

# Infrastruktur und Netzwerke für Repositorien



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



# Overview

- General
- DRIVER (Europe)
  - Organisation / Data
  - Data / Software
  - DRIVER-II
- OA-Netzwerk (DE)
  - Rahmenprojekt
  - Zusatzdienste



# *The current situation for digital repositories*

- More than 850 institutional repositories worldwide
  - Many others: disciplinary, national, ...
  - Many types: Primary data, textual documents, learning materials, multimedia objects, Code etc.
  - Documents: incl. pre-prints, postprints, technical papers, dissertations, theses
  - Various repository software



# *What is known about repositories?*

- Many have the OAI-PMH implemented
  - small but relevant local specialties
- Some international registries exist
  - OpenDOAR, ROAR ...
- Some search engines exist
  - BASE, OAIster, Google Scholar ...



# *Collaboration in repositories*

- Very few mature national repository organisations/collaborations
- No trans-national repository organisation/collaboration
- Lack of data harmonization, orchestration of services



# *From the user point of view*

*[ talking about researchers ]*

- Fragmented, obscure information landscape
- content can be (partly) searched and found
- quality and re-use differs from repository to repository



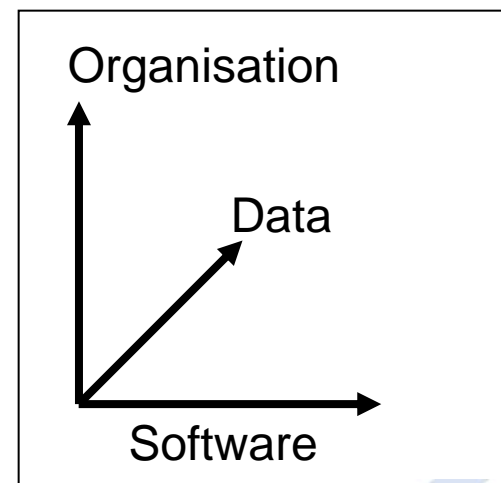
# Overview

- General
- DRIVER (Europe)
  - Organisation / Data
  - Data / Software
  - DRIVER-II
- OA-Netzwerk (Deutschland)
  - Rahmenprojekt
  - Zusatzdienste



# Need for Repository Infrastructure

- Organisational Networking
  - Forum for stakeholders
- Data harmonization
  - Cleaned and enriched „information space“
  - Offered for re-use by service providers
- Software orchestration
  - Open, scalable technical architecture
  - Re-use of existing services
  - Facilitate interactions between services



# *DRIVER Vision*

- Build a **pan-European Digital Repository Infrastructure**
- and provide the **European Hub in a Global Repository Network**

for

- any type of document, data and object,
- of any format,
- involving all European countries and
- covering all academic disciplines



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



# *DRIVER now!*

## A Digital Repository Infrastructure Test-bed

- Textual documents
- from 5 Countries (BE, FR, GE, NL, UK)
- Distributed, high-quality content provision
- Distributed software deployment
- All open and re-usable



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



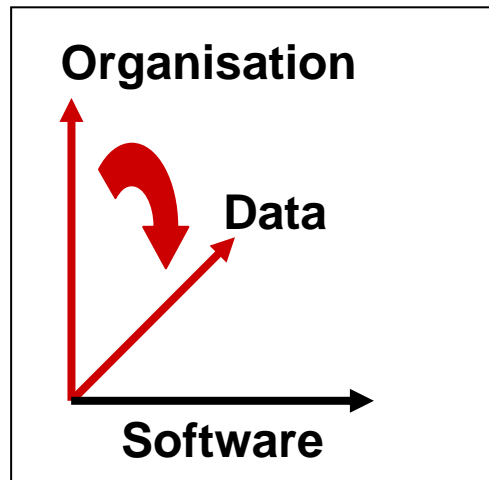
# Overview

- General
- DRIVER (Europe)
  - Organisation / Data
  - Data / Software
  - DRIVER-II
- OA-Netzwerk (Deutschland)
  - Rahmenprojekt
  - Zusatzdienste



# *Repository networking*

*Organisation being synergetic  
with data infratstructure*



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



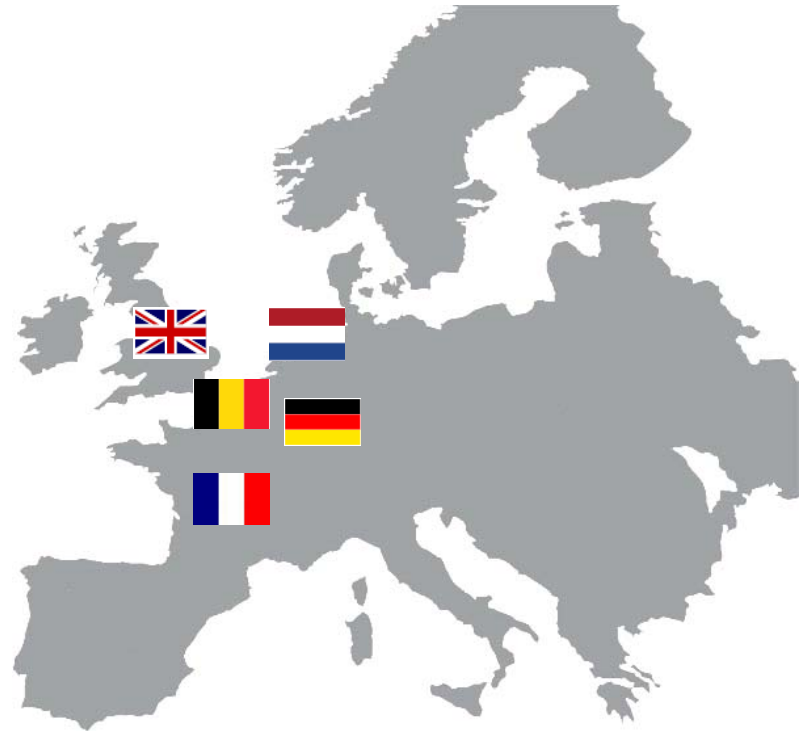
# “Make it workable”

- Focus on **existing** repositories and services
- Focus on **Institutional** Repositories
  - Rapid **progress** over the last years
  - Inherent **sustainability** (e.g. libraries)
  - Adequate technical **homogeneity** (OAI-PMH)
- Focus on **textual** materials
- Focus on specific (test-bed) **countries**



# Networking test-bed

- 5 countries
  - DRIVER exemplars
  - BE, DE, FR, NL, UK
- Different
  - Maturity
  - Policies
  - Technologies
  - ...

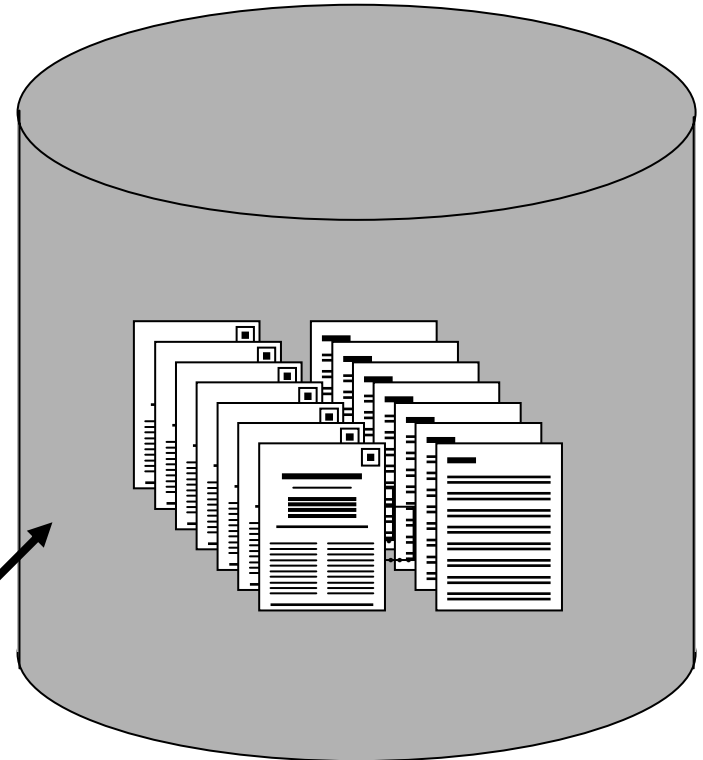
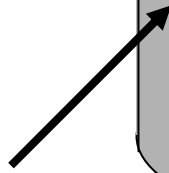
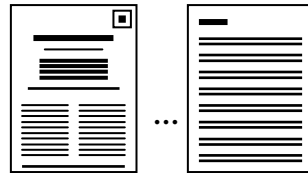
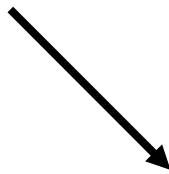


# Networking test-bed

- 50+ “organized” repositories
- *Is there a simple technical recipe for networking these repositories?*



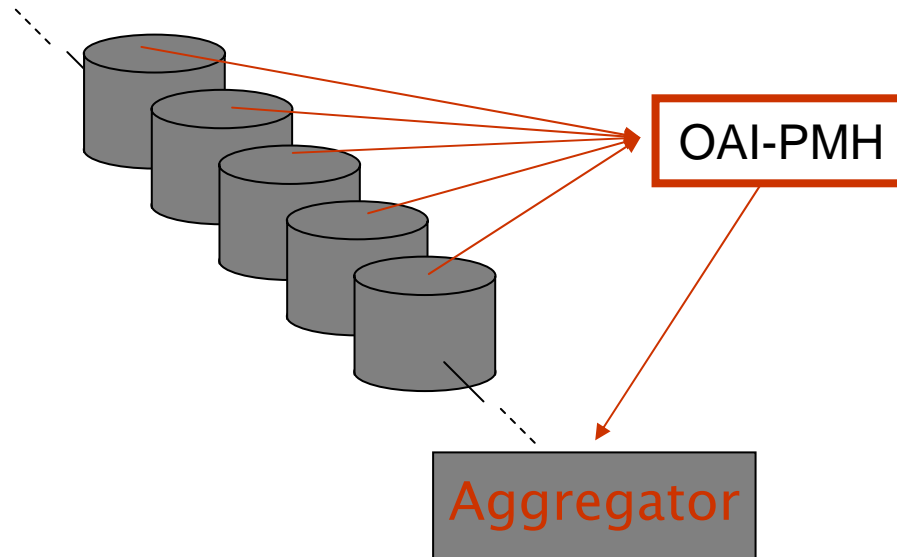
# All publications in a repository



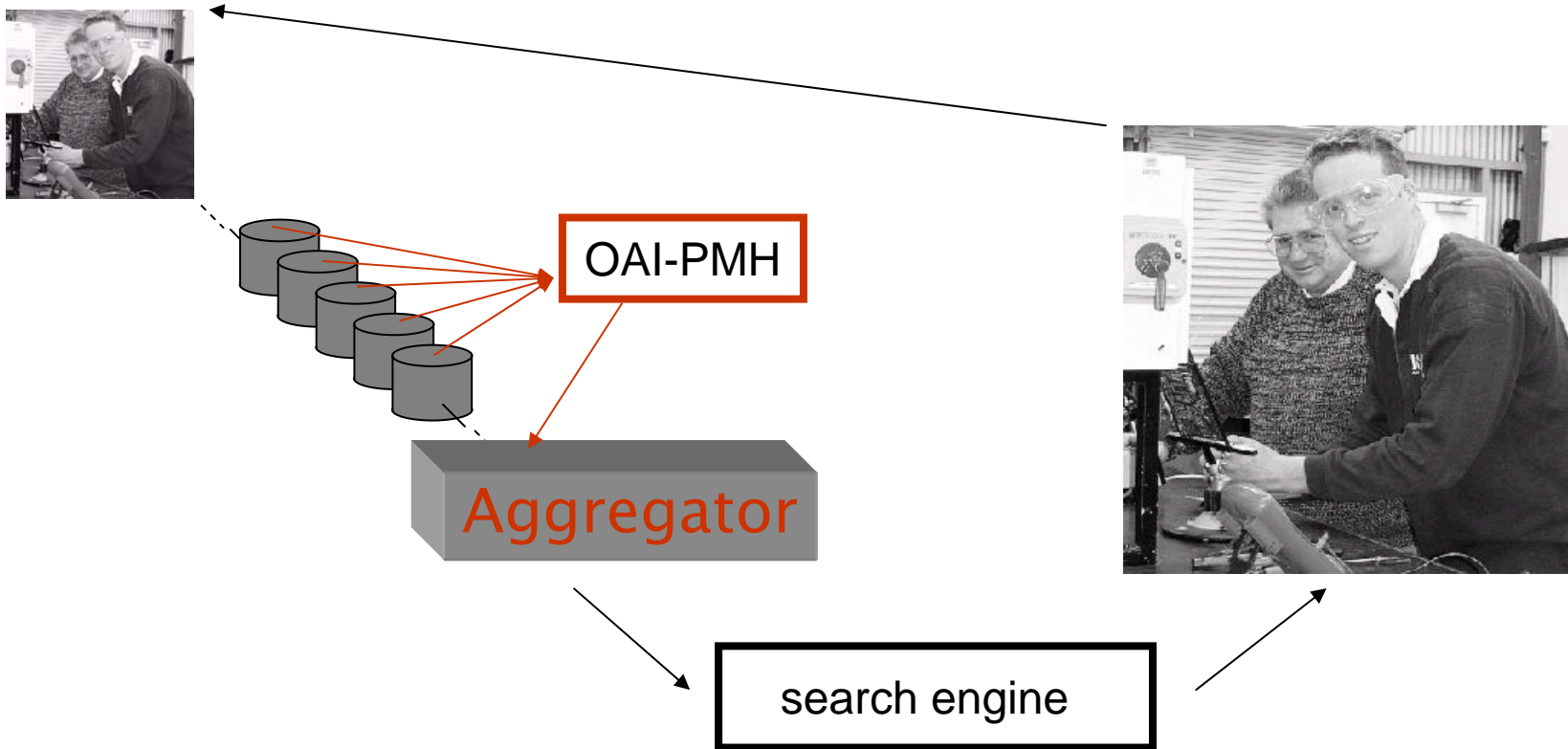
Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



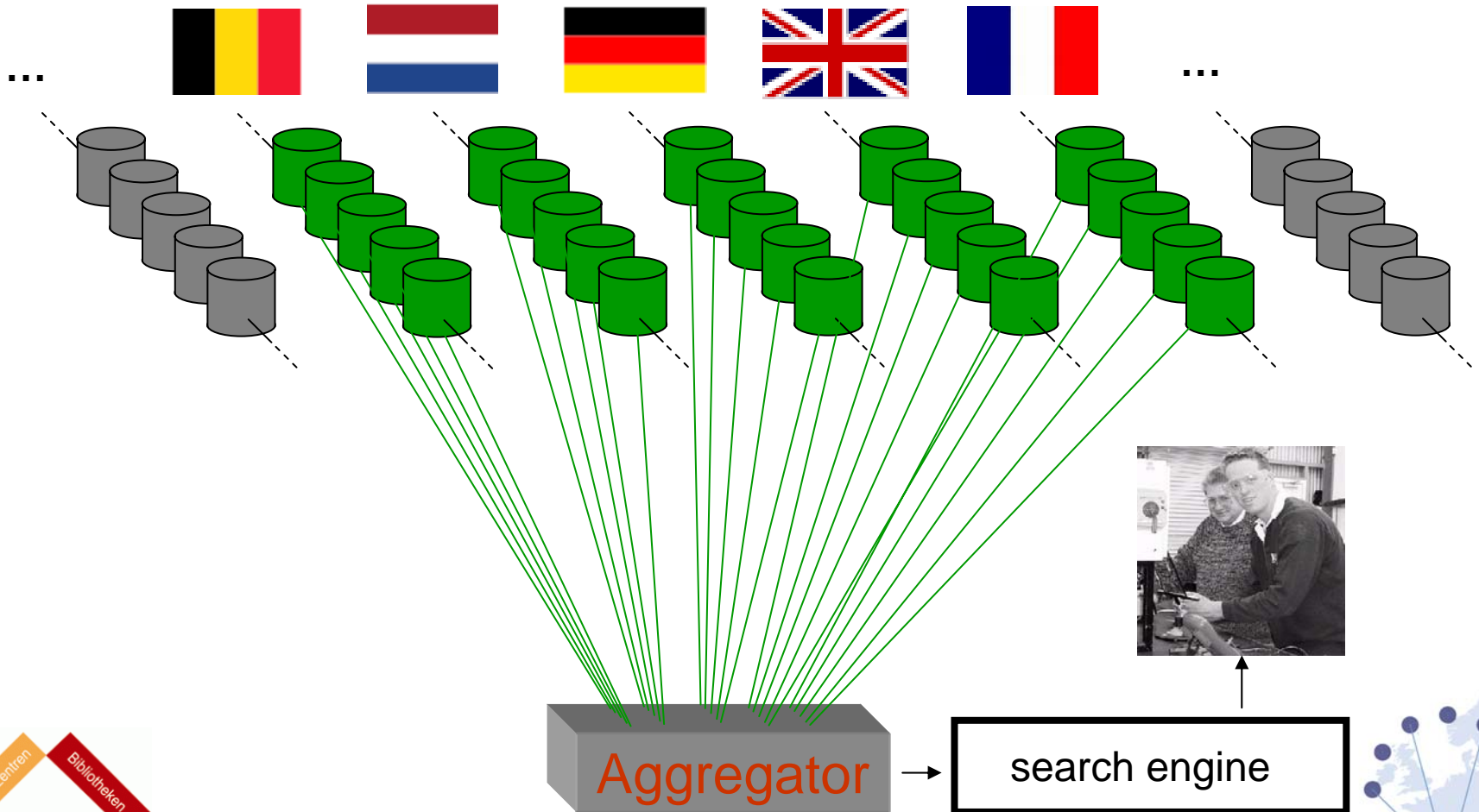
# Aggregate with OAI-PMH



# Feed back to research



# Lift to an international scale



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



# *Problems remain*

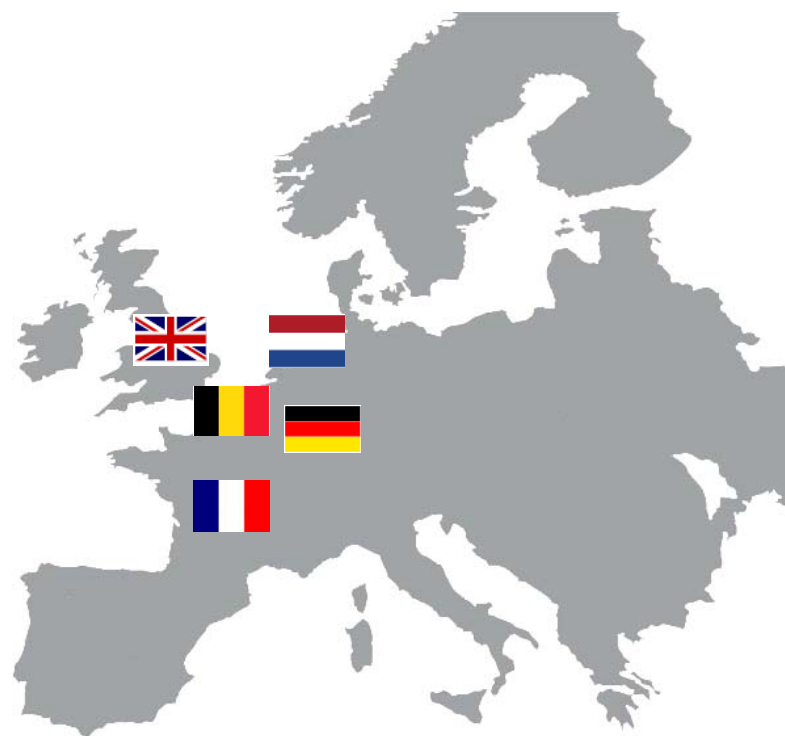
## *Building of aggregations is not trivial!*

- An **example**: Full-text availability
  - Full-text not homogeneously offered
    - metadata only, jump-off pages, authorization
    - full-text URLs in varying locations



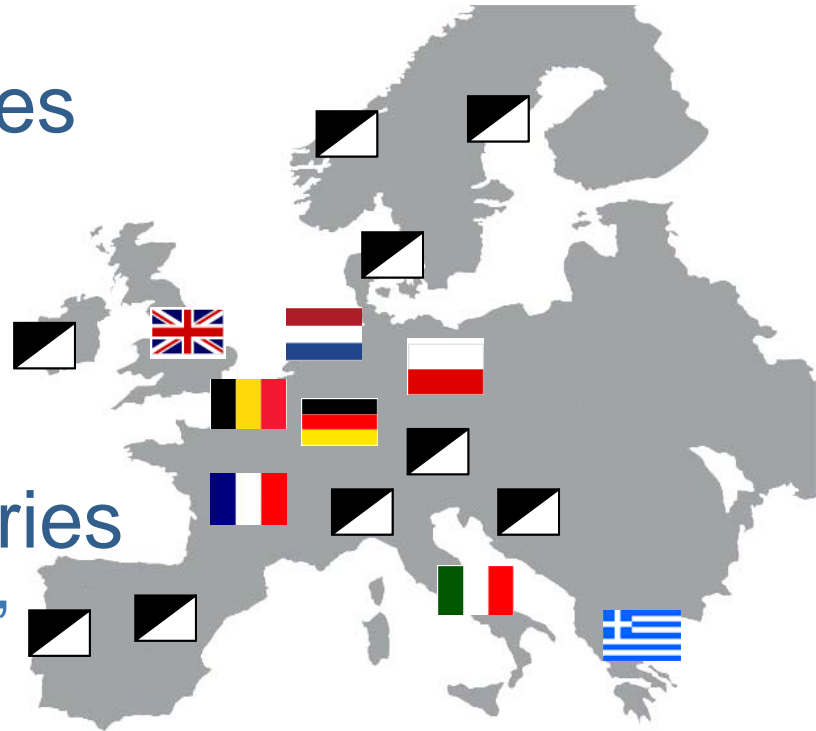
# Full-text guidelines

- Help to **improve** the repository landscape
  - Support of “sets”
  - Clear location of URL
  - Resource harvesting
- Studies say:
  - International retrieval
  - Guidelines



# Next Steps

- Help **50+** repositories
  - Guidelines release
  - Help Desk (WP8)
  - Validator [\[1\]](#) [\[2\]](#)
- Help **more** repositories
  - “DRIVER countries”
  - other countries
  - ...

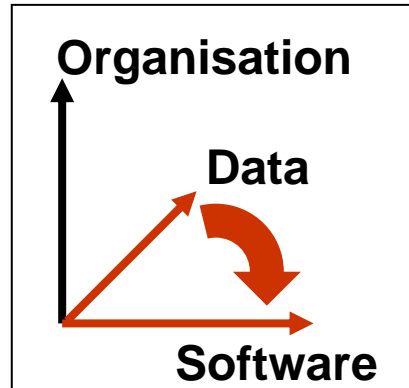


# Overview

- General
- DRIVER (Europe)
  - Organisation / Data
  - Data / Software
  - DRIVER-II
- OA-Netzwerk (Deutschland)
  - Rahmenprojekt
  - Zusatzdienste



# *DRIVER Software*



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



# *EU Knowledge Infrastructural 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 means:
  - **Maintainability, Scalability, Reusability**

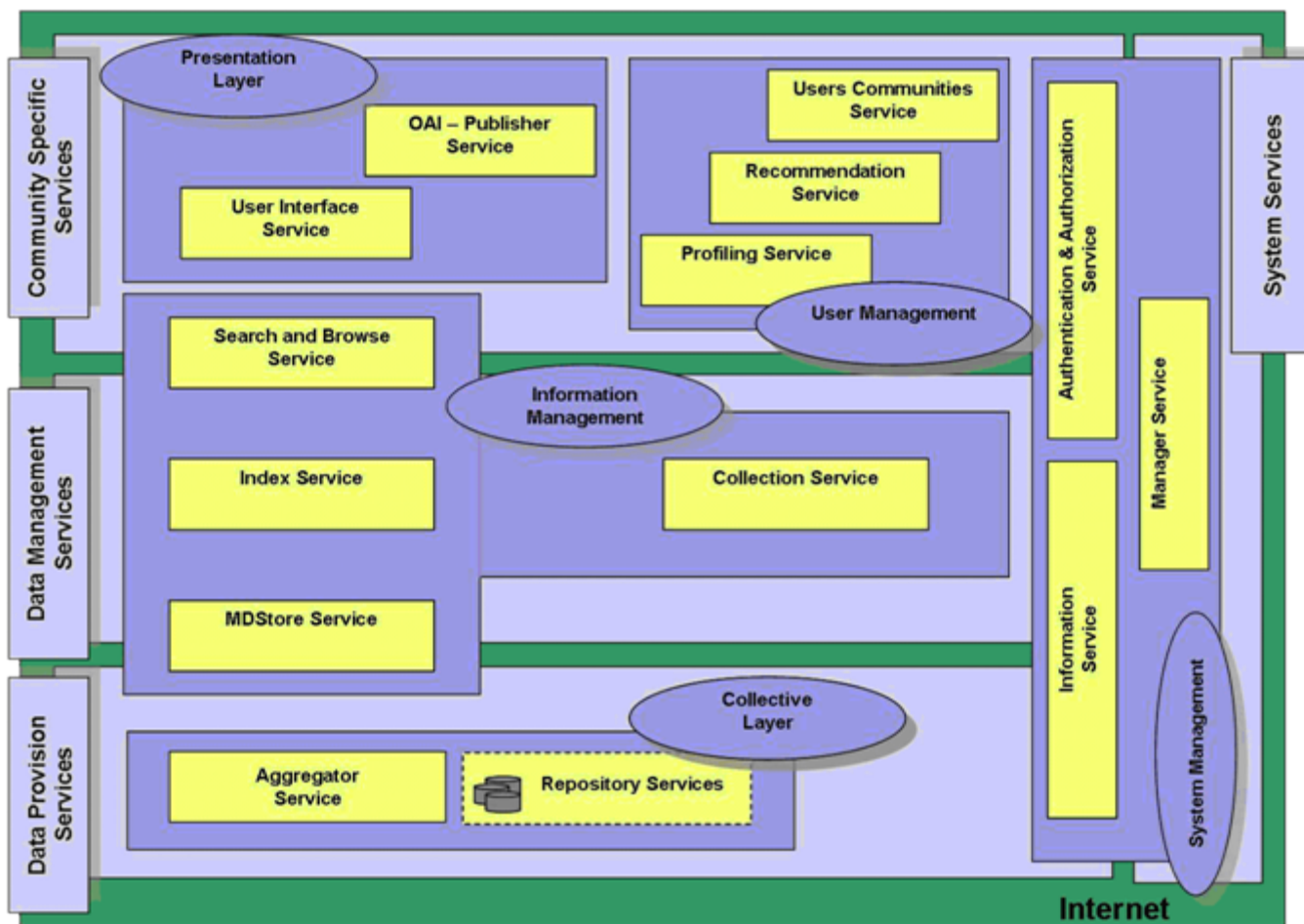


# *Service-based Repository Systems*

- Environment and tools for building *service-based Repository Systems*
  - Sets of **services** running at different **network sites**, possibly in **multiple** instances, **interacting**, **dynamic**, **sharable**, **open**



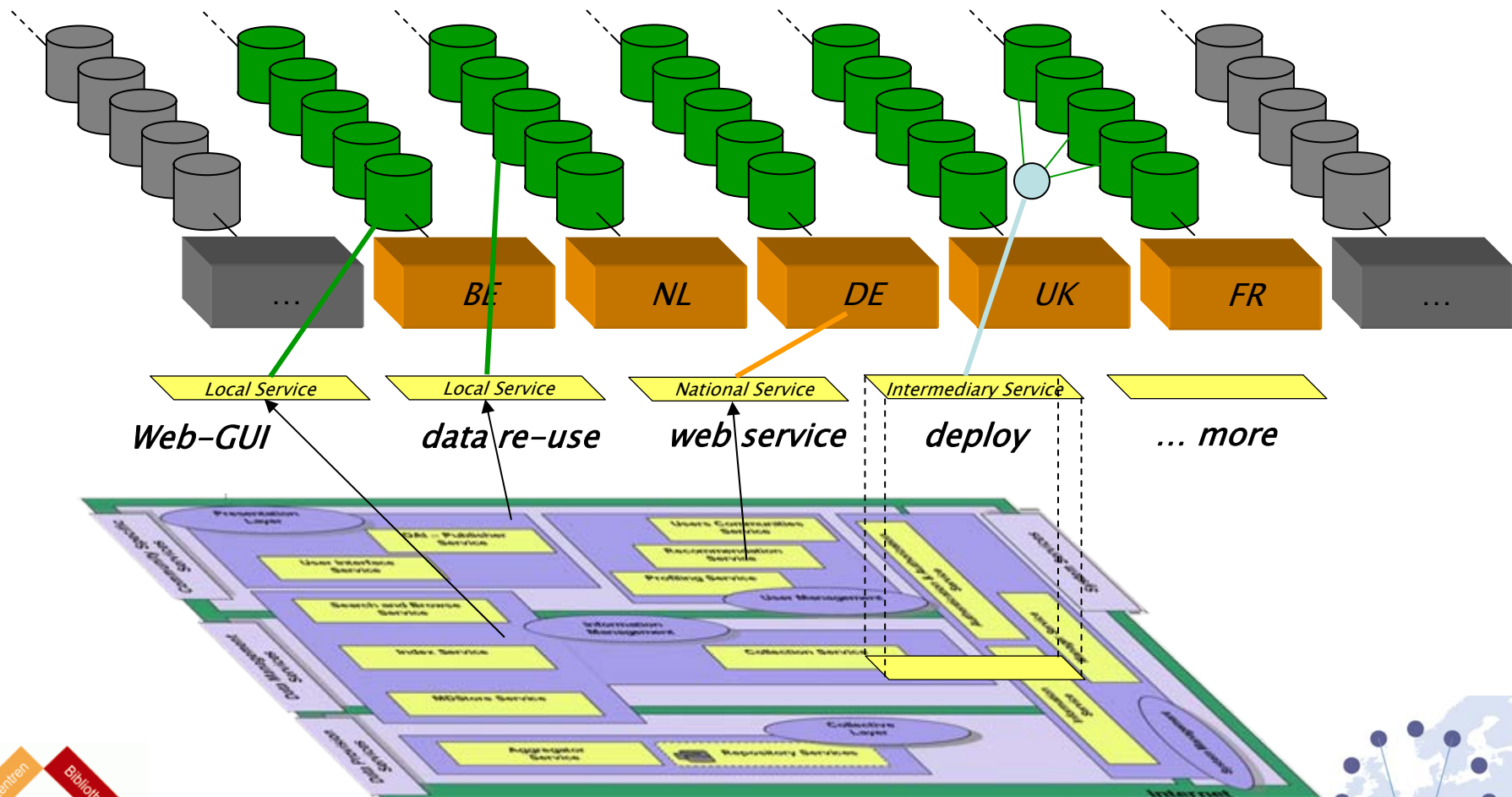
# Software Infrastructure Test-Bed



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



# Software Infrastructure Test-Bed



Frank Scholze & Wolfram Horstmann  
 Vascoda | Kiel | 30-OKT-2007



# *DRIVER and standards*

- Service Resources are implemented as **Web Services** and accessed through the corresponding Web Service Interface
  - Parameters calls are enveloped into **SOAP** messages
  - The Enabling Services are also compatible with **REST**
- XML is the lingua-franca for the whole system
  - Resource internal status, i.e. Resource profiles
  - Profiles in Information Service use **Exist XML engine**



# *DRIVER and standards*

- Vocabularies
  - Names of Languages: **ISO 639 – 2** (three letters, B/T)
  - Names of Countries: **ISO 3166** (two letters)
  - Date format: **ISO 8601: 1988 (E)**



# *DRIVER and standards*

- DRIVER Aggregation
  - Harvesting according to **OAI-PMH protocol**
  - Adopting **OAI-Provenance best practice** (OAI-about)
  - To be extended to other object models and harvesting protocols
- Queries to Search Service and Index Service obey to **SRW/CQL** standard



# *DRIVER and standards*

- Subscription and Notification Service
  - Any Service can subscribe to events regarding any DRIVER Resource: creation, deletion, and specific action accomplished by a resource
  - The Subscription and Notification mechanism is compliant with the **OASIS Standards WS Base Notification 1.3** and **WS Topics 1.3**
- Authorization and Authentication Service
  - Access Control Markup Language, **XACML**



# „Human“ portal for access



Search Results (7 documents)

Pages 1 of 1 << < 1 > >>

## THE STRUCTURE AND USE OF SHAPE-BASED NOUN CLASSES IN MIRAÑA (NORTH WEST AMAZON)

**CREATOR(S):** Seifart, Frank  
**DESCRIPTION:** Miraña, an endangered Witotoan language spoken in the Colombian Amazon region, has an inventory of over 60 noun class markers, most of which denote the shape of nominal referents. Class markers in thi ...  
**REPOSITORY:** DSpace at Radboud Univ. Nijmegen  
Repository Info | Repository's web site  
**LANGUAGE(S):** English  
**DOCUMENT:** View this document ...

## NOMINALIZATION AND ITS VARIOUS USES IN THULUNG.

**CREATOR(S):** Lahaussais, Aimee  
**DESCRIPTION:** This paper looks at the various types of nominalization in Thulung Rai, an endangered language of Eastern Nepal, and traces the use of the nominalizer for a variety of functions throughout the languag ...  
**REPOSITORY:** HAL - Hyper Article on Line  
Repository Info | Repository's web site  
**LANGUAGE(S):** English  
**DOCUMENT:** View this document ...

## NOMINALIZATION, RELATIVIZATION AND GENITIVIZATION IN THULUNG RAI

**CREATOR(S):** Lahaussais, Aimee

### Start new Search

MODIFY YOUR SEARCH REFINER YOUR SEARCH

All Fields all endangered languages +

#### More Preferences

- Date of Publication
- Document type
- Document Language
- Repository
- Community
- Collection

Reset Search

SELECTED FIELDS *hide details*

SEARCH HISTORY

- (endangered AND languages)

# Infrastructural Tools



**driver** Digital Repository Infrastructure Vision for European Research

[Home](#) [Repositories](#) [Network](#) [Collections](#)

Welcome

Login

Home > Welcome

## Welcome to DRIVER



Repository map



Network map

Copyright © 2007 Progetto Driver | All Rights Reserved

Original design by [Driver team](#)  



# Managing Aggregation



driver Digital Repository Infrastructure Vision for European Research

Home Services Repositories Network Data Structures Collections Schemas Aggregator

Welcome admin Logout

Aggregator > Aggregator

## DRIVER Aggregator Manager - Admin Control Panel

### List of Repositories

Nr.	Repository	Status	Last Harvesting Date	Harvestingtype	Harvesting schedule
1	<a href="#">University of Technics Hamburg, GERMANY, TUBdok</a>	active	2007-05-28T23:34:44Z	REFRESH	WEEKLY
2	<a href="#">Dissertations of the Universiteit van Amsterdam</a>	inactive	2007-06-01T10:36:14Z	REFRESH	WEEKLY
3	<a href="#">SciDok, der Wissenschafts-Server der Universitaet des Saarlandes</a>	active	2007-05-28T23:45:10Z	REFRESH	WEEKLY
4	<a href="#">Royal Holloway Research Online</a>	active	2007-05-28T23:37:38Z	REFRESH	WEEKLY
5	<a href="#">University Digital Archive of the University of Groningen, The Netherlands.</a>	active		REFRESH	WEEKLY
6	<a href="#">DSpace at Open Universiteit Nederland</a>	active	2007-05-29T13:13:52Z	REFRESH	WEEKLY
7	<a href="#">DSpace at Vrije Universiteit Amsterdam</a>	inactive			
8	<a href="#">OAI-Repository SUB Goettingen</a>	active	2007-05-28T23:49:49Z	REFRESH	WEEKLY
9	<a href="#">SOAS Eprints</a>	active	2007-05-28T23:51:31Z	REFRESH	WEEKLY



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



# Test of Compliance



driver Digital Repository Infrastructure Vision for European Research

Home Services Repositories Network Data Structures Collections Schemas Repository

Welcome admin Logout

Data structures List Aggregator > Repository

## DRIVER Aggregator Manager - OAI Admin Panel

### Repository Form

[List of Repositories](#)

Repository Information: [Identify](#) - [ListSets](#) - [ListMetadataFormats](#)

Record Information: [View ListRecords](#) [Test Mapping](#)

[Start Harvesting](#) (Current settings)

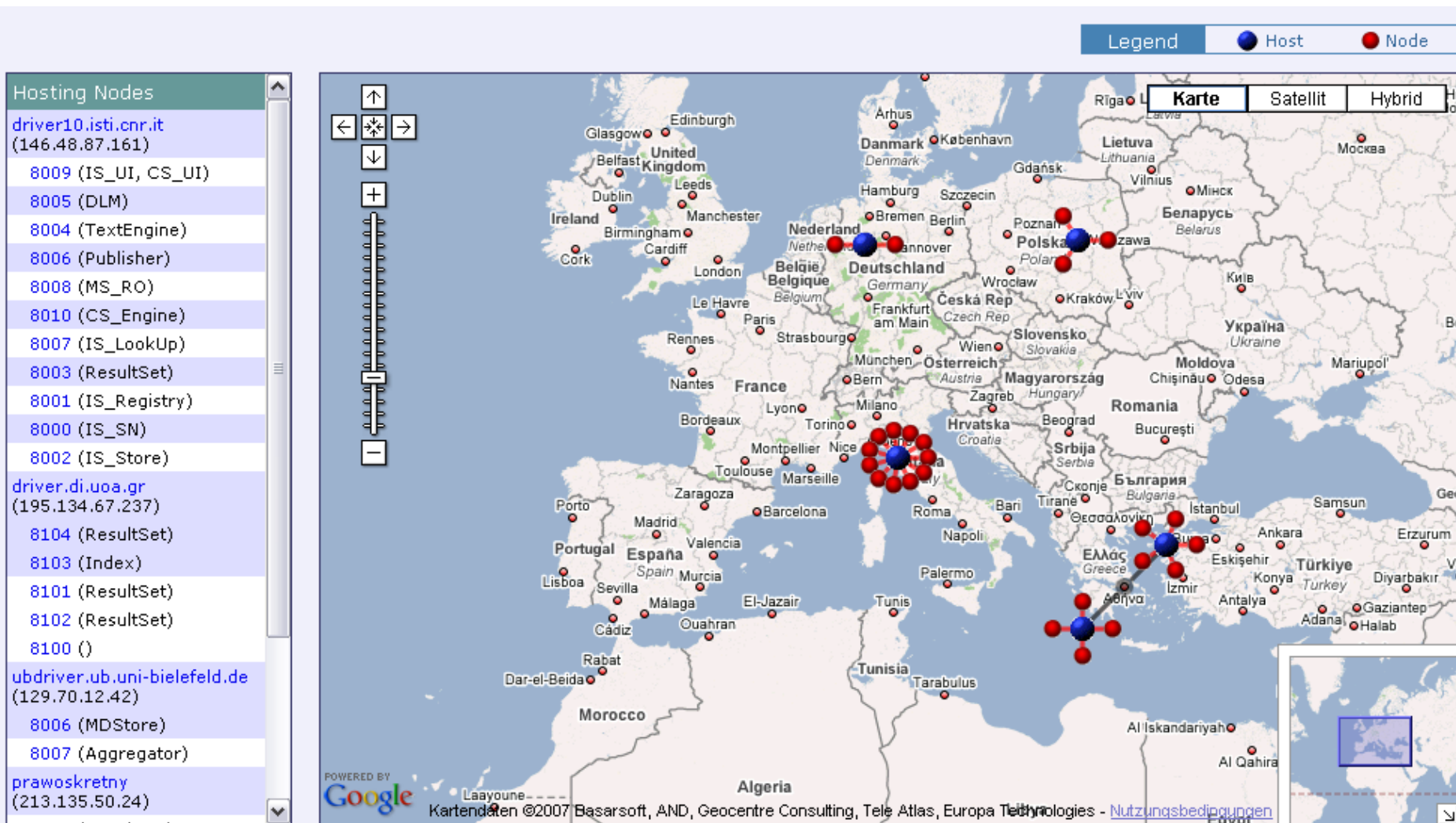
Repository:	SciDok, der Wissenschafts-Server der Universitaet des Saarlandes
Repository Identifier:	99-83409cf6-0d36-11dc-9ade-000347f19e46_UmVwb3NpdG9yeVNlcnZpY2VSZXNvdXJjZXMvUmVwb3NpdG9yeVNlcnZpY2VSZXNvdXJjZVR5cGU=
Harvesting instance	111-9de560c8-0d36-11dc-9ade-000347f19e46_SGFydmVzdGluZ0luc3RhbmNlRFNSZXNvdXJjZXMvSGFydmVzdGluZ0luc3RhbmNlRFNSZXNvdXJjZ
Status:	active



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



# Software Infrastructure



# *DRIVER-I Summary*

- Researchers need full-text access
- International cooperation needed
- Current repository system on the move
- But full-text access is hampered
- DRIVER tackles at European scale
  - Organizational framework
  - Data infrastructure
  - Software infrastructure



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007



# DRIVER-II – Overall Goals

IR Community

- Organization

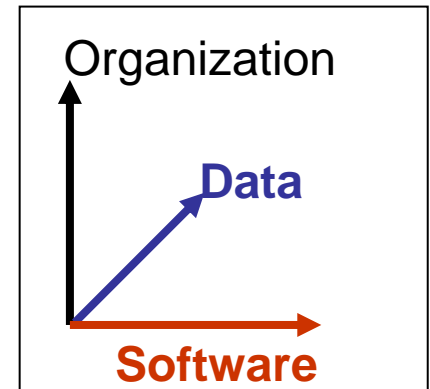
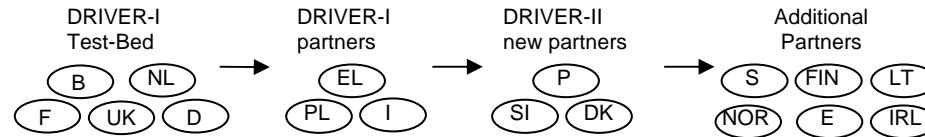
- Building a **Confederation** of European Digital Repositories

- Data

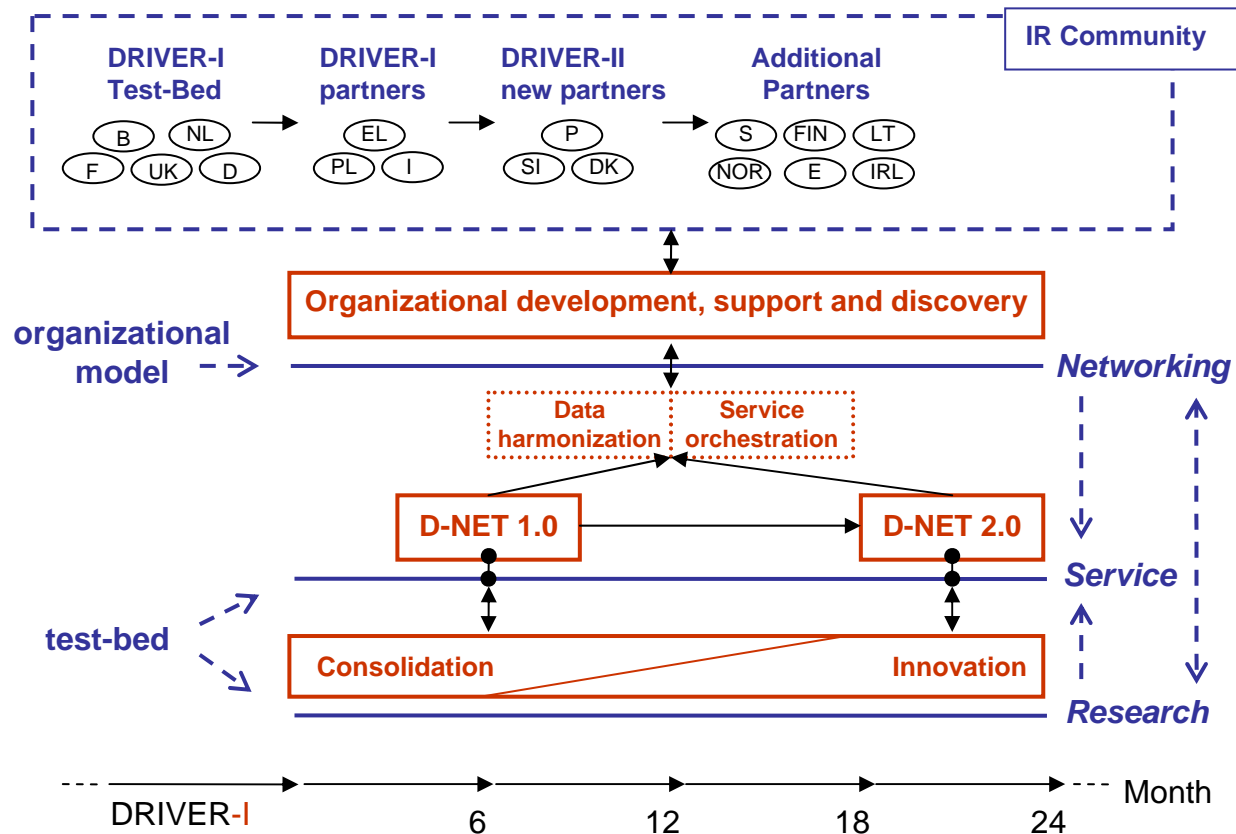
- Extending the virtual Information Space over multiple and **heterogeneous repositories**

- Software

- Consolidating and enhancing the Infrastructure to operate a **production quality** system with **advanced user functionalities**



# DRIVER-II – Process



# Overview

- General
- DRIVER (Europe)
  - Organisation / Data
  - Data / Software
  - DRIVER-II
- OA-Netzwerk (DE)
  - Rahmenprojekt
  - Zusatzdienste

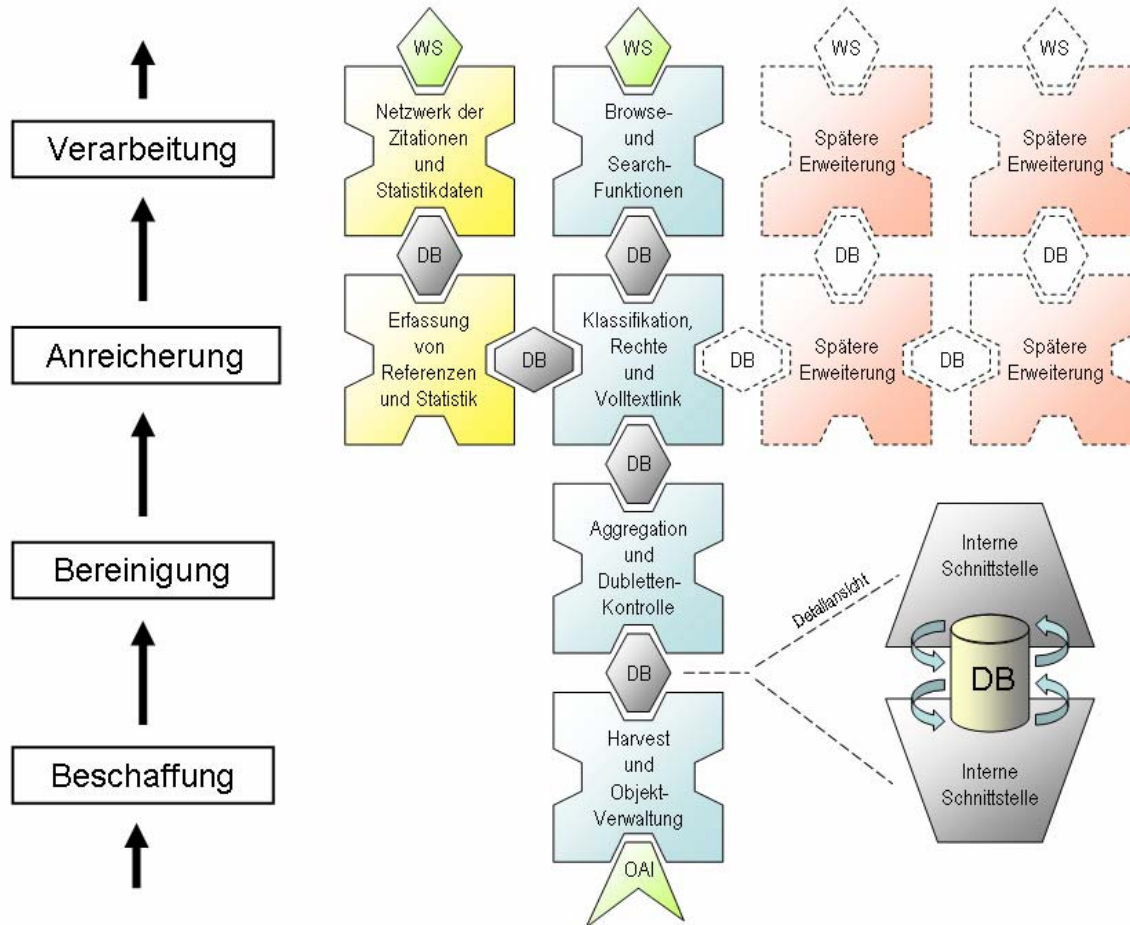


# OA-Netzwerk

- Netzwerk und Infrastruktur für deutsche Rep.
- z.Zt. in der Startphase
- Innovative Zusatzdienste
  - Nutzungsstatistik
  - Zitationsanalyse
- Beziehung zu DRIVER
  - Deutscher Knoten für DRIVER
  - Höhere Qualitätsstandards
  - Zusatzdienste gehen über DRIVER hinaus



# Architektur



# Zitationsanalyse



- Zitationsextraktion
  - Welche Literatur wird in einem Dokument zitiert
- Zitationsnetzwerk
  - Aufbauend auf Citebase
- Alternative Kennzahlen
  - Erweiterung des Dokumentraumes
  - Alternative Algorithmen (Citation Page Rank)



# Zitationsanalyse

[Reinhard PG, 1998, European Physical Journal, v 3, p 175](#)

[Field Amplification in Na Clusters](#)

von [PG. Reinhard](#) und [E. Suraud](#)

Auswahl eines Zitationsgraphen durch Klick auf den gewünschten Graphen.

Darstellung eines komplexeren Graphen durch Klick auf den Pfeil nach rechts. Zur Darstellung eines weniger komplexen Graphen klicken Sie auf den Pfeil nach links.

Im Graph: 1

[Reinhard PG, 1999, Comptes Rendus De L Academie Des Sciences Serie Ii Fascicule B-Mecanique Physique Astronomie, v 327, p 893](#)

[Excitation of Metal Clusters by Intense Lasers](#)

Im Graph: 2

[Brack M, 1999, European Physical Journal, v 9, p 111](#)

[Frequencies, Times, and Forces in the Dynamics of Na Clusters](#)

Im Graph: 3

[Giglio E, 2001, Journal Of Physics, v 34, p 0](#)

[Impact of Two-Body Collisions on Explosion Dynamics of Irradiated Clusters](#)

Im Graph: 4

[Reinhard PG, 1999, Journal Of Cluster Science, v 10, p 239](#)

[On Electron Dynamics in Violent Cluster Excitations](#)

Im Graph: 5

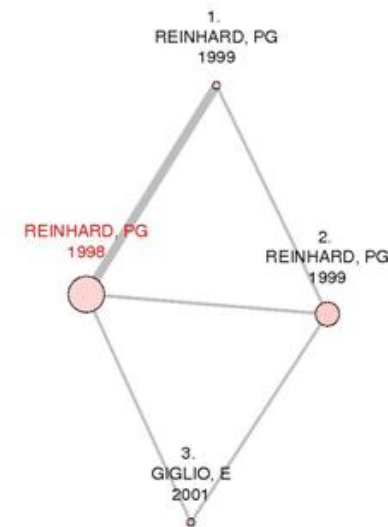
[Giglio E, 2000, Journal Of Physics, v 33, p 0](#)

[Semiclassical Approach to Metal Cluster Dynamics](#)

Im Graph: 6

[Giglio E, 2003, Physical Review A, v 67, p 0](#)

[Angular Distribution of Emitted Electrons in Sodium Clusters: a Semiclassical Approach](#)



Knotengröße = Zitierungszahl

# Nutzungsstatistik



- Lokale und aggregierte Nutzungsdaten
- Transparente Zugriffszahlen
  - z.B. COUNTER, IFABC, LogEc
- Damit nachvollziehbar, wie die Daten berechnet wurden
  - Klickspannen, Eliminierung automatisierte Zugriffe etc.



# Kennzahlen

- Kennzahlen basierend auf der Häufigkeit oder auf strukturellen Zusammenhängen von Nutzungszahlen
- Ein Beispiel für eine quantitative Kennzahl ist der **Usage Factor** (Durchschnittswert normalisierter und aggregierter Nutzungszahlen über einen definierten Zeitraum)
- Ein Beispiel für eine strukturelle Kennzahl ist der **Usage Page Rank** (reziprokes Abstimmungsverhalten)

*Danke!*



Frank Scholze & Wolfram Horstmann  
Vascoda | Kiel | 30-OKT-2007

