

Universität Duisburg-Essen, Campus Duisburg

Institut für Informatik und interaktive Systeme

Fachgebiet Informationssysteme

Blockseminar zu Studienprojekt

**Yahoo für das Invisible Web:**

Scatter/Gather-Clustering für semistrukturierte Daten

Prof. Dr. Norbert Fuhr  
Dipl.-Inform. Gudrun Fischer  
(Wintersemester 2003/04)

# **Die Open Archives Initiative und Metadaten**

Ting Li

04. 02. 2004

# Die Open Archives Initiative (OAI) und Metadaten



Die „Open Archives Initiative“ hat sich den weltweiten Austausch von elektronischen Veröffentlichungen zum Ziel gesetzt und entsprechende Standards definiert. In diesem Papier werden die Geschichte der OAI – ihr Ursprung bei der Förderung von E-Print, das Erweitern ihres Fokus (Kap.1), die Details ihres technischen Standards für Metadaten (Kap.2), die Anwendungen dieses Standards (Kap.3) und anschließend die aktuellen Probleme und zukünftige Aussichten (Kap.4) beschrieben.

## 1. Einführung der OAI

### 1.1 Gründe der Entstehung der OAI:

Die Open Archives Initiative ist ein Ergebnis von vier Entwicklungen Anfang der 90er Jahre, aus denen sie auch ihre Zielsetzung definiert hat.

- Die Wurzeln der Open Archives Initiative liegen in den Bemühungen, Zugang zu den neuesten Forschungsergebnissen zu verschaffen. Z.B: Bei der Vorbereitung der Diplomarbeit hat man entweder aufmerksam die Fachpresse verfolgt oder beginnt intensiv in Bibliotheken und Büchereien nach geeigneter Literatur zu recherchieren. Diese Wissenbasis ist jedoch begrenzt und ihre Ergründung ist recht *zeitaufwändig*, insbesondere auf internationaler Ebene. Mit dem Internet sind weitere Möglichkeiten aufgekommen, aber man findet jedoch nur *begrenzte Angebote*.
- Forscher und Autoren von Zeitschriftenartikeln wurden zunehmend unzufrieden mit der zeitlichen Verzögerung zwischen Abgabe eines eingereichten Aufsatzes bis zur Veröffentlichung, mit den hohen Zeitschriftenpreisen, die viele Bibliotheken veranlasst haben, Abonnements abzubestellen und Bestände zu reduzieren, sowie mit der üblichen Übertragung der Autorenrechte an den Verlag, der dann unter anderem das alleinige Vertriebsrecht hatte und durch seine Publikations- und Preispolitik die Verbreitung seiner Forschungsergebnisse einschränken konnte.
- Um die schnelle Verarbeitung zu erreichen, gewinnt die Interoperabilität

sowie Standardisierung bei Metadatenformaten und Datenaustauschprotokollen zunehmend an Bedeutung.

- Weil jeder Dokumentenserver eine eigene Suchoberfläche anbot, die sich natürlich in der Bedienung unterschieden, wurde eine standardisierte Suchschnittstelle über alle Archive gewünscht.

Zur Lösung dieser Problematik ist die „Open Archives Initiative“ (OAI) entstanden. Sie beschreibt sich wie folgt:

*“The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication.”*

Die OAI entwickelt und fördert Interoperabilitätsstandards, die darauf abzielen, die leistungsfähige Verbreitung von Inhalten zu erleichtern. Die „Open Archives Initiative“ hat sich also den weltweiten Austausch von elektronischen Veröffentlichungen zum Ziel gesetzt und entsprechende Standards definiert.

Ergebnis ist eine netzwerkartige Struktur, der sich jeder anschließen kann. In diesem Rahmen gibt die OAI vor:

- Mindestvorgaben zur Beschreibung von elektronischen Dokumenten (Dublin Core) (siehe Kap.2).
- ein Protokoll zur Anbindung an das internationale OAI-Netzwerk zum Austausch von elektronischen Dokumenten (siehe Kap.3).

Durch das definierte Austauschprotokoll werden nicht die elektronischen Publikationen selbst, sondern lediglich ihre standardisierten Beschreibungen ausgetauscht. Der Bezug des Dokuments erfolgt über einen Verweis auf den lokalen Dokumentenserver im OAI-Netzwerk.

### **1.3 Kurze Geschichte der OAI:**

Nach einer Erläuterung der Ereignisse und Gründe, die zur Entwicklung der Open Archives Initiative (OAI) geführt haben, wird die Entstehungsgeschichte der OAI kurz dargestellt.

Das Gründungstreffen fand am 21.- 22. Oktober 1999 in Santa Fe im Bundesstaat New Mexiko (USA) mit der Unterstützung mehrerer richtungsgebender bibliothekarischer Fachgesellschaften und Veröffentlichungsorgane statt. Ziel des Treffens war es, die technischen und organisatorischen Voraussetzungen für einen Service zu schaffen, der die Metadaten aller Preprint-Server, fachlich unabhängig von einander, in ein virtuelles, universelles Preprint-System einbindet und dabei ein neues wissenschaftliches Kommunikationsmodell zusammenführt.

Innerhalb von wenigen Monaten (im Jahr 2000) entstand das erste Protokoll, das diese Interoperabilität für das Alpha-Test-Format für die Frühanwender anbot. Während dieser Zeit entwickelte sich der Name UPS (Universal Preprint Service) zur „Open Archives Initiative“.

Die erste offizielle Version wurde im Januar 2001 vorgestellt. Um die Version 1.0 im Januar mit Erfahrungsberichten der Erstanwender in der Öffentlichkeit vorzustellen, waren dafür zwei „Open Meetings“ geplant: eines am 23. Januar 2001 in Washington, D.C. (USA) und ein europäisches Treffen am 26. Februar 2001 in Berlin (Deutschland). Darauf folgten zwei Workshops in den USA und Europa zur Darlegung des Protokolls und Vorstellung exemplarischer Implementierungen des Protokolls. Das OAI - Metadata - Harvesting - Protokoll wurde in seiner Zielsetzung und Möglichkeiten detailliert dargestellt und auch die vom Protokoll erforderlichen Metadaten wurden erläutert. Die Basisbefehle zur Abfrage in verschiedenen OAI-kompatiblen Repositories, auch mit Unterscheidung zwischen besonderen Sammlungen innerhalb der Repositories, wurden erklärt.

## 2. Metadaten

Nachdem die Geschichte der OAI erklärt wurde, wird die Beschreibung eines elektronischen Dokuments (Metadaten) bzw. ihre verschiedenen Formate (Metadatenformate) in diesem Kapitel kurz erläutert.

### 2.1 Was sind Metadaten und Metadatenformate?

**Metadata** (Metadaten) sind Beschreibungen eines elektronischen Dokuments. Eine solche Beschreibung muss durch ein entsprechendes Schema[1] definiert sein - das Metadata Format. Um den Austausch von elektronischen Dokumenten weltweit zu ermöglichen, ist es empfehlenswert, ein offizielles und möglichst auch international anerkanntes Kategorisierungssystem zu benutzen.

Die minimale Anforderung im Rahmen von OAI[5] ist, dass jedes elektronische Dokument mindestens durch den **Dublin Core**[2] beschrieben wird. Der Dublin Core enthält sowohl optionale als auch wiederholbare Elemente.

*Zusätzlich* zum Dublin Core ist meist eine detailliertere Einordnung im weiteren Fachgebiet durch angemessene Schemata sinnvoll. Nachfolgend seien beispielhaft einige fachspezifische Schemata aufgeführt:

- ETD-ms (Theses / Dissertations)
- OLAC (Open Language Archive Community)
- MARC (Libraries)
- IMS (Education)
- RFC 1807 (Bibliographies)

## 2.2 Dublin Core Metadaten:

Das **Dublin Core Metadata Element Set** ist ein Standard für die Beschreibung von Informationsressourcen aus unterschiedlichen Bereichen. Hier werden die Informationsressourcen als „alles, das Identität hat“ definiert. Dieses ist die Definition, die in Internet-RFC 2396 verwendet wird. Es gibt keine grundlegenden Beschränkungen zu den Arten von Ressourcen, denen Dublin Core zugewiesen werden kann.

Dublin Core enthält 15 Elemente, die optional und wiederholbar sind. Die folgende Webseite enthält einen Überblick zu den Spezifikationen aller Metadata-Elemente:

<http://dublincore.org/usage/documents/overview/>.

### Dublin Core Metadata Element Set:

1. Title:  
Titel der Quelle, der vom Verfasser, Urheber oder Verleger vergebene Name der Ressource.
2. Creator:  
Die Person(en) oder Organisation(en), die hauptsächlich für den intellektuellen Inhalt verantwortlich sind.
3. Subject:  
Ein Thema, Schlagwort oder Stichwort der Ressource.
4. Description:  
Eine Textbeschreibung des Ressourceninhalts.
5. Publisher:  
Die Einrichtung, die dafür verantwortlich ist, dass diese Ressource in der Form zur Verfügung steht.
6. Contributor:  
Zusätzliche Person(en) und Organisation(en) zu jenen, die im Element „Creator“ genannt wurden, die einen bedeutsamen intellektuellen Beitrag zur Ressource geleistet haben, deren Beitrag aber sekundär im Verhältnis zu denen im Element „Creator“ zu betrachten ist.
7. Date:  
Das Datum, an dem die Ressource in der gegenwärtigen Form zugänglich gemacht wurde.
8. Resource Type:  
Die Art der Ressource.

9. Format:  
Hier wird das datentechnische Format der Ressource eingetragen.
10. Identifier:  
Eine Zeichenkette oder Zahl, die diese Ressource eindeutig identifiziert.
11. Source:  
Das gedruckte oder elektronische Werk, aus dem diese Ressource stammt.
12. Language:  
Die Sprache(n) des intellektuellen Inhalts dieser Ressource.
13. Relation:  
Eine Möglichkeit, Verbindungen unter verschiedenen Ressourcen darzustellen.
14. Coverage:  
Hier werden Angaben zur räumlichen Bestimmung und zeitlichen Gültigkeit eingetragen.
15. Rights:  
Vermerk über die rechtlichen Bedingungen oder ggf. zu einem Server, der solche Informationen dynamisch erzeugt.

Das folgende Beispiel ist der Metadatenteil eines OAI-Records in **Dublin Core:**

```
<metadata>
  <oai_dc:dc
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    <dc:title>Using Structural Metadata to Localize Experience of Digital Content
    </dc:title>
    <dc:creator>Dushay, Naomi</dc:creator>
    <dc:subject>Digital Libraries</dc:subject>
    <dc:description>With the increasing technical sophistication of both information
      consumers and providers, there is increasing demand for more meaningful
      experiences of digital information. We present a framework that separates digital
      object experience, or rendering, from digital object storage and manipulation, so
      the rendering can be tailored to particular communities of users.
    </dc:description>
    <dc:description>Comment: 23 pages including 2 appendices, 8 figures
    </dc:description>
    <dc:date>2001-12-14</dc:date>
    <dc:type>e-print</dc:type>
```

```
<dc:identifizier>http://arXiv.org/abs/cs/0112017</dc:identifizier>
</oai_dc:dc>
</metadaten>
```

In diesem Beispiel kommen die folgenden Elemente von Dublin Core vor, die als Beschreibungen dieses elektronischen Dokument gespeichert wurden: title, creator, subject, description (2 Mal), date, type und identifizier. An diesem Beispiel kann man sehen, dass die Elemente von Dublin Core optional und wiederholbar sind.

### 2.3 OLAC Metadaten (Open Language Archive Community):

Im Vergleich mit dem Dublin Core ist das OLAC[3] Metadatenformat ein alternativer Standard. Das **OLAC Metadata Format** wird von OLAC verwendet, um Sprachressourcen zu beschreiben und die zugehörigen Dienstleistungen anzubieten. OLAC verwendet das XML-Format, um Sprachressourcen auszutauschen.

Das OLAC Metadata Set basiert auf dem Dublin Core Metadata Set und verwendet alle 15 Elemente von DC. Um größere Präzision in der Beschreibung der Ressource für andere nutzbar zu machen, folgt OLAC der DC-Empfehlung für qualifizierende Elemente mittels „element refinements“ oder „encoding schemas“ (**DCQ - Qualified Dublin Core**). OLAC ist ein Anwendungsprofil, das die Elemente von den Basic DC und DC Qualifiers verwendet.

Ein Element von DC Qualifiers kann eine Verfeinerung (verwendet ein in dcterms definiertes Element) oder ein Kodierungsschema (verwendet ein in dcterms definiertes Schema als den Wert des Attributes xsi:type) oder beide spezifizieren. „dcterms“ namespace des Metadata Records muss so deklariert werden:

```
xmlns:dcterms=http://purl.org/dc/terms/
<dcterms:created xsi:type="dcterms:W3C-DTF">2002-11-28</dcterms:created>
```

Das Attribut „xsi:type“ ist eine Methode, die im XML-Schema-Standard [ XMLS ] verwendet wird, um den Typ des aktuellen Elements durch ihren Wert zu überladen. In diesem Beispiel wurde das Datum in einem durch W3C-DTF definierten Format spezifiziert.

### 2.4 MARC 21 Concise Format for Bibliographic Data:

**MARC 21 Concise Format**[4] ist ein weit verwendeter Standard für die Representation und den Austausch von Berechtigung, Bibliographie, Klassifikation, Gemeinschaftsinformationen. Es besteht aus einer Familie von 5 koordinierten Formaten:

- MARC 21 Format for Authority Data
- MARC 21 Format for Bibliographic Data

- MARC 21 Format for Classification Data
- MARC 21 Format for Community Information
- MARC 21 Format for Holdings Data

**MARC 21 Format for Bibliographic Data** wurde als Träger von bibliographischer Information über das gedruckte und Text-Manuskript, Dateien, Diagramme, Musik, visuelles und gemischtes Material entworfen. Bibliographische Daten umfassen allgemein title, names, subjects, notes, publication data und Information über die Beschreibung eines Einzelteils. Das bibliographische Format enthält Datenelemente für die folgenden Arten von Material:

Bücher, weitere Ressourcen, Computer Datei, Karte, Musik, Tonaufnahmen, visuelles Material, gemischtes Material.

Die folgende Webseite enthält einen Überblick zu den Spezifikationen zu **MARC 21 Format for Bibliographic Data**:

<http://www.loc.gov/marc/bibliographic/ecbdintr.html>

### 3. OAI - PMH

#### (The Open Archive Initiative Protocol for Metadata Harvesting )

Im vorigen Kapitel wurden Metadaten und verschiedene Metadatenformate vorgestellt. Das Protokoll zum Austausch der Metadaten und die detaillierte Prozedur des Austausches werden in diesem Kapitel dargestellt.

#### 3.1 Was ist OAI-PMH?

*“The Open Archives Initiative Protocol for Metadata Harvesting provides an application-independent interoperability framework based on metadata harvesting[5].”*

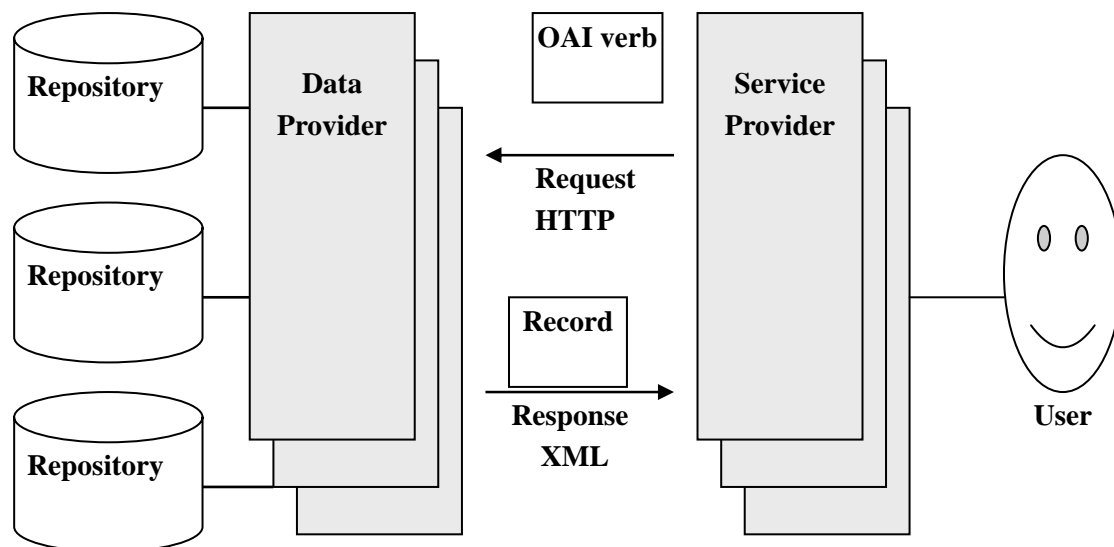
*Zielsetzung* des OAI Metadata Harvesting Protokolls ist es, einen applikationsunabhängigen Rahmen für Interoperabilität zu bieten, der in verschiedenen Fachcommunities für das Auffinden und die weitere Bearbeitung webbasierter wissenschaftlicher Veröffentlichungen anwendbar ist.

Es gibt zwei Arten von Teilnehmern im Rahmen von OAI-PMH:

**Data Providers** - die die Inhalte so aufbereiten, dass ihre Metadaten vom OAI Protokoll erfasst werden können.

**Service Providers** - die verschiedene Dienstleistungen und Mehrwert-Services auf der Basis des OAI-Netzes und enthaltener OAI-Metadaten aufbauen und anbieten.

## Aufbau des OAI-PMH Protokolls:



## 3.2 Wichtige Definitionen:

Da die folgenden Begriffe für das weitere Verständnis wichtig sind, werden sie hier kurz beschrieben.

### 3.2.1 Harvester:

Client-Anwendung, die OAI-PMH-Anfragen stellt. Ein Service Provider enthält in der Regel eine Harvester - Komponente zum Sammeln von Metadaten.

### 3.2.2 Repository:

Netzwerkfähiger Server, welcher OAI-PMH-Anfragen beantworten kann. Um verschiedene Konfiguration des Repositorys zu erlauben, unterscheidet OAI-PMH zwischen drei eindeutigen Entitäten und zwar *resource*, *item* und *record*.

### 3.2.3 Item:

Komponente eines Repository, von welchem Metadaten über eine Ressource zur Verfügung gestellt werden, sie hat einen eindeutigen Identifier.

### 3.2.4 Unique Identifier:

Eindeutiger Schlüssel für ein Item in einem Repository, wird bei OAI-PMH-Anfragen zum Extrahieren verwendet.

„Unique Identifier“ spielt zwei Rollen im Protokoll:

Response: Identifier werden durch die Anfrage **ListIdentifiers** und **ListRecords** zurückgegeben.

Request: Identifier in Verbindung mit einem **metadataPrefix** werden in der Anfrage **GetRecord** zur Identifikation bei der Abfrage eines Records in einem spezifischen Metadataformat von einem Item benutzt.

### 3.2.5 **Record:**

Die Metadaten, die in einem einzelnen spezifischen Metadatenformat ausgedrückt werden. Ein Record wird eindeutig durch die Kombination des eindeutigen Identifiers des Items und des Metadataformats und des Datestamps des Records gekennzeichnet. Die XML-Kodierung des Records besteht aus folgenden Teilen:

- **header (notwendig):** enthält den eindeutigen Bezeichner des Items und die Eigenschaften, die für selektives Harvesten notwendig sind.
  - Identifier (1)
  - Datestamp (1)
  - setSpec Elemente (\*)
  - status attribute (optional)
- **metadata (notwendig):** das einzelne Format von Metadata eines Items. Das OAI-PMH unterstützt mehrere Formate von Metadaten. Aber das Repository muss mindestens in der Lage sein, den Record mit Metadaten in einfachem Dublin Core zurückzuliefern.
- **about (optional):** ein optionaler und wiederholbarer Behälter für Daten über den Metadaten Teil des Records.
  - Rights Angaben: Angaben zu Rechten.
  - Provenance Angaben: Angaben zur Herkunft der Metadaten.

Beispiel: OAI Record

```
<record>
  <header>
    <identifier>oai:physnet.de:tut1</identifier>
    <datestamp>2003-03-27</datestamp>
    <setSpec>tut</setSpec>
  </header>
  <metadata>
    <oai_dc>
      <title>OAI Tutorial at OA Forum III</title>
      <creator>Heinrich Stamerjohanns</creator>
      <language>ger</language>
```

```

    </oai_dc>
  </metadata>
  <about>
    <metadataID>oai:physnet.de:tut1md</metadataID>
  </about>
</record>

```

### 3.2.6 **Set:**

Ein optionales Konstrukt, um Items in einem Repository anzuordnen. Wenn ein Repository eine Set-Struktur definiert, müssen die gesetzten Zugehörigkeitsinformationen im Header des Items enthalten sein, die bei der Anfrage von *ListIdentifiers*, *ListRecords* und *GetRecord* zurückgeliefert werden.

### 3.2.7 **Selective Harvesting:**

Selective Harvesting erlaubt einem Harvester, die Anfrage auf Teile der Metadaten zu begrenzen, die im Repository vorhanden sind. Das OAI-PMH unterstützt Harvester mit zwei Arten von Kriterien, die in einer Anfrage kombiniert werden können: *datestamps* und *set membership*.

## 3.3 Protocol Features:

### 3.3.1 HTTP Anfrageformat:

OAI-PMH-Anfragen müssen mit der Benutzung der HTTP-„*Get*“ oder „*Post*“ Methoden geschickt werden. Repositories müssen beide Methoden unterstützen. Die „*Post*“-Methode hat den Vorteil, dass sie keine Beschränkungen der Länge der Argumente hat. Es gibt eine einzige Basis-URL für alle Anfragen. Die Basis-URL spezifiziert „*host*“ und „*port*“. Repositories beschreiben ihre Basis-URL als Wert des „*baseUrl*“ Elements bei der Antwort der Anfrage „*Identify*“.

Zusätzlich zur Basis-URL bestehen alle Anfragen aus einer Liste von Schlüsselargumenten. Jede OAI-PMH-Anfrage hat mindestens ein *key=value* Paar, wobei *key='verb'*, *value=AnfrageTyp (s.u.)* sind. Weitere *key=value* Paare hängen von der Anfrage ab (mehrere Argumente sollten durch & getrennt werden).

z.B: [http://an.oa.org/OAI-script?](http://an.oa.org/OAI-script?verb=GetRecord&identifier=oai%3AarXiv.org%3Ahep-th%2F9901001&metadataPrefix=oai_dc)

```
verb=GetRecord&identifier=oai%3AarXiv.org%3Ahep-th%2F9901001&metadataPrefix=oai_dc
```

In diesem kleinen Beispiel wird ein Typ von Anfrage benutzt und zwar *GetRecord*, wobei die Identifizierung eindeutig gegeben ist.

### 3.3.2 HTTP Antwortformat:

Antworten werden als HTTP-Antworten, mit passendem HTTP-Header formatiert.

*Content type* muss „text/xml“ sein. Status Codes unterscheiden sich von OAI-PMH Fehlern. Z.B.:302(*redirect*), 503(*service unavailable*). Kompression ist optional bei OAI-PMH.

### 3.3.3 XML Antwortformat:

Das Antwortformat ist wohlgeformtes XML mit Markup, das die folgenden Teile enthält:

1. XML Declaration

(z.B.:<? xml version="1.0" encoding="UTF-8" ?>)

2. Root Element namens OAI-PMH mit drei Attributen

(xmlns,xml:xsi,xsi:schemaLocation)

3. drei Kind-Elemente

- **responseDate** (UTC datetime)

- **request** (Anfrage, die diese Antwort generiert)

- a) **error** (im Fehlerfalle)

b) Element mit dem Namen der OAI-PMH-Anfrage

## 3.4 Implementierung des Protokolls

Es folgt eine kurze Einführung zur Implementierung des OAI-Protokolls. Eine detaillierte Beschreibung, wie das OAI-Protokoll umzusetzen ist, kann unter folgender URL gefunden werden:

<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>

Das OAI-Protokoll erfordert die Implementierung von insgesamt 6 Funktionen: *Identify*, *ListSets*, *ListMetadataFormats*, *ListRecords*, *ListIdentifiers* und *GetRecord*.

### **Identify:**

- Funktion:

Liefert allgemeine Informationen über das Repository

- Parameter:

Keine

- Beispiel-URL:

<http://memory.loc.gov/cgi-bin/oai?verb=Identifier>

- mögliche Fehler:

- badArgument: Diese Anfrage enthält unzulässige Parameter.

### **ListMetadataFormats:**

- Funktion:

Gibt die DTDs der vom Data Provider verwendeten Metadata Formats zurück.

- Parameter:

Identifizier: für einen spezifischen Record (optional)

- Beispiel-URL:

<http://www.perseus.tufts.edu/cgi-bin/pdataprov?verb=ListMetadataFormats&identifier=oai:perseus.tufts.edu:Perseus:text:1999.02.0119>

- mögliche Fehler:

- badArgument: Diese Anfrage enthält unzulässige Parameter.
- idDoesNotExist: Identifizier ist unbekannt oder unzulässig in diesem Repository.
- noMetadataFormats: Es gibt keine Metadatenformate, die für das spezifizierte Item vorhanden sind

### **ListSets:**

- Funktion:

Gibt hierarchisch alle beim Data Provider definierten Sets zurück

- Parameter:

resumptionToken: Flusskontrolle (exclusive)

- Beispiel-URL:

<http://an.oa.org/OAI-script?verb=ListSets>

- mögliche Fehler:

- badArgument: Diese Anfrage enthält unzulässige Parameter.
- badResumptionToken: Der Wert von resumptionToken Argument ist unzulässig oder abgelaufen.
- noSetHierarchy: Dieses Repository unterstützt keine Sets.

### **ListRecords:**

- Funktion:

gibt eine Liste mehrerer Records zurück

- Parameter:

- from: Startdatum (optional)

- until: Enddatum (optional)
- metadataPrefix: Metadatenformat, für welches die Identifier gelistet werden sollen (required)
- set: Menge aus welcher eingesammelt werden soll (optional)
- resumptionToken: Flusskontrolle (exclusive)
- Beispiel-URL:  
[http://physnet.de/oai/oai2.php?verb=ListRecord&metadataPrefix=oai\\_dc&from=2001-01-01](http://physnet.de/oai/oai2.php?verb=ListRecord&metadataPrefix=oai_dc&from=2001-01-01)
- mögliche Fehler:
  - badArgument: Diese Anfrage enthält unzulässige Parameter.
  - badResumptionToken: Der Wert von resumptionToken Argument ist unzulässig oder abgelaufen.
  - cannotDisseminateFormat: Der Wert von „metadataPrefix“ wird nicht vom Repository unterstützt.
  - noRecordsMatch: Die Kombination der Parameter ergibt eine leere Liste.
  - noSetHierarchy: Dieses Repository unterstützt keine Sets.

### **ListIdentifiers:**

- Funktion:

Listet die Header für alle Records, die den Parametern entsprechen, auf.
- Parameter:
  - from: Startdatum (optional)
  - until: Enddatum (optional)
  - metadataPrefix: Metadatenformat, für welches die Identifier gelistet werden sollen (required)
  - set: Menge von welche eingesammelt werden soll (optional)
  - resumptionToken: Flusskontrolle (exclusive)
- Beispiel-URL:  
<http://an.oa.org/OAI-script?verb=ListIdentifiers&from=1998-01-15&metadataPrefix=oldArXiv&set=physics:hep>
- mögliche Fehler:
  - badArgument: Diese Anfrage enthält unzulässige Parameter.
  - badResumptionToken: Der Wert von resumptionToken Argument ist unzulässig oder abgelaufen.

- cannotDisseminateFormat: Der Wert von „metadataPrefix“ wird nicht vom Repository unterstützt.
- noRecordsMatch: Die Kombination der Parameter ergibt eine leere Liste.
- noSetHierarchy: Dieses Repository unterstützt keine Sets.

### **GetRecord:**

#### - Funktion:

GetRecord wird verwendet, um einen einzelnen Metadaten-Record aus einem Repository zu holen.

#### - Parameter:

- Identifier: eindeutige ID für den Record(required)
- metadataPrefix: Metadatenformat(required)

#### - Beispiel-URL:

[http://arXiv.org/oai2?verb=GetRecord&identifier=oai:arXiv.org:cs/0112017&metadataPrefix=oai\\_dc](http://arXiv.org/oai2?verb=GetRecord&identifier=oai:arXiv.org:cs/0112017&metadataPrefix=oai_dc)

#### - mögliche Fehler:

- badArgument: Diese Anfrage enthält unzulässige Parameter.
- cannotDisseminateFormat: Der Wert von „metadataPrefix“ wird nicht vom Repository unterstützt.
- idDoesNotExist: Die id existiert nicht.

## **4. Aussichten: Was kann man von OAI in der Zukunft erwarten?**

Welche Mehrwertdienste können auf dem Protokoll und der einfachen Abfrage sinnvoll aufgebaut werden? Weitere Fragen stellen sich ebenfalls, die durch Erfahrungswerte im Laufe der Jahre gelöst oder beobachtet werden können. Wie vertragen sich z.B. verschiedene Metadatensätze miteinander?

Im Rahmen der ersten Anwendungen des OAI-Metadaten Harvesting Protokolls müssen auch Fragen nach dem geistigen Eigentum und nach den Autorenrechten untersucht werden. Wenn mehrere Metadatensätze von verschiedenen Stellen erhältlich sind, wo und inwieweit wird ein Metadatensatz bevorzugt, oder haben sie gegebenenfalls in verschiedenen wissenschaftlichen Communities unterschiedliche Stellenwerte? Im Gegensatz zur bibliothekarischen Katalogisierung nach bestimmten Regelwerken, können Metadaten als geistiges Eigentum erkannt werden. Wie konkurrieren verschiedene Metadatensätze miteinander? Wie wird eine solche Frage gerade bei einem Rechercheprozess über das OAI-Protokoll abgewickelt? Könnten Kosten und Urheberrechtsfragen zwischen Abfrage und Anzeige der ersten

Ergebnisse ein großes Hindernis sein? Indem das OAI-Protokoll das „About“-Feld für Informationen über die Metadaten selbst definiert hat, besteht eine Möglichkeit, auch diese Rechte sowie die Zugangsrechte zu den Servern darzustellen.

## 5. Wichtige Ressourcen:

[1] Fallside (ed.), D.C., *XML Schema Part 0: Primer*. 2000, World Wide Web Consortium.

<http://www.w3.org/TR/xmlschema-0/>

[2] Dublin Core Metadata Initiative (DCMI):

<http://dublincore.org/>

[3] OLAC Metadata (Open Language Archive Community):

<http://www.language-archives.org/OLAC/metadata.html>

[4] MARC 21 Concise Format:

<http://www.loc.gov/marc/bibliographic/ecbdhome.html>

[5] OAI web site und OAI-PMH specification:

<http://www.openarchives.org/>

[6] Diann Rusch-Feja : Die Open Archives Initiative (OAI)

[http://www.bibliothek-saur.de/2001\\_3/291-300.pdf](http://www.bibliothek-saur.de/2001_3/291-300.pdf)