

# **Projekt DRIVER:** *Wizja infrastruktury repozytoriów cyfrowych dla europejskich badań naukowych*

Alek Tarkowski, ICM UW



W prezentacji zostały wykorzystane materiały informacyjne projektu DRIVER – w szczególności fragmenty prezentacji Norberta Lossau, Sophii Jones, Paolo Manghiego, Natalii Manoli, Wolframa Horstmanna oraz Michele Artini z zespołem.

Wszystkie prezentacje są dostępne na stronie:

<http://www.driver-support.eu/linkspubs/driverslides.html>

# Wymogi komunikacji naukowej

1. **Wyczerpujący, globalny** dostęp do wszelkiej informacji naukowej
2. **Minimalny** czas i wysiłek
3. **Łatwe** wyszukiwanie, opracowywanie i dalsza dystrybucja informacji
4. **Maksymalna** widoczność w społeczności
5. **Długoterminowy** dostęp i przechowywanie wyników badań

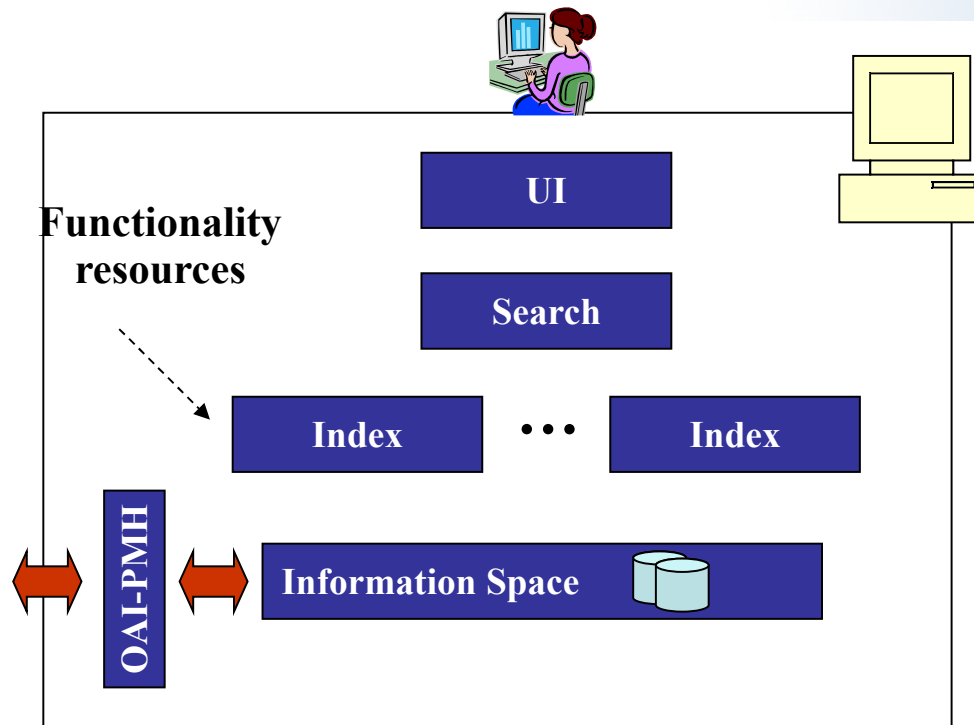
# Wizja europejskiej e-infrastruktury

“Obszar e-Infrasktruktury obejmuje szereg połączonych zagadnień związanych z wyłanianiem się nowego środowiska badawczego, w którym “wirtualne społeczności” dzielą się i wykorzystują zbiorowy potencjał europejskiego zaplecza naukowego i inżynieryjnego.

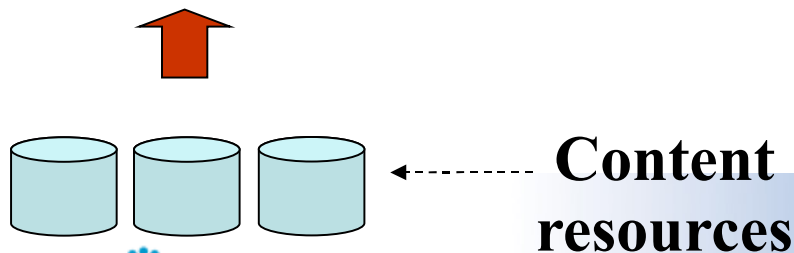
Celem zadania “e-Infrastruktury oparte na ICT” jest stworzenie we wszystkich obszarach wiedzy i technologii nowych infrastruktur badawczych, potrzebnych europejskiemu środowisku badawczemu i pomagające przemysłowi wzmocnić bazę wiedzy i technologicznego know-how.”



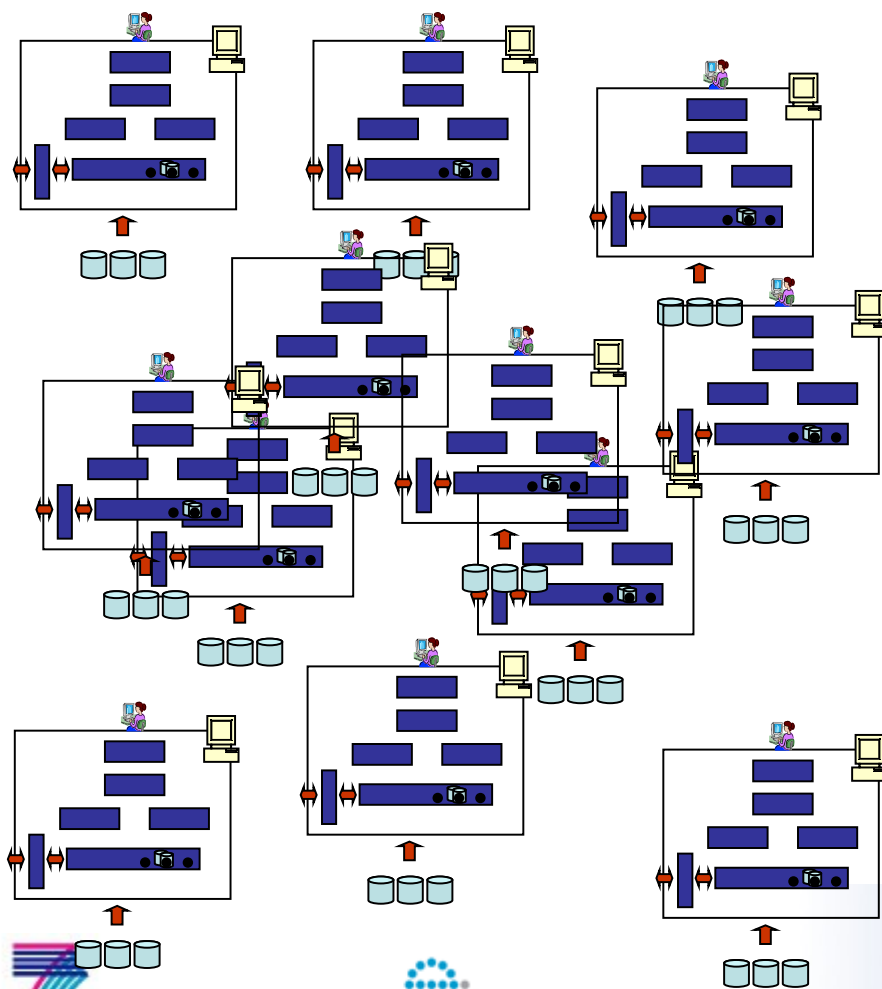
# Systemy repozytoryjne: indywidualna instytucja



- Scentralizowany system
- Wysokie koszty instalacji i obsługi sprzętu i oprogramowania
- Ograniczona skalowalność
- Ponowne wykorzystanie danych i usług – jedynie przez powielanie

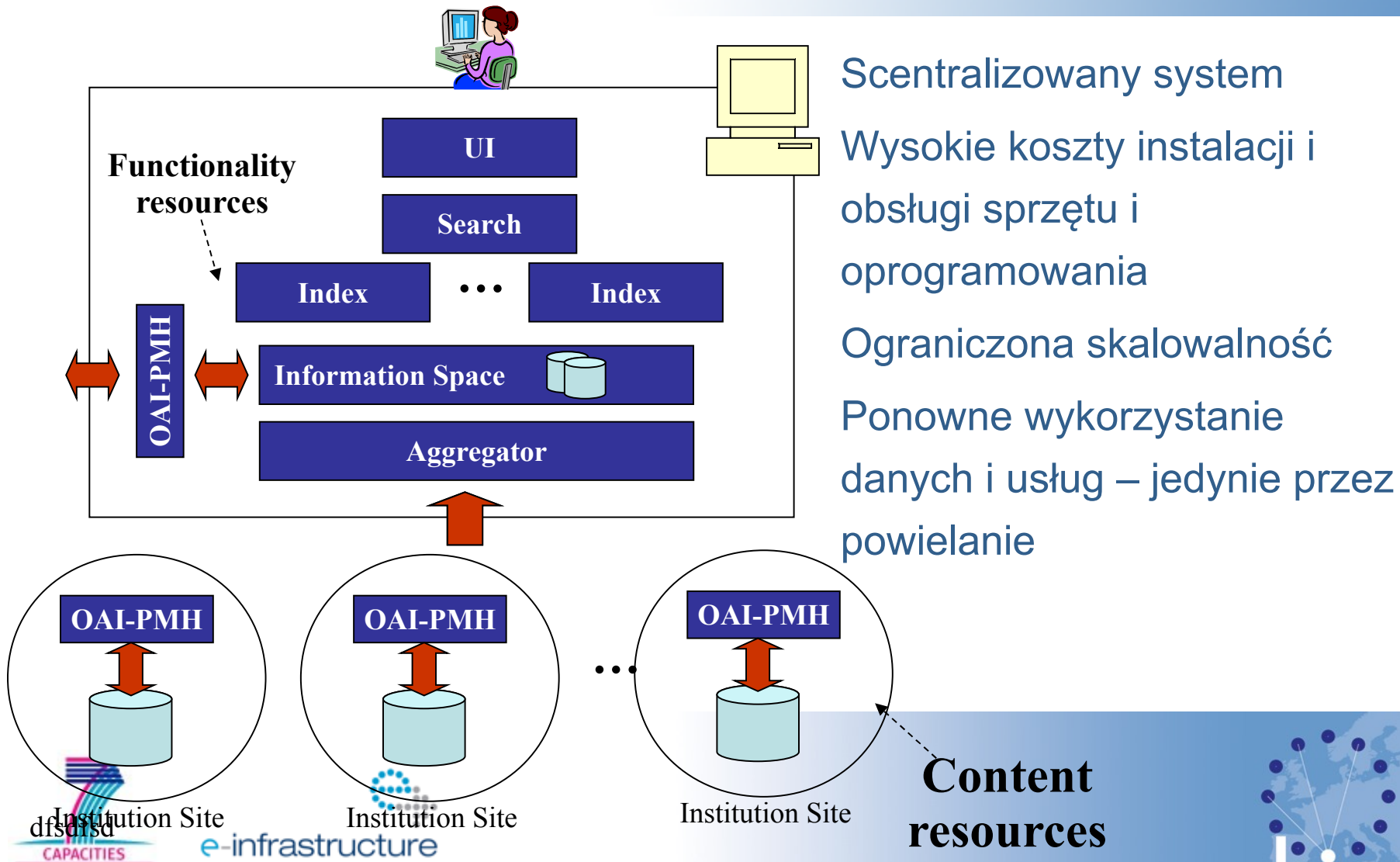


# Systemy repozytoryjne: wiele instytucji



- Powielane wysiłki
- Osobne repozytoria

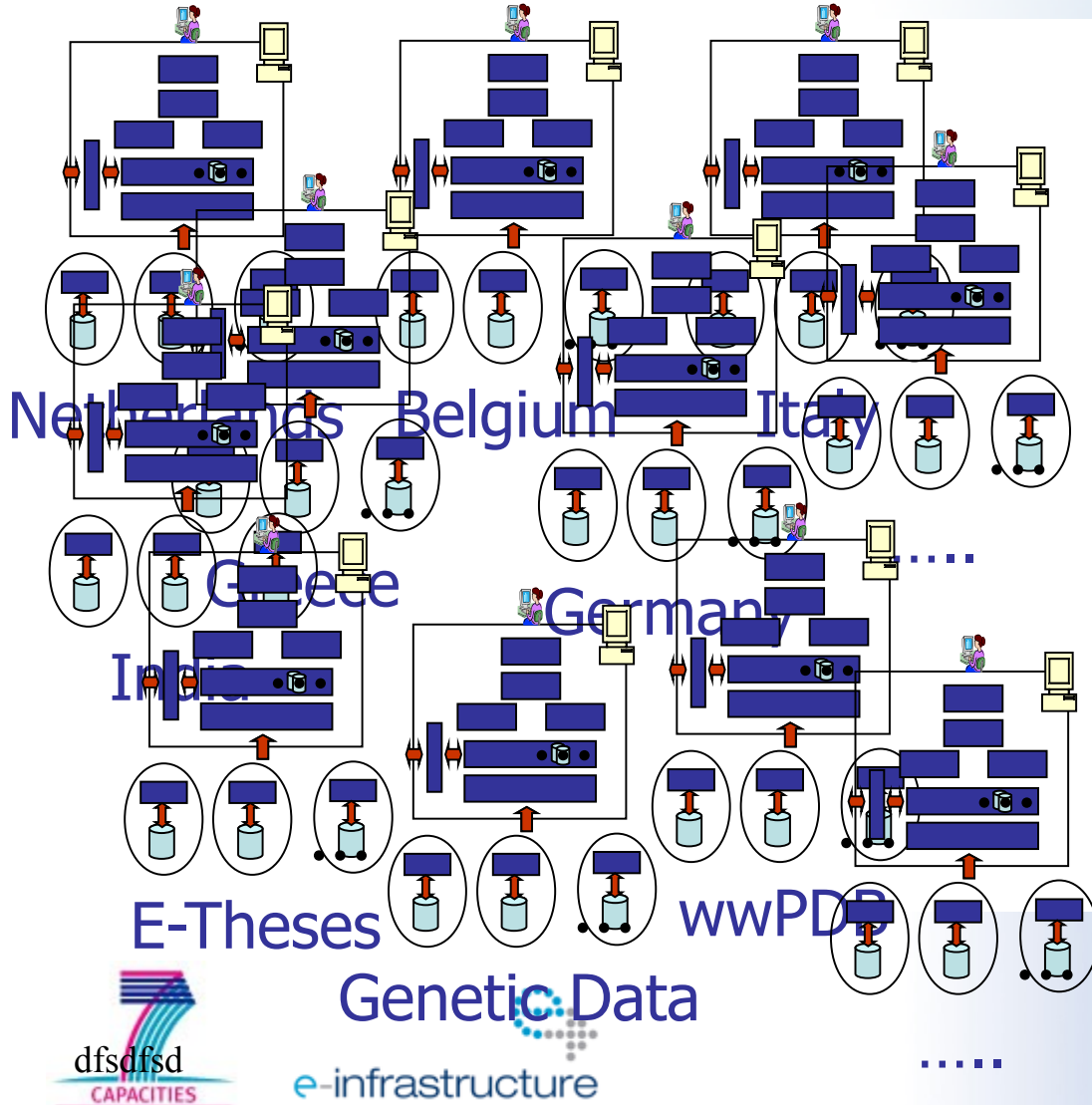
# Systemy repozytoryjne: wymiana i ponowne wykorzystanie



- Scentralizowany system
- Wysokie koszty instalacji i obsługi sprzętu i oprogramowania
- Ograniczona skalowalność
- Ponowne wykorzystanie danych i usług – jedynie przez powielanie

# Systemy repozytoryjne: wymiana i ponowne wykorzystanie

Powielane wysiłki  
Osobne repozytoria  
(narodowe lub tematyczne)





# Wizja infrastruktury DRIVER

Przejdźcie od

Budowania indywidualnych  
Repozytoriów lub ich klastrów

Do budowania

„silnika generującego”

- INFRASTRUKTURY

Ułatwiającej powyższe działania  
dzięki standardowym serwisom  
nadającym się do wielokrotnego użycia

# Wizja DRIVER

- Wszystkie instytucje badawcze w Europie i na świecie udostępniają w sposób otwarty publikacje naukowe poprzez repozytoria instytucjonalne
- Wytyczne, certyfikaty i technologia gwarantują wspólnie zautomatyzowaną agregację wszystkich publikacji w jedną, wirtualną i otwartą bazę wiedzy
- Treści (indeks danych) są dostępne dla dostawcy usług, aby zapewniali szeroki zakres usług dla użytkowników końcowych (wyszukiwanie, przeglądanie, wizualizacja, miary cytowalności, itd.)
- API umożliwia łączenie i interakcję publikacji z innymi typami danych i obiektami (dane źródłowe, filmy, nagrania, materiały edukacyjne, itd.)



# Inwentaryzacja repozytoriów, 2007

Badanie „The European Repository Landscape”

Znacząca liczba instytucji posiada repozytoria w 15 państwach. 5 państw jest w fazie początkowej. W 7 pozostałych (w tym Polska) brak repozytoriów.

Okolo miliona obiektów ok. 200 000 w 2005 roku. Średnia liczba obiektów repozytorium: ok. 9000.

90% to obiekty tekstowe; 68% z nich to jedynie metadane. 54% obiektów tekstowych to artykuły, 19% książki, 15% rozprawy, 10% publikacje konferencyjne, 5% teksty robocze.

Eprints (24%), DSpace (20%), OPUS (11%), lokalne oprogramowanie (17%) - i 14 innych narzędzi.

70-75% repozytoriów nadaje dokumentom identyfikatory. 48% stosuje zestandaryzowane słowa kluczowe, a 44% dowolne.

Brak jednolitych usług dla repozytoriów.



# Cele / wyzwania

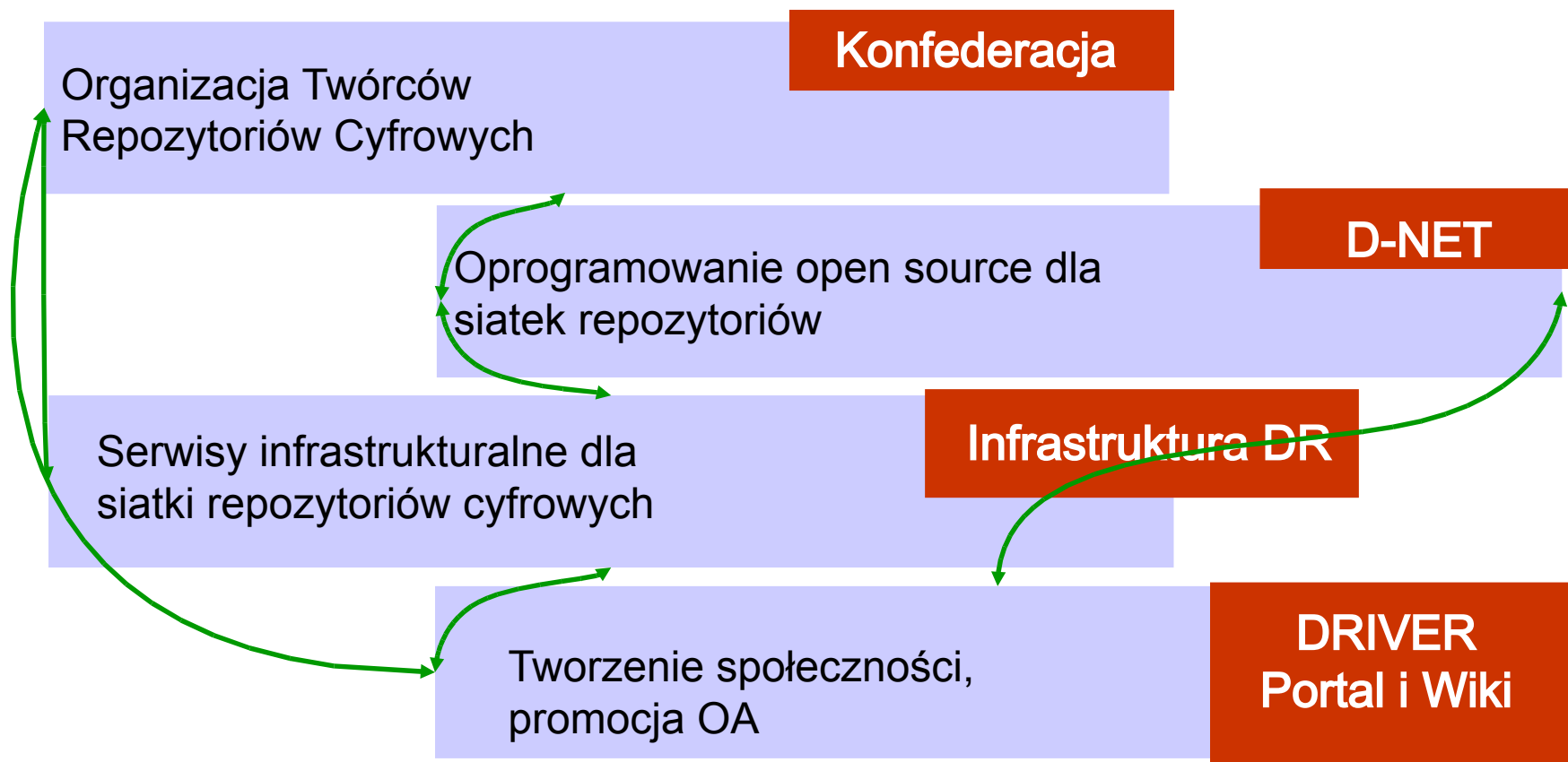
- Stworzyć środowisko integrujące istniejące repozytoria: narodowe, regionalne i tematyczne
- Stworzyć europejską infrastrukturę repozytoriów cyfrowych
- Identyfikacja i promocja standardów
- Zwiększenie świadomości w społecznościach naukowych

# DRIVER II

- 24 miesiące
- Budżet 2,7m EUR
- Grudzień 2007 – listopad 2009
- Produkty:
  - Infrastruktura DR
  - Europejska Konfederacja DR
- Finansowane przez KE, FP 7
- Partnerzy konsorcjum
  - Univ.of Athens (GR)
  - Univ. of Bielefeld (GE)
  - CNR-ISTI (IT)
  - STICHTING SURF (NL)
  - Univ. of Nottingham (UK)
- Partnerzy cd.
  - Univ. of Bath (UK)
  - Univ. of Warszawski (PO)
  - Univ. of Gent (BE)
  - Univ. of Goettingen (GE)
  - Danish Technical University (DK)
  - Universidade do Minho (PT)
  - Narodna in univerzitetna knjižnica (SLO)



# DRIVER II: obszary działania



# DRIVER: architektura

Agregacja repozytoriów wzorowana na projektach krajowych, działających w oparciu o OAI-PMH (BASE, DARE, OAster).

E-Infrastruktura oparta o serwisy sieciowe (web services)

Zasoby są wykorzystywane jako serwisy sieciowe działające w sposób rozproszony

Aplikacje powstają jako połączenia serwisów  
Różnorodne aplikacje mogą współistnieć i wykorzystywać te same serwisy.

Specjalne serwisy śledzą dostępność serwisów i aplikacji.

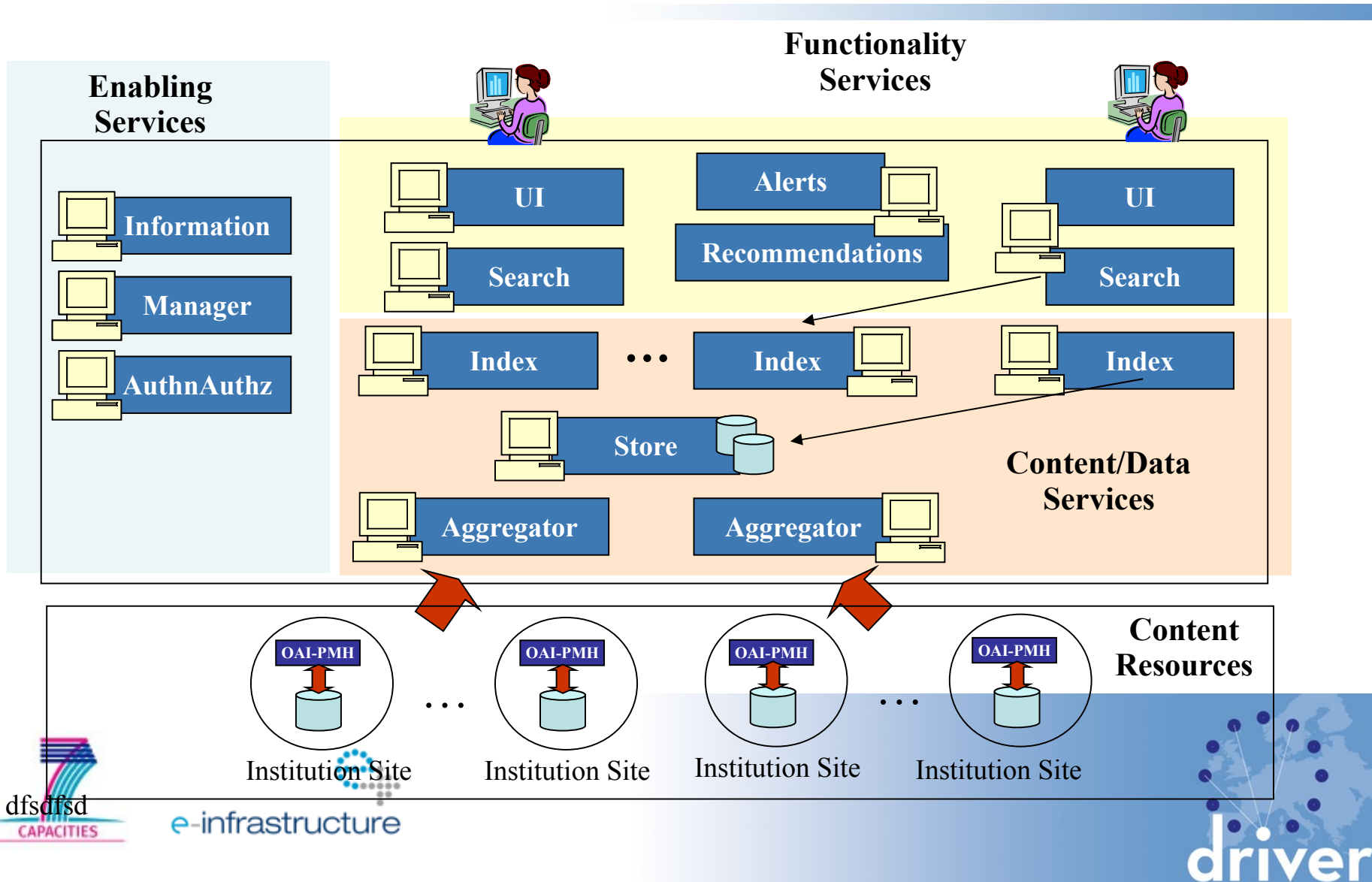


# Zalety technologiczne

- Skalowanie i dynamiczność
  - Repozytoria mogą być dodawane dynamicznie
- Skalowanie wraz ze wzrostem obciążenia / wykorzystania
- Nowe funkcjonalności / usługi mogą być łatwo dodane
- System w pełni dystrybuowany
  - Dzięki architekturze typu Web Services



# Infrastruktur DRIVER



# DRIVER: Aktorzy

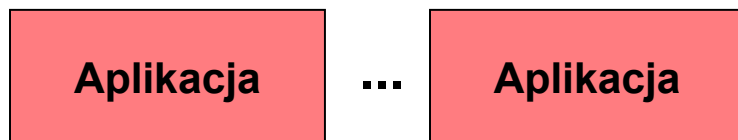
Konsorcjum  
DRIVER



Spoleczności użytkowników

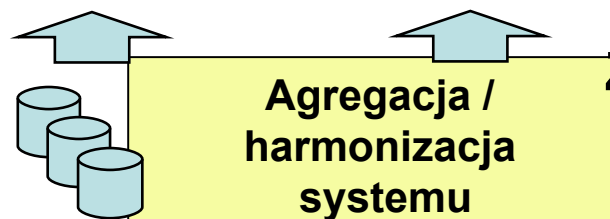


“Operating”  
System



**Dostawcy usług**

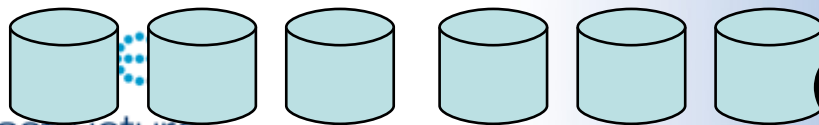
Konsorcjum DRIVER  
Spół. narodowe  
Spół. tematyczne.



**Zarządzający DR / sieciami DR**

Konsorcjum DRIVER  
Spół. narodowe

Spoleczności  
narodowe



**Dostawcy treści**  
(repozytoria europejskie)



e-infrastructure



# DRIVER: stan infrastruktury

DRIVER: infrastruktura sieciowa + zagregowane dane + aplikacje DRIVER tworzone na bazie danych i infrastruktury (+ aplikacje zewnętrzne)

## Dane

- 150 zagregowanych repozytoriów (30 kolejnych zarejestrowanych) z 21 krajów
- Ponad 700 000 dokumentów w modelu open access•

## Usługi

- 25 różnych serwisów sieciowych



## *Do czego służy DRIVER?*

Przeszukiwanie zagregowanych repozytoriów

Portal agregujący treści (np. Recolecta) –  
wyspecjalizowane kolekcje treści

Weryfikacja repozytoriów

- Wytyczne odnośnie metadanych

- Sieciowy system weryfikujący metadane (validator)

- Rejestracja repozytorium – zezwala na pobieranie i indeksowanie danych



# *Konfederacja DRIVER: społeczność europejska*

Współpraca z czołowymi ośrodkami (DARENET / Surf, SHERPA/ JISC, itd.)

Konfederacja skupia – początkowo w skali europejskiej – społeczności repozytoryjne, dostawcy systemów repozytoryjnych, dostawcy usług, organizacje polityczne, badawcze i finansujące badania.

Wsparcie w warstwie społecznej integracji repozytoriów.

Członkowie zgadzają się co do podstawowych zasad:

Otwarta dostępność publikacji naukowych

Partnerzy w sieci usług repozytoryjnych

Stosowanie wytycznych dla wyższej interoperacyjności danych i usług

Zapewnienie długoterminowego dostępu do publikacji



# DRIVER i społeczności krajowe

## Społeczności krajowe reprezentowane przez “korespondentów”

Instytucja lub grupa instytucji odpowiedzialna za budowanie narodowej społeczności repozytoryjnej (np. DARENet-NL, SHERPA-UK, OA-Netzwerk-GE, RECOLECTA-ES, HAL-FR)

- “Korespondenci”
  - Utrzymują informacje o danym kraju na wiki DRIVER
  - Organizują imprezy we własnym kraju
  - Tłumaczą i promują wytyczne oraz inne kluczowe materiały
  - Agregują dane, czyszczą dane, ew. Oferują dodatkowe usługi



# *D-NET: DRIVER Network-Evolution-Toolkit*

Konsorcjum DRIVER jako pierwszy dostawca usług. Wesja 1.0 udostępniona 20 czerwca 2008 na licencji Apache.

Przykład aplikacji tworzonych na bazie dostępnych serwisów sieciowych.

*„zestaw narzędzi pozwalający stworzyć dystrybuowany system zawierający narzędzia służące pobieraniu i agregacji różnorodnych źródeł danych. Pakiet zawiera szereg funkcjonalności dla użytkowników końcowych, takich jak wyszukiwanie, rekomendacje, kolekcje, profilowanie czy innowacyjne narzędzia dla administratorów repozytoriów”*



# *D-NET: DRIVER Network-Evolution-Toolkit*

Pakiet D-NET zawiera m.in.:

- Oprogramowanie do zarządzania siecią repozytoriów (*Repository Network Manager, Resource Monitoring*)
- Usługi dla użytkowników końcowych (*Search, Browse, Profiling*)
- Usługę wspierającą lokalnych administratorów i agregatorów repozytoriów (*Validation Tool*)

*Stosowanie D-NET dobrowolne – ale gwarantuje interoperacyjność.*





# *D-NET: DRIVER Network-Evolution-Toolkit*

## Scenariusze wykorzystania:

Nowe usługi / serwisy działające w oparciu o infrastrukturę DRIVER (np. hiszpański portal Recolecta)

Wyspecjalizowane portale tworzone w oparciu o istniejące usługi i dane, dostosowane do potrzeb danej społeczności

Uruchomienie nowej instancji infrastruktury DRIVER



# Wytyczne DRIVER

Oparte na „The Investigative Study of Standards for Digital Repositories and Related Services”

Wskazówki dla administratorów repozytoriów, służące dobraniu odpowiednich polityk zarządzania danymi.

Uzupełnienie standardów sieciowych (np. OAI PMH)

Kolekcje: definiowanie kolekcji pełnotekstowych w ramach repozytorium

OAI-PMH: standaryzacja interpretacji protokołu

Metadane: obowiązkowe i dobrowolne wytyczne związane z implementacją Dublin Core

# *Dalsze działania*

Lobbowanie na rzecz mandatów OA

Renegocjacja kontraktów z wydawcami – tworzenie konsorcjów OA (promocja modelu “green road”, zautomatyzowany transfer tekstów do repozytoriów)

Zautomatyzowany transfer metadanych z baz danych do repozytoriów

Treści inne niż tekstowe; “publikacje o podwyższonej jakości” - łączące w spójną całość powiązane obiekty informacyjne

*Dalsze działania*

Teraz Polska?



# Dziękuję.

Więcej informacji na stronach:

*<http://driver.icm.edu.pl/>*

*<http://www.driver-repository.eu/>*

