

**„Jakość danych w systemach informatycznych ZU”
Warszawa, 19 IV 2012r.**

**„Klasyfikacja systemów i usług
długotrwałego przechowywania zasobów
cyfrowych – problemy jakości”**

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

„Jakość danych w systemach informatycznych ZU” - (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 danych w systemach informatycznych ZU” – (przypomnienie z 2011r. ...)

Nowe zagadnienie – powszechne długotrwałe (wieczyste) przechowywanie danych cyfrowych

Nowe zagrożenia - ciągłość utrzymania jakości, dostępności i bezpieczeństwa zasobów cyfrowych podlegających długotrwałemu (wieczystemu) przechowywaniu danych.

„Jakość danych w systemach informatycznych ZU”

„ –(przypomnienie z 2011r. ...)

- 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)
- **Kryzys „nadmiernych oczekiwań w zakresie jakości systemów „archiwów” cyfrowych**, (prywatnie tego też doświadczyliśmy kopiując zdjęcia lub filmy z taśmy filmowej na nośniki cyfrowe)?
 - rozwój inżynierii materiałowej (nośniki) nie nadąża za potrzebami,
 - rozwój metod diagnozy nie nadąża za potrzebami,
 - rozwój metod rekonstrukcji nie nadąża za potrzebami,
 - rozwój metod „ostrzegania” o degradacji zapisów (nośników) nie nadąża za potrzebami,
 - rozwój modeli usług (prawo, organizacja, koszty) długotrwałego przechowywania zasobów cyfrowych jest niewystarczający

Przypomnienie z 2011r. - Tło

1. Cele eAdministracji

2. Governance 2.0+ , reużycie informacji publicznych

- otwarta (informacyjnie) administracja, aktywny, partnerski i współuczestniczący obywatel, powszechna partycypacja obywateli i podmiotów w prawnie dozwolonych procesach informacyjnych państwa,

3. Europejska Agenda Cyfrowa

- podejście integralne, kompleksowe, infrastruktura informacyjna państwa, to znacznie więcej niż infrastruktura informacyjna administracji: modele korzyści różne ale wiele problemów wspólnych

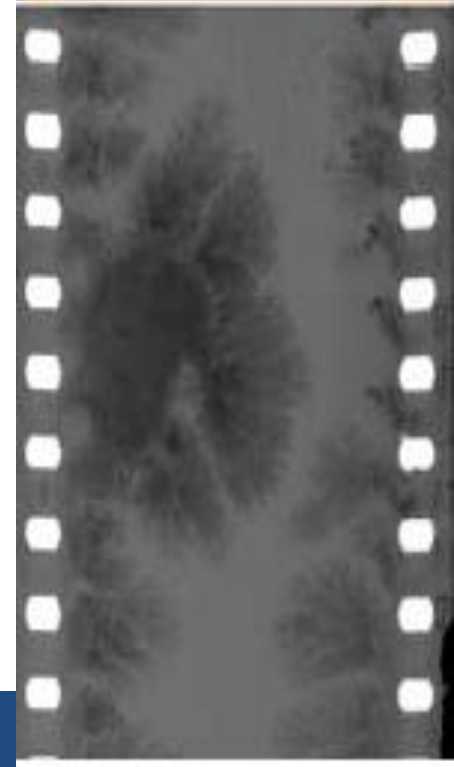
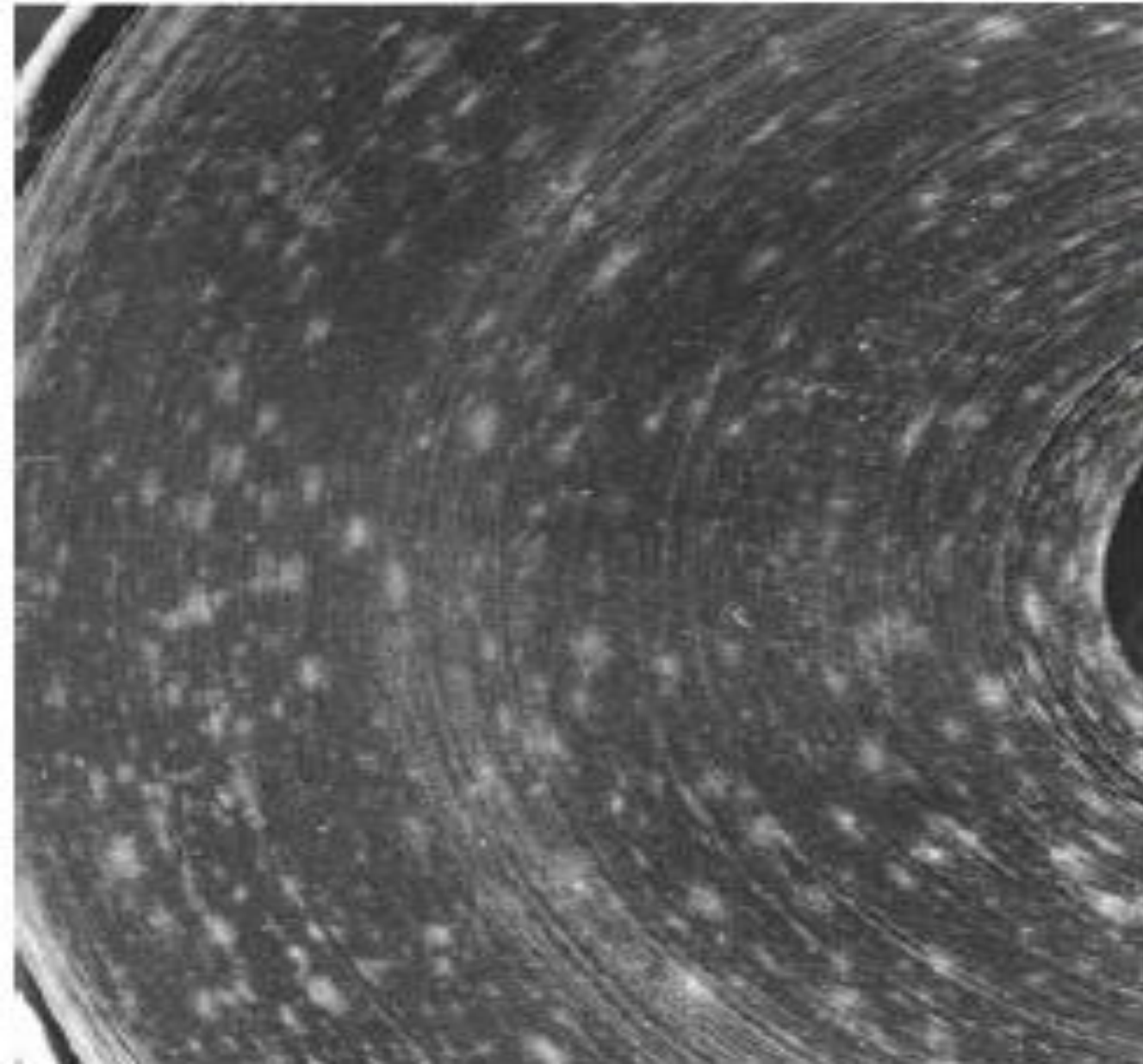
4. Cyberprzestępczość i cyberterroryzm

- bezprawne zamachy lub ich groźby w celu zastraszenia/ zmuszenia władz do spełniania określonych żądań politycznych lub społecznych.

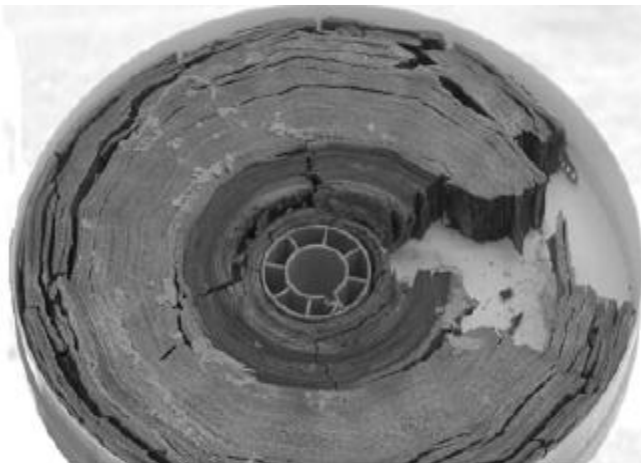
Przypomnienie z 2011r - ePUAP

- Zgodnie ze znowelizowaną ustawą o informatyzacji działalności podmiotów realizujących zadania publiczne ePUAP, czyli elektroniczna platforma usług administracji publicznej jest systemem teleinformatycznym, w którym instytucje publiczne udostępniają usługi przez pojedynczy punkt dostępowy w sieci Internet. ePUAP w swej istocie jest przedsięwzięciem, którego najważniejszymi celami są:
 - stworzenie jednolitego, bezpiecznego i w pełni zgodnego z obowiązującym prawem cyfrowego kanału udostępniania usług publicznych obywatelom, przedsiębiorcom i podmiotom publicznym,
 - zapewnienie interoperacyjnego dostępu do infrastruktury informacyjnej administracji publicznej (przede wszystkim rządowej) wszystkim podmiotom wytwarzającym i udostępniającym usługi publiczne drogą elektroniczną (*a więc i do przechowywanych długotrwale*

USZKODZENIA BIOLOGICZNE – (źródło TVP)



USZKODZENIA CHEMICZNE – (źródło TVP)



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 (projekt przyjęty przez Radę Ministrów)

- 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 [...]

[J. Deminet – "Jak przekazywać, publikować i przechowywać przez 150 lat akty prawne", seminarium PWPW]

**ZALECENIA KOMISJI Z DNIA 24 SIERPNIA 2006 R. W SPRAWIE
DIGITALIZACJI I UDOSTĘPNIENIA W INTERNECIE DOROBKU
KULTUROWEGO ORAZ W SPRAWIE OCHRONY ZASOBÓW CYFROWYCH
(2006/585/WE).**

Zaleca się, aby państwa członkowskie UE między innymi:

- wspierały partnerstwo między instytucjami kulturalnymi a sektorem prywatnym w celu tworzenia nowych sposobów finansowania cyfryzacji dorobku kulturowego (zalecenie 3);**
- tworzyły duże ośrodki cyfryzacji i zapewniały ich funkcjonowanie jako elementów ośrodków kompetencji na rzecz rozwoju digitalizacji w Europie lub w ścisłej współpracy z nimi (zalecenie 4);**
- zapewniły stosowanie wspólnych standardów digitalizacji przez instytucje kulturalne i – jeśli dotyczy – podmioty prywatne w celu zagwarantowania interoperacyjności zasobów cyfrowych na szczeblu europejskim oraz ułatwienia wyszukiwania w różnych językach (zalecenie 5);**

Klasyfikacja zasobów cyfrowych

1. Ze względu na rodzaj zasobów

- Zasoby kultury narodowej (dorobek kulturowy),
- Zasoby gospodarcze,
- Zasoby administracji publicznej,
- Zasoby prywatne.

[Komentarz: zasoby faktograficzne (faktury, testamenty, projekty), zasoby formalne (traktaty, ustawy, akta sądowe), zasoby dóbr kultury (filmy, przedstawienia, obrazy, książki), zasoby prywatne, zasoby ochrony zdrowia (ciągi znaków, filmy, wykresy, zdjęcia, dźwięk; narady medyczne, dokumenty związane → zdjęcie-opis, przebieg operacji, ...]

2. Ze względu na przeznaczenie

- Proces dowodowy,
- Proces badawczy,
- Proces informacyjny (poznawczy), ...

Klasyfikacja systemów długotrwałego przechowywania zasobów cyfrowych

3. Ze względu na okres przechowywania (przykłady)
- 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),
 - przechowywanie długoterminowe– na okres powyżej 10 lat, ale z określonym końcowym terminem tego okresu (np. do 50 lat dane ubezpieczeń społecznych, nagrania rozpraw sądowych itp.),
 - przechowywanie bezterminowe (wieczyste) – na okres powyżej 10 lat, bez oznaczenia końcowego terminu tego okresu (np. dane historyczne, dane rządowe, nagrania audiowizualne, rodzinne archiwa osób prywatnych, itp.).

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

Klasyfikacja systemów długotrwałego przechowywania zasobów cyfrowych

4. Ze względu na sposób dostępu

- **Systemy działania operacyjnego (systemy podręczne)**
- **Systemy działania analitycznego (systemy buforowe)**
- **Systemy działania archiwalnego (systemy głębokiego składowania)**

Cechy systemów długotrwałego przechowywania zasobów cyfrowych a problemy jakości

1. Wielopoziomość i rozproszenie systemów przechowywania zasobów cyfrowych
 - podręczne, buforowe, głębokie
2. Interoperacyjność systemów przechowywania zasobów cyfrowych
 - przekazywanie zasobów między systemami na różnych poziomach,
 - kompatybilność typów i usług (tabele transformacji formalnej i semantycznej)
3. Różne zbiory usług
 - systemy podręczne - *usługi b. szybkiego i selektywnego dostępu i prezentacji (emisji), usługi pracy grupowej (identyfikacja aktorów),*
 - systemy buforowe – *usługi analizy (opisu) treści i usługi „transformacji” postaci i jakości,*
 - systemy głębokiego składowania - *usługi trwałego i niskokosztowego (energooszczędne) przechowywania, zasoby niedostępne dla bezpośredniego działania operacyjnego, usługi udostępniania na zamówienie, usługi transformacji do potrzeb systemów buforowych (zmiany parametrów i sposobu kodowania), usługi „uodporniające” na zmiany nośników*

Cechy systemów długotrwałego przechowywania zasobów cyfrowych a problemy jakości

- 4. Bezpieczeństwo – różne modele ochrony dla różnych rodzajów systemów (zwiększona rola integralności, znaczenie długości okresu przechowywania)**
- 5. Koszty – tradycyjne sposoby wyceny kosztów i analizy ryzyk nie są odpowiednie (stosowane modele nie zostały jeszcze zweryfikowane, zbyt krótki okres czasu. Jak ująć w kosztach technologie przyszłości?)**
- 6. Umowy – różne modele dla różnych systemów, brak doświadczeń**
 - metody oceny (pomiaru i interpretacji)**
 - tożsamości (wierności)**
 - czy**
 - równoważności informacyjnej zasobów cyfrowych**
 - kwestia specyfikacji punktów odniesienia i punktów kontrolnych, problem buldoga**
- 7. Autentyczność (zgodność z deklarowaną zawartością, oryginał a kopia), integralność zapisu (nienaruszalność i kompletność), integralność informacyjna – nie ma skutecznych, zweryfikowanych modeli postępowania**
- 8. Treści wrażliwe – jak je opisać i rozpoznawać (zdjęcie medyczne, ...)**

Klasyfikacja usług

- 1. Usługi cyfryzacji zasobów**
- 2. Usługi opisu i zarządzania zasobami (katalogu metadanych, w tym specyfikacja kryteriów selekcji, specyfikacja parametrów jakości, bezpieczeństwa, wydajności, interoperacyjności)**
 - wewnętrzne,
 - zewnętrzne (udostępniane zewnętrznemu dysponentowi zasobów).
- 3. Usługi manipulowania zasobami (dostępu do zasobów) – ziarnistość treściowa zasobu a usługi selekcji**
 - pośredniego,
 - bezpośredniego (udostępniane zewnętrznemu dysponentowi zasobów).
- 4. Usługi transformacji (przekształcania) zasobów**

3. Usługi oznaczania i pomiaru jakości

Jakość zapisu rozpatrywana jest zwykle (źródło: K. Marasek, J. Walczak) w dwóch aspektach: strumienia danych binarnych (bit preservation) oraz zapewnienia użyteczności danych (long-term preservation).

- Na poziomie strumienia danych zachowanie jakości winno być interpretowane jako pewność uzyskania po zadanym czasie dokładnie tego samego strumienia bajtów (bitów), który został tam zapisany.
- Na poziomie struktury zachowanie jakości powinno być rozumiane jako pewność odtworzenia dokładnie tej samej informacji

5. Usługi wzorca (np. dla renowacji i weryfikacji jakości analogowych zasobów oryginalnych)

6. Usługi diagnozy stanu

Klasyfikacja usług

7. Usługi wsparcia rekonstrukcji cyfrowej i analogowej
8. Usługi pielęgnacji zasobów
9. Usługi statystyczne
10. Usługi spedycyjne