

eco System für semantisch ausgezeichnete Publikationsdaten

eco4r

Jochen Schirrwagen,
Universitätsbibliothek Bielefeld

Überblick

- Projektinhalte und –ziele
- Ausgangsmotivation
- Bestandsaufnahme
- Datenmodell und Anwendungsszenario
- Herausforderungen und Perspektiven

Projekthalte und -ziele

- Mehrteilige Publikationen in Repositorien
 - Maschinenlesbare Repräsentation, Verarbeitung, Austausch, Wiederverwendung
 - semantische Auszeichnung
 - Vorbereitung für die Langzeitarchivierung
- unter Anwendung der Standards OAI-ORE und METS
- Demonstrator
- nachnutzbare Softwarekomponenten

Ausgangsmotivation

- das Web ist **die** Publikationsplattform für Volltexte und Forschungsdaten
- erfordert von Repositories Kompatibilität mit der Web-Architektur
 - maschinenlesbare Repräsentation
 - Eindeutige Identifizierbarkeit von Entitäten
 - Relationen

Bestandsaufnahme - Typklassifizierung

- Typ 1 – einteilige Publikationen mit mehreren Manifestationen
- Typ 2 – mehrteilige Publikationen ohne Zusatzmaterial
- Typ 3 – mehrteilige Publikationen mit Zusatzmaterial

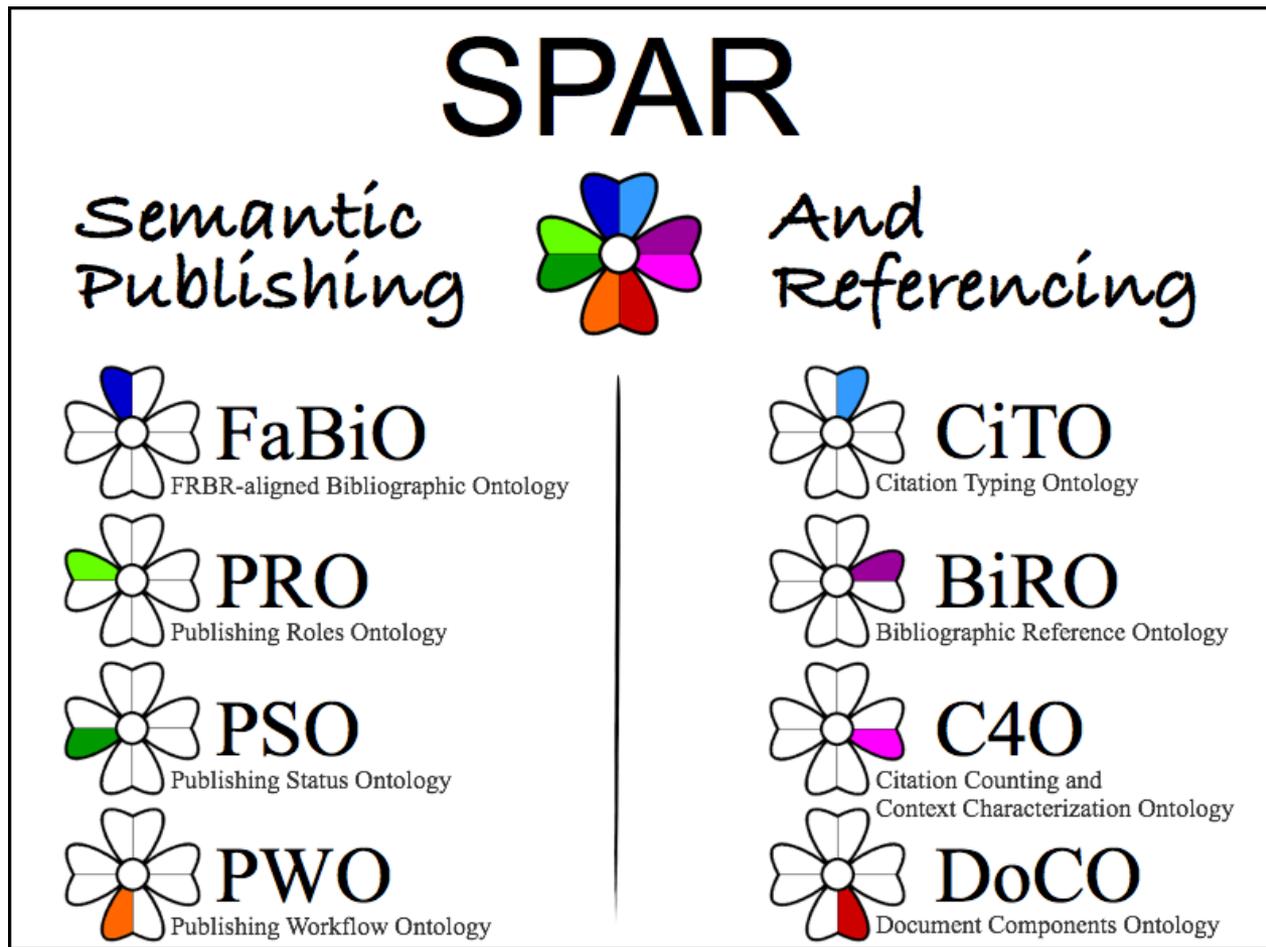
Bestandsaufnahme

Plattform/ System	Publikation insgesamt (Stand)	Art der Publikation	Typ 1	Typ 2	Art der Mehrteiligkeit	Typ 3	Art des Zusatzmaterials
DiPP/ Fedora	1221 (11/2010)	OA- eJournal- Artikel	1201	2	zweiteiliges eBook	22	Video, Audio, Excel Tabellen, ZIP, TXT, RTF, Java- Applet
BiPrints/ OPUS v3	3899 (05/2010)	Post-Prints	-	-	-	3	2x PDF-Supp. 1x ZIP-Suppl.
BieSO _n / OPUS v2	1250 (05/2010)	Diss.	18 (ps/pdf)	63	58x Kapitel und Anhang 4x Proceedings 1x Monographie	10	2x Softwaredownloads 4x Statistiken 1x Elearning-Material 1x Videosequenzen 1x Studie 1xCD-ROM als –Zip- Datei
BieCol/ OPUS v3	414 (05/2010)	Konferenz- beiträge	-	10	Kollektion/ Konferenzband mit Beiträgen	-	-

Datenmodell

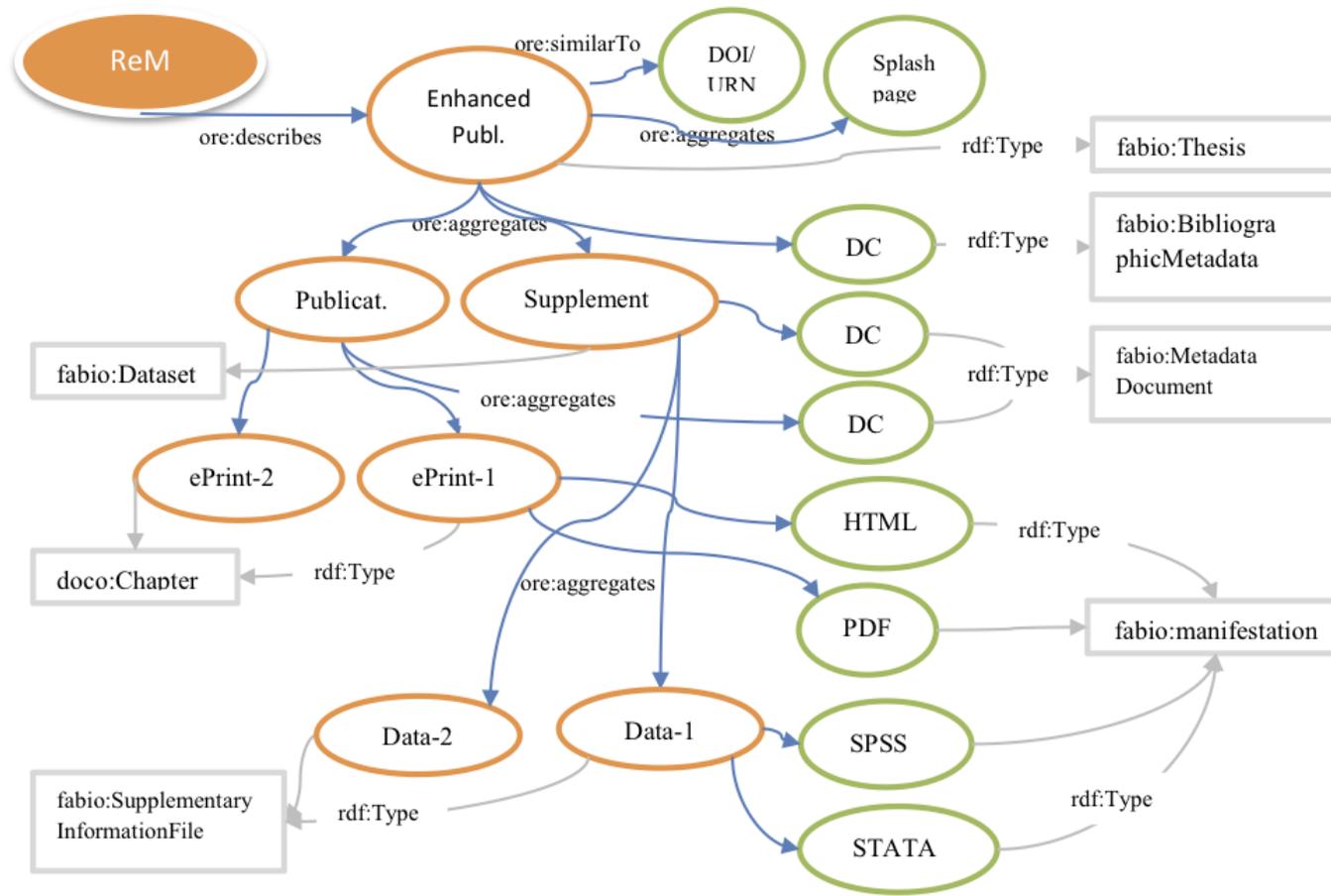
- basiert auf der Idee der „Enhanced Publications“ (DRIVER-II)
- OAI-ORE Abstract Data Model
 - zur Modellierung aggregierter Ressourcen
- SPAR-Ontologien zur Auszeichnung bibliographischer Entitäten
 - FaBiO – FRBR aligned Bibliographic Ontology
 - DoCO – Document Component Ontology

Bibliographisches Modell



Flower diagram of the SPAR component ontologies, Quelle: Shotton et.al.

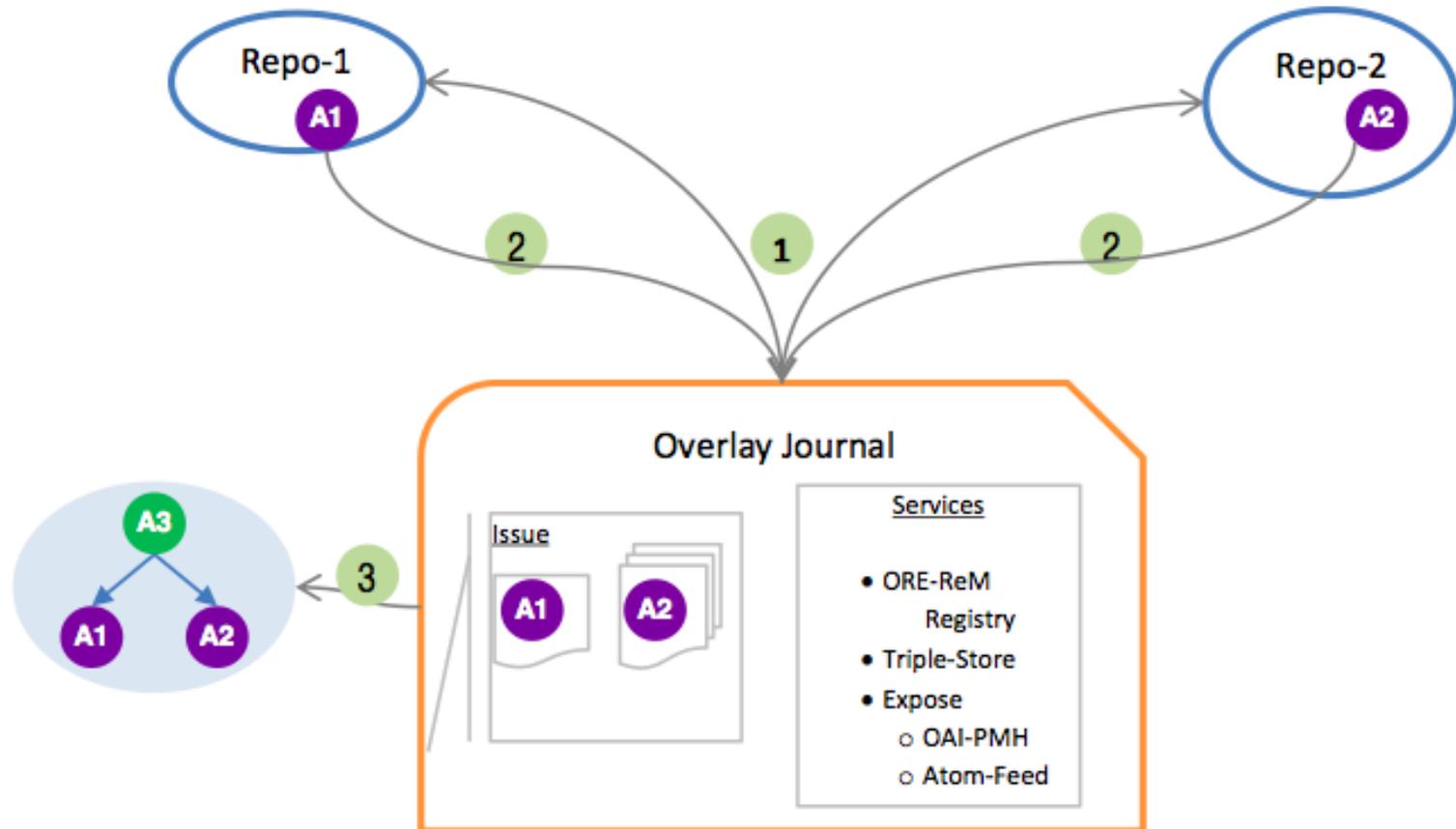
eco4r Datenmodell - Beispiel



Umsetzung und Anwendungsszenario

- „Dissertationsmagazin“ als Overlay Journal
 - Aggregation von verteilt vorliegenden Publikationsdaten
 - Repräsentation von Kollektionen nach fachspezifischen Themen
 - Exponieren derart angereicherter Aggregationen
- Nutzung von Linked Data Services (DDC, PND)
- Repository-Plugins für
 - Fedora, basierend auf OAI-ORE Implementierungen
 - OPUS, basierend auf ARC RDF-Framework für PHP

Overlay-Journal Skizze



Herausforderungen

- Mehraufwand bei der Metadatenerstellung
 - Perspektivisch qualitativ bessere und feingranulare Metadateninformationen
- (global) eindeutige Identifikatoren für Publikationen, Datasets, Autoren, Organisationen ...

Perspektivische Anwendungen

- Crawling- und Harvestingaktivitäten
 - unterstützt durch semantisch angereicherte Daten
- Verlinkung zwischen verteilten Publikations- und Forschungsdateninfrastrukturen
- Synchronisation bei verteiltem Publikationsmanagement
- Verknüpfung von Forschungsinformationen und Publikationsdaten (CRIS-OAR, OpenAIRE)

Inhaltliche Querverbindungen

- Komplementäre Entwicklungen im mORE-Projekt
 - unter Nutzung des Europeana Data Model
- Praxisfokussierte Anwendungen von „Enhanced Publications“ bei SURF (NL)
- „Library Linked Open Data“ Community

Im Team mit

- Anouar Boulal, Andres Quast (hbz Köln)

Mit Unterstützung von

- Martin Iordanidis, Jan Schnasse (hbz Köln)
- Wolfram Horstmann, Friedrich Summann (UB Bielefeld)

Weitere Informationen

- Projekt-Webseite: www.eco4r.org
- Projekt-Wiki: trac.eco4r.org