

PLATFORMA OBSŁUGI NAUKI **PLATON**



www.platon.pionier.net.pl



**Usługi przechowywania danych
KMD/PLATON-U4
dla bibliotek cyfrowych**

**Maciej Brzeźniak, Norbert Meyer,
Rafał Mikołajczak, Maciej Stroiński**



**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

PLATON-U4 Plan prezentacji

- **KMD/PLATON-U4:**
 - Motywacje, cel, założenia
 - Zastosowania i potencjalni użytkownicy
 - Kim jesteśmy
- **Usługa PLATON-U4/Usługa Powszechnej Archiwizacji:**
 - Ważne cechy usługi
 - Architektura systemu
 - Środowisko realizacji usługi
- **PLATON-U4: Zastosowanie dla bibliotek cyfrowych**

Motywacje, cel

- Dane są bardzo wartościowe:
 - Bezcenne zbiory cyfrowe
 - Wyniki kosztownych badań
 - Społeczeństwo Informacyjne....



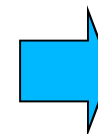
==



Cele i założenia

- **Cele nadrzędne:**

- „Usługa Powszechnej Archiwizacji”
- Pomoc użytkownikom i instytucjom w ZABEZPIECZENIU ich danych



- **Cele szczegółowe:**

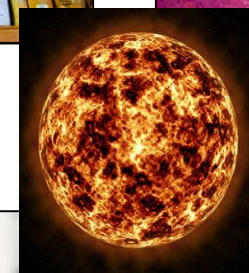
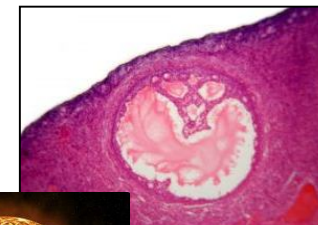
- Zabezpieczenie fizyczne danych
- Zapewnienie i kontrola integralności logicznej danych
- Poufność danych
- Długoterminowe przechowywanie i udostępnianie wykonanych kopii
- Dostarczenie narzędzi wspierających wykonywanie kopii danych



Zastosowania i użytkownicy

- **Potencjalni użytkownicy:**

- **Biblioteki cyfrowe** 😊
- Wirtualne laboratoria
- Uczelnie, pracownicy uczelni, naukowcy
- Instytuty PAN i ich pracownicy
- Akademickie centra obliczeniowe
- Szpitale kliniczne
- Operatorzy sieci miejskich MAN w ramach konsorcjum PIONIER
- Inni



Kim jesteśmy

- **KMD vs PLATON-U4:**

- **KMD** – Krajowy Magazyn Danych – projekt rozwojowy (2007-2009)
- **PLATON-U4** – „Usługa powszechnej archiwizacji” - wdrożenie KMD dla użytkowników EDU (2010)



- Partnerzy:

- Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS)
- Akademickie Centrum Komputerowe CYFRONET AGH
- Politechnika Gdańska, Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej
- Politechnika Białostocka (BIAMAN)
- Politechnika Częstochowska (PCz)
- Uniwersytet Marii Curie-Skłodowskiej w Lublinie, LubMAN UMCS (LUBMAN)
- Politechnika Łódzka, Centrum Komputerowe, (LODMAN)
- Politechnika Wrocławska, Wrocławskie Centrum Sieciowo Superkomputerowe (WCSS)
- Uniwersytet Warszawski, ICM
- Politechnika Świętokrzyska



Zastosowania

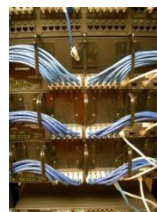
- **Docelowe zastosowanie**
 - Długoterminowa **archiwizacja danych**
 - **Kopie zapasowe**
 - Przechowywanie **2-go poziomu!!** (ang. *Secondary Storage*)
 - Zakładamy, że użytkownicy mają lokalne systemy przechowywania



Użytkownik



LOKALNY
SYSTEM
PRZECHOWYWANIA
(1-go poziomu)



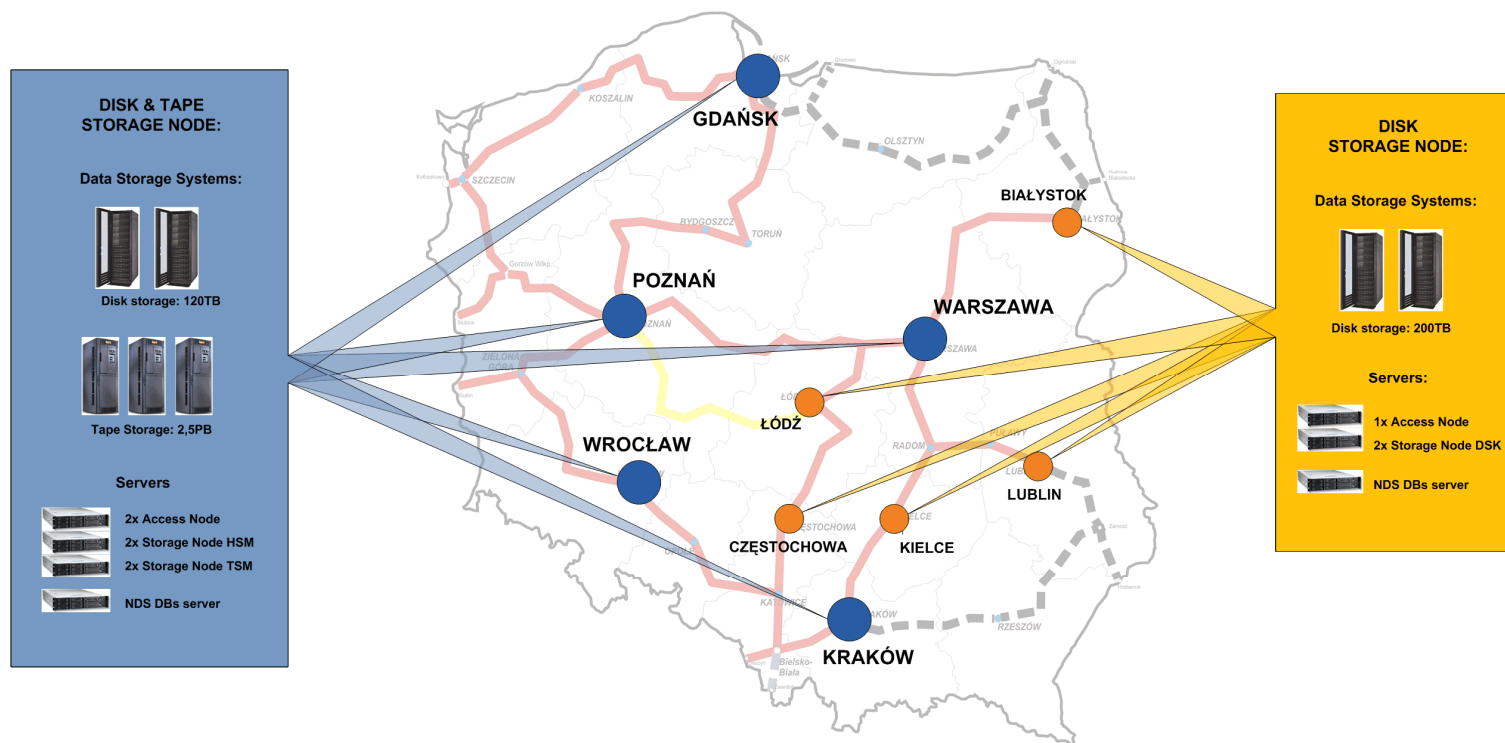
Sieć
PIONIER



Usługa PLATON-U4
KOPIA ARCHIWALNA/ZAPASOWA
(2-go poziomu)

Cechy usługi

- **Zabezpieczenie fizyczne danych:**
 - Replikacja geograficzna



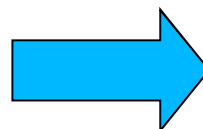
Cechy usługi - replikacja



Użytkownik



Dane
użytkownika



Usługa PLATON-U4

REPLIKACJA



Replika 1



Centrum
Danych 1



Replika 2



Centrum
Danych 2



Replika 3



Centrum
Danych 3

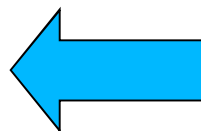
Cechy usługi - replikacja



Użytkownik



Dane
użytkownika



Usługa PLATON-U4

Dane dostępne!

ODTWARZANIE



Replika 1



Replika 2



Replika 3



Centrum
Danych 1



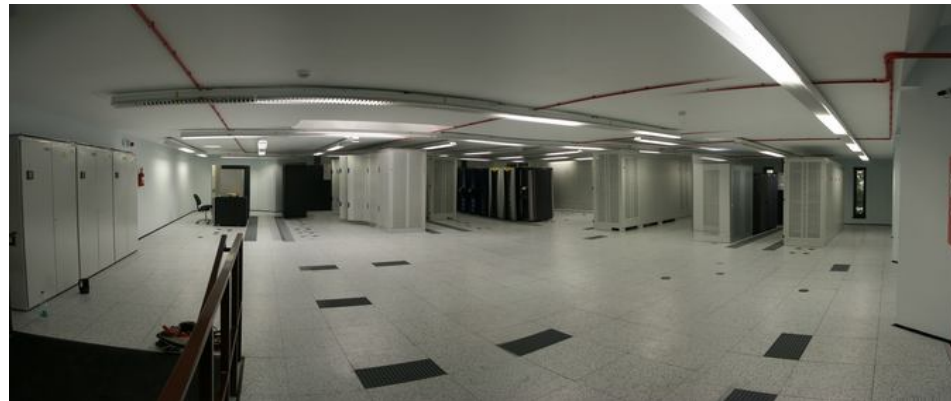
Centrum
Danych 2



Centrum
Danych 3

Cechy usługi

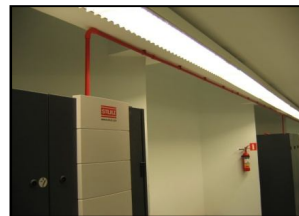
- **Zabezpieczenie fizyczne danych:**
 - Bezpieczne centra danych



Serwerownia w PCSS



Wiele linii zasilania



Redundantna klimatyzacja



System wczesnego ostrzegania



Monitoring wizyjny

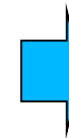
Cechy usługi

- **Zapewnienie i kontrola integralności logicznej danych:**
 - Wyliczanie skrótów kryptograficznych danych umieszczanych i składowanych
- **Poufność danych:**
 - Dane **szyfrowane** w drodze do systemu:
 - Wsparcie dla SSH, HTTPS
 - Dane **szyfrowane** wewnątrz systemu:
 - Łącza zabezpieczone kryptograficznie
 - Sprzętowe szyfrowanie w technologii taśmowej LTO4 / LTO5
- **Kontrola dostępu**



Cechy usługi

- Długoterminowe przechowywanie
 - Automatyczna kontrola i „odświeżanie” mediów:
 - *Disk-scrubbing* w macierzach dyskowych
 - Technologia S.M.A.R.T.
 - Kontrolne odczytywanie taśm
 - Automatyczne przepisywanie danych pomiędzy taśmami – oprogramowanie zarządzające np. *Tivoli Storage Manager*
 - Kontrola skrótów kryptograficznych danych – cyklicznie lub na żądanie (plany)
 - Porównywanie replik (plany)



43981352983132...



Cechy usługi

- Długoterminowe przechowywanie
 - Migracja między technologiami dyskowymi -> outsourcing



IBM 350
(1956)



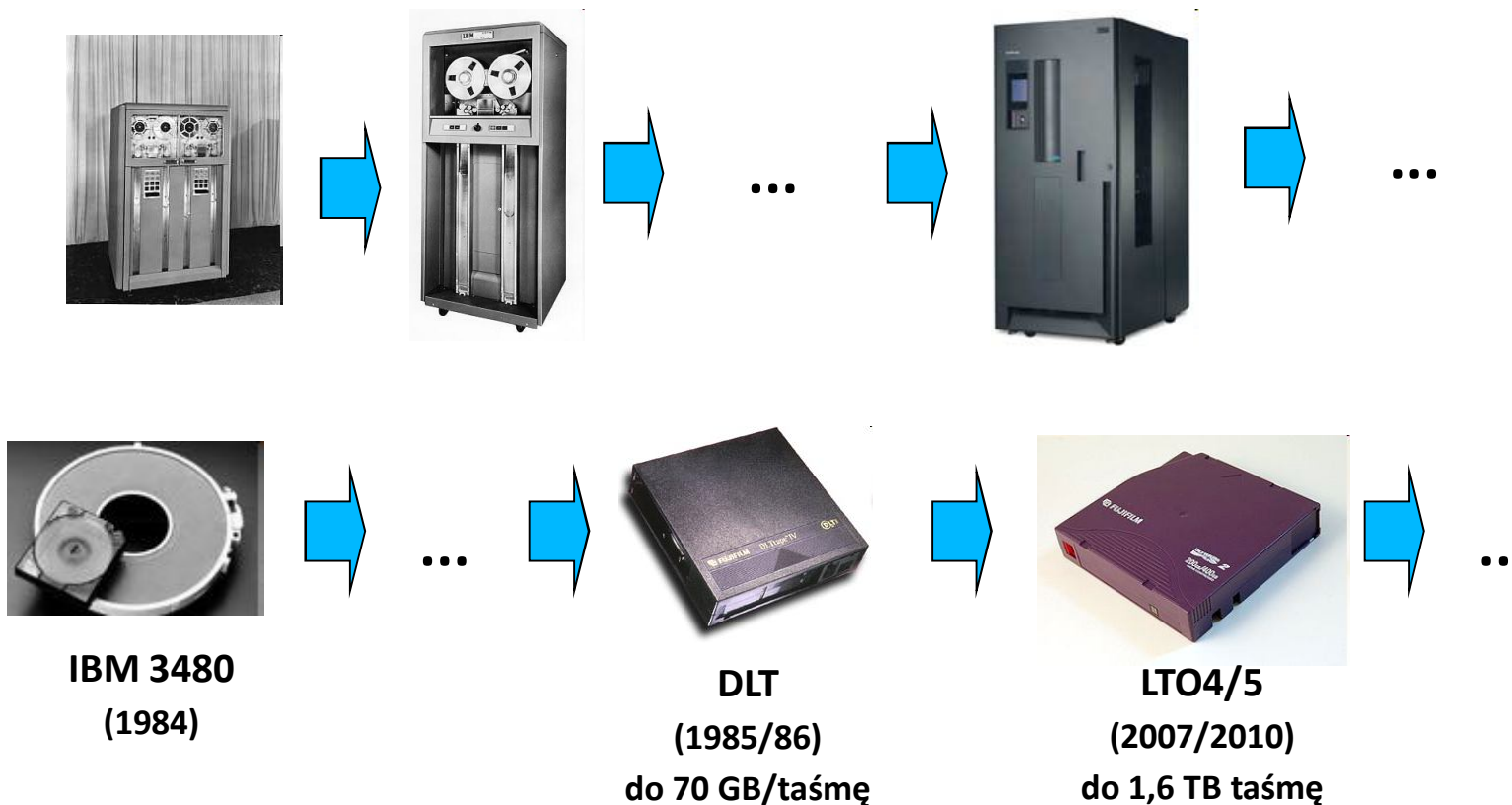
Dysk twarde
3,5"



SSD

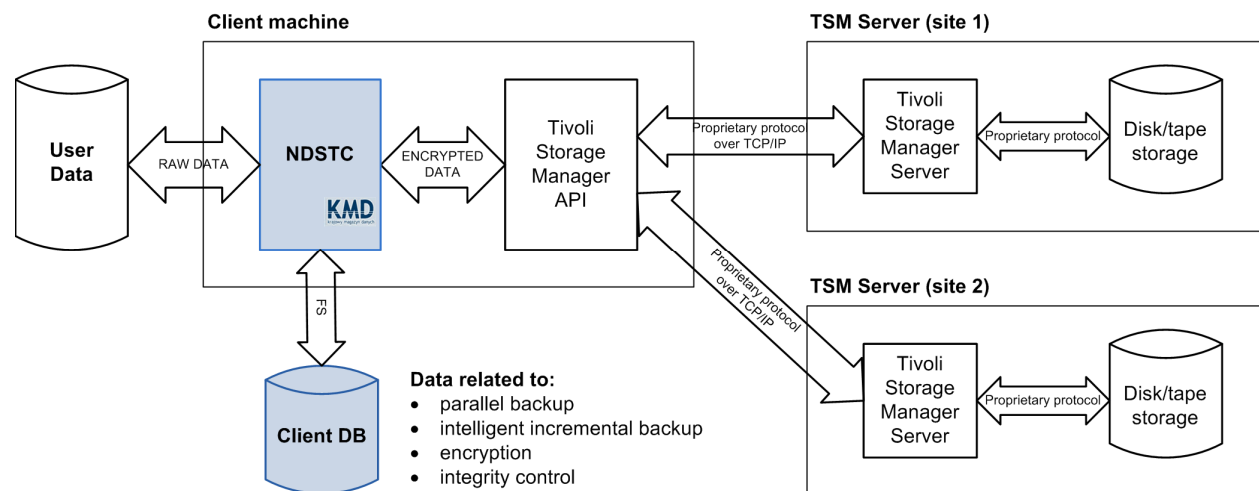
Cechy usługi

- **Długoterminowe przechowywanie**
 - Migracji między technologiami taśmowymi -> outsourcing

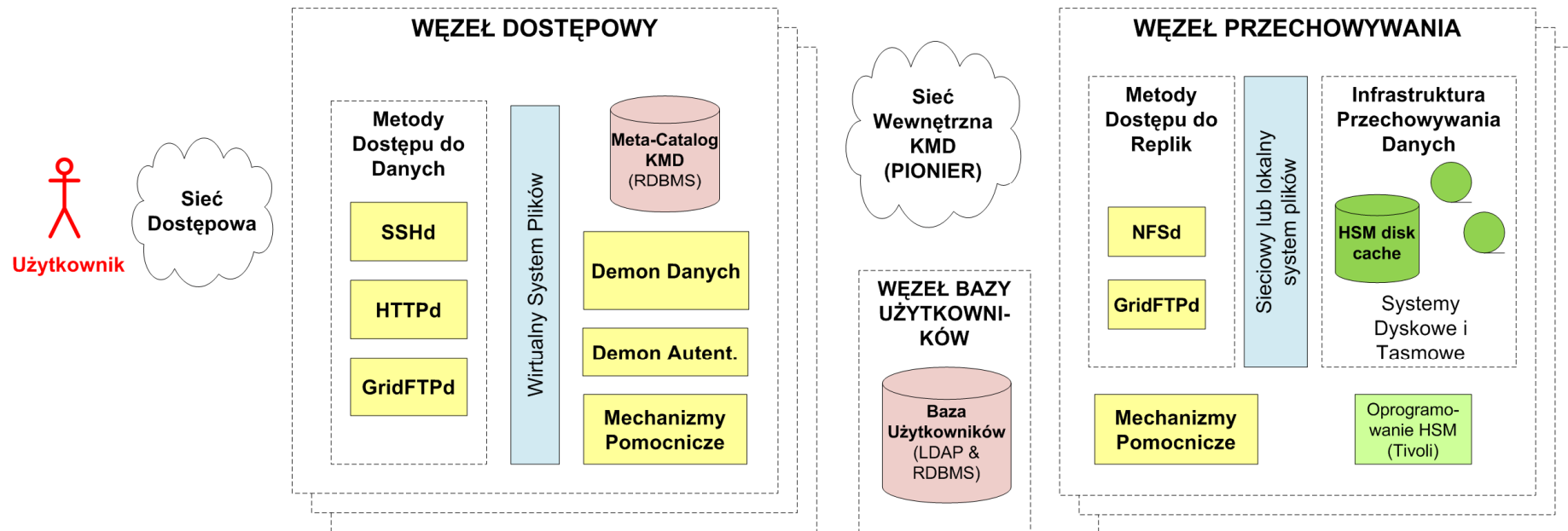


Cechy usługi

- Dostarczenie narzędzi wspierających wykonywanie kopii:
 - Szablony wdrożeń dla popularnego oprogramowania do wykonywania kopii zapasowych i archiwalnych m.in. *open source*
 - Tzw. *appliance* kopiująco-archiwizujące dla dużych instytucji (plany)
 - Aplikacja – zaawansowany klient kopii zapasowych / archiwizacji



PLATON-U4: Architektura usługi



- **Węzły Dostępowe (AN):**
 - Oferują różne **Metody Dostępu do Danych**: SSH, HTTP/WebDAV i GridFTP
- **Węzły Przechowywania (SN):**
 - Przechowują fizyczne repliki danych użytkownika,
 - Na różnego typu mediach: systemy dyskowe, biblioteki taśmowe,
 - Wykorzystują oprogramowanie TSM

PLATON-U4: Architektura usługi (4)

- **Zalety architektury**

- **De-centralizacja, rozproszenie danych** i usług:

- » Wiele punktów świadczenia usługi (Węzły Dostępowe),

- » Wiele punktów przechowywania danych (Węzły Przechowywania),

- **Automatyczna replikacja** danych (synchr. & asynchr.)
i replikacja meta-danych systemu KMD

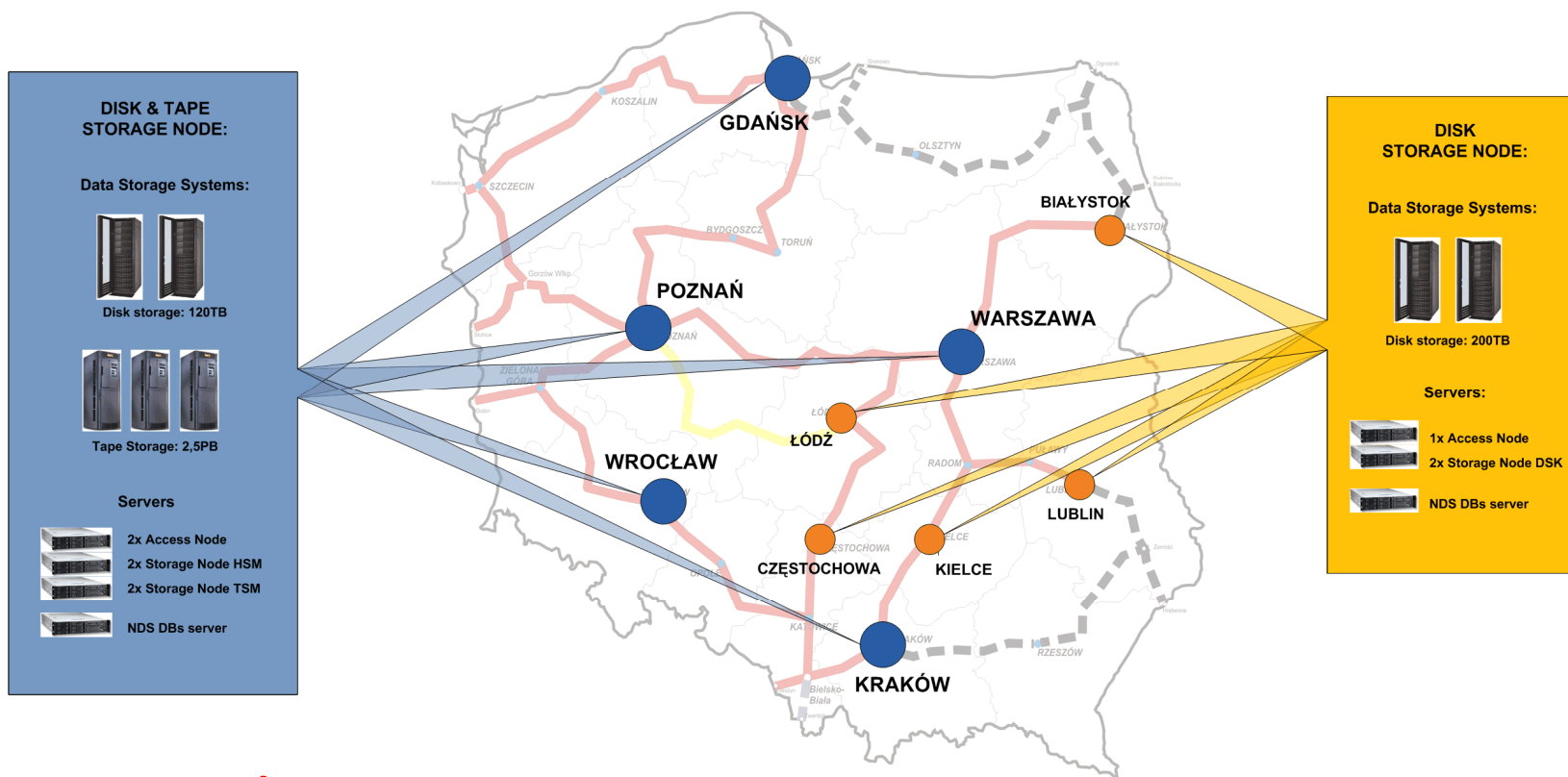
- **Skalowalność w wielu wymiarach:**

- » Liczba użytkowników i instytucji klienckich

- » Przestrzeń przechowywania, liczba obiektów danych

- » Wielkość ruchu od klientów

PLATON-U4: Realizacja usługi

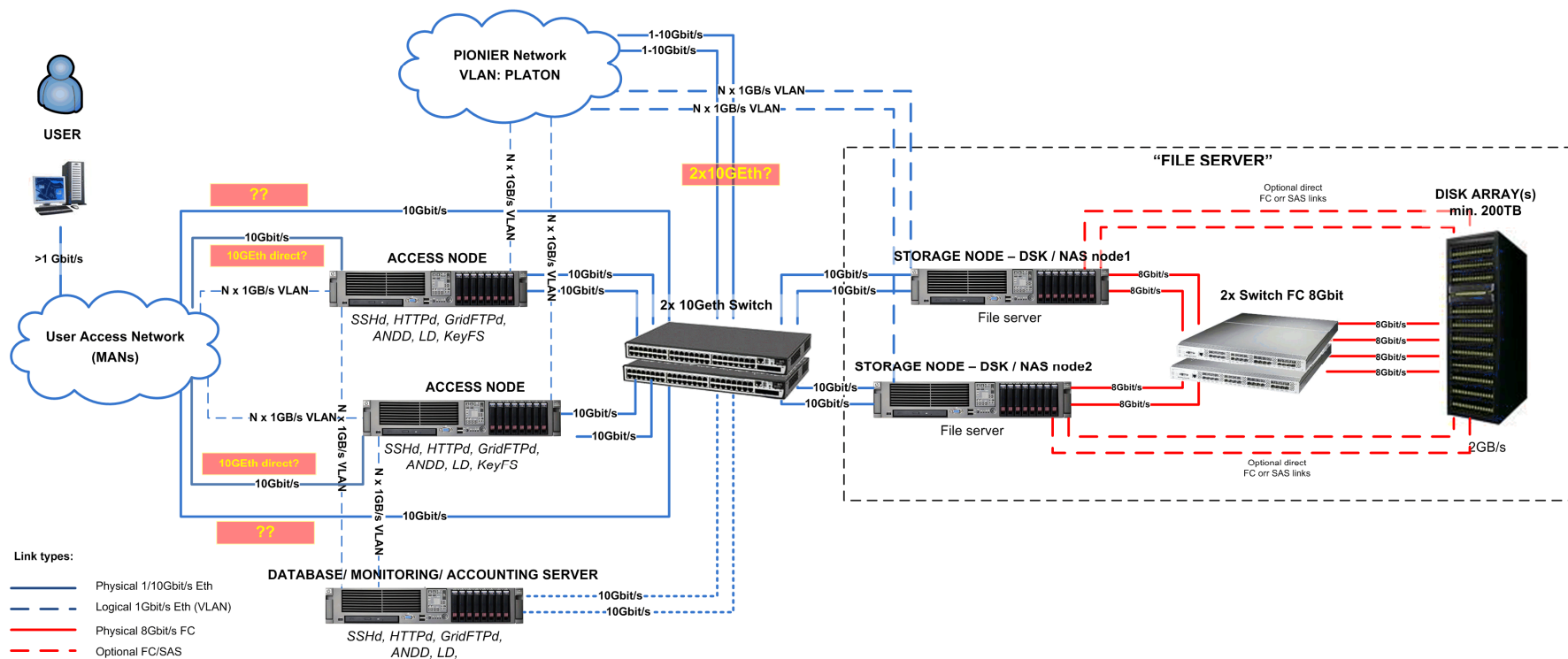


- **W sumie**

- 12,5 PB pamięci taśmowych
- 2 PB pamięci dyskowych

Węzły przechowywania dyskowego

NDS/PLATON DISK STORAGE NODE v2





Zastosowanie PLATON-U4 / KMD dla bibliotek cyfrowych

PLATON-U4: Zastosowanie dla bibliotek cyfrowych

Interfejs do danych oraz formaty danych:

- a. Standardowe interfejsy – **OK**.
 - Dostęp: *SCP, HTTP/WebDAV*, usługa B/A
 - Użytkownik widzi zdalny system plików (jak zdalny *site FTP* lub *HTTP/WebDAV*)
- b. Format danych nie gra roli – **OK**

Kompatybilność:

- a. dLibra - **OK**
- b. Inne – nie powinno być problemu

PLATON-U4: Zastosowanie dla bibliotek cyfrowych

Długoterminowe zabezpieczenie danych:

- a. Fizyczne - przeciw kataklizmom, aktom wandalizmu etc. – **OK**
- b. Trwałość danych i mediów – **OK**
- c. Migracja między technologiami – **OK**

=> OUTSOURCING PROBLEMU – OK!

Masowość danych (TB... PB... ?):

- a. Duża przestrzeń przechowywania – niskie koszty – **OK.**
- b. **Usługa jest darmowa – OK.**
- c. Wydajność dostępu do danych – Sieć PIONIER + MANy – **OK**

=> EFEKT SKALI – DUŻY SYSTEM

Plany

- 3Q2009-1Q2010: testy usługi:
 - wewnętrzne – w ramach konsorcjum KMD
 - zewnętrzne – **ZAPRASZAMY DO UDZIAŁU!!!**
- 1-2Q2010: Zakupy i instalacja sprzętu
- 2-3Q2010: Wdrożenie usługi powszechnej archiwizacji



Usługa Powszechnej Archiwizacji

Dziękujemy za uwagę

Kontakt: maciekb@man.poznan.pl
kmd.pcass.pl



PLATFORMA OBSŁUGI NAUKI **PLATON**



www.platon.pionier.net.pl



INSTYTUT CHEMII BIOORGANICZNEJ PAN

POZNAŃSKIE CENTRUM SUPERKOMPUSEROWO - SIECIOWE

ul. Noskowskiego 12/14, 61-704 Poznań, (+48 61) 858 20 00, fax: (+48 61) 852 59 54,

e-mail: office@man.poznan.pl, www: <http://www.man.poznan.pl>