



[www.kim-forum.org](http://www.kim-forum.org)



# Interoperable Metadatenmodelle und Repositorien

Stefanie Rühle

Niedersächsische Staats- und Universitätsbibliothek Göttingen

Mirjam Keßler

Deutsche Nationalbibliothek



# Ausgangslage

---

- Metadatenutzung und Heterogenität
- standardisierte Schnittstellen
- OAI-PMH
- OAI-ORE



# Ausgangslage

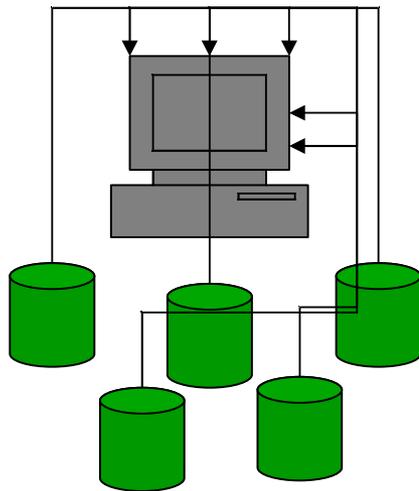
---

- **Vernetzung von Repositories für übergreifende Services**
  - Bereitstellung von elektronische Publikationen
  - Austausch von Metadaten der Publikationen
- **Voraussetzungen für Metadatenaustausch**
  - standardisierte Schnittstellen
  - standardisierte Metadatenformate
  - Interoperabilität von Metadatenmodellen



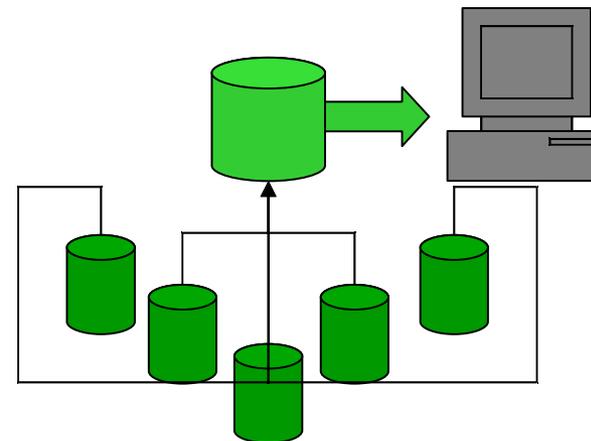
# Ausgangslage: Metadatenutzung

## Nutzung von Metadaten in einem gemeinsamen Suchraum



Metasuche

Cross searching



Suchmaschinenindex

Harvesting



# Ausgangslage: Heterogenität

---

- **Heterogenität im Bereich Repositories**
  - Objekte
    - ⇒ unterschiedliche Typen (Volltext, Filme, Audio, Bilder, ...)
  - Metadaten
    - ⇒ bibliographische, Langzeitarchivierung, Rights, „administrative“ = lokale/domainspezifische
    - ⇒ unterschiedliche Metadatenformate



# Ausgangslage: heterogene Metadaten

---

MPEG-7

IPTC-NAA

BibTeX

EAD

Dublin Core

MPEG-21

MARC 21

MAB

METS

LOM

MODS

ONIX

EXIF

MOF

FGDC

XMP

XBMF

VRA Core

PICA

PREMIS

ID3-Tags

OLAC

USW.



## Ausgangslage: Bedingungen

---

- Austausch von Metadaten bei Vernetzung von Repositories für übergreifende Services benötigt:
  - standardisierte Schnittstellen  
⇒ OAI-PMH
  - Object Re-Use and Exchange  
⇒ OAI-ORE
  - Standardisierung von Metadaten  
⇒ Dublin Core Metadata Element Set
  - Interoperable Metadatenmodelle  
⇒ Singapore Framework, DCAM



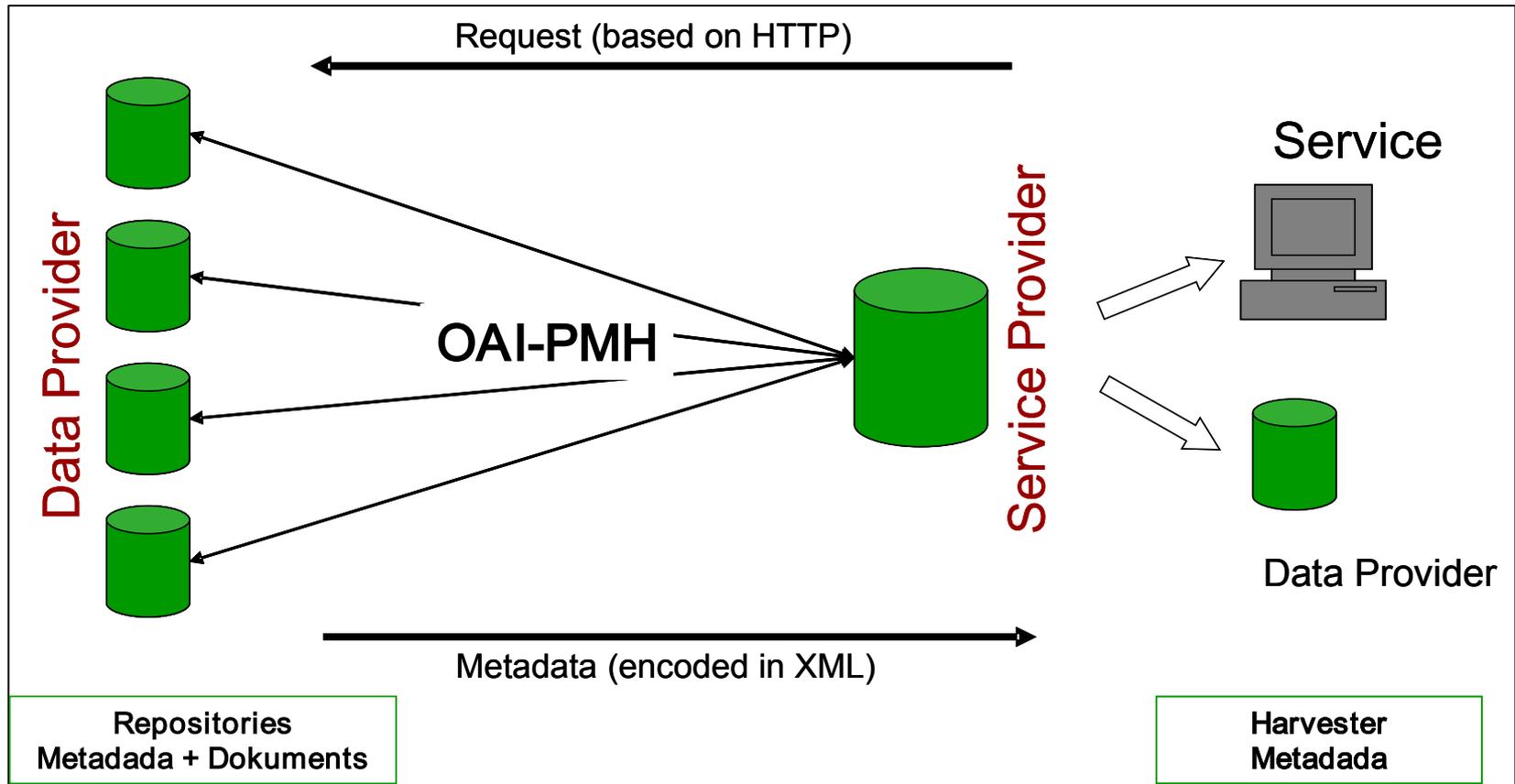
## Standardisierte Schnittstellen: OAI-PMH

---

- **Ziel**
  - „Low-barrier Interoperability“ => Interoperabilität durch geringst möglichen Aufwand
  - anwendungsunabhängiges Framework für Interoperabilität, das auf *Metatdata-Harvesting* beruht
- **Standards**
  - Hypertext Transfer Protocol (HTTP)
  - eXtensible Markup Language (XML)
- **Interoperabilität des OAI-PMH**
  - basiert auf den vorgesehenen Metadaten: Dublin Core simple



# OAI-PMH: Modell





# Object Re-Use: OAI-ORE

---

- **Ziel**
  - Abbildung der Binnenstruktur und der Verknüpfungen von einzelnen Aggregaten eines digitalen Objektes zur Nutzung zum Austausch
    - ⇒ Binnenstruktur (Buch => Kapitel => Seiten)
    - ⇒ Verknüpfung logisch zusammenhängender Aggregate (PDF, MP3)
    - ⇒ externe Verknüpfung logisch zusammenhängender Ressourcen (Zitate)
- **Bestandteile eines digitalen Objekts variieren in:**
  - semantischer Typ (Buch, Video)
  - Medien-Typ (Text, Image)
  - Medien-Format (PDF, MP3)
  - Ort (World Wide Web, institutionelle Repositories)
  - Relationships (extern - intern)



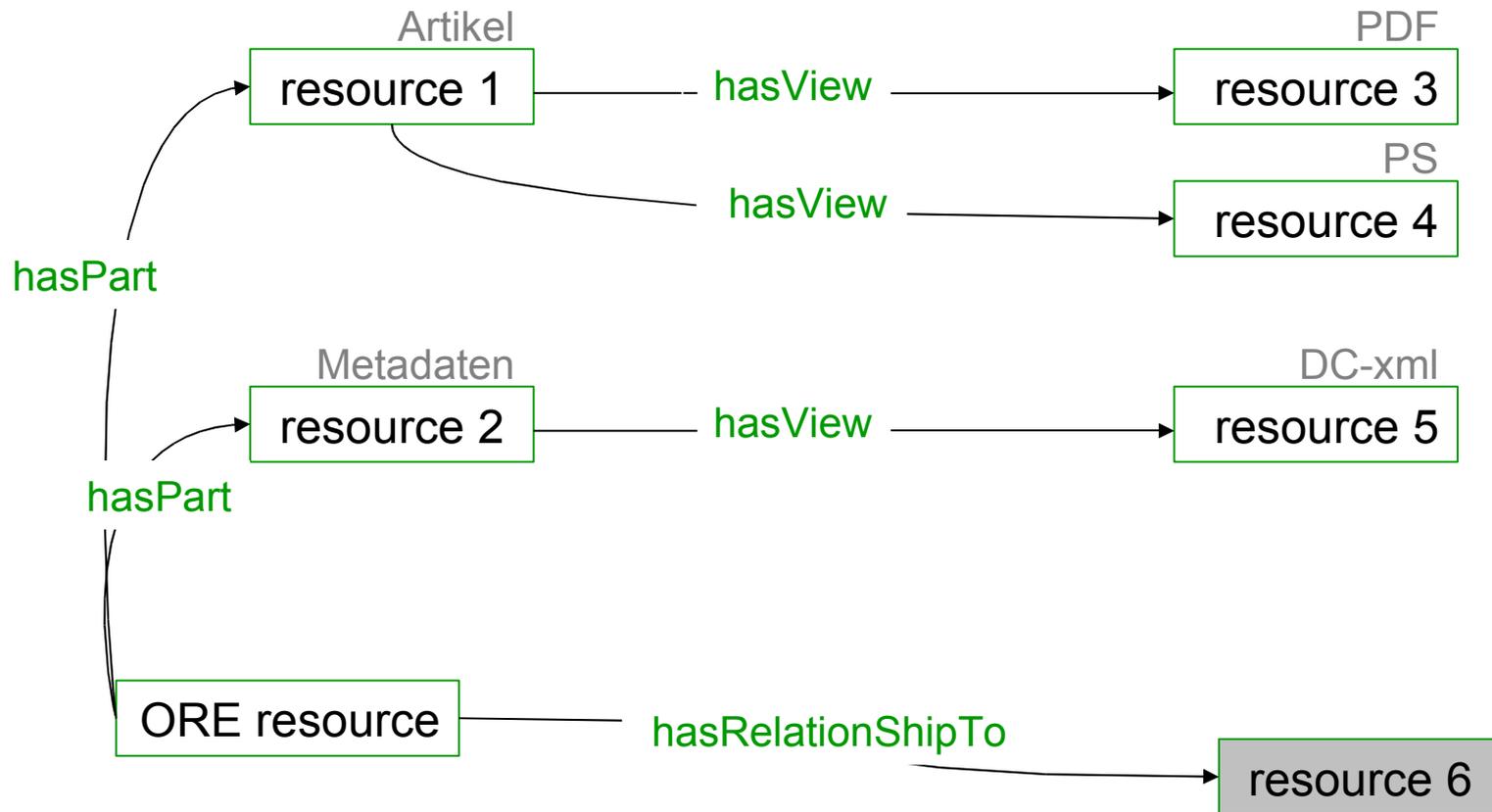
# ORE PMH

---

- **OAI-ORE Prinzipien**
  - Interoperabilität basiert auf Prinzipien der W3C Web-Architektur  
⇒ OAI-PMH basiert auf Metadaten-Interoperabilität
  - Interoperabilität auf Objektebene  
⇒ OAI-PMH auf Metadaten-Ebene
  - Ziel ist es, mittels URIs den Object Re-Use von digitalen Objekten und Aggregaten zu ermöglichen  
⇒ OAI-PMH Metadata Harvesting

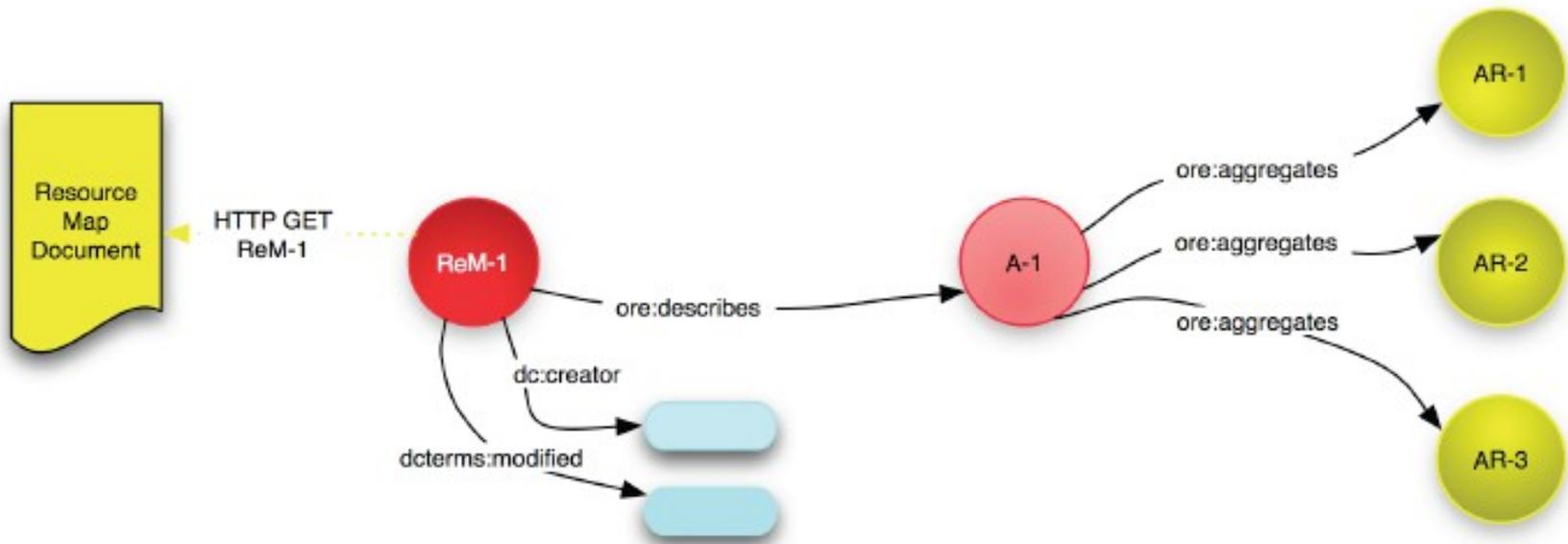


# ORE - Modell





# ORE: Resource Map



Quelle: H. Van de Sompel: Thoughts about Repositories, Use, and Re-Use



---

# Medaten-Harvesting

---

- Ziele und Methoden am Beispiel Geo-Leo
- Metadaten Interoperabilität
- Singapore Framework
- Zusammenfassung



# Metadaten - Harvesting

---

- **Ziel Datenanbieter:**
  - Sichtbarkeit verbessern durch Integration der Metadaten in relevante Portale
- **Ziel Serviceanbieter:**
  - Angebot um relevante Ressourcen erweitern



## zum Beispiel Geo-Leo

---

- **Ziel:**
  - Bestände der Hochschulschriftenserver im Portal nachweisen.
  - Nur fachlich relevante Bestände berücksichtigen
  - Metadaten der Server in eine übergreifende Suche integrieren
- **Verfahren:**
  - Harvesting der Metadaten mit OAI-Diggr
  - Regelmäßige automatische Abfrage ausgewählter Server (60 Repositories)
  - Auswahl der records nach Content in dc:subject



# Interoperabilität

---

- Prüfung möglicher Hochschulschriftenserver ergibt:
  - OAI-Schnittstellen sind vorhanden
  - Mit DC-Simple existiert ein allen gemeinsames Metadatenformat für den Austausch der Daten
  - DC-Subject ist in der Mehrzahl der Fälle belegt und ermöglicht somit die fachliche Selektion
  
- Ergebnis: Hochschulschriftenserver sind interoperabel

???



# Interoperabilität von Metadaten ?

---

- **Unterschiedliche encoding schemas**
  - dc:subject: unterschiedliche Thesauri, Klassifikationen, Schlagwortnormdaten
  - dc:date: unterschiedliches encoding der Datumsangaben
  - dc:language: unterschiedliches encoding der Sprachangaben
  - ...
- **Keine encoding schemas**
  - dc:subject: freie Stich- und Schlagworte anstelle von kontrolliertem Vokabular
  - dc:date: unregelmäßiger Gebrauch von Datumsangaben
  - dc:language: unregelmäßiger Gebrauch von Sprachangaben
  - ...



# Interoperabilität der Metadaten ?

---

- **Semantische Erweiterung des element-content**
  - dc:title beinhaltet neben dem Titel auch den Verfasser
  - dc:creator wird ohne Unterschied für Verfasser und beteiligte Personen verwendet
  - ...
- **Element-content semantisch nicht korrekt.**
  - dc:relation enthält den Herausgeber
  - dc:coverage enthält Angaben zum Erscheinungsjahr
  - dc:source enthält Angaben zur relation



# Metadatenformate interoperabel gestalten

---

- Ein standardisiertes Metadaten-Anwendungsprofil muss die folgenden Fragen beantworten:
  - Was will ich mit den Metadaten beschreiben und welche Beziehungen bestehen zwischen den Entitäten, die ich beschreiben will?
    - ⇒ Online-Volltexte, audiovisuelle Ressourcen, Personen ...
  - Welches Ziel habe ich?
    - ⇒ Einbindung in eigene Suche, in nationale oder internationale Portale ...
  - Was für Anforderungen habe ich?
    - ⇒ Einfache Suche, Erweiterte Suche, thematisches Browsing ...
  - Was brauche ich, um meine Ziele und Anforderungen zu erreichen?
    - ⇒ Metadatenelemente
    - ⇒ Encoding schemas
    - ⇒ „Regeln“



---

# Singapore Framework for Dublin Core Application Profiles

---

- <http://dublincore.org/architecturewiki/SingaporeFramework/>
- „...a framework for designing metadata applications for maximum interoperability and for documenting such applications for maximum reusability.“



# Singapore Framework

---

- **Standard für die Dokumentation von Metadaten-Anwendungsprofilen**
  - definiert die Komponenten, die für die Beschreibung eines Metadatenanwendungsprofils notwendig sind.
  - Beschreibt die Rolle dieses Standards im Rahmen von Standard Domain Models und Semantic Web
- **Ermöglicht die Kombination von**
  - Verschiedenen Metadatenstandards
  - Metadatenstandards und domainspezifischen Metadatenelementen
  - Auf der Basis des Dublin Core Abstract Model (DCAM)
- **Unterstützt die syntaktische Interoperabilität von Metadatenanwendungen**



## Dublin Core Abstract Model (DCAM)

---

- <http://dublincore.org/documents/2007/02/05/abstract-model/>
- Referenzmodell auf der Grundlage von RDF
- Meta- Modell, das auch auf Terme anwendbar ist, die nicht zur DC-Domain gehören
- Fordert die Verwendung von URIs für die Beschreibung der Terme



# Singapore Framework – verpflichtende Komponenten

---

- “Functional requirements“
  - Was für Anforderungen habe ich?  
⇒ Einfache Suche, Erweiterte Suche, thematisches Browsing ...
- “Domain model“
  - Was will ich mit den Metadaten beschreiben und welche Beziehungen bestehen zwischen den Entitäten, die ich beschreiben will?  
⇒ Online-Volltexte, audiovisuelle Ressourcen, Personen ...
- “Description Set Profile“
  - Was brauche ich, um meine Ziele und Anforderungen zu erreichen?  
⇒ Metadatenelemente, Encoding schemas, „Regeln“
  - Spezifiziert die im Domain model benannten Entitäten



---

# Singapore Framework – optionale Komponenten

---

- **“Usage guidelines“**
  - In welcher Form ist das Anwendungsprofil zu verwenden, was muss bei der Implementierung beachtet werden?
- **„Encoding syntax guidelines“**
  - Ggf. Beschreibung der profilspezifischen Syntax



# Zusammenfassung

---

- Verwendung von Standards bei der Beschreibung von Ressourcen
- Nicht irgendeinen Standard, sondern im Kontext von
  - Zielsetzung
  - Anforderungen
  - Mögliche Nachnutzungsszenarien
- Standardisierte Dokumentation der domainspezifischen Verwendung von Standards
  - Möglichst unter Berücksichtigung des Singapore Framework
- Standards nicht nur verwenden, sondern sich auch daran halten
  - Abweichungen gut dokumentieren
  - Darauf achten, dass Abweichungen nicht im Widerspruch zu dem verwendeten Standard stehen



---

**Vielen Dank für Ihre Aufmerksamkeit**

**Fragen? Anmerkungen?**

**Stefanie Rühle**  
SUB Göttingen  
[sruehle@sub.uni-goettingen.de](mailto:sruehle@sub.uni-goettingen.de)

**Mirjam Keßler**  
Deutsche Nationalbibliothek  
[m.kessler@d-nb.de](mailto:m.kessler@d-nb.de)

**[www.kim-forum.org](http://www.kim-forum.org)**