lub zapisaną w formie elektronicznej lub zarejestrowaną w formie dźwiękowej, wizualnej albo audiowizualnej); b) każdą część tej treści.**
- "document" means: (a) any content whatever its medium (written on paper or stored in electronic form or as a sound, visual or audiovisual recording); (b) any part of such content.
- Konwencja Rady Europy w sprawie dostępu do dokumentów urzędowych Tromsø, 18.VI.2009 Council of Europe Convention on Access to Official Documents [Council of Europe Treaty Series - No. 205]
- Art. 1 B **"dokumenty urzędowe" oznacza wszystkie informacje utrwalone w dowolnej formie, sporządzone lub otrzymane i posiadane przez władze publiczne.**
- "official documents" means all information recorded in any form, drawn up or received and held by public authorities.

Polskie definicje ustawowe a europeizacja prawa

- Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne (Dz. U. Nr 64, poz. 565, z 2006 r., Nr 12, poz. 65, Nr 73, poz. 501, z 2008 r., Nr 127, poz. 817 oraz z 2009 r., Nr 157, poz. 1241):
- Stan na 1 marca 2010:
- **dokument elektroniczny - stanowiący odrębną całość znaczeniową zbiór danych uporządkowanych w określonej strukturze wewnętrznej i zapisany na informatycznym nośniku danych (art.3 pkt.2) definiowanym jako materiał lub urządzenie służące do zapisywania, przechowywania i odczytywania danych w postaci cyfrowej lub analogowej (art.3 pkt.1);**

[komentarze do aktualnej regulacji: G. Szpor, C. Martysz, K. Wojsyk. Ustawa o informatyzacji działalności podmiotów realizujących zadania publiczne. Komentarz. ABC z Wolters Kluwer, Warszawa 2007, s. 40 - 43; J. Janowski Elektroniczny obrót prawny Wolters Kluwer 2008, s 176 – 178]

Ustawa o zmianie ustawy o ogłaszaniu aktów normatywnych i niektórych aktów prawnych (z dn. 4 marca 2011, DU 2011 nr 117 poz. 676)

- Akty [...] ogłasza się w formie dokumentu elektronicznego
- Dzienniki urzędowe wydaje się w postaci elektronicznej
- Podstawą do ogłoszenia aktu [...] jest akt w formie dokumentu elektronicznego podpisany przez upoważniony do wydania aktu organ bezpiecznym podpisem elektronicznym [...]
- [...] wiza organu wydającego dziennik urzędowy [...] jest opatrzona bezpiecznym podpisem elektronicznym [...]

[J. Deminet – "Jak przekazywać, publikować i przechowywać przez 150 lat akty prawne", seminarium „Podstawy dokumentu cyfrowego” PWPWI

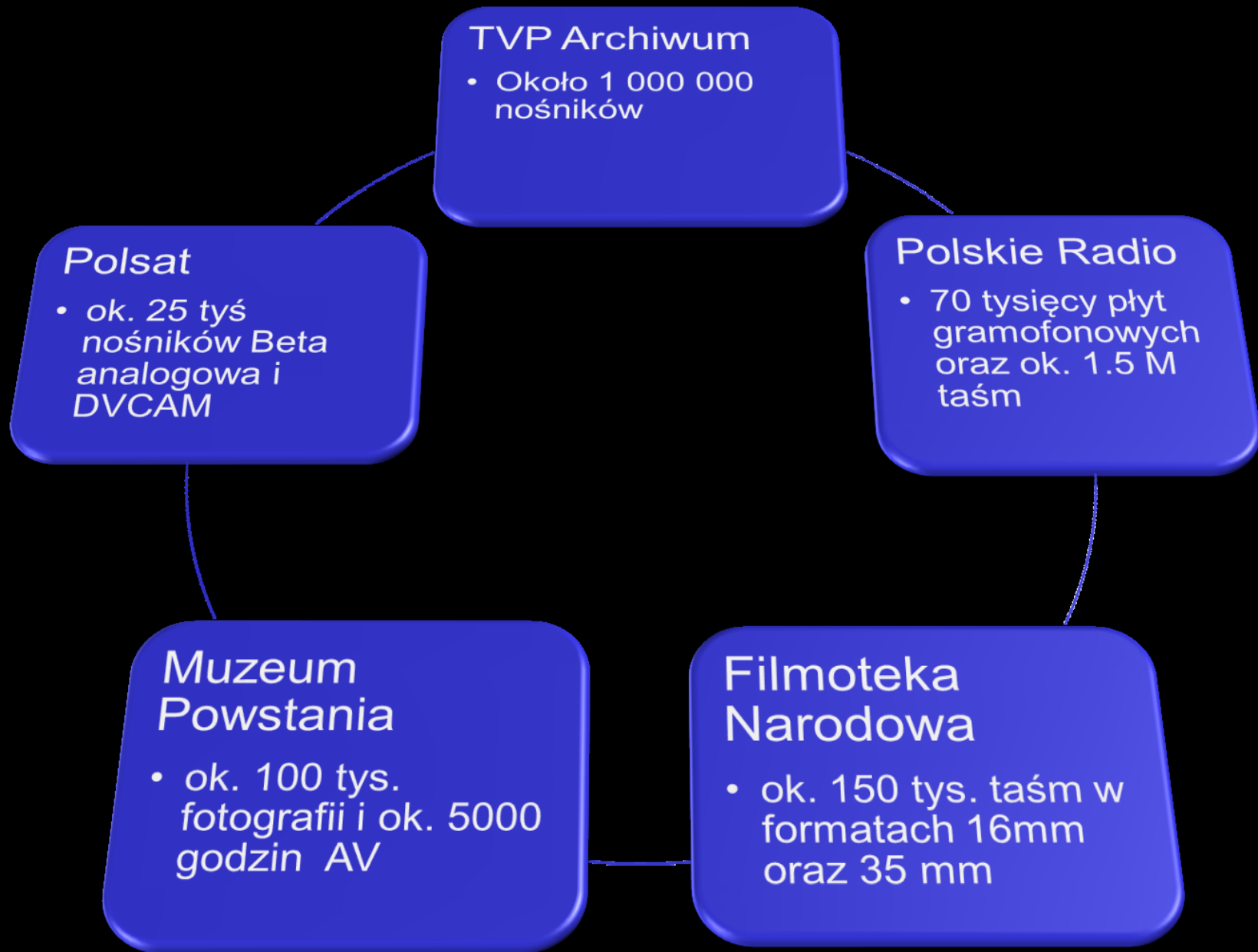
„Jakość danych w systemach informatycznych ZU”

Warszawa, 10 X 2011r.

Wnioski:

- _Pojęcie dokumentu przestało się wiązać się sztywno z jedną postacią dokumentu, bowiem zarówno po stronie potrzeb społeczeństwa informacyjnego (w tym wynikających ze współpracy międzynarodowej), jak i w realnie tworzonym prawie gwałtownie rozszerza się postrzeganie dokumentu jako bytu niematerialnego, elektronicznego, cyfrowego będącego ścisłym odpowiednikiem dokumentu tradycyjnego,**
- Powszechność stosowania dokumentów cyfrowych, masowa cyfryzacja zasobów należących do dziedzictwa kulturowego oraz innych rejestrowych i gospodarczych (np. w energetyce, w geodezji, w wymiarze sprawiedliwości, w sektorze finansowym, itd.) wskazują na pilną potrzebę powstania usług długotrwałego przechowywania zasobów cyfrowych oraz podjęcia badań dotyczących jakości świadczenia tych usług.**

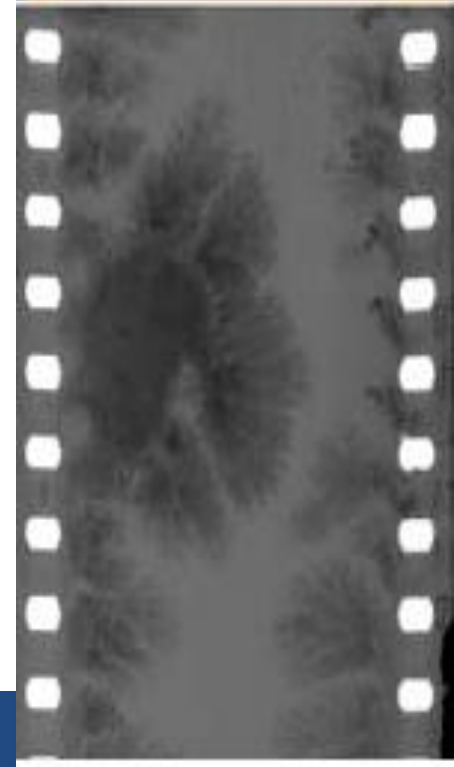
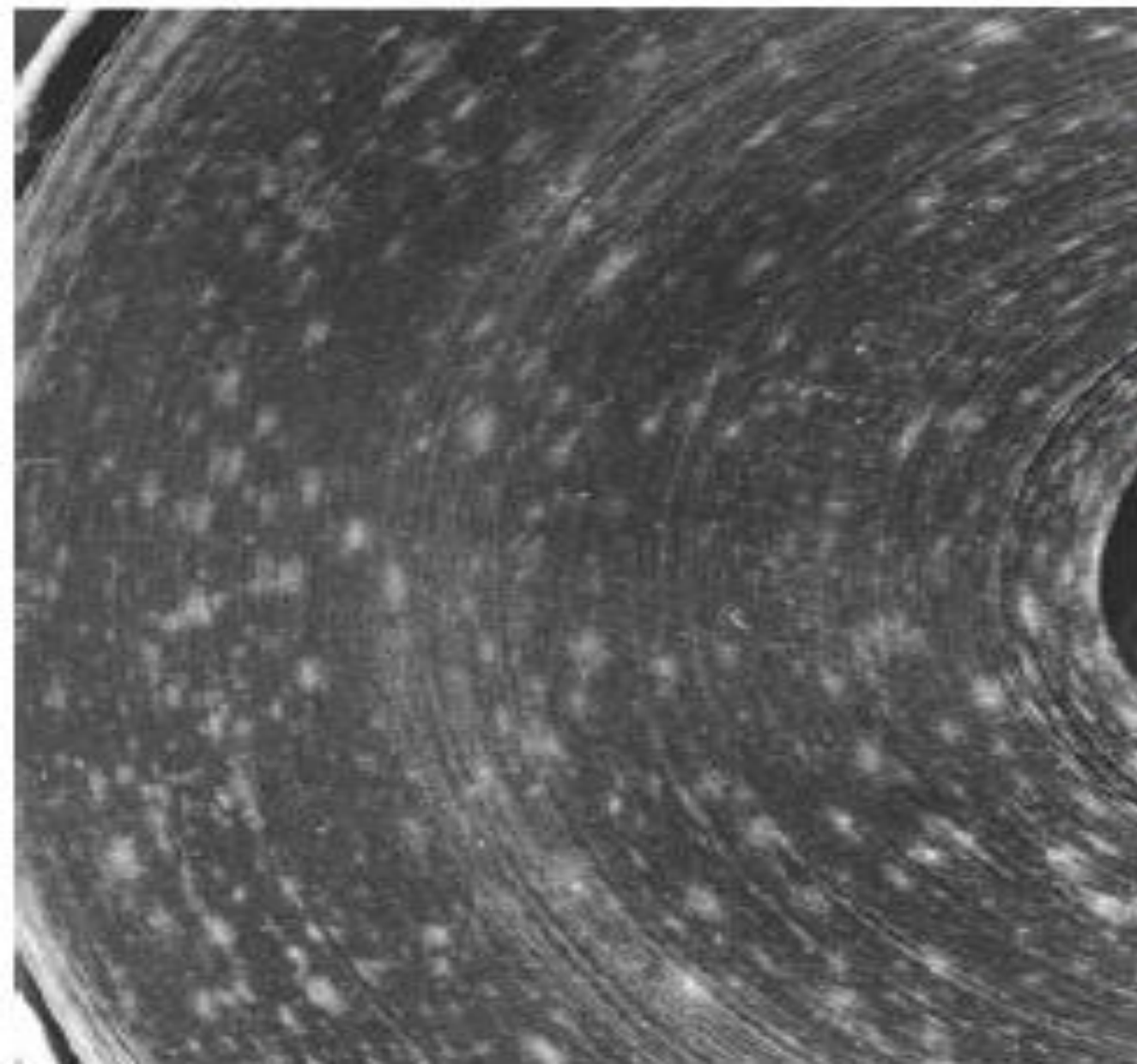
Wybrane archiwa multimedialne (źródło: K. Marasek – kwerenda własna, seminarium PWPW)



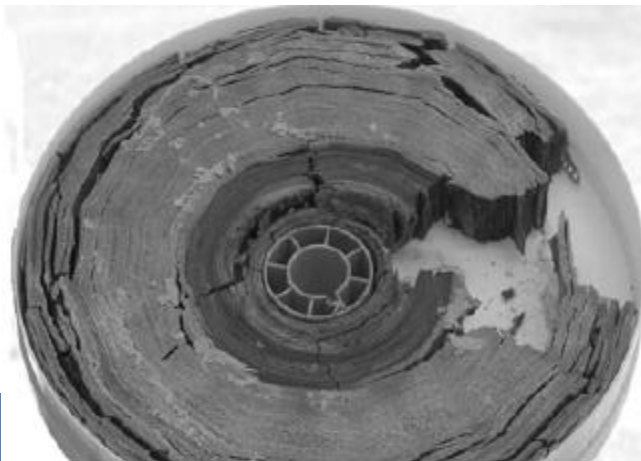
TVP

- *W magazynach Wideoteki znajduje się przeszło 300 tys. taśm magnetycznych z nagranych audycjami,*
 - *W Filmotece znajduje się 78 tys. pozycji katalogowych taśm światłoczułych, z czego 60.650 pozycji może zostać wykorzystanych dla potrzeb programowych TVP (8.476 godzin nagrań). Wśród tej dokumentacji są niezwykle atrakcyjne filmy, popularne seriale, teatry telewizji, dziecięce dobranocki, dokumenty filmowe,*
 - **Od początku lat 90. transfer audycji z taśm filmowych (światłoczułych) na kasety Beta Digital ...**
 - **Od 2006 nowoczesna linia digitalizacji ...**
 - **Najstarszy materiał pochodzi z 1954 roku.**
- [Źródło: „Cyfryzacja archiwalnych zbiorów filmowych”, prezentacja TVP SA, seminarium PWPW, dotyczy również slajdów nt. uszkodzeń zbiorów]*

USZKODZENIA BIOLOGICZNE (TVP)



USZKODZENIA CHEMICZNE (TVP)



dokument elektroniczny



Cyfrowy film kinowy



Telewizja i kino domowe

Audio
Syntetyczne

Obrazy syntetyczne
2D, 3D

oraz
dane pomiarowe i medyczne

Teksty tworzone elektronicznie
(XML)
oraz dokumenty cechowane

Fotografia cyfrowa
profesjonalna

Audio
CD/SACD/DVD
Formaty profesjonalne
384/24;192/24;96/24;DXD

Dane prywatne
(fotografia i filmy amatorskie,
dokumenty cechowane)

Wybrane problemy długotrwałego przechowywania danych

1. Baza pojęciowa

- W zależności od wartości danych, rodzaju danych i obowiązującego w tej mierze prawa lub innych uregulowań (wprowadzanych np. ze względów biznesowych), a także innych powodów obowiązują różne okresy przechowywania i udostępniania danych.
 - przechowanie tymczasowe (chwilowe)– tylko na okres ich przetwarzania (np. w pamięci operacyjnej komputera),
 - przechowywanie krótkoterminowe– na okres poniżej trzech lat (np. w systemach billingowych, Internecie),
 - przechowywanie średnioterminowe– na okres do 10 lat (np. dane podatkowe),

Wybrane problemy długotrwałego przechowywania danych

- przechowywanie długoterminowe – na okres powyżej 10 lat, ale z określonym końcowym terminem tego okresu (np. do 50 lat dane ubezpieczeń społecznych, nagrania audiowizualne rozpraw sądowych itp.),
- przechowywanie bezterminowe – na okres powyżej 10 lat, bez oznaczenia końcowego terminu tego okresu (np. dane historyczne, dane rządowe, nagrania audiowizualne, rodzinne archiwa osób prywatnych, itp.).
- Bezterminowe przechowywanie danych nazywane jest przechowywaniem wieczystym

- Rekomendacje NARA (National Archives and Records Administration) dotyczące migracji danych z nośnika na nośnik:
 - dyski magnetyczne: raz na 3 lata
 - taśmy magnetyczne: raz na 5 lat .

[Źródło: K. Marasek, „Metody i urządzenia długoterminowego przechowywania danych”, seminarium PWPW]

Problemy jakości

- Uszkodzenia: mechaniczne, biologiczne, chemiczne.
- 81% respondentów wskazuje, że już istnieje potrzeba przechowywania zasobów cyfrowych przez 50 lat i więcej (dotyczy to zarówno podmiotów gospodarczych, podmiotów publicznych, jak i potrzeb „archiwów” domowych → **internet a fotografia amatorska**)
- **Czy to jest możliwe przy zachowaniu wymaganych, b. różnych, poziomów jakościowych oraz formułowanej tezie o kryzysie systemów „archiwów” cyfrowych. Wcześniej bowiem wyrażano bezwzględne przekonanie o pewności archiwów cyfrowych. Okazało się w praktyce, że:**
 - rozwój inżynierii materiałowej (nośniki) nie nadążają za potrzebami,
 - rozwój metod diagnozy nie nadąża za potrzebami,
 - rozwój metod rekonstrukcji zapisów cyfrowych nie nadąża za potrzebami,
 - rozwój metod „automatycznego ostrzegania” utraty części zasobów nie nadąża za potrzebami,
 - rozwój modeli usług (prawo, organizacja, koszty) długotrwałego przechowywania zasobów cyfrowych niewystarczający

Problemy jakości

Rodzaje systemów archiwów cyfrowych:

- Podręczne (b. szybki i selektywny dostęp, praca grupowa, duża niezawodność),
- „Czasowe” (często traktowane jako pośrednie między podręcznym i głębokim, dostosowane do odbiorcy)
- Głębokie (trwałe i pewne przechowywanie, niedostępne dla działania operacyjnego, udostępnianie na zamówienie, możliwe przekodowywanie, możliwa zmiana nośnika, usługa tworzenia w buforze zasobu „czasowego” o różnych parametrach zapisu, ...)

Główny problem z punktu widzenia jakości: **gwarancja równoważności informacyjnej (po latach) przechowywanego materiału (zasobów cyfrowych) !!!!**

Problem związany – **energochłonność cyfryzacji i długotrwałego przechowywania zasobów cyfrowych**

Problemy jakości

Gwarancja tożsamości lub równoważności informacyjnej a degradacja nośnika (słabość narzędzi konwersji) – co to oznacza:

- **Czy - 100% wierności, 100% integralności wszystkich cech zasobu?**
- **Czy – 100% treści zasobu?**
- **Czy– x% wierności, integralności?**
- **Czy – x% treści zasobu lub elementów zasobu?**

Od czego to zależy:

- **od rodzaju zasobów: np. zasoby faktograficzne (faktury, testamenty, projekty), zasoby formalne (traktaty, ustawy, akta sądowe), zasoby dóbr kultury (filmy, przedstawienia, obrazy, książki), zasoby prywatne**
- **od przeznaczenia:**
 - **Proces dowodowy,**
 - **Proces badawczy,**
 - **Proces informacyjny (poznawczy)**

Problemy jakości

Kilka przykładów:

- Brak integralności w fakturze: może w liczbach, może w tekście, może w grafice faktury → **czy to ma znaczenie?**
- Brak integralności w filmie: jeden czy wiele bajtów, wiele bajtów sąsiednich czy rozproszonych, w zasobie czy w metadanych, itd. → **czy to znaczenie?**

Problemy jakości usług

Umowa SLA:

- Badanie tożsamości zasobów (metody oparte na podpisie cyfrowym, metody oparte na cechowaniu naturalnym),
- Alokacja kosztów – zjawisko degradacji zapisów istnieje obiektywnie (**czy ktoś poszedł do więzienia za niską jakość nośników z Kabaretem Starszych Panów? Nie nasza wina, to technologia nie nadążyła**), kto poniesie koszty ich rekonstrukcji i ryzyko braku sukcesu, jak to wycenić, jak to wyrazić, jak zagwarantować wymagane parametry jakościowe, np. po 30 latach (a może punkty kontrolne) → **problem buldoga,**
- **Rozwiązania prawne w klinczu między możliwościami technologii a potrzebami,**
- Bezpieczeństwo – długi okres, więc duże możliwości przygotowania i przeprowadzenia ataku; kto powinien ponieść koszty ochrony, jaki model kosztowy zastosować

Usługi dodane a problem jakości

Usługi dodane systemu archiwów głębokich (przykłady):

- usługi digitalizacji,
- usługi katalogu metadanych,
- usługi udostępniania zasobów zgodnie z zadanymi parametrami (standardami),
- usługi diagnozy i rekonstrukcji (dostarczonego i posiadanego materiału),
- usługi wzorca (np. dla renowacji i weryfikacji jakości renowacji obrazu, podwójna renowacja: cyfrowa i analogowa),
- usługi wsparcia procesów rekonstrukcji analogowych,
- ...

Paradoksy – pytania zawsze aktualne

Rozdzielczość zdjęcia

a

Możliwości poczty elektronicznej

Kolorowy dokument

a

Czarnobiałe urządzenie faxowe.

Czy tędy droga?