From Digital Libraries to e-Research in Europe: a Challenge for your Institutional Repository?

Matthias Razum
FIZ Karlsruhe

Library Fair
Yokohama, Japan
November 11, 2009

Agenda

- **Institutional Repositories, Digital Libraries, and their challenges**
- Data and Information Infrastructures in Europe
- Virtual Research Environments
- eSciDoc

Institutional Repositories and Digital Libraries

Characteristics of Institutional Repositories and Digital Libraries

- Content mostly traditional publications
- Maintained by librarians
- Selection and quality-ensurance processes in place
- Well-structured content
- High-quality metadata
- Data is fixed

Important IR Software

Usage of Institutional Repository software systems (November 2009, based on ROAR):

IR System	World-wide	Germany
DSpace	32,1 %	6,3 %
EPrints	22,9 %	13,5 %
OPUS	2,4 %	30,6 %
Fedora	1,0 %	0,1 %

Registry of Open Access Repositories (ROAR), http://roar.eprints.org/

What is e-Research?

The invention and exploitation of advanced IT:

- to generate, curate and analyse **research data**;
- to develop and explore models and simulations;
- to enable dynamic distributed virtual organisations

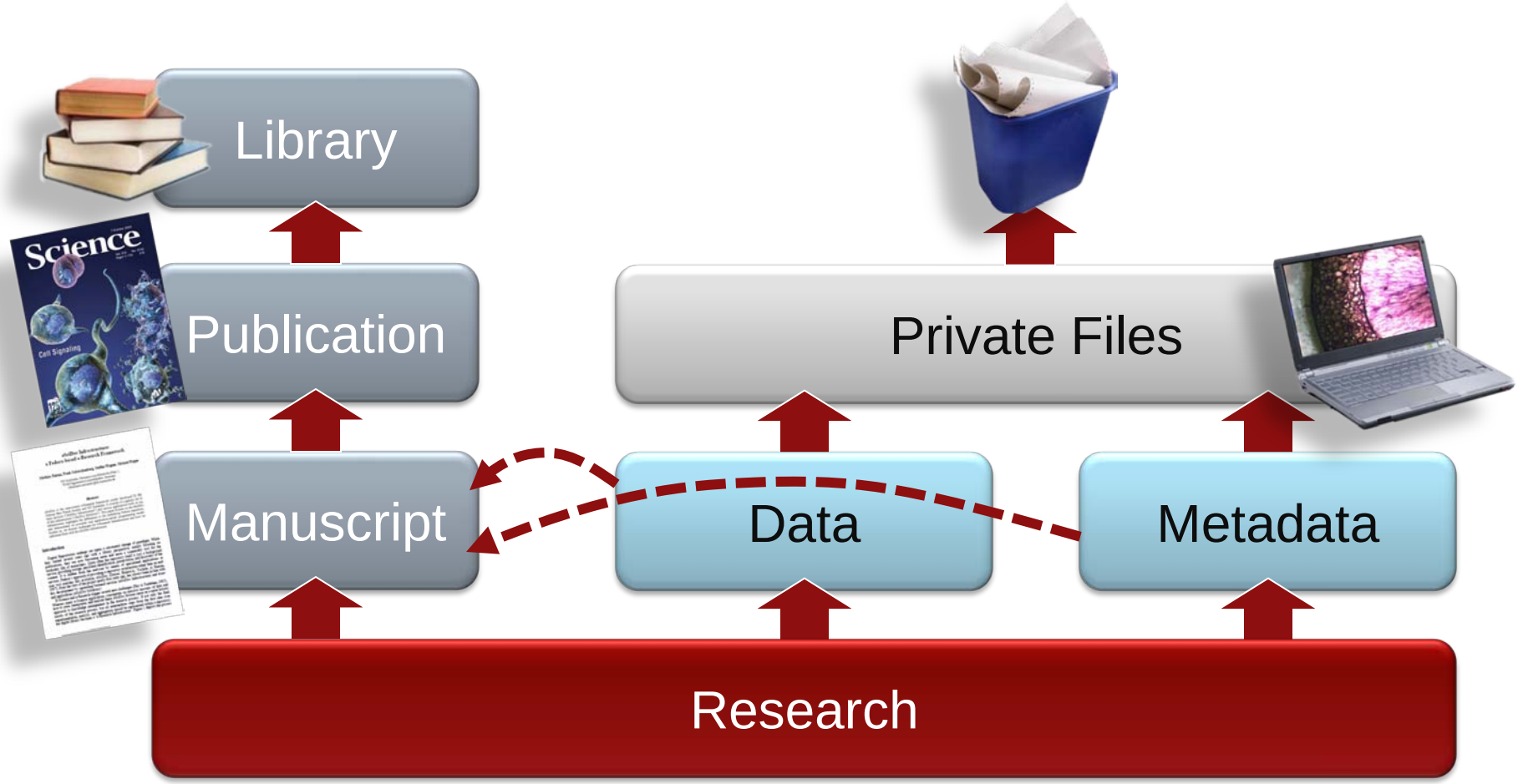
Tony Hey (former Director of UK e-Science Core Programme, UK)

Data – The Next Challenge

- Most research today is data-driven, even in the humanities
- Reference data should be accessible independently of scientific papers in a citable form, allowing attribution
- Scholarly publication
  - remains essential for describing and contextulizing findings,
  - but is inadequate as the only document of research activity
- Databases
  - are essential for comprehensively archiving both published and unpublished data,
  - but are not yet fully integrated into the scientific process

Editorial „Sharing Data“, Nature Cell Biology, 11 (11), 2009

Data in the Publication Process Today



Helly, J., H. Staudigel, and A. Koppers, Scalable models of data sharing in Earth sciences, Geochem. Geophys. Geosyst., 4(1),1010, doi:10.1029/2002GC000318, 2003

Agenda

- Institutional Repositories, Digital Libraries, and their challenges
- **Data and Information Infrastructures in Europe**
- Virtual Research Environments
- eSciDoc

Data and Information Infrastructures

"Equally, if not more important than its own data and information needs, today's research community must also assume **responsibility** for building a **robust data and information infrastructure** for the **future**."

International Council for Science, ICSU, 2004

Examples from Europe

- DRIVER
  - pan-European infrastructure for digital repositories
  - e-Infrastructure is used to store knowledge
- Enabling Grids for E-sciencE (EGEE)
  - Europe's leading grid computing project
  - providing a computing support infrastructure for over 10,000 researchers world-wide
- DILIGENT
  - advanced testbed for knowledge e-Infrastructure
  - integrating **grid** and **digital library** technology
- BRICKS
  - organizational and technological foundations of a digital library

Agenda

- Institutional Repositories, Digital Libraries, and their challenges
- Data and Information Infrastructures in Europe
- **Virtual Research Environments**
- eSciDoc

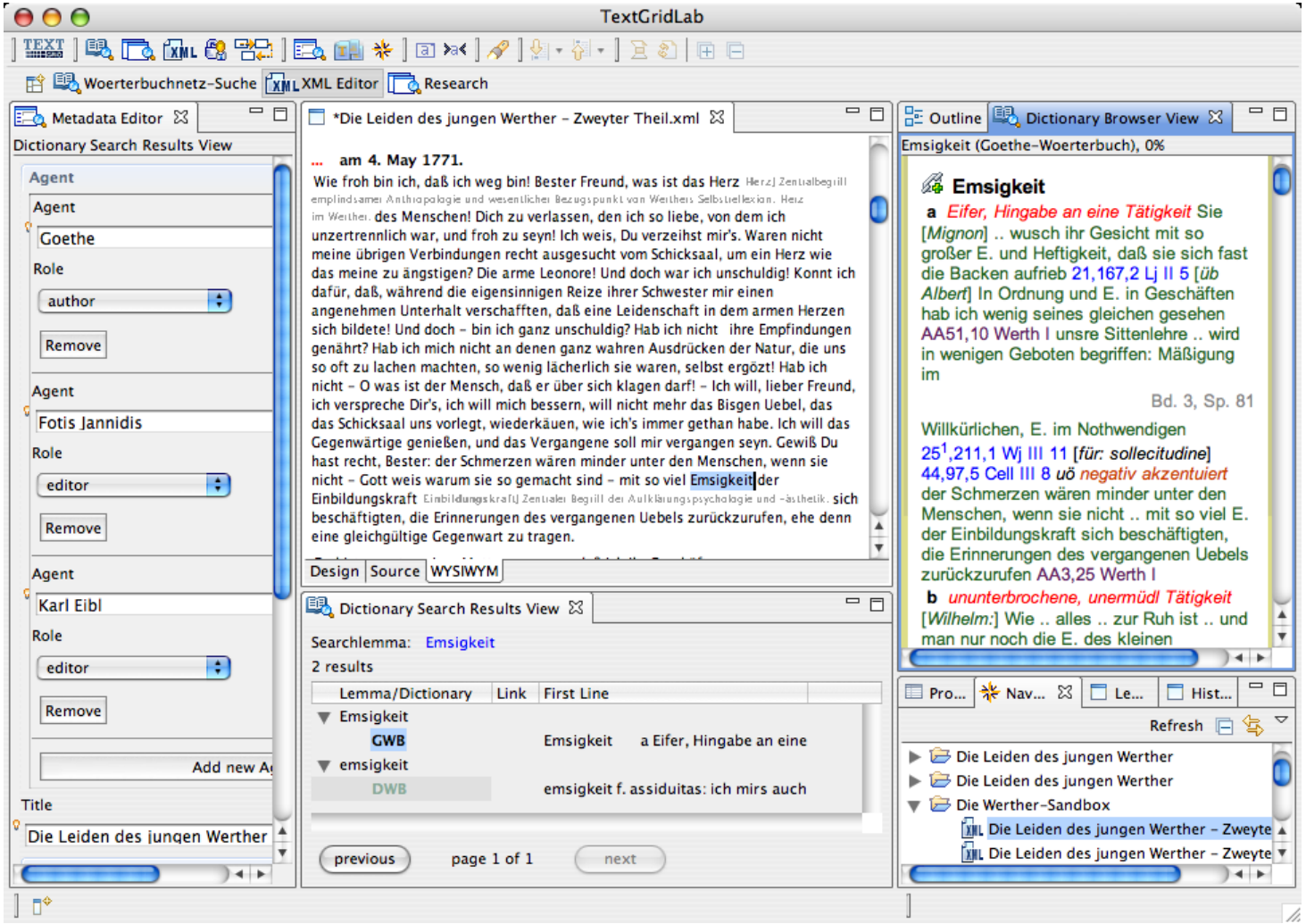
Virtual Research Environments

"The aim of a Virtual Research Environment (VRE) is to help researchers manage this complexity [of research] by providing an infrastructure specifically designed to support the activities carried out within research teams, on both small and large scales."

JISC VRE Roadmap (2004)

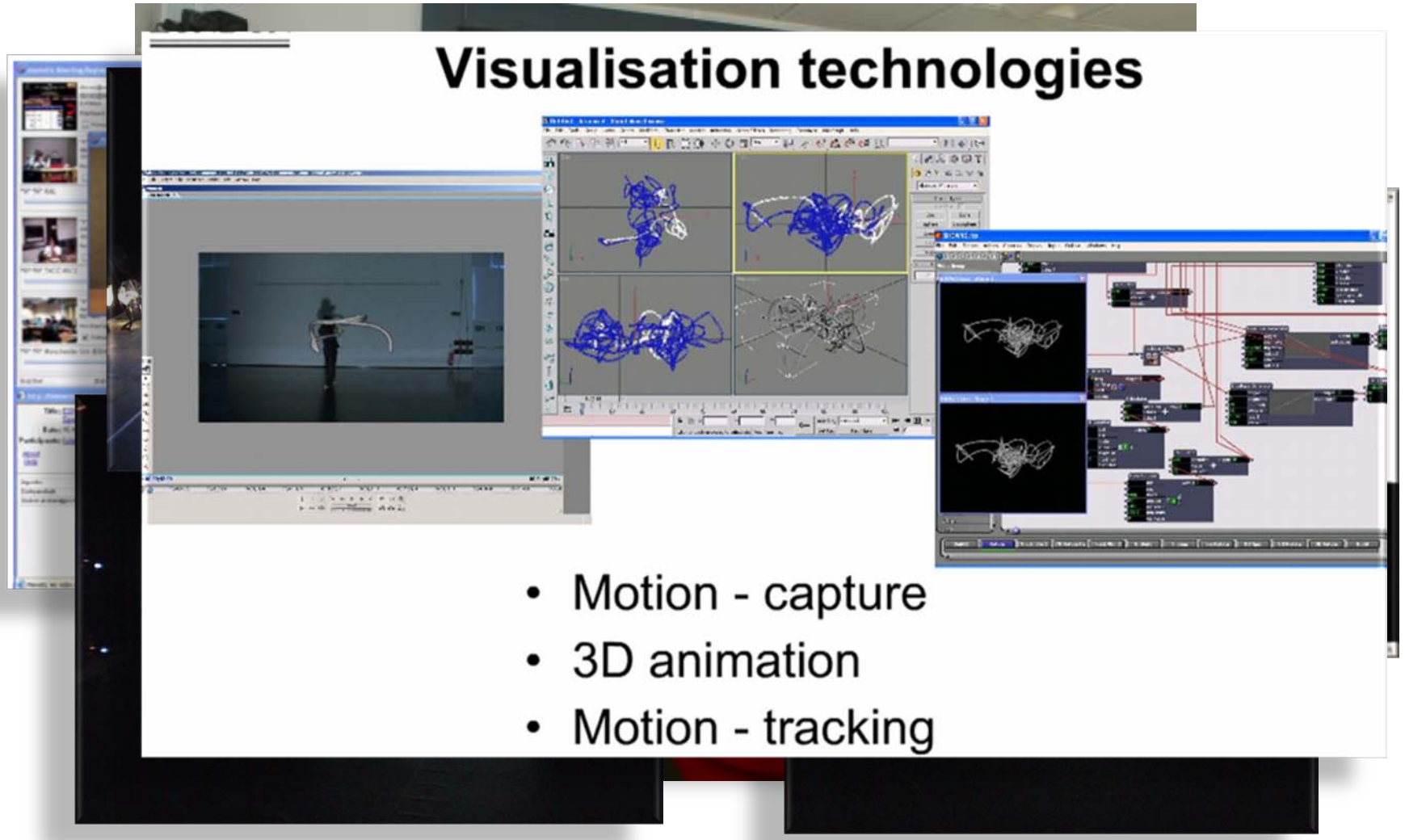
- Originally planned to widen the reach of e-Science
- VRE's are not just about grand challenge science, but everyday research

TextGridLab



Microsoft Research Information Centre

From Meetings to Dancing



http://laws1.reading.ac.uk - IADB Document Editor - Mozilla Firefox

Project SILCH: 124 - Timber Building 1

Close

Document | Metadata | Discussion | History

[Print | Edit Document]

Early Roman Timber Building 1A

Timber Building 1

Occupying what became the angle between the main north-south street and the subsidiary east-west street our building consisted of two rooms. The larger, Room 1 (Object 50043, Plan, Matrix 290), measures c.10 by 7m, giving an internal surface area of 70 square metres, was bounded by shallow slots (5393, 5304, 5375, 5357, 5356, 5346, 5396, 5371, 5323) filled with yellow clay. This is assumed to be material associated with the infilling of the wall-structure, which, over time, replaced the voids created by the decayed horizontal sill-beams. At the centre of the room was a rectangular, tiled hearth (1433), built up of horizontally laid, broken tile fragments and associated with dark, charcoal-rich soils (6038, 5919). At least three separate phases of use of the hearth can be identified, each associated with a re-laying or re-patching of the floor. The associated floor surfaces of the room were composed

http://laws1.reading.ac.uk -

Project SILCH: 184 - Ces

SILCH: Cont

Fill of cess

Small fragment of possible amber ring SF3516

Small fragment of possible amber ring SF3516

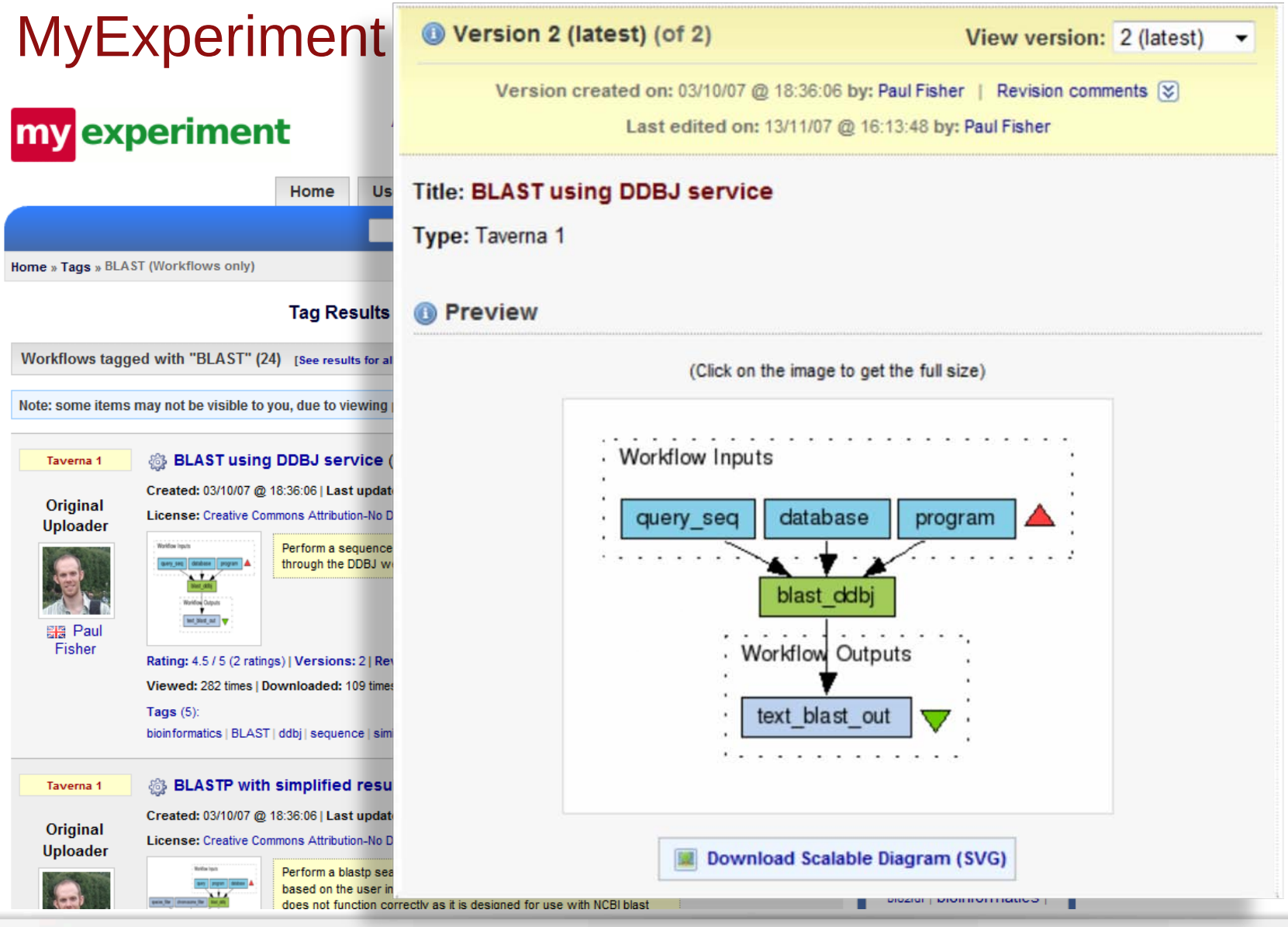
Design Principles for (Social) VREs

- Fit in, Don’t Force Change
- Act Local, think Global
- Enable Users to Add Value
- Design for Network Effects
- Keep your Friends Close
- Embed

- But: Researchers collaborate to compete!

De Roure, D. and Goble, C. "Software Design for Empowering Scientists," IEEE Software, vol. 26, no. 1, pp. 88-95, January/February 2009

MyExperiment



Social VREs, Collaboration, and Policies

- Researchers use informal and trusted sources of advice from colleagues, rather than institutional service teams, to help identify information sources and resources
- The use of social networking tools for scientific research purposes is far more limited than expected
- Data and information sharing activities are mainly driven by needs and benefits perceived as most important by life scientists rather than ‘top-down’ policies and strategies

Williams, R., Pryor, G. et al. “Patterns of information use and exchange: case studies of researchers in the life sciences”, Report by the Research Information Network and the British Library, November 2009

Lessons Learned: Data

- Lots of new digital-born objects (tags, relations, annotations)
- Clash of Digital Library models:
  Curated by experts (librarans) ←→ self-organizing systems
- New expertise is required to curate research data
- Librarians need to manage trust, while sharing control

Lessons Learned: VREs

- Standardization is difficult to achieve, even within a disciline
  - We need to avoid contraining researchers and their methodologies
  - Enable researchers to create their own solutions
- Talk to researchers: they understand their processes, data entities, and requirements best
  - Show them early prototypes to avoid misunderstandings
- Researchers are reluctant to change processes and tools that work for them
  - Try to re-use and integrate existing tools instead of re-implementing functionality

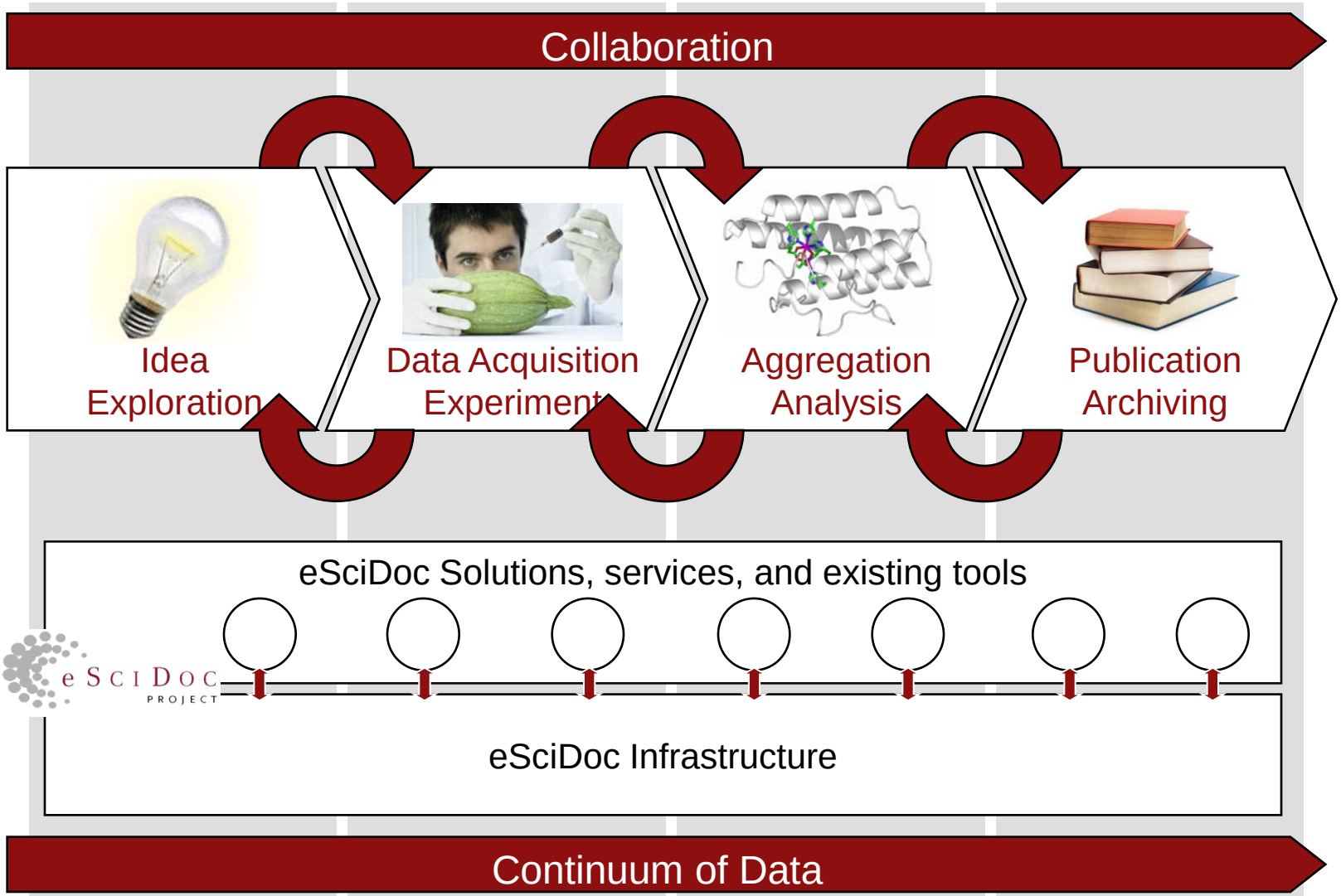
Agenda

- Institutional Repositories, Digital Libraries, and their challenges
- Data and Information Infrastructures in Europe
- Virtual Research Environments
- **eSciDoc**

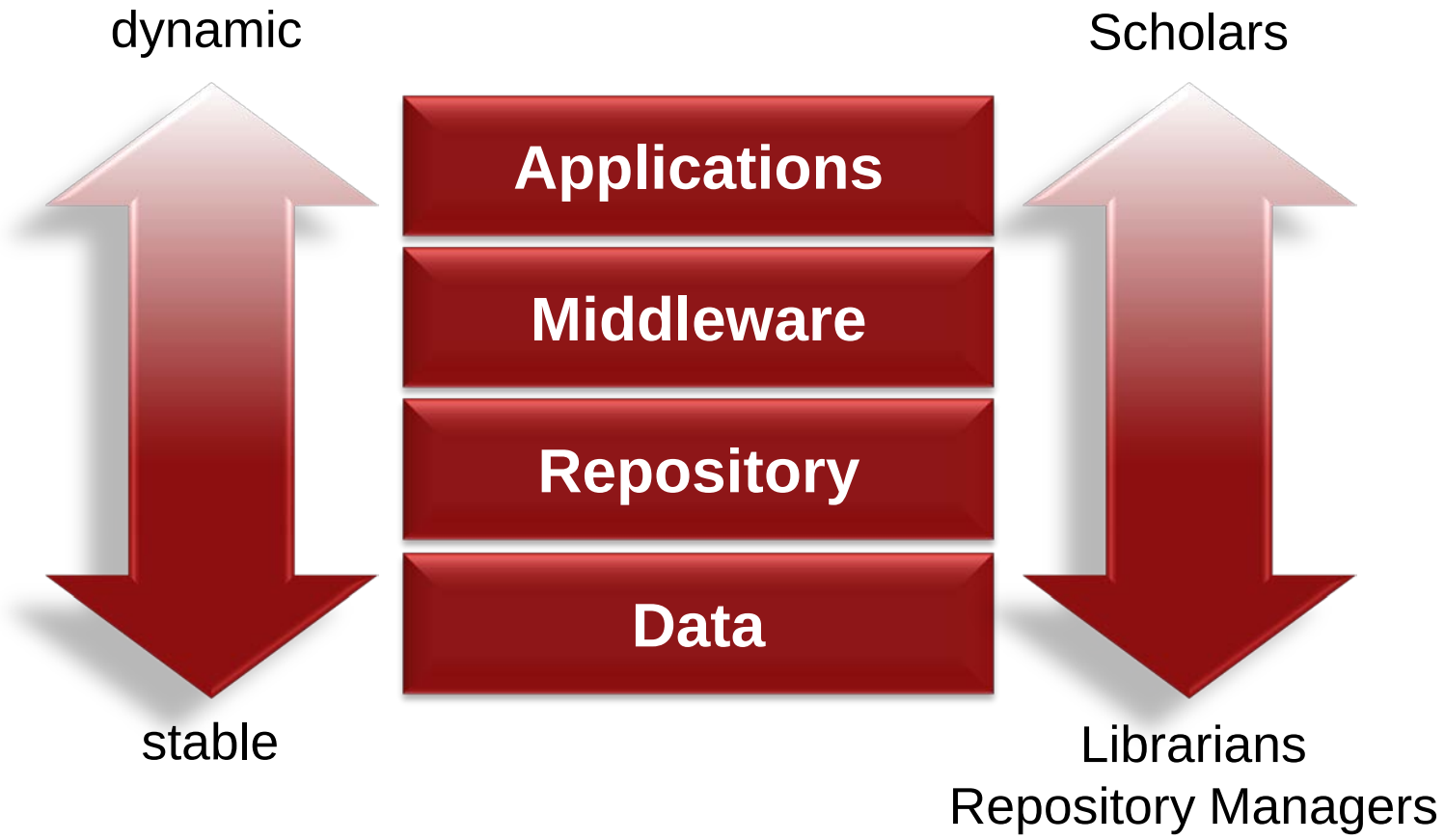
eSciDoc

- eSciDoc is as a joint project of the Max Planck Society and FIZ Karlsruhe, funded by the Federal Ministry of Education and Research (BMBF), aimed at building an e-Science platform for multi-disciplinary research organizations.

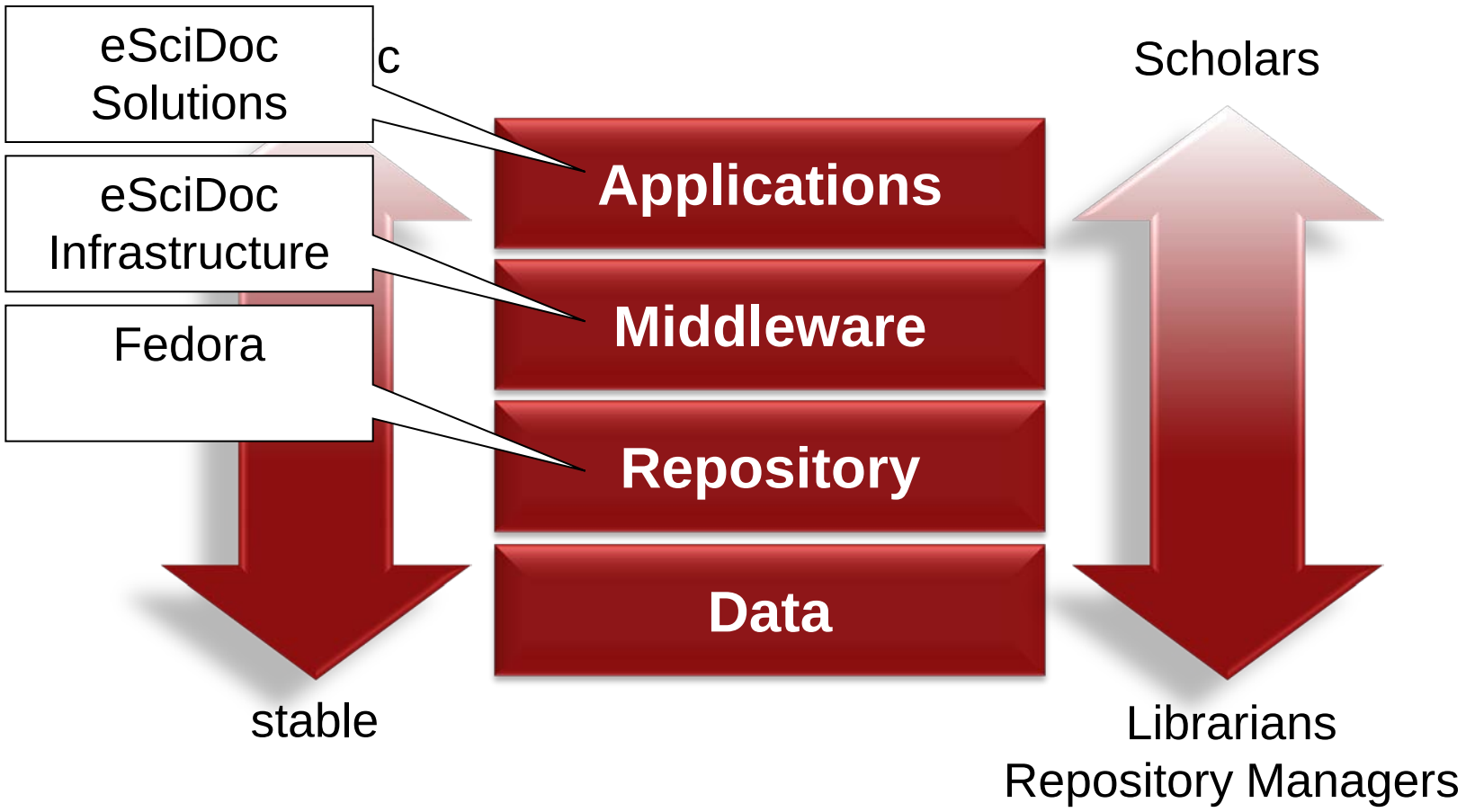
- eSciDoc
  - integrates research results and materials in an emerging e-research network,
  - provides effective and comprehensive access to data and information
  - supports collaboration and interdisciplinary research in future e-Science scenarios
  - increases the accountability of research
  - improve the visibility of research institutions and organizations


Collaboration
Idea
Exploration
Data Acquisition
Experiment
Aggregation
Analysis
Publication
Archiving
eSciDoc Solutions, services, and existing tools
eSciDoc Infrastructure
Continuum of Data

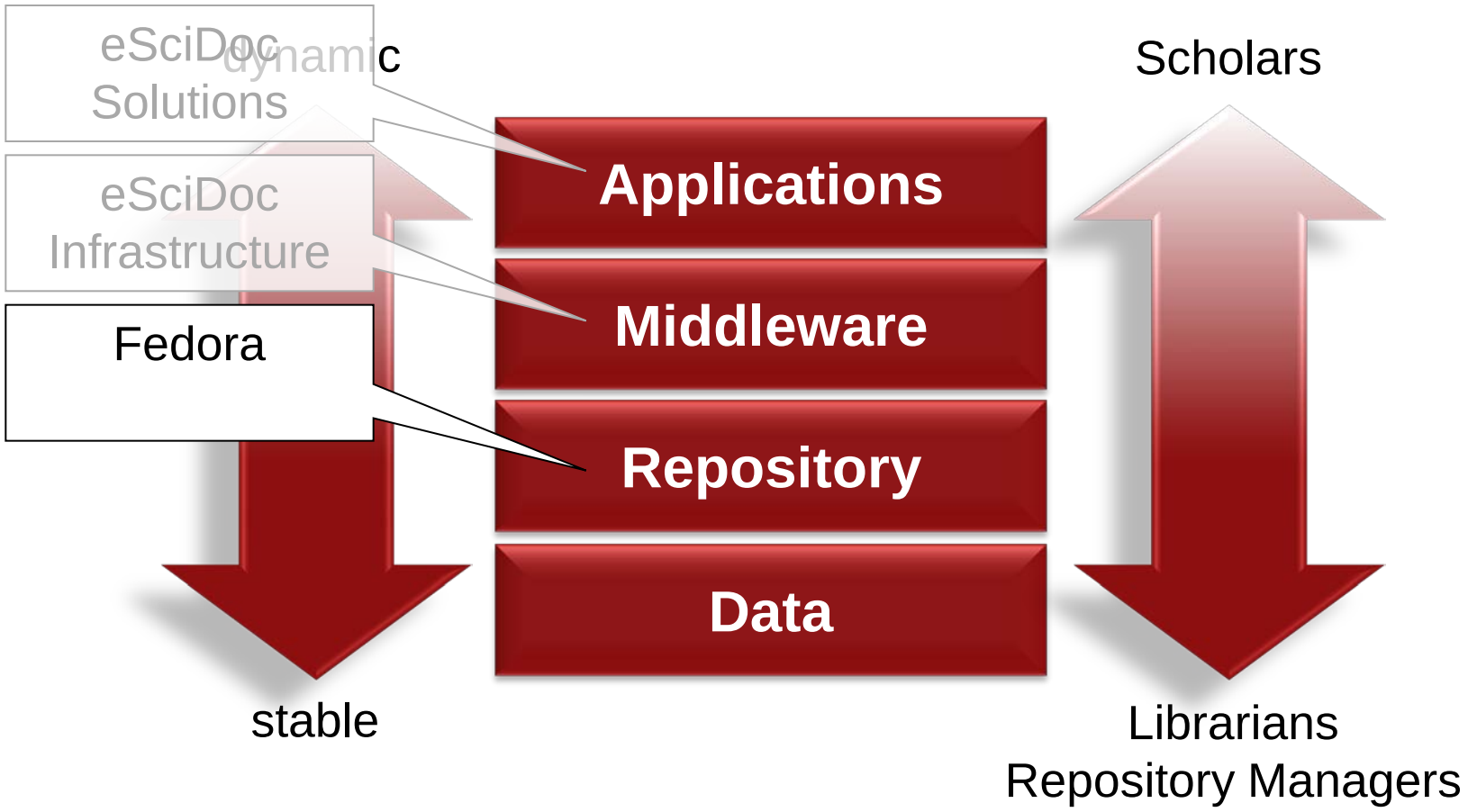
Hierarchy of Persistence



Hierarchy of Persistence



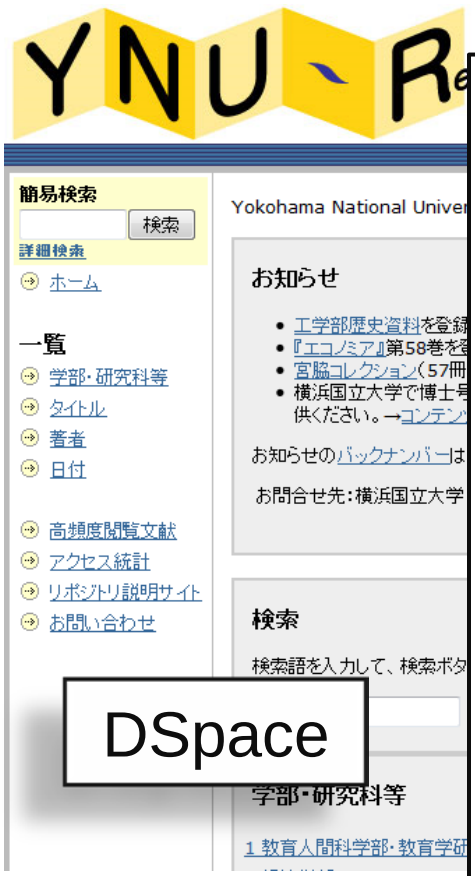
Hierarchy of Persistence



Fedora is an Acronym for

- **F**lexible
- **E**xtensible
- **D**igital
- **O**bject
- **R**epository
- **A**rchitecture

What means „Repository Architecture“?



DSpace

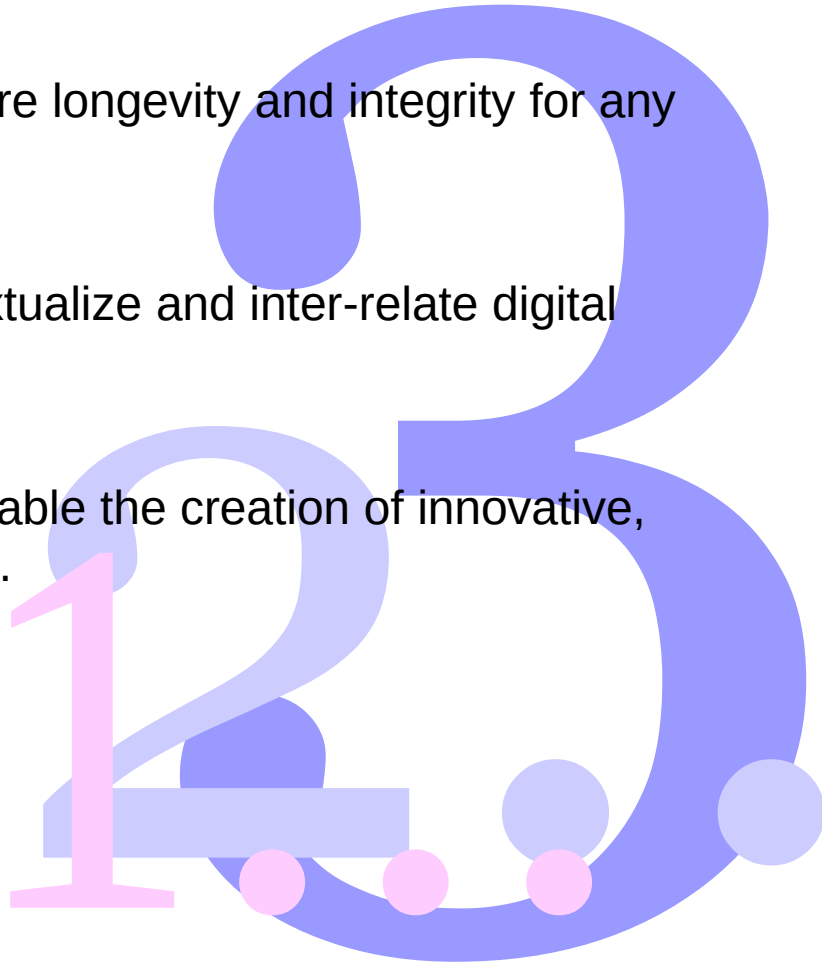
Yokohama National University

```
<?xml version="1.0" encoding="UTF-8"?>
<foxml:digitalObject VERSION="1.1" PID="escidoc:ex1"
xmlns:foxml="info:fedora/fedora-system:def/foxml#"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="info:fedora/fedora-system:def/foxml# http://www.fedora
<foxml:objectProperties>
<foxml:property NAME="info:fedora/fedora-system:def/model#state" VALUE="Ac
<foxml:property NAME="info:fedora/fedora-system:def/model#label" VALUE="Co
<foxml:property NAME="info:fedora/fedora-system:def/model#createdDate" VAL
<foxml:property NAME="info:fedora/fedora-system:def/view#lastModifiedDate"
</foxml:objectProperties>
<foxml:datastream ID="AUDIT" STATE="A" CONTROL_GROUP="X" VERSIONABLE="fals
<foxml:datastreamVersion ID="AUDIT.0" LABEL="Audit Trail for this object"
<foxml:xmlContent>
<audit:auditTrail xmlns:audit="info:fedora/fedora-system:def/audit#">
<audit:record ID="AUDREC1">
<audit:process type="Fedora API-M"/>
<audit:action>ingest</audit:action>
<audit:componentID></audit:componentID>
<audit:responsibility>fedoraAdmin</audit:responsibility>
<audit:date>2009-09-16T15:55:51.790Z</audit:date>
<audit:justification>eSciDoc object created</audit:justification>
</audit:record>
</audit:auditTrail>
</foxml:xmlContent>
</foxml:datastreamVersion>
```

Fedora

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.



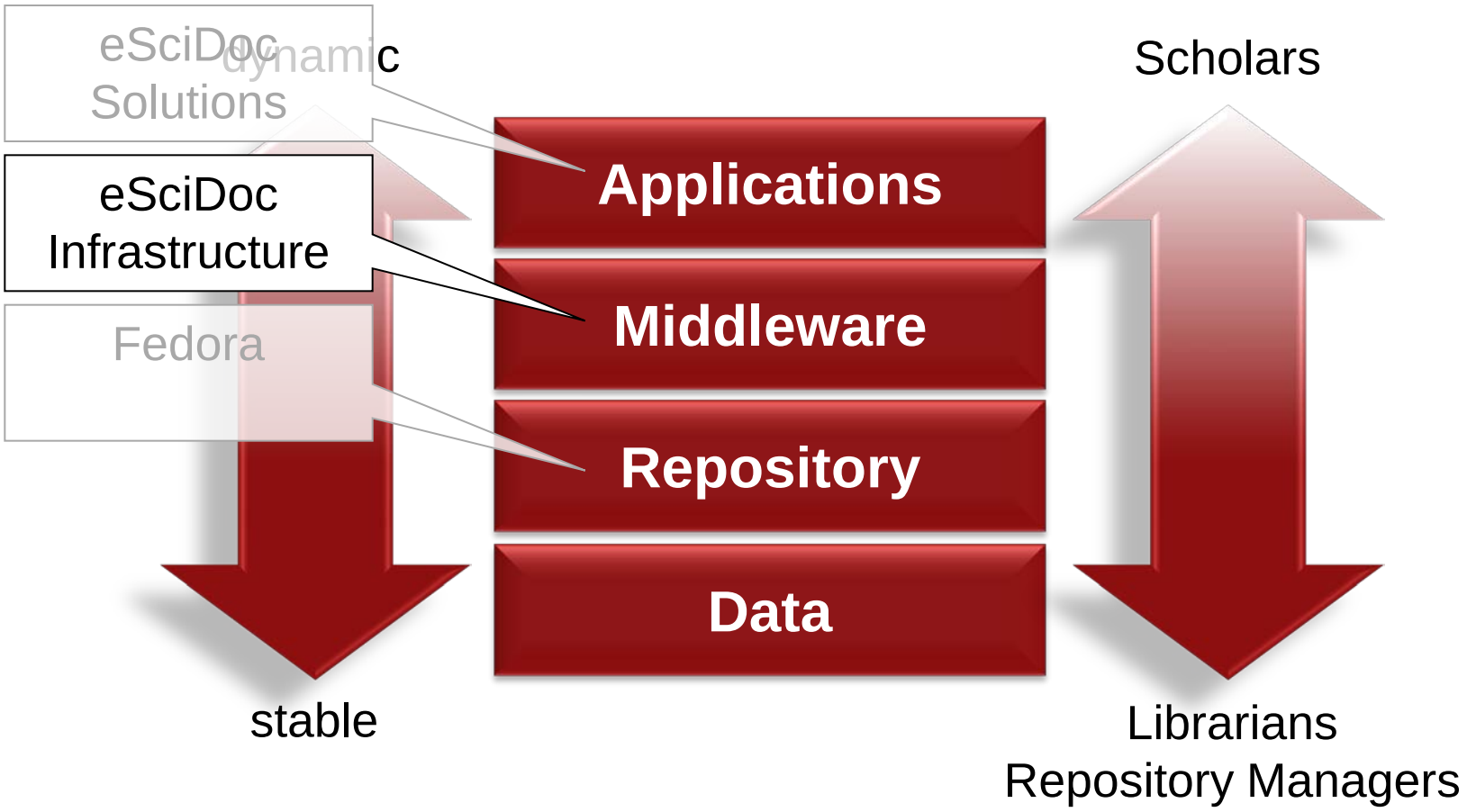
Fedora's Key Features

- 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

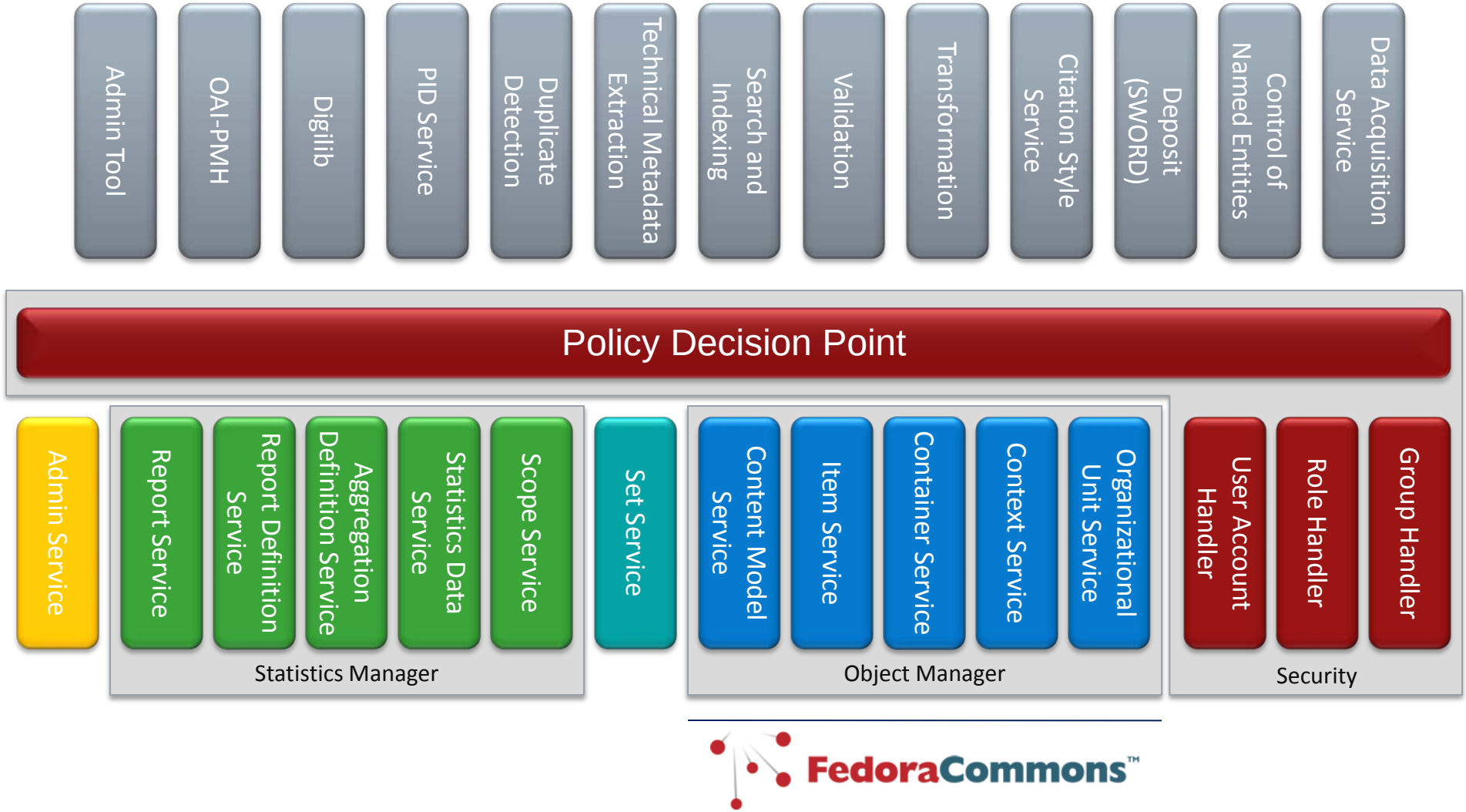
Fedora, Duraspace, and eSciDoc

- Fedora Commons and Dspace Foundation are joining their organizations to pursue a common mission.
- The new organization called Duraspace will sustain and grow both repository systems: Dspace and Fedora.
- eSciDoc benefits from the strengthend Duraspace foundation and the increased development and innovation power.
- We are Fedora committers, thus influencing the further development of the Fedora software.

Hierarchy of Persistence



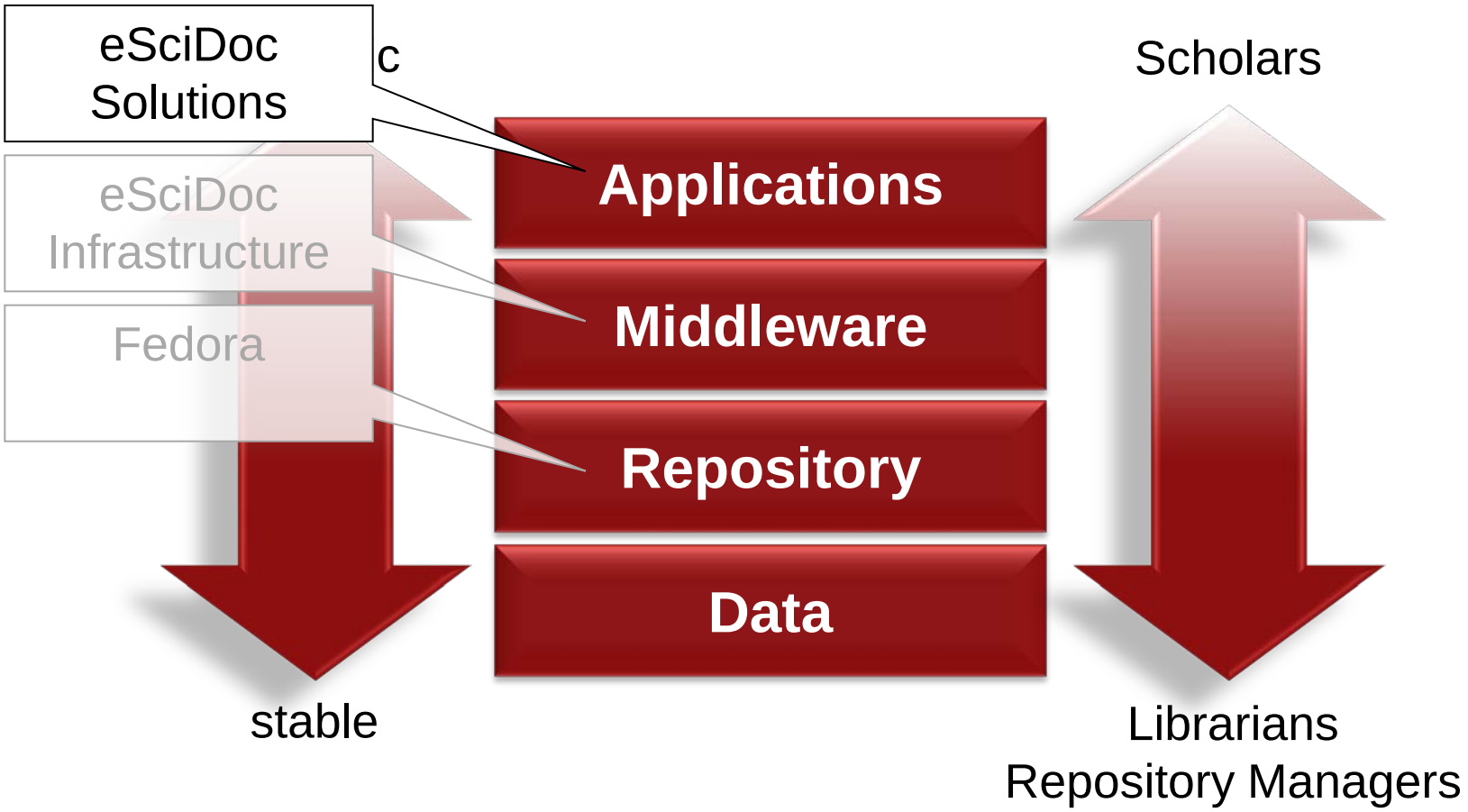
eSciDoc Services



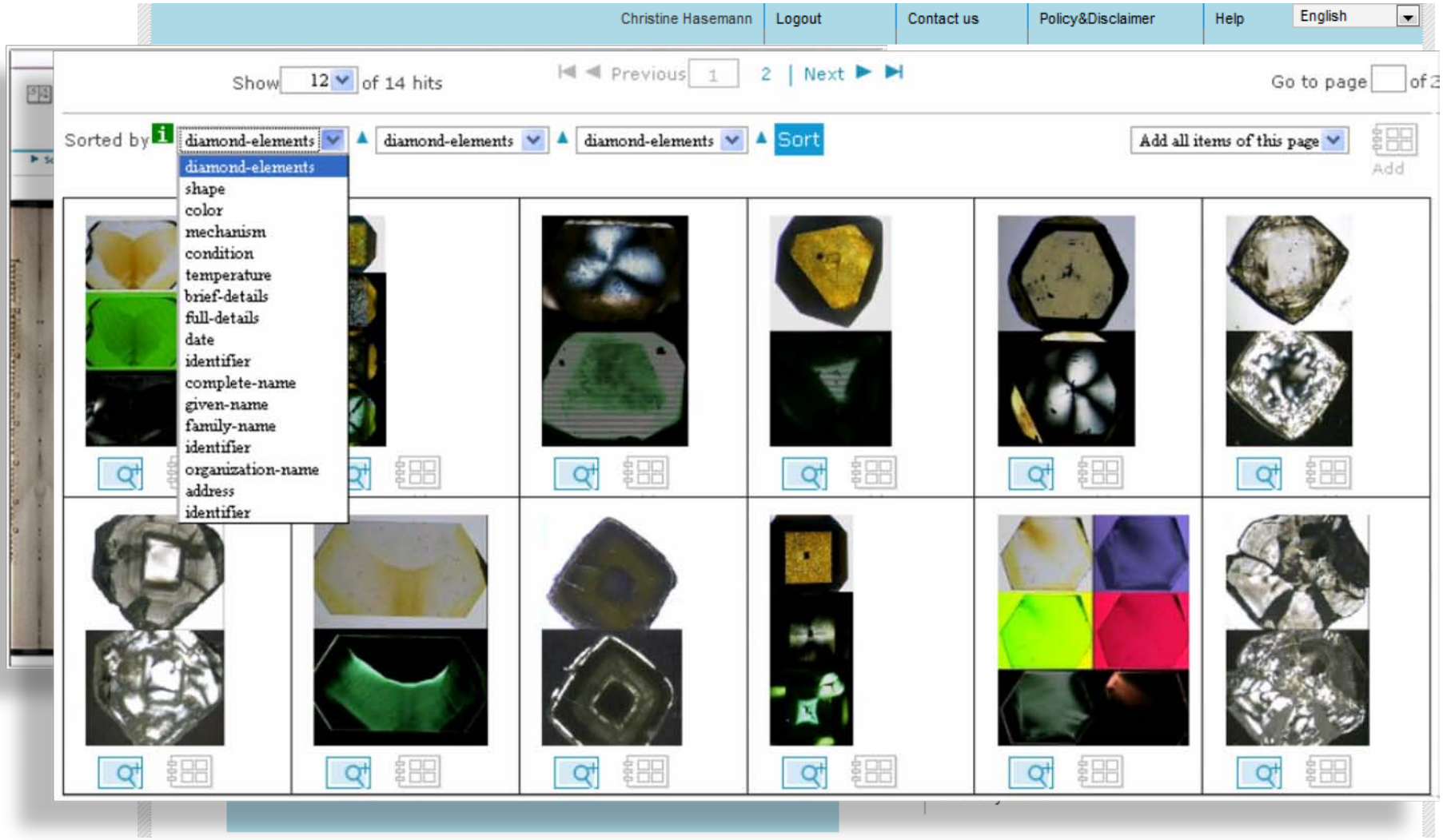
Key Features of the eSciDoc Infrastructure

- Service-oriented architecture
- Stable Core Services
- Flexible content models
- Arbitrary metadata profiles
- Application-independent design
- Support for object relations and multiple ontologies
- Search (OpenSearch, SRW/SRU)
- Distributed Authentication/Authorization (Shibboleth)

Hierarchy of Persistence



eSciDoc Solutions



Thank you!

Questions?

Matthias Razum
matthias.razum@fiz-karlsruhe.de

http://www.escidoc.org/