

OGF23 Outreach Seminar: Digital Repositories

e-Infrastructure for Research Repositories

Andreas Aschenbrenner
(SUB Goettingen, TextGrid/D-Grid)



Open Grid Services Architecture, v1.5

(p.57) OGSA is **not prescriptive** on the **data model** used to implement an information service or the language used to query for information. Current systems broadly fall into those that are based on XML and XPath/XQuery query languages (e.g., Globus MDS4) and those that use the relational model and the SQL query language (e.g., R-GMA).

Open Grid Services Architecture, v1.5

(p.57) OGSA is not prescriptive on the data model used to implement an information service or the language used to query for information. Current systems broadly fall into those that are based on

- XML and XPath/XQuery query languages (e.g., Globus MDS4)
- those that use the relational model and the SQL query language (e.g., R-GMA), and
- **object-centric systems** with related content models and metadata management (e.g. ...)

software
versioning

registry

package
mangement

institutional repository
trusted repository

job
scheduling

MyProxy
credentials

...

...



institutional repository

institutional - disseminate



institutional repository

OpenDOAR - disseminate

| Any Software |
|---------------------|
| DigiTool (5) |
| Diva-Portal (14) |
| dLibra (10) |
| DLXS (1) |
| DOKS (5) |
| Drupal (1) |
| DSpace (302) |
| Dynawen (1) |
| eDoc (6) |
| EPrints (224) |
| ePub (4) |
| ETD-db (18) |

**1131 registeres
repositories**



[Hauptseite](#)

CERN Document Server

Mehr als **800,000** Datensätze vorhanden, **360,000** volltext Dokumente in der Teilchenphysik und in den verbundenen Bereichen einbezogen. Die Datenbank enthielt Preprinte, Artikel, Bücher, Zeitschriften, Photographien und viel mehr.

Durchsuche 965,035 Datensätze nach:

alle Felder ▼

Suchen

Durchblättern

[Suchtipps](#) :: [Erweiterte Suche](#)

Einschränken nach Sammlungen:

- Articles & Preprints** (791,084)
 - [Published Articles](#) (300,949) [Preprints](#) (430,919) [Theses](#) (15,774)
 - [Reports](#) (5,569) [CERN Internal Notes](#) (12,003) [Committee Documents](#) (27,908)
- Books & Proceedings** (65,891)
 - [Books](#) (41,506) [Proceedings](#) (16,459) [Standards](#) (7,928)
- Presentations & Talks** (15,867)
 - [Conference Announcements](#) (14,660) [Academic Training Lectures](#) (564)
 - [Summer Student Lectures](#) (437) [General Talks](#) (133) [Videotapes](#) (296)
- Periodicals & Progress Reports** (3,244)

Hinweis auf:

- CERN Articles & Preprints** (96,182)
 - [CERN Published Articles](#) (49,540) [CERN Preprints](#) (14,783) [CERN Theses](#) (2,902)
 - [CERN Reports](#) (1,161) [Committee Documents](#) (27,908)
- CERN Series** (14,184)
 - [CERN Yellow Reports](#) (1,119) [CERN-TH](#) (12,095) [Academic Training Lectures](#) (564)
 - [Summer Student Lectures](#) (437) [General Talks](#) (133)
- CERN Departments** (67,294)
 - [Accelerator Technology \(AT\)](#) (4,990) [Accelerators & Beams \(AB\)](#) (16,079) [Finance \(FI\)](#) (899) [Human Resources \(HR\)](#) (28) [Information Technology \(IT\)](#) (3,732) [Physics \(PH\)](#) (36,553) [Secretariat-General \(SG\)](#) (8,479) [Technical Support \(TS\)](#) (1,240)



- Beschreiben
- Beschreiben
- Beschreiben
- Hochladen
- Überprüfen
- Lizenz
- Lizenz
- Fertig

:: Suche

- » Suche
- » Erweiterte Suche
- » Thematische Suche
 - Geowissenschaften
 - Bergbau
 - Geographie/Kartogr.
 - Thematische Karten

:: Auflistung nach

- » Bereichen & Sammlungen
- » Titel
- » Autoren
- » Datum

:: Mein GEO-LEO *e-docs*

Angemeldet als
 quast@sub.uni-goe...
 (Abmelden)

- » Mein GEO-LEO *e-docs*
- » Abonnement Neuerscheinungen
- » Benutzerprofil bearbeiten

Veröffentlichen: Beschreibung des Dokuments

Bitte geben Sie alle notwendigen Informationen zu Ihrer Veröffentlichung ein. In den meisten Browsern können Sie r der Tabulator-Taste von Feld zu Feld springen. ([Mehr Hilfe](#))

Geben Sie bitte einen oder mehrere Autoren ein.

Nachname

Vorname(n)

z. B. **Mustermann**

z. B. **Michal Max**

Autor(en)

Quast

Andres

Weitere eintragen

Bitte geben Sie den Titel ein.

Titel

Test

Geben Sie bitte die Kurzfassung ein.

Kurzfassung

Testdokument für den Workflow

Vervollständigen Sie hier bitte das Zitat mit den entsprechenden Seitenzahlen

Orginalzitat

In: Philipp, S., Leiss, B, Vollbrecht, A., Tanner, D. & Gudmunds:



Beschreiben Beschreiben Beschreiben Hochladen Überprüfen Lizenz Lizenz Fertig

:: Suche

- » Suche
- » Erweiterte Suche
- » Thematische Suche
 - Geowissenschaften
 - Bergbau
 - Geographie/Kartogr.
 - Thematische Karten

:: Auflistung nach

- » Bereichen & Sammlungen

Veröffentlichen: Beschreibung des Dokuments

Bitte machen Sie weitere Angaben zu Ihrer Veröffentlichung ([Mehr Hilfe](#))

Ordnen Sie Ihre Arbeit bitte einem oder mehreren Themengebieten der GOK zu.

Göttinger Online Klassifikation (GOK)

TM 000

GOK

GOK

Weitere eintragen

Geben Sie bitte Schlagworte ein

Schlagworte

Weitere eintragen

http://134.76.163.148:8080 - Classification System - Mozilla Firefox

Filter:

Trim...

Fill in a term to trim the concept tree

Classification system

- ⊕ [V](#) (Geowissenschaften)
 - ⊕ [TM 000](#) (Geophysik)
 - ⊕ [TT 000](#) (Meteorologie und Klimatologie)
 - ⊕ [UA 000](#) (Hydrologie Allgemein)
 - ⊕ [UD 000](#) (Regionale Hydrologie)
 - ⊕ [UG 000](#) (Ozeanologie)
 - ⊕ [UN 000](#) (Geodäsie)
 - ⊕ [V 000](#) (Geologie)
 - ⊕ [VF 000](#) (Mineralogie)
 - ⊕ [VJ 000](#) (Geochemie)
 - ⊕ [VK 000](#) (Petrologie)
 - ⊕ [VN 000](#) (Bodenkunde)
 - ⊕ [VQ 000](#) (Lagerstättenkunde)
 - ⊕ [VU 000](#) (Gesamte Paläontologie)
- ⊕ [Q](#) (Geographie)
 - ⊕ [QA 000](#) (Geographie)
 - ⊕ [QG 000](#) (Regionale Geographie)
 - ⊕ [QGA 000](#) (Europa)

Weiter >

Abbrechen/Speichern

edback

Beschreiben

Beschreiben

Beschreiben

Hochladen

Überprüfen


Lizenz

Lizenz

Fertig

Veröffentlichen: Benutzen der Creative Commons Lizenz

Um Ihre Veröffentlichung unter der Creative Commons Lizenz zu veröffentlichen, folgen Sie bitte den unten aufgeführten Anweisungen. Ihnen wird danach nochmal die Möglichkeit gegeben Ihre Entscheidungen zu überprüfen und zu korrigieren. Wenn Sie die Creative Commons Lizenz auslassen wollen klicken Sie auf "Creative Commons Lizenz auslassen".



provided by [Creative Commons](#)

Sie haben ein kreativen Inhalt geschaffen, auf das Sie stolz sind. Jetzt ist es an der Zeit, nach kreativen Wegen zu suchen, es anderen zugänglich zu machen.

Die Lizenzverträge von Creative Commons helfen Ihnen, Ihren Inhalt der Allgemeinheit zugänglich zu machen, ohne dass Sie dabei auf Ihr Urheberrecht gänzlich verzichten: Andere können Ihren Inhalt unter bestimmten Bedingungen vervielfältigen und verbreiten, vorausgesetzt, [Sie werden als Rechtsinhaber genannt](#). Die einzelnen Bedingungen bestimmen Sie selbst, und diese Seite hilft Ihnen dabei. Wenn Sie Ihren Inhalt unter Verzicht auf alle nach deutschem Recht bestehenden Rechte der Allgemeinheit zur Verfügung stellen wollen, dann wählen Sie die [public domain](#).

Kommerzielle Verwertung erlauben? ([Info](#) ⓘ)

Ja
 Nein

Bearbeitung Ihres Inhalts zulassen? ([Info](#) ⓘ)

Ja
 Ja, solange andere die gleichen Lizenzbedingungen verwenden ([Info](#) ⓘ)
 Nein

Bitte Beachten: Um einen Inhalt lizenzieren zu können, müssen Sie Inhaber aller an dem Inhalt bestehender Rechte sein.

< Zurück Creative Commons Lizenz auslassen > Abbrechen/Speichern

institutional repository

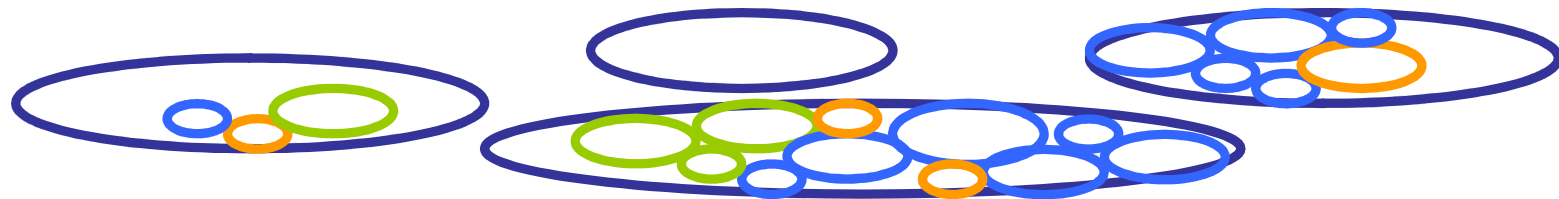
institutional - disseminate



open *

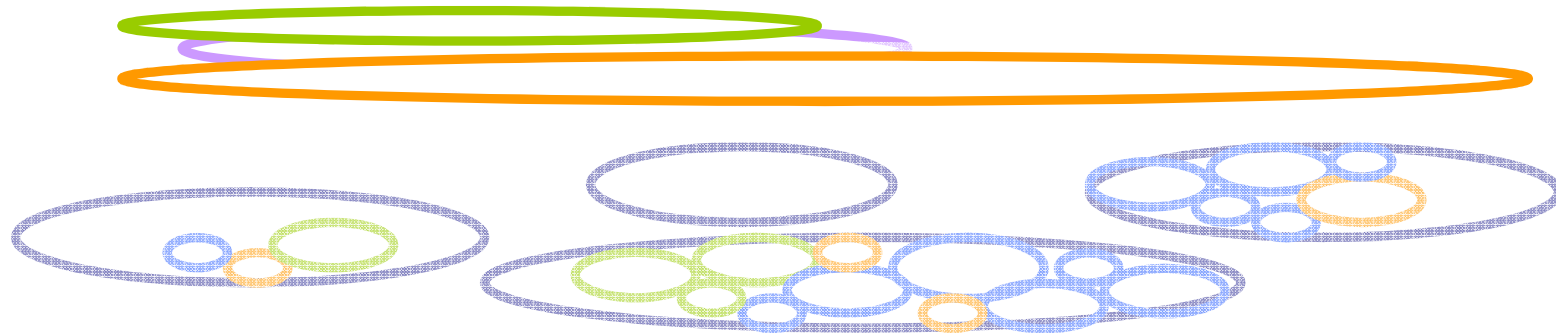
departmental repository

institutional - disseminate
departmental - promote



thematic repository

- institutional - disseminate
- departmental - promote
- thematic - communicate



**publications
open access**

**shared
infrastructure**

institutional - disseminate

departmental - promote

preservation

**shared
collections**

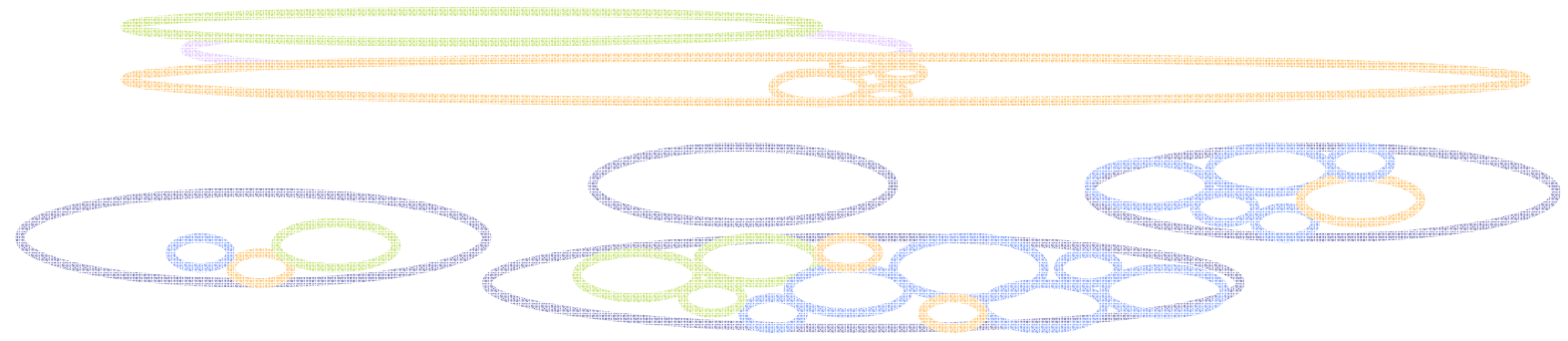
**primary
data**

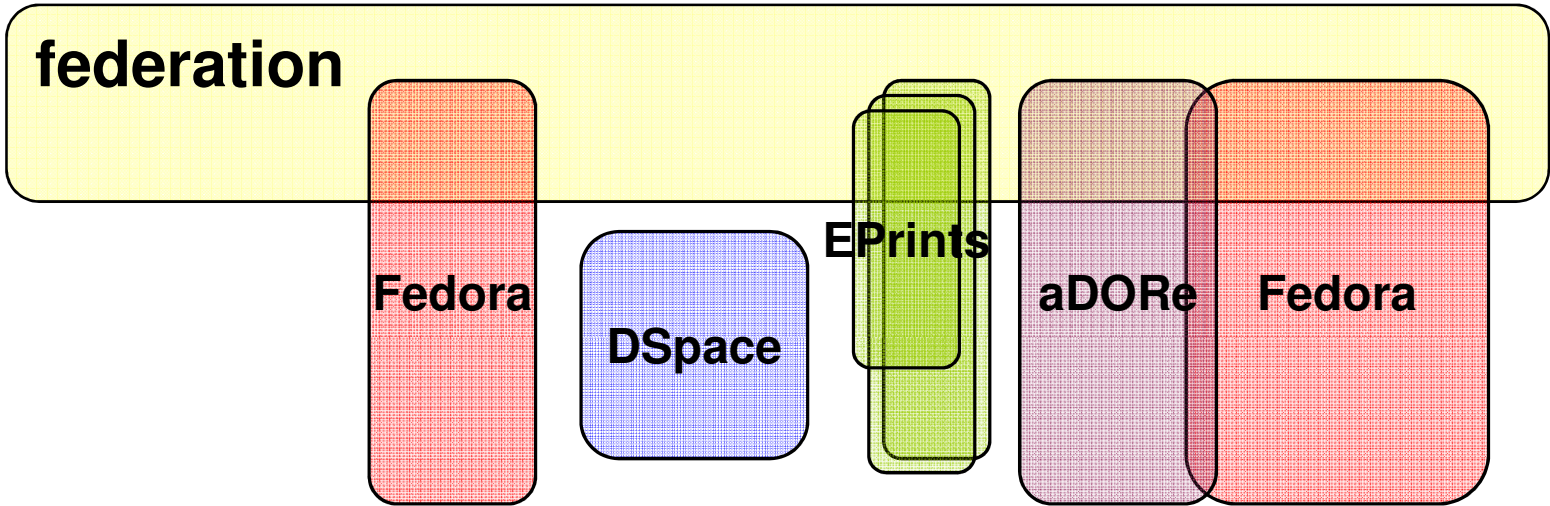
thematic - communicate

ad hoc / temporary - collaborate

**cross-repository
management
services**

**scientific
workflows**





Open Archives Initiative



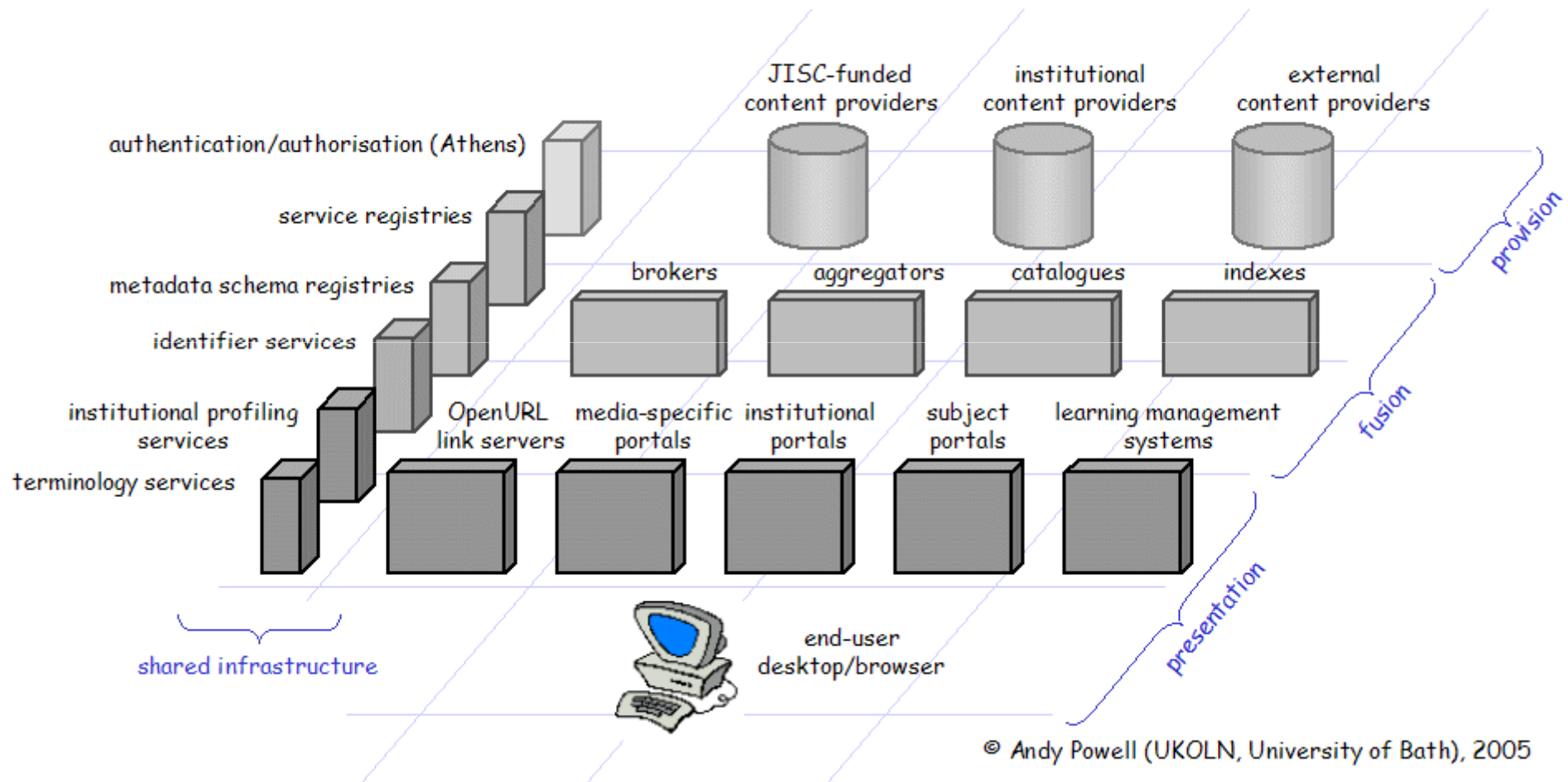
- **OAI-PMH**
Protocol for Metadata Harvesting
- **OAIster:** (Mai 2008)
15,622,344 records from 950 contributors
→ AstroGrid registry OAI-PMH-based
→ usage for content transfer

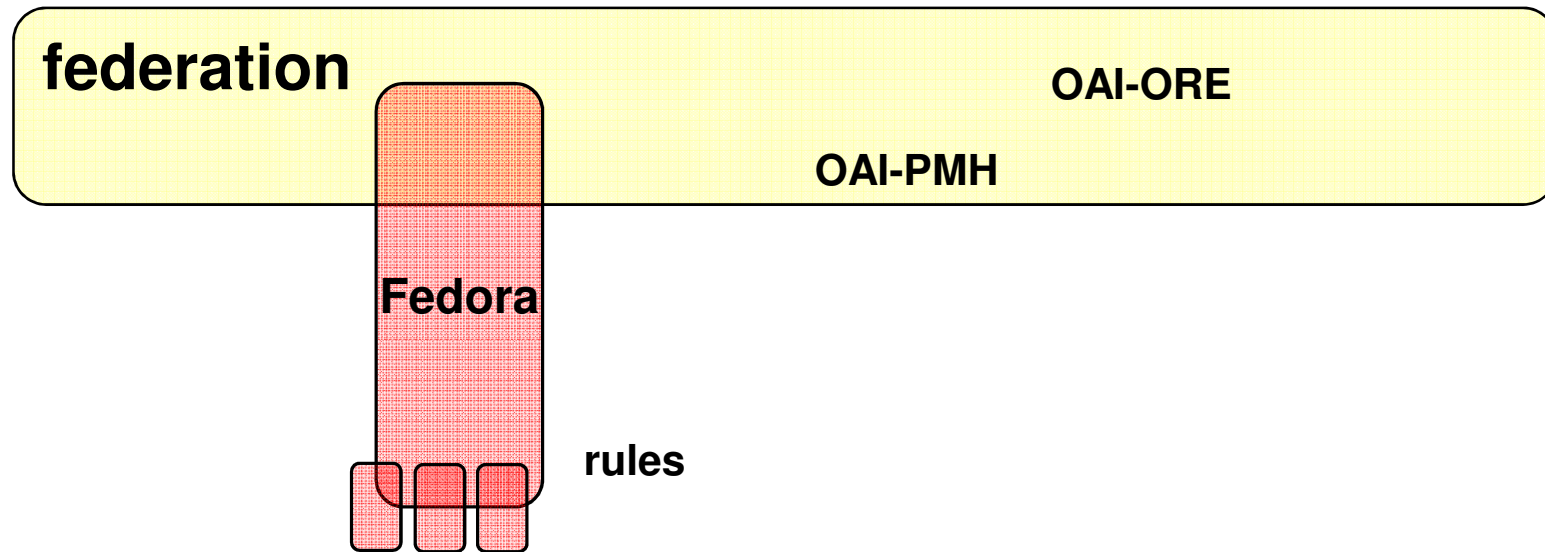
Open Archives Initiative

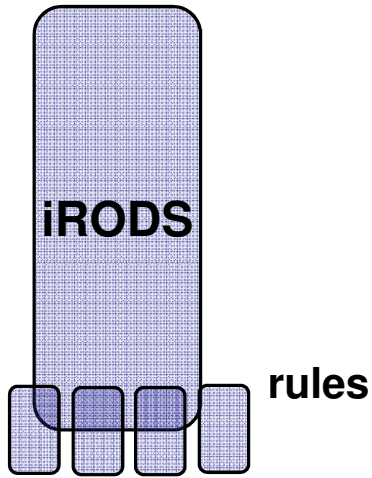
- **OAI-ORE** - Object Reuse and Exchange
- Object Models
- in RDF
- Atom-serialised
 - repository federation
 - object replication

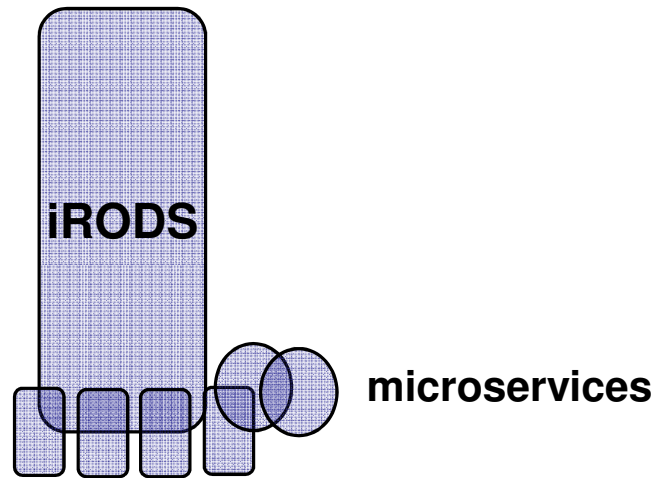


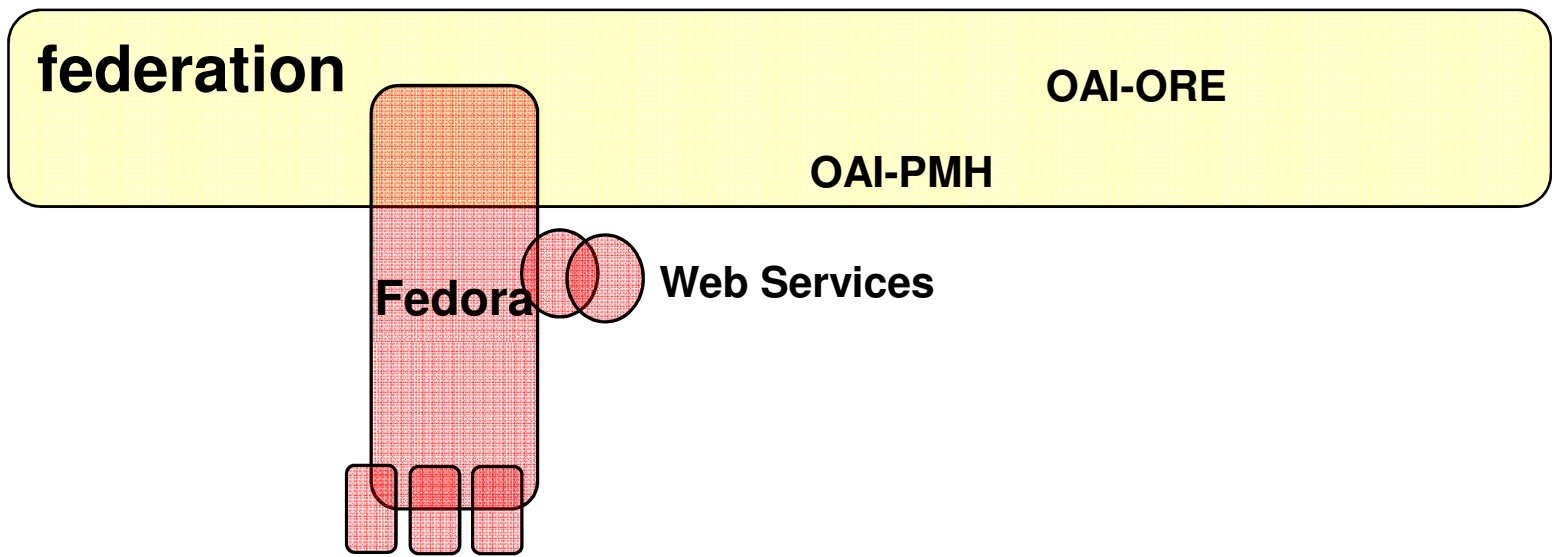
JISC information architecture

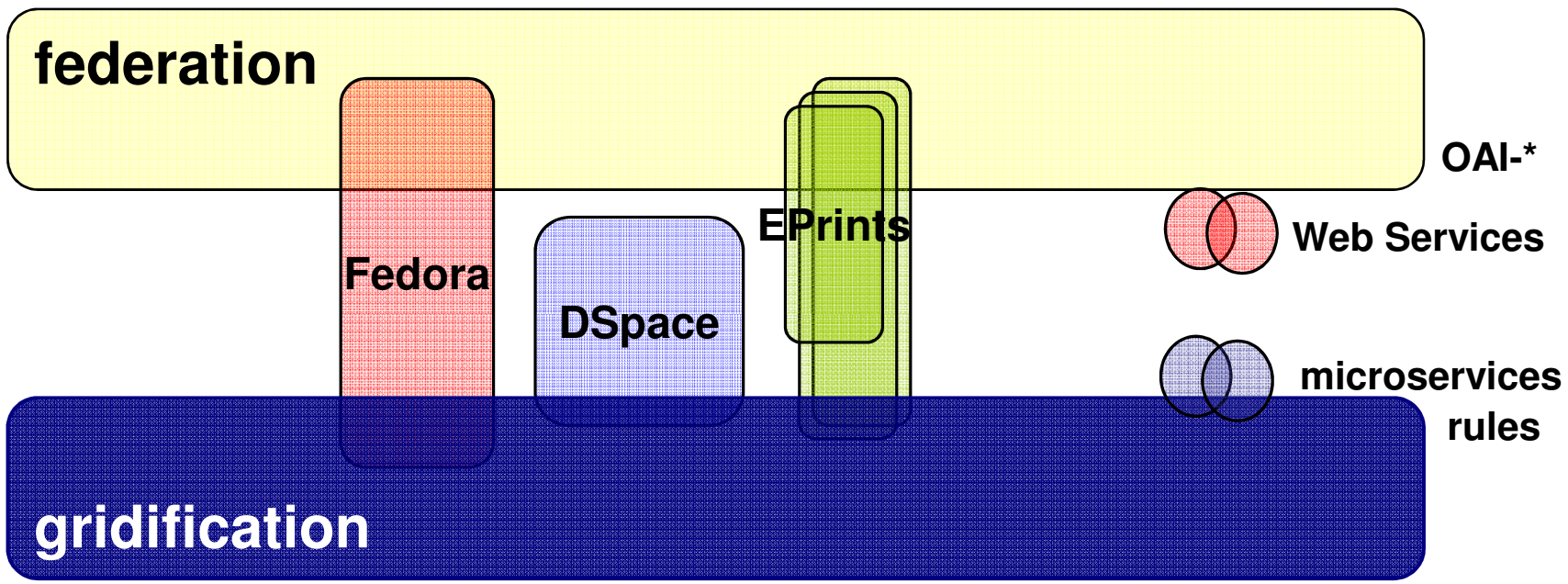












resources?!



GWDG Grid Resources

compute
storage

→ semantic services

→ digital objects ...



SOKU -
service oriented
knowledge utilities

Special Session: Repositories

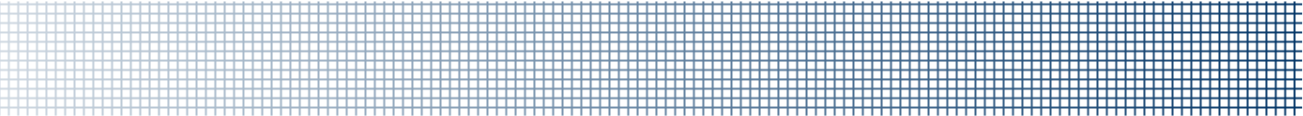
4th IEEE International Conference on e-Science
December 7–12, 2008

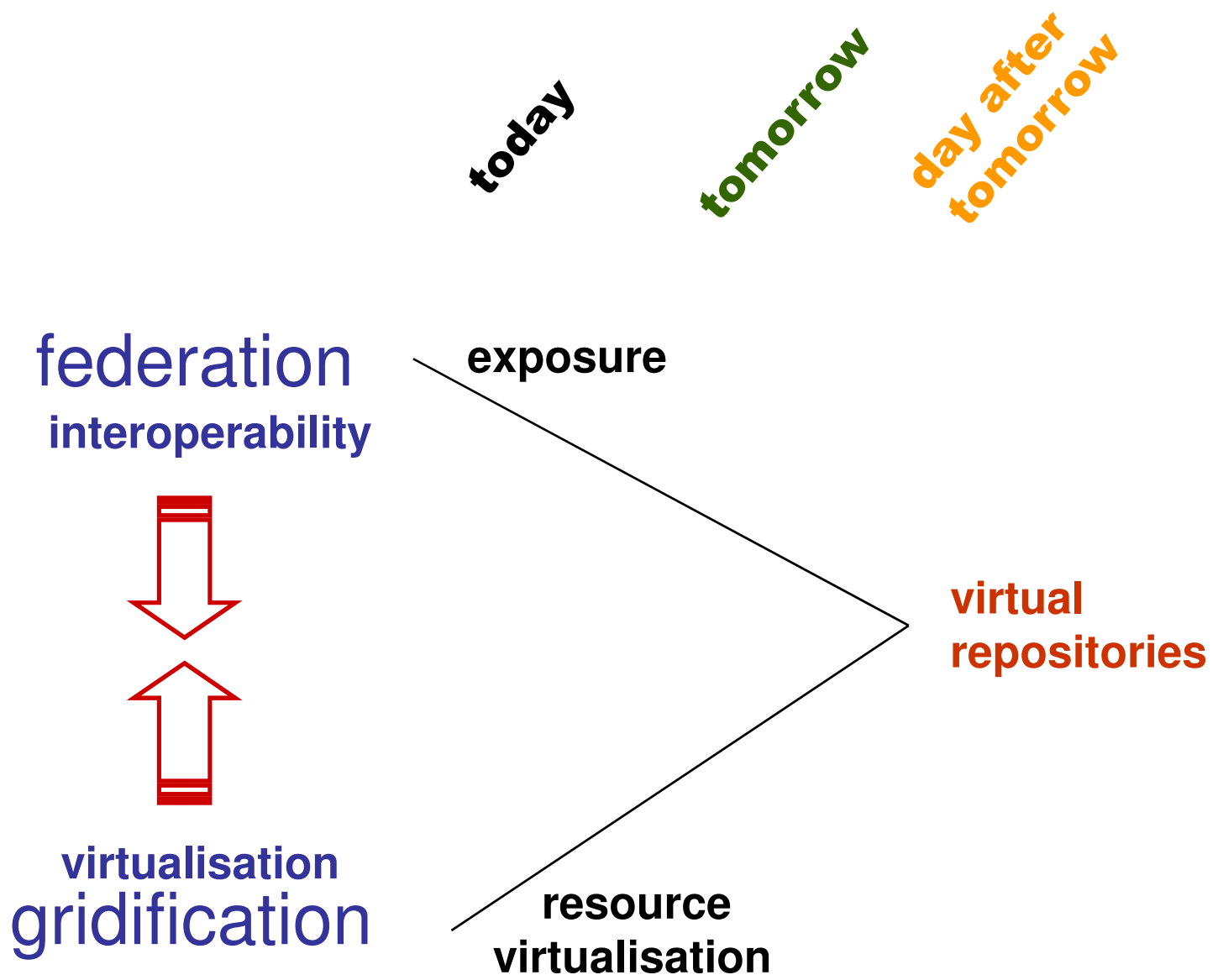
University Place Conference Center & Hotel,
Indianapolis, Indiana, USA

thanks ...
looking forward to your feedback ...

Andreas Aschenbrenner
aschenbrenner@sub.uni-goettingen.de







today

tomorrow

**day after
tomorrow**

metadata exchange, e.g. OAI-PMH
object exchange, e.g. OAI-ORE

dedicated utilities, e.g. Pronom

targeted exchange, shared collection management

ad hoc interoperability, e.g. inter-collections
nested repositories

on-demand repository services

distributed storage and object management

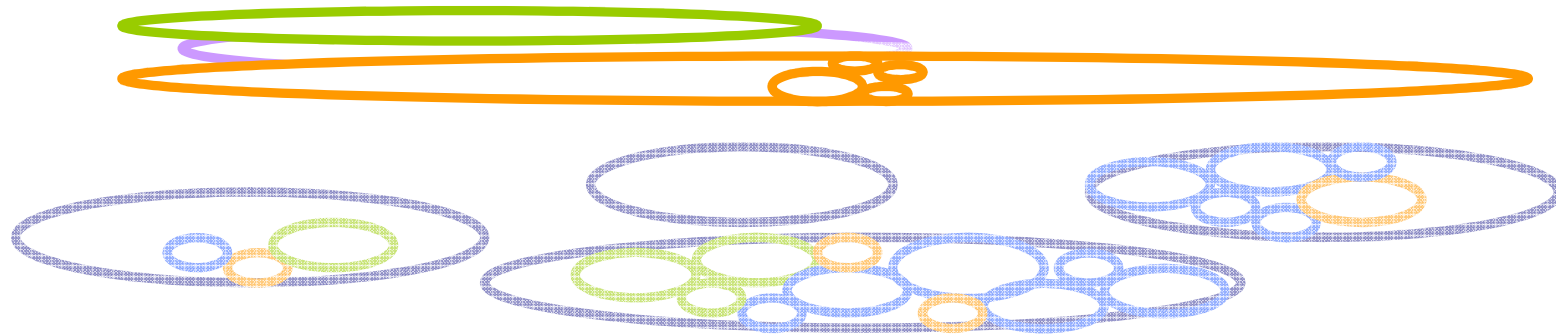
farming out auxiliary jobs, e.g. format conversion

handling raw data

virtualisation of resources (storage + compute)

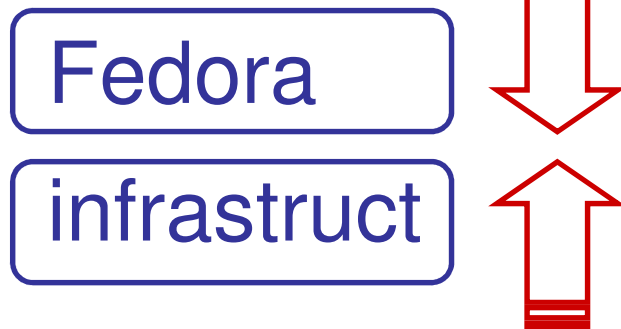
temporary / ad hoc

- institutional - disseminate
- departmental - promote
- thematic - communicate
- ad hoc / temporary - collaborate



technical environment

- shared infrastructure
 - to raise stability and lower costs (software and orgware)
- access and exposure
 - performance and scalability
 - interoperability, context and reference networks
- preservation
 - distributed, redundant storage
 - criteria for trusted repositories
- embed in scientific workflow
 - persistent identification
 - service networks and workflows



opportunities

- JISC RepositoryNet:
2006 JISC committed £15 million towards Digital
Repositories and Preservation activity



Digital Preservation

- deposit: no control over object creation
 - metadata: as much as necessary, as simple as possible
- organisational issue,
rather than technological one
- towards "trusted" digital repositories (2000)

DSpace format policy (2004)

- Supported formats
 - e.g. TIFF, SGML/XML, PDF...
- Known/unsupported formats
 - e.g. Microsoft Word, PowerPoint (common)...
 - e.g. Lotus 1-2-3, Visicalc, WordStar (less common)...
- Unknown/unsupported
 - Highly complex and rare formats
 - e.g. one-of-a-kind software programs...



preservation environment

- PRONOM - file format registry
UK National Archives
- GDFR - global digital format registry
- tools for format validation (JHove, DROID)
for migration, technology watch,
metadata crosswalks, etc
- service-based preservation environments:
PANIC (Monash), PLANETS (FP6), ...

outlook

- shared infrastructure
 - storage and compute on demand
- shared services
 - preservation
- shared collections / interoperability
 - institutional x thematic x personal
- linking environments
 - publications ↔ scientific data ↔ multimedia objects
- embedding repositories in scientific workflow