eSciDoc Infrastructure

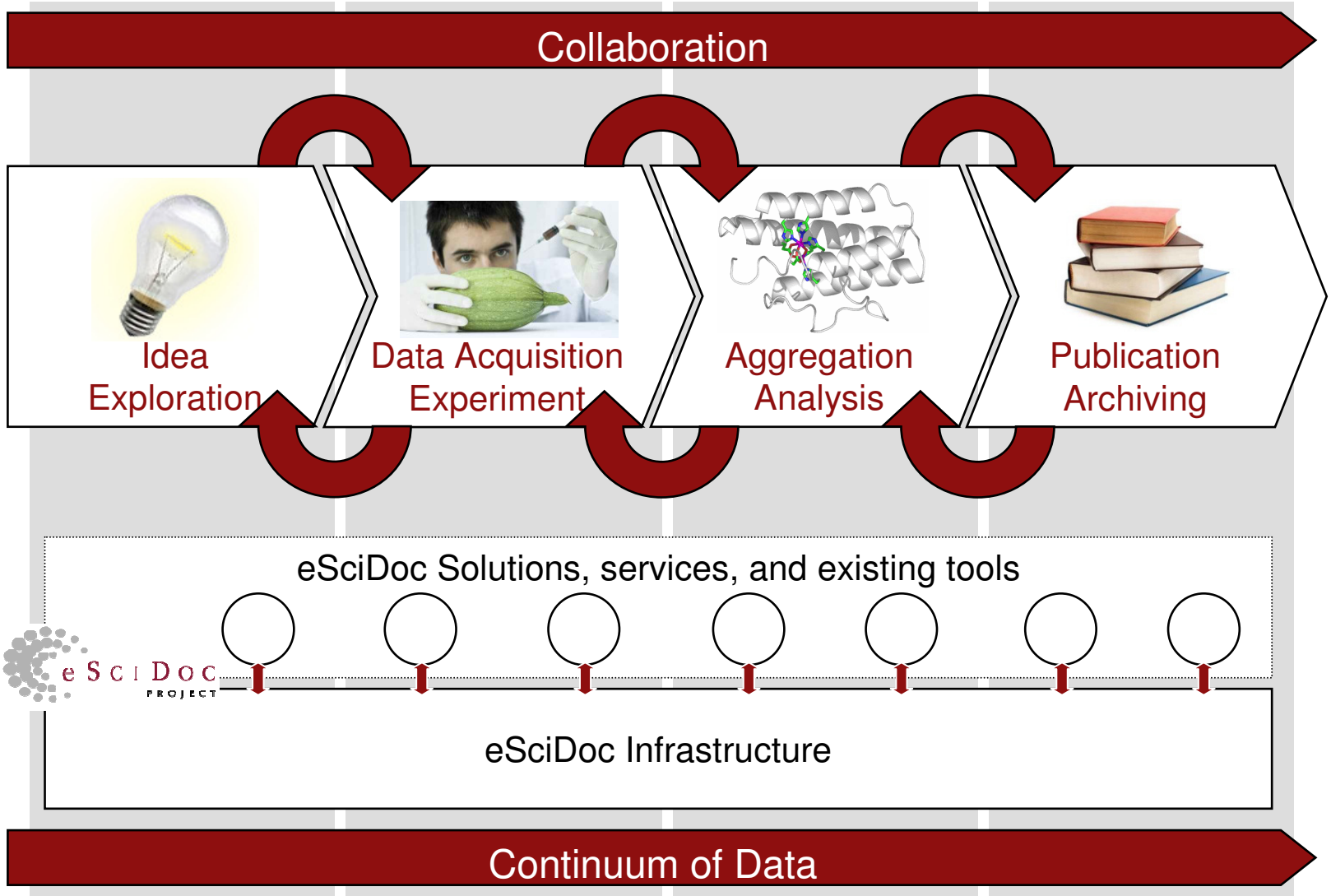
Matthias Razum
FIZ Karlsruhe

Berlin
6. November 2008

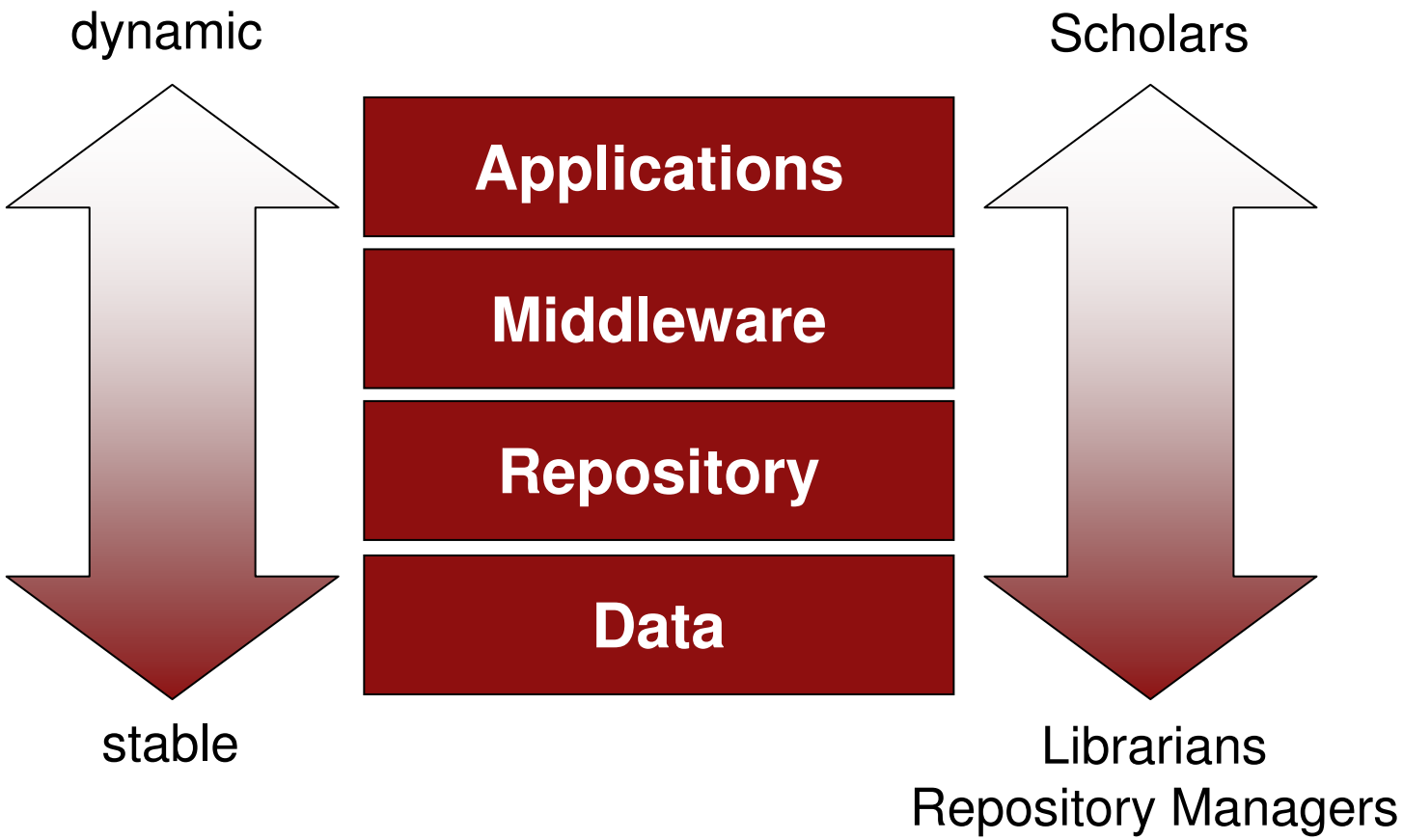
Data and Information Infrastructures

“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****.“*

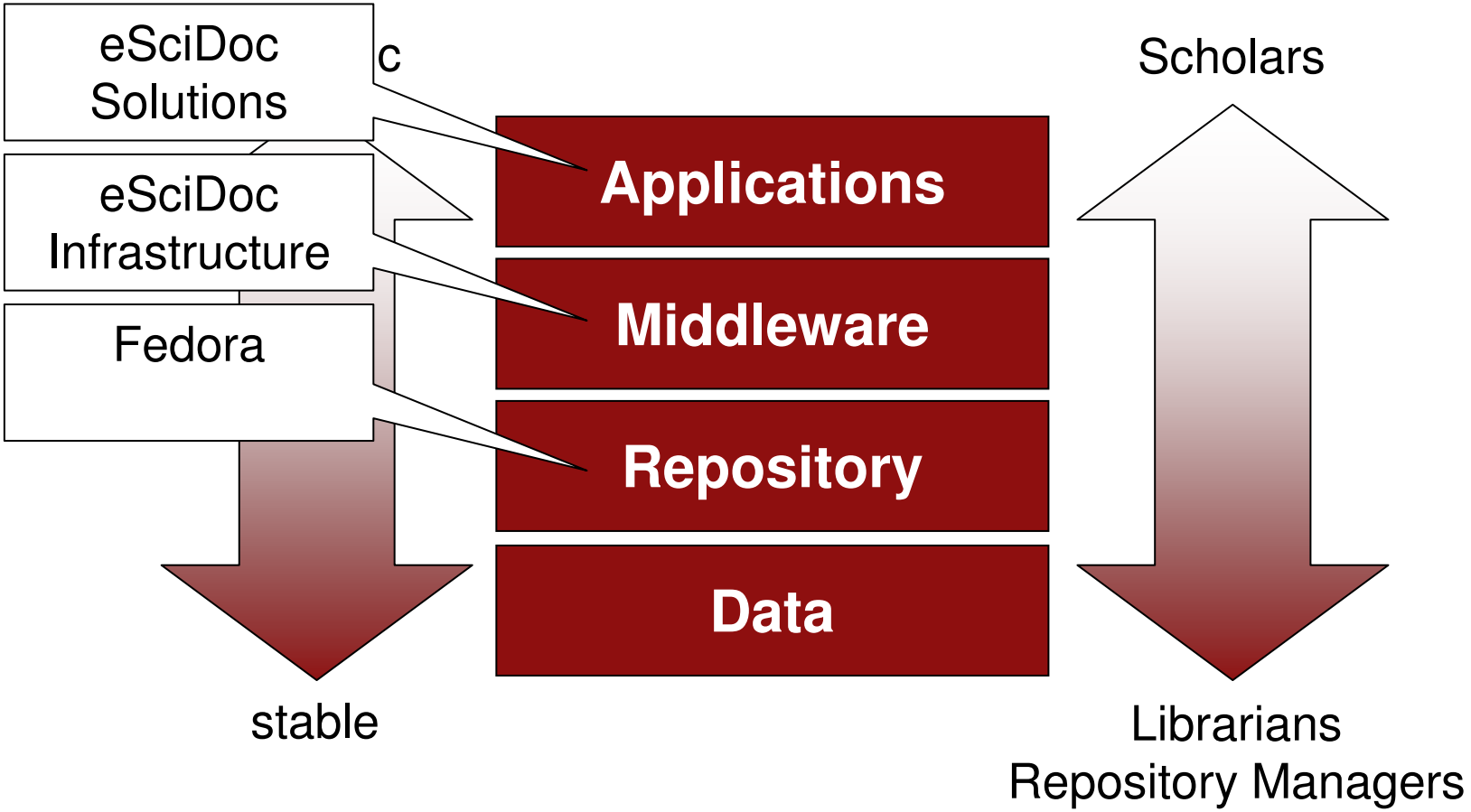
(International Council for Science, ICSU, 2004).


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



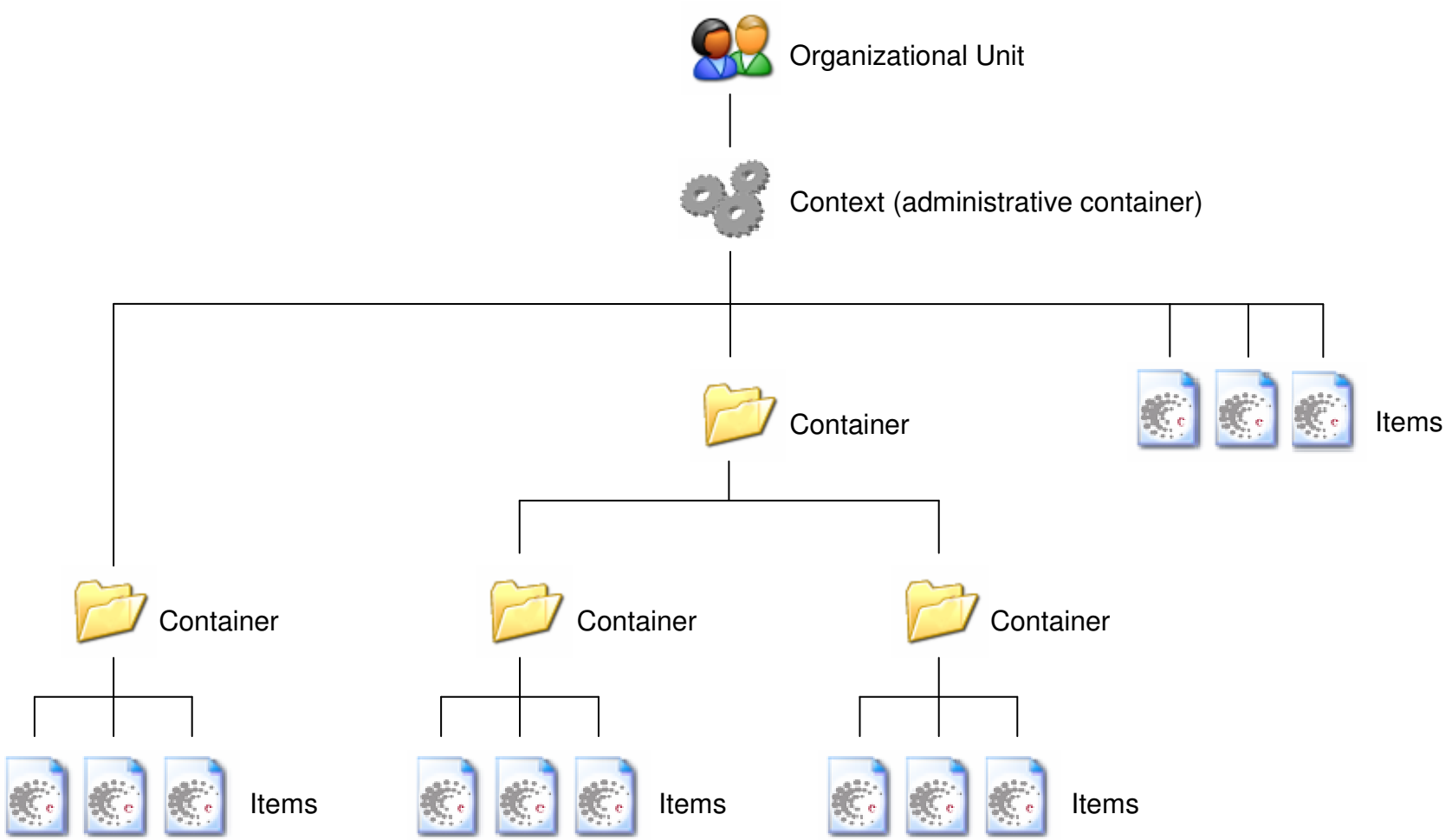
Challenges for an e-Research Infrastructure

- Maintain both data and publications
- Stable citations
- Integration with existing tools/instruments
- Re-use of data
- Accountability of research
- Collaboration across institutional bounds
- Mixture of open access and private material
- Support for (relational) databases

eSciDoc addresses these Requirements /1

- Maintain both data and publications
  - Flexible content models
  - Compound objects
  - Arbitrary metadata profiles

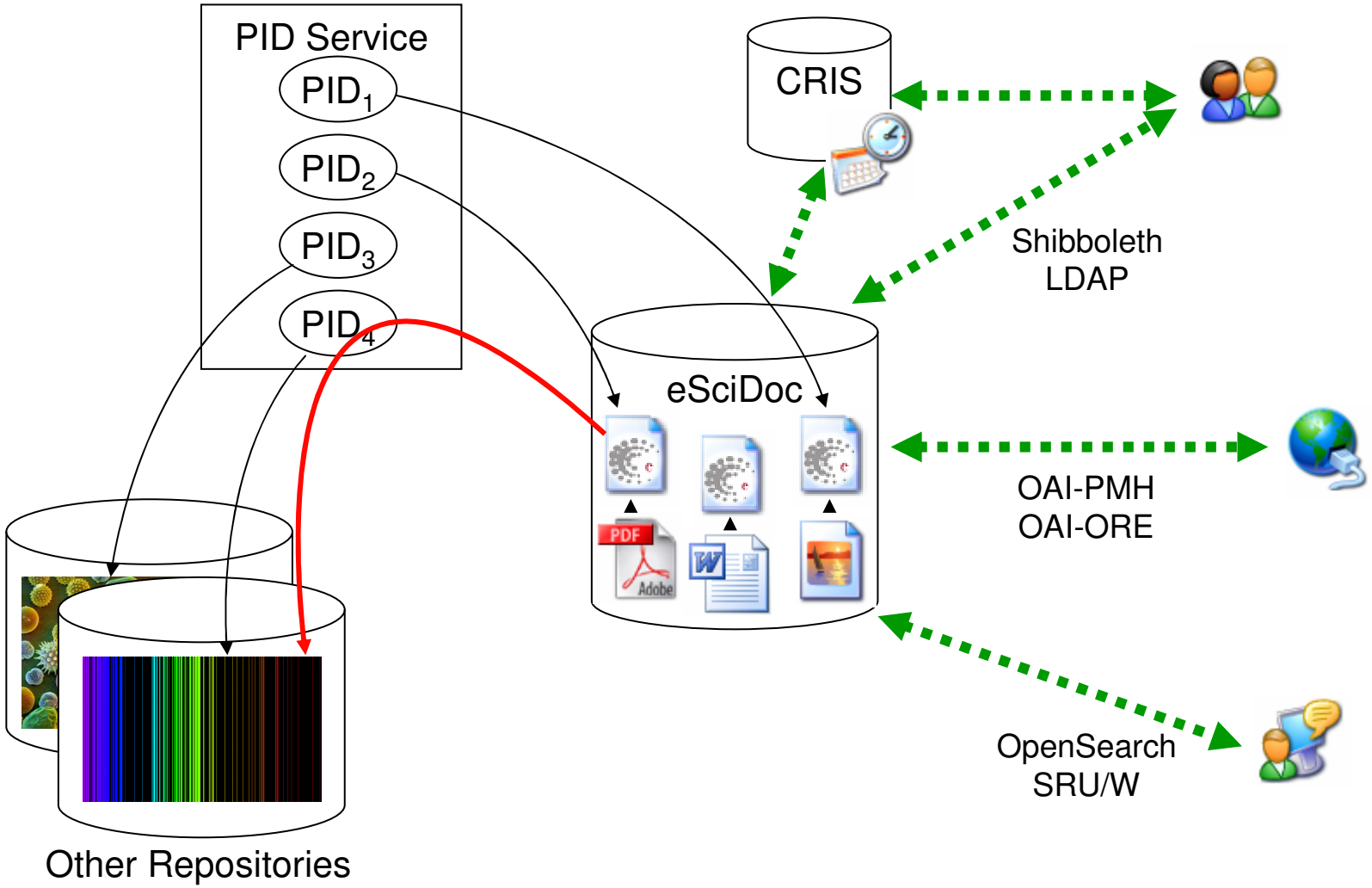
Conceptual Data Model



eSciDoc addresses these Requirements /1

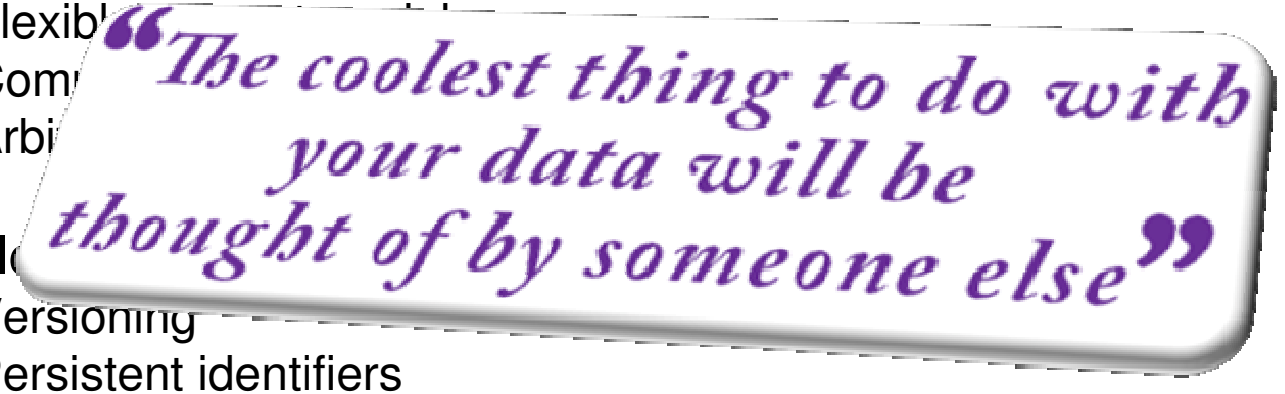
- Maintain both data and publications
  - Flexible content models
  - Compound objects
  - Arbitrary metadata profiles
- Stable citations
  - Versioning
  - Persistent identifiers

Identification, Linking, and Integration



eSciDoc addresses these Requirements /1

- Maintain both data and publications
  - Flexib
  - Com
  - Arbi
- Stabl
  - Versioning
  - Persistent identifiers



CRIG Logo

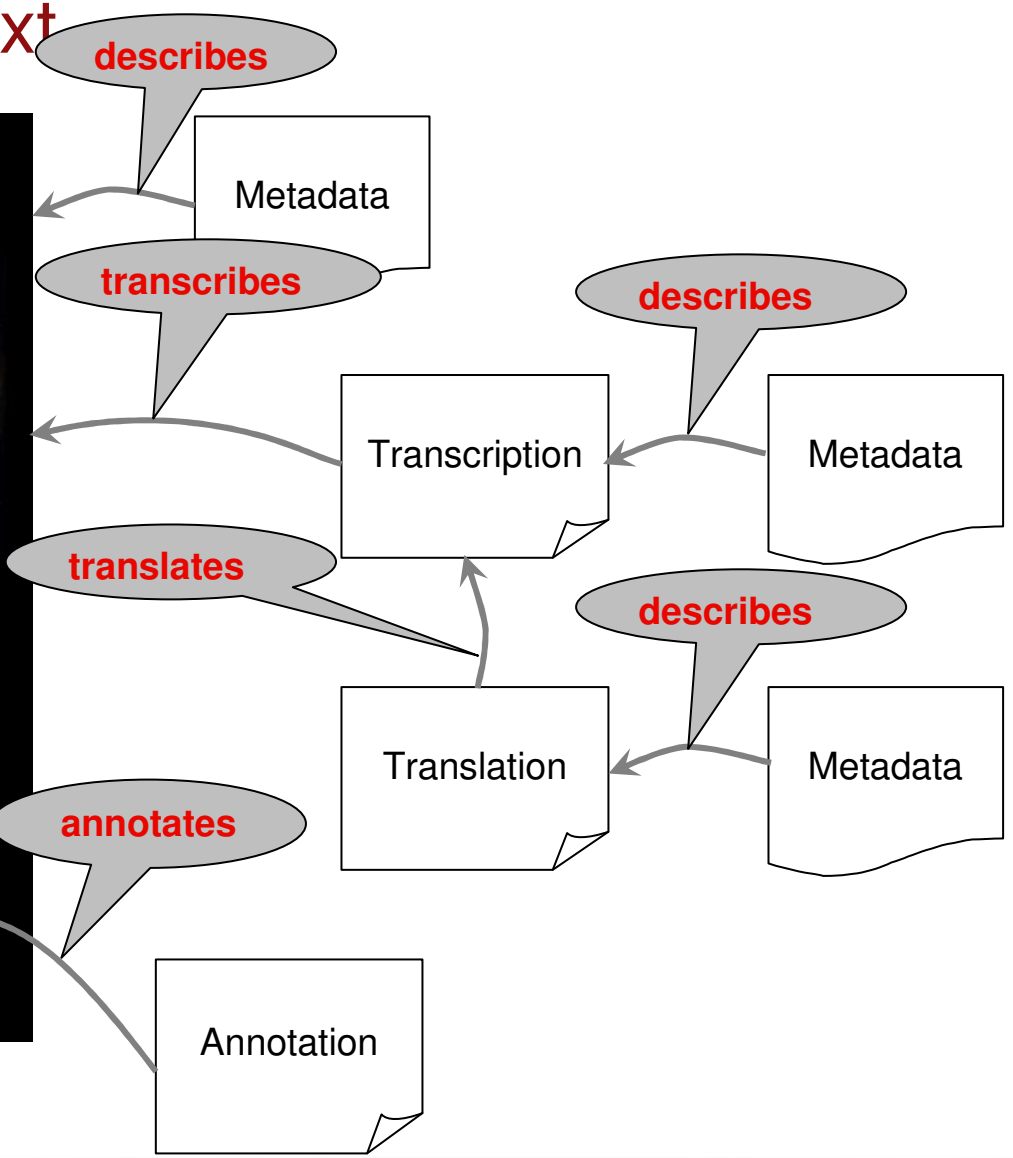
- Integration with existing tools/instruments
  - Open programming interfaces
  - Application-independent design
  - Extensibility (SOA)

eSciDoc addresses these Requirements /2

- Re-use of data
  - Provide context with data
  - Semantic linking / support for multiple ontologies
  - Support for multiple metadata records per object

Providing Context

© Institut Catholique, Paris, France



eSciDoc addresses these Requirements /2

- Re-use of data
  - Provide context with data
  - Semantic linking / support for multiple ontologies
  - Support for multiple metadata records per object
- Accountability of research
  - Preservation of data and methods
  - Stable links
  - Provenance metadata

eSciDoc addresses these Requirements /3

- Collaboration across institutional bounds
  - Distributed authentication (Shibboleth)
  - Virtual working groups
  - Annotations
- Mixture of open access and private material
  - fine-granular access rights

eSciDoc Infrastructure Release 1.1

- Group Policies
- Administrative Searches
- Setting States During Ingest
  Support for Japanese
- Unifying SOAP and REST Representations
- User Preference Handler
- Repository Reset Tool
- Ease of Use
  - Demo System
  - Documentation
  - Installer

eSciDoc Infrastructure Release 1.2

- Technical Metadata Extraction
- Organizational Unit Tree
- Searching in Albums
- OAI-PMH Sets
- Ontology Manager
- Searching for Tags
- Default Metadata

eSciDoc Infrastructure 2009-2011

- Simplified Policies for Role Handler
- Solr/Facetted Browsing and Proximities
- Ingest Tool
- Batch Updates of Items
- Interoperability
- Virus Checks
- Content Model Handler
- Table of Contents Handler
- Comment on Update
- Support for large XML corpora
- Automatic metadata extraction from full-texts
- File Format Manager
- Certifications / Process certificates (e.g., DINI, OAIS)
- Improved handling of large XML chunks in CRUD operations
- Service Authorization
- New resource Person
- Surrogate Objects

Vielen Dank!

Fragen?

Matthias Razum
matthias.razum@fiz-karlsruhe.de

http://www.escidoc.org/