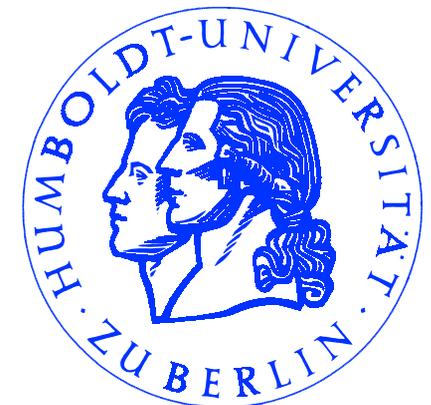




OAI – *Service Provider*

Uwe Müller
Humboldt-Universität zu Berlin
Rechenzentrum
AG Elektronisches Publizieren





Begriffe

Beispiele

Aufbau

Schritte

Ausblick

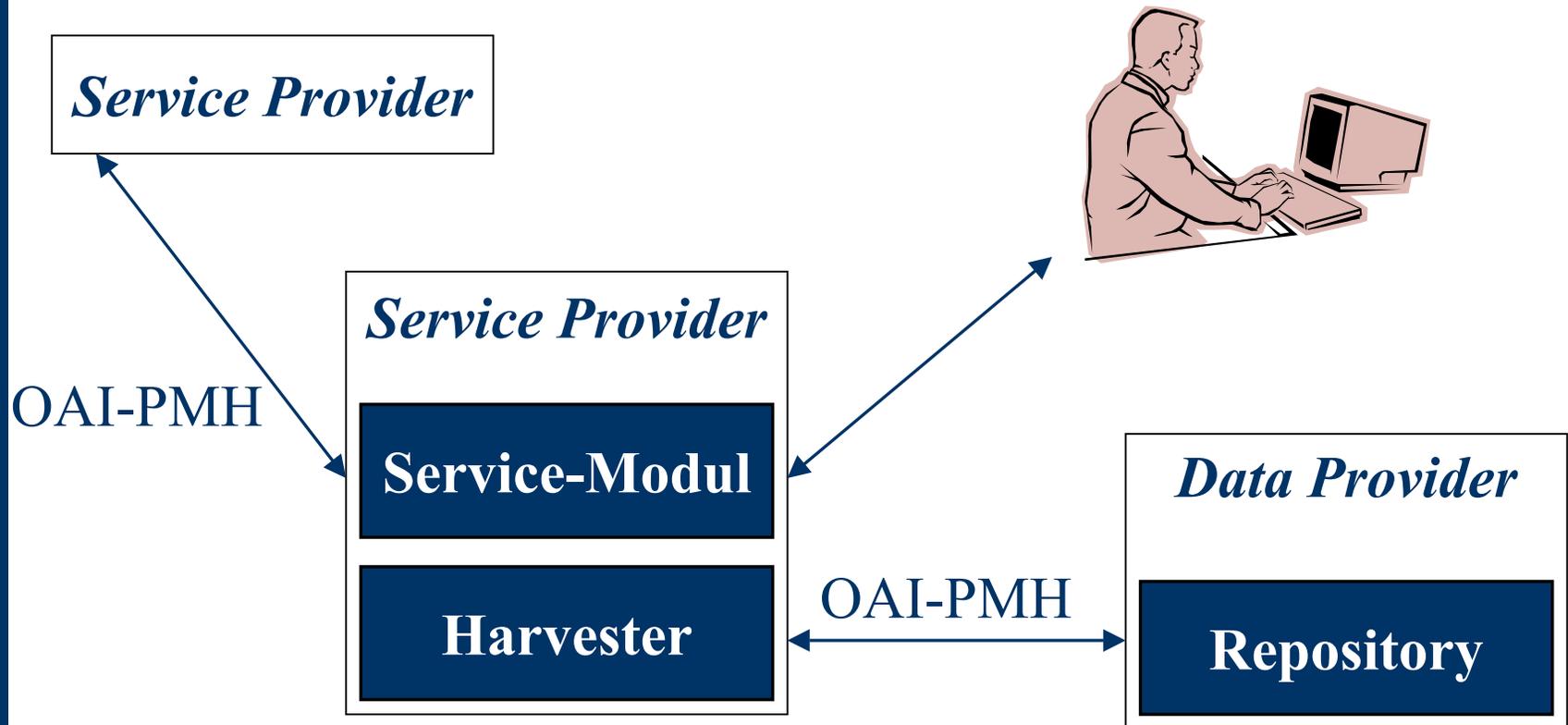


Was ist ein *Service Provider*?

- System, das einen Service bereit stellt, der auf OAI basiert
- verwendet die über das OAI-Protokoll gesammelten Metadaten
- besteht aus zwei Komponenten
 - Harvester (gemäß OAI-PMH)
 - nach „außen“ zur Verfügung gestellter Service
- Harvester: kann mit *Data Providern* kommunizieren
 - Anfragen stellen, Antworten auswerten
- Service-Modul kommuniziert mit „Nutzer“



Was ist ein Service Provider?





Was zeichnet einen *Service Provider* aus?

- stellen den eigentlichen Mehrwert zur Verfügung, der aus OAI resultiert
- Service kann sehr unterschiedlich sein
- Aufwand für Realisierung sehr viel höher als für *Data Provider*
 - Grund liegt in der Protokoll-Definition



Begriffe

Beispiele

Aufbau

Schritte

Ausblick



Beispiel: Repository Explorer

- bei Virginia Tech entwickelt
- Möglichkeit der Abfrage einzelner Archive
- Frontend für *Data Provider*
 - Eingabemaske für Abfrageargumente
- nach außen stark an der OAI-Syntax orientiert
- Testumgebung für Entwickler von *Data Providern*
- Veranschaulichung der grundsätzlichen Arbeitsweise eines *Service Providers*
- oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai



Beispiel: Suchmaschinen

- Vorhalten abgefragter Metadaten in eigener DB
 - Suche basiert auf lokal gespeicherten Daten
 - Anzeige von Metadatensätzen
- Dokumente nicht lokal gespeichert
 - Verweis auf Volltexte in den Archiven
- fachspezifisch
 - z.B. Physnet / Oldenburg
<http://physnet.uni-oldenburg.de/oai/query.php>
- fachübergreifend
 - z.B. DINI <http://www.dini.de/oaisuche/>
 - z.B. Tübingen <http://cm1.zdv.uni-tuebingen.de/portal/index.jsp>



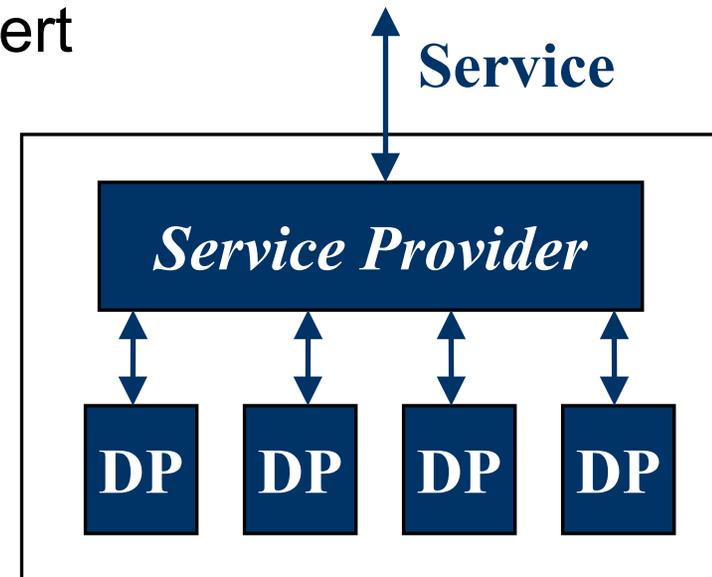
Beispiel: Suchmaschinen / DINI

- Programmiersprache: PHP4 unter Apache
- integrierter XML-Parser
- Datenbanksystem: Sybase / MySQL
- fachübergreifende Suche (Testversion)
- Unterstützung der OAI-Versionen 1.0, 1.1, 2.0
- Metadatenformat: Dublin Core
- Set-Hierarchien beliebig, aber nicht Archiv-
übergreifend



Beispiel: interner Datenaustausch

- Einsammeln von Metadaten von unterschiedlichen Standorten eines Systems
- *Data Provider* sind nicht unbedingt öffentlich zugänglich bzw. offiziell registriert
- Beispiele:
 - BSZ
 - ProPrint





Begriffe

Beispiele

Aufbau

Schritte

Ausblick



Aufbau

Archiv-Verwaltung

- Zusammenstellung der zu durchsuchenden Archive
- manuelles Eintragen
- automatisches Hinzufügen / Löschen von Archiven durch die offizielle Registry

Anfragekomponente

- stellt HTTP-Anfragen an OAI-Archive (*Data Provider*)
- fordert Metadaten unter Benutzung der zulässigen Verbs und der entsprechenden Argumente an
- u.U. fachliche Selektion über *set*-Argument



Aufbau

Scheduler

- realisiert die zeitlich bestimmte regelmäßige Abfrage der angeschlossenen Archive
- einfachster Fall: manuelles Auslösen der Aufträge
- sonst: z.B. Cron-Job

Flusskontrolle

- Resumption-Token: Aufteilung der Ergebnisliste in Teillieferungen; erneute Anfrage, um weitere Ergebnisse anzufordern
- HTTP-Fehler 503: temporäre Nicht-Verfügbarkeit



Aufbau

Update-Mechanismus

- realisiert das Zusammenfügen (Merge) mit früher eingesammelten Metadaten
- einfachster Fall: Vorheriges Löschen aller „alten“ Daten aus dem abgefragten Archiv
- sinnvoll: Benutzung des *from*-Arguments, um nur Änderungen von Metadaten abzufragen

XML-Parser

- wertet die Ergebnisse aus
- überführt die in XML gelieferten Metadaten in die interne Datenstruktur



Aufbau

Normalisierer

- überführt die Daten in eine homogene Struktur bzw. vereinheitlicht ihre Darstellung
 - Metadatenformate
 - Set-Hierarchien
 - Darstellungsformen (Autor, Datum, Sprache)

Datenbank

- Abbildung der XML-Struktur mit einer relationalen Struktur
- oder: XML-Datenbank



Aufbau

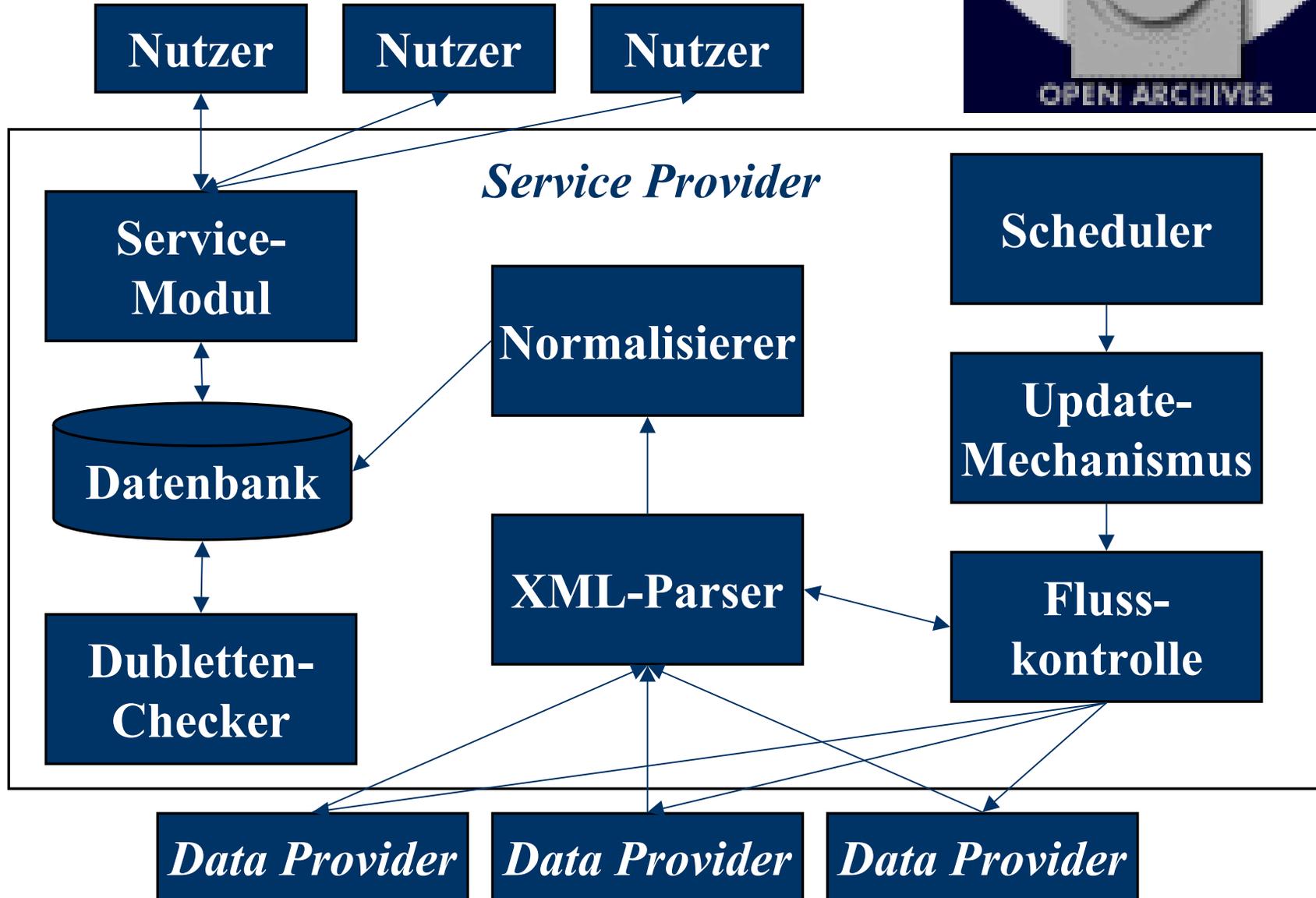
Dubletten-Checker

- führt gleiche Metadatenätze aus unterschiedlichen Archiven zusammen
- Möglichkeit: eindeutiger Identifikator (URN, ...)
- oftmals nicht ohne weiteres möglich bzw. risiko- und fehlerbehaftet

Service-Modul

- stellt auf der Basis der gesammelten und gespeicherten Metadaten einen Dienst nach außen zur Verfügung
- greift nur auf die lokal gespeicherten Daten zurück

OAI – Service Provider





Begriffe

Beispiele

Aufbau

Schritte

Ausblick



Empfehlungen / Schritte

- Welcher Service soll zur Verfügung gestellt werden?
- Welche *Data Provider* sollen integriert werden?
 - sinnvolle Auswahl
- Welcher Metadatensatz wird verwendet?
 - Mindeststandard: Dublin Core
- Welche Set-Definitionen werden verwendet?
 - Empfehlungen von DINI zu OAI (www.dini.de)
- Auswahl des DB-System
- Auswahl der Programmiersprache
- technische Details



Begriffe

Beispiele

Aufbau

Schritte

Ausblick



Ausblick

- „Low Level“-Standard
 - geringer Implementationsaufwand
 - niedrige Schwelle insbesondere für kleinere Archive
 - leichte Erweiterbarkeit
- Integration von OAI-Schnittstellen findet weite Verbreitung
- *Service Provider* als Portal für virtuelle Fachbibliotheken
 - Definition eigener Metadatensätze
 - Definition verbindlicher Set-Hierarchien



Vielen Dank

Fragen?

Rechenzentrum der Humboldt-Universität
Arbeitsgruppe Elektronisches Publizieren
u.mueller@rz.hu-berlin.de
+49 . 30 . 20 93 – 25 46

www.dini.de/dinioai/dinioai.php3