

PLATFORMA OBSŁUGI NAUKI **PLATON**

www.platon.pionier.net.pl



Usługa archiwizacji danych w PLATON U4

Siódme Warsztaty „Biblioteki Cyfrowe”

Maciej Brzeźniak, PCSS



**INNOWACYJNA
GOSPODARKA**
NARODOWA STRATEGIA SPÓJNOŚCI



UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO



Projekt nr. POIG.02.03.00-00-028/08

DOTACJE NA INNOWACJE

Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego

Plan prezentacji

- Wprowadzenie
- Zagrożenia i wyzwania związane z archiwizacją danych
- Rozwiązanie - outsourcing do PLATON-U4
- **Usługi PLATON-U4 (Usługa powszechnej archiwizacji)**
 - Cechy i funkcje
 - Interfejsy użytkownika
 - Status projektu
- **Konsorcjum PLATON-U4:**



Wprowadzenie

Dane = wartość = problem

- Dane są bardzo wartościowe:
 - Bezcenne zbiory cyfrowe
 - Wyniki kosztownej digitalizacji oryginałów



=



- Producenci danych:
 - Biblioteki cyfrowe:
 - 80-100 TB rocznie / bibliotekę



- W skali kraju - PB rocznie?

Dane = wartość = problem

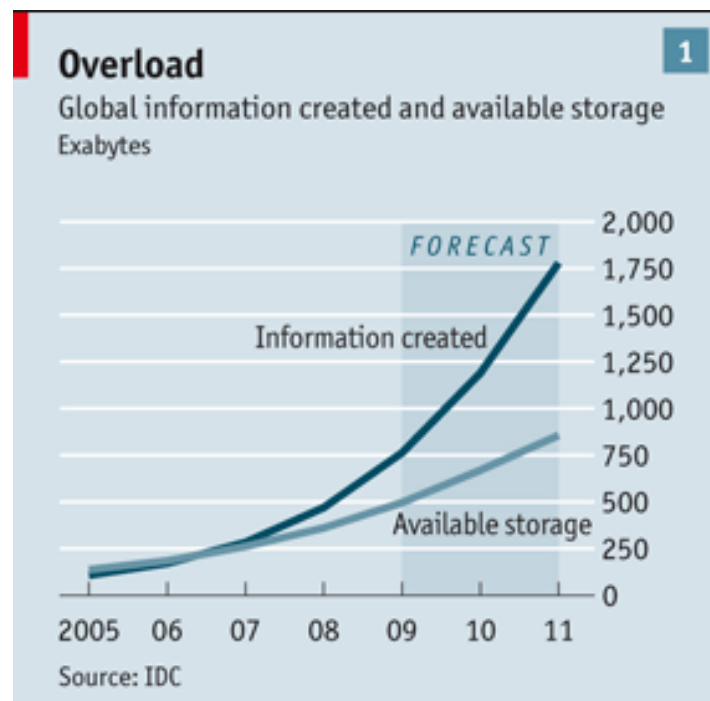
- Dane są bardzo wartościowe:
 - Bezcenne zbiory cyfrowe
 - Wyniki kosztownej digitalizacji oryginałów



=



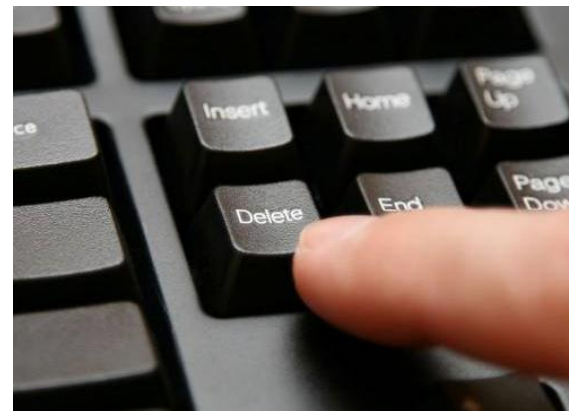
- Producenci danych:
 - W skali świata:



Zagrożenia i wyzwania

Zabezpieczenie danych – zagrożenia i wyzwania (1)

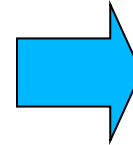
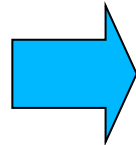
- Bezpieczeństwo fizyczne danych:



=> Warto przechowywać dane w co najmniej 1 dodatkowej lokalizacji...

Zabezpieczenie danych – zagrożenia i wyzwania (2)

- Integralność logiczna i poufność danych w archiwum
 - Problem 1 – integralność logiczna



-
- The diagram illustrates a data flow process. It begins with a user (a man sitting at a desk with a laptop) on the left. An arrow points from the user to a central image of a hard drive. Another arrow points from the hard drive to a second user (a man sitting at a desk with a laptop) on the right. A third arrow points from the hard drive down to a security threat (a person in a balaclava running) at the bottom, with a background of red binary code.

Zabezpieczenie danych – zagrożenia i wyzwania (4)

- **Długoterminowe archiwizacja danych i udostępnianie przechowywanych kopii:**
 - **Problem 1: Ograniczona trwałość nośników**



Dysk twardy – 5-10 lat
(w zależności od obciążenia)



CD/DVD – praktycznie 5-10 lat



Pamięć flash – 10 lat



Taśma LTO – 15-30 lat
(5000 montowań)



Blu-ray - ???

Zabezpieczenie danych – zagrożenia i wyzwania (5)

- Problem 2: Starzenie się technologii:



Dysk twardy
IBM 350 (1956)



Dysk twardy 3,5"
(Rodime, 1986)



SSD (1995, M-Systems
i później)



IBM 3480
(1984)



...



DLT
(1985/86)
do 70 GB/taśmę

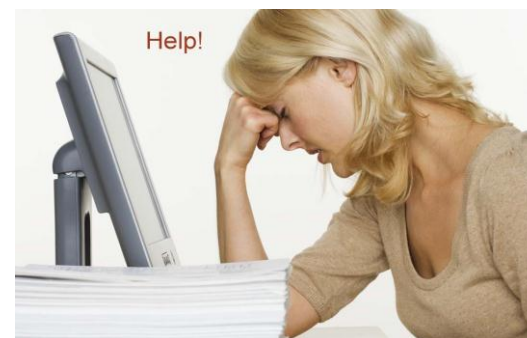
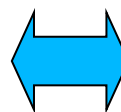


LTO4/5
(2007/2010)
do 1,6 TB taśmę



Zabezpieczenie danych – zagrożenia i wyzwania (6)

- Złożoność procesu archiwizacji danych i dostępu do nich:
 - Problem 1: złożoność procesu



Zabezpieczenie danych – zagrożenia i wyzwania (7)

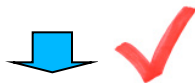
- **Złożoność procesu** archiwizacji danych i dostępu do nich:
 - **Problem 2: potrzeba automatyzacji**



Dzień 1



Kopia



Dzień 2



Kopia



Dzień 3



Kopia



Dzień 4



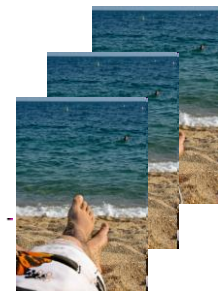
Kopia



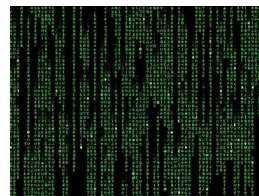
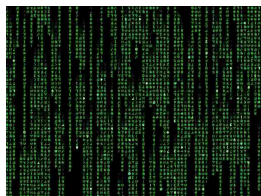
Dzień 5 - wakacje



Brak kopii

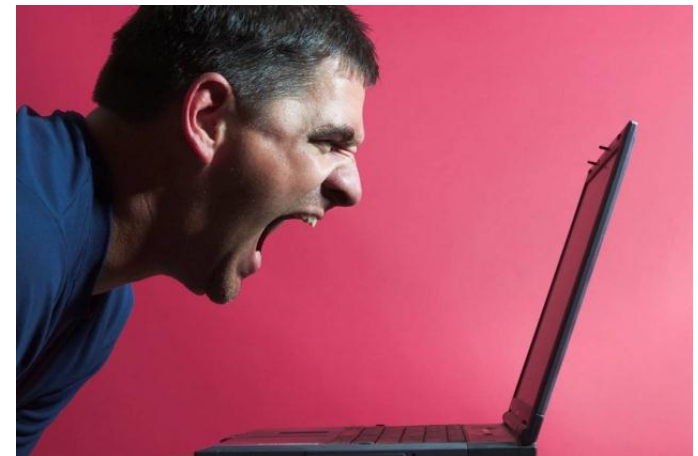


Wakacje
już na zawsze



Zabezpieczenie danych – zagrożenia i wyzwania (8)

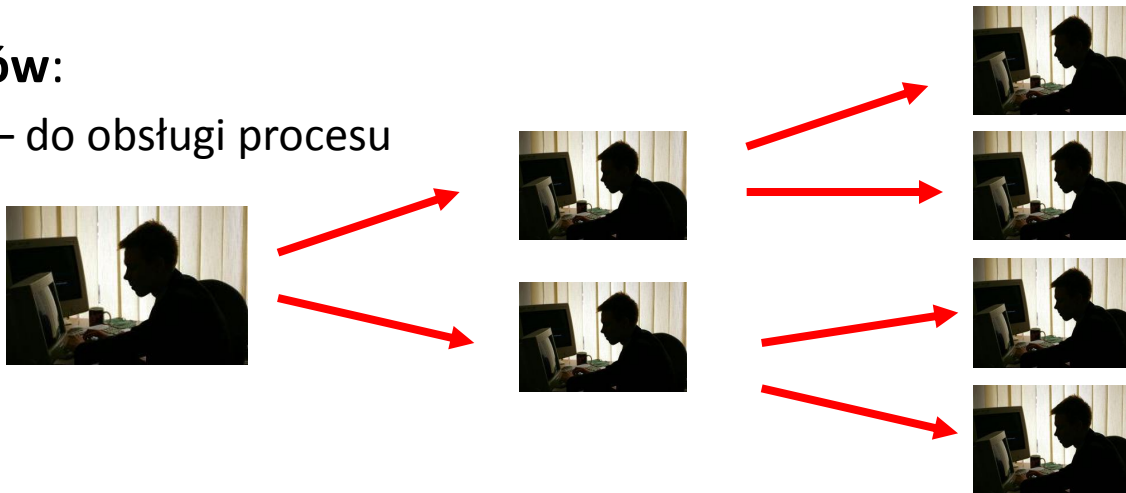
- **Brak doświadczenia i *know-how* w dziedzinie:**
 - Wiele instytucji nie ma dedykowanych działów IT
 - lub są one „niedoszacowane” (za mało pracowników)
 - Działy IT w wielu instytucjach:
 - zaniedbują proces wykonywania kopii danych, który:
 - jest uciążliwy – jeśli nie zautomatyzowany
 - nie jest regularny
 - obejmuje wycinek danych organizacji
 - nie mają doświadczenia z zarządzaniem dużymi ilościami danych i cyklem życia danych



Zabezpieczenie danych – zagrożenia i wyzwania (9)

- **Brak zasobów:**

- Ludzkich – do obsługi procesu



- Sprzętu i oprogramowania – do automatyzacji procesu



Dysk 1TB,
Koszt: 300 zł
brak redundancji (np. RAID)



Macierz SAS-SATA: 12-48x dysk 1TB,
Koszt: 30 000 – 90 000 zł
redundancja (RAID)



Macierz FC-SATA: 16 x dysk 1TB,
koszt: 280 000 zł (tzw. *List Price*)
redundancja (RAID)



Sieć SAN
koszt: 100 000 zł



Taśma LTO5, 1.5 TB
Koszt: 800 zł
Wymaga napędu ☹



Napęd IBM TS2350: 1 taśma,
1 napęd, koszt:



Autoloader Quantum Superloader 3;
16 taśm, 1 napęd
koszt:



Oprogramowanie:
Koszt: 4 000 - zł
Wymaga serwera ☹

Problemy z zabezpieczeniem danych (10)

=> Jest wiele problemów z zabezpieczeniem danych

=> Potrzebne jest solidne i wiarygodne
rozwiązanie tych problemów

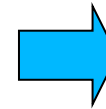
Outsourcing problemu – realne rozwiązanie

- Outsourcing do PLATON-U4 jest:

- Prostym i skutecznym rozwiązaniem
- Korzystny z pkt. widzenia kosztów



Dane



Usługa PLATON-U4

- Alternatywy:

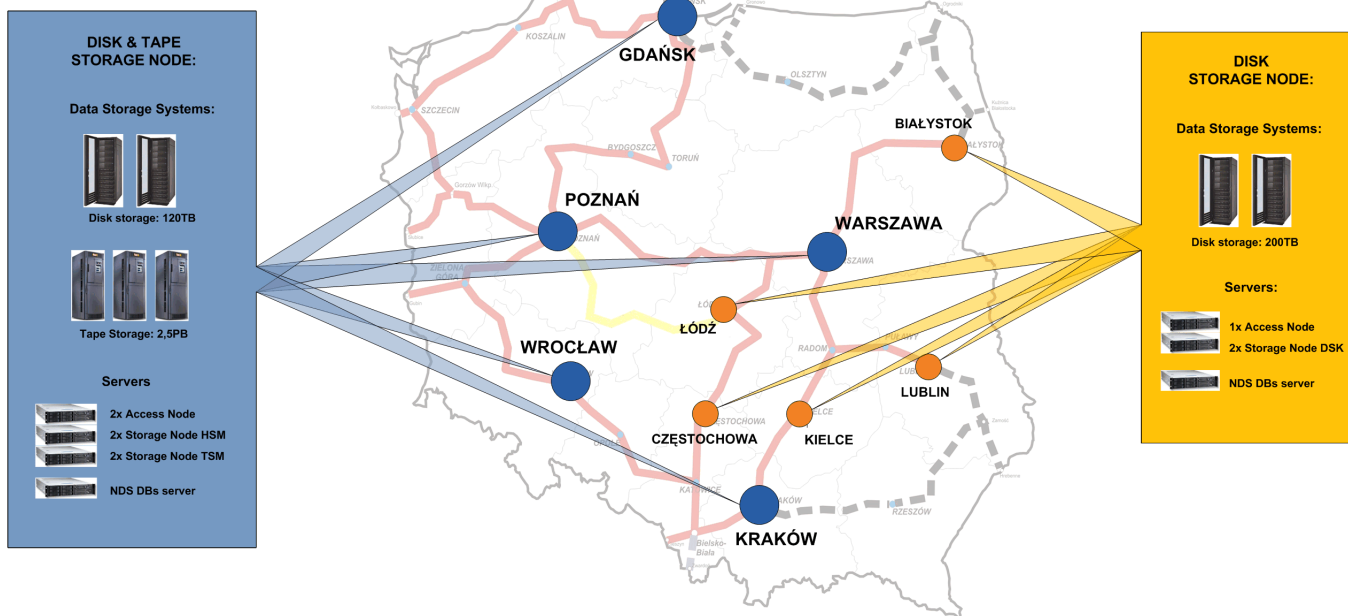
- Inne możliwości outsourcingu:
 - Usługi kopii zapasowych w sieci:
 - np. iBard24, AutoBackup.pl
 - Przechowywanie chmurowe:
 - Amazon S3 (Simple Storage Service)
- Rozwiązania producenckie – w kolejnych prezentacjach

PLATON-U4

– cechy i funkcje

PLATON-U4: 10 lokalizacji, konsorcjum 10 jednostek

- **Zabezpieczenie fizyczne danych:**
 - Replikacja geograficzna
 - Rozproszona infrastruktura:
 - **12,5 PB** pamięci taśmowych – w 5 lokalizacjach – automatyczne biblioteki taśmowe
 - **2 PB** pamięci dyskowych – w 10 lokalizacjach – macierze dyskowe i serwery plików
 - **70** serwerów, sieci SAN (ang. Storage Area Network) i 10Gbit Ethernet



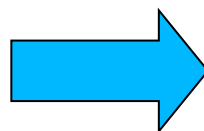
- Zapewnienie faktycznej redundancji danych:



Użytkownik



Dane
użytkownika



Usługa PLATON-U4



Replika 1



Centrum
Danych 1



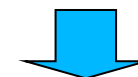
Replika 2



Centrum
Danych 2



Replika 3



Centrum
Danych 3

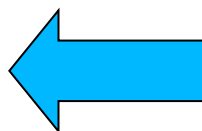
- Zapewnienie faktycznej redundancji danych:



Użytkownik



Dane
użytkownika



Usługa PLATON-U4

Dane dostępne!

ODTWARZANIE



Replika 1



Replika 2



Replika 3



Centrum
Danych 1



Centrum
Danych 2



Centrum
Danych 3



PLATON-U4 – cechy usługi

- **Zabezpieczenie fizyczne danych:**
 - Bezpieczne centra danych



Serwerownia w PCSS



Wiele linii zasilania



Redundantna klimatyzacja



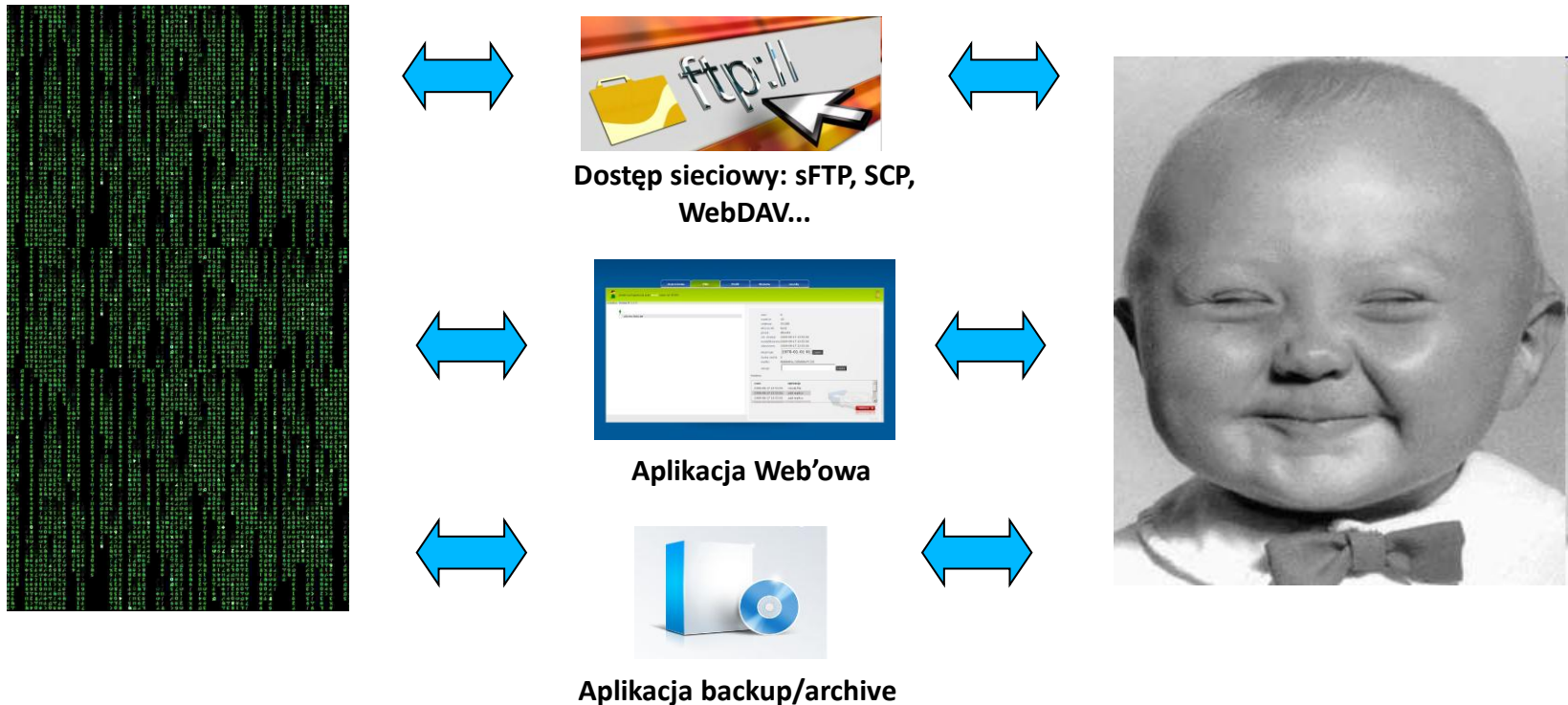
System wczesnego ostrzegania



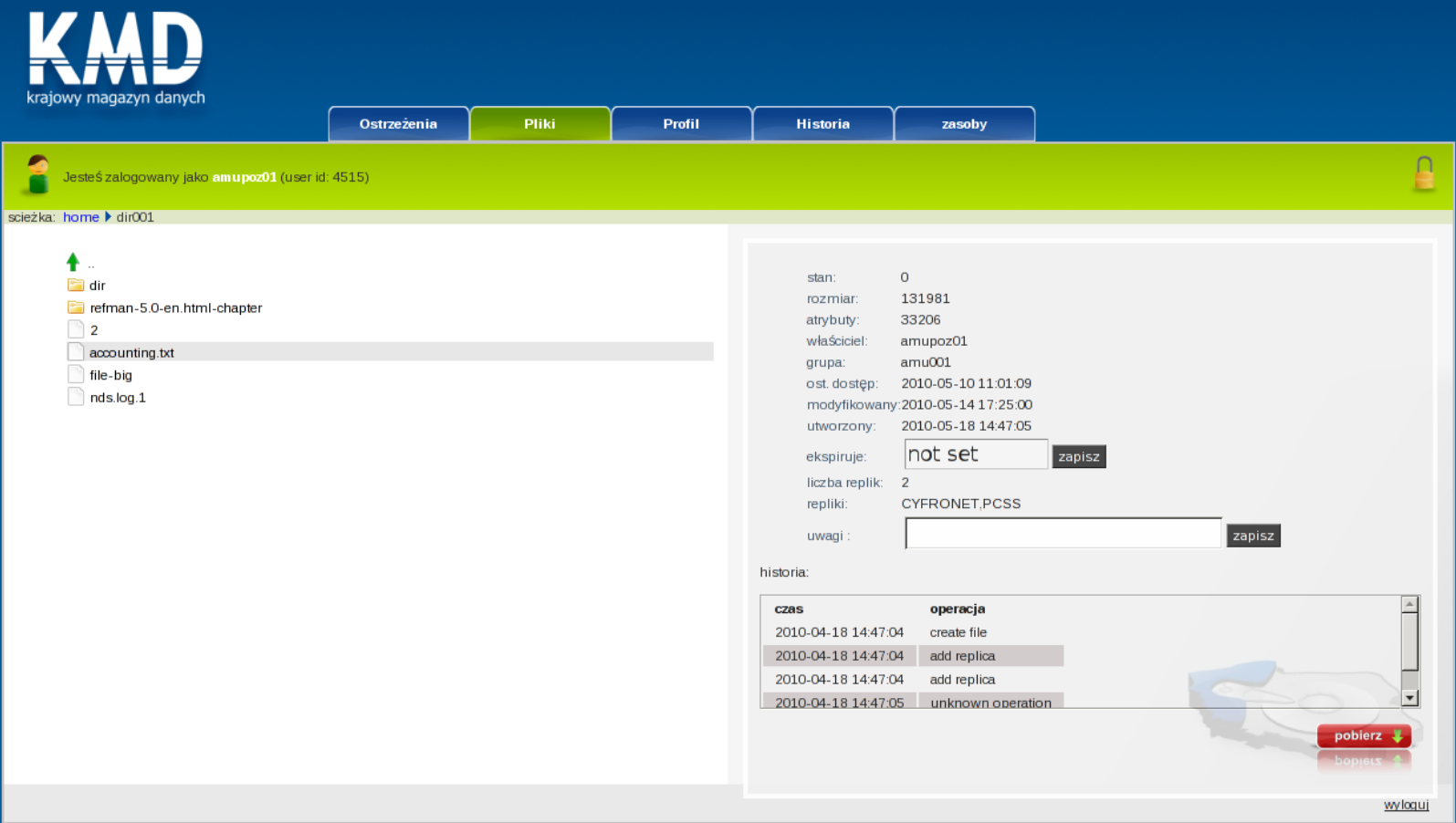
Monitoring wizyjny

PLATON-U4 – cechy usługi c.d.

- Prosty, standardowy interfejs użytkownika



PLATON-U4 – Interfejs użytkownika (Web)



KMD
krajowy magazyn danych

Ostrzeżenia Pliki Profil Historia zasoby

Jesteś zalogowany jako **amupoz01** (user id: 4515)

ścieżka: [home](#) ▶ [dir001](#)

- ↑ ..
- dir
- refman-5.0-en.html-chapter
- 2
- accounting.txt
- file-big
- nds.log.1

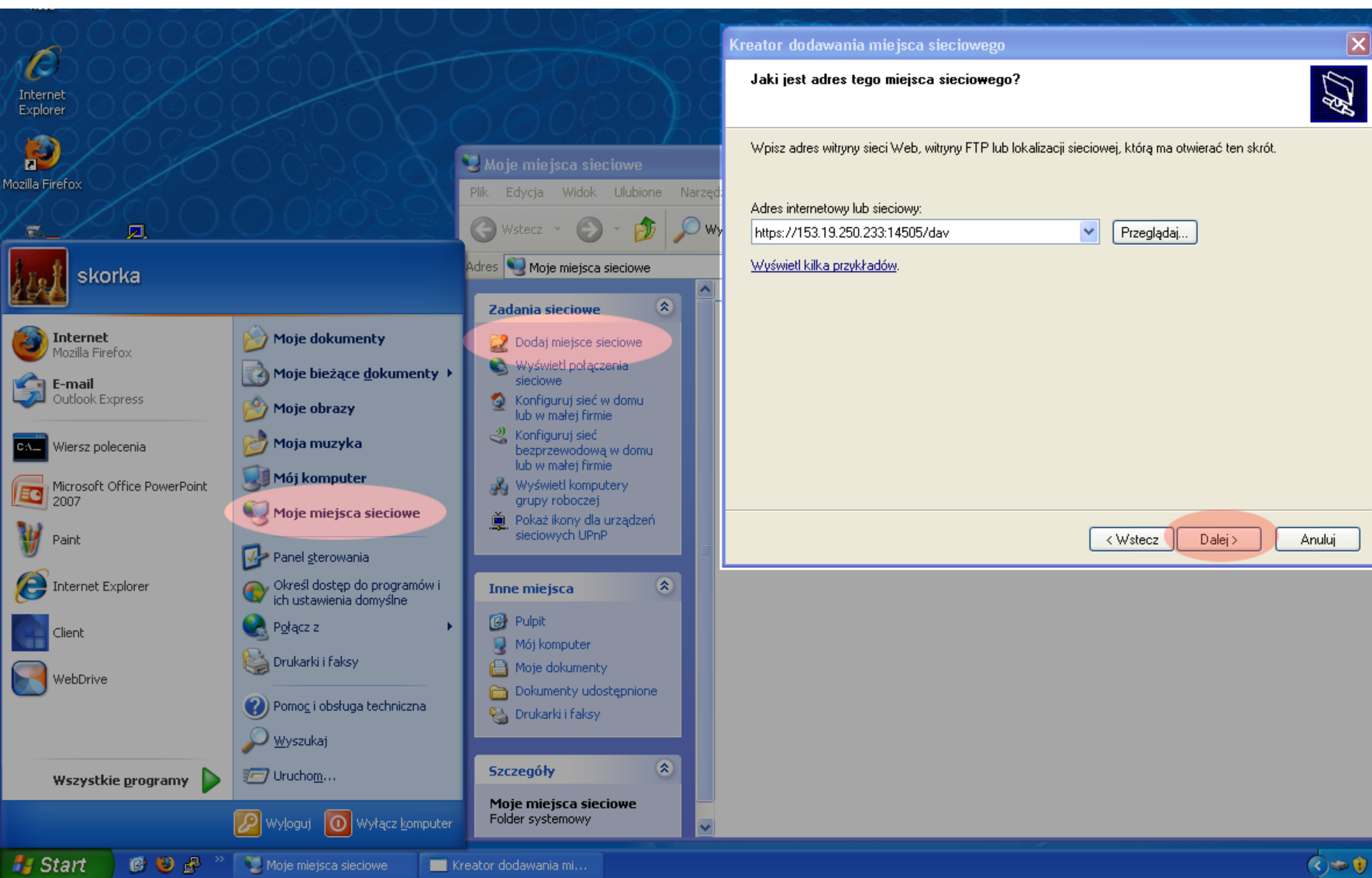
stan: 0
rozmiar: 131981
atrybuty: 33206
właściciel: amupoz01
grupa: amu001
ost. dostęp: 2010-05-10 11:01:09
modyfikowany: 2010-05-14 17:25:00
utworzony: 2010-05-18 14:47:05
ekspiruje:
liczba replik: 2
repliki: CYFRONET,PCSS
uwagi:

historia:

czas	operacja
2010-04-18 14:47:04	create file
2010-04-18 14:47:04	add replica
2010-04-18 14:47:04	add replica
2010-04-18 14:47:05	unknown operation

[wyloguj](#)

PLATON-U4 – Interfejs użytkownika (WebDAV)



The screenshot displays a Windows XP desktop environment. The Start menu is open, showing various applications and folders. The 'Moje miejsca sieciowe' (My Network Places) window is open, displaying a list of network locations. The 'Kreator dodawania miejsca sieciowego' (Add Network Location Wizard) is running, showing the 'Jaki jest adres tego miejsca sieciowego?' (What is the address of this network location?) step. The wizard is configured with the address 'https://153.19.250.233:14505/dav' and the 'Przeglądaj...' (Browse...) button is visible. The 'Dalej >' (Next >) button is highlighted with a red circle.

Start menu items:

- Internet Explorer
- Mozilla Firefox
- skorka
- Internet Mozilla Firefox
- E-mail Outlook Express
- Wiersz poleceń
- Microsoft Office PowerPoint 2007
- Paint
- Internet Explorer
- Client
- WebDrive
- Wszystkie programy
- Wyloguj
- Wyłącz komputer

Moje miejsca sieciowe window:

- Plik Edycja Widok Ulubione Narzędzia
- Wstecz
- Wyświetl
- Adres: Moje miejsca sieciowe
- Zadania sieciowe
 - Dodaj miejsce sieciowe
 - Wyświetl połączenia sieciowe
 - Konfiguruj sieć w domu lub w małej firmie
 - Konfiguruj sieć bezprzewodową w domu lub w małej firmie
 - Wyświetl komputery grupy roboczej
 - Pokaż ikony dla urządzeń sieciowych UPnP
- Inne miejsca
 - Pulpit
 - Mój komputer
 - Moje dokumenty
 - Dokumenty udostępnione
 - Drukarki i faksy
- Szczegóły
 - Moje miejsca sieciowe
 - Folder systemowy

Kreator dodawania miejsca sieciowego:

Jaki jest adres tego miejsca sieciowego?

Wpisz adres witryny sieci Web, witryny FTP lub lokalizacji sieciowej, którą ma otwierać ten skrót.

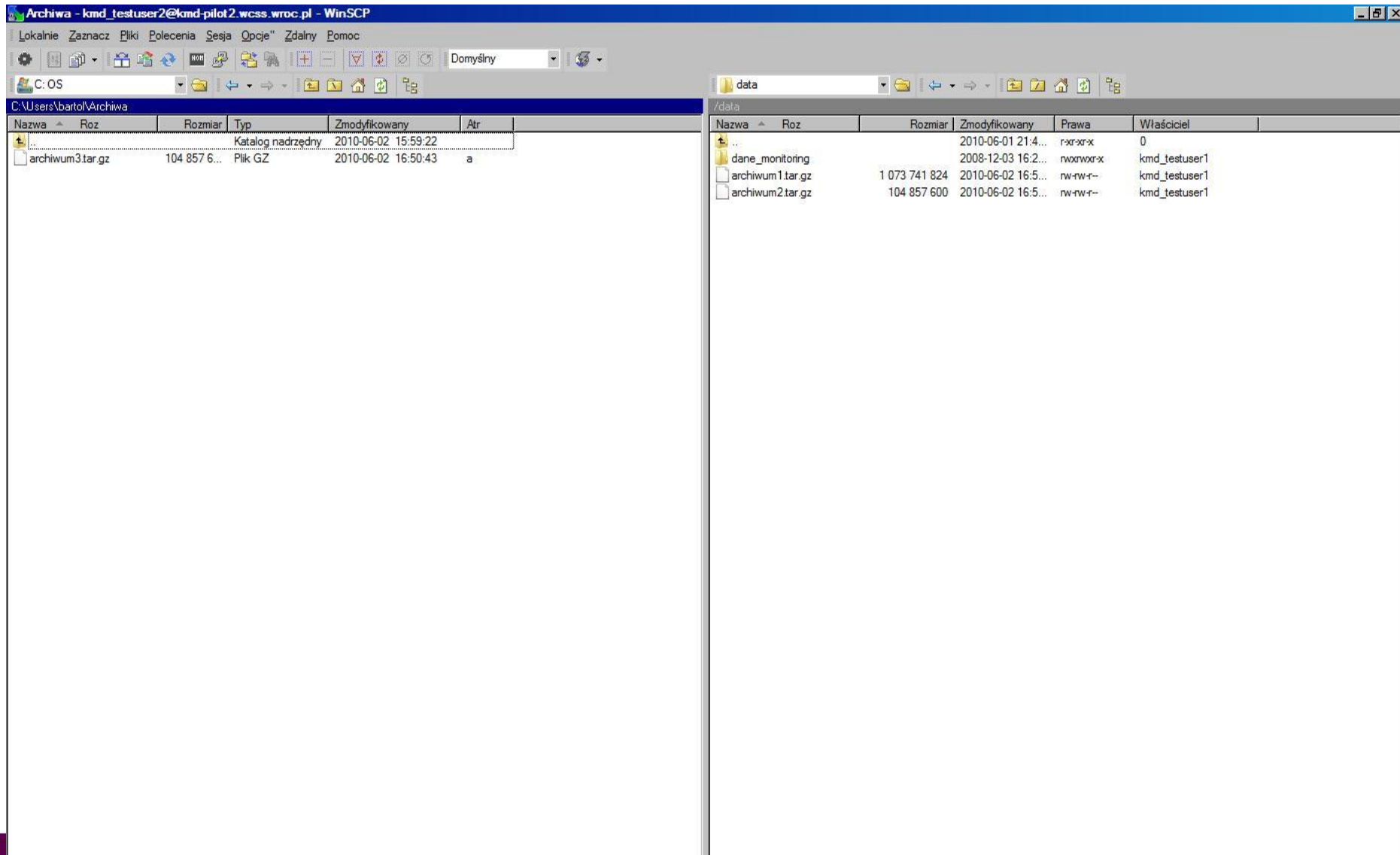
Adres internetowy lub sieciowy:

https://153.19.250.233:14505/dav Przeglądaj...

[Wyświetl kilka przykładów.](#)

< Wstecz Dalej > Anuluj

PLATON-U4 – Interfejs użytkownika (SCP)



The screenshot displays the WinSCP interface with two panes. The left pane shows the local file system at C:\Users\bartol\Archiwa, and the right pane shows the remote file system at /data.

Local File System (C:\Users\bartol\Archiwa):

Nazwa	Roz	Rozmiar	Typ	Zmodyfikowany	Atr
..			Katalog nadrzędny	2010-06-02 15:59:22	
archiwum3.tar.gz		104 857 6...	Plik GZ	2010-06-02 16:50:43	a

Remote File System (/data):

Nazwa	Roz	Rozmiar	Zmodyfikowany	Prawa	Właściciel
..			2010-06-01 21:4...	rxr-xr-x	0
dane_monitoring			2008-12-03 16:2...	rw-rw-r--	kmd_testuser1
archiwum1.tar.gz		1 073 741 824	2010-06-02 16:5...	rw-rw-r--	kmd_testuser1
archiwum2.tar.gz		104 857 600	2010-06-02 16:5...	rw-rw-r--	kmd_testuser1

Użytkownik

Dane użytkownika

Aplikacja kopii zapasowych / archiwalnych

Kopia danych

Serwery PLATON-U4

PLATON-U4

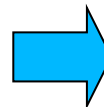
– warunki korzystania

Outsourcing do PLATON-U4 – koszty

- Składowanie kopii archiwalnych w PLATON-U4:
 - Przez 3 lata za darmo
 - Po 3 latach (od 2012) przez 5 lat:
 - Utrzymanie usługi w ramach kosztów własnych – „po kosztach”
 - Efekt skali pozwala na optymalizację jednostkowych kosztów przechowywania
 - Po tym okresie należy się spodziewać projektów kontynuujących PLATONa



Dane



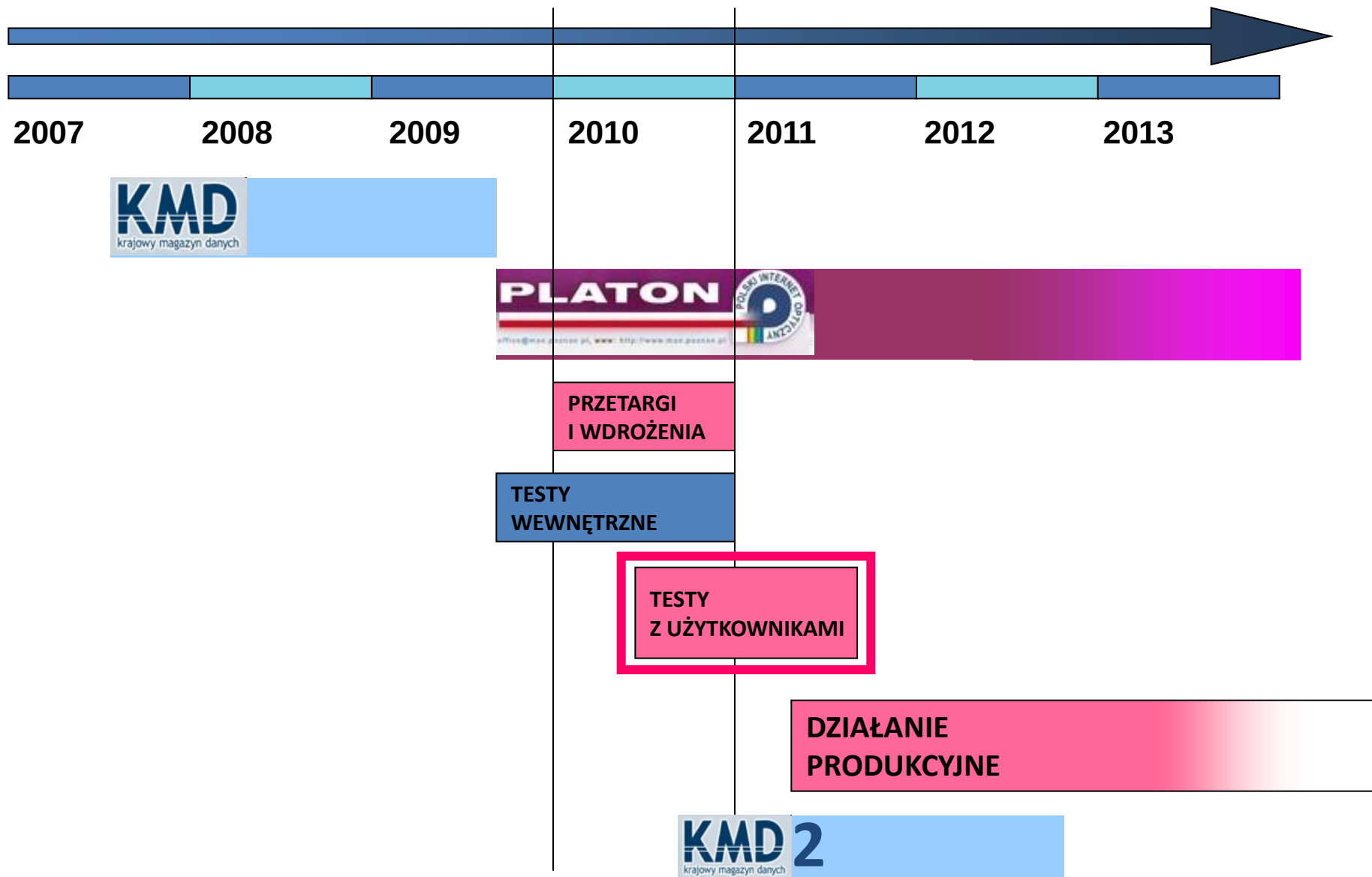
Usługa PLATON-U4

- Użytkownicy:
 - Instytucje akademickie i naukowe
 - Konsorcja mieszane? (sprawdzamy możliwości w MNiSW)

PLATON-U4

– status projektu
oraz dalsze plany

PLATON-U4 – status projektu



PLATON-U4 – plany

- **Najbliższe działania:**
 - **Warsztaty „Usługa powszechnej archiwizacji” w 5 miastach Polski:**
 - 1/2.XII Wrocław
 - 9.XII Kielce
 - 17.XII Białystok
 - + Lublin, Łódź i Częstochowa – daty do ustalenia
 - **Testy z użytkownikami – 4Q2010/1Q2011**
 - **Zapraszamy do testowania!**
 - Zainteresowanych prosimy o wypełnienie ankiety:
 - http://www.platon.pionier.net.pl/online/archiwizacja_ankieta.php
 - Dostęp: użytkownik: ankieta, hasło: warsztaty_platon-u4
 - **Wdrożenie produkcyjne – 1Q2011...**

Podsumowanie

Podsumowanie (1)

- **Dane = wartość = problem**

- Dane są bardzo wartościowe / bezcenne
- Dane są „wrażliwe”
- Danych jest coraz więcej:
 - **Biblioteki cyfrowe...**
 - Inne instytucje



- Efektywna archiwizacja dużych ilości danych:
 - jest trudne
 - ... i kosztowne

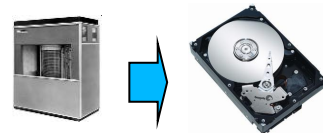


Podsumowanie (2)

- **Sensownym wyjściem wydaje się outsourcing:**

- „Ktoś inny” martwi się:

- O bezpieczeństwo fizyczne
 - Poufność i integralność danych
 - Trwałość nośników
 - Migrację między technologiami

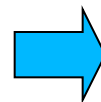


- Wartości dodane – wsparcie dla:

- Automatyzacji procesu archiwizacji
 - Zabezpieczenia kryptograficznego danych

- **Proponujemy outsourcing do usługi PLATON-U4:**

- Rozproszony system archiwizacji z replikacją geograficzną
 - Znany dostawca usługi
 - Niskie koszty dla użytkownika



Dane

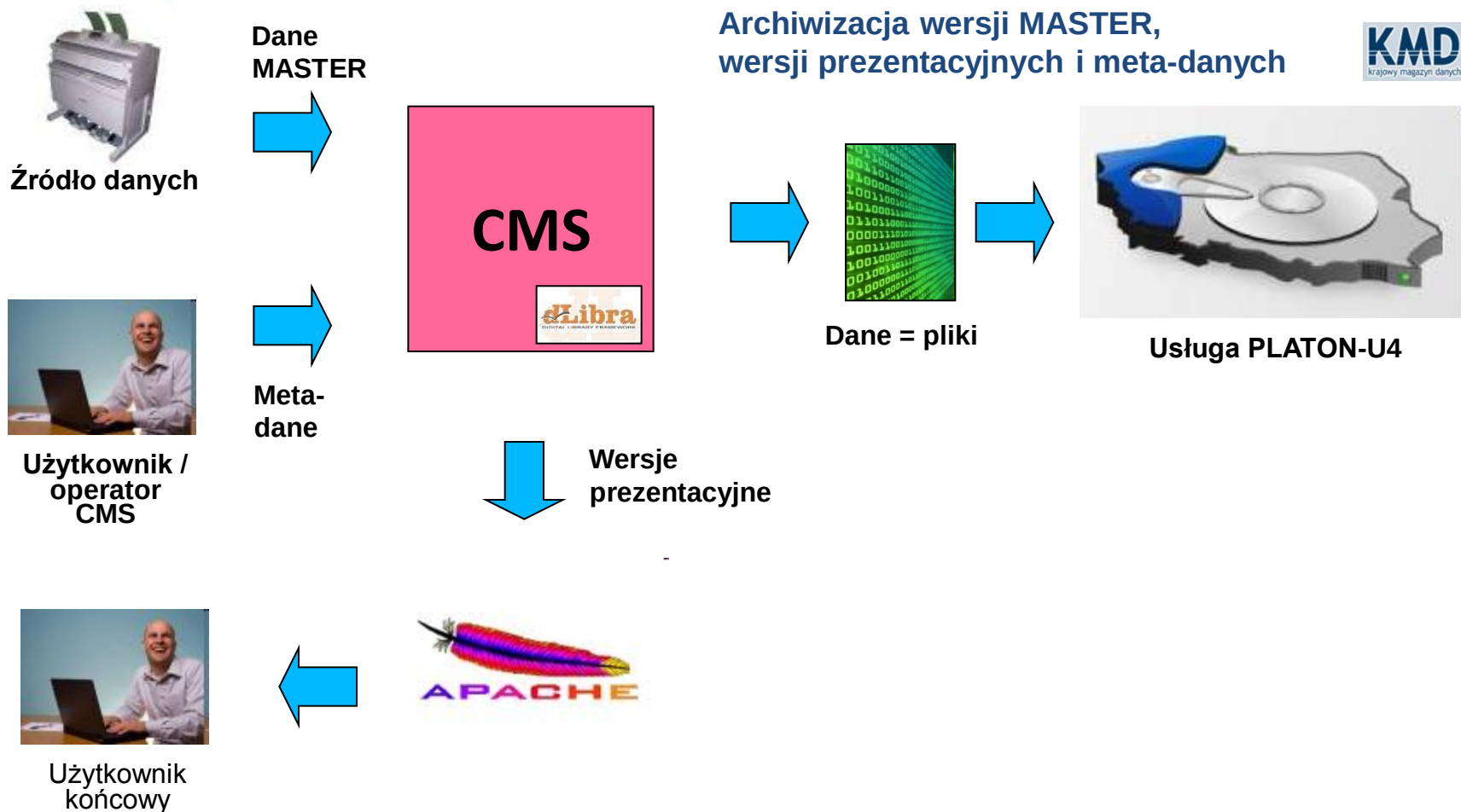
Usługa PLATON-U4

Podsumowanie (3)

- Usługa PLATON-U4 a archiwizacja zasobów cyfrowych:
 - PLATON-U4/KMD zapewnia:
 - **bezpieczeństwo fizyczne danych**
(m.in. dzięki replikacji geograficznej)
 - **trwałość danych**
 - **integralność danych – na poziomie ciągu bitów (pliku):**
 - zapewnienie **niezmienności** treści danych w czasie
 - **kompletność** danych – zabezpieczenie przed usunięciem całości lub części pliku
 - **poufność danych – niedostępność treści dla „obcych” systemów**
=> z punktu widzenia usługi składowany obiekt cyfrowy jest plikiem
 - **System Zarządzania Treścią (ang. CMS) powinien zapewniać:**
 - Interpretowalność danych
 - Autentyczność rozumianą jako pewność co do:
 - Pochodzenia danych
 - Niezafałszowania danych (na wysokim poziomie)
 - Poufność na wysokim poziomie

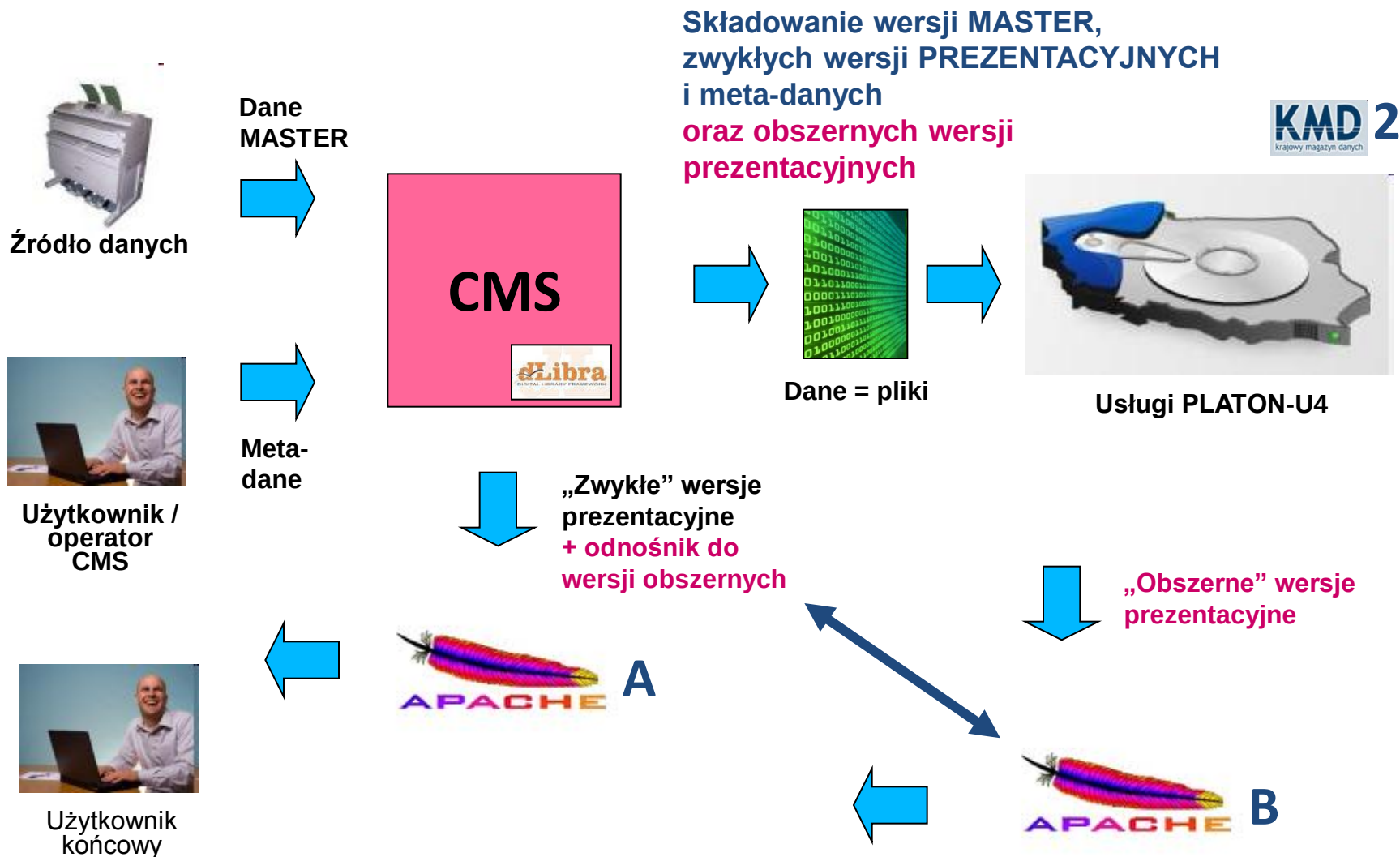
Podsumowanie (4)

- Możliwe sposoby wykorzystania usługi:



Podsumowanie (5)

- Przyszłe sposoby wykorzystania usługi:



Dalsze informacje

- **Artykuły:**

- Usługa Powszechnej Archiwizacji i jej zastosowanie w bibliotekach naukowych do zabezpieczenia i archiwizacji danych
 - <http://www.ebib.info/2010/115/a.php?brzezniak>
 - EBIB: Nr 6/2010 (115) „Nowy wymiar zabezpieczania zasobów”
- Popular backup/archival service and its application for the archival of the network traffic in the academic network PIONIER
 - Computational Methods in Science and Technology (2010)

- **Strony www:**

- <http://kmd.pcass.pl>
- <http://www.platon.pionier.net.pl/online/archiwizacja.php>



Usługa Powszechnej Archiwizacji

Dziękujemy za uwagę

Kontakt: maciekb@man.poznan.pl



Slajdy zapasowe

Przykładowe koszty alternatywnego przechowywania danych

Przyczyny problemów z zabezpieczeniem danych (5)

- **Wysokie koszty:**

- „tanie” dyski



dysk 1TB,
koszt: 300 zł
brak redundancji (np. RAID)

Założmy: 12 TB – 12 dysków



koszt: 3 600 zł
Nadal brak redundancji!

Dołożymy redundancję – RAID



Dysk dekstop:
koszt: 300 zł



+ kontroler RAID
karta PCI do serwera
Koszt 1 200 zł



Dysk
serwerowy:
koszt: 1 000 zł



+ serwer
12 wnęk na dyski
Koszt ok. 20 000 zł



+



+



=

RAZEM:
12 TB fizycznie
10 TB logicznie (RAID)
REDUNDANCJA!
Koszt ok. 34 000 zł
=> 3 400 zł / TB



Macierz SAS-SATA: 12 dysków 1TB
12 TB fizycznie
10 TB logicznie
REDUNDANCJA (RAID +2 kontrolery)
Koszt ok. 30 000 zł / 12 TB
=> 2 500 zł / TB

Przyczyny problemów z zabezpieczeniem danych (7)

- Wysokie koszty:

- taśmy



Taśma LTO5, 1.5 TB

koszt: 800 zł

wymaga napędu

Założmy: 16 taśm – 24 TB (LTO5)



Koszt 16 taśm: 12 800 zł

- brak redundancji!

- wymagana zmieniaraka

Dołożmy autoloader



Np. zmieniaraka Quantum Superloader 3

24 TB fizycznie

12 TB logicznie z REDUNDANCJĄ

Koszt ok. 22 000 zł

Dołożmy oprogramowanie



Oprogramowanie:

Koszt: 4 000 zł

Wymaga serwera ☺

oraz serwer...



1 serwer – 2 x 4-core CPU

6 dysków wewn. (baza oprogramowania)

Koszt ok. 20 000 zł



+



+



+



=

+ SERWISY... (typowo po 3 latach ok. 10% ceny listowej / rok)

RAZEM:

24 TB logicznie, 12 TB fizycznie (RAID)

Koszt ok. 58 800 zł

• 2 450 zł / TB (bez redundancji)

• 4 900 zł / TB (zakładając redundancję 1:1)

Przyczyny problemów z zabezpieczeniem danych (10)

- Koszty 1 TB danych – zakup i utrzymanie

Rozwiązanie	Koszt zakupu /TB	Koszt zakupu /TB / rok / przez 3 lata amort.	Koszt zakupu /TB / miesiąc / 3 przez lata amort.	Koszt serwisu /TB / rok / przez kolejne 3 lata	Koszt serwisu /TB / miesiąc / przez kolejne 3 lata	Średni koszt / miesiąc / TB / przez 6 lat
	[PLN/TB]	[PLN/TB]	[PLN/TB]	[PLN/TB]	[PLN/TB]	[PLN/TB]
Serwer z dyskami	3 400	1 133	94	408	34	64
	3 400	1 133	94	408	34	64
	3 400	1 133	94	408	34	64
	3 400	1 133	94	408	34	64
Macierz dyskowa	3 000	1 000	83	360	30	57
	2 550	850	71	306	26	48
	2 400	800	67	288	24	45
	2 325	775	65	279	23	44
Autoloader taśmowy	2 450	817	68	294	25	46
bez redundancji	1 950	650	54	234	20	37
Autoloader taśmowy	4 900	1 633	136	588	49	93
redundancją	3 900	1 300	108	468	39	74

Outsourcing problemu – alternatywy dla PLATON-U4

• Koszty składowania w innych systemach...

• iBard24 ¹⁾: *osoby prywatne* *firmy*

- 1 GB - 0,00 zł / rok 0,00 zł / rok
- 10 GB - 98,99 zł / rok 249,99 zł / rok

• **100 GB - 798,99 zł / rok 1399,99 zł / rok**

- > 100 GB - brak cen

• AutoBackup.pl ²⁾:

- 1 GB - 29,80 zł / rok (taryfa <3 GB)
- 10 GB - 154,80 zł / rok (taryfa 3-30 GB)

• **100 GB - 468,00 zł / rok (?) – „bez limitu pojemności”** (taryfa >30 GB)

• Amazon S3 ³⁾: *20% out* *50% out* *100% out*

- 1 GB - 6 zł / rok 7 zł / rok 7 zł / rok
- 10 GB - 68 zł / rok 69 zł / rok 71 zł / rok

• **100 GB - 684 zł / rok 689 zł / rok 715 zł / rok**

• **1 TB - 6 840 zł / rok 6 892 zł / rok 7 154 zł / rok**

• kruczki:

- PUT, COPY, POST, or LIST \$0.01 per 1,000 Requests
- GET and All Other Requests*** \$0.01 per 10,000 Requests
- XX% out - założenie: użytkownik pobiera do XX% danych składowanych uprzednio w systemie