

Kolloquium Wissensinfrastruktur

Neue Methoden für wissenschaftliche Online-Server:
SWORD und OAI-ORE

Jochen Schirrwagen

Universitätsbibliothek Bielefeld

11. Dez. 2009

Stark vereinfachtes Konzept eines typischen Dokumentenspeichers:
Dokument → Metadaten → Datenbank → Ausgabeschnittstellen
typische Eigenschaften eines Dokuments

- Versionen: V_1, \dots, V_n , z.B. Autorenmanuskript, Gutachterversion, Entwurf, Printfassung
- Bestandteile: P_1, \dots, P_n , z.B. Zusatzmaterialien, Artikel einer eJournal-Ausgabe, Primärdaten
- Autoren und Mitwirkende: A_1, \dots, A_n ; M_1, \dots, M_n , z.B. korresp. Autor, weitere Autoren, Übersetzer, Programmierer, ...
- ... Lizenzen, etc.pp.

Wissenschaftliche Kommunikation erzeugt weitere, vielfältige Abhängigkeiten, z.B.

- ähnliche Dokumente - z.B. in bezug auf eine Thematik
- Zitationen, Verlinkungen
- aber auch Collagen/Überlagerungen (Overlay)
- dynamische Dokumente (Annotationen, Kommentare, etc.)
- ER-Beziehungstypen, z.B. alle Dokumente eines bestimmten Autors, etc.

- Sind einfache bibliografische Metadatenformate (wie DublinCore) unzureichend für maschinelle Be- und Verarbeitung ?
- Übertragung / Austausch / Wiederverwendung von - komplexen - Dokumenten → Interoperabilität ?

XML-basiertes Atom Format

- Konsumieren: Atom Syndication Format → plattformunabhängiger Austausch von Informationen
- Produzieren: Atom Publishing Protocol → Erstellung und Bearbeitung von Webressourcen

Überblick

- Ein "Profile" des Atom Publishing Protocol
- Sender implementiert SWORD-Client (z.B. MS-Word), Empfänger implementiert SWORD-Interface (z.B. Repository)
- zweistufiger Prozess
 - 1 Anfordern eines Service-Dokument; liefert Informationen über verfügbare Collections und Packages in die ein authentifizierter Benutzer Dokumente ablegen darf
 - 2 Deposit von Dokumenten

Zur Erinnerung

- PEER untersucht Auswirkungen des „Grünen Weges“ auf die Verfügbarkeit, Sichtbarkeit und Nachhaltigkeit von Zeitschriftenliteratur in Kollaboration mit Wissenschaftsverlagen
- Ansätze zur Übertragung von Publikationen und Metadaten von Verlagen in das PEER-Depot und vom PEER-Depot in OA-Repositories
 - SFTP
 - Authenticated OAI-PMH
 - SWORD

PEER Workflow

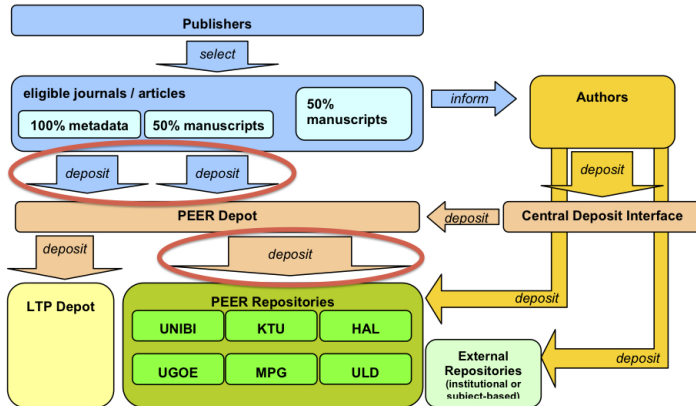


Figure 1: PEER workflow

Abbildung: Quelle: PEER-Projekt, D2.2, 10/2009

- Deposit von zip-Files mit TEI-Metadaten PDF als Stage-2 Dokument
- automatisiertes Entpacken, Validieren, Parsen und Hochladen in das Repository
- SWORD-Interface sendet Feedback als <atom:entry>, darin u.a. vom Repository vergebene Dokument-Id

Deposit im PEER-Kontext

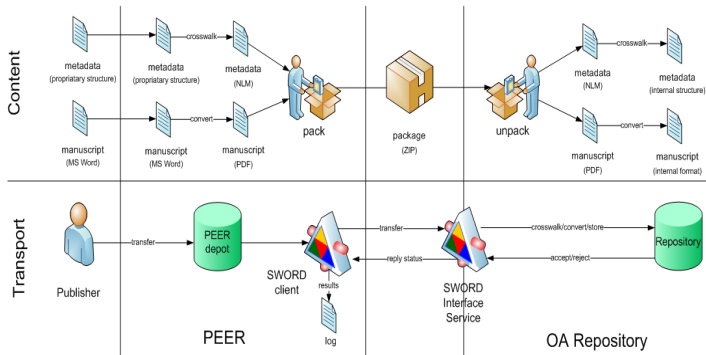


Figure 15: SWORD use in PEER for PEER Depot

Abbildung: Quelle: PEER-Projekt, D2.2, 10/2009

- Plugin für WORD-2007
- PublicationsList.org

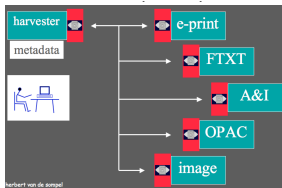
Anforderungen

- Austausch von Informationen über verteilte heterogene Softwaresysteme
- Identifizierbarkeit von Informationen

- Annahme: Repository und Metadaten-Records stehen im Mittelpunkt von Interoperabilitätsüberlegungen

paßt jedoch nicht recht zur Definition einer Web-Architektur

- Annahme: es gibt **Ressourcen**, die durch **URIs** identifizierbar sind und **Repräsentation(en)** haben



Keep dreaming!

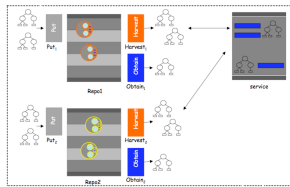


Abbildung: Quelle: Herbert van de Sompel, OAI-ORE slides, Inforum 2009, Prague, Czech Republic

Resource Description Framework

- Triples von Subject, Prädikat, Objekt
- <Paper><hatAutor><Manfred Mustermann>

Linked Data

- <http://www.w3.org/DesignIssues/LinkedData.html>
- analog zu Links in Hypertexten
- z.B. Linked Data aus den Datensätzen der DBLP Bibliography Database für Computer Science Journals und Konferenzbände
<http://www4.wiwiss.fu-berlin.de/dblp/>

Cool URIs

- <http://www.w3.org/TR/cooluris>
- Verfahren zur Sicherstellung der Stabilität von URIs

Termini im Kontext von OAI-ORE

- **Aggregation**: Menge von Webressourcen, die eine bestimmte Sache beschreiben
- eine Aggregation kann aber auch selbst wieder eine Webressource darstellen (Aggregation von Aggregationen)
- **Resource Map**: ein RDF-Dokument, welches eine Aggregation beschreibt → Kandidat für den Austausch
- Publikation von Resource Maps im Web
- **Relationen** und **Typen**

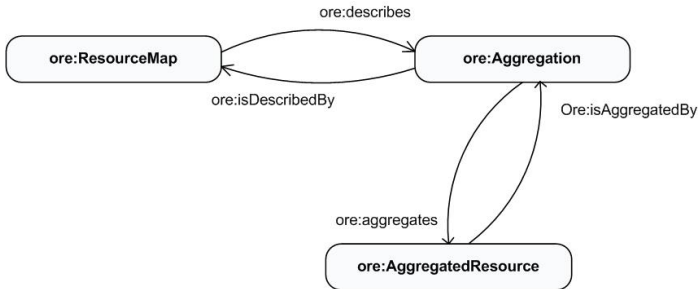


Figure 4 Basics of the OAI-ORE model

Abbildung: Quelle: DRIVER-Projekt, D4.2 Report on Object Models and Functionalities, 2008

Enhanced Publications - Modell für

- Paper als Textdokument mit
 - Forschungs-/Primärdaten
 - Zusatzmaterialien
 - Annotationen, Kommentare von Dritten

Demonstrator für Enhanced Publications

DRIVER Enhanced Publications Model

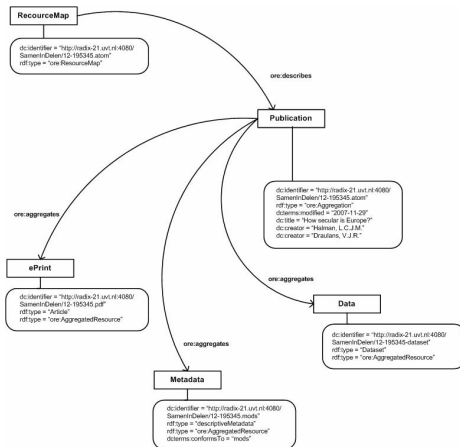


Figure 5 Serialisation of an enhanced publication (Example 1)

Abbildung: Quelle: DRIVER-Projekt, D4.2 Report on Object Models and Functionalities, 2008

DRIVER im Kontext der Repositorien - Resource Map Discovery

Batch Discovery

- Exponieren von OAI-ORE records über OAI-PMH, z.B. DSpace-Repository der Universität Utrecht, NL
- Atom Feeds
- SiteMaps

Einbettung von Discovery-Links in Web-Ressourcen:

- mit HTML <link>Element, z.B. DRIVER Technology Watch Report beschrieben als RDF

- myexperiment.org - zur Beschreibung von „research objects“
- Zentity - Repository Platform von Microsoft Research
- Übertragung digitaler Ressourcen für die Langzeitarchivierung

- Validierung von ResourceMaps
- OAI-ORE Wordpress-Plugin

- SWORD
- PEER-Project - D2.2 Final report on the provision of usage data and manuscript deposit procedures for publishers and repository managers
- OAI-ORE
- DRIVER Technology Watch Report
- DRIVER Enhanced Publications

Vielen Dank für Eure Aufmerksamkeit!