

Ein Java Repository für digitalen Content in Bibliotheken

9. Sun Summit Bibliotheken
13.-14. November 2007, Deutsche Nationalbibliothek

Christof Mainberger, BSZ

Retrodigitalisierung

Bild-Archiv **BLOG** Web-Site

Hochschulschriftenserver **Mediensever**

Kataloganreicherung **Lehr- Lernplattform**

Institutionelles Repository **Langzeitarchivierung**

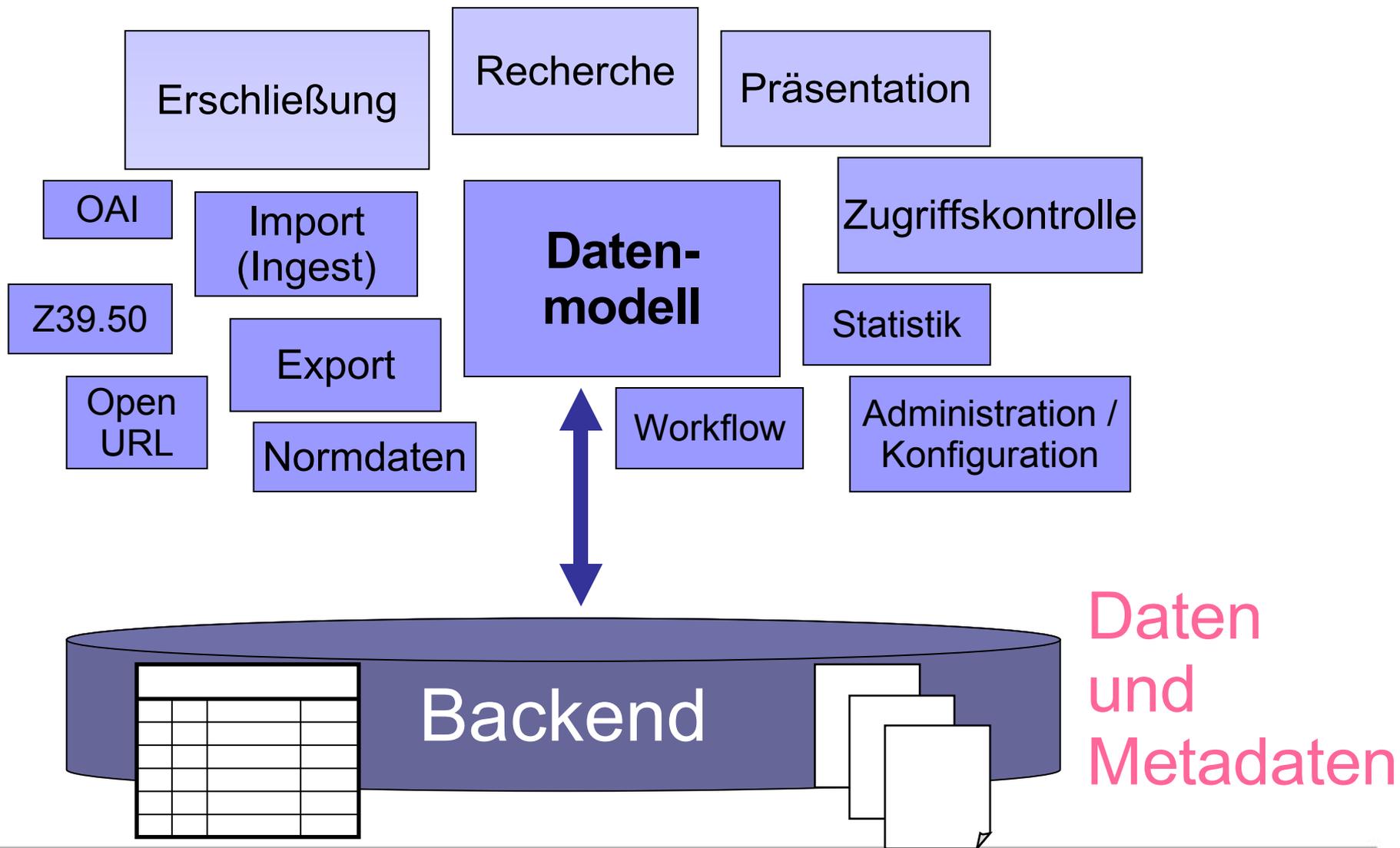
Elektronischer Semesterapparat Back-Files

Virtuelle Auskunft **Elektronische Zeitschrift**

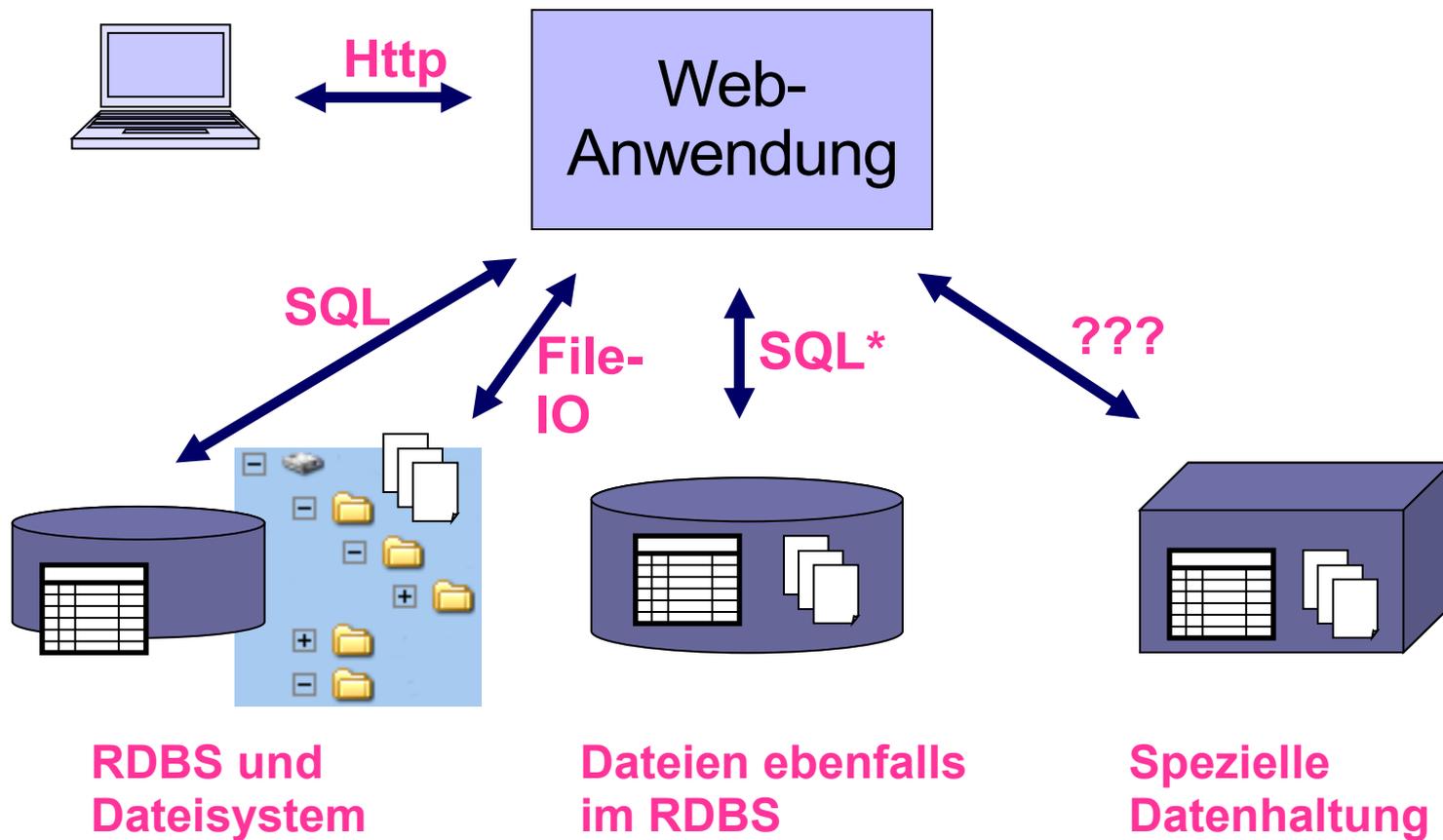
Multimedia **Wiki**

Museums-Dokumentation

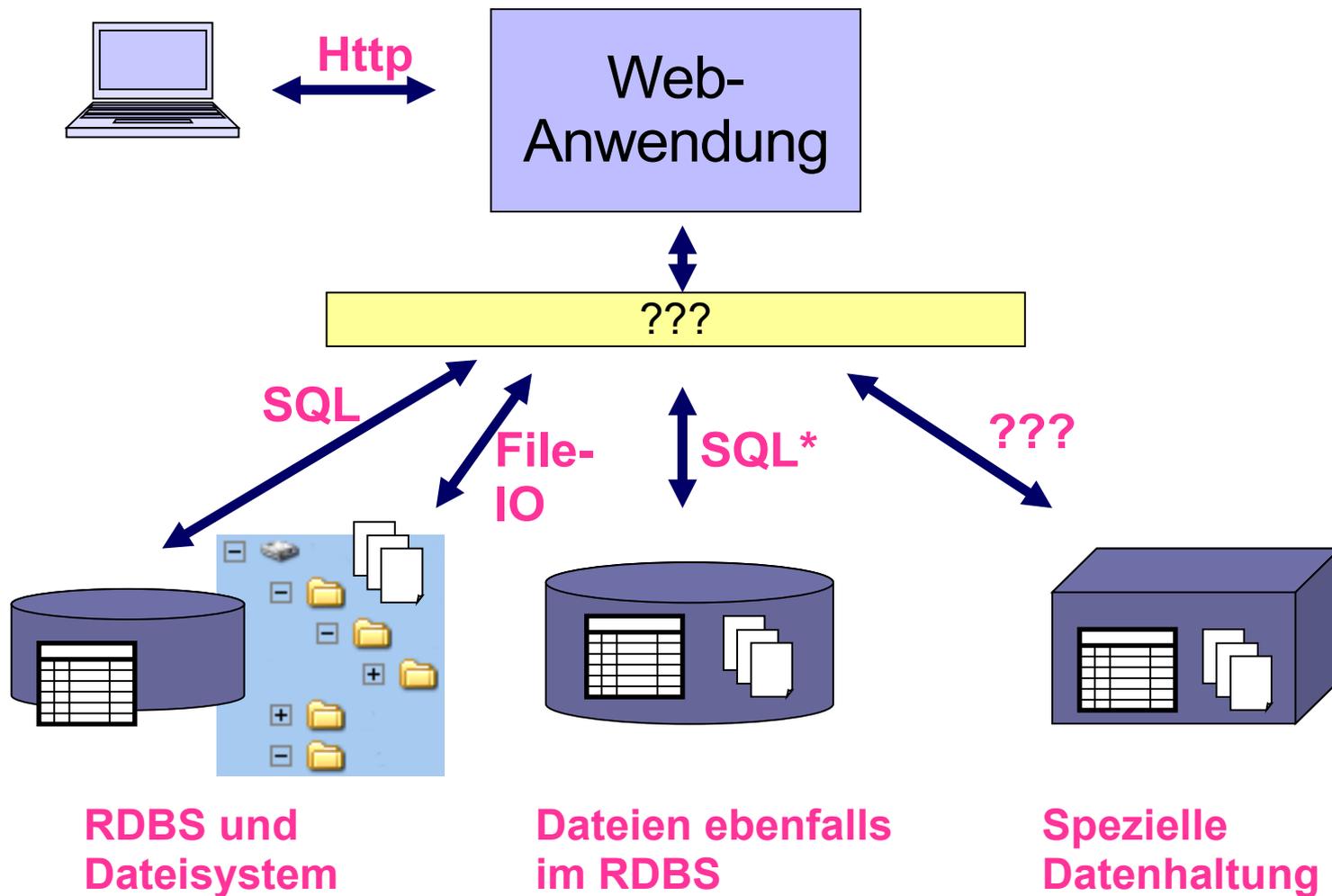
... und benötigen dazu Werkzeuge!



Varianten des Backends von Repositorien



Unterschiedliche Schnittstellen zwischen Frontend und Backends



EINE Schnittstelle, die die Varianten der Datenhaltung abstrahiert.

- Beliebige Medienobjekte, beliebige Metadaten
- Verbindung von Medienobjekt und Metadaten
- Hierarchische Sammlungsstruktur
- Unterstützung von XML, ...
- feingranulare Rechteverwaltung
- Versionierung
- Volltextsuche, ...
- leistungsfähige Programmierschnittstelle (API)

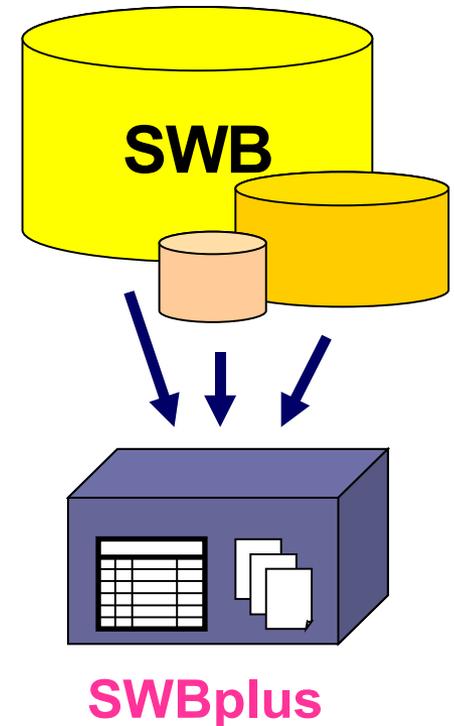
- Standardisierung, breite Unterstützung und erschwingliche Implementierungen



- Resultat des **JCP170-Prozesses**, Juli 2005
- Apache, BEA, Day Software, Fujitsu, HP, IBM, Novell, Oracle, SAP, Sun Microsystems, u.v.a.m.
- JCR definiert ein abstraktes **hierarchisches Datenmodell** mit **XPath** als Anfragesprache
- **XML-Protokoll** für Im- und Export von Datenbeständen
- **Transaktionen, Versionierung, Observation, Locking und SQL**
- Open Source Implementierung:  **Jackrabbit**
Apache Jackrabbit
- wird fortgesetzt im JCP283:
Zugriffskontrolle, Mehrfachhierarchien, ...

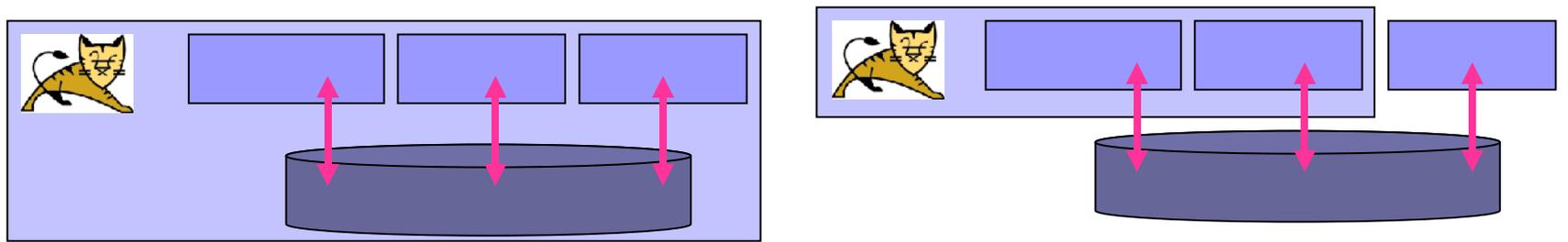
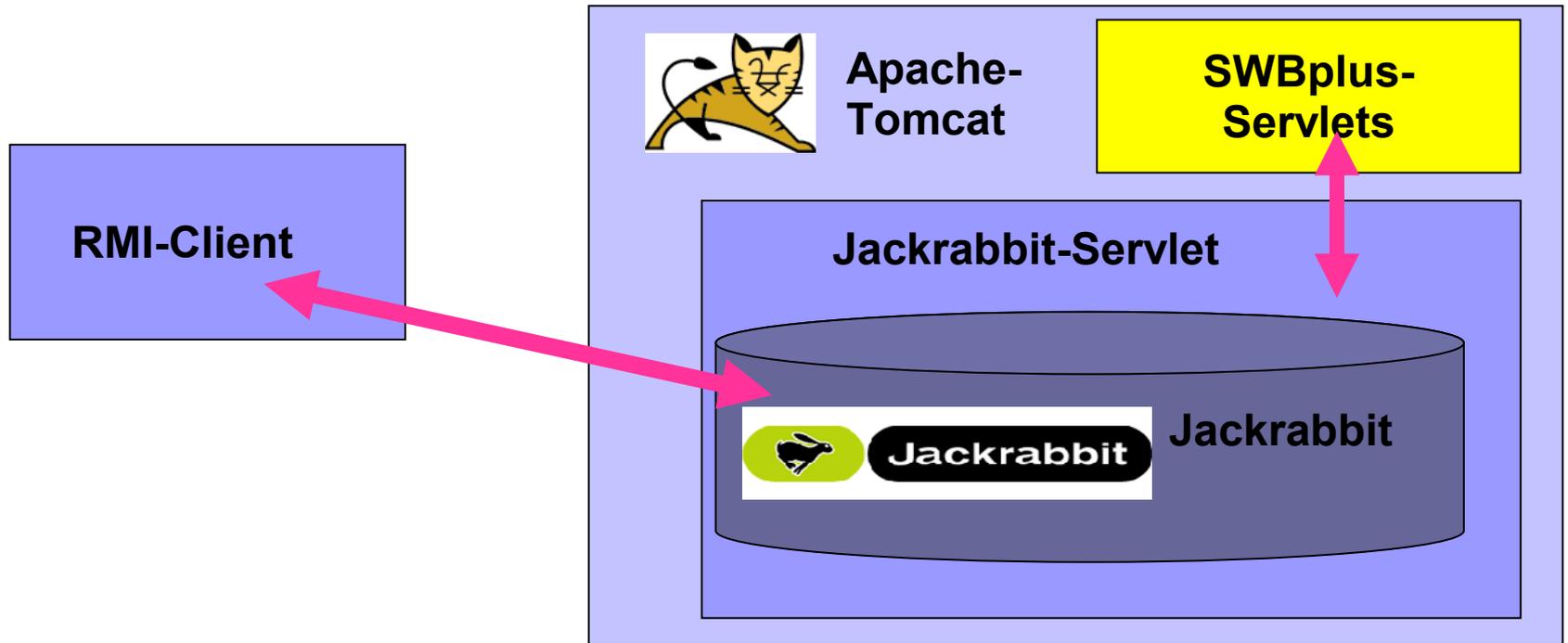
SWBplus verwaltet die Kataloganreicherungen des BSZ:

- Inhaltsverzeichnisse, Abstracts, Klappentexte, etc.
- derzeit ca. 155.000 Anreicherungen aus 12 Bibliotheken und 14 Verlagen
- Dezentrales Scannen und zentrale OCR
- Keine „eigene Rechercheoberfläche“; Abruf der Anreicherungen als Original, Text oder PDF über URLs
- Export der Anreicherungen in lokale OPACs



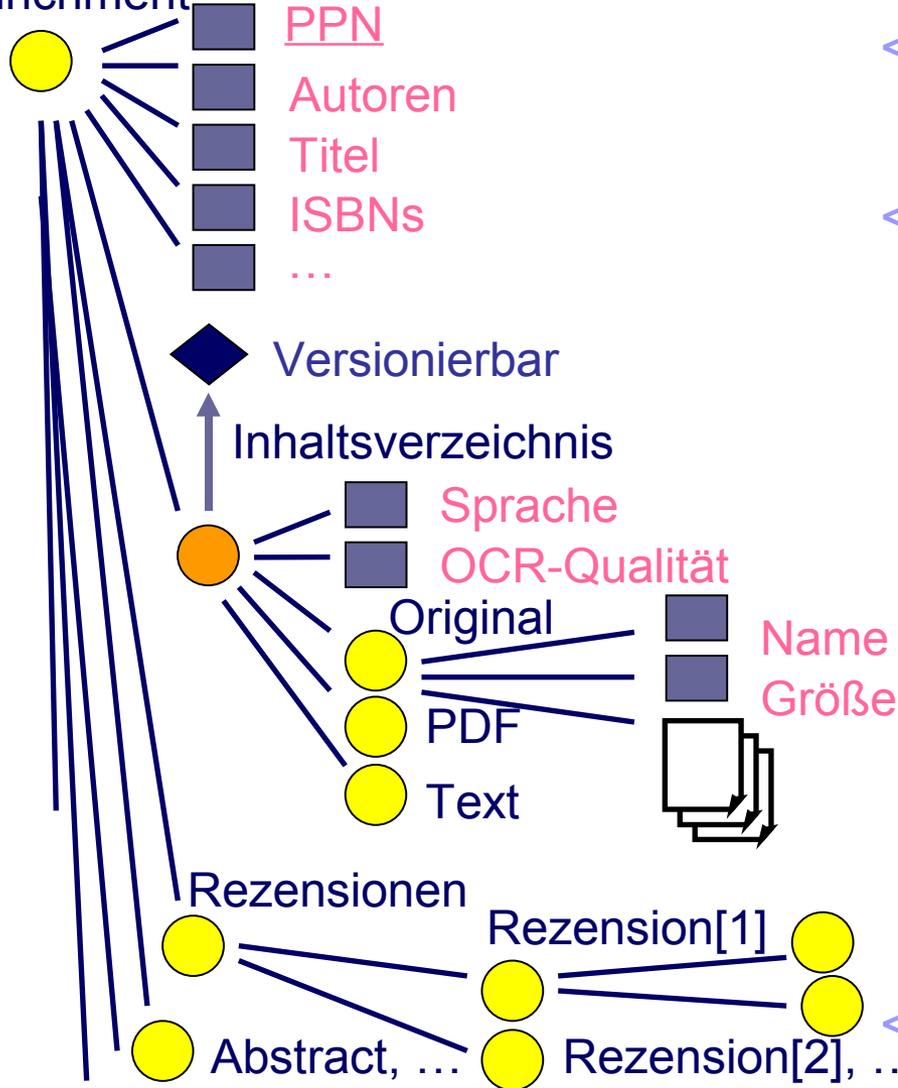
Anfang 2007 wurde für SWBplus von ORACLE auf eine APACHE-Jackrabbit Installation migriert.

Architektur des SWBplus-Repository



(Vereinfachtes) Datenmodell von SWBplus

Enrichment



```

<enrichment id="258311665" type="rez">
  <authors><author>Romain..</author></authors>
  <titles><title>Catalogue des...</title></titles>
  <publishers><publisher>...
</publisher></publishers>
  <years><year>2006</year></years>
  <places><place>Diet...</place></places>
  <isbns> <isbn>3-85...</isbn></isbns>
  <expression type="inh" label="Inhaltsverzeichnis">
    <manifestations>
      <lang>fre</lang>
      <ocr-quality>orig</ocr-quality>
      <manifestation type="orig">
        <filename>/orig/bsz1234.pdf</filename>
      </manifestation>
      <manifestation type="ocr">
        <filename>/ocr/bsz1234.pdf</filename>
      </manifestation>
      <manifestation type="txt">
        <filename>result/txt/bsz1234.txt</filename>
      </manifestation>
    </manifestations>
  </expression>
</enrichment>
  
```

Beispiel:

<http://swbplus.bsz-bw.de/bsz08762818xinh.pdf>

```
//enrichment[swbplus:ppn= '08762818x']  
/swbplus:inhaltsverzeichnis/swbplus/PDF
```

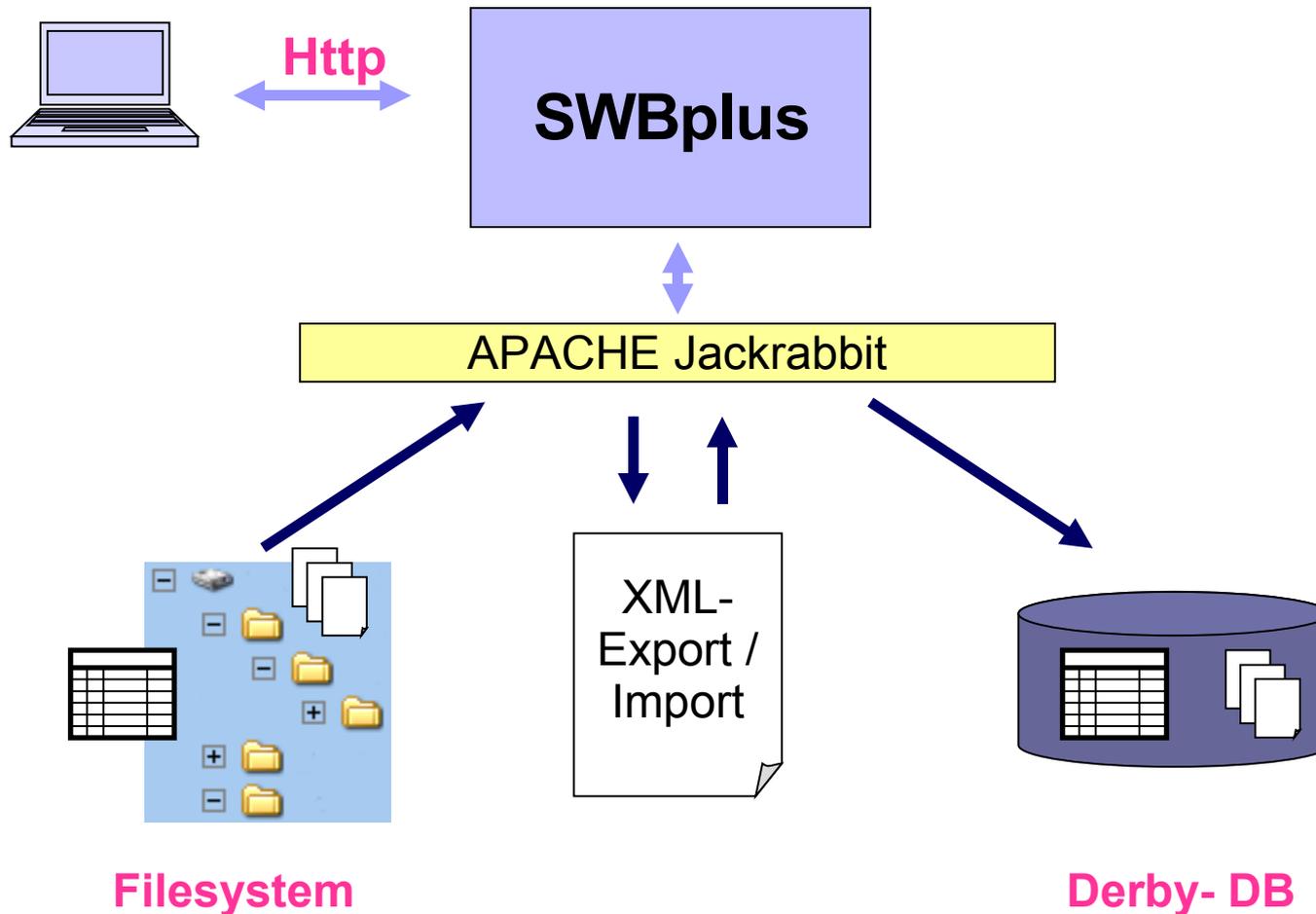
oder

```
SELECT ./swbplus:inhaltsverzeichnis/swbplus:PDF  
FROM swbplus:enrichment  
WHERE swbplus:ppn= '08762818x'
```

oder

```
getEnrichment(„08762818x“).getInh().getPdf()
```

mit Hilfe einer Abbildung auf ein JAVA-Objektmodell



Migration der Daten in eine (beliebige) SQL-Datenbank

It works!

Und nun:

- Anwendung des JCR für ‚kleinere‘ Anwendungen
- Warten auf JCP283
- Warten auf „industrielle“ Implementierungen:
IBM, Oracle, ..., SUN,... ?

Vielen Dank für Ihre Aufmerksamkeit!

christof.mainberger@bsz-bw.de

Tel.: 0 75 31 / 88 – 4468