



Europeana Cloud: Wykorzystanie technologii chmurowych do współdzielenia on-line baz danych dziedzictwa kulturowego

Cezary Mazurek, Aleksandra Nowak, Maciej Stroiński, Marcin Werla
Poznańskie Centrum Superkomputerowo-Sieciowe

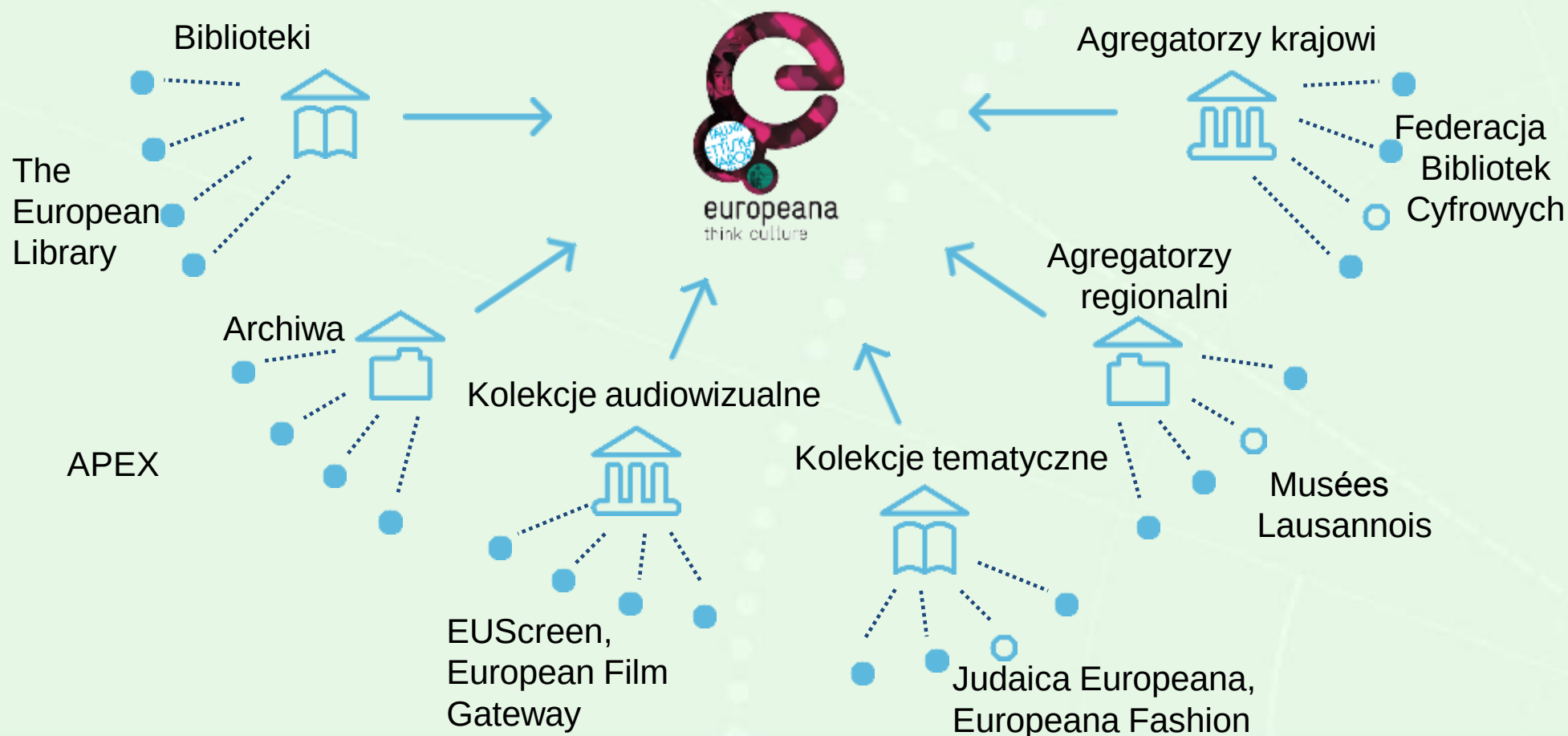
Agregacja jako model współdzielenia danych

Wprowadzenie

- Model działania wyrastający z protokołu OAI-PMH
 - podział na dostawców danych i dostawców usług
- Z czasem coraz bardziej wielowarstwowy
 - Dostawcy usług coraz częściej stają się równocześnie dostawcami danych, które agregują
- Przykłady:
 - Federacja Bibliotek Cyfrowych
 - Deutsche Digitale Bibliothek
 - Europeana
 - Digital Public Library of America

Agregacja jako model współdzielenia danych

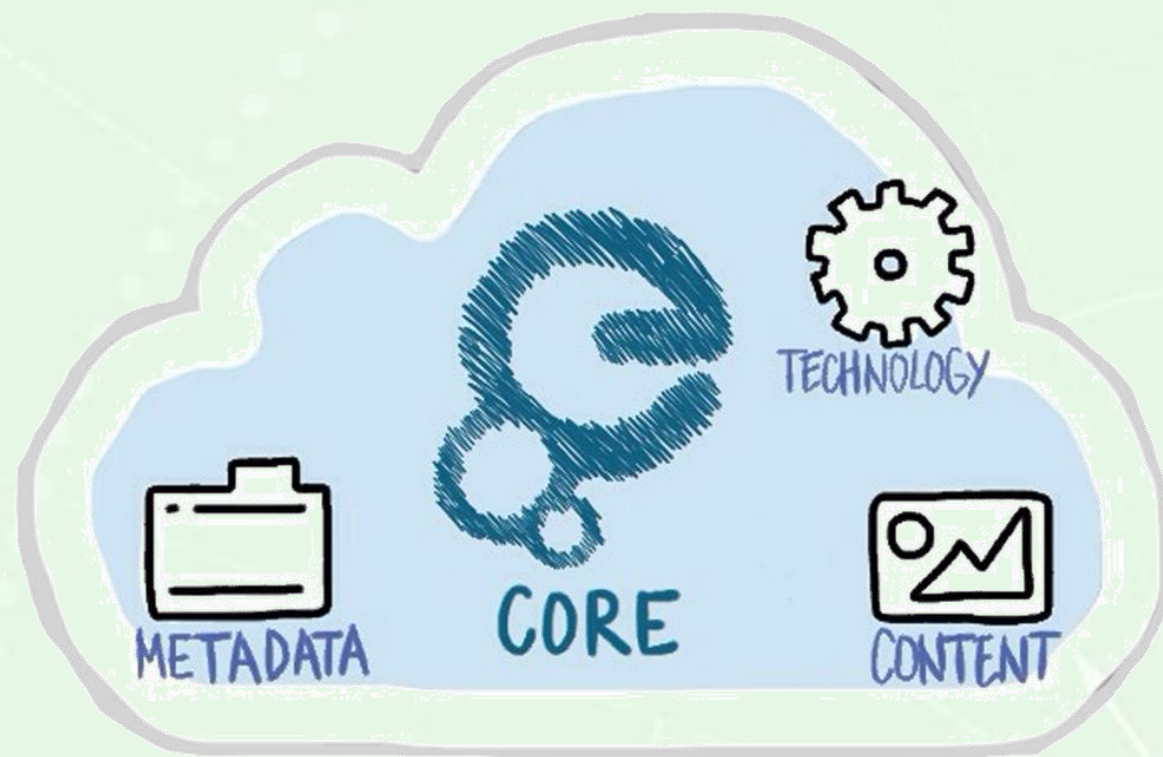
Stan obecny



Strategia Europeany na lata 2015-2020

Plany działania

- 3 główne priorytety:
 - Poprawa jakości danych
 - Otwarcie danych
 - Tworzenie wartości dla partnerów
- Proces „od portalu do platformy”
- Kluczowym zadaniem jest budowa rdzenia, który będzie dostarczał funkcjonalność niezbędną do składowania danych i zarządzania nimi
- Rdzeń powinien być otwartą podstawą do efektywnego budowania dalszych usług



Dostęp on-line, również w języku polskim:

<http://strategy2020.europeana.eu/>

Europeana Cloud

Opis projektu

<http://pro.europeana.eu/web/europeana-cloud>

- 3 letni projekt finansowany przez Komisję Europejską w ramach programu CIP ICT-PSP
 - Luty 2014 – Styczeń 2016
 - Budżet: 4,75 mln EUR
- Koordynator: The European Library
- 35 partnerów z 16 krajów UE
- Polski partner: Poznańskie Centrum Superkomputerowo-Sieciowe - główny partner technologiczny w obszarze zbudowania współdzielonego, chmurowego systemu przechowywania metadanych i treści
 - Odpowiedzialny za zaprojektowanie i rozwój architektury systemu
 - Uczestniczący w rozwoju systemu
 - Odpowiedzialny za hosting systemu (jedno z 3 data center)
 - Zaangażowany w zdefiniowanie zasad operacyjnego utrzymania systemu po projekcie



Europeana Cloud

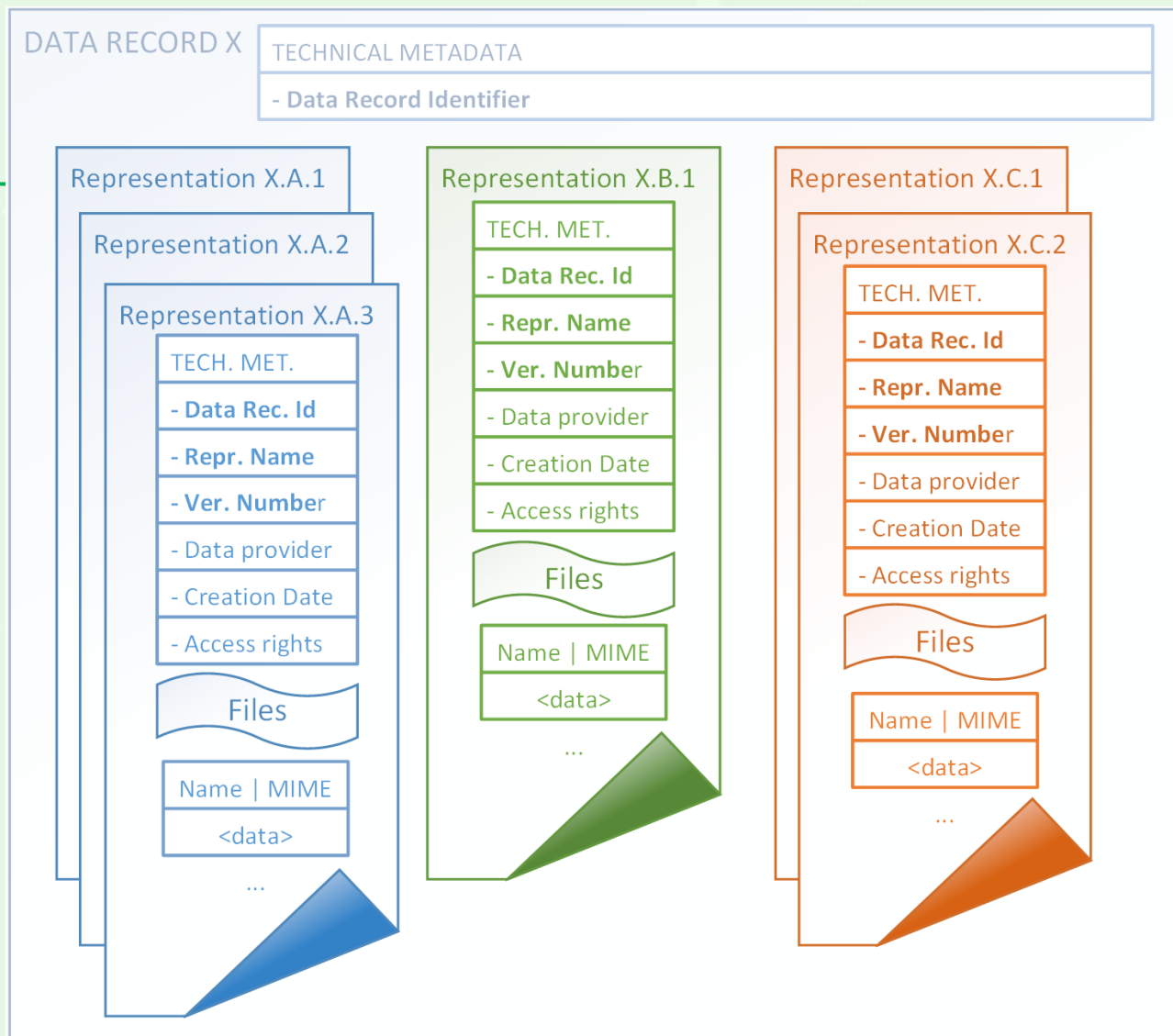
Zakres funkcjonalny systemu

- Możliwość tworzenia globalnie unikalnych identyfikatorów dla rekordów danych kulturowych z różnych źródeł
- Możliwość składowania i dostępu do rekordów danych kulturowych, składających się ze strumieni danych i metadanych w wielu różnych formatach i wersjach
- Możliwość adnotowania rekordów danych kulturowych
- Możliwość śledzenia zmian w rekordach danych kulturowych
- Możliwość przetwarzania rekordów danych kulturowych
- Wszystkie powyższe funkcje powinny być zrealizowane w bezpieczny, niezawodny i skalowalny sposób
 - Plan: wykorzystanie systemu Europeana Cloud jako infrastruktury dla aplikacji i systemów informacyjnych związanych z dziedzictwem kulturowym w skali Europy

Europeana Cloud

Model danych

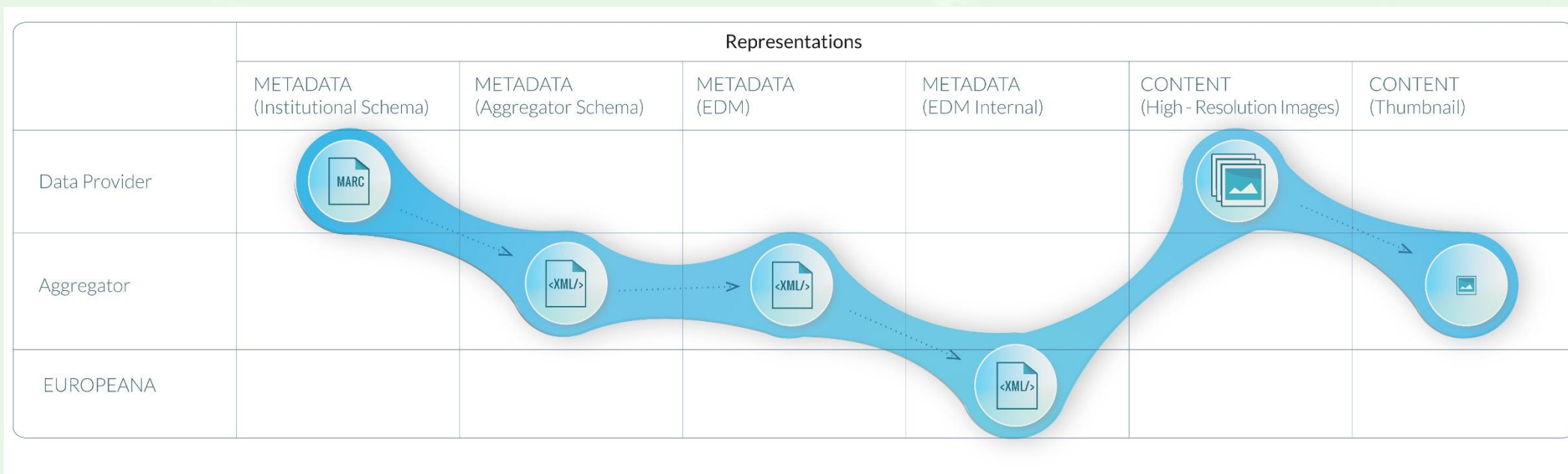
- Podstawowa jednostka: Rekord danych
- Rekord dzieli się na reprezentacje
- Każda reprezentacja może mieć wiele wersji
- Każda wersja może składać się z jednego lub więcej plików
- Otwartość techniczna:
 - Uprawnienia definiowane są na poziomie wersji
 - Dostęp ograniczać można tylko do plików zawartych w danej wersji – metadane techniczne o strukturze rekordu danych są otwarte



Europeana Cloud

Model danych

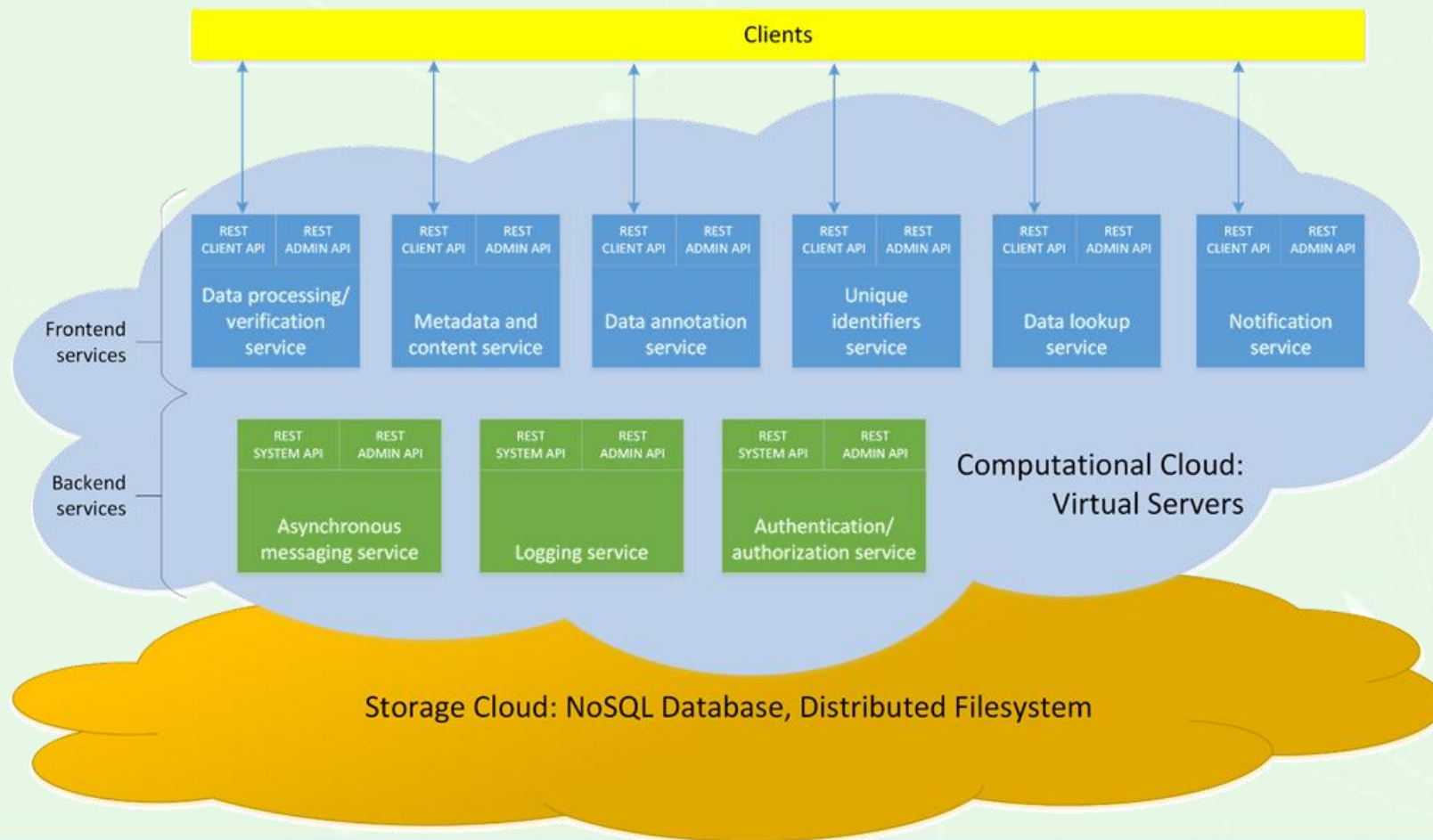
- Model danych został zaprojektowany tak, aby odzwierciedlać logikę procesu agregacji i realizowanego w ramach agregacji przetwarzania danych



Europeana Cloud

Architektura usługowa

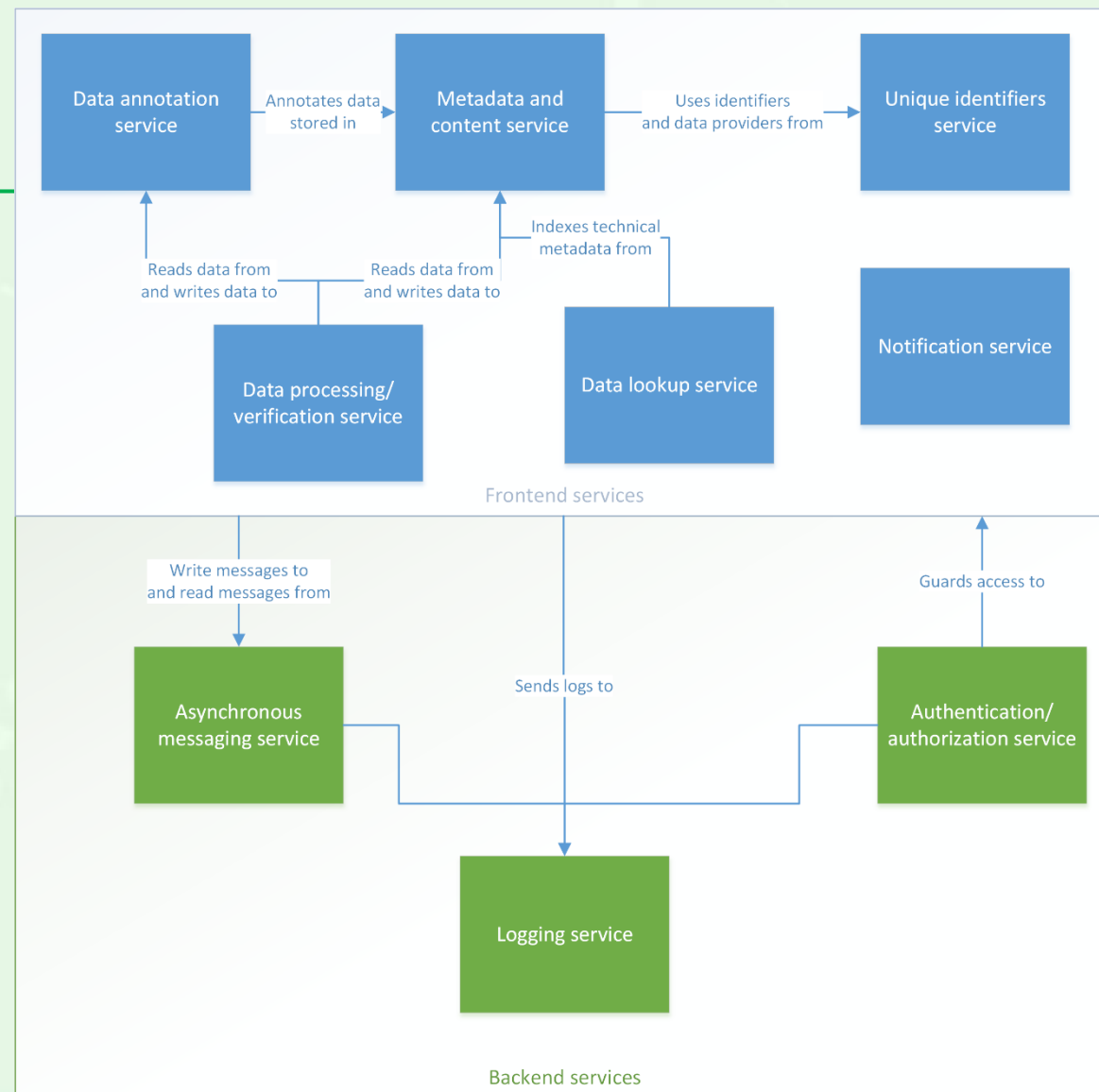
- Z punktu widzenia użytkownika końcowego Europeana Cloud będzie zestawem usług dostępnych w modelu SaaS
- Usługa realizowana będzie w środowisku chmurowym
- Dwie chmury:
 - Składowanie danych
 - Obliczenia
- Dwie warstwy usług:
 - Zewnętrzna (6 usług)
 - Wewnętrzna (3 usługi)



Europeana Cloud

Podstawowe zależności pomiędzy usługami

- Początkowo 3 data center
- Technicznie również możliwość skalowania systemu z chmury prywatnej na chmurę publiczną (Amazon itp.)
- Model biznesowy funkcjonowania systemu jest w trakcie opracowania
 - Finansowanie z programu CEF Europeana
 - Składki krajów członkowskich UE
 - Składki użytkowników systemu



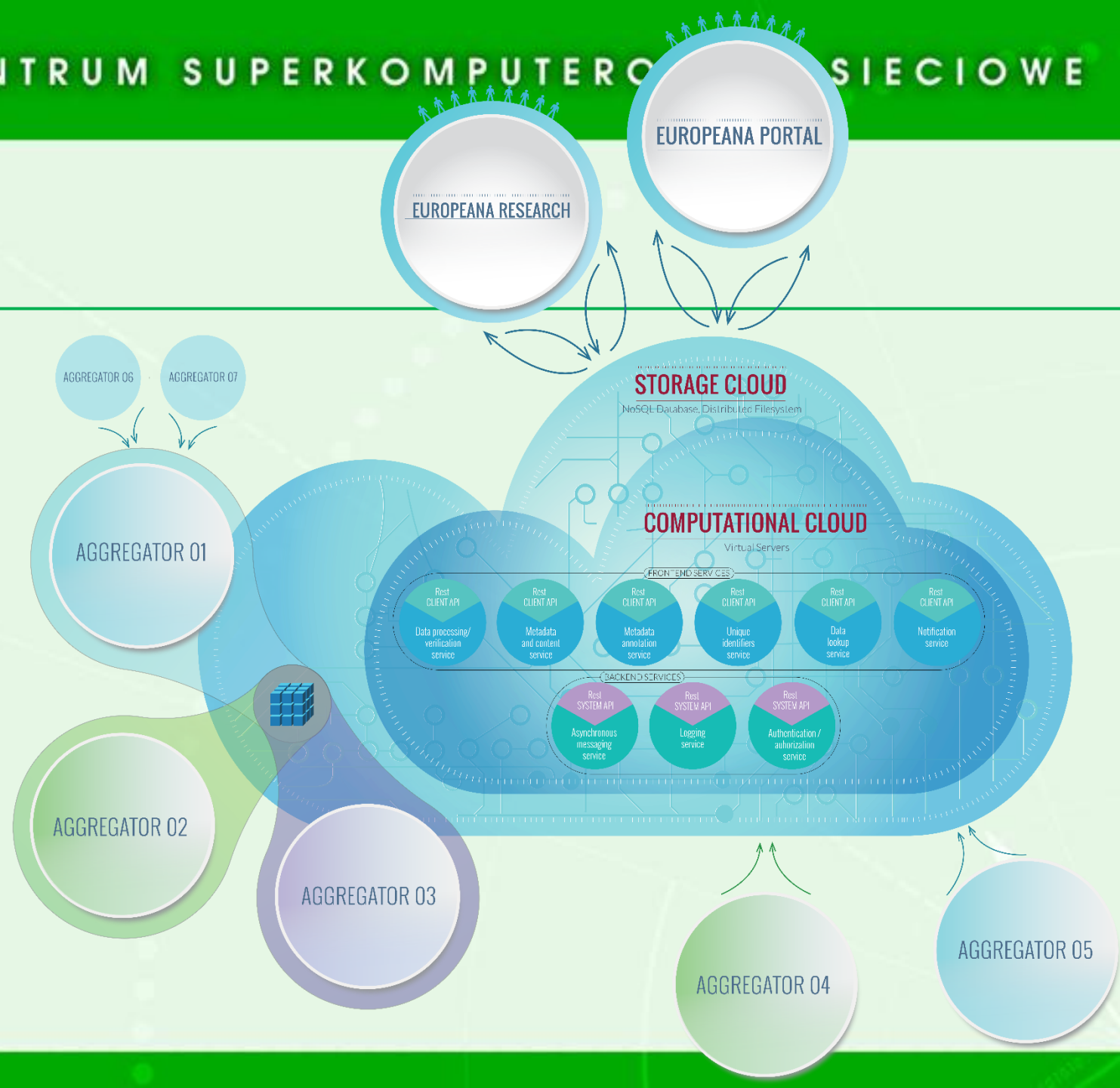
Cyfrowa humanistyka w Polsce i Europie

- Jednym z partnerów projektu Europeana Cloud jest DARIAH ERIC – Cyfrowa Infrastruktura Badawcza dla Nauk Humanistycznych
 - Polskie konsorcjum DARIAH jest obecnie zawiązywane
 - Koordynatorem jest DHLab Uniwersytetu Warszawskiego, a wśród członków założycieli są m.in. PCSS oraz IBL PAN
 - Plany działania zakładają przystąpienie Polski do DARIAH jeszcze w tym roku
- W ramach projektu Europeana Cloud powstaje również platforma Europeana Research, która ma umożliwiać rozwój narzędzi wykorzystujących dane zebrane w Europeanie
- Kluczowym elementem z punktu widzenia rozwoju tych narzędzi jest zwiększenie interoperacyjności danych

Europeana Cloud

Wizja efektu projektu

- Chmurowy rdzeń Europeany pozwalający agregatorom metadanych współdzielić dane i korzystać z chmury jako podstawowego miejsca składowania danych
- Plan realizacji:
 - 2014-2016 – Rozwój systemu w ramach projektu Europeana Cloud, główni użytkownicy: Europeana, TEL, FBC
 - 2016-2018 – Stopniowe włączanie do Europeana Cloud kolejnych agregatorów metadanych



Możliwości rozwoju narzędzi humanistyki cyfrowej

Interoperacyjność treści w Europeana Cloud

- Obiekt cyfrowy zawierający skany starodruku wraz z metadanymi jest udostępniany w Wielkopolskiej Bibliotece Cyfrowej (WBC).
- Metadane obiektu są pobierane (kopiowane) przez FBC (operacja wykonywana raz na dobę) i składowane w systemie Europeana Cloud.
- Europeana jest automatycznie powiadamiana o nowych danych z FBC i niezwłocznie aktualizuje swoje indeksy na podstawie danych już dostępnych w Europeana Cloud.
- Użytkownik końcowy – naukowiec – poszukując w Europeanie materiałów do pracy odnajduje rekord metadanych obiektu udostępnionego w WBC.
- Użytkownik przechodzi z Europeany poprzez FBC do stron WBC. Tam wyświetla mu się strona WWW z opisem obiektu cyfrowego oraz możliwość wyświetlenia i pobrania obiektu.
- Po zapoznaniu się z treścią obiektu użytkownik rozpoczyna pracę z nim w Wirtualnym Laboratorium Transkrypcji (<http://wlt.synat.pcss.pl>).
 - Dzięki dostępności danych opisujących obiekt w Europeanie Cloud, system WLT jest w stanie automatycznie dokonać importu i konwersji obiektu do pożądanego formatu.
- Po zakończonej pracy użytkownik eksportuje wyniki transkrypcji na swój lokalny dysk w celu prowadzenia dalszych prac badawczych.
 - Ponadto wyniki te (za zgodą użytkownika) są również zapisane jako nowa reprezentacja obiektu w Europeana Cloud, dzięki czemu mogą być automatycznie użyte przez inne narzędzia i systemy. Publiczna dostępność tych danych zależy od decyzji użytkownika.



**Dziękuję za uwagę!
Pytania?**

Aleksandra Nowak
anowak@man.poznan.pl