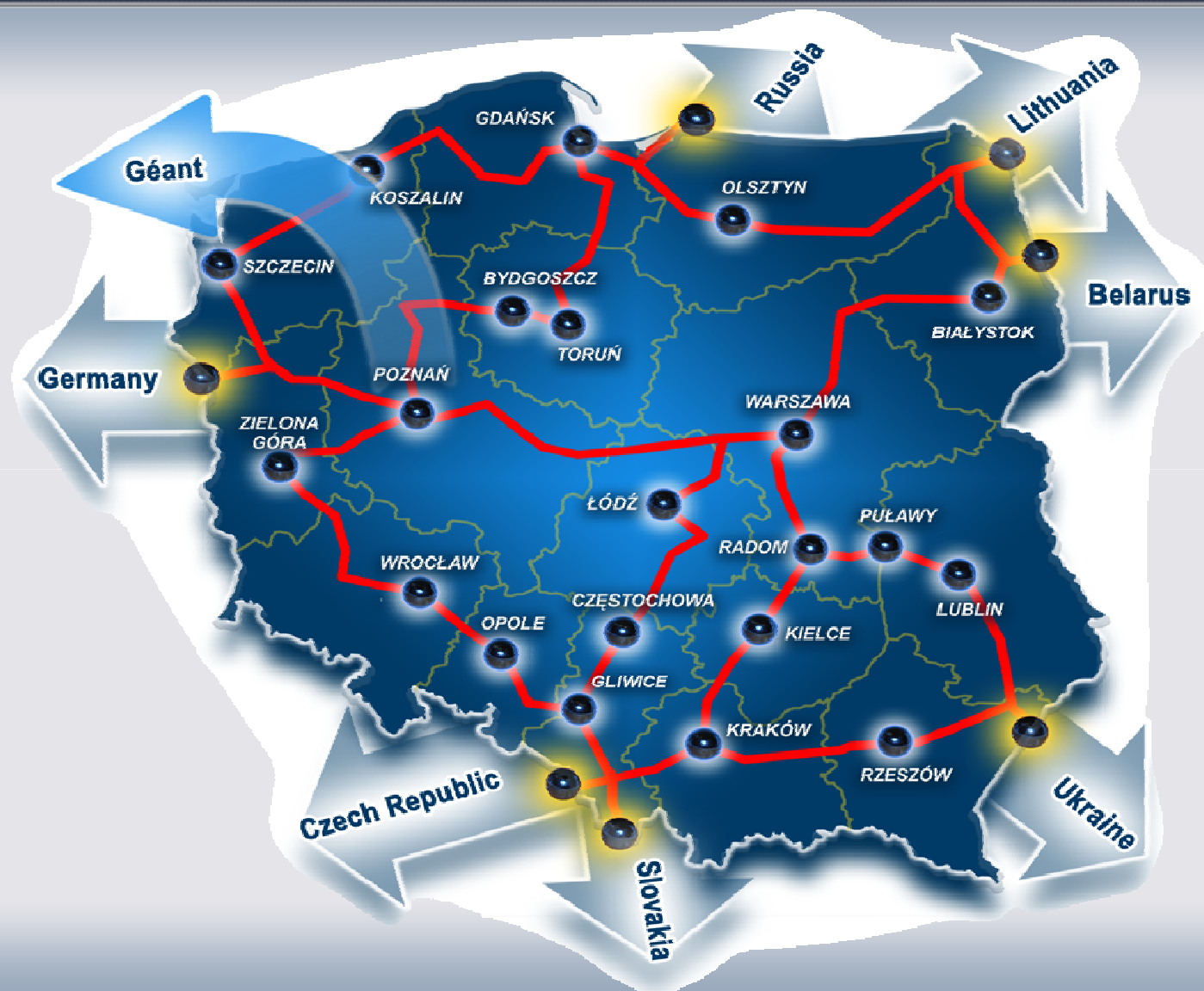




Infrastruktura bibliotek cyfrowych w sieci PIONIER

C. Mazurek, M.Stroiński, M.Werla, J.Węglarz

Poznańskie Centrum Superkomputerowo-Sieciowe



- Program PIONIER umożliwił realizację nowoczesnej, dobrze zorganizowanej i rozproszonej **infrastruktury bibliotek cyfrowych**
- Treści cyfrowe dostępne w środowisku bibliotek cyfrowych są przetwarzane przez **narzędzia i usługi** stanowiące element tej infrastruktury i określające funkcjonalność biblioteki cyfrowej, a także zakres odbiorców
- W takim kontekście można mówić o globalnej infrastrukturze bibliotek cyfrowych jako o zasobie wartościowym jednocześnie pod względem **naukowym** i **społecznym**
- Wyróżnienie narzędzi i usług na poziomie pojedynczej biblioteki cyfrowej umożliwia ich **federowanie** na poziomie globalnej infrastruktury bibliotek cyfrowych
- Przykładową implementacją zbioru narzędzi i usług biblioteki cyfrowej jest oprogramowanie **dLibra**
- **Federacja Bibliotek Cyfrowych** w sieci PIONIER jest przykładem składnika ogólnopolskiej infrastruktury informatycznej, która zapewnia dostęp do zdigitalizowanych zasobów dziedzictwa kulturowego, dokumentów i druków regionalnych oraz materiałów edukacyjnych

- Główne modele organizacyjne
 - **Regionalne biblioteki cyfrowe**
 - Współtworzone przez wiele instytucji
 - Najczęściej zasób związany z regionem, ale także materiały edukacyjne i dziedzictwo kulturowe
 - **Instytucjonalne biblioteki cyfrowe**
 - Tworzone przez pojedyncze instytucje
 - Najczęściej zasób związany z działalnością i historią danej instytucji
 - **Konsorcjalne biblioteki cyfrowe**
 - Tworzone przez konsorcja/federacje instytucji o zbliżonym profilu
 - Zasób analogiczny do inicjatyw instytucjonalnych
- Za zaplecze informatyczne bibliotek cyfrowych często odpowiadają lokalne/uczelniane centra sieciowe czy obliczeniowe



- Ogłoszenie przez Komisję Europejską w 2005 roku inicjatywy **i2010 Digital Libraries** oraz plany budowy Europejskiej Biblioteki Cyfrowej
 - wirtualne „połączenie” polskich rozproszonych bibliotek cyfrowych w jedną spójną platformę umożliwiającą automatyczną wymianę informacji pomiędzy poszczególnymi bibliotekami
- Takie polskie podejście umożliwiło rozwiązanie części problemów sygnalizowanych we wspomnianej powyżej inicjatywie i2010, poprzez
 - ułatwienie wykorzystania polskich zasobów cyfrowych,
 - zwiększenie ich widoczności w Internecie
 - stworzenie automatycznych mechanizmów wspierających koordynację digitalizacji w rozproszonych ośrodkach

- Najistotniejsze cechy infrastruktury bibliotek cyfrowych
 - Architektura silnie rozproszona
 - BC: 38 instalacji bibliotek cyfrowych
 - Systemy heterogeniczne
 - BC: Otwarty standard komunikacji
 - Jeden punkt dostępu dla użytkowników i innych systemów
 - BC: Federacja Bibliotek Cyfrowych
 - Skoordynowane zarządzanie zasobami
 - BC: Koordynacja digitalizacji, automatyczne zapobieganie powstawaniu duplikatów
 - Wykorzystanie usług wspólnych
 - BC: Sieciowy profil użytkownika



- Najistotniejsze cechy infrastruktury bibliotek cyfrowych
 - Architektura silnie rozproszona
 - BC: 38 instalacji bibliotek cyfrowych
 - Systemy heterogeniczne
 - BC: Otwarty standard komunikacji
 - Jeden punkt dostępu dla użytkowników i innych systemów
 - BC: Federacja Bibliotek Cyfrowych
 - Skoordynowane zarządzanie zasobami
 - BC: Koordynacja digitalizacji, automatyczne zapobieganie powstawaniu duplikatów
 - Wykorzystanie usług wspólnych
 - BC: Sieciowy profil użytkownika

Liczba obiektów cyfrowych

✓ Ok. 170 tysięcy

Liczba aktywnych bibliotek cyfrowych:

✓ 13 regionalnych

✓ 25 instytucjonalnych

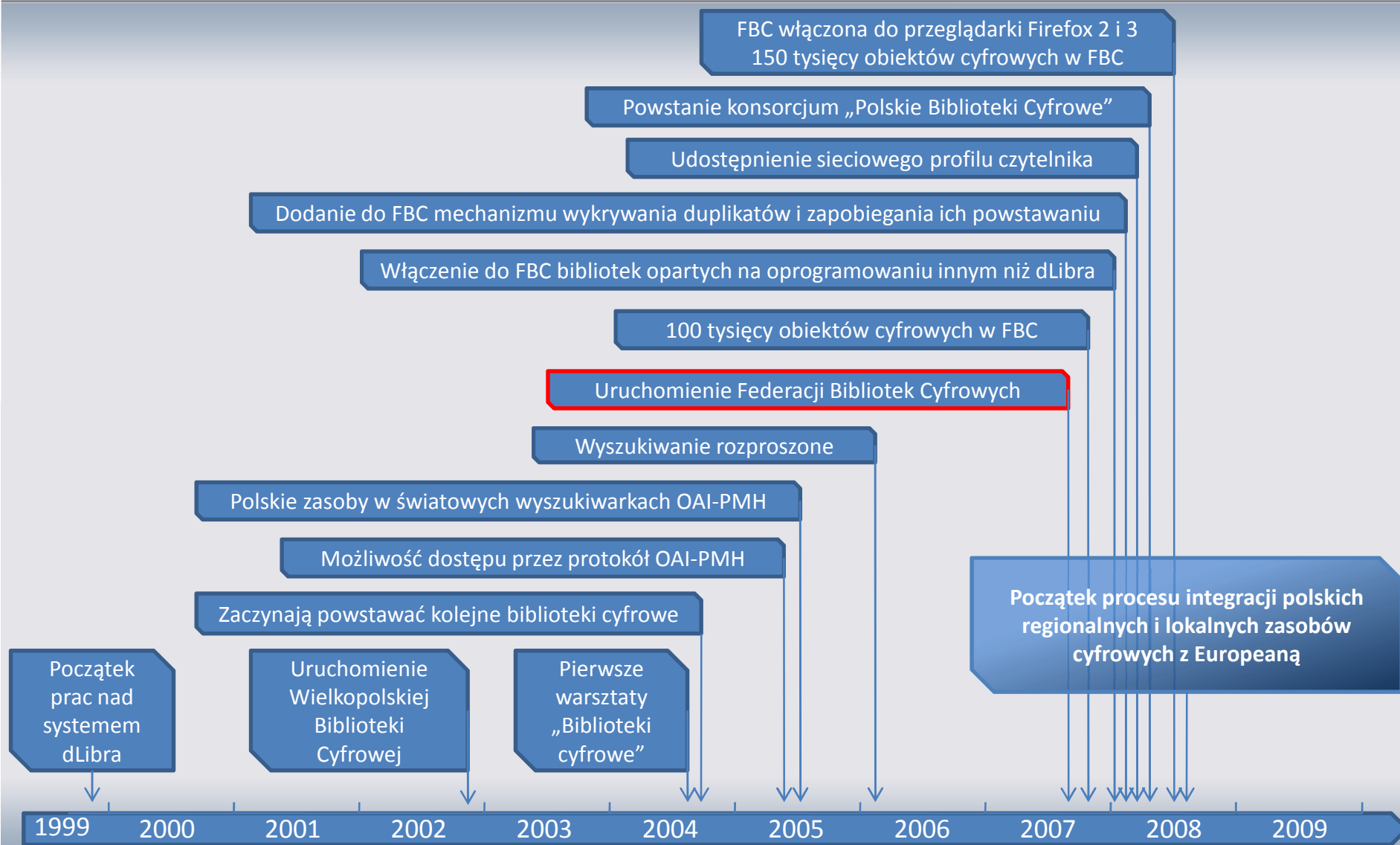
Ok. 200 instytucji zaangażowanych
w zarządzanie zawartością

łącznie : **38 instalacji**





- Najistotniejsze cechy infrastruktury bibliotek cyfrowych
 - Architektura silnie rozproszona
 - BC: 38 instalacji bibliotek cyfrowych
 - Jeden punkt dostępu do zasobów i usług dla użytkowników i innych systemów
 - BC: Federacja Bibliotek Cyfrowych
 - Systemy heterogeniczne
 - BC: Otwarty standard komunikacji
 - Skoordynowane zarządzanie zasobami
 - BC: Koordynacja digitalizacji, automatyczne zapobieganie powstawaniu duplikatów
 - Wykorzystanie usług wspólnych
 - BC: Sieciowy profil użytkownika

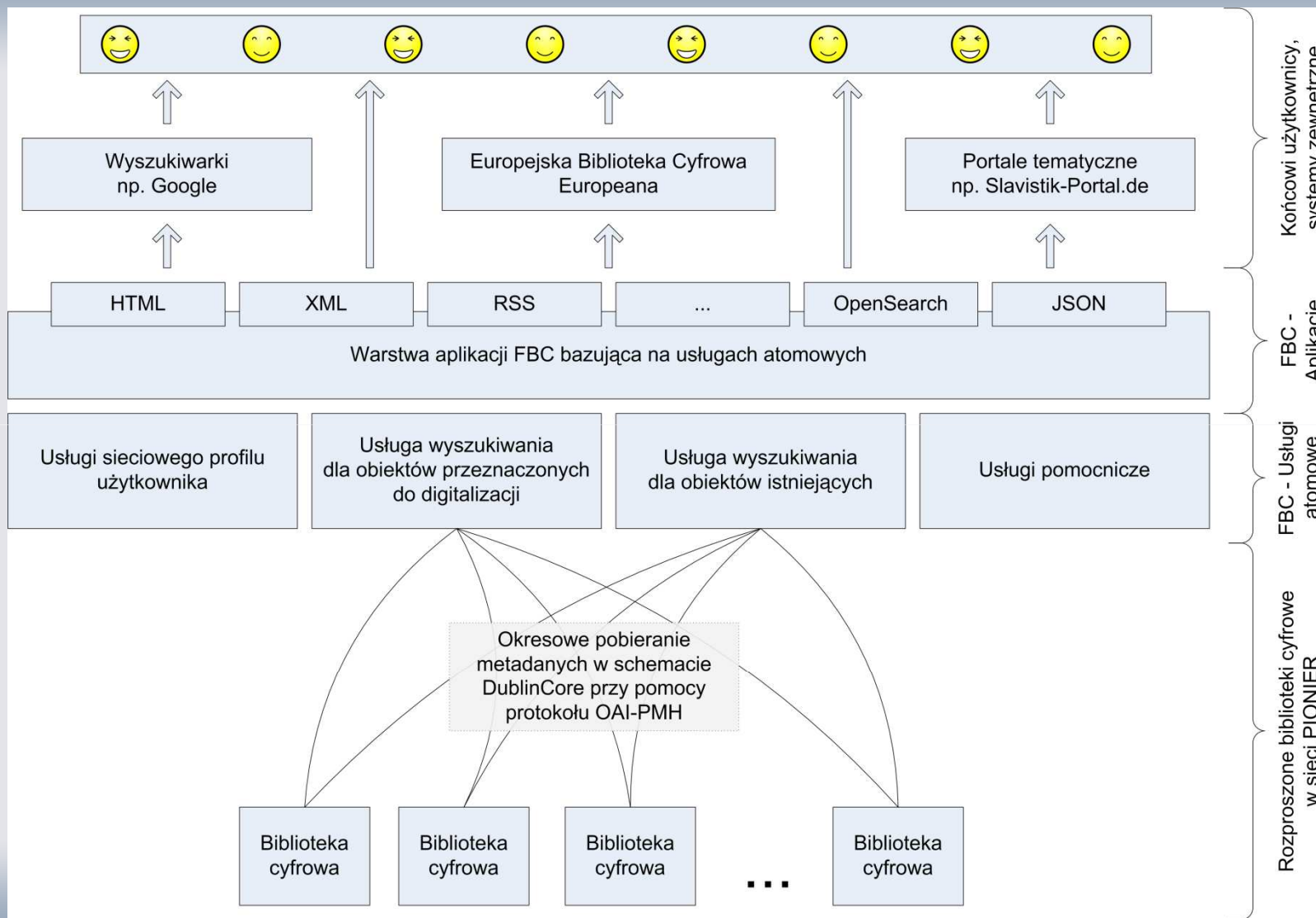


- Cele Federacji
 - **Ułatwienie wykorzystania** zasobów polskich bibliotek cyfrowych i repozytoriów
 - **Zwiększenie widoczności i popularyzacja** zasobów polskich bibliotek cyfrowych i repozytoriów w Internecie
 - **Udostępnienie** użytkownikom Internetu i twórcom bibliotek cyfrowych **nowych, zaawansowanych usług sieciowych** opartych na zasobach polskich bibliotek cyfrowych i repozytoriów



- Podstawowe założenia
 - **Nie ma** potrzeby i konieczności **przekazywania** swoich **zasobów na rzecz FBC**
 - **Nie ma opłat** za korzystanie z FBC i przynależność do FBC
 - **Podstawą** działania są **otwarte standardy**
 - Możliwość użycia różnych rozwiązań technicznych przez poszczególne biblioteki cyfrowe

- Podstawowe funkcje
 - Przeszukiwanie dostępnych publikacji cyfrowych
 - Proste
 - Zaawansowane
 - Plany digitalizacji
 - Przeszukiwanie
 - Zestawienie (raport)
 - Lokalizowanie publikacji cyfrowych na podstawie identyfikatora OAI
- Informacje w FBC aktualizowane są codziennie





- Najistotniejsze cechy infrastruktury bibliotek cyfrowych
 - Architektura silnie rozproszona
 - BC: 38 instalacji bibliotek cyfrowych
 - Jeden punkt dostępu do zasobów i usług dla użytkowników i innych systemów
 - BC: Federacja Bibliotek Cyfrowych
 - Systemy heterogeniczne
 - BC: **Otwarty standard komunikacji**
 - Skoordynowane zarządzanie zasobami
 - BC: Koordynacja digitalizacji, automatyczne zapobieganie powstawaniu duplikatów
 - Wykorzystanie usług wspólnych
 - BC: Sieciowy profil użytkownika



- Udział w FBC wymaga używania otwartych standardów – podstawą jest uznany na świecie protokół OAI-PMH
- Do stycznia 2008 protokół ten w Polsce stosowały tylko biblioteki cyfrowe oparte na systemie dLibra
- W 2008 roku protokół zaczęły implementować kolejne instytucje:
 - **Biblioteka Cyfrowa Politechniki Krakowskiej**
(w FBC od 06.02.2008 r.)
 - **Akademicka Biblioteka Cyfrowa KRAKÓW**
(w FBC od 21.05.2008 r.)
 - **ICM – DIR – Zasoby Polskie**
(w FBC od 25.07.2008 r.)
 - **Biblioteka Cyfrowa Archiwum Diecezjalnego w Tarnowie**
(w FBC od 22.10.2008 r.)



- **Najistotniejsze cechy infrastruktury bibliotek cyfrowych**
 - Architektura silnie rozproszona
 - BC: 38 instalacji bibliotek cyfrowych
 - Jeden punkt dostępu do zasobów i usług dla użytkowników i innych systemów
 - BC: Federacja Bibliotek Cyfrowych
 - Systemy heterogeniczne
 - BC: Otwarty standard komunikacji
 - **Skoordynowane zarządzanie zasobami**
 - **BC: Koordynacja digitalizacji, automatyczne zapobieganie powstawaniu duplikatów**
 - Wykorzystanie usług wspólnych
 - BC: Sieciowy profil użytkownika

- Wsparcie poprzez
 - Przeszukiwanie planów digitalizacji
 - Raport z planami digitalizacji
 - Raport z listą i zestawieniem ilościowym potencjalnie istniejących już duplikatów
 - Zapobieganie powstawaniu duplikatów na etapie planowania digitalizacji
 - FBC umożliwia automatyczne sprawdzenie czy znajdują się w niej (zdigitalizowane bądź planowane) publikacje będące potencjalnymi duplikatami publikacji o podanym opisie

- Zapobieganie powstawaniu duplikatów na etapie planowania digitalizacji
 1. Redaktor biblioteki cyfrowej przygotowuje opis publikacji przeznaczonej do digitalizacji
 2. System biblioteki cyfrowej sprawdza ten opis w FBC – określa w ten sposób, czy planowana publikacja może już być przeznaczona do digitalizacji/zdigitalizowana w innym ośrodku
 3. Redaktor dostaje informację zwrotną i na tej podstawie zatwierdza publikację jako planowaną bądź też usuwa ją z planów digitalizacji



- **Najistotniejsze cechy infrastruktury naukowej**
 - Architektura silnie rozproszona
 - BC: 38 instalacji bibliotek cyfrowych
 - Jeden punkt dostępu do zasobów i usług dla użytkowników i innych systemów
 - BC: Federacja Bibliotek Cyfrowych
 - Systemy heterogeniczne
 - BC: Otwarty standard komunikacji
 - Skoordynowane zarządzanie zasobami
 - BC: Koordynacja digitalizacji, automatyczne zapobieganie powstawaniu duplikatów
 - Wykorzystanie usług wspólnych
 - **BC: Sieciowy profil użytkownika**

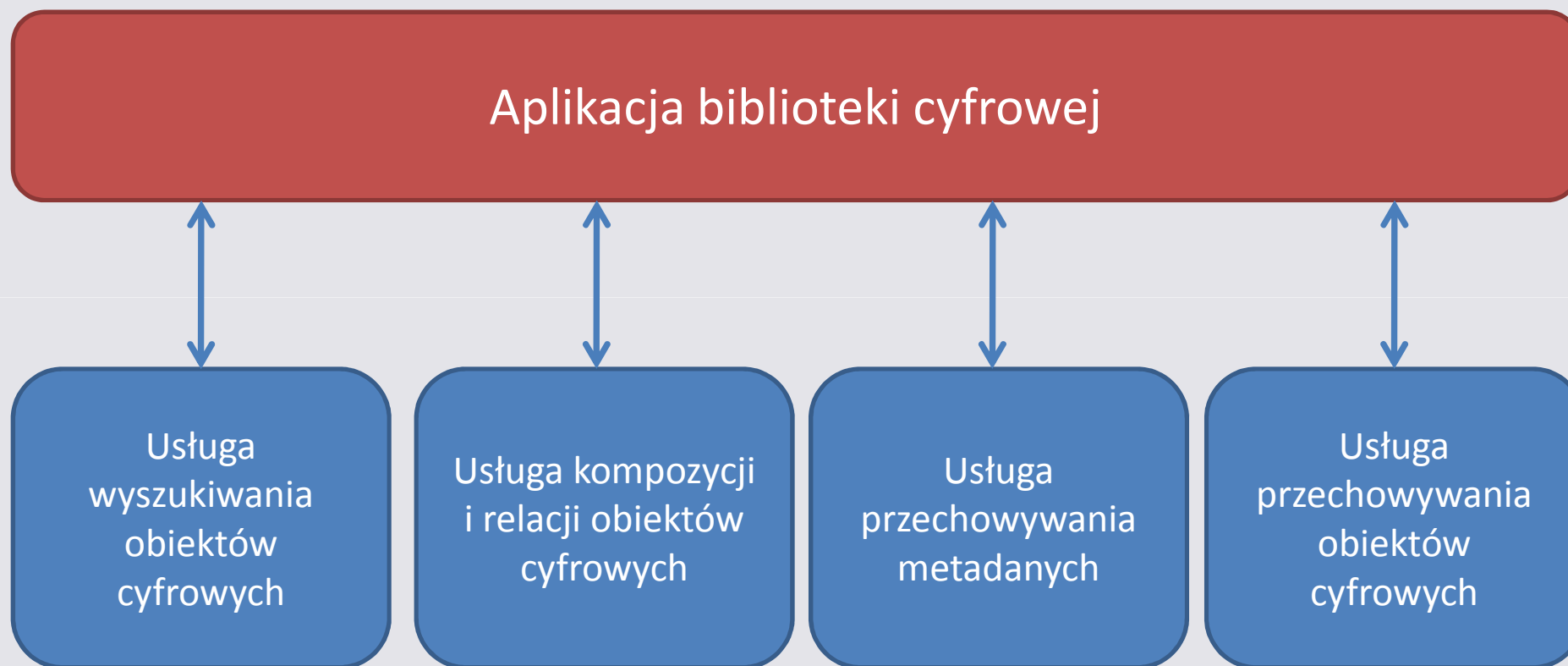


- Biblioteki należące do FBC „ufają sobie” i mogą w bezpieczny sposób wymieniać się informacjami o profilach czytelników
- Czytelnicy mogą przy pomocy konta w jednej bibliotece cyfrowej logować się do innych bibliotek cyfrowych Federacji
- Czytelnicy posiadają dzięki temu tylko jedno hasło i we wszystkich bibliotekach mają do dyspozycji jednolitą informację zapisaną w swoim profilu
 - Np. jedna lista ulubionych publikacji zawierająca obiekty z wielu bibliotek cyfrowych

- Realizacja techniczna
 - Opiera się na projekcie Shibboleth rozwijanym w ramach programu Internet2 w USA oraz podobnym projekcie Guanxi realizowanym w Europie
 - Podstawą jest otwarty protokół komunikacyjny SAML
 - Model Shibboleth został w ramach opracowywania sieciowego profilu czytelnika rozszerzony o dodatkowe funkcje związane z profilem użytkownika
 - Szczegóły prezentowano na konferencjach naukowych
 - Pełna specyfikacja mechanizmu profilu sieciowego będzie wkrótce opublikowana jako otwarta na stronach Zespołu Bibliotek Cyfrowych PCSS oraz FBC

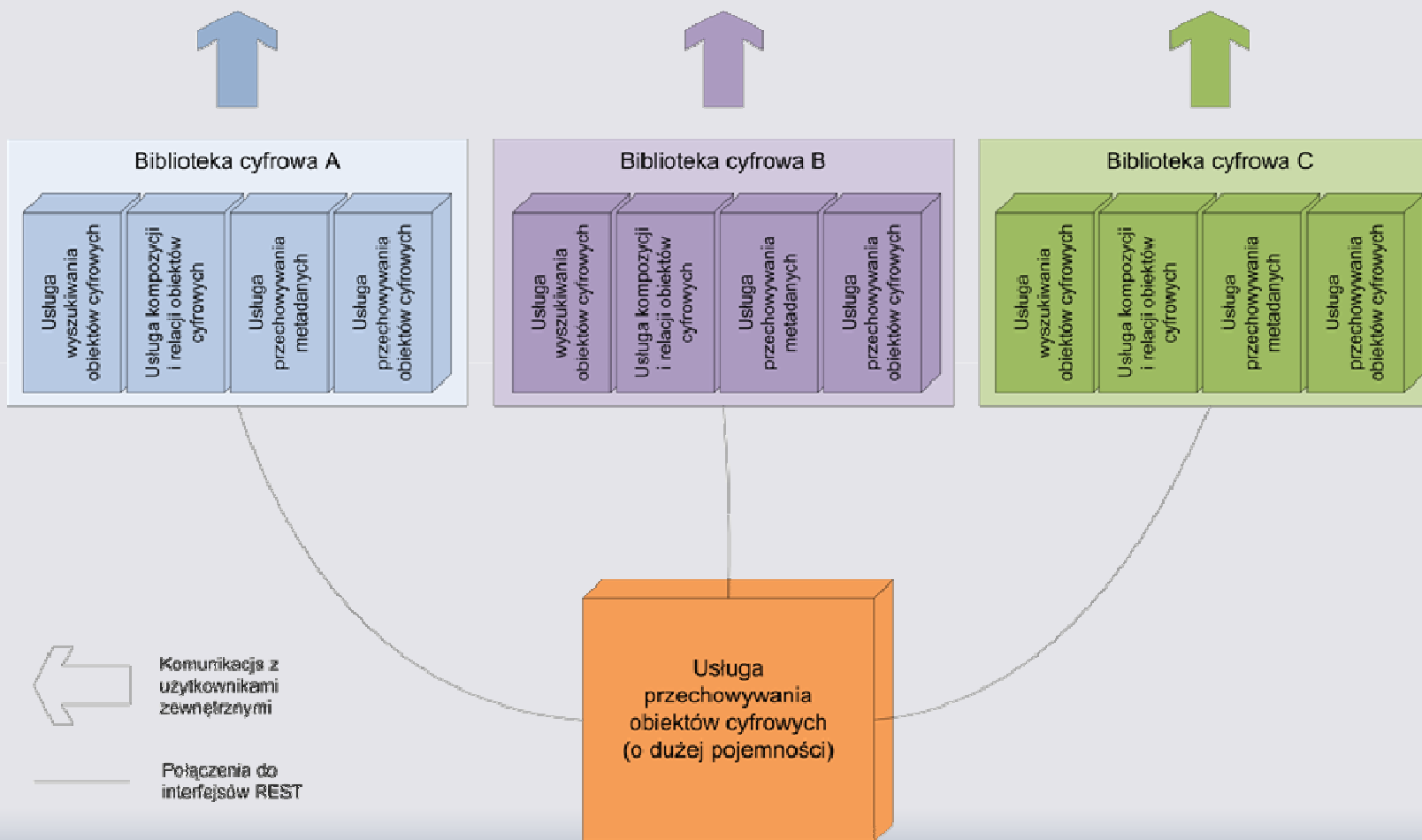
- Podstawą realizacji zaawansowanych scenariuszy w środowisku rozproszonych bibliotek cyfrowych było zdefiniowanie w tym środowisku **usług atomowych**
- Grant badawczy MNiSW „Mechanizmy usług atomowych dla rozproszonych bibliotek cyfrowych” (07.2006-06.2008)
- Cel grantu:
 - Analiza funkcjonalności bibliotek cyfrowych
 - Podział funkcjonalny bibliotek cyfrowych na atomowe usługi
 - Opracowanie modelu danych i protokołów dostępowych dla poszczególnych usług
 - Prototypowe implementacje i testy

- Usługa przechowywania obiektów cyfrowych
 - Przechowywanie i wersjonowanie obiektów cyfrowych dowolnego typu
- Usługa przechowywania metadanych opisujących obiekty cyfrowe
- Usługa kompozycji i relacji obiektów cyfrowych
 - Tworzenie relacji pomiędzy obiektami cyfrowymi oraz tworzenie grup powiązanych obiektów
- Usługa wyszukiwania obiektów cyfrowych
 - Przeszukiwanie zbioru obiektów cyfrowych danego typu lub zbioru metadanych o określonym formacie



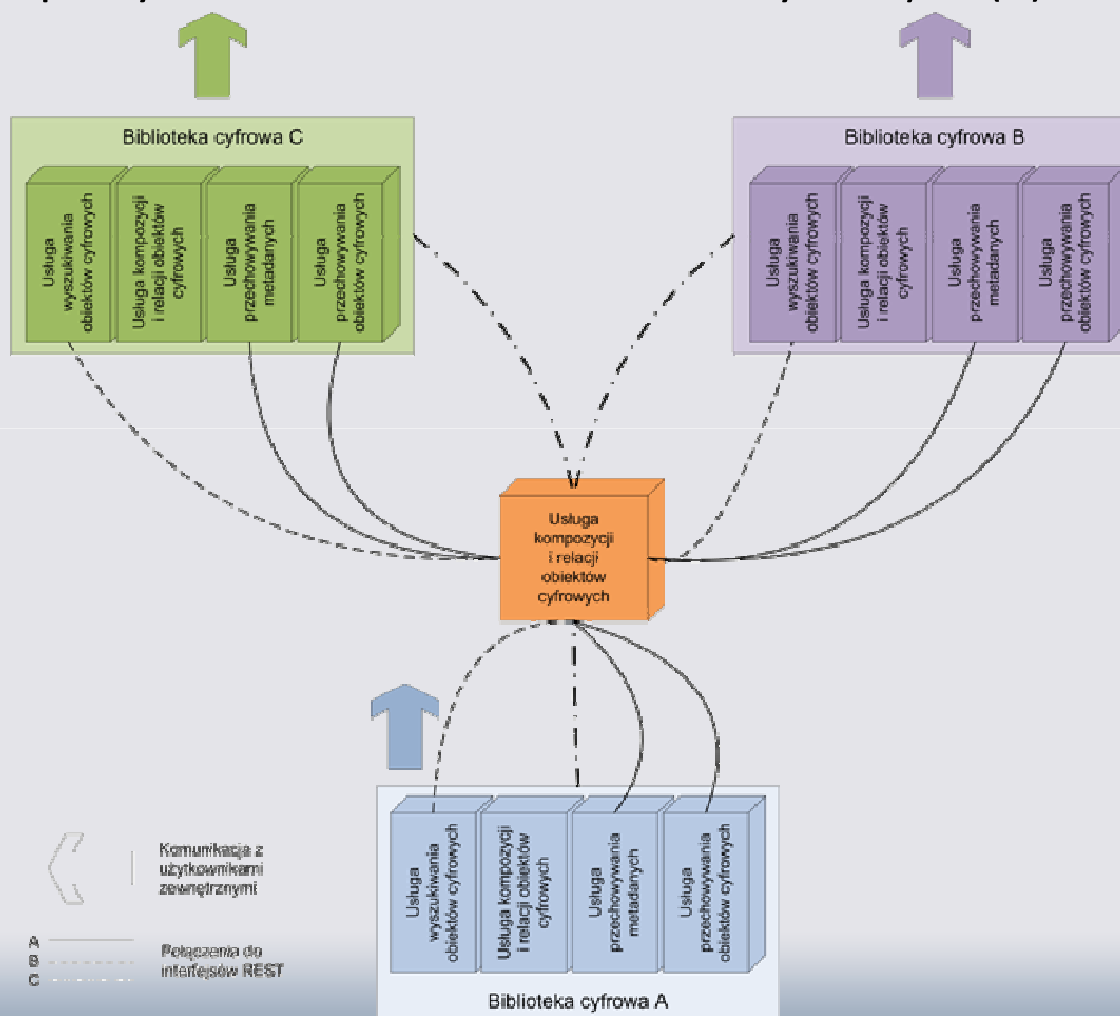
- Model danych oraz bazujące na nim interfejsy, zaprojektowane są zgodnie z podejściem ROA (*ang. Resource-oriented architecture*).
- Mechanizm wyszukiwania usług wykorzystuje rejestr semantyczny zrealizowanym z użyciem ontologii OWL-S
 - Umożliwia on prowadzenie rejestru instancji usług atomowych i udostępnienie ich semantycznego opisu, który jest następnie podstawą do wyszukiwania usług spełniających parametry zadane przez aplikacje korzystające z tych usług.

- Scenariusze pracy w infrastrukturze bibliotek cyfrowych (1)

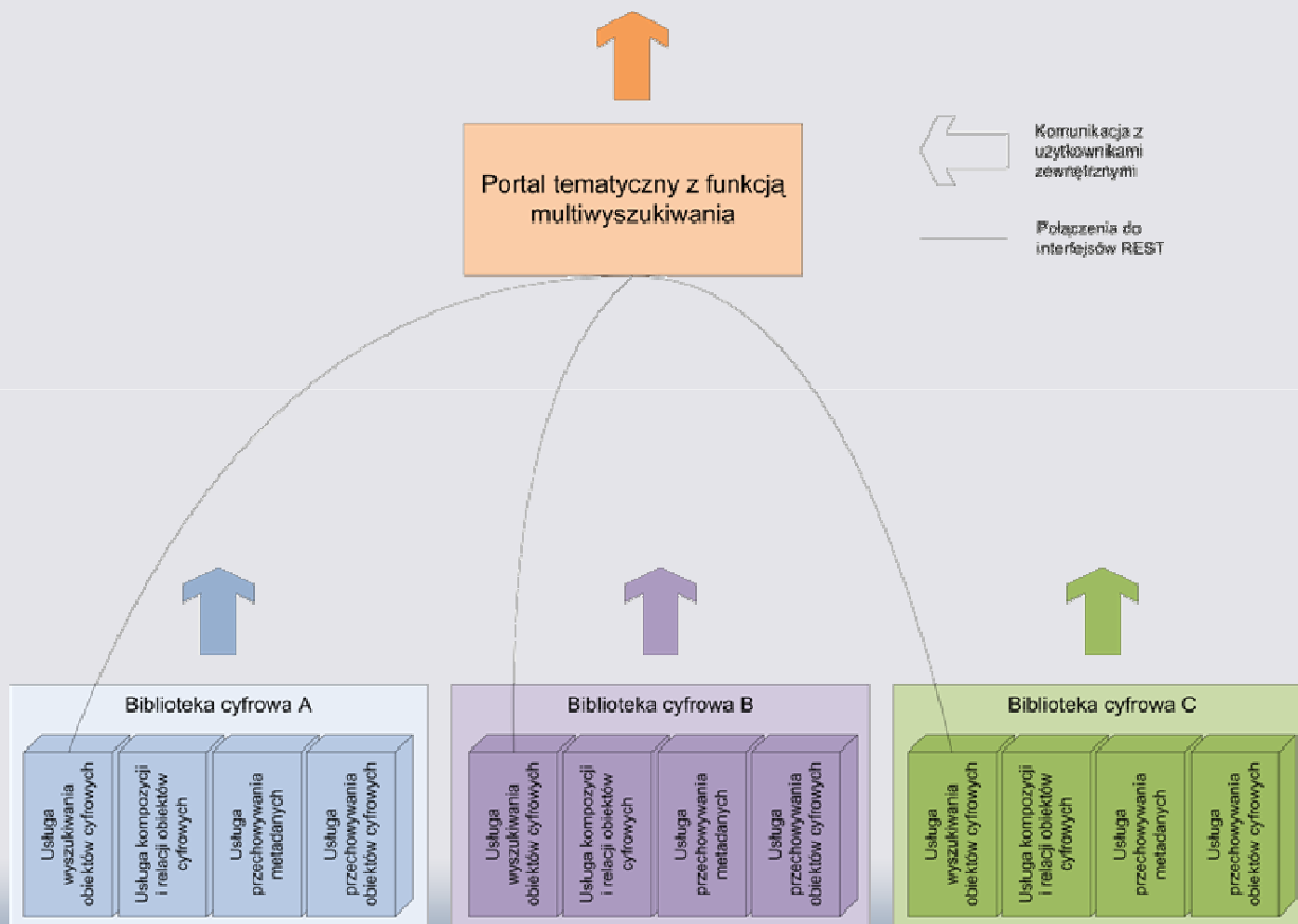




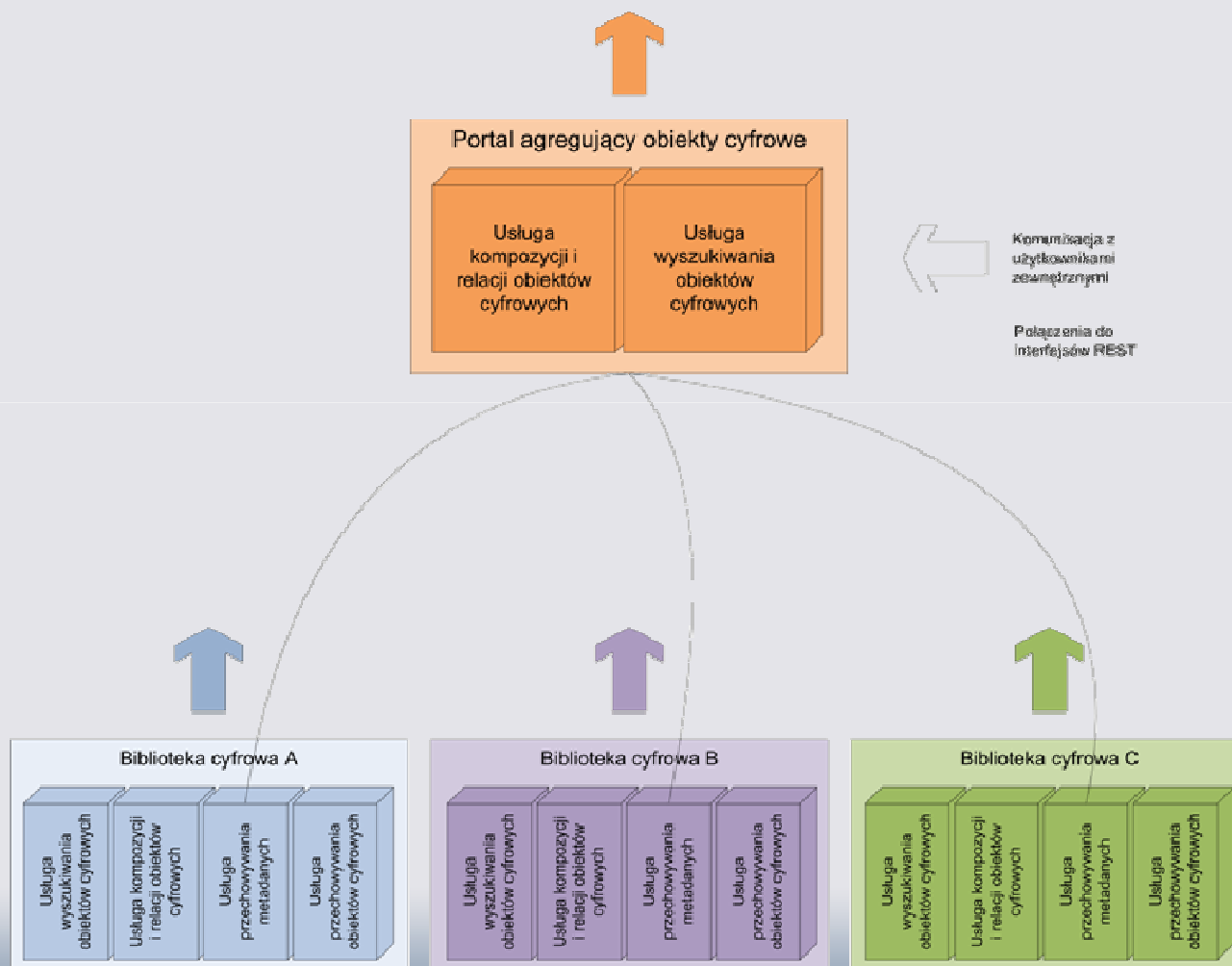
- Scenariusze pracy w infrastrukturze bibliotek cyfrowych (2)



- Scenariusze pracy w infrastrukturze bibliotek cyfrowych (3)



- Scenariusze pracy w infrastrukturze bibliotek cyfrowych (4)



- Zawartość bibliotek cyfrowych stanowi przedmiot badań wielu naukowców nauk humanistycznych na świecie
- Polski model realizacji infrastruktury bibliotek cyfrowych pozwala na **synergiczny rozwój zasobów cyfrowych oraz usług**, które są niezbędne do pracy w ramach nowoczesnego warsztatu naukowego osadzonego w infrastrukturze informatycznej nauki
- Wyłącznie cechą tego składnika infrastruktury informatycznej nauki jest również społeczne znaczenie zgromadzonych zasobów informacyjnych, a w przyszłości również powszechne wykorzystanie tych narzędzi i usług



DZIĘKUJĘ ZA UWAGĘ !

Infrastruktura bibliotek cyfrowych w sieci PIONIER

C. Mazurek, M.Stroiński, M.Werla, J.Węglarz

Poznańskie Centrum Superkomputerowo-Sieciowe