DRIVER – Building an Infrastructure of European Scientific Repositories

Norbert Lossau

Göttingen State and University Library, Germany

at

IPRES 2007, Beijing

Topics

- Why do we need a DRIVER project at all?
- DRIVER basics
- What happens today with repositories?
- Where DRIVER can make a difference!
- DRIVER products
- DRIVER II – outlook

The scientific/political framework and vision Berlin Declaration, Oct/2003

Free and unrestricted access to sciences and human knowledge representation worldwide, incl. cultural heritage

⇨Needs an interoperable, trusted, long-term repository infrastructure

Topics

- Why do we need a DRIVER project at all?
- DRIVER basics
- What happens today with repositories?
- Where DRIVER can make a difference!
- DRIVER products
- DRIVER II – outlook

General Information: DRIVER I

- Duration: 18 months
- Budget: 1.8M
- Timeplan: 06/'06 -11/'07
- Main Deliverables:
  - DRIVER Public Test-bed
  - European Awareness Repository Point
- Website:

http://www.driver-repository.eu/

- Funded by the European Commission, "Research Infrastructure" Unit. FP 6

- Partners
  - Univ.of Athens (GR)
  - Univ. of Bielefeld (GE)
  - CNR-ISTI (IT)
  - STICHTING SURF (NL)
  - Univ. of Nottingham (UK)
  - CNRS-CCSD (FR)
  - Univ. of Bath (UK)
  - Univ. of Warszawski (PO)
  - Univ. of Gent (BE)
  - Univ. of Gottingen (GE)

driver

DRIVER Overall Vision

To build a pan-European Digital Repository Infrastructure Hub in a Global Repository Network for

- ⇨ any type of document, data and object,
- ⇨ of any format,
- ⇨ involving all European countries and
- ⇨ collaborating with disciplinary repositories, built up by scientific communities

DRIVER

« The first step towards a pan-European Infrastructure for digital Repositories »

Content

Network of “Content-Providers”

- Institutional, national, European, thematic networking
- Currently documents, second phase primary data etc.

Infrastructure

“Test Bed” for Repository-Services

- Open Service-Architecture with core functionalities
- Fundament for a pan-European repository infrastrucutre

Planning

“Focussed Studies”

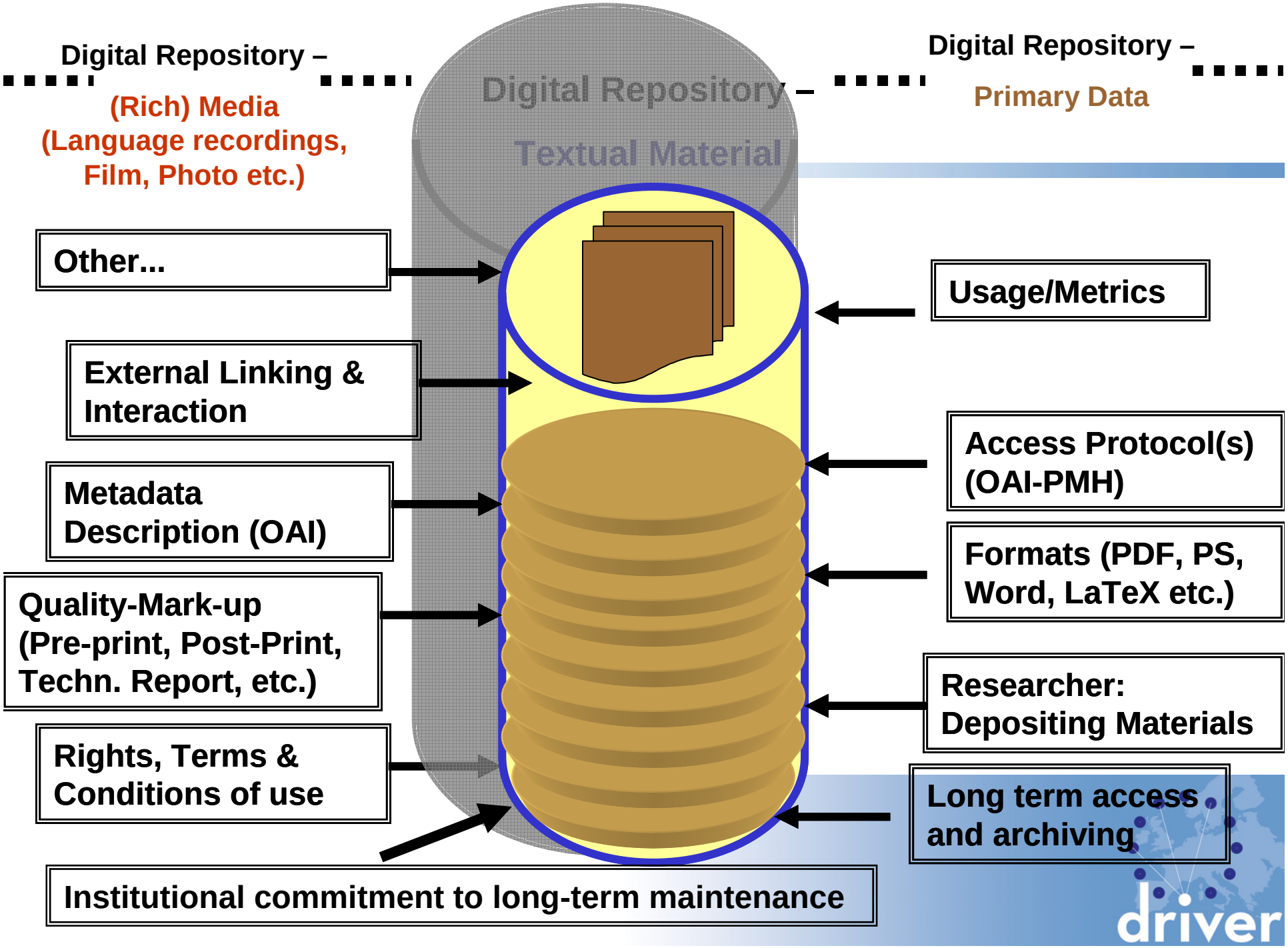
- Ex. Most comprehensive study on digital repositories in Europe

Outreach

“Advocacy and Awareness Raising”

- A European point of information
- Information- und training activities

driver


Digital Repository –
(Rich) Media
(Language recordings,
Film, Photo etc.)
Digital Repository –
Textual Material
Digital Repository –
Primary Data
Other...
External Linking &
Interaction
Metadata
Description (OAI)
Quality-Mark-up
(Pre-print, Post-Print,
Techn. Report, etc.)
Rights, Terms &
Conditions of use
Institutional commitment to long-term maintenance
Usage/Metrics
Access Protocol(s)
(OAI-PMH)
Formats (PDF, PS,
Word, LaTeX etc.)
Researcher:
Depositing Materials
Long term access
and archiving
driver

Topics

- Why do we need a DRIVER project at all?
- DRIVER basics
- What happens today with repositories?
- Where DRIVER can make a difference!
- DRIVER products
- DRIVER II – outlook

Task 7.1 Inventory Study of EU repositories

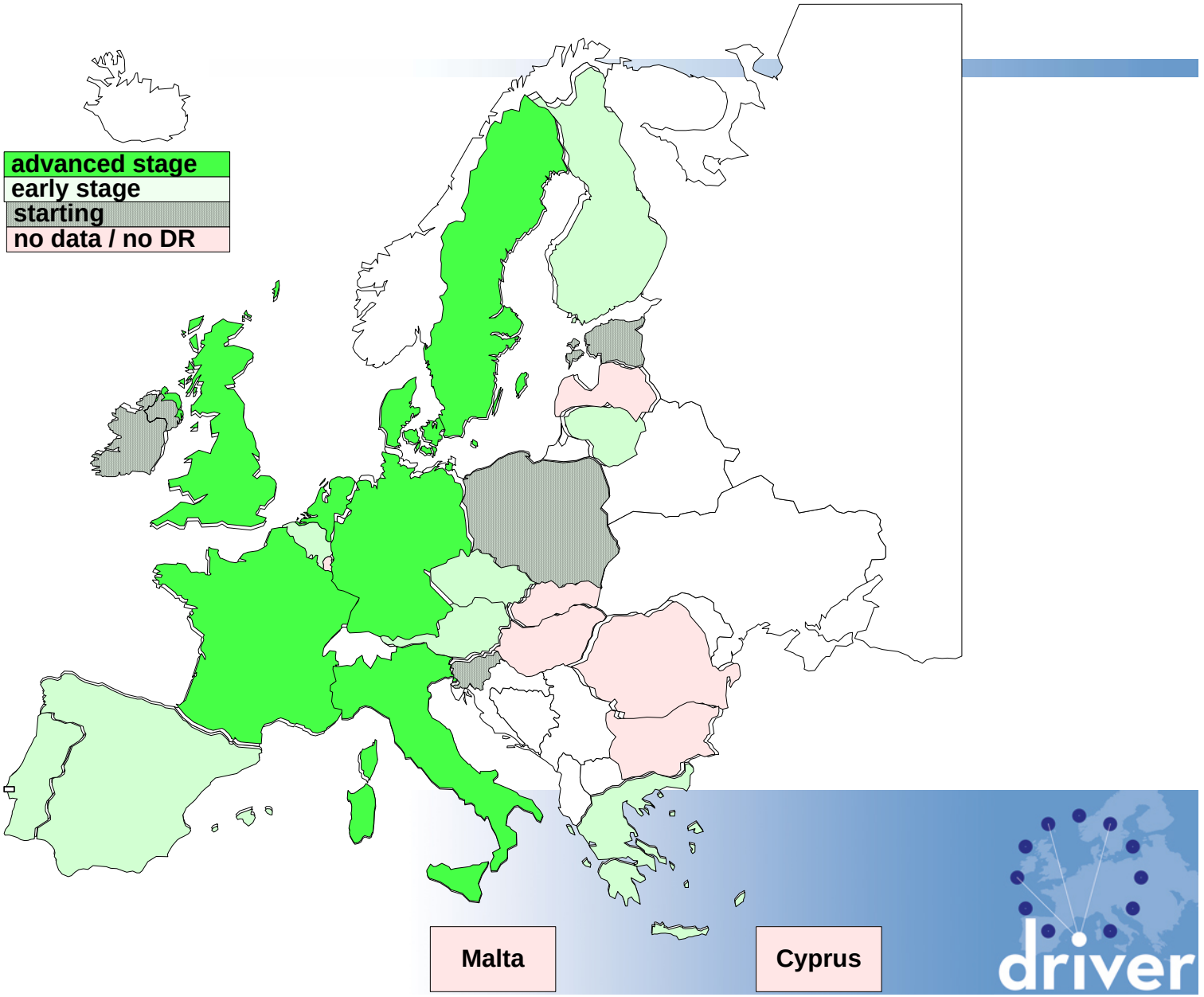
Selection;

- 230 institutions with DR's (with contact addresses) identified
- of which 50% participated in the web survey / telephone interviews

Results:

- 7 EU countries do not appear to have Research Institutes with DR's
- 5 EU countries seem to be in a starting phase
- in 15 Countries a sizeable part of the Research Universities has implemented a DR for research output; in 7 countries more than half of the Universities have done so

Task 7.1 Inventory Study. Results per country



Task 7.2 A Driver's guide to European Repositories. Inventory study of important DR related issues and good practices

Five specific, complex and longer term issues that are essential to the establishment, development or sustainability of a DR;

- Business Models (Alma Swan)
- Stimuli for depositing materials into DR's (Vanessa Proudman, UvT)
- Intellectual Property Rights (Wilma Mossink, SURF)
- Data Curation (Rene van Horik, DANS)
- Long Term Preservation (Barbara Sierman, KB)

Task 7.2 A Driver's guide to European Repositories

Practical approach reflected in chapter on Populating Repositories;

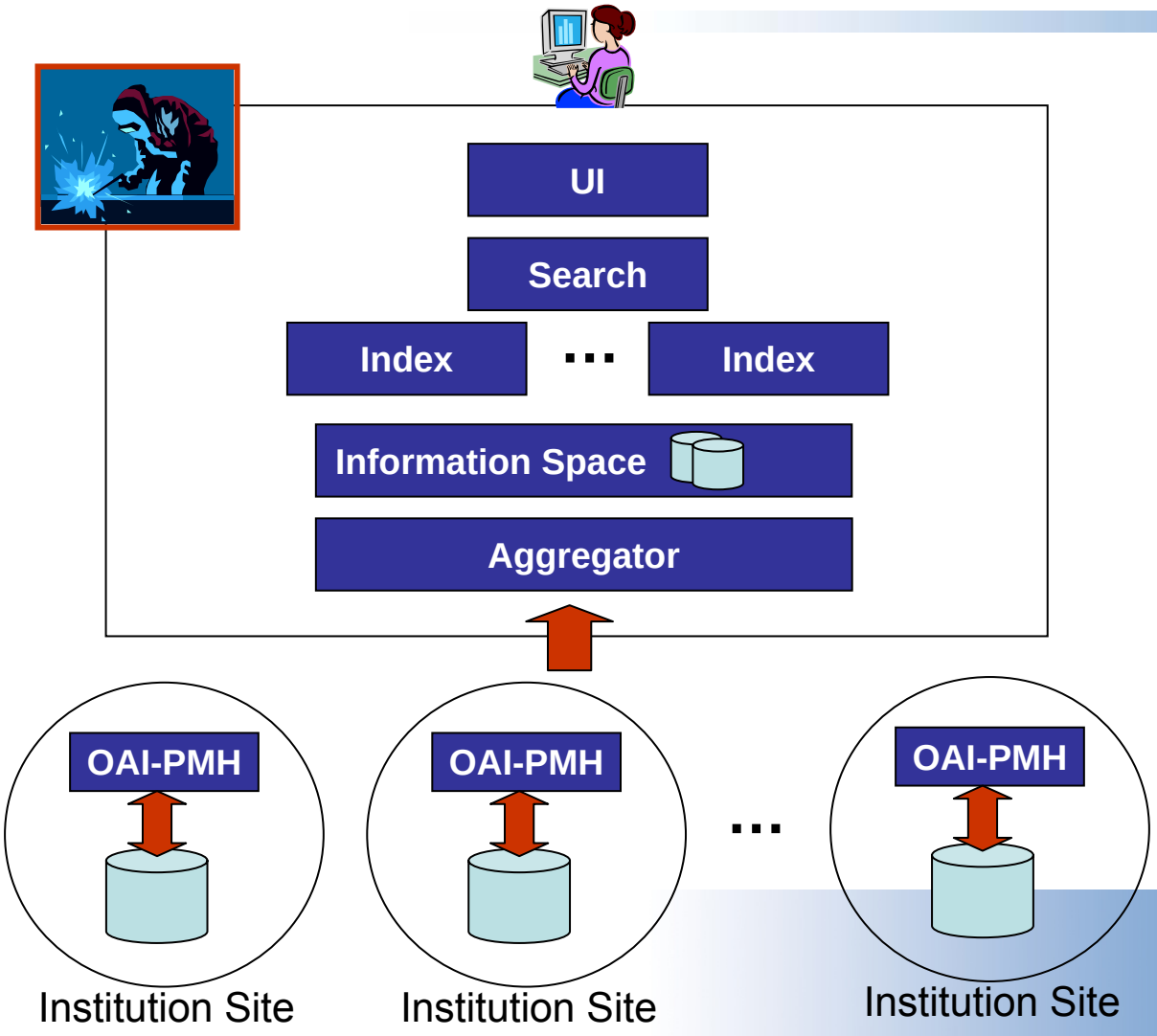
6 case studies that reflect different types of DR's and service models:

- A university DR (Minho, Portugal)
- A university School Repository + campus wide IR (Southampton)
- A central archive Repository which brings together national Research results (HAL, France)
- An international research organisation IR (CERN, Switzerland)
- A subject specific service model built on IR content (Connecting Africa, the Netherlands)
- A service which increases quality IR (Cream of Science, the Netherlands)

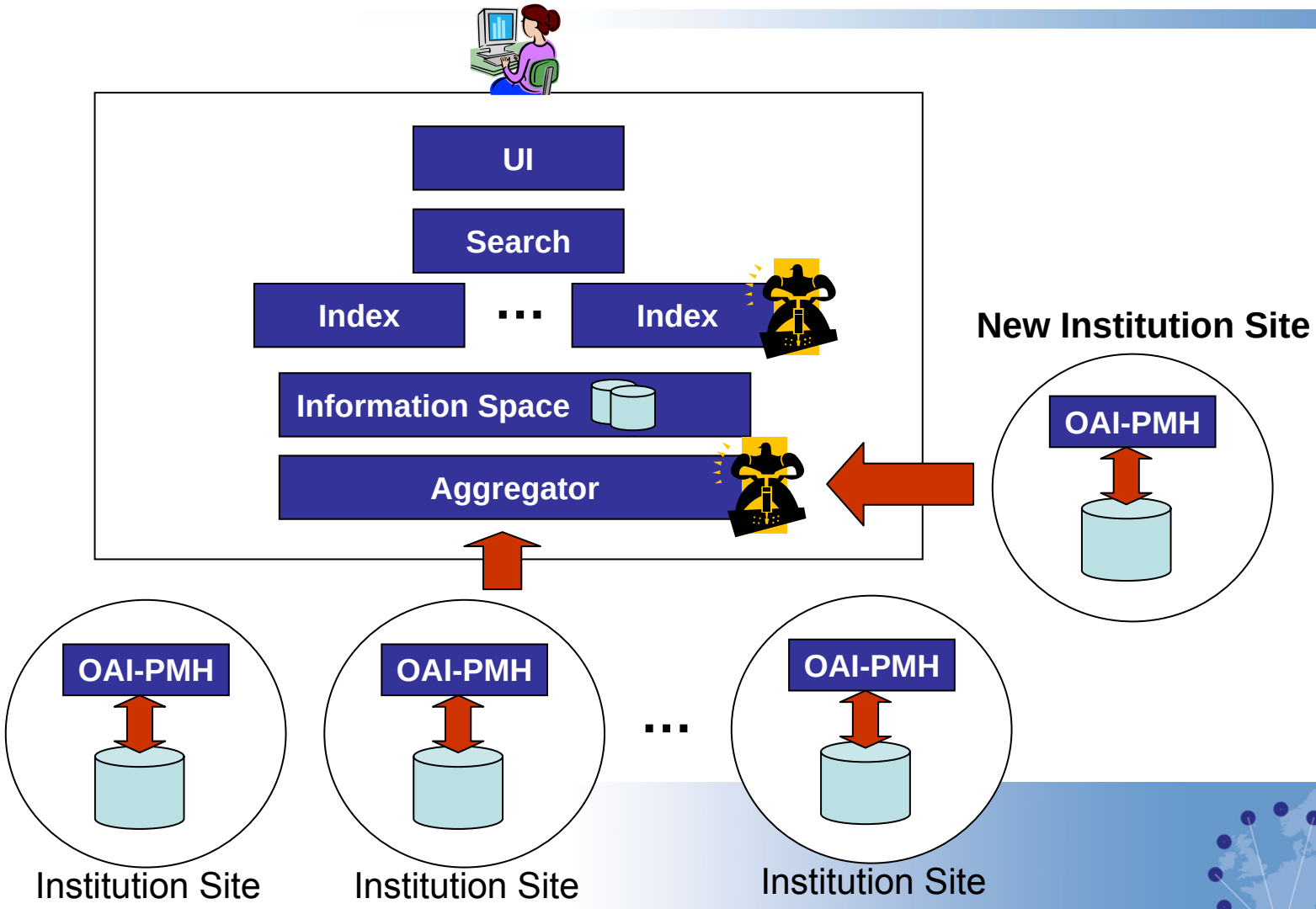
Current Solutions are not sustainable

- EU features a multitude of *Repository Systems*, providing content and functionality over Repositories
- Based on different technologies; e.g. ePrints, DSpace, Fedora, simple Web Servers, proprietary solutions
- Heterogeneous content, i.e. metadata formats, object models, file formats

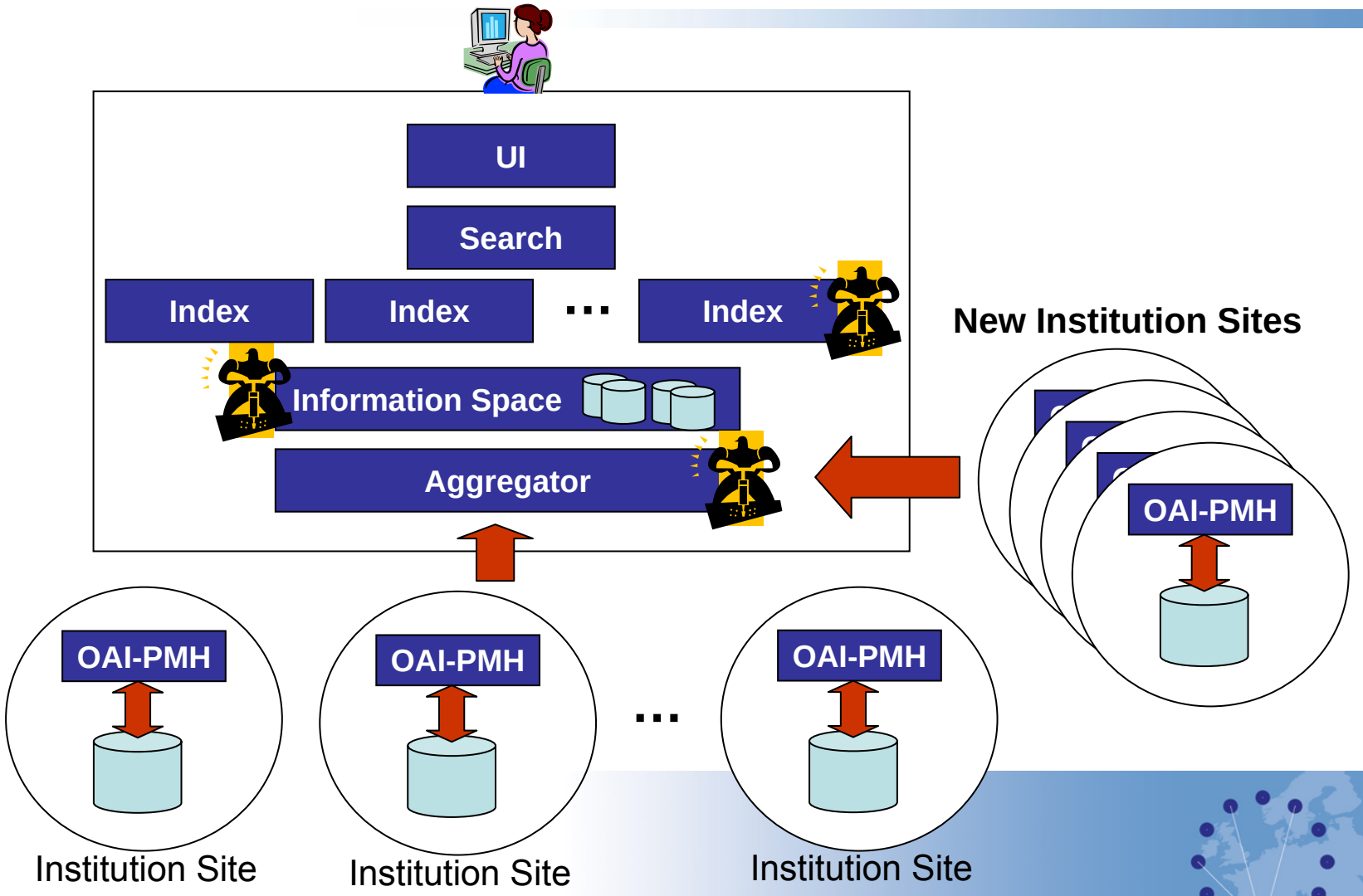
Repository Systems



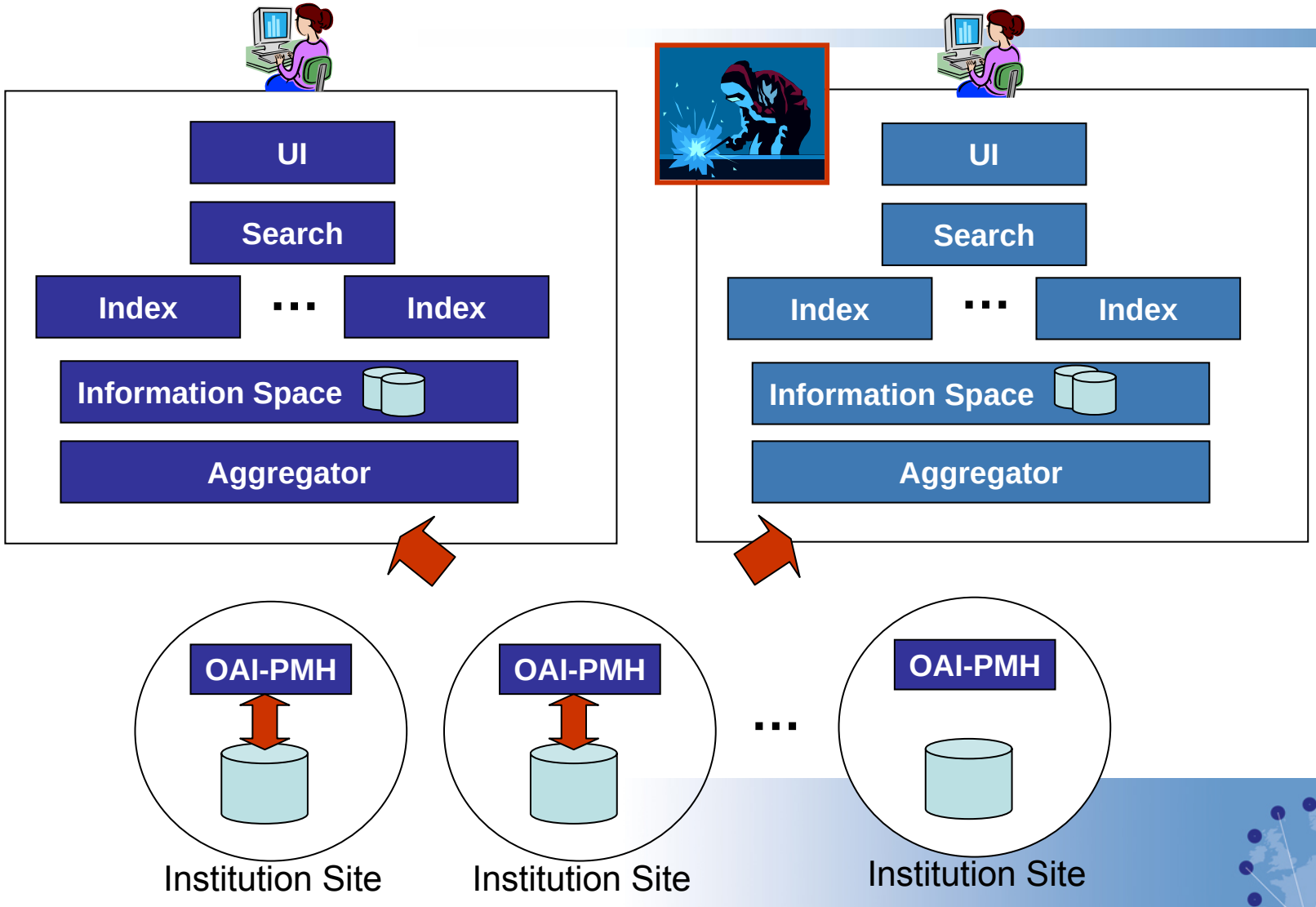
High maintenance costs



Hardly scalable

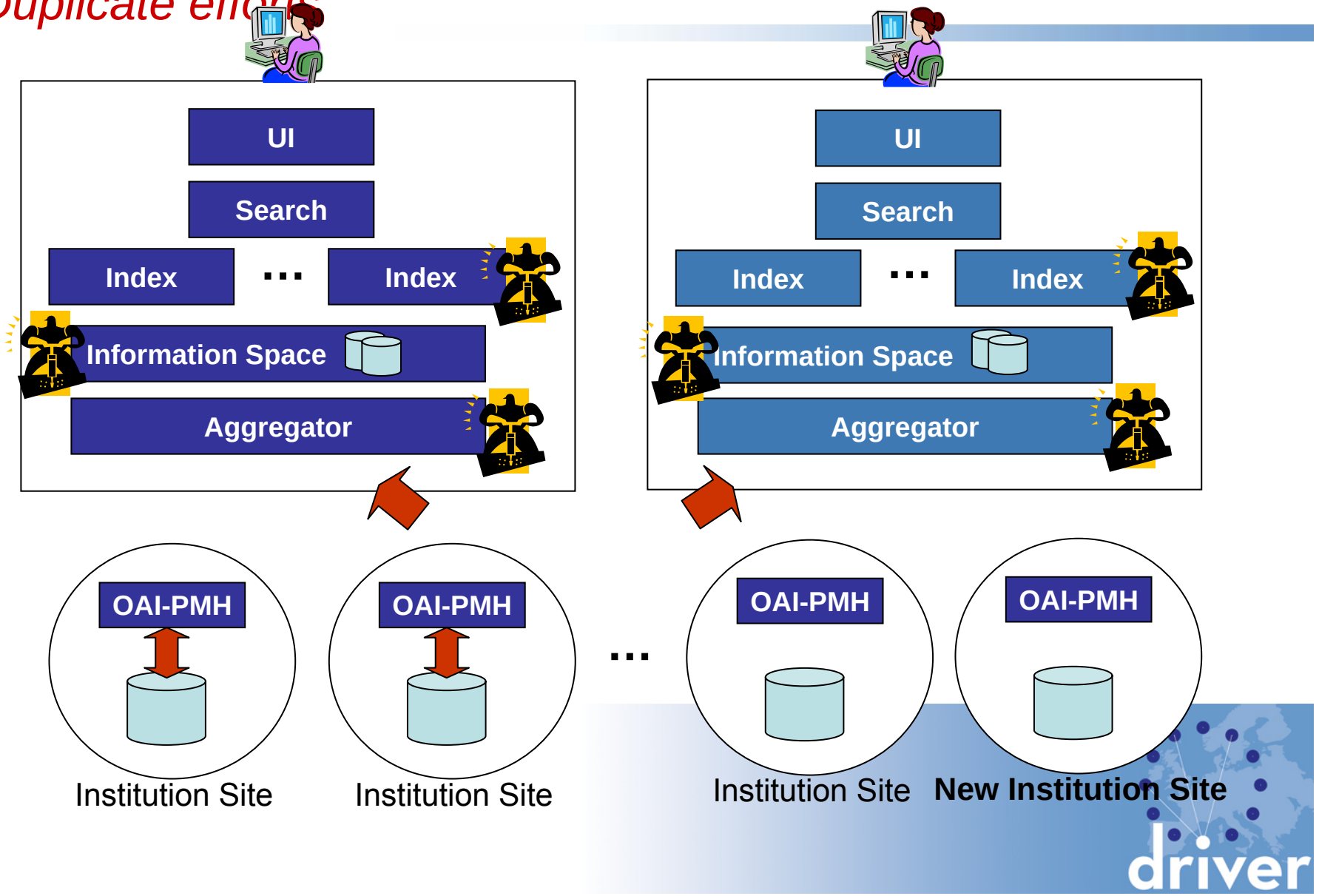


Not reusable



Not reusable

Duplicate efforts



Topics

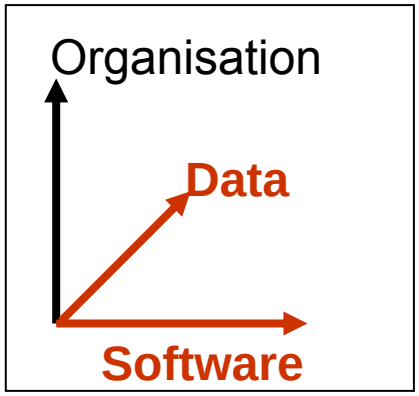
- Why do we need a DRIVER project at all?
- DRIVER basics
- What happens today with repositories?
- Where DRIVER can make a difference!
- DRIVER products
- DRIVER II – outlook

DRIVER Infrastructure Vision

- Build and maintain a **sustainable** European **environment** where **content** and **functionality** resources can be openly **shared** and **integrated** for use by any Application
- Sustainability
  - **Maintainability**
  - **Scalability**
  - **Reusability**

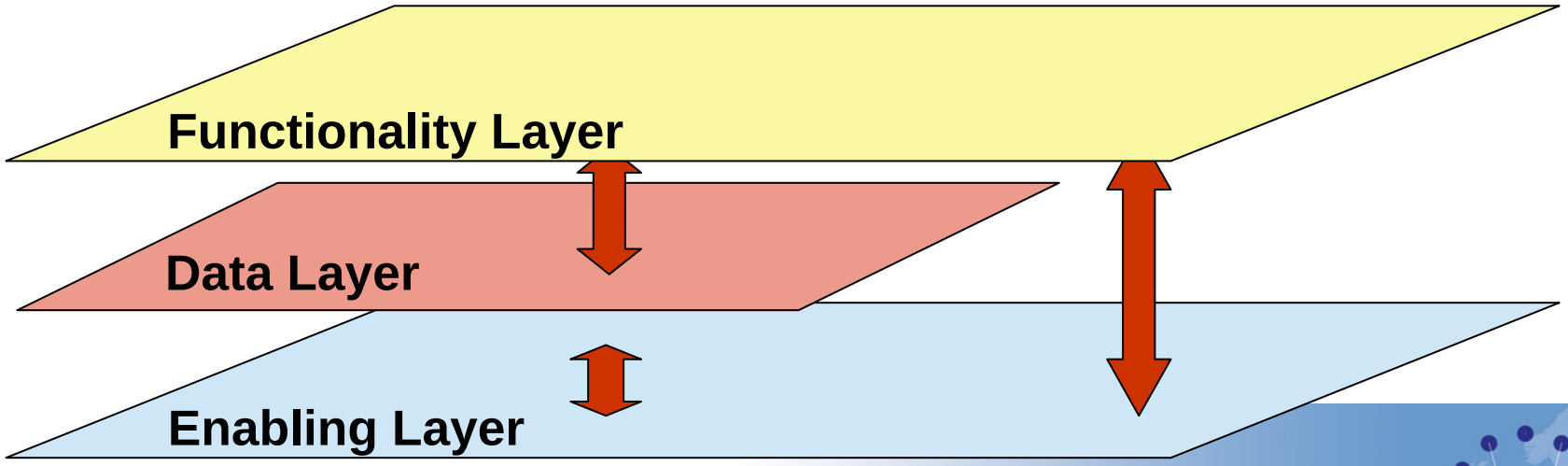
DRIVER Infrastructure and European Information Space

Technical details

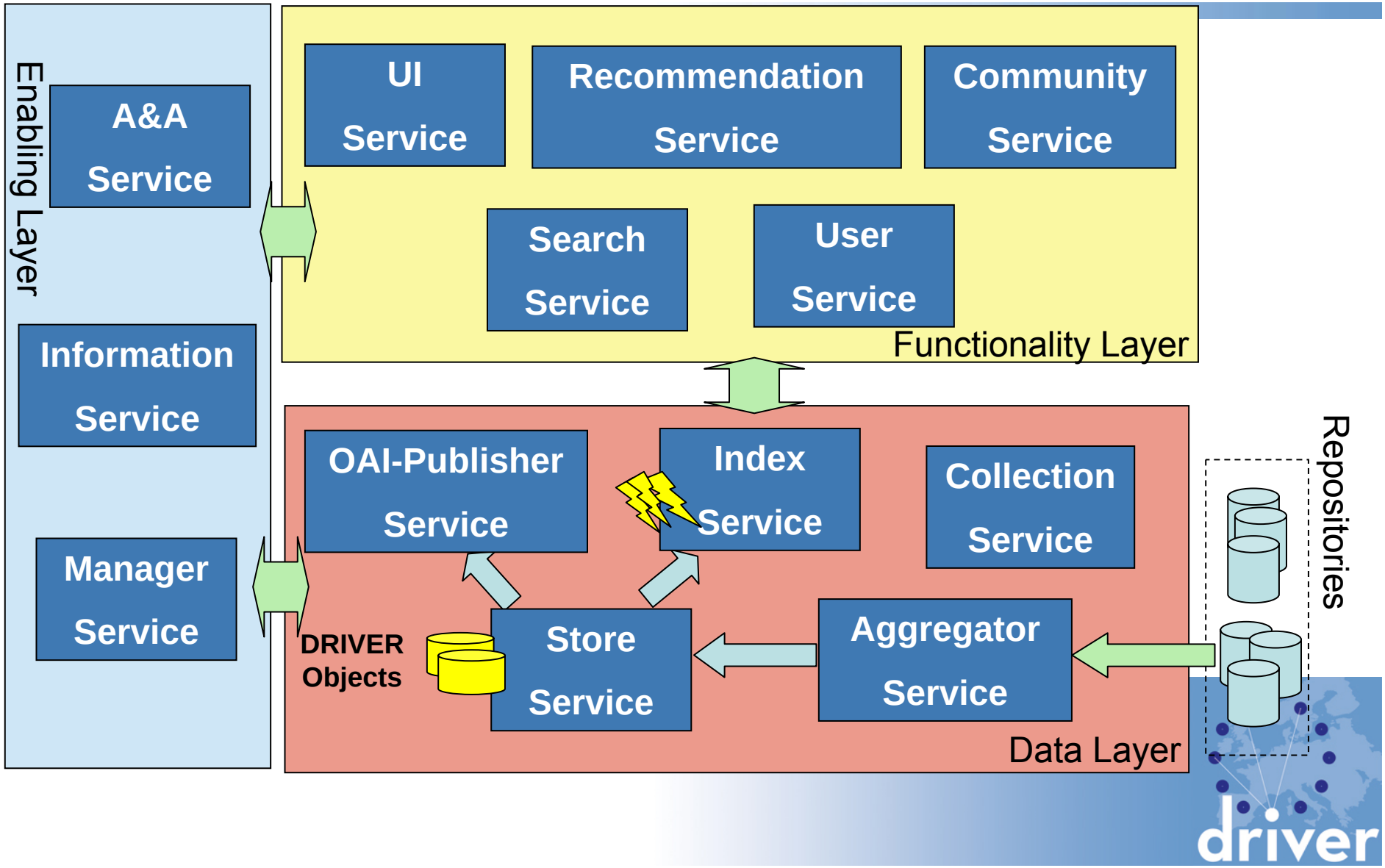


DRIVER Technology Layers

- Enabling Layer
- Service selection
- *European Information Space* Repository System
  - Scalable, reusable, self-maintained

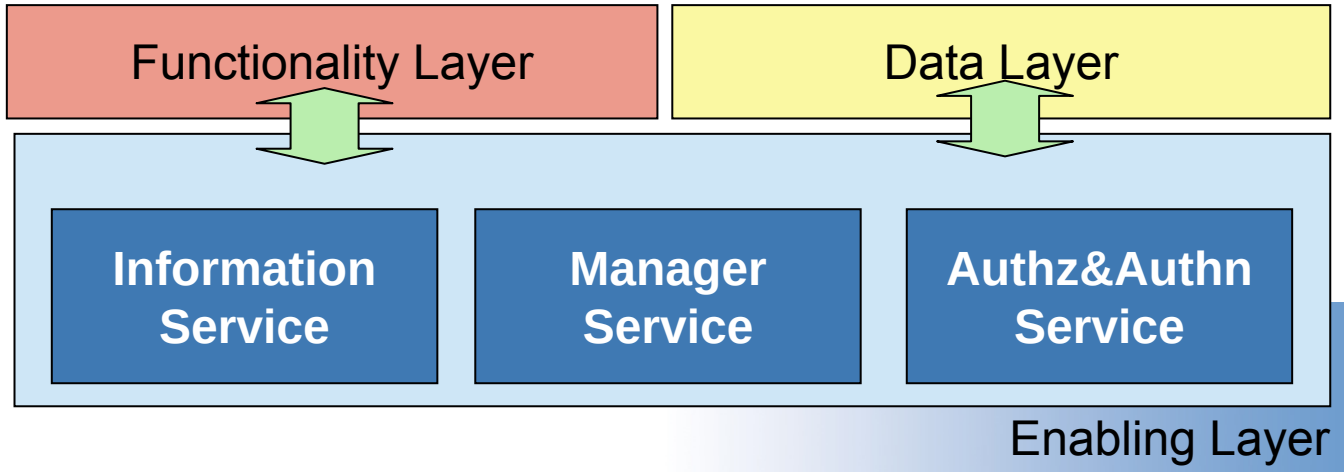


Driver Technology Layers



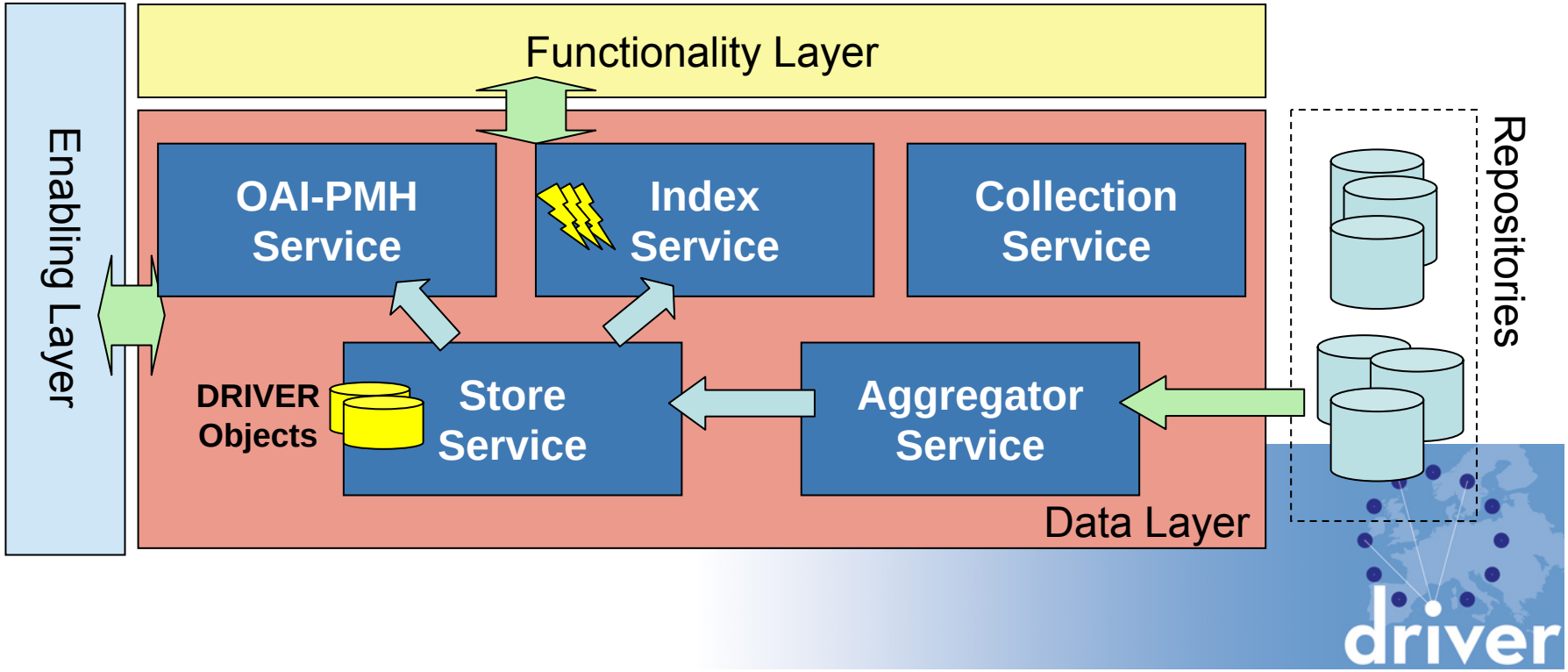
Enabling Layer

- Infrastructure management: common to all Repository Systems
  - Service Registration (P2P-like) and Discovery (by functionalities and capabilities)
  - Subscription and notification on Service-related events
  - Authentication and Authorization Service
  - Orchestration of Services: coordination of Service interactions



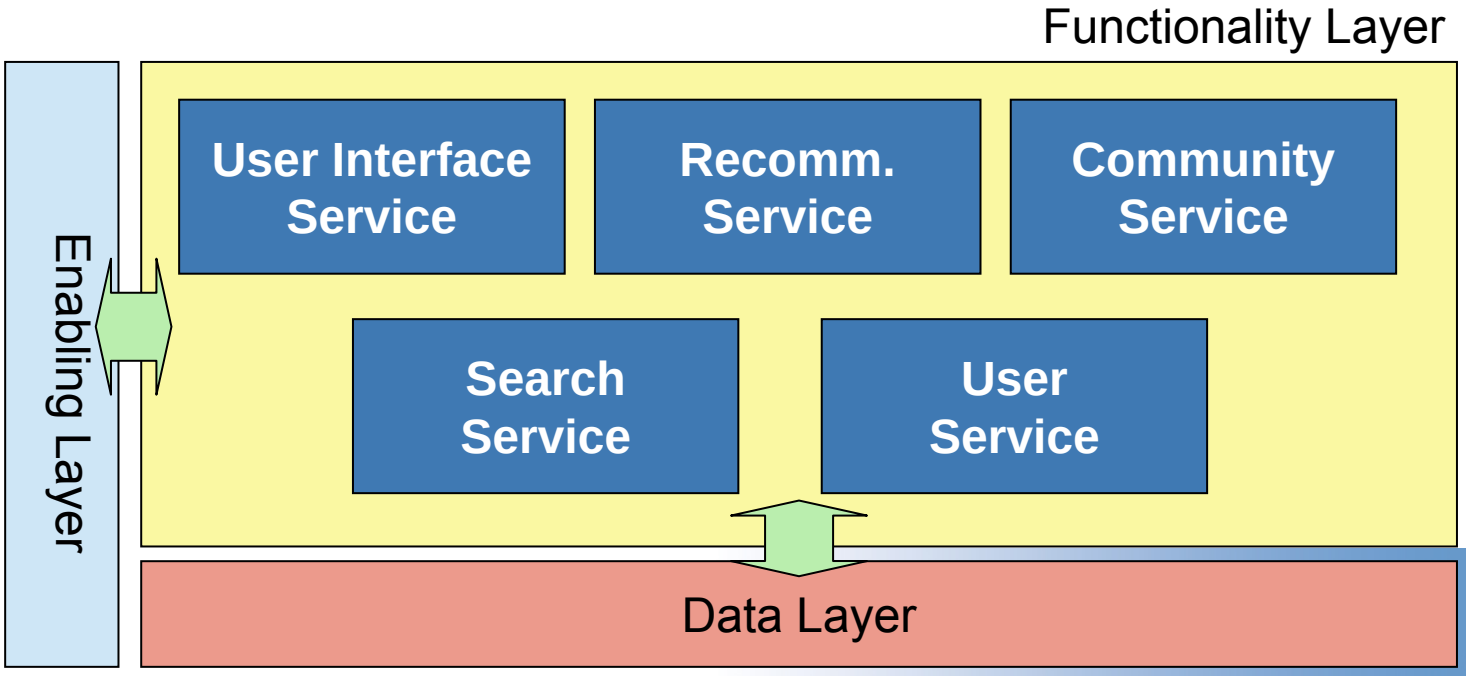
Data Layer

- DRIVER Information Space Management
  - Population: constructing DRIVER Objects from Objects harvested from external Repositories (only metadata)
  - Storage, indexing, OAI-Publishing of DRIVER Objects
  - Virtualization of the Information Space: collections



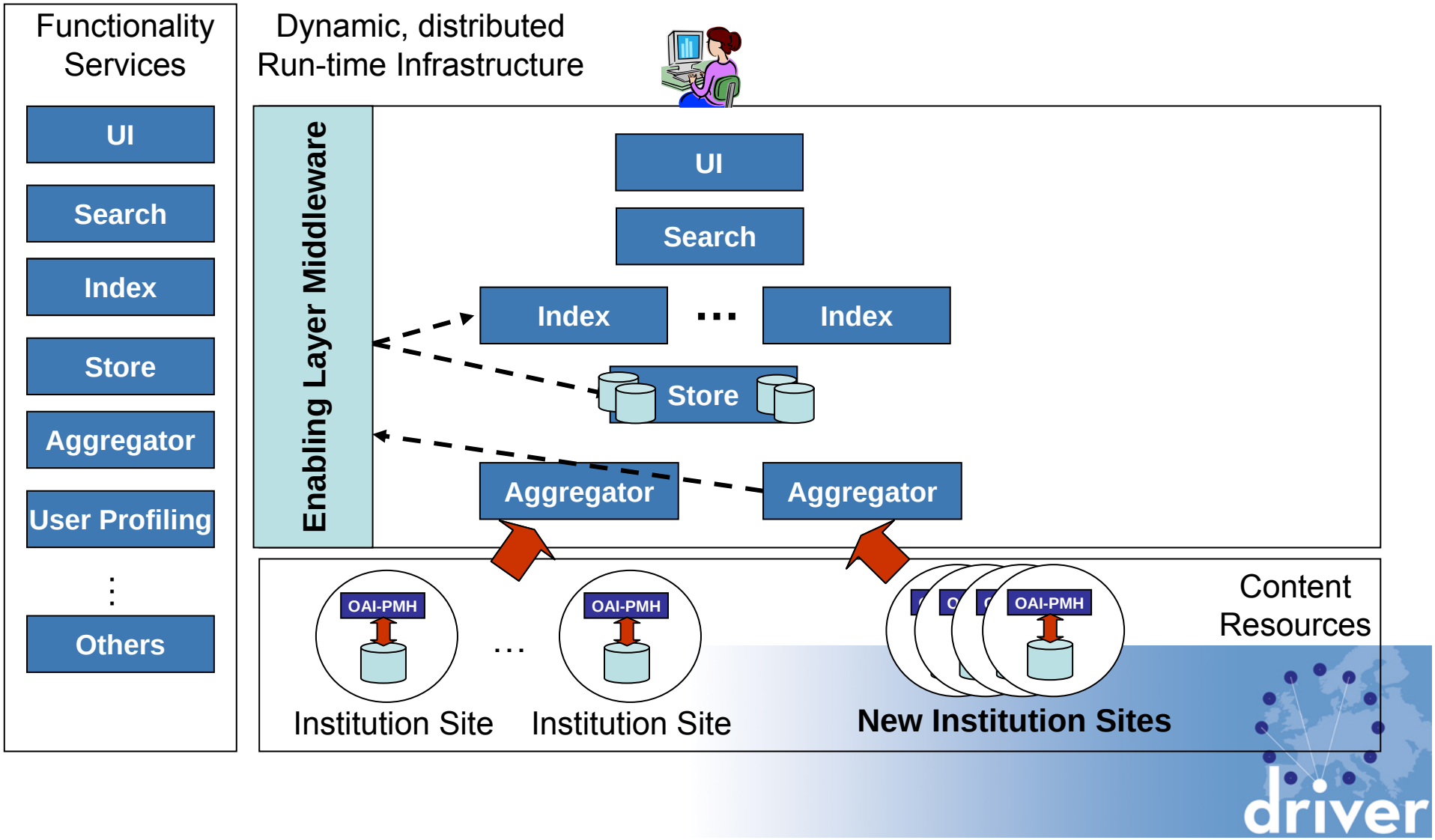
Functionality Layer

- User-oriented functionalities
  - User and Communities profiling
  - User recommendations
  - User Interfaces: Information Space Search



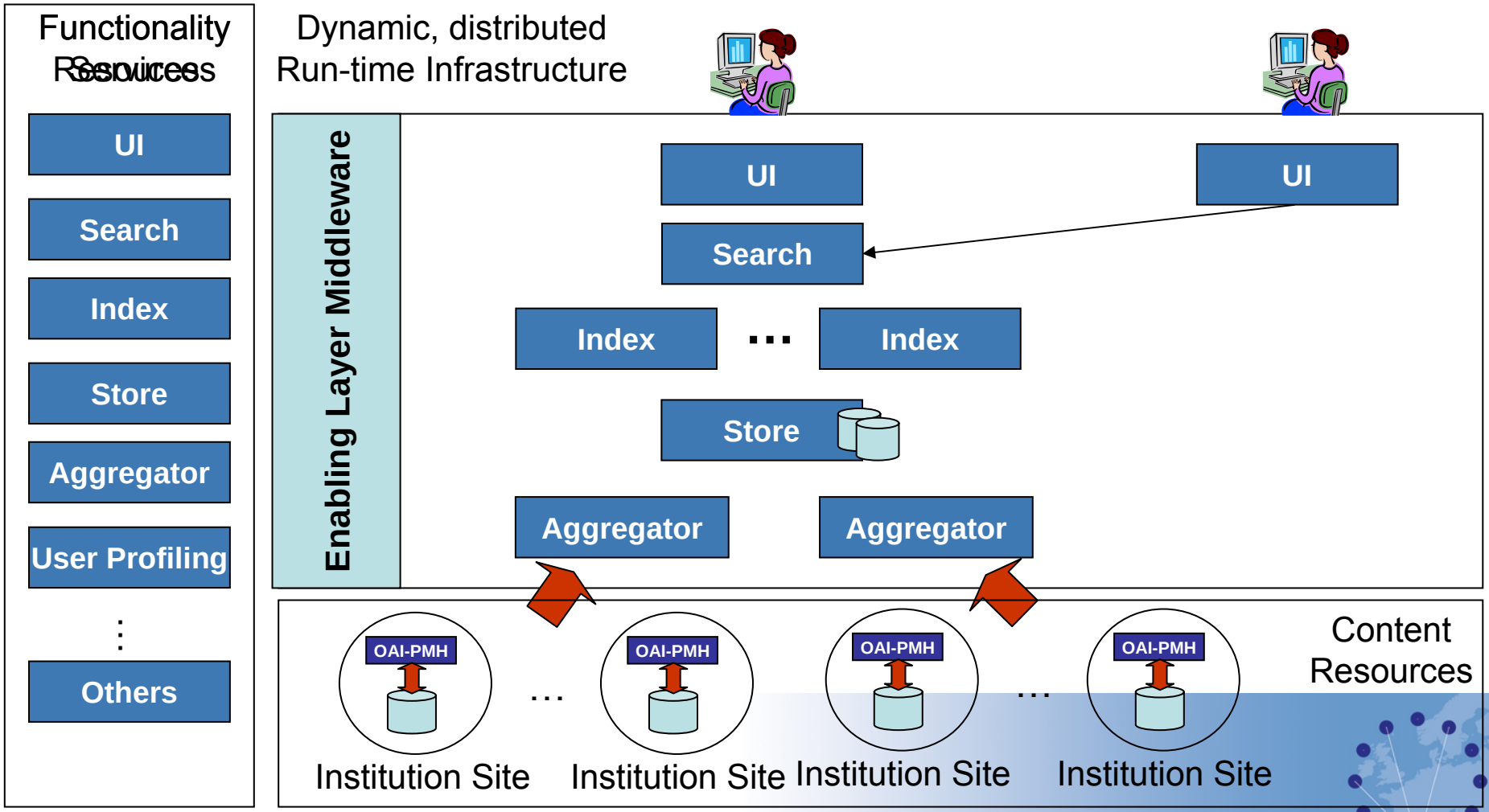
Repository System scalability

Infrastructure Service Orchestration



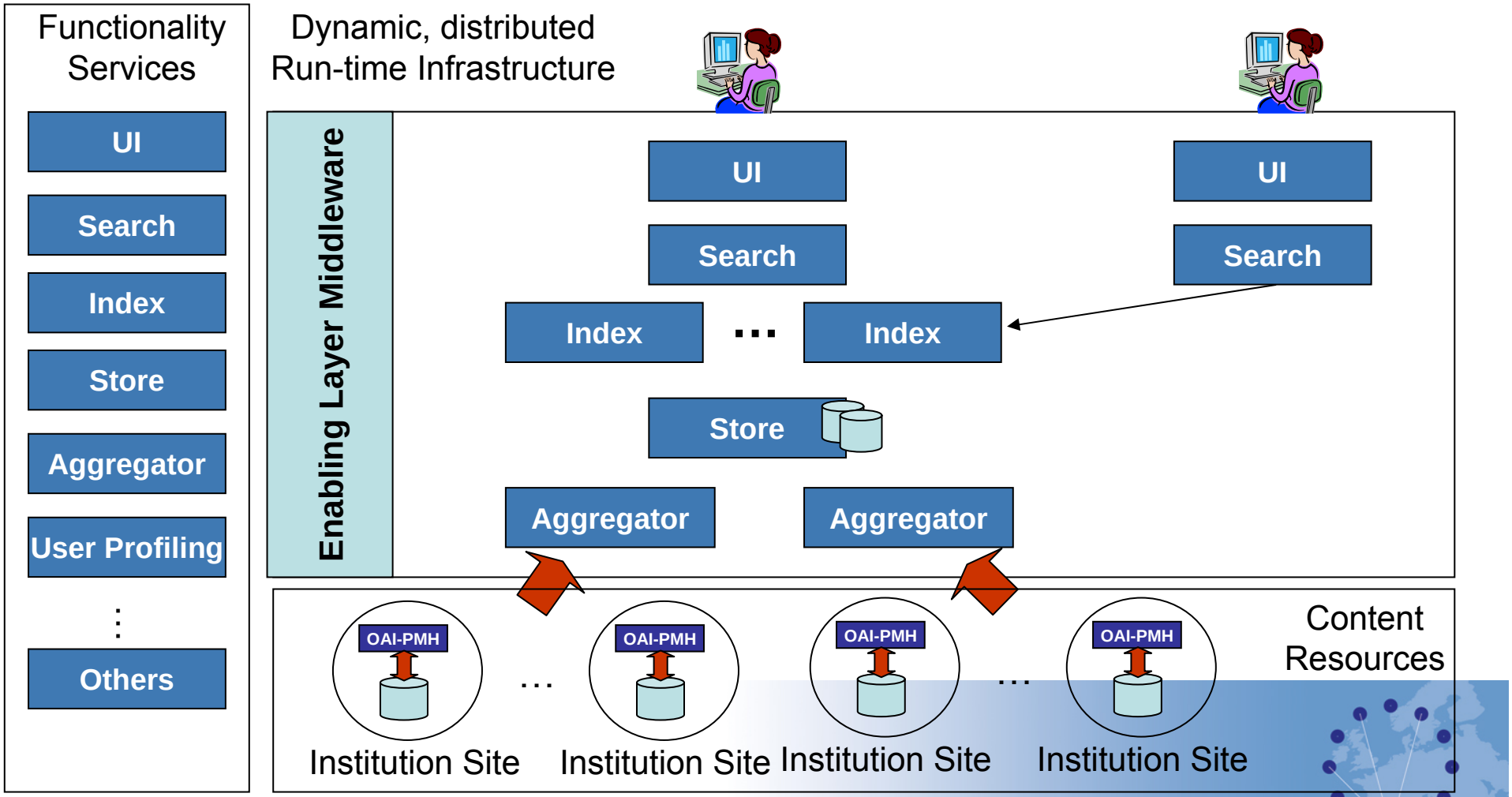
Repository System reuse

Functionality sharing #1



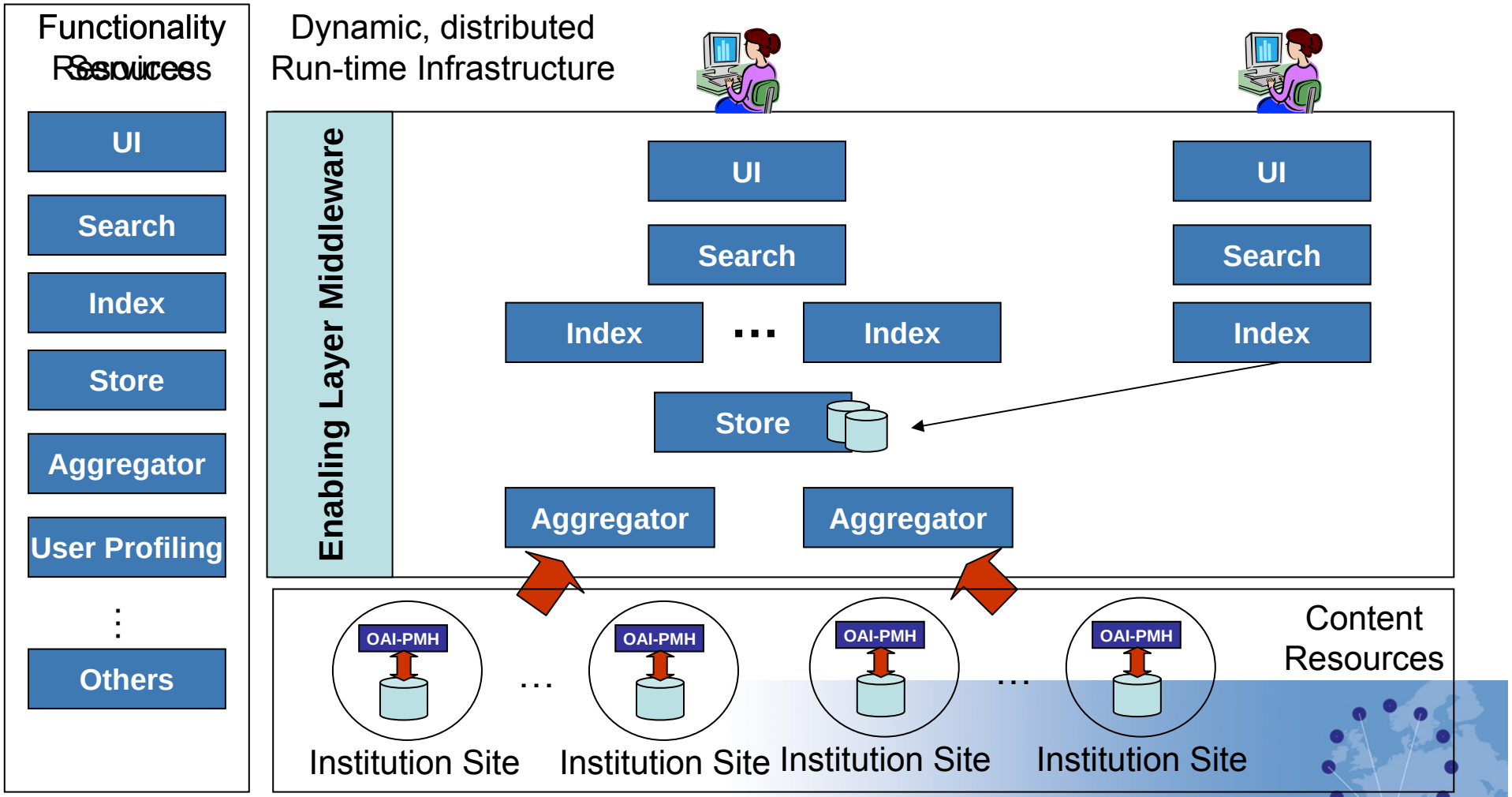
Repository System reuse

Functionality sharing #2

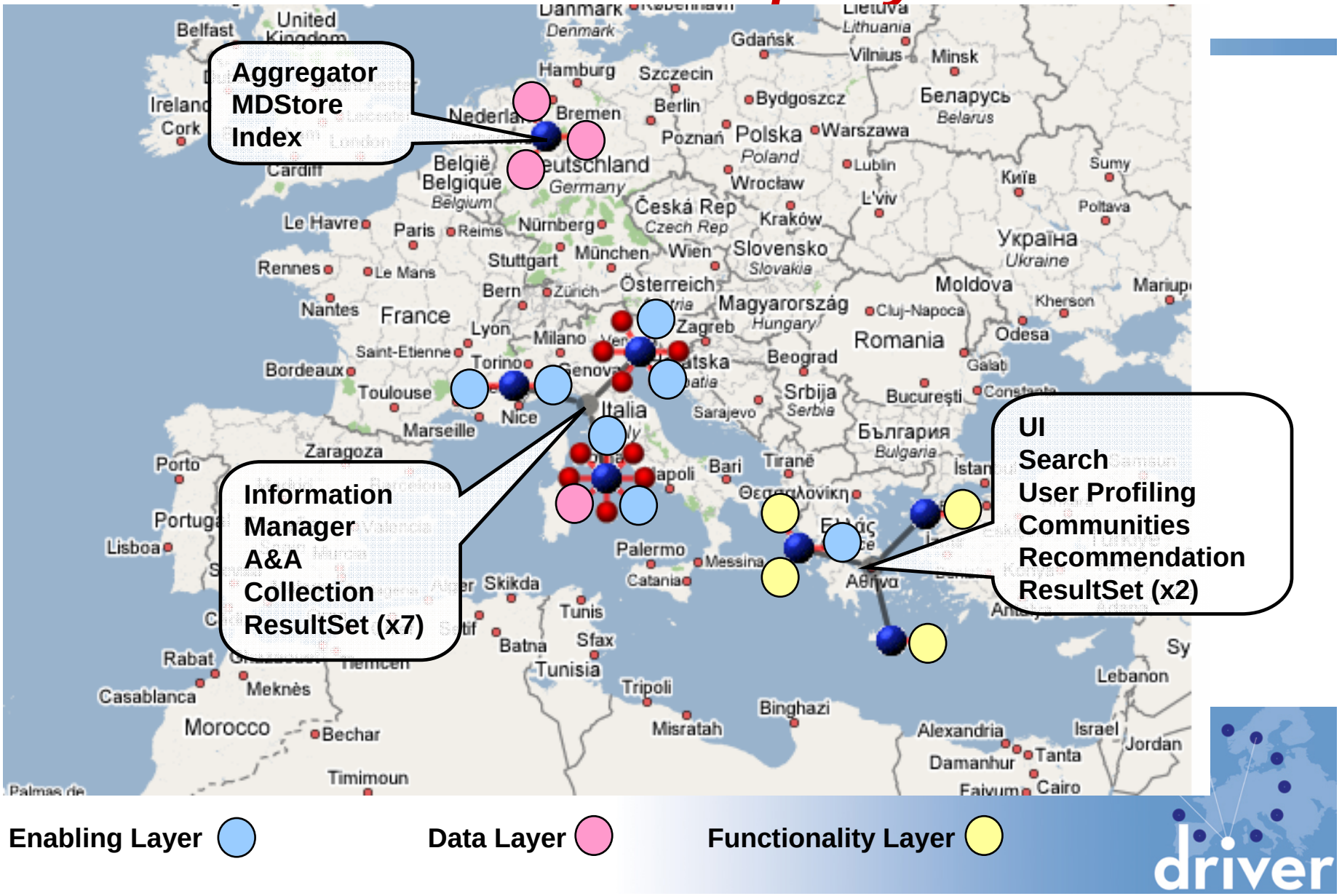


Repository System reuse

Functionality sharing #3



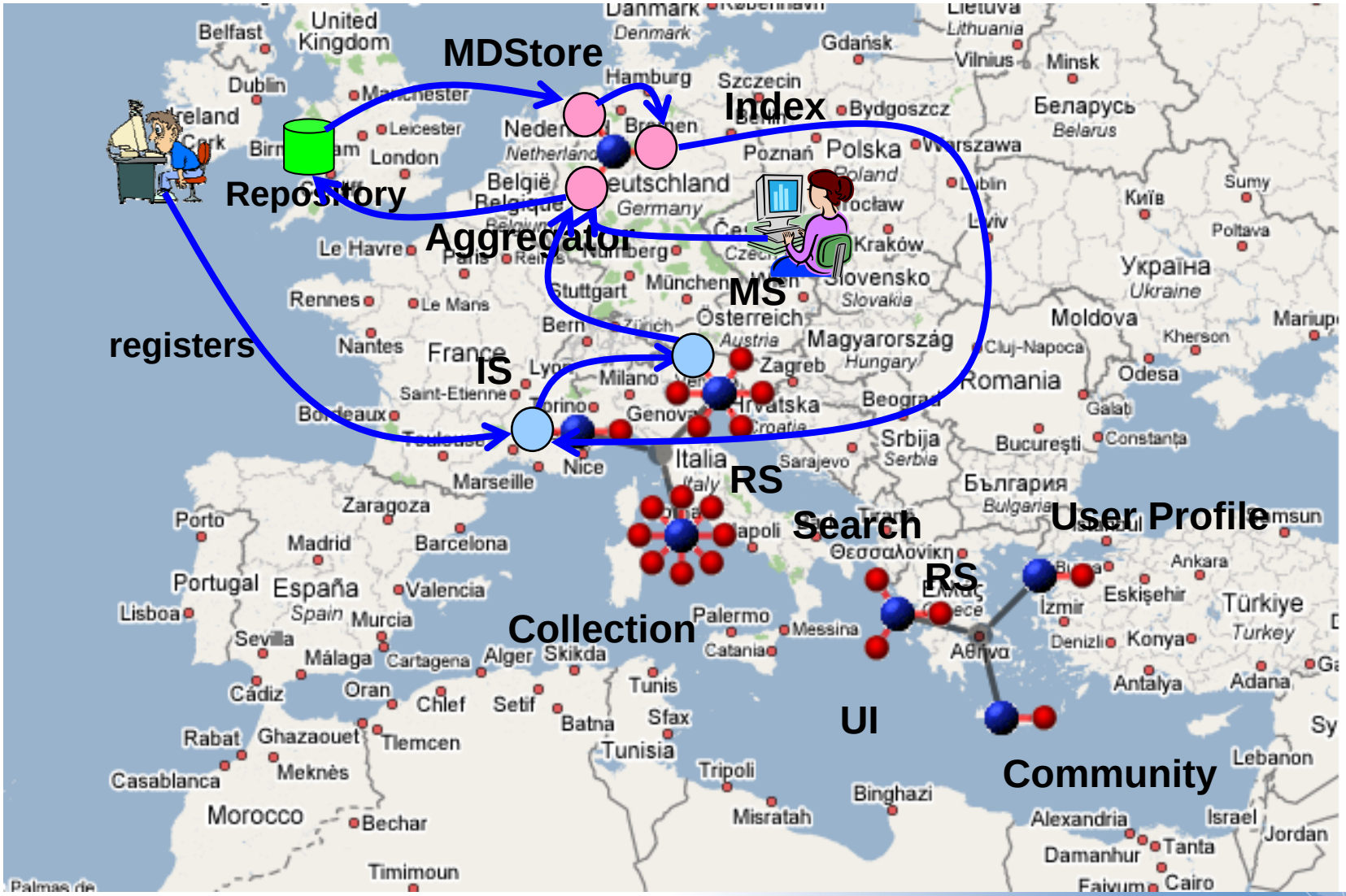
Current DRIVER deployment



Demo Scenario #2: Repository Managers

- Repository registration
- Harvesting
- Aggregating
- Indexing

Repository Registration-Harvesting-Indexing



Topics

- Why do we need a DRIVER project at all?
- DRIVER basics
- What happens today with repositories?
- Where DRIVER can make a difference!
- DRIVER products
- DRIVER II – outlook

DRIVER Products/Deliverables

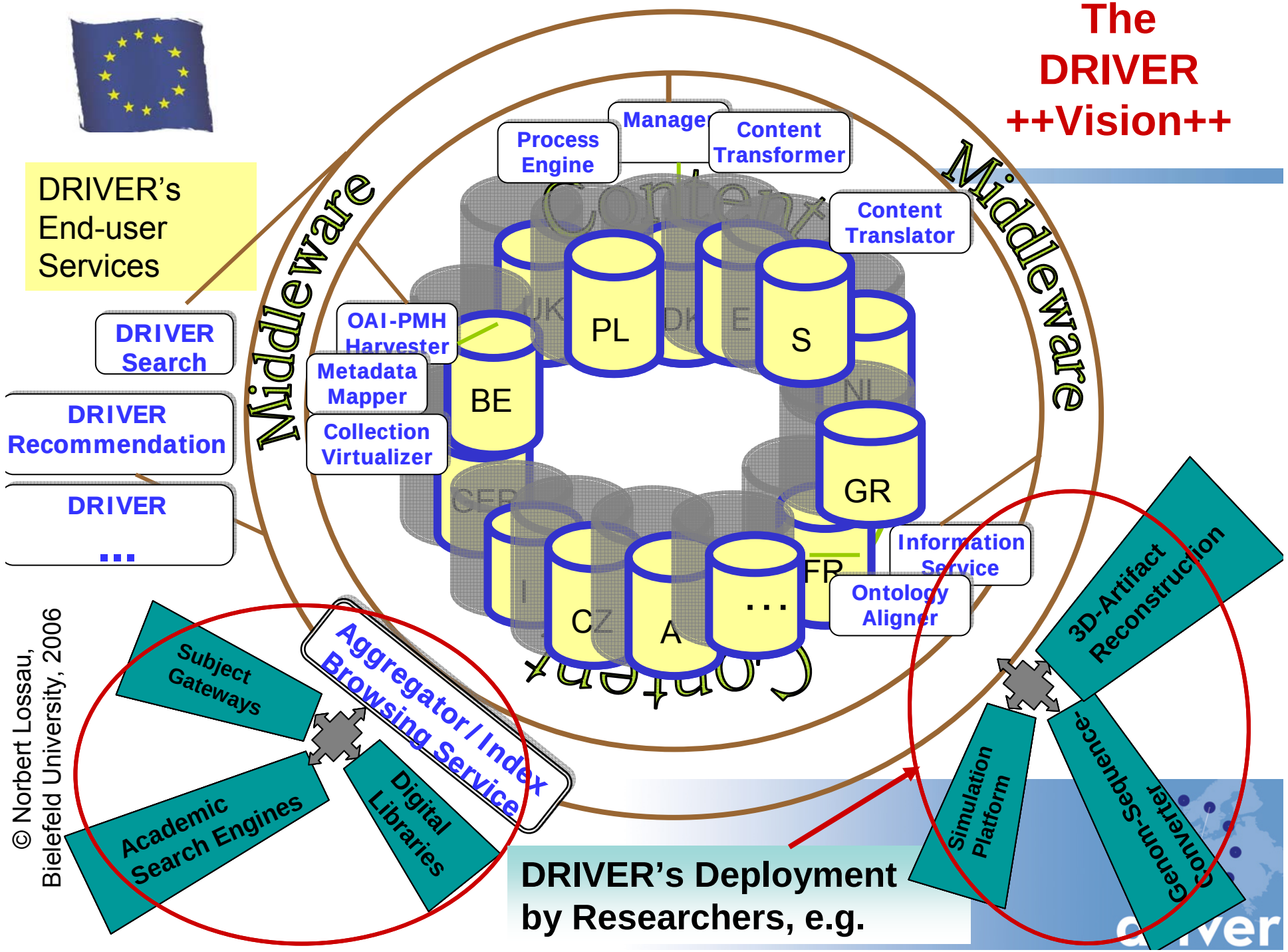
1. An organisational framework for repository providers
2. An established network of repository providers from 5 countries
3. Praxis guidelines to make your local repository fit for the European Network
4. A harmonised data index of more than 51 institutional repositories for open re-use
5. Content: multidisciplinary documents, all open access, various categories (technical papers, pre-/postprints, dissertations etc.)

DRIVER Products/Deliverables

1. A core set of software to aggregate, manage, distribute repository data, open-source, for re-use
2. A production quality testbed which runs the software modules
3. A website to support the building of local, regional, national repository infrastructures (ERAP)
4. An updated, comprehensive overview of the repository situation in Europe (study)
5. Further studies (incl., „how to populate repositories“, technical standards)

DRIVER success factors as infrastructure project

- Service Providers deploying the data index (or portions of it), like SUMMA, Digital Library System, Aarhus, DK
- National, regional repository aggregators re-using DRIVER software or parts of it (like Germany-planned, South Africa through SURF, India – planned)
- Countries, regions, building repository networks facilitated by DRIVER, re-using DRIVER software (like Belgium)


The DRIVER ++Vision++
DRIVER's End-user Services
DRIVER Search
DRIVER Recommendation
DRIVER ...
Middleware
Middleware
Process Engine
Manager
Content Transformer
Content Translator
OAI-PMH Harvester
Metadata Mapper
Collection Virtualizer
Content
PL
S
BE
GR
CZ
A
...
FR
Information Service
Ontology Aligner
Aggregator / Index Browsing Service
Subject Gateways
Academic Search Engines
Digital Libraries
3D-Artifact Reconstruction
Simulation Platform
Genom-Sequence-Converter
DRIVER's Deployment by Researchers, e.g.
© Norbert Lossau, Bielefeld University, 2006

Topics

- Why do we need a DRIVER project at all?
- DRIVER basics
- What happens today with repositories?
- Where DRIVER can make a difference!
- DRIVER products
- DRIVER II – outlook

General Information: DRIVER II

- Duration: 24 months
- Budget: 2.7 MEUR
- Timeplan: 12/'07 -11/'09
- Main Deliverables:
  - Digital Repository Infrastructure
  - European Digital Repository Organisation
- Funded by the European Commission, "Research Infrastructure" Unit, FP 7
- Partners
  - Univ.of Athens (GR)
  - Univ. of Bielefeld (GE)
  - CNR-ISTI (IT)
  - STICHTING SURF (NL)
- Further Partners
  - CNRS-CCSD (FR)
  - Univ. of Bath (UK)
  - Univ. of Warszawski (PO)
  - Univ. of Gent (BE)
  - Univ. of Gottingen (GE)
  - Danish Technical University (DK)
  - Universidade do Minho (PT)
  - Narodna in univerzitetna knijznica (SLO)

Outlook: DRIVER II Products (Planning)

1. An organisation of repository providers
2. An established network of repository providers from 15 (= 5+10) countries
3. A harmonised data index of more than 150 (50+100) institutional repositories for open re-use
4. Content: multidisciplinary documents + other data

Outlook: DRIVER II Products (Planning)

5. An expanded set of software to aggregate, manage, distribute repository data, open-source, for re-use
6. A production quality infrastructure which runs the software modules
7. An expanded website to support the building of local, regional, national repository infrastructures (ERAP)
8. Connectors to long term preservation systems
9. Further studies (e.g. primary data in repositories)

DRIVER is liaising with…

- National, regional organisations of digital repositories providers
- International repository projects and initiatives
- Research organisations and research funders
- National and international university organisations
- Existing repository service providers (such as OAIster, Terminology & Classification Services)

DRIVER will liaise with…

- Communities/projects building disciplinary repositories (such as NEREUS, FP 7 projects)

Pro-active approaches to DRIVER from…

- National initiatives in “non-DRIVER“ countries
  - Czech Republic, Finland, Ireland, Lithuania, Norway, Portugal, Slovenia, Sweden
  - South Africa (through eIFL)
  - India…
- Projects such as ORE, DPE, DELOS, BELIEF …
- Organisations such as LIBER, EUA …
- Companies such as Microsoft, Google-Scholar
- …

Summary

DRIVER II;

from an European focal point to an established DR organisation and infrastructure,

actively building a global network of repositories, with international partners

DRIVER Consortium Contacts:

mike@di.uoa.gr

lossau@sub.uni-goettingen.de

whorstmann@sub.uni-goettingen.de

+ Project website:
  www.driver-repository.eu