

# EVENTOS | 2007

VII WORKSHOP REBIUN PROYECTOS  
DIGITALES. LA ESTRATEGIA DIGITAL



eSciDoc: UNA PLATAFORMA DE NUEVA  
GENERACIÓN PARA LA INFORMACIÓN  
Y LA COMUNICACIÓN CIENTÍFICA



crue

Universidades  
Españolas

Red de Bibliotecas  
REBIUN

# eSciDoc: una plataforma de nueva generación para la información y la comunicación científica

Matthias Razum  
FIZ Karlsruhe

VII Workshop REBIUN sobre proyectos digitales  
Madrid, October 18<sup>th</sup>, 2007



# Agenda

- Motivation
- Fedora
- eSciDoc Infrastructure
- eSciDoc Solutions

# Agenda

- **Motivation**
- Fedora
- eSciDoc Infrastructure
- eSciDoc Solutions

## Project Setup and Mission

- eSciDoc is a joint project of the Max-Planck-Society (MPS) and FIZ Karlsruhe
- 6 million € five-year grant (2004 – 2009) from the German Federal Ministry of Education and Research
- It aims to build an integrated information, communication and publishing platform for web-based scientific work, exemplarily demonstrated for multi-disciplinary applications in the MPS
- eSciDoc is not a mere research project, but aims at establishing an innovative productive system



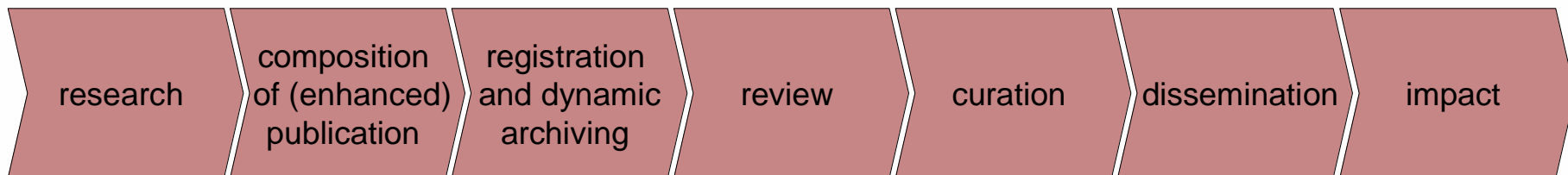
# Project Setup and Mission

- eSciDoc is a joint project of the Max-Planck-Society (MPS) and FIZ Karlsruhe
- 6 million € five-year grant (2004 – 2009) from the German Federal Ministry of Education and Research
- It aims to build an integrated **information, communication and publishing** platform for web-based scientific work, exemplarily demonstrated for **multi-disciplinary** applications in the MPS
- eSciDoc is not a mere research project, but aims at establishing an innovative productive system

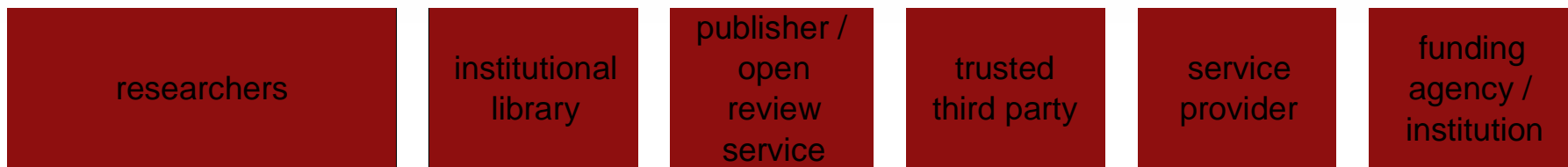


# Infrastructure for Scholarly Communication

## Tasks



## Actors



## Infrastructure



Source: SURFoundation

# What is e-Science?

*“e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it.”*

John Taylor

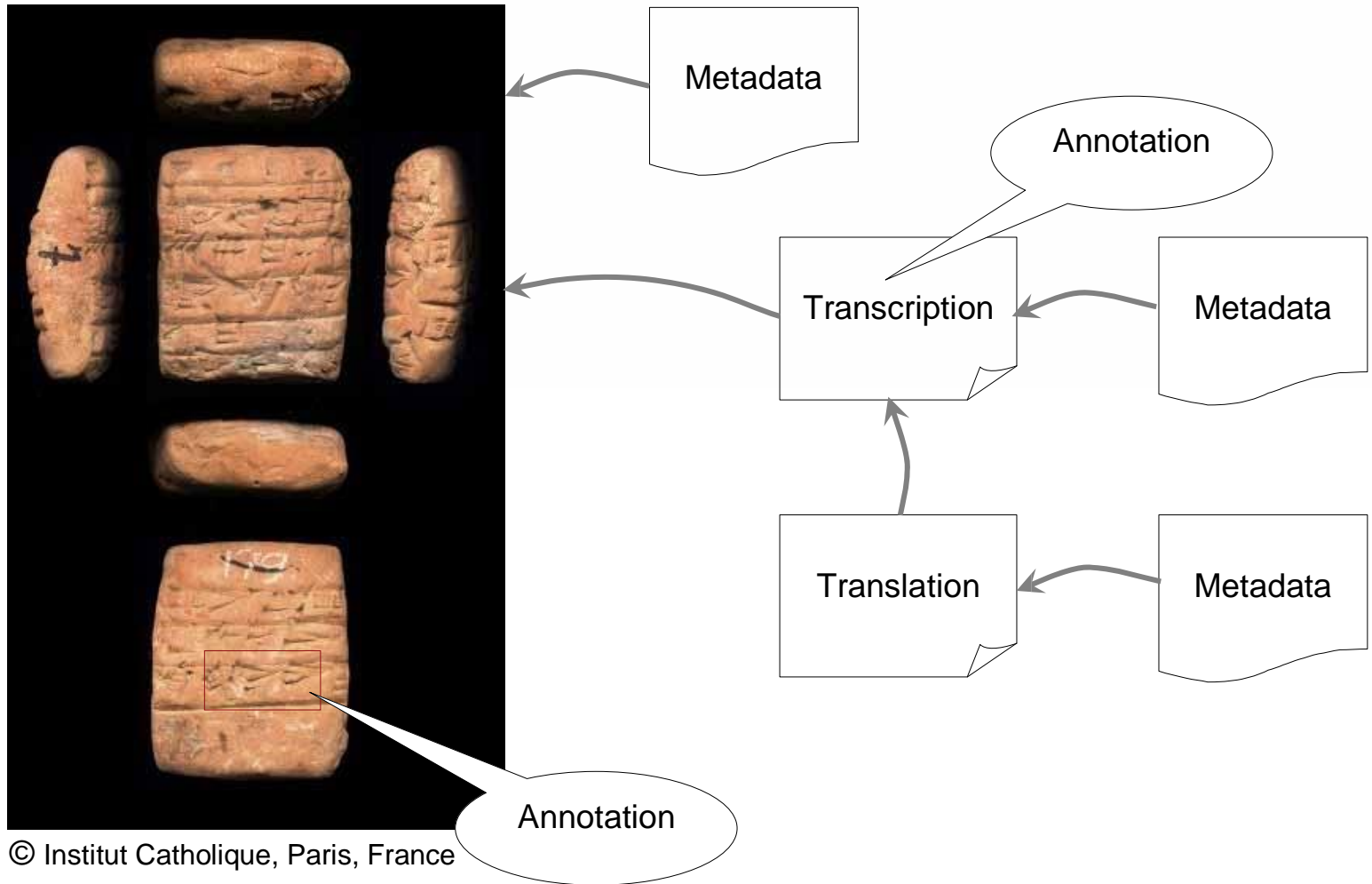
(former Director General of Research Council, UK)

## Turning Static Objects into ‘Living’ Knowledge

- e-Science allows to publish all intermediate results of knowledge generation from first ideas, theories, discussions with peers to final results
- Direct access to primary data and intermediate work results is essential
- Thinking a step further leads to interactive authoring environments with support for collaboration and annotations

- The contents of such repositories form the “institutional memory” of an work group, organization, or company
- Just like human memory, they should allow for associating information objects in novel contexts, thus creating new ways of research

## Example of a Knowledge Network







Repository



Repository



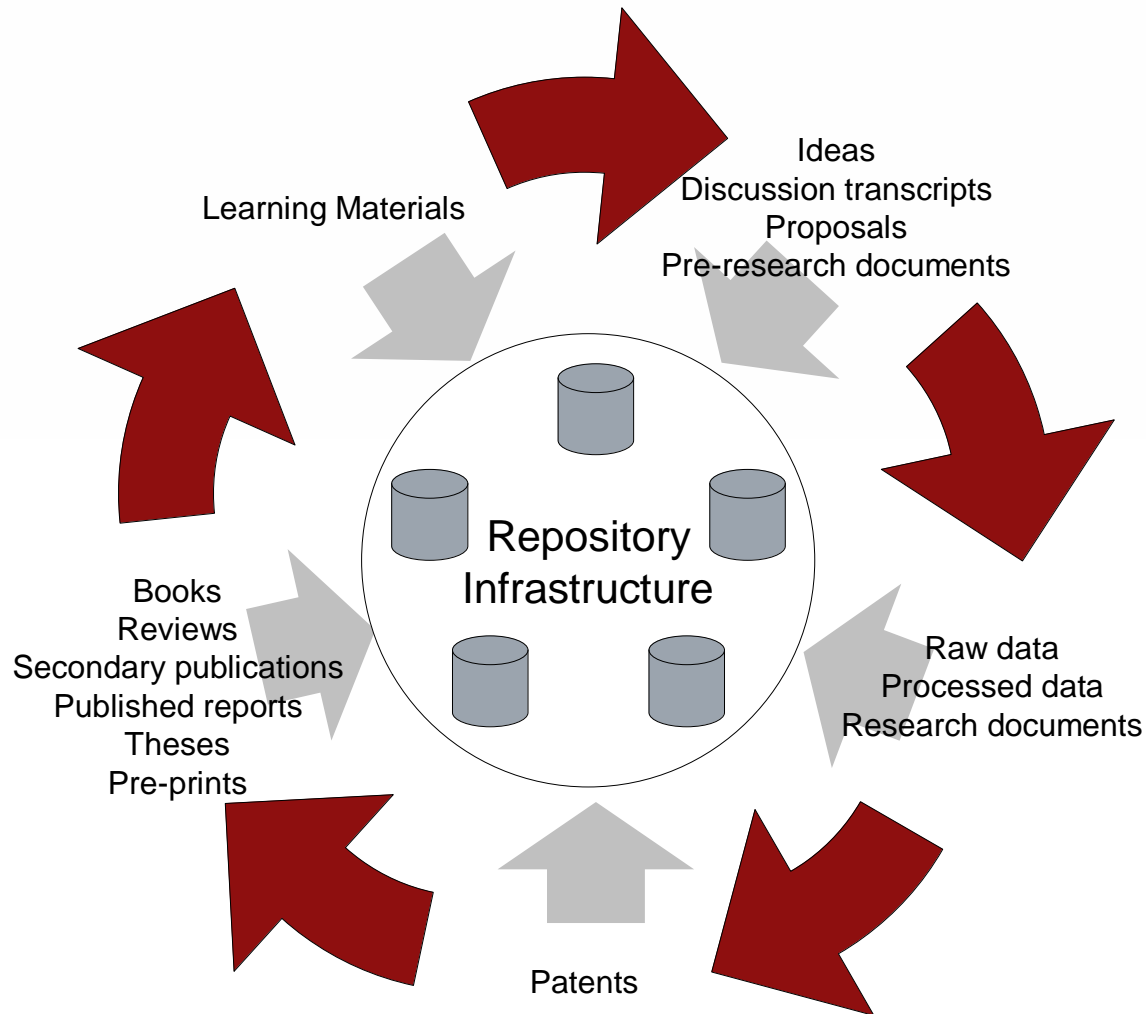
Repository

  
Repository

  
Repository

D

# Scholarly Communication Cycle



Source: e-SciDR Study/ A. Macdonald, P. Lord

# Information Management is a Key Issue

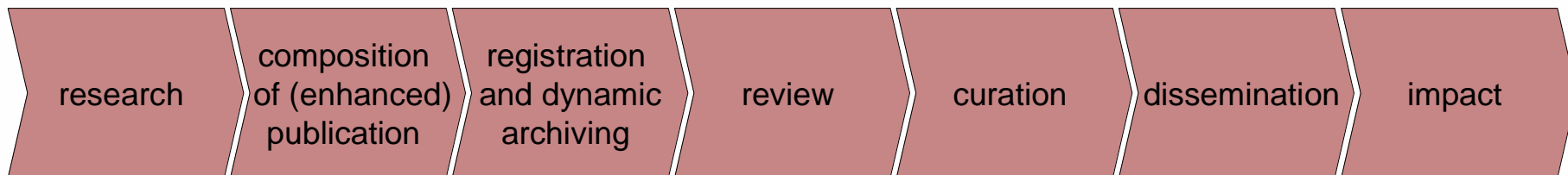
E-Science imposes new requirements onto existing information management processes and solutions:

- Ever-growing amounts of information
- Instant access without barriers
- Integration of heterogeneous information sources within and outside of an organization
- Manifold of object types and metadata formats
- Need to integrate information sources with existing and new applications

Librarians have to expand their traditional role and become scientific information managers!

## Infrastructure for Scholarly Communication

## Tasks



## Actors



## Infrastructure



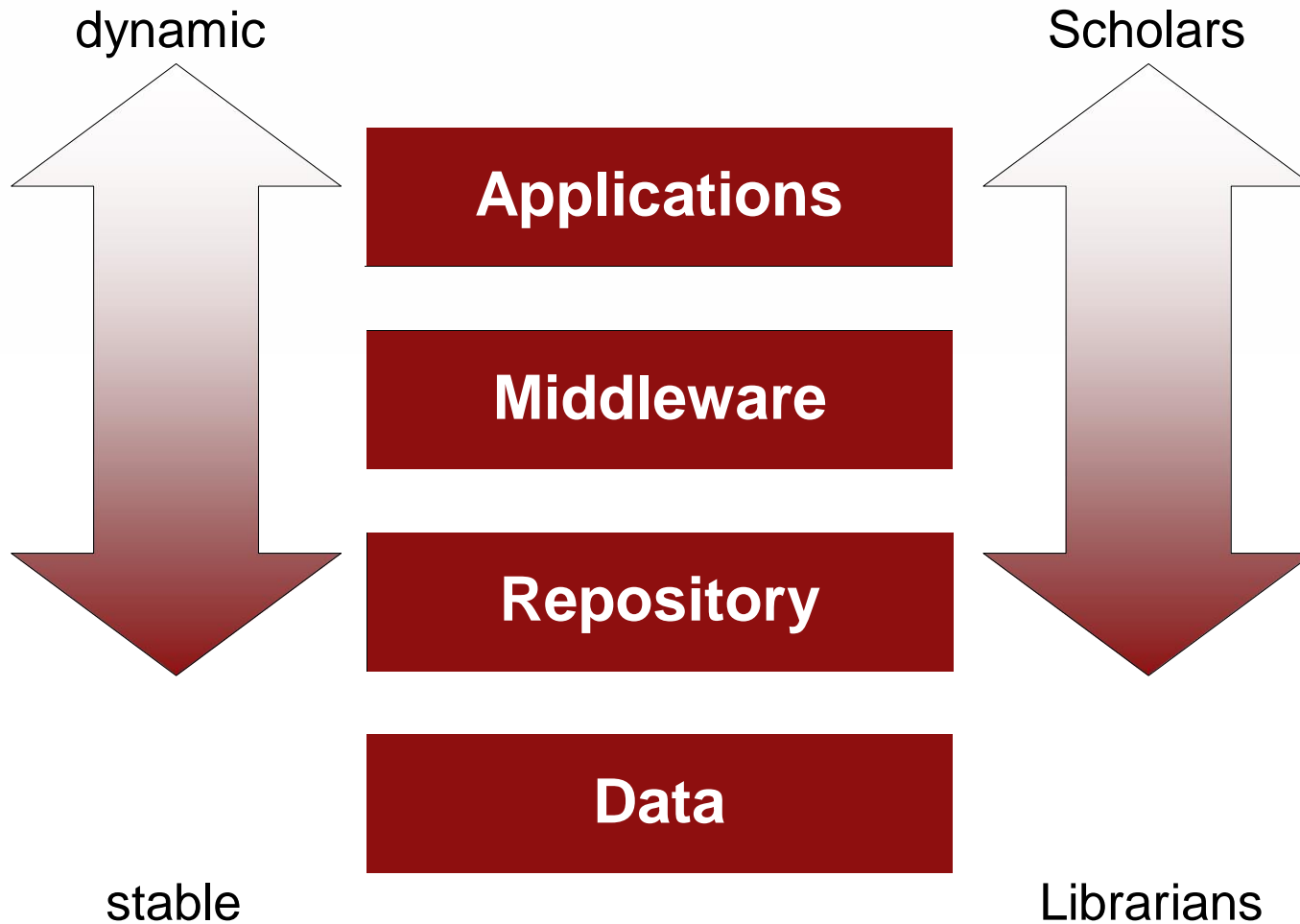
Source: SURFfoundation

# Agenda

- Motivation

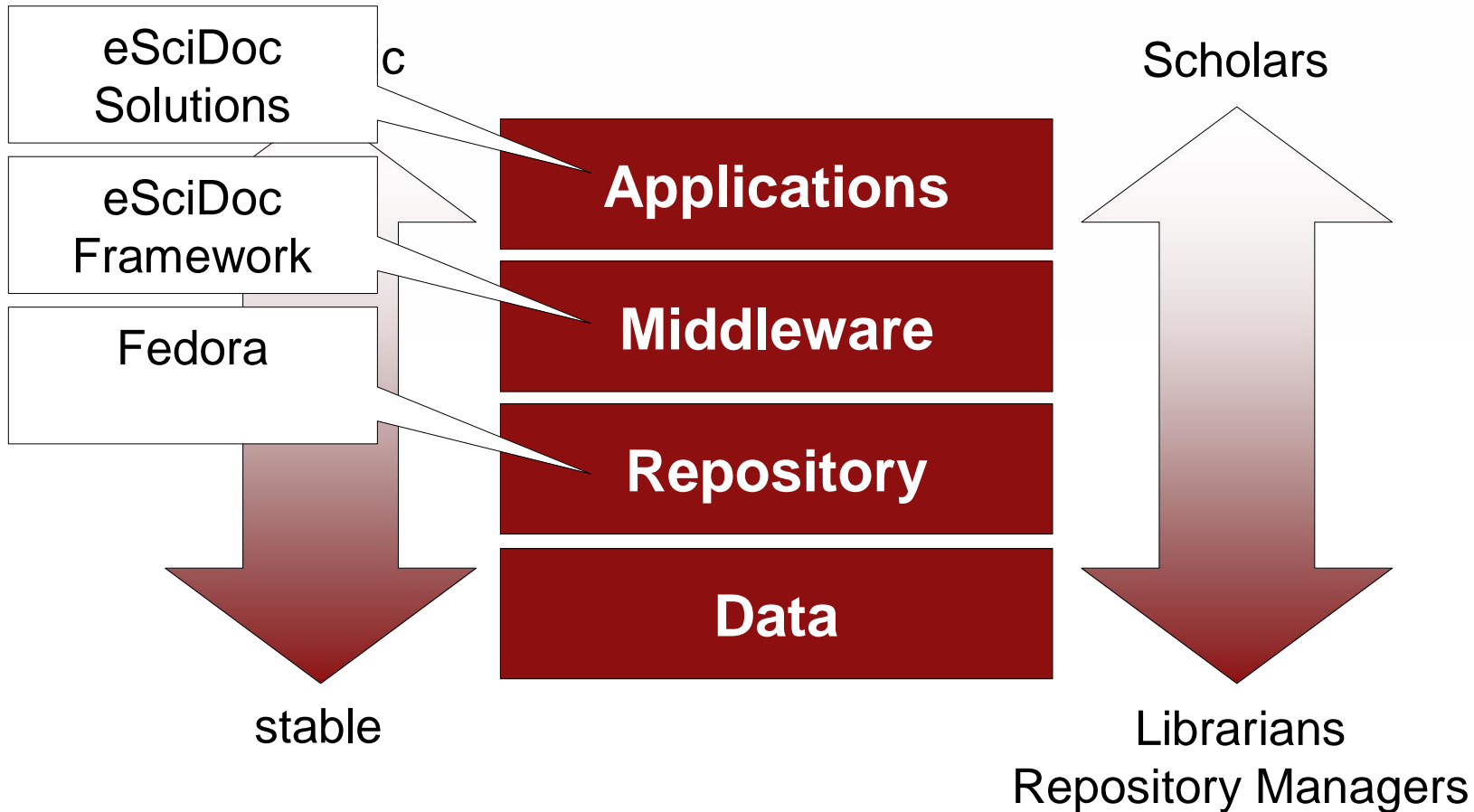
- **Fedora**
- eSciDoc Infrastructure
- eSciDoc Solutions

# Hierarchy of Persistence

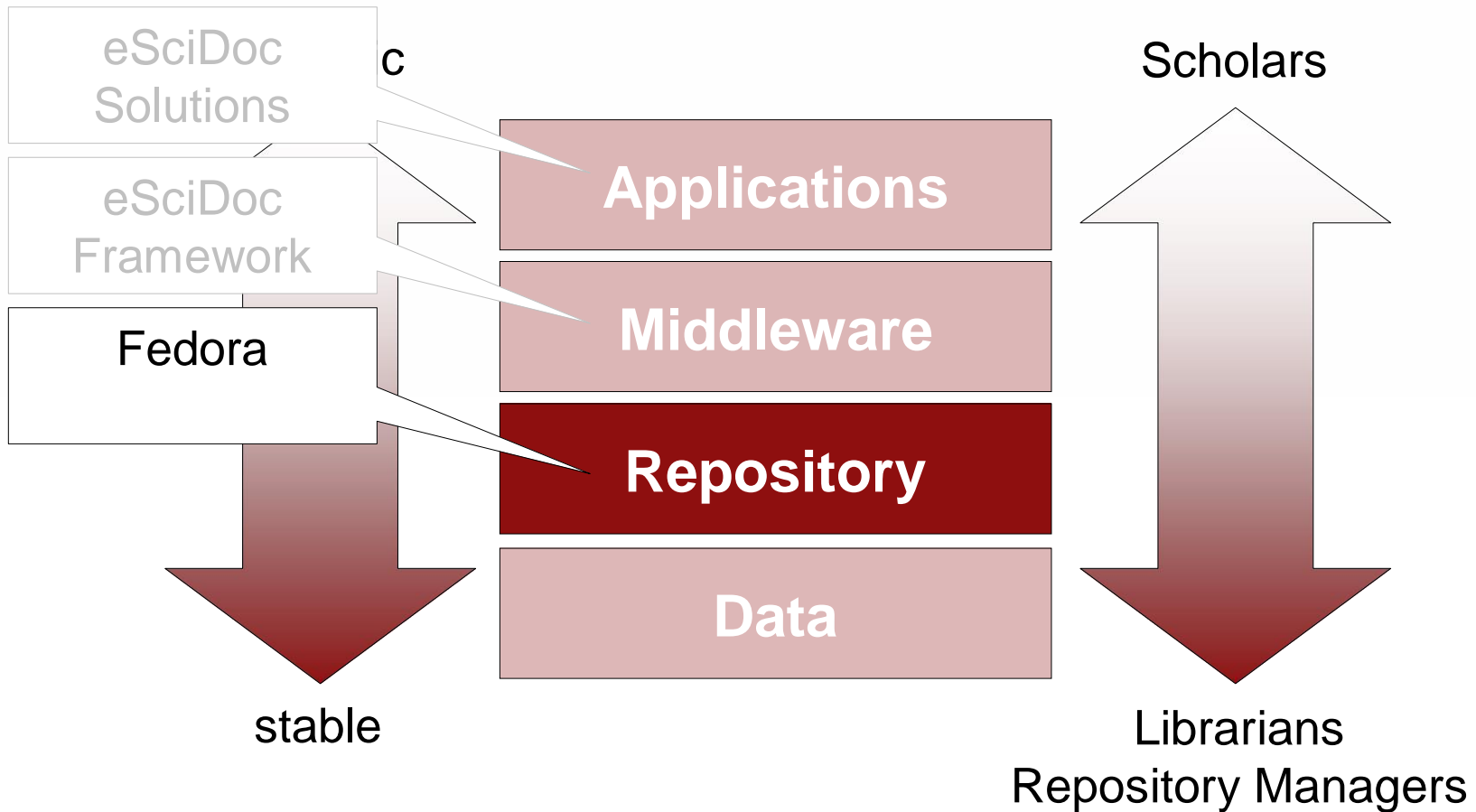


Repository Managers

## Hierarchy of Persistence



# Hierarchy of Persistence



## Fedora Stands for

- Flexible
- Extensible
- Digital
- Object
- Repository
- Architecture

## What is Fedora?

- Fedora Commons provides sustainable technologies to
  - create,
  - manage,
  - publish,
  - share and



- preserve digital content as a basis for intellectual, organizational, scientific and cultural heritage.

## The Fedora Triple Play

- Storage
  - Repository technologies to ensure longevity and integrity for any kind of digital content.
- Semantics
  - Semantic technologies to contextualize and inter-relate digital content from many sources.
- Services
  - Collaborative technologies to enable the creation of



innovative, collaborative information spaces.

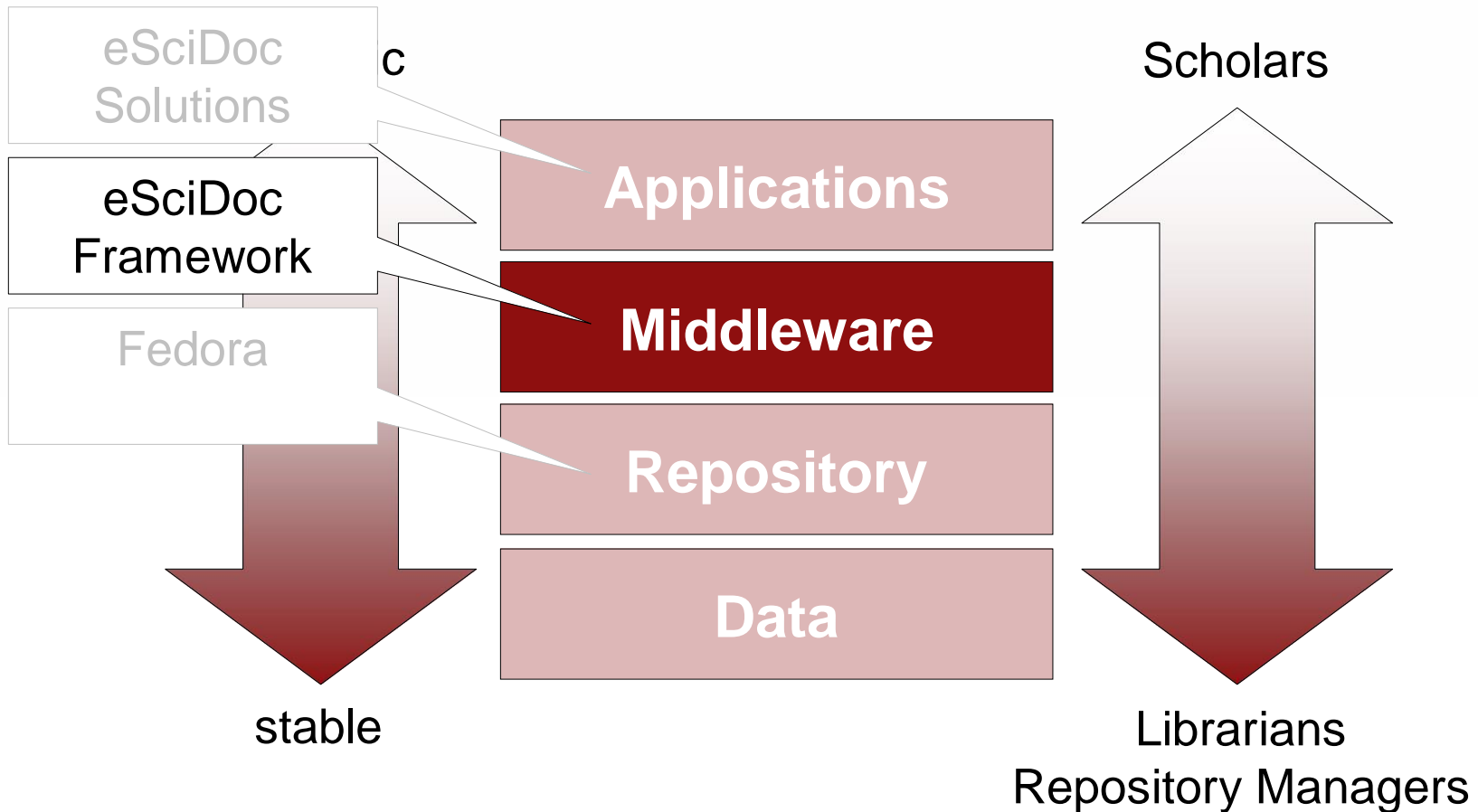
## Fedora's Key Features

- Store whatever you want
- Easy access
- Express relationships
- Free open source software
- Enables permanence
- A supportive community
- Extensibility
- Maturity
- Scalability, both small and large
- An engine for innovation

# Agenda

- Motivation
- Fedora
- **eSciDoc Infrastructure**
- eSciDoc Solutions

# Hierarchy of Persistence



## eSciDoc is Flexible

- Service-oriented architecture
- Object model and content models
- Arbitrary metadata profiles
- Application-independent design
- Many solutions
- Support for object relations and multiple ontologies
- Search (OpenSearch, SRW/SRU)
- Distributed Authentication/Authorization (Shibboleth)
- Integrates into a broader e-Science infrastructure

## eSciDoc is Re-usable

- The eSciDoc infrastructure generalizes functions

- Compliance with important standards fosters interoperability
- The multi-disciplinary nature of the MPS ensures the coverage of a broad range of generic and discipline-specific requirements
- Solutions like Publication Management are highly configurable
- The service-oriented architecture fosters the creation of autonomous services, which can be re-used independently from the rest of the infrastructure

## eSciDoc is an Enabling Technology

- eSciDoc supports knowledge management in the scientific domain
- eSciDoc can present heterogeneous and distributed data in a integrated and adaptive workspace
- eSciDoc is not a silo, but an open system
- eSciDoc cares about long-term preservation

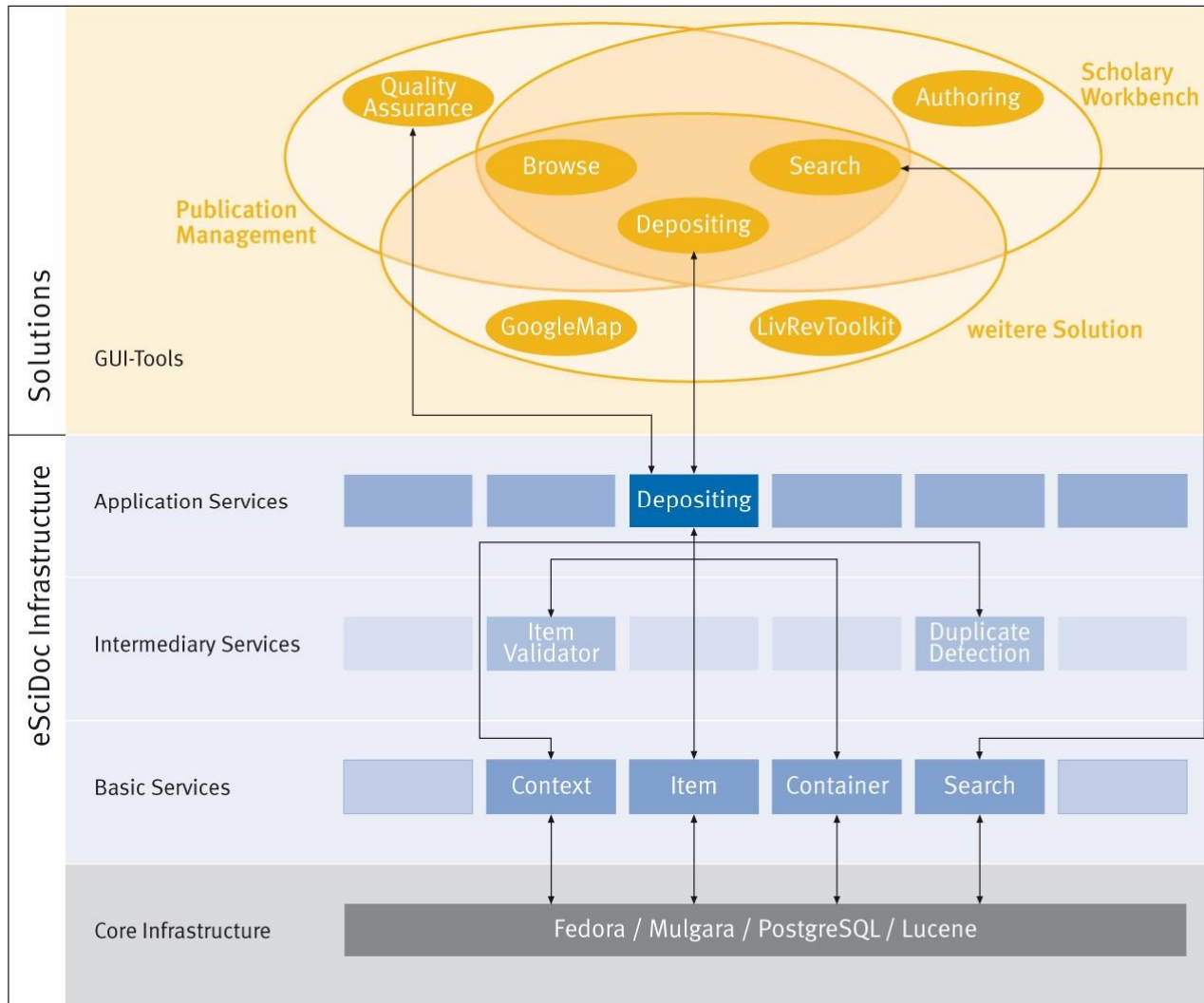
- eSciDoc covers a manifold of digital object types (e.g., experimental and primary data, persons, organizational units, traditional and new forms of publications)
- eSciDoc integrates into distributed authentication systems
- eSciDoc provides a fine-granular authorization











## However, eSciDoc is not a DBMS

- A typical design methodology for a framework is the use of an OR mapper (e.g., Hibernate)
- The eSciDoc infrastructure follows a different design pattern, which is mainly defined by the underlying repository Fedora
- Fedora focuses on
  - standard-compliant object container (METS)
  - simple validation and versioning of objects
  - disseminations (comparable to methods)
  - object relations

- Fedora's weaknesses include
  - lack of transactions
  - poor retrieval times for whole objects
  - non-standardized API

# Service-oriented Architecture



- Object Manager 
- Context Handler  
- Container Handler 
- Item Handler 
- Organizational Unit Handler 
- Authentication & Authorization   
  

- User & Role Manager
- PID Handler
- Validation Service
- Workflow Manager
- Statistics Manager
- Semantics Handler
- Search & Indexing Service 

## Services of the eSciDoc Infrastructure

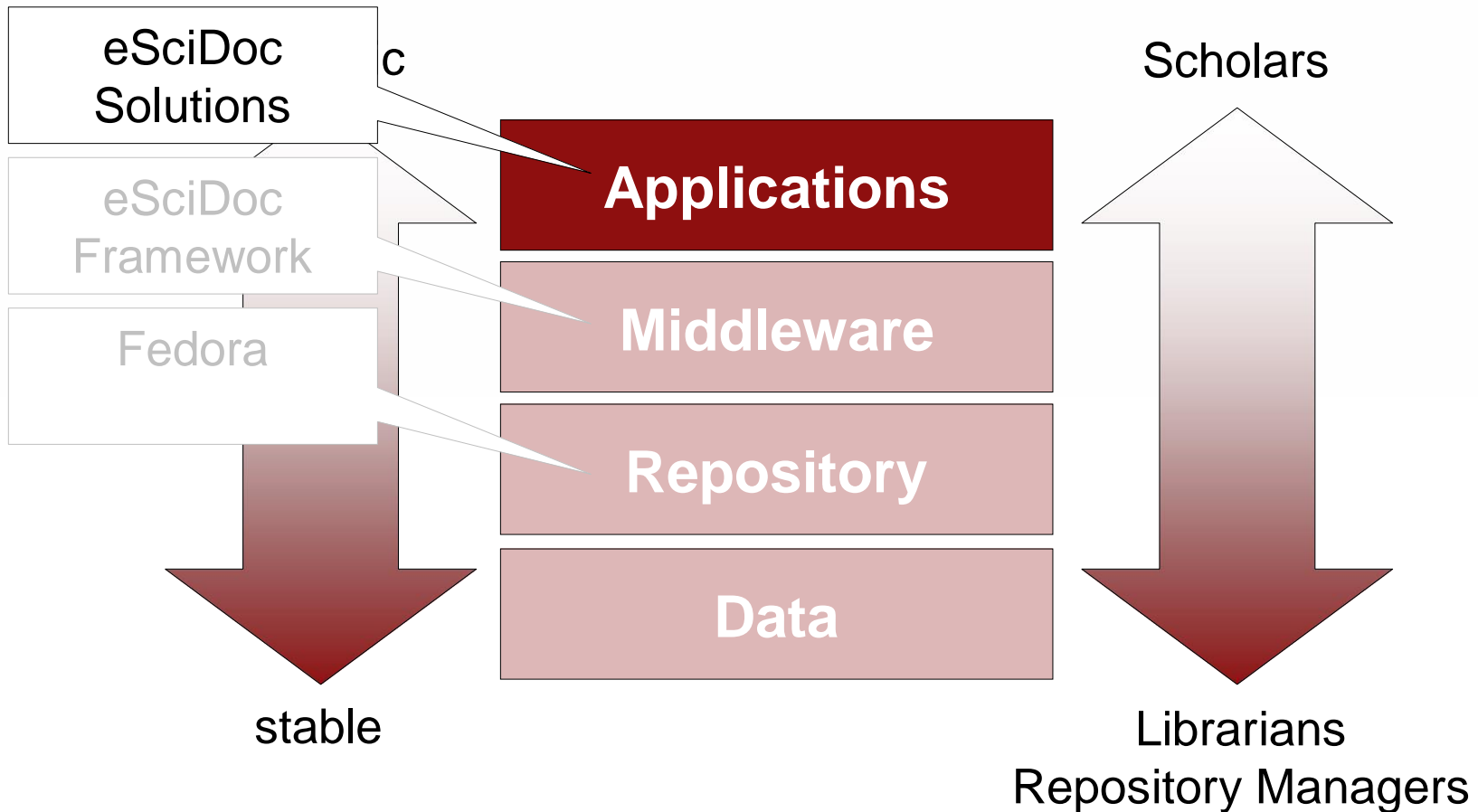


## Agenda

- Motivation

- Fedora
- eSciDoc Infrastructure
- **eSciDoc Solutions**

# Hierarchy of Persistence



# Solutions

- Publication Management
  - Group of pilots– Early adopters
  - Cluster (law institutes)
- Faces
  - MPI für Bildungsforschung
- ViRR
  - MPI für Europäische Rechtsgeschichte
- WALs Online
  - MPI für Evolutionäre Anthropologie
- LAMUS Integration
  - MPI für Psycholinguistik

# Solutions

- Publication Management
  - Group of pilots
  - Early adopters
  - Cluster (law institutes)

- Faces
  - MPI für Bildungsforschung

- ViRR
  - MPI für Europäische Rechtsgeschichte

- WALs Online
  - MPI für Evolutionäre Anthropologie

- LAMUS Integration
  - MPI für Psycholinguistik

“Scholarly  
Workbench”

# Publication Management

- Publication Management (“PubMan”) supports scientists and institutes in the management and the digital curation of their publications.
- This solution addresses all disciplines and focuses on the target groups
  - scientists, – local librarians and – local IT.
- Due to the heterogeneous policies, cultures and workflows of the Max-Planck-Institutes, customization and configuration options throughout the solution features are of major importance.
- See <http://colab.mpd.l.mpg.de/mediawiki/PubMan>

# Publication Management Features

- PubMan offers functionality for
  - Submitting content
  - Managing content
  - (Re-)Using content
- Important aspects of PubMan include the following concepts:
  - Versioning
  - Controlled vocabulary
  - Persistent Identification
  - Collections
  - Affiliations
- PubMan is an ongoing development (started in December 2006)

# Faces

- Faces is a lifespan database of adult emotional facial stimuli.
- It contains a set of images of naturalistic faces of 171 persons, separated in three different age groups (young, middle-aged and old) displaying each of 6 facial expressions:
  - neutrality,
  - sadness,
  - disgust,
  - fear, – anger, and – happiness.
- Development of Faces will start in October 2007

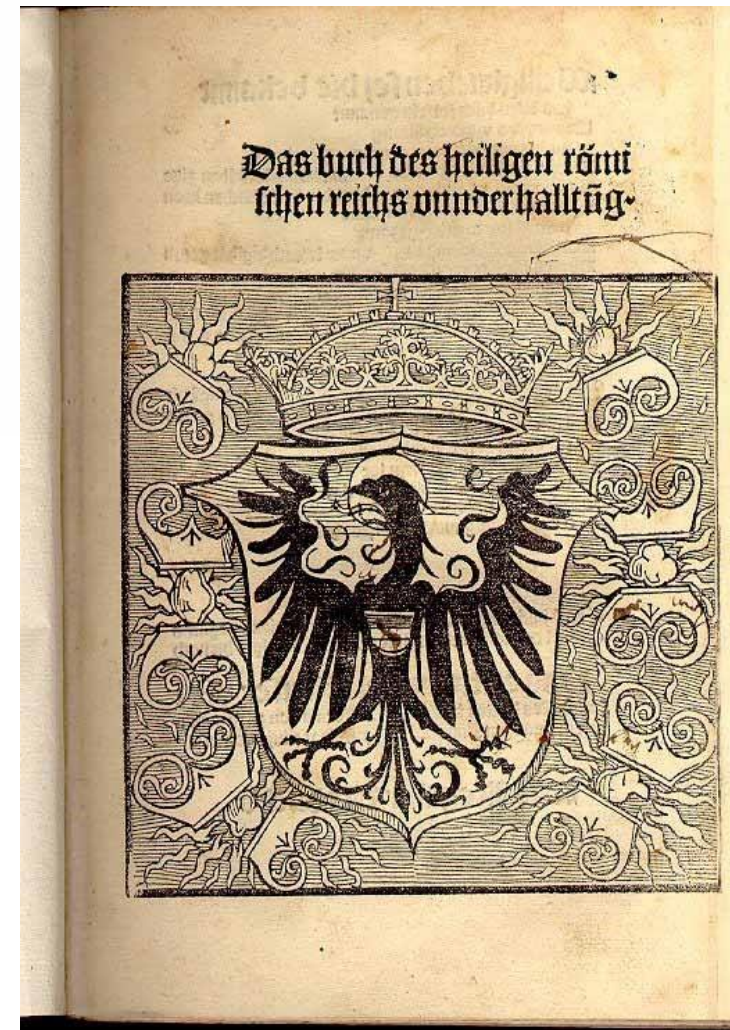
# Faces

 ID: 5445 Age: 19 Gender: F Emotion: <b>Happiness</b> <input type="checkbox"/> Include in Selection <a href="#">View Photo and all Attributes</a>	 ID: 5421 Age: 20 Gender: M Emotion: <b>Happiness</b> <input type="checkbox"/> Include in Selection <a href="#">View Photo and all Attributes</a>
 ID: 5421 Age: 20 Gender: M Emotion: <b>Happiness</b> <input type="checkbox"/> Include in Selection <a href="#">View Photo and all Attributes</a>	 ID: 5445 Age: 19 Gender: F Emotion: <b>Happiness</b> <input type="checkbox"/> Include in Selection <a href="#">View Photo and all Attributes</a>

## ViRR (Virtueller Raum Reichsrecht)

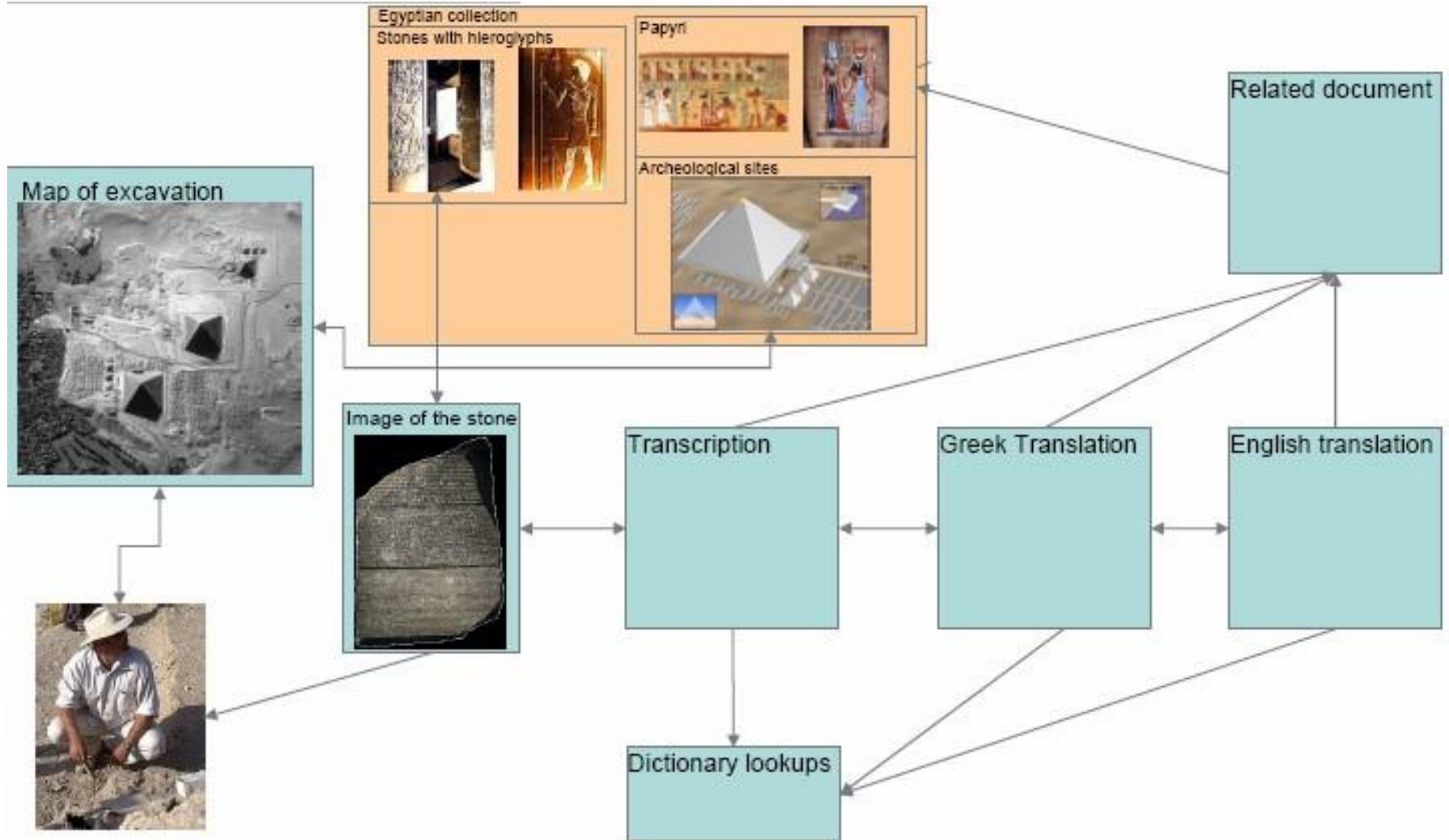
- ViRR covers information about the law of the Holy Roman Empire.
- It will start with digital scans of 15 books (together more than 16000 pictures).
- ViRR will offer online browsing and navigation functionalities.
- The collection will be expanded with transcriptions of the table of contents and transcriptions of all pages.

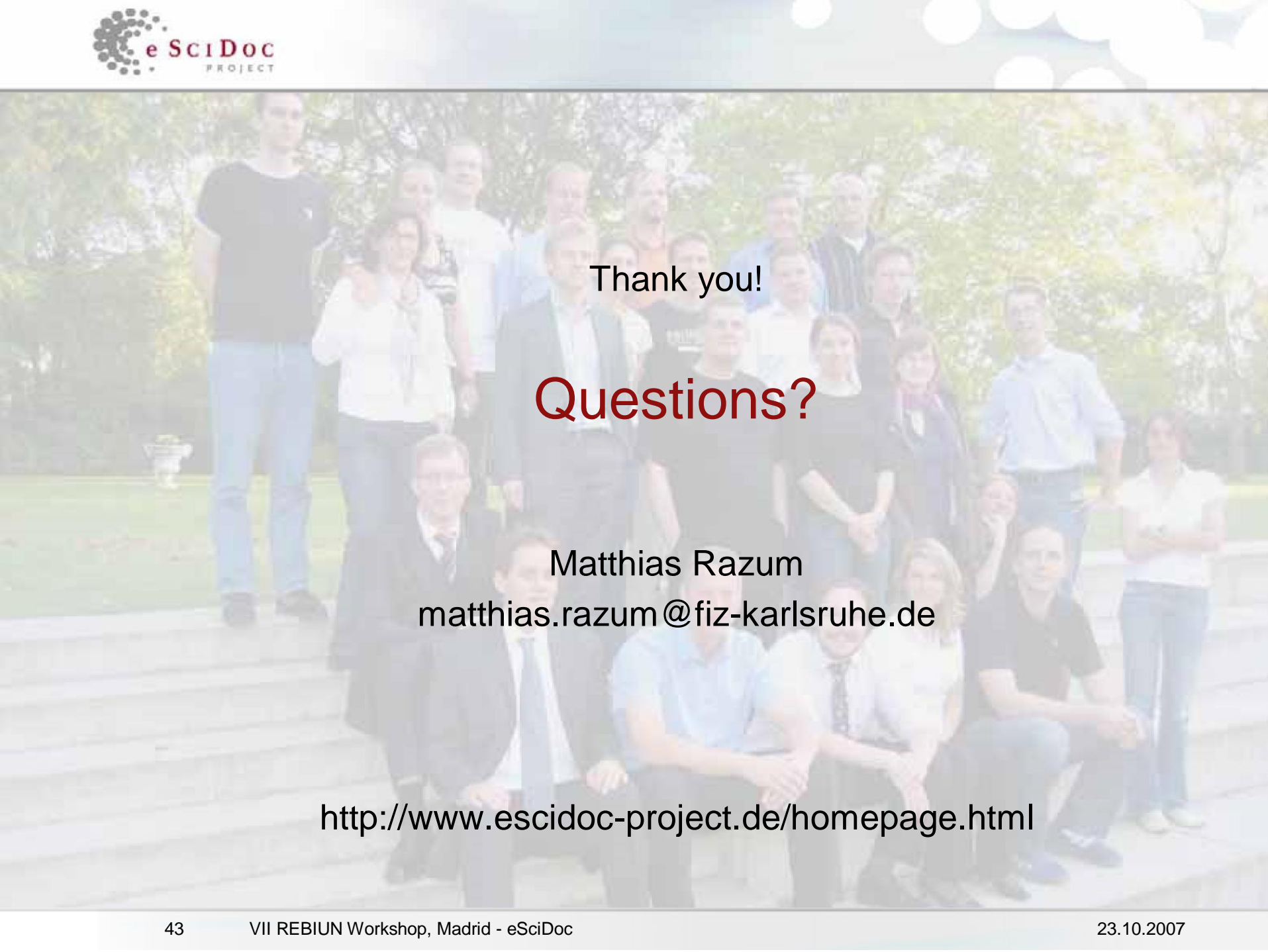


# ViRR

- The "Virtueller Raum Reichsrecht" will be a tool to work with.
- Envisioned working scenarios are:
  - Assignment of keywords and synonyms to the digitized work
  - Transcription and markup of the digitized work
  - Submission of documents
  - Annotations (with different visibilities) of the digitized work
- Development of ViRR will start in 2008

# LAMUS





Thank you!

Questions?

Matthias Razum

matthias.razum@fiz-karlsruhe.de

<http://www.escidoc-project.de/homepage.html>