



# Edutella - Digital Archive Integration with Peer-to-Peer and Semantic Web Technology

Wolf Siberski





## Edutella: General Introduction

- Open source project (<http://edutella.jxta.org>)
- A multi-staged effort to scope, specify, architect, and implement an RDF-based meta-data P2P infrastructure
- Main Goal: Achieve interoperability between heterogeneous metadata-driven (e-learning) systems
- Principal contributors
  - *Germany*: University of Hanover, University of Karlsruhe
  - *Sweden*: KTH Stockholm, University of Uppsala
  - *USA*: Stanford University
- Foundations
  - Semantic Web Technology
  - Peer-to-Peer Technology





## Overview

- Semantic Web
- Peer-to-Peer (P2P)
- Edutella Query Service
- OAI-P2P: a P2P network for Open Archives



## Semantic Web

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." --

*Tim Berners-Lee, James Hendler, Ora Lassila:*

*The Semantic Web*

- Metadata-based
- Uniform information format (RDF)
- Ontologies
  
- Vision: allow to infer new information from available information + rules



## RDF

- Resource Description Framework
- allows annotation of resources with metadata
- „A language for transmitting pieces of collaborative databases“
- „The bytes within the Semantic Web processor (but where you don't know what the processor will be)“  
*(from RDF newsgroup)*
- Descriptions consist of simple sentences
  - Form: <subject> <predicate> <object>
  - Sample: „book1234 hasTitle ‚Introduction to Digital Libraries‘ “



## RDF Sample

COURSE LIST

TREE VIEW

TRAIL

RELATIONS

- Introduction
- Intelligent Agents
  - Definitions
  - Structure of Intelligent Agents
- Agent Types
  - Reflex agent
  - Agents that keep track of the world
  - Goal based agent**
  - Utility based agent
  - Anwendungsbeispiele für Softwareagenten
  - Ameisenbeispiel
  - Der Lego-Roboter als Intelligenter Agent
  - MIT Agent Projects
  - Robocup Homepage
- Agent Environment
  - Europe's Network of Excellence for Agent-Based Computing

### Goal based agent

- autonomous
- using the goal and the result of its possible actions it is able to choose the action that will make it achieve the goal

**Agent**

Sensors

State

How the world evolves

What my actions do

Goals

What the world is like now

What it will be like if I do action A

What action I should do now

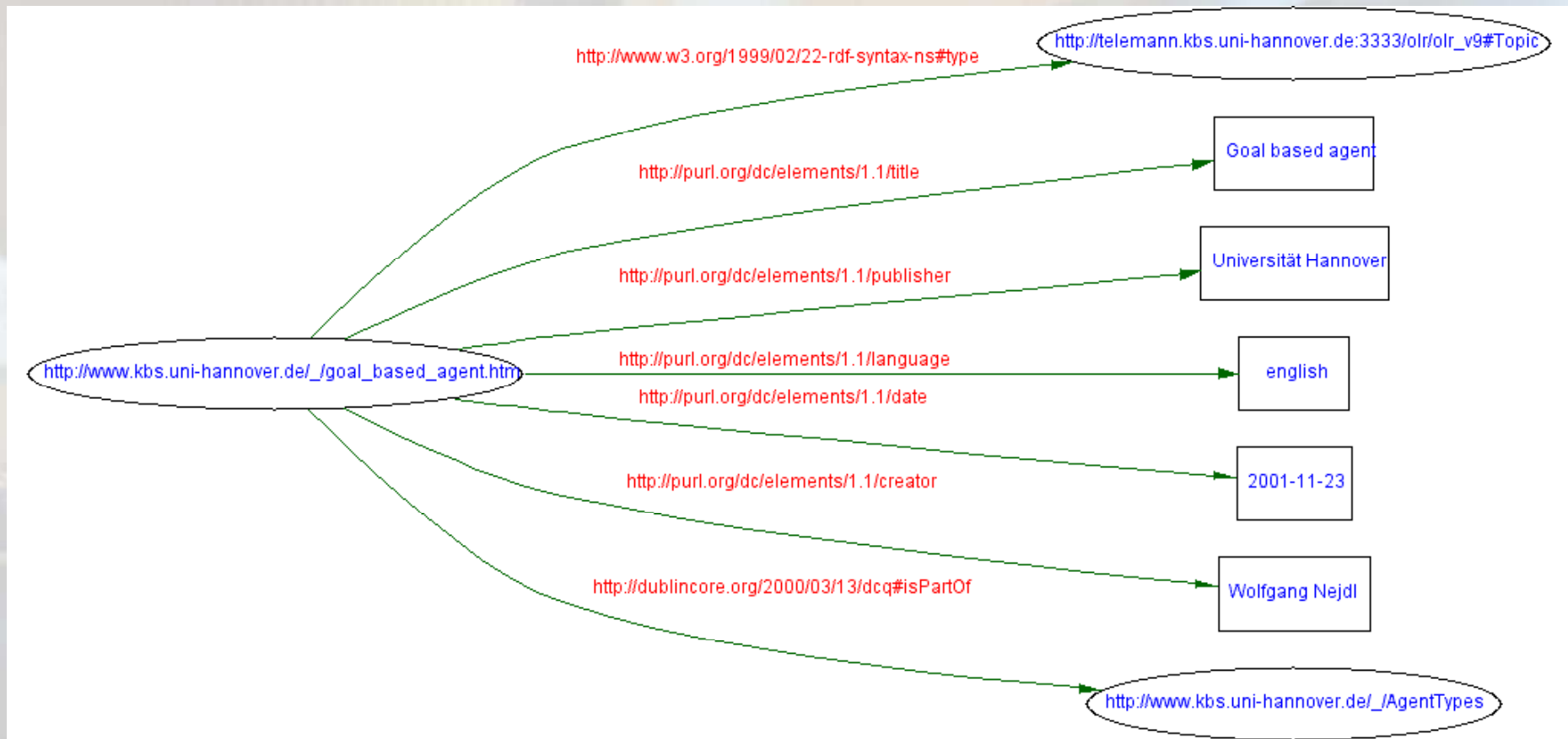
Effectors

**Environment**

Agent with explicit goals



## RDF Sample - Graph





## RDF vs. XML

- XML: ‚Closed‘ description (document contains complete resource description, no way to add information from outside)
- RDF: ‚Open‘ description (Information about one resource may be found in many places)
- XML: Description with predefined metadata structure
- RDF: Descriptions can be combined and extended
- XML: sender and receiver have agreed about semantics
- RDF: programs can reason about resources and their descriptions, semantics need not be hard-coded (vision)



## Peer-To-Peer-Networks

- Advantages
  - No central administration
  - New information providers can join immediately
  - information control and responsibility remains local
  
- JXTA
  - P2P Framework by Sun (<http://www.jxta.org>)
  - Open Source
  - Provides basic P2P network services
    - Rendezvous service: integration of new peers
    - Discovery service: discovery of peers with custom services
    - Pipe service: message transport between peers
    - ...



## Edutella - Basic Assumptions

- Provide content descriptions (metadata) only
- No own repository, but adapters for several repositories
- Metadata is provided in RDF format
- Peers can take several roles:
  - Provider: provides information
  - Consumer: requests information
  - Mediator: ,translates' requests
  - ...
- Query Exchange Language (QEL) as ,lingua franca'



## Edutella Query Service

- Edutella specifies and implements several services
- Query service provides standardized query and retrieval of RDF metadata stored in distributed RDF repositories

- Query Exchange Language

- Based on Datalog (allows expression of rules)
- For exchange only
- Example:

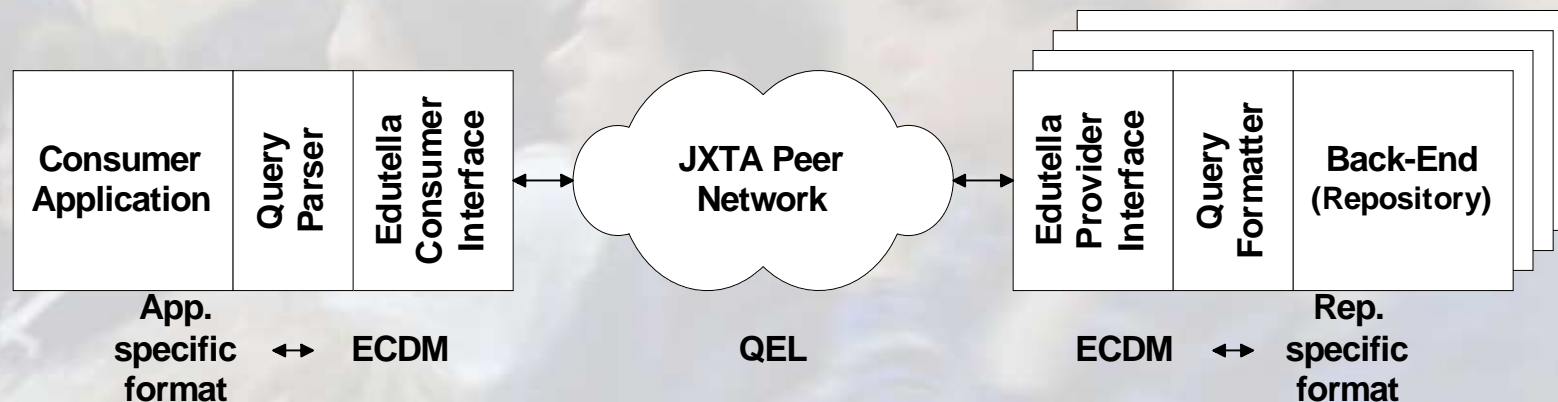
```
dlbook(X) :- subject(X, 'Digital Library'), type(X, 'Book').  
dlbook(X) :- subject(X, 'Digital Archive'), type(X, 'Book').  
?- dlbook(X).
```

- Adapters to translate QEL queries for several Backends:
  - File, RDBMS, Rule Database, ...



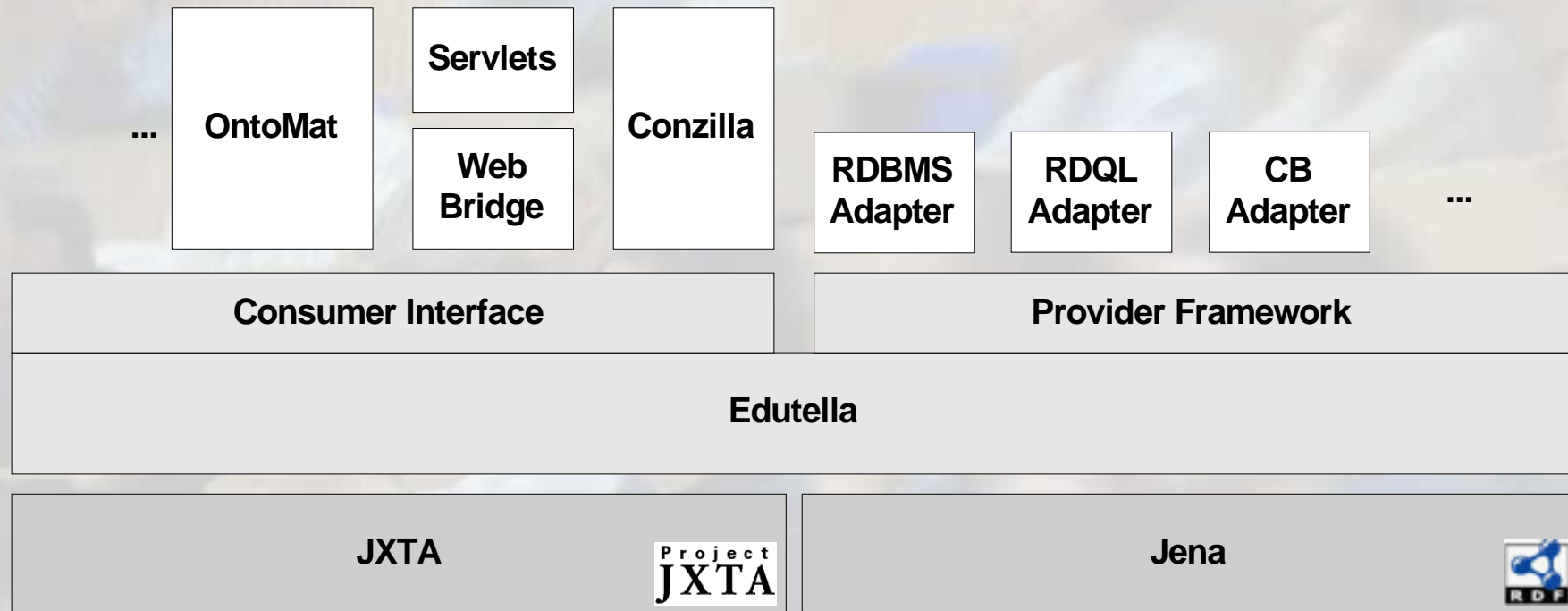
## Query processing

- Parsers/Formatters convert between query languages
- Applications and backends are shielded from communication layer
- Allows usage of different query languages in applications and backends





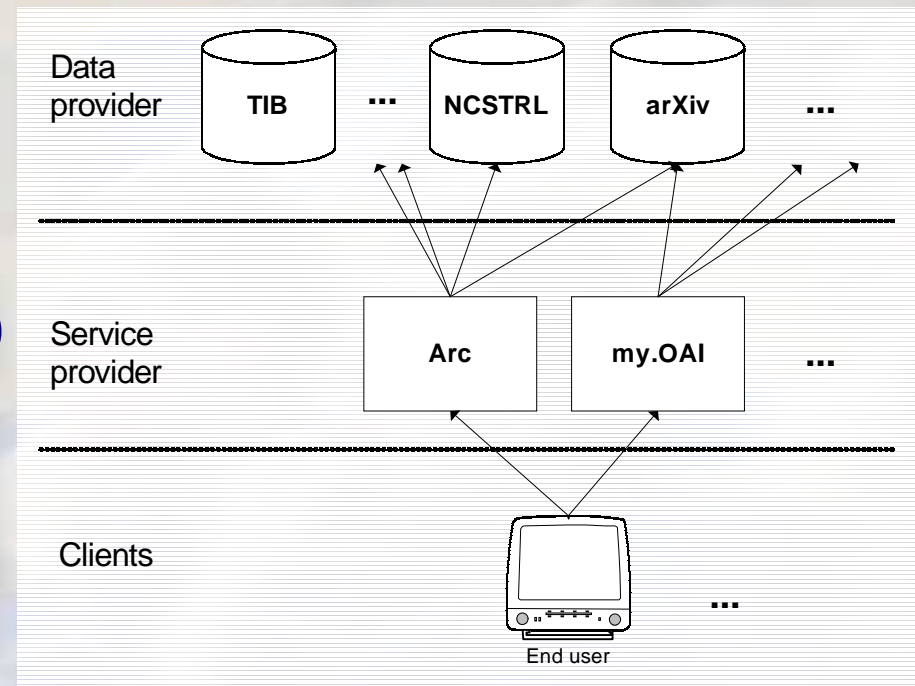
## Edutella Architecture





## Open Archive Initiative

- Set of standards for access to digital archives
- Uses Dublin Core metadata standard
- Repository (Data Provider)
  - provides metadata about documents
  - access via HTTP-based protocol (Harvesting Protocol)
- Harvester (Service Provider)
  - collects metadata from repositories
  - provides end-user interface for queries





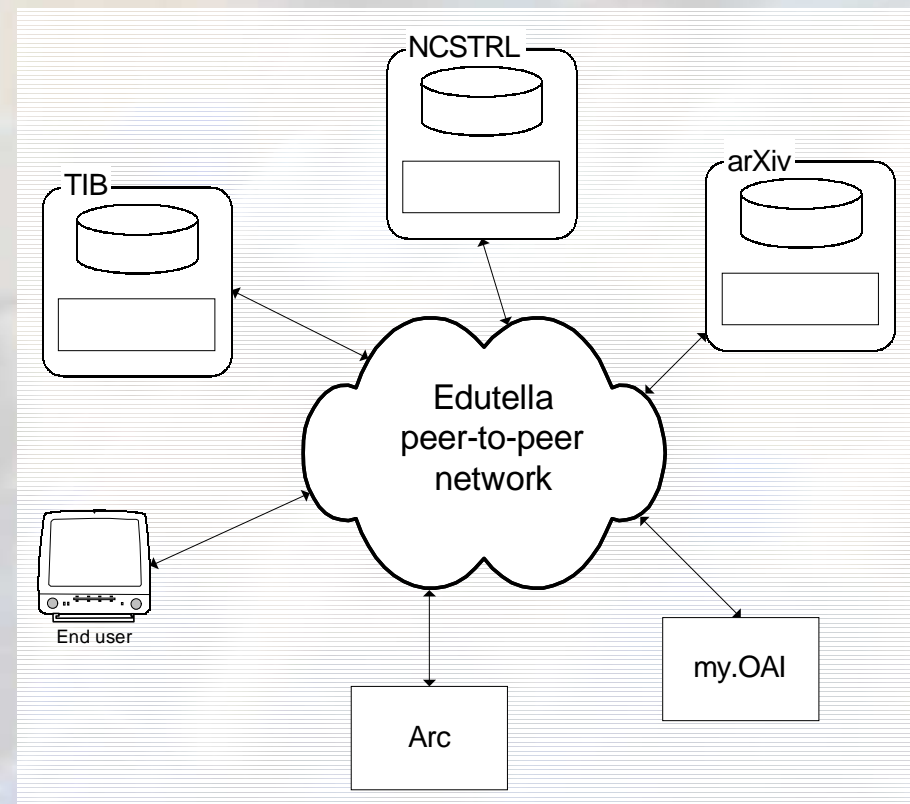
## OAI Characteristics

- low-adoption-barrier approach (harvesting protocol simple to implement)
- clear architecture (separation of data management and data access)
- data provider need service provider(s) to make its data visible
- adding new repositories to harvester needs administration (Exception: Kepler project)
- no service provider available which allows to search through *all* OAI repositories



## OAI-P2P - Approach

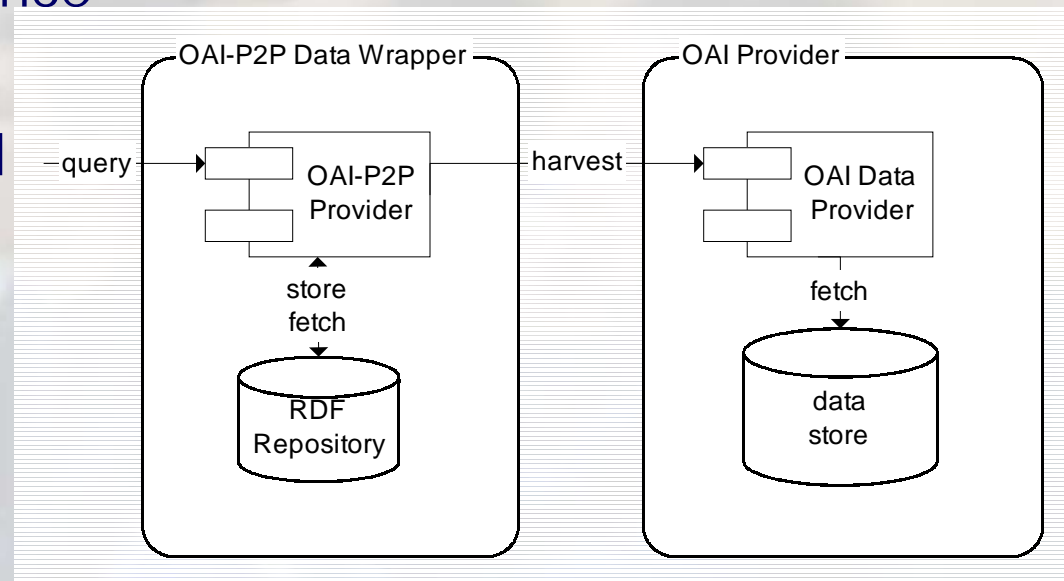
- OAI data provider become Edutella provider
- End-User frontends connect directly to Edutella network (no service provider as mediator)





## OAI-P2P - Data Wrapper Implementation

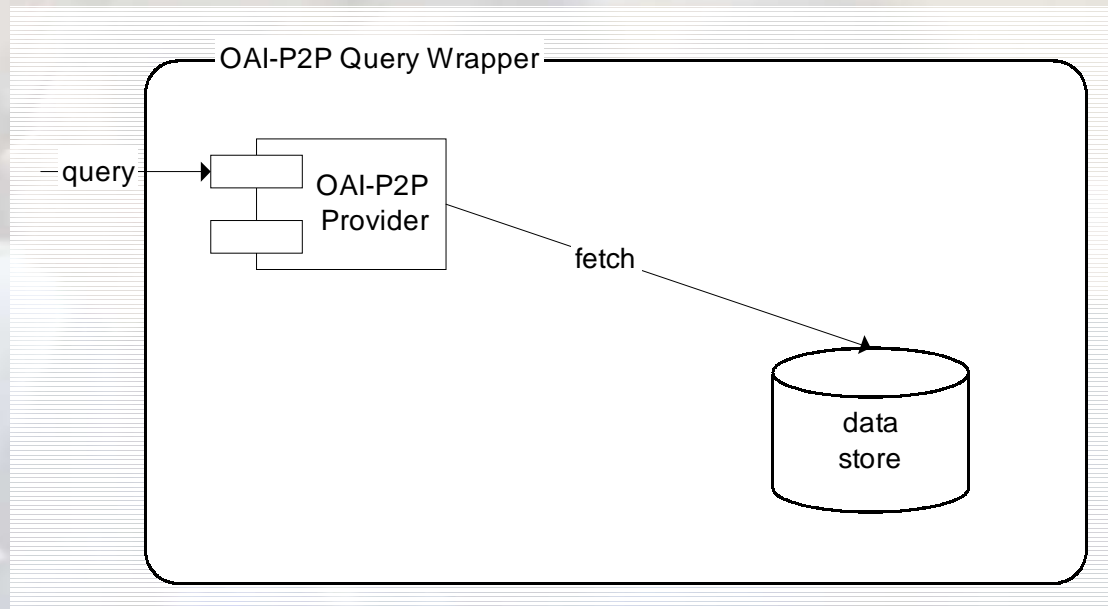
- Wrapper uses harvesting protocol to get content
- acts as OAI harvester and Edutella provider
- no modification of OAI repository necessary
- can be implemented once
- metadata is replicated
- not always up-to-date





## OAI-P2P - Query wrapper implementation

- Wrapper executes Edutella query directly at the data store
- data store specific implementation necessary
- no replication
- always up-to-date





## Advantages and Issues

### ■ Advantages

- New data providers become visible immediately
- Query reaches all providers
  - but closed groups of providers possible
- Little effort to integrate existing OAI repositories into Edutella

### ■ Issues

- Content quality uncontrolled
  - could be done by certifying providers
- Scalability
  - currently tackled



Current Status

Demo





## Future Work

- Improved provider implementation
  - Goal: Out-of-the-box solution for data wrapper
- ‚Real world‘ client interface
  - also web-based
- Efficient Query Distribution
  - Semantic Query Routing