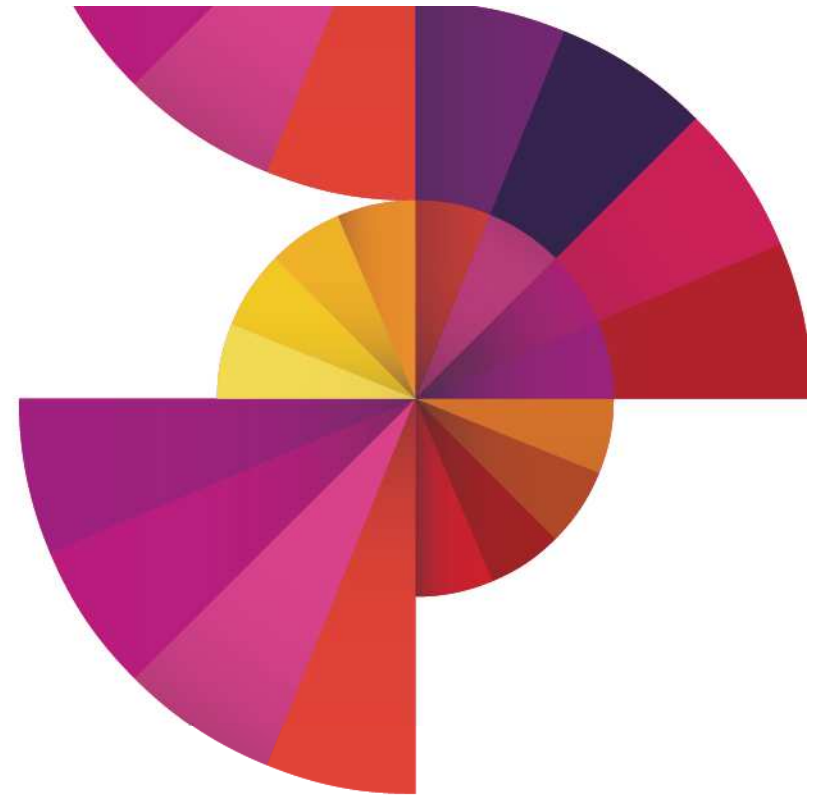




Inteligentne Przetwarzanie



Konferencja

Inteligentne Technologie w Praktyce

18 września 2013 | Hotel Sofitel Victoria





Inteligentne Przetwarzanie

Konferencja

Inteligentne Technologie w Praktyce



Długoterminowe przechowywanie danych cyfrowych

Zdjęcia cyfrowe za lat 50+

Piotr Beńke

Manager of STG Client Architects



Master
Certified IT Architect



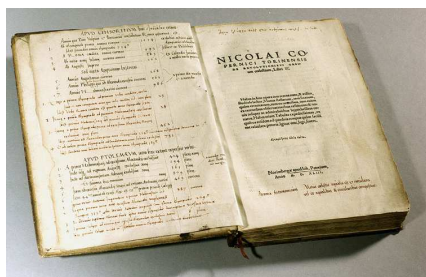


Wyzwania dla długoterminowego przechowywania danych

*Skrypt: ~2400 BC.
Nośnik: Tablica gliniana.
Język: sumeryjski.*



*Zwoje znad Morza Martwego, ~70AD.
Nośnik: Miedź.
Język: hebrajski.*



*„O obrotach ciał niebieskich”, XVI wiek.
Nośnik: papier.
Język: łaciński.*



Czy informacja utworzona kilka lat temu będzie dostępna, interpretowalna lub możliwa do zaprezentowania za 20, 50 a może 100 lat?



Paradoks

W czasach, kiedy świat staje się coraz bardziej **cyfrowy** nasze dane są coraz bardziej **zagrożone**

Nasze zdolności do **gromadzenia informacji w postaci cyfrowej** wzrastają przy **jednoczesnym obniżaniu zdolności do jej długoterminowego przechowywania**

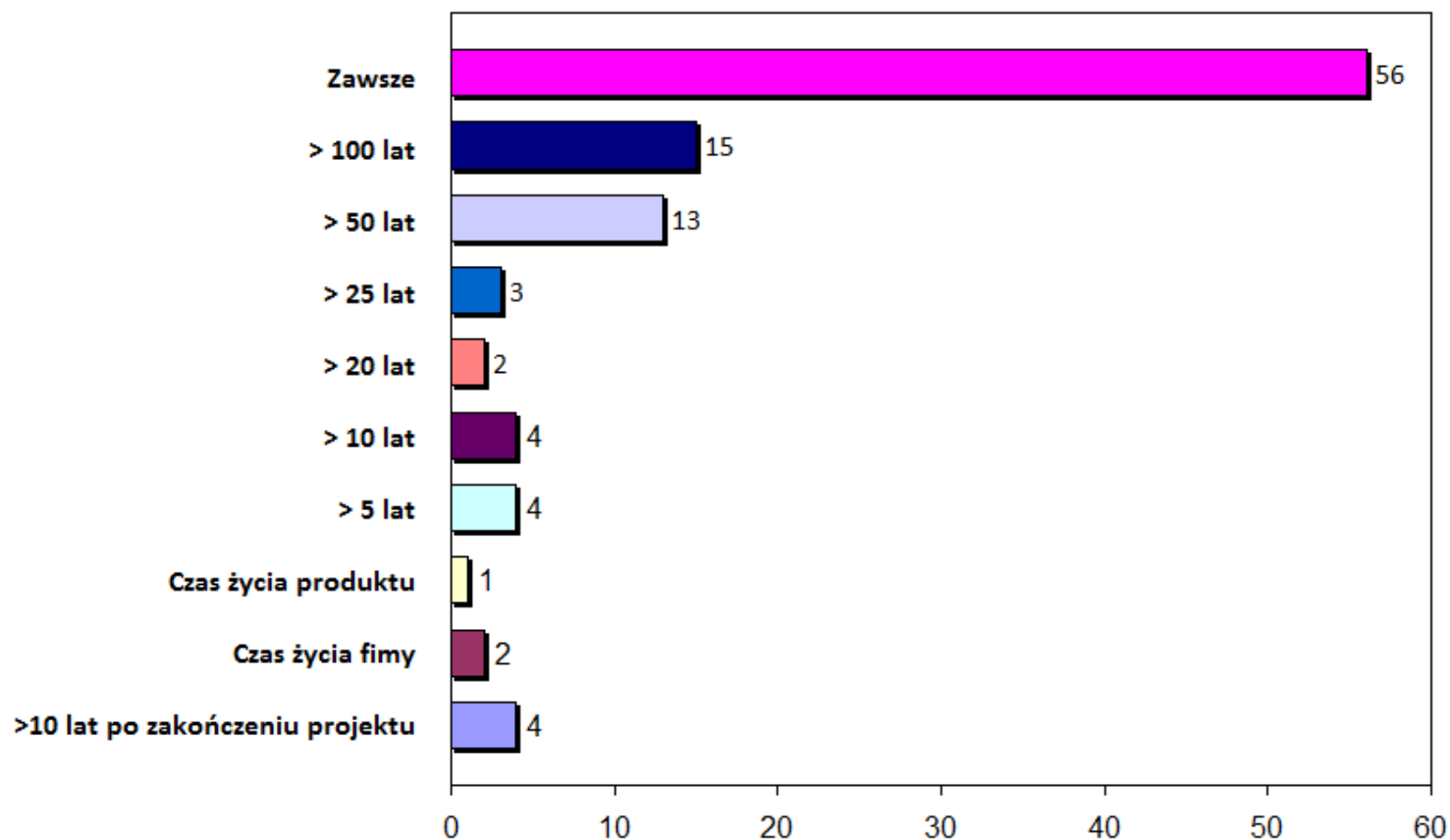


Długoterminowe przechowywanie danych ?

- Dłużej niż 10 lat,
- Dłużej niż czas „życia” nośnika/technologii,
- Dłużej niż czas „życia” firmy,
- Dłużej niż czas życia formatu zapisu informacji,
- Bezterminowo.



Wymagania dla długoterminowego przechowywania danych

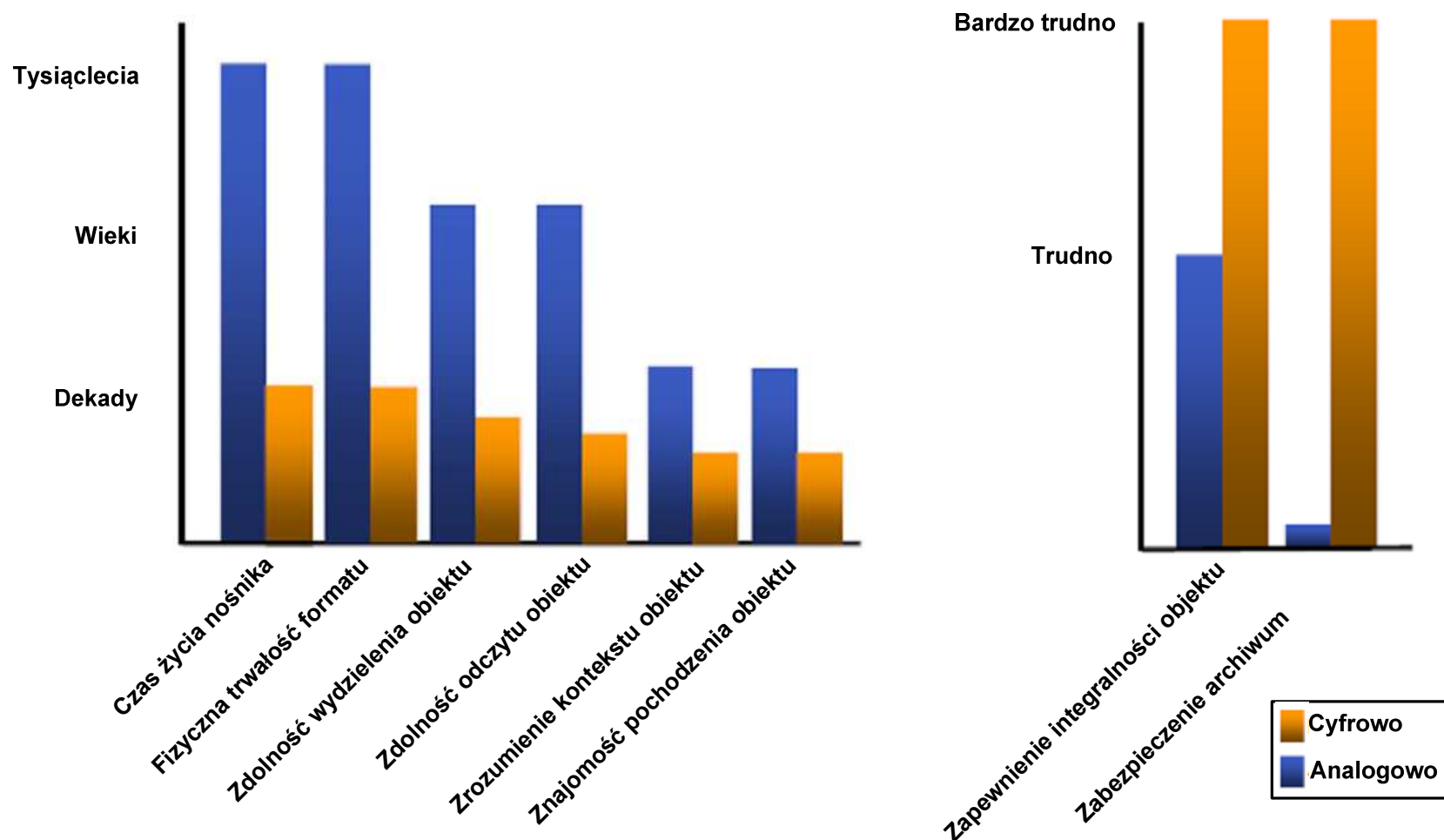


Źródło : IBM

Liczba respondentów : 104



Analogowe vs. Cyfrowe przechowywanie informacji





Dlaczego cyfrowy format danych jest potrzebny

- **Za**
 - Szybkie wyszukiwanie informacji
 - Łatwy dostęp
 - Ta sama jakość przez lata (starzenie się jakości obiektu wraz z nośnikiem)
 - Mniej fizycznej przestrzeni na nośniki (20 000 stron na 1 CD)
 - Dane powstające w formie cyfrowej (zdjęcia, muzyka, maile, dokumenty, itp.)
- **Przeciw**
 - Autentyczność informacji
 - Koszt zabezpieczenia informacji



Jaka jest przyszłość zdjęć cyfrowych (i nie tylko) zrobionych w ostatnich latach

- Powszechność aparatów fotograficznych
- Dostępność serwisów internetowych oferujących składowanie danych
 - Czy zauważysz brak twoich zdjęć z przed kilku/kilkunastu lat?
 - Czy utrata zdjęć będzie dotkliwa?
- Ile jesteśmy gotowi zapłacić za przechowywanie zdjęć cyfrowych?

Zabezpieczenie na poziomie bitów



Bit preservation – zdolność do zabezpieczenia informacji na poziomie bitów w obliczu degradacji/starzenia się nośników/technologii, katastrof środowiskowych typu pożar, powódź itp.

- Żywotność nośników : dysków 3-5 lat, taśm 5-10 lat, CD/DVD : 10-20 lat.
- Istnieją pewne klasy rozwiązań realizujące zabezpieczenia na poziomie bitów jak usługi kopiujące, odświeżanie technologii, korekcja błędów.

Fakty z życia

- Dyskietki 5.25” były bardzo popularne 15 lat temu ale dzisiaj jest trudno znaleźć komputer z napędem dla tych dyskietek.
- NASA straciła wiele danych z misji na Marsa z powodu zbyt późnego odświeżenia taśm z danymi.



Zabezpieczenie logiczne

Logical preservation – ochrona „rozumialności” i użyteczności danych w przyszłości

- Obecne technologie sprzętowe i software’owe mogą nie istnieć w przyszłości. Pojawią się nowe technologie
- Czy obecnie tworzone dane będą interpretowane i prezentowane we właściwy sposób w przyszłości?
- Co z autentycznością informacji i z autoryzowanym dostępem do niej?

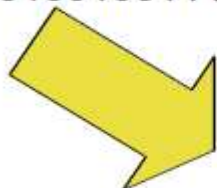
Fakty z życia

- 20 lat temu WordPerfect był standardem dla przetwarzania tekstu. Czy dzisiejsze procesory tekstu prezentują dokumenty utworzone w tamtych czasach w ten sam sposób.
- Inne ułożenie tekstu może zmienić całkowicie wartość informacji niesionej przez tekst.



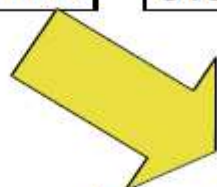
Interpretacja danych

01000010010011110100000101010100 **Strumień bitów**



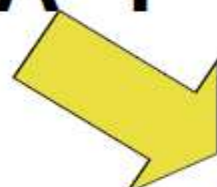
Klastrowanie bitów w bajty

01000010 01001111 01000001 01010100



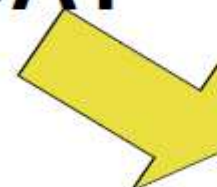
Mapowanie bajtów w znaki

B O A T



słowo w języku angielskim

BOAT



Obiekt fizyczny





Renderowanie danych

I	II	IIIb	IVb	Vb	VIb	VIIb	VIIIb	Ib	IIb	III	IV	V	VI	VII	0		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
H																	He
Li	Be										B	C	N	O	F		Ne
Na	Mg										Al	Si	P	S	Cl		Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La*	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac**	Rf	Db	Sg	Bh	Hs	Mt	110	111	112	113					
Lanthanides *				Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Actinides **				Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

I II IIIb IVb Vb VIb VIIb VIIIb Ib IIb III IV V VI VII 0

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

H He
 Li Be B C N O F Ne
 Na Mg Al Si P S Cl Ar
 K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr
 Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe
 Cs Ba La* Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn
 Fr Ra Ac** Rf Db Sg Bh Hs Mt 110 111 112 113

Lanthanides * Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu

Actinides ** Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr



Zabezpieczenie danych – podejście typu muzeum

Utrzymywanie urządzeń składających i renderujących dane

Za

- Brak utraty danych

Przeciw

- Koszty
- Ograniczenia czasowe
- Brak skalowalności
- Gwarancja + części zamienne



Zabezpieczenie danych – emulacja



Emulacja pierwotnego urządzenia renderującego

Za

- redukcja problemu do zabezpieczenia platformy emulacyjnej
- Koszt proporcjonalny do ilości renderowanych formatów

Przeciw

- Duża inwestycja początkowa
- Zastosowanie dla danych związanych z oprogramowaniem
- Nie pozwala na nowe interpretacje posiadanych danych





Zabezpieczenie danych – migracja na nowe nośniki

Migracja danych ze starzejących się nośników/technologii do nowych

Za

- Mniejsza inwestycja w przypadku przetworzonych danych
- Pozwala na nowe zastosowanie danych



Przeciw

- Może wprowadzić zniekształcenia
- Koszt proporcjonalny do rozmiaru danych
- Nieustanny koszt



Zabezpieczenie danych – podejście opisowe

Składowanie danych wraz z metadanymi opisującymi ich reprezentację

Za

- Brak utraty informacji
- Minimalne założenia dotyczące przyszłości
- Opóźnienie kosztów do momentu zapotrzebowania

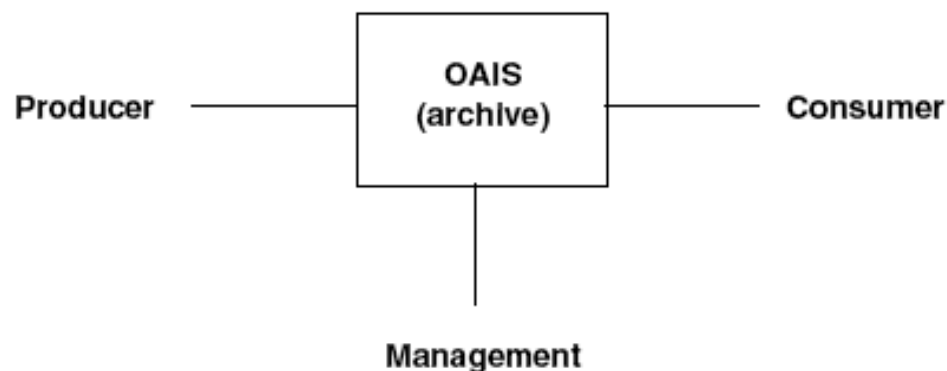
Przeciw

- Ograniczone wsparcie dla formatów własnych
- Możliwe duże koszty w przyszłości



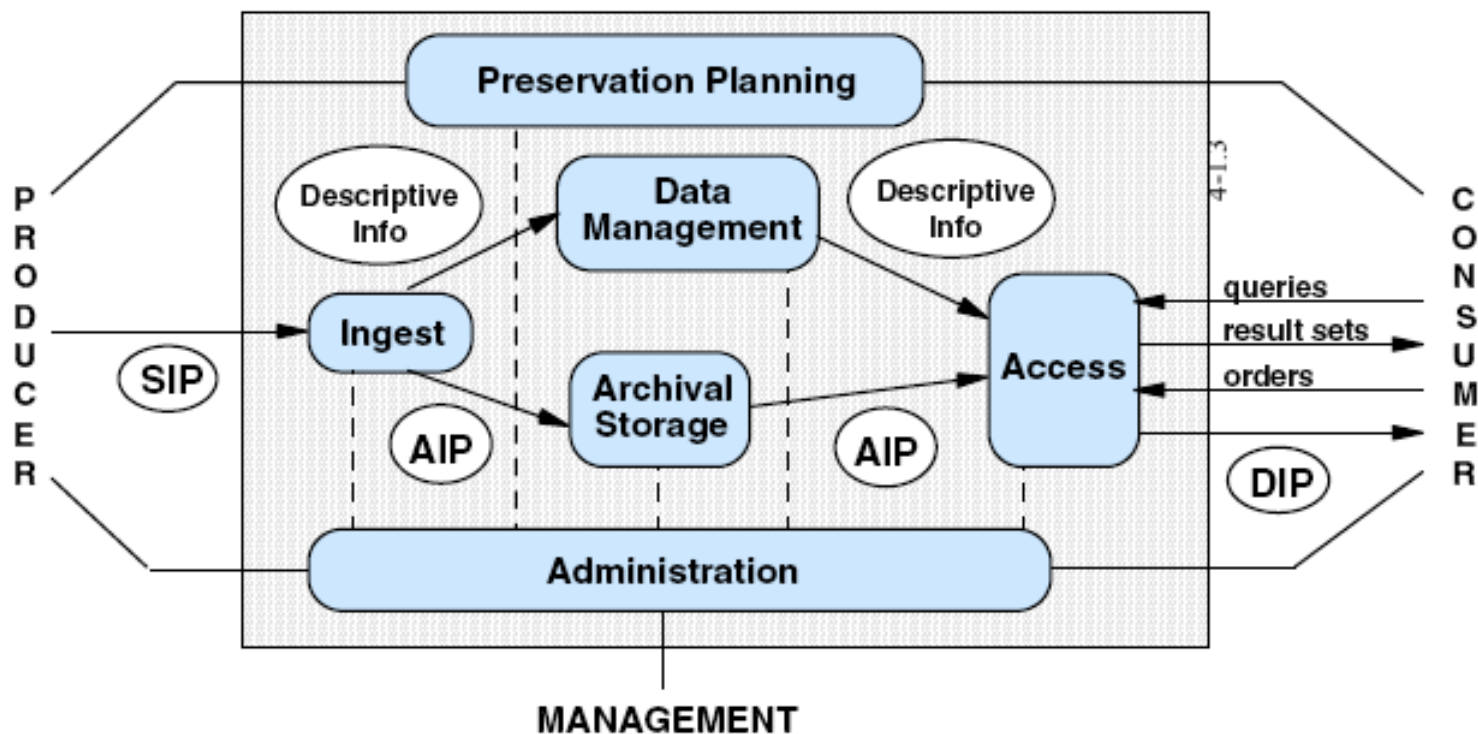
Open Archival Information System - OAIS

- Model archiwum rozwinięty przez Consultative Committee for Space Data Systems (CCSDS)
- ISO:14721:2002
- Dostarcza koncepcje i model odniesienia w archiwach długookresowych





OAIS – model funkcjonalny



- Submission Information Package (SIP)
- Archival Information Package (AIP)
- Dissemination Information Package (DIP)



Skalowalność archiwum cyfrowego

- Archiwa rosną/rozwijają się w sposób organiczny
- Archiwum powinno funkcjonować bez względu na ilość zgromadzonych danych
- Możliwość zarządzania dziesiątkami/setkami urzędzeń
- Standaryzacja dostępu do archiwum



Ekonomiczne aspekty archiwum długoterminowego

- Nieustanne koszty składowania danych cyfrowych
 - Cykl odświeżania dla formatów danych i nośników.
 - Zarządzanie kosztami pamięci masowych, monitoring technologii i standardów.
 - Wykorzystanie przestrzeni archiwum spada dramatycznie wraz ze starzeniem się danych.
 - Kompromis pomiędzy kosztami początkowymi budowy archiwum, a kosztami jego utrzymania.



Podsumowanie



Inteligentne Przetwarzanie

Konferencja

**Inteligentne Technologie
w Praktyce**



Dziękuję

Piotr Beńke

piotr.benke@pl.ibm.com

