

## **„Das Open Archives Initiative Protocol for Metadata Harvesting: Zielsetzung, Funktionalität, Einsatzgebiete“**

Authors: Katja Diederichs (<https://orcid.org/0000-0003-3409-9902>) und Ulrike Wuttke (<https://orcid.org/0000-0002-8217-4025>) (2018).

DOI: 10.5281/zenodo.1253735.

This paper is licensed under a Creative Commons ShareAlike 4.0 International license (<https://creativecommons.org/licenses/by-sa/4.0/>).

Grundlage dieser Publikation ist eine Hausarbeit (term paper), die im Rahmen des Masters Library and Information Science (TH Köln) erstellt wurde (Summer Term 2017).

Part of the original term paper was a poster (in English) with the title: „OAI-PMH and its role for scholarly communication: Past, Present, Future“ by Katja Diederichs und Ulrike Wuttke. The poster can be downloaded in a slightly modified version from Zenodo: DOI: 10.5281/zenodo.1253731.



## **Inhaltsverzeichnis**

1. Einführung	2
2. Hintergrund OAI-PMH	2
2.1. Die Open Archives Initiative	2
2.2 OAI-PMH	2
3. Konzept und Funktionalität des OAI-PMH	3
3.1 Das OAI-PMH-Konzept	3
3.2 Protokollspezifikationen und technische Eigenschaften	4
3.3 Metadatenbeschreibung	5
3.4 Infrastruktur des implementierten OAI-PMH	7
4. Einsatzgebiete	8
5. Zukunft	10
Bibliographie	13
Anhang	15
Glossar	15

# 1. Einführung

In der vorliegenden Hausarbeit beschreiben die Autorinnen das Open Archives Protocol for Metadata Harvesting (kurz: OAI-PMH). Nach der Darstellung des historischen Hintergrunds, inklusive der Zielsetzung von OAI-PMH, gehen die Autorinnen auf die Funktionalität und die Einsatzgebiete ein und geben abschließend einen Ausblick auf mögliche zukünftige Entwicklungen. Ein besonderes Augenmerk liegt auf dem Stellenwert von OAI-PMH für bibliothekarische Dienstleistungen.

## 2. Hintergrund OAI-PMH

### 2.1. Die Open Archives Initiative

Die Open Archives Initiative<sup>1</sup> (kurz OAI) ist eine Initiative von Organisationen, die Server für elektronische Publikationen (wie z. B. Preprints, Postprints) oder andere Dokumentenserver betreiben. Sie wurde mit dem Ziel gegründet, im Internet frei veröffentlichte wissenschaftliche elektronische Publikationen durch die Entwicklung technologieunabhängiger und interoperabler Standards besser auffindbar und besser nachnutzbar zu machen. Über den anfänglich auf wissenschaftliche Inhalte bezogenen Fokus hinaus steht die Arbeit der OAI nun allgemein für die Öffnung des Zugangs zu einer breiten Palette unterschiedlicher digitaler Materialien, unabhängig von ihrem Inhalt oder eventuellen wirtschaftlichen Implikationen.<sup>2</sup> Der Kern ihrer Arbeit ist jedoch weiterhin der freie Zugang zu wissenschaftlichen Publikationen.<sup>3</sup>

### 2.2 OAI-PMH

Das Open Archives Initiative Protocol for Metadata Harvesting (im Folgenden kurz OAI-PMH) wurde von der OAI als interoperables Standardprotokoll entwickelt, um der freien Wissenschaftskommunikation durch die Verbesserung der Auffindbarkeit von in unterschiedlichen Repositorien befindlichen elektronischen Publikationen (im Folgenden kurz E-Prints) mittels Metadatenaustausch den Weg zu bahnen.<sup>4</sup>

Die Anfänge der OAI und des OAI-PMH liegen in der seit Anfang der neunziger Jahre wachsenden Verbreitung von E-Prints, im Sinne von durch die Autorinnen und Autoren auf E-Print-Repositorien von Institutionen (den sogenannten Repositorien oder Archiven) selbst archivierten, digitalen

---

<sup>1</sup> "OAI | Homepage", letzter Zugriff 25.07.2017.

<sup>2</sup> Siehe "OAI | Organization", letzter Zugriff 25.07.2017.

<sup>3</sup> Siehe "OAI | Homepage", letzter Zugriff 25.07.2017.

<sup>4</sup> Siehe Van de Sompel & Nelson (2015), Kap. 2 The Open Archives Initiative Protocol for Metadata Harvesting – OAI-PMH (1999).

wissenschaftlichen Publikationen. Durch diese Praxis wurde es möglich, aktuelle Forschungsergebnisse zeitnah und unabhängig von Peer-Review-Prozessen oder Zeitschriftenveröffentlichungen ohne (kommerzielle) Zugangsbeschränkung an die Wissenschaftsgemeinde zu kommunizieren, entweder vor der Veröffentlichung (*pre-print*) oder im Nachhinein.<sup>5</sup>

Der Zugriff auf die Inhalte der Repositorien war per Internet über unterschiedliche Datenbankoberflächen und die jeweiligen Findmittel der einzelnen Repositorien, sowie später bestenfalls über verschiedene Web-Interfaces möglich. Dies war aus Nutzersicht problematisch, da eine Nutzung bzw. Suche die Kenntnis und Anwendung dementsprechend verschiedener Protokolle voraussetzte. Daher fand 1999 erstmals ein Treffen in Santa Fe mit dem Ziel statt, interoperable Standards und ein technologisches Framework für die bessere Auffindbarkeit von E-Prints in Repositorien und die bessere Verbreitung und den Austausch von Informationen über E-Prints zu entwickeln.<sup>6</sup> Das sogenannte Santa Fe-Treffen wurde zur Geburtsstunde des OAI-PMH.

### 3. Konzept und Funktionalität des OAI-PMH

In diesem Kapitel wird das Konzept sowie die Funktionalität des OAI-PMH näher beleuchtet.<sup>7</sup> Zunächst werden das OAI-PMH-Konzept (Kap. 3.1) sowie die darauf basierenden Protokollspezifikationen und technischen Eigenschaften beschrieben (Kap. 3.2). Daraufhin werden die Anforderungen an die Metadatenbeschreibung für die sinnvolle Austauschbarkeit von Inhalten aufgezeigt (Kap. 3.3). Abschließend wird eine beispielhafte Infrastruktur des implementierten OAI-PMH skizziert (Kap. 3.4).

#### 3.1 Das OAI-PMH-Konzept

Die Frage, ob mehrere verteilte Repositorien (Data Provider) mittels eines Suchprotokolls wie Z39.50<sup>8</sup> für jede einzelne Suchanfrage übergreifend durchsucht werden sollen, oder erst die Daten verschiedener Provider geharvestet und in einem Datenspeicher zusammentragen und für Suchen aufbereitet und bereitgestellt werden sollen (Service Provider), wurde bei der Entwicklung des OAI-PMH zugunsten des Harvesting-Ansatzes entschieden, weil zum damaligen Zeitpunkt übergreifende

---

<sup>5</sup> Siehe "Open Archives Forum | Tutorial: 2. History and development of OAI-PMH", letzter Zugriff 25.07.2017.

<sup>6</sup> Siehe "Open Archives Forum | Tutorial: 2. History and development of OAI-PMH", Letzter Zugriff 25.07.2017.

<sup>7</sup> Einige zum besseren Verständnis des OAI-PMH dienliche Definitionen werden im Glossar (siehe Anhang) näher erläutert: Harvesting, Archive, Protocol, Metadata, Repository, Data Provider, Service Provider, Aggregator, Interoperability.

<sup>8</sup> Z39.50 ist ein 1984 in den USA unter Beteiligung der Library of Congress entwickeltes, noch heute genutztes Netzwerkprotokoll aus dem Bereich des Bibliothekswesens. Es wird u. a. von Literaturverwaltungsprogrammen genutzt und für den Abruf von Daten aus Bibliothekskatalogen. Siehe Kostädt (2015), S. 515 und Lauber-Reymann (2017), S. 26.

verteilte Suchen sehr zeitaufwendig bzw. schlecht-skalierbar waren und z. B. vom jeweils langsamsten Server ausgebremst wurden.<sup>9</sup>

Um die Aufbewahrung und Suche zu gewährleisten, wurde für eine zweiteilige logische Provider-Struktur, bestehend aus Data Provider und Service Provider, optiert. Das grundlegende Prinzip dieser Struktur ist es, dass Data Provider für die Archivierung und Speicherung der digitalen Ressourcen sowie das Anbieten technisch frei zugänglicher Metadatenbeschreibungen verantwortlich sind. Die Service Provider fragen diese Metadatenbestände in regelmäßigen Abständen ab, sammeln sie ein (Harvesting) und bieten sie anschließend mittels eines Services an Dritte an, z. B. als Suchinterface oder zum Peer-Review. Hiermit war ein wichtiger Paradigmenwechsel vollzogen: von menschnennutzbaren zu maschinennutzbaren Daten-Repositories-Interfaces mit denen sowohl der Zugriff auf maschinenlesbare strukturierte Daten als auch deren Austausch zwischen verschiedenen Anwendungen möglich wurde.<sup>10</sup> Menschennutzbare Interfaces, mit deren Hilfe Nutzer z. B. die Daten durchsuchen können, blieben ebenfalls erhalten und wurden als Komponente der Service Provider konzipiert.<sup>11</sup>

## 3.2 Protokollspezifikationen und technische Eigenschaften

Das OAI-PMH liegt zurzeit in der 2002 zuletzt aktualisierten, stabilen *Protocol Version 2.0*<sup>12</sup> vor. Die folgende Abbildung (Abb. 1) verdeutlicht die technische Umsetzung des Harvesting per OAI-PMH. Für Anfragen des Service Providers an die Data Provider werden die HTTP-Methoden GET und POST genutzt. Genauer gesagt werden bei Harvesting-Anfragen sechs auf OAI-PMH beruhende Methoden bzw. Verben genutzt, um die strukturierten Metadaten gezielt abzufragen.<sup>13</sup> Die Antworten werden wiederum in XML beschrieben zurückgesendet.<sup>14</sup> In beide Richtungen werden die Kommunikationsinhalte per HTTP (Hypertext Transfer Protocol), einem für die Nutzung im www (Word Wide Web) weit verbreiteten, allgemeinen Datenübertragungsprotokoll, transportiert. Die geharvesteten Inhalte können schließlich vom Nutzer mittels eines vom Service Provider bereitgestellten Interfaces und Mehrwert-Dienstes durchsucht werden.

---

<sup>9</sup> Siehe "Open Archives Forum | Tutorial: 2. History and development of OAI-PMH", letzter Zugriff 25.07.2017.

<sup>10</sup> Siehe Borst (2014), S. 257-258.

<sup>11</sup> Siehe "Open Archives Forum | Tutorial: 2. History and development of OAI-PMH", letzter Zugriff 25.07.2017.

<sup>12</sup> "OAI | The Open Archives Initiative Protocol for Metadata Harvesting V. 2.0", letzter Zugriff 25.07.2017.

<sup>13</sup> Ebd.

<sup>14</sup> Ebd.

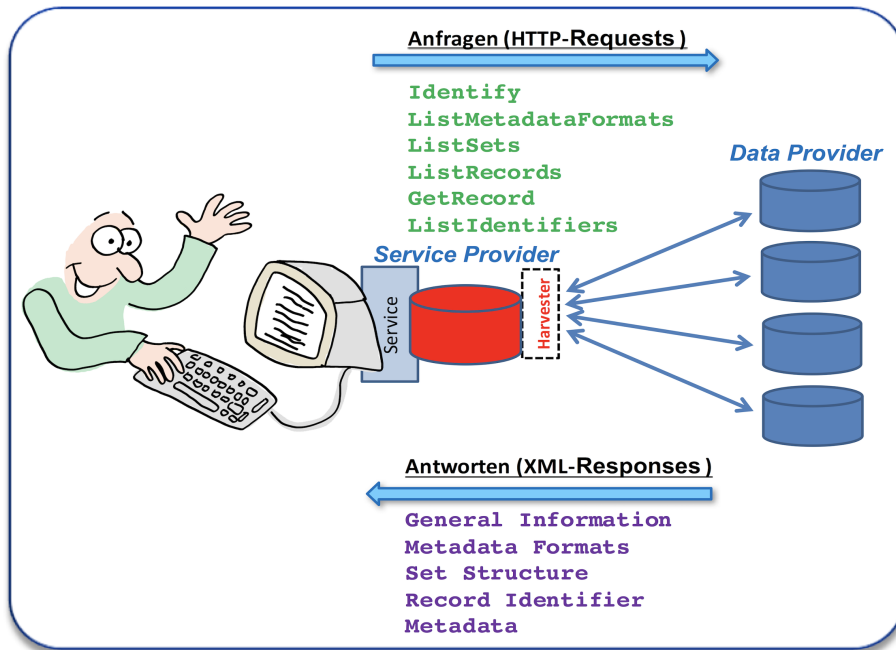


Abb. 1: Harvesting per OAI-PMH

### 3.3 Metadatenbeschreibung

Wie obenstehend (Kap. 3.2) beschrieben, werden nach einer Abfrage per OAI-PMH-Verben generelle Informationen zu den angefragten Ressourcen in XML-kodiert geliefert. Abb. 2 veranschaulicht das Aussehen des Codes eines solchen XML-Dokuments. Es wird ausschnittsweise gezeigt, wie insbesondere der Bereich **Metadata** zu einem sogenannten **Record** aussehen kann.

```

<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. (...)
    </dc:description>
    <dc:description>Comment: 23 pages including
      2 appendices, 8 figures</dc:description>
    <dc:date>2001-12-14</dc:date>
  </oai_dc:dc>
</metadata>

```

Abb. 2: Ausschnitt aus einem XML-Dokument: Dublin Core Beschreibung der bibliothekarischen Metadaten

Im Gegensatz zum Vorläufer der aktuellen Version 2.0 des OAI-PMH können seitens der Data Provider digitale Ressourcen beschrieben werden, nicht wie zuvor nur dokumentartige Objekte.

Die Beschreibung einer Ressource soll laut den Empfehlungen des OAI-PMH durch den Metadatenstandard Dublin Core (kurz: DC), mindestens DC-Simple (oder Unqualified), erfolgen, Erweiterungen mit zusätzlichen Daten sind möglich.<sup>15</sup> Das bedeutet, dass bis zu 15 mögliche DC-Elemente des Elementsets den Record, d. h. die laut DC-Standard strukturierten Metadaten einer Ressource, konstituieren.<sup>16</sup> Der Record besitzt Eigenschaften (Properties) und ihnen zugewiesene Werte (Values).<sup>17</sup> Weil DC sich dadurch auszeichnet, dass per *Identifizier* auf so genannte *Ressourcen* verwiesen wird, und eine Ressource im Sinne des DC alles sein kann, was eine Identität besitzt, solange es über einen eindeutigen Identifizier benannt und beschrieben werden kann, können Ressourcen im Sinne des OAI-PMH-Konzepts nicht nur elektronische Dokumente, sondern auch Bilder, Web-Services oder Sammlungen verschiedener Ressourcen sein. Es ist auch keine Bedingung, dass eine Ressource in einem Netzwerk digital auffindbar sein muss, da sich der Begriff ebenfalls auf analoge Objekte, Menschen oder Institutionen beziehen kann, d. h. auch ein gebundenes Buch wie eine Gutenberg-Bibel kann eine Ressource darstellen.<sup>18</sup>

Dublin Core ist ein Metadatenstandard zur digitalen Beschreibung von Internetressourcen unterschiedlichster Art. Daher hat er sich international sowie disziplinübergreifend etabliert, d.h. wird genutzt von z. B. Museen, Archive, Behörden, in der Wirtschaft etc.,<sup>19</sup> und wird daher auch in der Wissenschafts- und Bibliothekswelt als digitaler Standard für bibliografische Metadaten genutzt. Wie in Abb. 2 verdeutlicht, werden anhand der 15 möglichen und wiederholbaren DC-Elemente inhaltliche Beschreibungen zu einer (bibliothekarischen) Ressource geliefert. Hierbei können formale Inhalte erschlossen werden, wie der Titel, Autor oder das Datum einer Ressource, aber auch eine sachliche Erschließung ist möglich, indem beispielsweise der Inhalt beschrieben oder kommentiert wird. Die Metadaten der strukturiert nach DC beschriebenen Inhalte können somit per OAI-PMH ausgetauscht, genutzt und angeboten werden, um eine effektive Repositorien-übergreifende Suche zu gewährleisten.<sup>20</sup>

---

<sup>15</sup> Siehe "Forschungsdaten.org | OAI-PMH", letzter Zugriff 25.07.2017.

<sup>16</sup> Siehe "Dublin Core Metadata Initiative | Dublin Core Metadata Element Set, Version 1.1", letzter Zugriff 25.07.2017.

<sup>17</sup> Siehe "Dublin Core Metadata Initiative | Guidelines for implementing Dublin Core in XML", letzter Zugriff 25.07.2017.

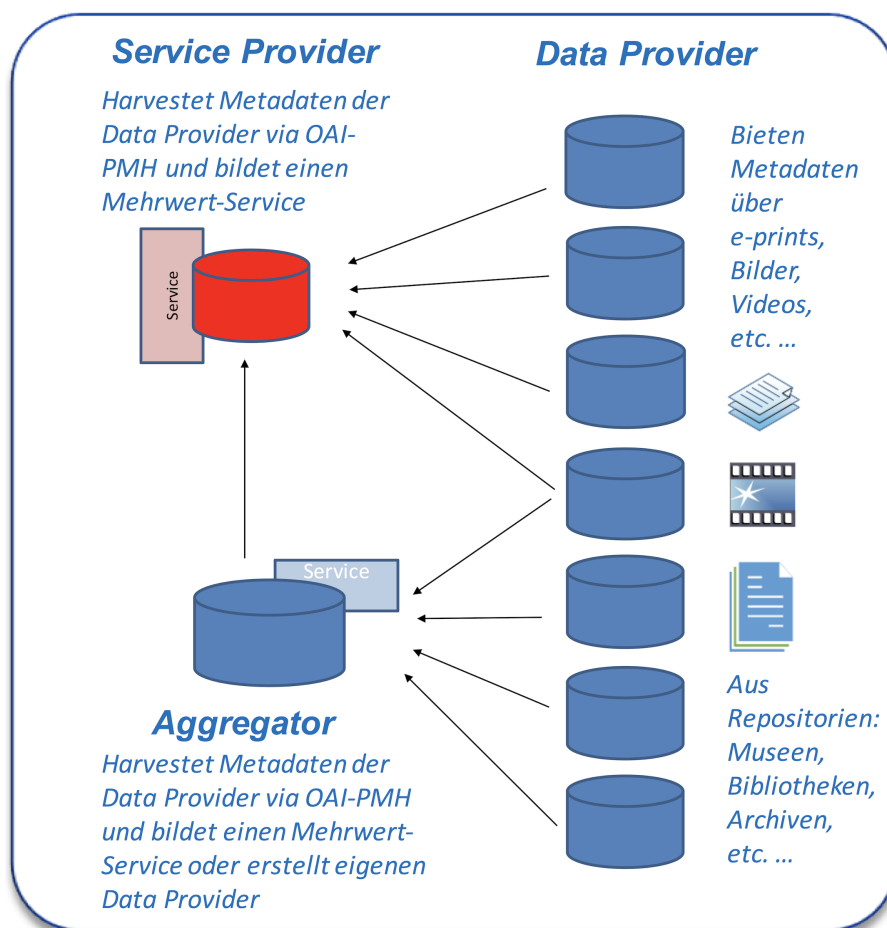
<sup>18</sup> Ebd.

<sup>19</sup> Siehe Rühle (o.J.), S. 10.

<sup>20</sup> Siehe "Dublin Core Metadata Initiative | Background" für vier aufsteigende Level von Interoperabilität, DC befähigt zum ersten Interoperabilitätslevel (*Level 1: Shared term definition*).

### 3.4 Infrastruktur des implementierten OAI-PMH

Die *Implementation Guidelines* (Richtlinien zur Implementierung) des OAI-PMH sind ergänzend zu den Spezifikationen des Protokolls zu nutzen.<sup>21</sup> In ihnen wird beschrieben, wie Repositorien oder Services zu implementieren sind, um gemäß den OAI-Standards miteinander interoperabel zu kommunizieren.



**Abb. 3: OAI-PMH gestütztes Framework**

Die obenstehende Abbildung (Abb. 3) stellt schematisch ein niedrigschwelliges, flexibles und leicht zu implementierendes OAI-PMH basiertes Framework für den interoperablen Austausch strukturierter Metadaten zu verteilten Ressourcen dar. Repositorien (Data Provider) bieten ihre Metadaten für Aggregatoren und Service Provider zum Harvesting an.<sup>22</sup> Die Service Provider (oder Aggregatoren mit Service-Komponente) bieten dabei nutzerseitig Mehrwert-Dienste zu den von verschiedenen Aggregatoren oder Data Providern eingesammelten Daten an.

<sup>21</sup> Siehe "OAI | Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting".

<sup>22</sup> Ein typisches Beispiel für einen Data Provider ist zum Beispiel *res doctae* (erreichbar über <https://rep.adw-goe.de/services>), das Publikationsrepositorium der Göttinger Akademie der Wissenschaften, das DINI-zertifiziert ist und somit über eine OAI-PMH Schnittstelle verfügt.



## 4. Einsatzgebiete

OAI-PMH hat sich inzwischen als grundlegender Standard für Repositorien durchgesetzt und zählt somit in Folge der zunehmenden Rolle von Repositorien für die Arbeit wissenschaftlicher Bibliotheken, z. B. als Anbieter bzw. Betreiber institutioneller Open Access-Repositorien, zu den bibliothekarischen Standardanwendungen im Bereich der digitalen Bibliothek. Ein Grund für die steigende Popularität des OAI-PMH war nicht zuletzt seine enge Verbindung und Unterstützung der aufkommenden Open Access-Bewegung, die sich durch die Budapester Erklärung (2002)<sup>23</sup> und weitere selbstverpflichtende Erklärungen für den offenen Zugang zu wissenschaftlichen Inhalten, ihren Weg durch die Wissenschaft und darüber hinaus bahnte.<sup>24</sup>

Durch die generelle Unterstützung des OAI-PMH-Protokolls im Bereich aller gängigen Repositorien-Systeme als *de facto* Standard ist momentan in diesem Bereich eine breite Grundlage für Interoperabilität, Austausch und Wiederverwendung von strukturierten Metadaten gelegt.<sup>25</sup> So schreibt zum Beispiel die bundesdeutsche Initiative für Netzwerkinformation (DINI e.V.) OAI-PMH im Rahmen ihres Zertifikats für Repositorien verbindlich vor<sup>26</sup> und ist OAI-PMH momentan die Grundlage für übergreifende nutzergerichtete Systeme zum Auffinden digitaler akademischer Ressourcen durch das Einsammeln von Metadaten, wie z. B. durch den Service Provider und Aggregator BASE (Bielefeld Academic Search Engine)<sup>27</sup>.

Der Vorteil des offenen OAI-Protokolls ist es, dass er von Anfang an auf bekannten und weit verbreiteten Standards beruht:

“Fest definierte Beschreibungssprachen, die ein Objekt beschreiben (Metadatenformate), werden in der eXtensible Markup Language (XML) kodiert und dann über das vom World Wide Web bekannten Hypertext Transfer Protocol (HTTP) über das Internet transportiert. Dabei können Metadaten in beliebigen Metadatenformaten ausgetauscht werden. Aus Gründen der Interoperabilität wird jedoch das bekannte und sehr einfache Metadatenformat Dublin Core als kleinster gemeinsamer Nenner von allen Data Providern unterstützt.”<sup>28</sup>

Das OAI-PMH 2.0 erfuhr wegen seiner Niederschwelligkeit und einfachen Nutzbarkeit große Popularität und wird bis in die heutige Zeit nach denselben Prinzipien und Spezifikationen genutzt. Trotz seiner Vorteile sind mit dem OAI-PMH auch Nachteile verbunden, vor allem der “Fluch”<sup>29</sup> des Minimalkonsens DC-Simple und sein im Grunde genommen nicht mehr zeitgemäßes Grundkonzept.

Bei der Entwicklung von OAI-PMH wurde bewusst für DC-Simple als Metadatenformat optiert, weil es sich dabei um einen relativ niedrighschwelligem bzw. leicht erfüllbaren Standard handelt.

---

<sup>23</sup> “Budapest Open Access Initiative | Homepage”, letzter Zugriff 25.07.2017.

<sup>24</sup> Siehe Van de Sompel & Nelson (2015), Kap. Introduction.

<sup>25</sup> Siehe Borst (2014), S. 257-258.

<sup>26</sup> Siehe DINI (2016).

<sup>27</sup> “BASE | Homepage”.

<sup>28</sup> DINI (2016), S. 6.

<sup>29</sup> Siehe “OAN | DINI-OAI-Validator verfügbar”, letzter Zugriff 25.07.2017.

DC-Simple verfügt zwar über normierte Felder, die Ausfüllung der Inhalte der Felder ist jedoch kaum normiert, d. h. sie ist sehr offen und unterliegt nur fachlichen, nicht aber formalen Vorgaben.<sup>30</sup>

“Angefangen von der einfachen Nennung des Autors, finden sich in den dc:author-Feldern beispielsweise Angaben wie ‚Vorname Nachname‘, ‚Nachname, Vorname‘ bis hinzu ‚Nachname Vorname‘. Das ganze dann noch kombiniert mit Kürzeln, Titeln und zusätzlichen Vornamen einer Person wird dann maschinell doch zu einer Herausforderung, wenn diese Feldinhalte nicht nur als Ganzes erfasst werden sollen.”<sup>31</sup>

Die hieraus resultierende Heterogenität stellt die Service Provider bei der finalen Homogenisierung der aggregierten Metadaten vor eine große Herausforderung, weshalb aggregatorseitig zusätzliche unterstützende Guidelines für Data Provider geschaffen wurden, wie z. B. das bereits erwähnte DINI-Zertifikat, um durch Vereinheitlichung “einen optimalen maschinellen Dialog zu ermöglichen” sowie Hilfsmittel zur Überprüfung und Steigerung der Metadatenqualität von Repositorien.<sup>32</sup>

Das Grundkonzept von OAI-PMH baut auf einer Metadatensammlung und Speicherung auf und basiert auf einem repositorienbezogenen Ansatz (engl.: *repository centric approach*). Heutzutage haben sich die technischen Möglichkeiten jedoch enorm vergrößert und können Suchen viel größeren Umfangs direkt über verschiedene einzelne Webseiten und Dokumente durchgeführt werden, wie Suchmaschinen und Discovery Services demonstrieren. Um für BASE oder OAIster<sup>33</sup> sichtbar zu sein, müssen die Ressourcen erst in ein Repository geladen werden (das setzt die Einrichtung und den Unterhalt eines Repositoriums voraus) und die Metadaten (insbes. die DC-Simple Records der jeweiligen Ressourcen) eingegeben und gepflegt werden. Kritisch ist bei diesem, dass die Vorteile der Netzstruktur nicht genutzt werden, die grundsätzlich einen flexiblen und granularen Zugriff ermöglichen, und stattdessen (Meta-)datensilos gefördert werden, die lediglich mit OAI-PMH erreichbar und abgreifbar sind.<sup>34</sup> Dazu kommt als zweiter Kritikpunkt, dass keine selbst erstellten Abfragen möglich sind (da OAI-PMH keine Abfragesprache ist) und sich auf Grund verschiedener der Funktionalität des Protokolls inhärenten Eigenschaften ein nicht-selektives Harvesten der Metadaten aller Ressourcen seitens der Aggregatoren und Serviceprovider eingebürgert hat, auf deren Basis erst im zweiten Schritt flexible Suchen und Filterungen ermöglicht werden können.<sup>35</sup> Dieser Weg mag auf den damaligen Technologiestand rückblickend sinnvoll gewesen sein, er ist jedoch auch sehr umständlich.<sup>36</sup> In Anbetracht der technisch in hohem Maße erweiterten Möglichkeiten der Informationsverarbeitung scheint er nicht mehr zeitgemäß zu sein. Auch wird durch OAI-PMH keine Interoperabilität im engeren Sinn einer semantischen Beschreibung gewährleistet, da es keine

---

<sup>30</sup> Ebd.

<sup>31</sup> Ebd.

<sup>32</sup> Ebd., Zitat ebd.

<sup>33</sup> “OCLC | OAIster database”, letzter Zugriff 25.07.2017.

<sup>34</sup> Siehe Borst (2014), S. 258.

<sup>35</sup> Ebd.

<sup>36</sup> Ebd.

Informationen dazu und daher auch keine Konsensmöglichkeit darüber gibt, wie die Werte der DC-Datenfelder zu interpretieren sind.

Nutzerseitig ist bei Aggregatoren bzw. Service Providern wie BASE sicherlich zu bemängeln, dass eine Suchanfrage zur Anzeige verschiedener Records der gleichen Ressource führt, und eine Zusammenführung der Daten, welche in Sets zusammenhängen, schwierig bis unmöglich zu bewerkstelligen ist, wenn die Metadaten von verschiedenen Data Providern angeboten werden. Dies macht die Anzeige der Suchergebnisse sehr unübersichtlich.

## 5. Zukunft

Im Laufe der letzten Jahre gab es einige Entwicklungen und Trends in der digitalen wissenschaftlichen Kommunikation und den digitalen Technologien, die angesichts der am Ende des vorherigen Kapitels (Kap. 4) kurz angerissenen Nachteile von OAI-PMH die Frage nach seiner Zukunft bzw. Zukunftstauglichkeit laut werden lassen. Hat ein Konzept wie OAI-PMH angesichts des rasanten Eroberungszuges von Services wie der Web-Suchmaschine Google Scholar, die sowohl Metadaten als auch die darin beschriebenen Inhalte *harvesten* und dazu in der Lage sind, direkt auf die Originalressourcen (z. B. Textdokumente) zu verweisen bzw. Kopien dieser anzubieten,<sup>37</sup> noch eine Zukunft? Wird sich OAI-PMH angesichts der neuen Semantic Web Technologien behaupten können? Wird es weiterhin zur Etablierung neuer Drittanbieter kommen, die die Inhalte der offenen Repositorien wiederverwenden; der erwartete Boom, der bei der Erstellung der Richtlinien zur Implementierung von Aggregatoren durch die OAI gemutmaß wurde,<sup>38</sup> ist bisher u. a. auch wegen des Aufkommens von Suchmaschinen und Discovery-Systemen ausgeblieben.

Aus heutiger Sicht ist der größte Schwachpunkt des OAI-PMH – neben des sehr niedrigschwelligen Metadatenstandards DC-Simple – sicherlich der repositorienbezogene Ansatz von OAI-PMH (der sehr stark auf Praktiken der digitalen Bibliothek beruht, siehe oben, Kap. 4), dem heute ein webbezogener und ressourcenbasierter Ansatz (*web centric, resource centric approach*) gegenübersteht. Deshalb hat die OAI OAI-ORE (Object Reuse & Exchange) entwickelt, die diesen neuen Ansatz verfolgt.<sup>39</sup> Ohne zu sehr ins Detail zu gehen, kann man OAI-ORE als ein weiteres Protokoll beschreiben, dass jedoch im Gegensatz zu OAI-PMH der Netzwerkstruktur des WWW angepasst ist. Das heißt, es findet URIs, die die Ressourcen beschreiben (d. h. die Ressourcen werden durch URIs beschrieben, nicht durch DC). Dadurch ist es möglich, einzelne Ressourcen in Repositorien unabhängig anzusteuern und einzeln auszusuchen und sie z. B. mit anderen Ressourcen repositorienübergreifend in anderen Kontexten miteinander zu verbinden. Noch scheinen sich jedoch

---

<sup>37</sup> Siehe Kelly et al. (2012), Kapitel 6.

<sup>38</sup> Siehe "Dublin Core Metadata Initiative | Guidelines for Aggregators, Caches and Proxies".

<sup>39</sup> Siehe Van de Sompel & Nelson (2015), Kap. 3 The Open Archives Initiative Object Reuse and Exchange - OAI-ORE (2006).

neuere und besser maschinenlesbare Interoperabilitätsansätze für die internetbasierte wissenschaftliche Kommunikation nicht durchgesetzt zu haben, nicht zuletzt weil sich sehr viele unterschiedliche Stakeholdergruppen an der Umsetzung beteiligen müssten.<sup>40</sup>

In diesem Zusammenhang sind auch die Möglichkeiten des *Semantic Web* mit seiner Tripelstruktur der Linked Open Data (LOD) zu sehen:

“Einer der Vorteile, URIs zu verwenden, ist, dass diese über das Protokoll HTTP aufgelöst werden können und oft ein Dokument zurückgeben, welches der gegebene URI weiter beschreibt (das ist das sogenannte Prinzip Linked Open Data). Im gegebenen Beispiel etwa kann man die URIs der Knoten und Kanten (z. B. <http://schema.org/Person>, <http://schema.org/birthPlace>, <http://www.wikidata.org/entity/Q1731>) alle auflösen und erhält dann weitergehende Beschreibungen, z. B. dass Dresden eine Stadt in Deutschland ist, oder dass eine Person auch fiktiv sein kann.”<sup>41</sup>

OAI-PMH und DC basieren nicht auf URIs, d. h. es ist bei der Angabe “Dresden” in einem Metdatenfeld nicht hinterlegt, dass es sich um einen Ort, z. B. einen Verlagsort, handelt. Im Prinzip wäre es möglich, DC-basierte Metadaten mit URIs “nachzurüsten”, es würde jedoch einen sehr großen Aufwand bedeuten. Auch ist es für “uneingeweihte Benutzer” schwierig gute – den Möglichkeiten des *Semantic Web* gerecht werdende – Abfragen zu gestalten. Da das *Semantic Web* jedoch sehr interessante, die Möglichkeiten von OAI-PMH weit übersteigende Abfragemöglichkeiten bietet, gehen einige Repositorien inzwischen dazu über, für die breite Nutzung OAI-PMH-Schnittstellen anzubieten und gleichzeitig LOD (z. B. als SPARQL-Endpoints) für die avancierte Nachnutzung bereitzustellen (zum Beispiel das Projekt Alexander von Humboldt auf Reisen (AvH) der BBAW).<sup>42</sup>

Zusammenfassend ist festzustellen, dass OAI-PMH im Zusammenhang mit den momentan u. .a. im Zusammenhang mit dem europäischen Förderprogramm Horizon 2020 viel diskutierten sogenannten FAIR-Prinzipien<sup>43</sup> vor allem das F von Findable (Auffindbarkeit) grundlegend sichert, auch wenn an der Granularität der Suche gerade in generischen Suchdiensten<sup>44</sup> bzw. der Qualität der Metadaten (siehe oben) einiges anzumerken ist und weitere Vereinheitlichungen erstrebenswert sind.<sup>45</sup>

---

<sup>40</sup> Siehe Van den Sompel and Nelson 2015, Kap. 7 Conclusion.

<sup>41</sup> “Wikipedia | Semantic Web”, letzter Zugriff 25.07.2017.

<sup>42</sup> Diese Einsichten beruhen auf einem Gespräch mit Stefan Dumont, dem Entwickler des Projekts AVH (<http://www.bbaw.de/forschung/avh-r>).

<sup>43</sup> Siehe Pampel und Kindling (2017), S. 19.

<sup>44</sup> Pampel und Kindling (2017), S. 23.

<sup>45</sup> Pampel und Kindling (2017), S. 23: “Ein FDR orientiert sich bei der Aufbereitung von Forschungsdaten an in der jeweiligen Fachcommunity verfügbaren Standards sowie an übergreifenden Qualitätsstandards für Informationsinfrastrukturangebote wie etwa dem Data Seal of Approval, der DIN-Norm 3164412 oder der ISO-Norm 16363:2012 ”Audit and certification of trustworthy digital repositories“. Ein FDR bietet eine metadatenbasierte Suche über die gespeicherten Inhalte und erlaubt den Zugang zu diesen über eine Weboberfläche. Darüber hinaus stellt das FDR die nachgewiesenen Metadaten über standardisierte Schnittstellen bereit und erlaubt so Aggregation der Metadaten durch Dienste Dritter (z. B. durch Suchmaschinen und Harvesting-Dienste).”

Es ist davon auszugehen, dass OAI PMH als Standard nicht so schnell verdrängt werden wird, nicht zuletzt weil das Semantic Web und die dazugehörigen Technologien noch voll in der Entwicklung sind, ohne dass momentan abzusehen ist, wohin diese Entwicklungen führen werden.

# Bibliographie

- Bielefeld Academic Search Engine (BASE), "BASE | Homepage", <https://www.base-search.net/>, letzter Zugriff 25.07.2017
- Borst, Timo (2014), "Repositorien auf ihrem Weg in das Semantic Web: Semantisch hergeleitete Interoperabilität als Zielstellung für künftige Repository-Entwicklungen", in: *Bibliothek. Forschung und Praxis* 38, 2, S. 257-265, <https://www.degruyter.com/view/j/bfup.2014.38.issue-2/bfp-2014-0034/bfp-2014-0034.xml> (DOI: <https://doi.org/10.1515/bfp-2014-0034>), letzter Zugriff 25.07.2017.
- Budapest Open Access Initiative, "Budapest Open Access Initiative | Homepage", <http://www.budapestopenaccessinitiative.org/>, letzter Zugriff 25.07.2017.
- DINI (2016), *DINI-Zertifikat 2016 „Open-Access-Repositorien und -Publikationsdienste“*, <http://nbn-resolving.de/urn:nbn:de:kobv:11-100239432>, letzter Zugriff 25.07.2017.
- DINI Arbeitsgruppe Elektronisches Publizieren (2005), Elektronisches Publizieren an Hochschulen: Inhaltliche Gestaltung der OAI-Schnittstelle – Empfehlungen (2005), <https://edoc.hu-berlin.de/bitstream/handle/18452/2131/2-de2.pdf?sequence=1&isAllowed=y> (DOI: 10.18452/1479), letzter Zugriff 25.07.2017.
- Dublin Core Metadata Initiative, "Dublin Core Metadata Initiative | Background", <http://dublincore.org/metadata-basics/>, letzter Zugriff 25.07.2017.
- Dublin Core Metadata Initiative, "Dublin Core Metadata Initiative | Dublin Core Metadata Element Set, Version 1.1", <http://dublincore.org/documents/dces/>, letzter Zugriff 25.07.2017.
- Dublin Core Metadata Initiative, "Dublin Core Metadata Initiative | Guidelines for implementing Dublin Core in XML", <http://dublincore.org/documents/2002/12/02/dc-xml-guidelines/>, letzter Zugriff 25.07.2017.
- Dublin Core Metadata Initiative, "Dublin Core Metadata Initiative | Guidelines for Aggregators, Caches and Proxies", <https://www.openarchives.org/OAI/2.0/guidelines-aggregator.htm>, letzter Zugriff 25.07.2017.
- Forschungsdaten.org, "Forschungsdaten.org | OAI-PMH", <http://www.forschungsdaten.org/index.php/OAI-PMH>, letzter Zugriff 25.07.2017.
- Kelly, B., N. Sheppard, J. Delasalle, M. Dewey, O. Stephens, G. Johnson und S. Taylor (2012), "Open Metrics for Open Repositories", (Conference Paper: OR2012: the 7th International Conference on Open Repositories, 2012-07-09 – 2012-07-13), <http://opus.bath.ac.uk/30226/>, letzter Zugriff 25.07.2017.
- Kostädt, Peter, "Suchportale, Discovery-Systeme und Linkresolver", in: Rolf Griebel, Hildegard Schäffler und Konstanze Söllner (Hg.), *Praxishandbuch Bibliotheksmanagement*, Band 1, Berlin, S. 513-523.
- Lauber-Reymann, Margrit (2017), *Informationsressourcen: Ein Handbuch für Bibliothekare und Informationsspezialisten*, 2., vollständig überarbeitete und erweiterte Auflage, Bibliotheks- und Informationspraxis, 49, Berlin.
- OCLC, "OCLC | OAIster database", <http://www.oclc.org/en/oaister.html>, letzter Zugriff 25.07.2017.
- Open-Access-Netzwerk (OAN), "OAN | DINI-OAI-Validator verfügbar", <https://oanetzwerk.wordpress.com/2011/11/10/dini-oai-validator-verfuegbar/>, letzter Zugriff 25.07.2017.
- Open Archives Forum, "Open Archives Forum | Tutorial: OAI for Beginners: Overview", <http://www.oaforum.org/tutorial/english/page1.htm>, letzter Zugriff 25.07.2017.
- Open Archives Forum, "Open Archives Forum | Tutorial: 2. History and development of OAI-PMH", <http://www.oaforum.org/tutorial/english/page2.htm>, letzter Zugriff 25.07.2017.
- Open Archives Initiative (OAI), "OAI | Homepage", <http://www.openarchives.org>, letzter Zugriff 25.07.2017.
- Open Archives Initiative (OAI), "OAI | Organization", <https://www.openarchives.org/organization/>, letzter Zugriff 25.07.2017.
- Open Archives Initiative (OAI), "OAI | The Open Archives Initiative Protocol for Metadata Harvesting V. 2.0", <http://www.openarchives.org/OAI/openarchivesprotocol.html>, letzter Zugriff 25.07.2017.
- Open Archives Initiative (OAI), "OAI | Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting", <https://www.openarchives.org/OAI/2.0/guidelines.htm>, letzter Zugriff 25.07.2017.
- Pampel, Heinz und Maxi Kindling (2017), "Informationsinfrastrukturangebote für digitale Forschungsdaten", in: Boris Jacob, Maxi Kindling und Uwe Müller (Hg.), *E(Hren)-Journal für Peter Schirmbacher anlässlich seiner Emeritierung als Professor am Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin*, Berlin, <http://ehrenjournal.ib.hu-berlin.de/>, letzter Zugriff 25.07.2017.
- Rühle, Stefanie (o.J.), "Kleines Handbuch zu Metadaten - Metadaten", online verfügbar: [http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Handbuch/metadaten.pdf?\\_\\_blob=publicationFile](http://www.kim-forum.org/Subsites/kim/SharedDocs/Downloads/DE/Handbuch/metadaten.pdf?__blob=publicationFile), letzter Zugriff 25.07.2017.

Van de Sompel, Herbert und Michael L. Nelson (2015), "Reminiscing About 15 Years of Interoperability Efforts", *D-Lib Magazine*, 21, 11/12, <http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html>, (DOI: 10.1045/november2015-vandesompel), letzter Zugriff 25.07.2017.  
Wikipedia, "Wikipedia | Semantic Web", [https://de.wikipedia.org/wiki/Semantic\\_Web](https://de.wikipedia.org/wiki/Semantic_Web), letzter Zugriff 25.07.2017.

**Abbildungsverzeichnis:**

Die Abbildungen stammen von unserem Poster (siehe Anmerkung Deckblatt, wurden jedoch ins Deutsche übersetzt und noch einmal bearbeitet).

# Anhang

## Glossar

- **HARVESTING**

Der Begriff **Harvesting** (dt. Ernten) bezieht sich im OAI-PMH-Kontext auf das regelmäßige asynchrone Suchen<sup>46</sup> und Einsammeln bzw. Abfragen von Metadaten von verteilten Repositorien,<sup>47</sup> sowie deren Zusammenführung in einem eigenen Datenspeicher, innerhalb dessen die Inhalte durch Dienste durchsuchbar gemacht werden.<sup>48</sup> Ein Harvester ist technisch gesehen eine Client-Anwendung, die OAI-PMH-Anfragen bearbeitet und von so genannten **Service Providern** betrieben wird.

- **ARCHIVE**

Der Begriff **Archive** (dt. Archiv) bezieht sich im OAI Kontext auf Repositorien mit Forschungspublikationen.<sup>49</sup> *Open* zeigt hierbei an, dass die Inhalte Open Access verfügbar sind.

- **PROTOCOL**

Das OAI-PMH ist ein technisches **Protokoll**, wobei ein Protokoll eine bestimmte Menge von Regeln festlegt, mit deren Hilfe Computer oder computerbasierte Systeme kommunizieren. Ein bekanntes Beispiel ist das HTTP Protokoll, das die Kommunikation von Systemen via Internet ermöglicht.<sup>50</sup>

- **METADATA**

**Metadaten** (engl. Metadata) sind Daten, die Daten beschreiben, im OAI-Kontext insbesondere Dokumente. Hierbei handelt es sich um übergeordnete Informationen, welche auffindbare Ressourcen für eine sinnvolle Durchsuchbarkeit technisch, administrativ und beschreibend auszeichnen. Insbesondere die beschreibenden Metadaten bilden die formale und sachliche Erschließung ab. Im Kapitel 3.4 wird auf diese Art der Metadatenbeschreibung näher eingegangen.<sup>51</sup>

- **REPOSITORY**

Ein **Repositorium** (engl. *Repository*) ist im OAI-Kontext ein durch ein Netzwerk oder das Internet frei erreichbarer Dokumente-Server mit Sammlungen elektronischer Ressourcen. Der Server ist in der Lage OAI-PMH-Anfragen (in der Version 2.0) verarbeiten und wird von einem sogenannten Data Provider gesteuert, um Metadaten austauschbar zu machen.<sup>52</sup>

---

<sup>46</sup> Siehe DINI Arbeitsgruppe Elektronisches Publizieren (2005), S. 6.

<sup>47</sup> Siehe "Open Archives Forum | Tutorial: OAI for Beginners: Overview", letzter Zugriff 25.07.2017.

<sup>48</sup> Siehe DINI Arbeitsgruppe Elektronisches Publizieren (2005), S. 6.

<sup>49</sup> Siehe "Open Archives Forum | Tutorial: OAI for Beginners: Overview", letzter Zugriff 25.07.2017.

<sup>50</sup> Ebd.

<sup>51</sup> Siehe DINI (2016), S. 59.

<sup>52</sup> Siehe "OAI | The Open Archives Initiative Protocol for Metadata Harvesting V. 2.0", letzter Zugriff 25.07.2017.



- **DATA PROVIDER**

Ein **Data Provider** (dt. Datenanbieter) unterhält Systeme, die OAI-PMH unterstützt Metadaten zu bestimmten Ressourcen anbieten.<sup>53</sup> Hierfür unterhält er ein oder mehrere Repositorien.<sup>54</sup>

- **SERVICE PROVIDER**

Ein **Service Provider** (dt. Dienstanbieter) stellt OAI-PMH-Anfragen an **Data Provider** oder **Aggregatoren**, um die dort befindlichen Metadaten zu harvesten. Die Metadaten werden dann weiterverarbeitet und wiederum in einem Mehrwert-Service (*value added service*) z. B. durchsuchbar gemacht und zur weiteren Nutzung angeboten.<sup>55</sup>

- **AGGREGATOR**

Ein **Aggregator** wird konzeptuell – ähnlich wie z. B. Proxies – als ein Drittanbieter von Metadaten betrachtet. Er vereint die Eigenschaften von Data Providern und Service Providern in sich, d. h. die Daten werden von Data Providern geharvestet, aber sie können auch verarbeitet und sowohl zum weiteren Harvesting angeboten als auch mittels eines Mehrwert-Services angeboten werden. Aggregatoren sind dazu in der Lage, ausgereifere Harvesting-Methoden zu nutzen und komplexere Metadaten anzubieten als Data oder Service Provider. Sie können theoretisch z. B. die Provenienz einzelner bibliographisch beschriebener Quellen (*records*) zurückverfolgen.<sup>56</sup>

- **INTEROPERABILITY**

**Interoperabilität** (engl. *Interoperability*) gewährleistet die anwendungsübergreifende Nutzung, die Verlinkung und den Austausch von Metadaten. Dadurch soll gewährleistet werden, Daten zwischen verschiedenen Systemen wie digitalen Anwendungen bestenfalls ohne großen Informationsverlust auszutauschen. Interoperabilität ist bezogen auf Metadaten hierbei dreigeteilt:

- 1) Die strukturelle Interoperabilität erfordert für verschiedene Metadatenformate ein gemeinsames Datenmodell (z. B. das Dublin Core Abstract Model). Ebenso erzeugt die gemeinsame Nutzung eines Protokolls (wie OAI-PMH) strukturelle Interoperabilität.
- 2) Die syntaktische Interoperabilität erfordert eine gemeinsame Syntax für die Daten (z. B. wird im OAI-PMH die Syntax XML genutzt).
- 3) Die semantische Interoperabilität erfordert die Nutzung von gleichen Metadatermen, also eine gemeinsame Nutzung desselben Metadatenstandards, so dass verschiedene Systeme die Daten inhaltlich auf dieselbe Weise interpretieren können. Eine grundlegende semantische Interoperabilität der Inhalte wird z. B. im OAI-PMH durch die Nutzung der 15 DC-Simple-Felder vorgegeben.<sup>57</sup>

---

<sup>53</sup> Siehe "OAI | The Open Archives Initiative Protocol for Metadata Harvesting V. 2.0", letzter Zugriff 25.07.2017.

<sup>54</sup> Siehe "Open Archives Forum | Tutorial: OAI for Beginners: Overview", letzter Zugriff 25.07.2017.

<sup>55</sup> Ebd.

<sup>56</sup> Siehe "OAI | Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting", letzter Zugriff 25.07.2017.

<sup>57</sup> Siehe Rühle (o.J.), S. 4-5.