

# Enabling Services in Knowledge Infrastructures: The DRIVER Experience

L. Candela   D. Castelli   P. Manghi   P. Pagano

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" - CNR  
Via G. Moruzzi, 1 - 56124 PISA - Italy  
{candela|castelli|manghi|pagano}@isti.cnr.it

IRCDL 2007, January 29-30<sup>th</sup>



# Abstract

**Knowledge Infrastructures** are the next big challenge of Digital Library research community

- large-scale
- world-wide, heterogeneous, evolving communities

In this talk:

- 1 **Background**
  - Knowledge Infrastructures
- 2 **DRIVER**
  - Overview
  - The Architecture
  - Enabling Services

# Definition and Imperatives

## Definition

A **Knowledge Infrastructure** is a basic physical and organizational framework which provides scientists with facilities needed to share, use and create knowledge

- **Comprehensive, global** access to any type of information
- **Minimum** time and resources effort to access and use this information
- **Easy** search/navigation, handling, manipulation, and re-dissemination of information
- **Maximum** visibility to and communication with the research community, research impact
- **Long-term** access and preservation of research results

# Experiences and Initiatives

Focused on content sharing and reuse (OAI-PMH based)

- the National Science Digital Library (NSDL)
- The European Library (TEL)
- the Dutch network of Digital Academic REpositories (DARE)

Large-Scale DL applications rather than Knowledge Infrastructures

- fixed view over the information space
- fixed pool of functionality
- difficulty to evolve and meet different requirements

Knowledge Infrastructures must be **sustainable**

- scalability, independency, maintainability, openness

DRIVER Digital Repository Infrastructure Vision for European Research



# Experiences and Initiatives

Focused on content sharing and reuse (OAI-PMH based)

- the National Science Digital Library (NSDL)
- The European Library (TEL)
- the Dutch network of Digital Academic REpositories (DARE)

Large-Scale DL applications rather than Knowledge Infrastructures

- fixed view over the information space
- fixed pool of functionality
- difficulty to evolve and meet different requirements

Knowledge Infrastructures must be **sustainable**

- scalability, independency, maintainability, openness

DRIVER Digital Repository Infrastructure Vision for European Research



# Experiences and Initiatives

Focused on content sharing and reuse (OAI-PMH based)

- the National Science Digital Library (NSDL)
- The European Library (TEL)
- the Dutch network of Digital Academic REpositories (DARE)

Large-Scale DL applications rather than Knowledge Infrastructures

- fixed view over the information space
- fixed pool of functionality
- difficulty to evolve and meet different requirements

Knowledge Infrastructures must be **sustainable**

- scalability, independency, maintainability, openness

DRIVER Digital Repository Infrastructure Vision for European Research



# Experiences and Initiatives

Focused on content sharing and reuse (OAI-PMH based)

- the National Science Digital Library (NSDL)
- The European Library (TEL)
- the Dutch network of Digital Academic REpositories (DARE)

Large-Scale DL applications rather than Knowledge Infrastructures

- fixed view over the information space
- fixed pool of functionality
- difficulty to evolve and meet different requirements

Knowledge Infrastructures must be **sustainable**

- scalability, independency, maintainability, openness

**DRIVER** Digital **R**epository **I**nfastructure **V**ision for **E**uropean  
**R**esearch



# DRIVER Facts

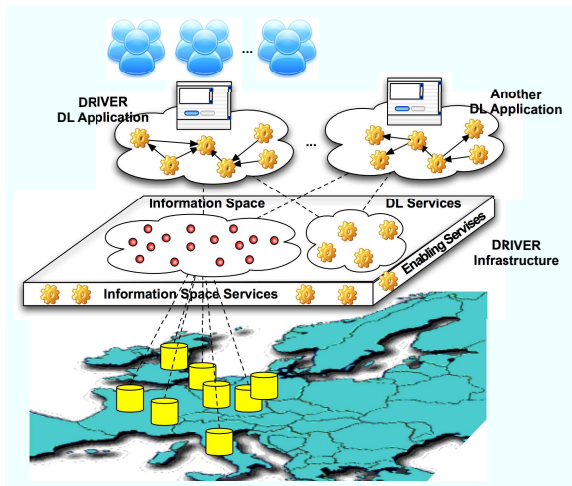
- **Ongoing** IST project (Jun '06 – Nov '07)
- Developing an **organization** and a **system**
- **Two** public **releases**: May '07, Nov '07
- **10 partners** (technology - organization & administration - content owners & users)
  - Univ. of Athens (Hellas), Univ. of Gottingen (Germany), CNR-ISTI (Italy), SURF (The Netherlands), Univ. of Nottingham (UK), CNRS (France), UKOLN (UK), ICM (Poland), Univ. of Gent (Belgium), Univ. of Bielefeld (Germany)
- provide access to **51 Institutional Repositories**
  - various European countries (Netherlands, Germany, France, Belgium and the UK)
  - half million of documents (will grow)
  - various topics, media, format, languages

# High-level Objectives

- Develop a test-bed for **integrating** existing national, regional, or thematic repositories
- Create a production-quality **European** DR infrastructure
- Prepare the future expansion and upgrade of **the DR infrastructure** across Europe
- Identify and promote the use of a relevant set of **standards**
- Raise **awareness** among user communities



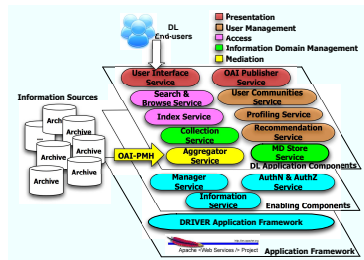
# High-level View



# Principles

**Layered approach** based on loosely-coupled component oriented approach (SOA)

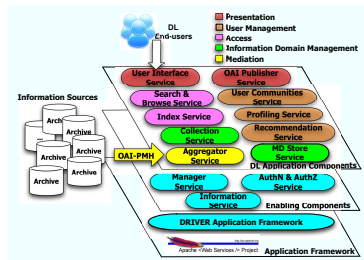
- **supporting components operations**
- supporting component to component co-operation
- providing end-user functionality
- supporting evolving requirements through component selection and replacement
- promoting reuse
- promoting distributed development and deployment



# Principles

**Layered approach** based on loosely-coupled component oriented approach (SOA)

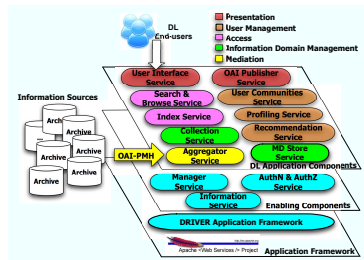
- supporting components operations
- **supporting component to component co-operation**
- providing end-user functionality
- supporting evolving requirements through component selection and replacement
- promoting reuse
- promoting distributed development and deployment



# Principles

**Layered approach** based on loosely-coupled component oriented approach (SOA)

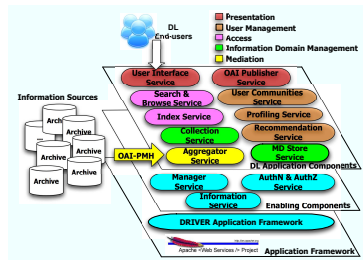
- supporting components operations
- supporting component to component co-operation
- **providing end-user functionality**
- supporting evolving requirements through component selection and replacement
- promoting reuse
- promoting distributed development and deployment



# Principles

Layered approach based on **loosely-coupled component oriented approach** (SOA)

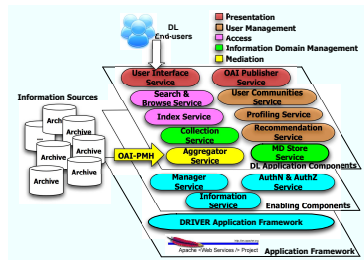
- supporting components operations
- supporting component to component co-operation
- providing end-user functionality
- **supporting evolving requirements through component selection and replacement**
- promoting reuse
- promoting distributed development and deployment



# Principles

Layered approach based on **loosely-coupled component oriented approach** (SOA)

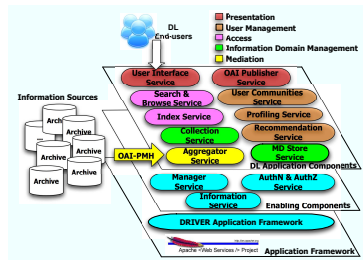
- supporting components operations
- supporting component to component co-operation
- providing end-user functionality
- supporting evolving requirements through component selection and replacement
- **promoting reuse**
- promoting distributed development and deployment



# Principles

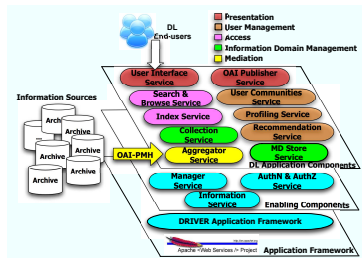
Layered approach based on **loosely-coupled component oriented approach** (SOA)

- supporting components operations
- supporting component to component co-operation
- providing end-user functionality
- supporting evolving requirements through component selection and replacement
- promoting reuse
- **promoting distributed development and deployment**



# DL Application Components and Functionality

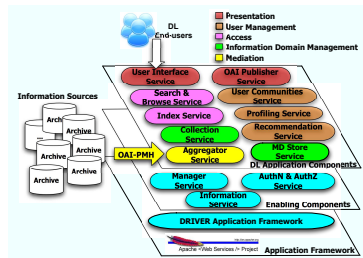
- Gathering content
- Storing collected content
- Collection management
- Support content discovery
- Content discovery
- User and Group management
- User profiling
- User notifiers
- Graphical user interface
- OAI-PMH content provider



# Role, Components and Functionality

## Core services to promote sustainability

- regulates resources usage (XACML)
- supports resources discovery and monitoring
- implements resources orchestration



# DRIVER Information Service: The Resource Model

$\forall$  resource  $\exists$  **Resource profile**, i.e. the minimal information to describe a resource

**Resource type**  $\in$  Profile - identify a class of services, i.e. the functionality and the logic

Two types of resources

- **service**, i.e. resources delivering a functionality
  - **factory**, creates functionality or data structures and manages them
  - **functionality**, operates on data structures
- **data structure**, i.e. information packages representing “the status” services act on



# DRIVER Information Service: The Functionality

Provide services with the updated **map of the infrastructure**

- scalability, independency, maintainability, and openness imply highly dynamic pool of resources

Two interaction models

- **push**-model, services issue queries
- **pull**-model, subscription/notification mechanism

Virtually centralized service implemented as **highly distributed and replicated** to avoid single point of failure

- the DRIVER Application Framework provides a lightweight software library supporting access while hiding distribution details



# DRIVER Manager Service

Orchestrates the resources forming each DL Application in order to **ensure their operation** and **guarantee QoS**

- create, update, and remove resources dynamically
- monitor resources through the IS

The goal is to maintain

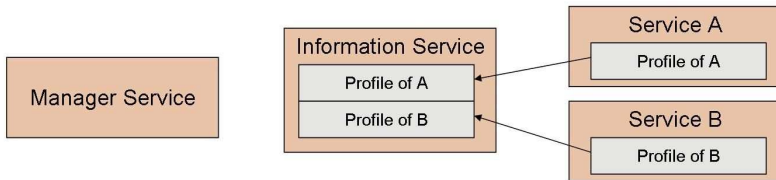
- **robustness**, guarantee the availability of minimal number of Resource Type resources
- **liveness**, guarantee the existence of certain resources for others to work

by interpreting **orchestration rules** defined at DL Application level, e.g.

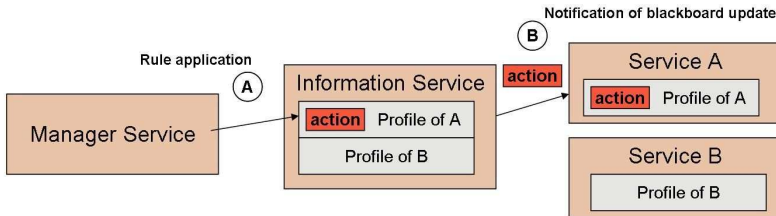
- create a new MD Store Service  $\forall$  Repository (liveness)
- notify application administrator whenever a fired rule cannot be applied (robustness)

# Orchestration Protocol

Combines subscription/notification mechanism and blackboard



Registration of A and B - Subscription to "own blackboard update" event



Orchestration: message through profile *blackboard* update

# Summary

## Knowledge Infrastructures emerged as the next generation Digital Library systems

- Goal: provide research community with the data and the tools they need
- Imperative: sustainability
  - ease the access to comprehensive pool of resources
  - ease the adjunction new resources
  - ease the creation of new DL applications
- Need: Enabling services implementing sustainability

**DRIVER** is a first step towards a **Pan-European Knowledge Infrastructure**

[www.driver-repository.eu](http://www.driver-repository.eu)