

Herbert Hrachovec

# OAI-PMH

## Grundstein und Prüfstein der Open-Access-Bewegung

Handbuch Repositorienmanagement, Hg. v. Blumesberger et al., 2024, S. 121–132  
<https://doi.org/10.25364/97839033742328>



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz, ausgenommen von dieser Lizenz sind Abbildungen, Screenshots und Logos.

Herbert Hrachovec, Universität Wien, Institut für Philosophie, [herbert.hrachovec@univie.ac.at](mailto:herbert.hrachovec@univie.ac.at) |  
ORCID iD: 0000-0003-4778-014X

## Zusammenfassung

Die technischen Voraussetzungen des freien Austausches wissenschaftlicher Fachliteratur, den die „Budapester Erklärung“ 2002 promulgierte, wurden bereits 1999 durch eine Forschungsgruppe angeregt. Ein HTML-Übertragungsprotokoll (OAI-PMH) sollte die Verbreitung bibliographischer Metadaten standardisieren. Dieser Regelsatz erwies sich als nachhaltiger Erfolg und wurde zum Kernstück der „Open Access Initiative“ (OAI). Er wird skizzenhaft dargestellt und in seiner Geschichte verfolgt. In der weiteren Entwicklung digitaler bibliographischer Informationshelfe zeigten sich Schwachstellen und Erweiterungsmöglichkeiten des Protokolls. Durch die Entwicklung mächtiger Suchmaschinen und leistungsfähiger Big-Data-Algorithmen wird es in eine neue, in ihren Auswirkungen noch ungewisse, sozopolitische Umgebung versetzt.

**Schlagwörter:** Metadaten; technische Infrastruktur; Bibliotheksparadigma; Suchmaschinen

## Abstract

### OAI-PMH. Cornerstone and Touchstone of the Open-Access Movement

The technical requirements of the free exchange of scientific literature, which the “Budapest Declaration” promulgated in 2002, had already been suggested by a research group in 1999. An HTML transfer protocol (OAI-PMH) was to standardise the dissemination of bibliographic metadata. This rule set proved to be a lasting success and became the core of open-access technology. It will be outlined and its history traced. As it progressed, weaknesses and the extensibility of the protocol became apparent. The development of powerful search engines and powerful big-data algorithms put it in a new socio-political environment, still uncertain in its implications.

**Keywords:** Metadata; technical infrastructure; library paradigm; search engines

## 1. Einleitung

Das „Open Archives Initiative Protocol for Metadata Harvesting“<sup>1</sup> legt Prozesse fest, die zwischen digitalen Archiven und Servicestellen, welche deren Metadaten sammeln, ablaufen. Es wurde ursprünglich entworfen, um den Austausch zwischen bibliographischen Datenquellen unterschiedlichen Formates zu standardisieren und eine einheitliche Grundlage zur Weiterverarbeitung ihrer Inhalte anbieten zu können. Softwarepakete, die im Anschluss an die „Open Access Initiative“ (OAI) entwickelt wurden, implementieren die sechs Abfragemuster, aus denen das knappe Protokoll besteht. Zunächst ist ein Blick auf diese Festlegungen zu werfen. Zwanzig Jahre nach seiner Veröffentlichung ist OAI-PMH einerseits nicht aus der Infrastruktur des offenen Datenaustausches wegzudenken, andererseits aber durch nachfolgende Entwicklungen teilweise überholt worden. Das heißt jedoch keineswegs, es wäre von bloß historischem Interesse. Gerade seine Rolle im Verlauf der letzten beiden Dezennien lässt erkennen, wo wir heute stehen. Diesem Aspekt gilt der zweite Abschnitt des vorliegenden Beitrags. Im dritten Teil wird, anknüpfend an die Geschichte dieser technischen Mustervorgabe, ein tiefgreifendes Dilemma der Politik von Open Access angesprochen. Institutionell vorgegebene Regeln zur Förderung des gemeinschaftlichen Zugangs zu Ressourcen stehen in Konkurrenz zu weniger kanonisch festgelegten Prozeduren, deren Effektivität ihr Defizit an Gemeinnutzen aufzuwiegen verspricht. Im Klartext: Die Informationsbürokratie der Archivverwaltung steht gegen kommerzielle Big-Data-Algorithmen.

## 2. Technik

Ein OAI-Archiv sammelt Publikationen und deren Metadaten, um sie zur Abfrage über das Internet zur Verfügung zu stellen. Das dazu verwendete Protokoll gliedert sich in Anfragen zur Archivstruktur und Anweisungen zur Übermittlung der angebotenen Inhalte. Konzeptuell wird zwischen einer Ressource („resource“), deren Eintragung im Archiv („item“) und den mit ihr verbundenen Datensätzen („records“) unterschieden. Einträge erfassen ein Objekt in Datensätzen, die möglicherweise verschiedenen metasprachlichen Standards folgen. Der digitale Informationstransfer besteht in Kommandos, die intuitiv gebräuchlichen „Verben“ nachgebildet sind. Drei Eingaben ermitteln die Archivstruktur. Dem in HTML abgesandten Anforderungstyp „identify“ antwortet eine Kurzbeschreibung des Archivs. „ListMetadataFormats“ gibt an, in welchen Formaten es seine Metadaten bereitstellt, und „ListSets“ dokumentiert, in welche inhaltlichen Gruppen sie allenfalls gegliedert

---

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

sind. Der eigentlichen Datenübertragung dienen drei weitere Ausdrücke. „ListIdentifiers“ bietet eine Kurzfassung der archivierten Ressourcen, während sie durch „ListRecords“ vollständig angeführt werden. „GetRecord“ liefert schließlich die Metadaten einer einzelnen Ressource. Alle Verben besitzen optionale Parameter, die es erlauben, die Abfragen unter anderem nach Kategorien, Datum und Zeitspanne der Archivierung einzuschränken.<sup>2</sup>

Der geordneten Erfassung von Metadaten kommt im Wissenschaftsbetrieb besondere Bedeutung zu. Das weltweite Netz akademischer Institutionen bedarf einer Infrastruktur, welche den digitalen Austausch der Information über vorhandene wissenschaftliche Ressourcen steuert. Aus diesem Grund haben sich gegen Ende des vergangenen Jahrhunderts diverse Zitiermuster zur informationstechnischen Erfassung wissenschaftlicher Publikationen entwickelt. Ein frühes Beispiel ist das seit 1999 etablierte MARC21, ein Digitalformat zur Erfassung computerisierter Bibliothekskataloge.<sup>3</sup> Der „Metadata Encoding and Transmission Standard“ (METS)<sup>4</sup> wiederum dient der Beschreibung unterschiedlicher Perspektiven auf eine bestimmte Ressource (z. B. jener der Drucktechnik und der eines Inhaltsverzeichnisses). Die „Digital Item Declaration Language“ (DIDL)<sup>5</sup>, um noch ein Beispiel zu nennen, definiert eine Struktur von Mengeninklusionen zwischen „Containern“, „Items“, „Components“ und „Resources“. Diese Formate sind, entsprechend lokalen Bedürfnissen, noch immer in Gebrauch. Die bedeutsame Neuerung von OAI-PMH besteht nun darin, solche Spezifikationen optional zuzulassen (sie werden durch „ListMetadataFormats“ abgefragt), jedoch für eine obligatorische Datenstruktur zu sorgen. Dazu wird die Erfassung der Bestände mittels einer Basisversion von „Dublin Core“<sup>6</sup> vorgeschrieben. Dieses weit verbreitete Format stellt 15 gebräuchliche bibliographische Attribute zur Verfügung. Zu diesen Angaben gehören u. a. Titel und Autor:innen der Ressource, eine Kurzbeschreibung des Inhalts, der Verlag und das Publikationsdatum. Die genannten Metadaten sind durch OAI-PMH für den allgemeinen Gebrauch über das WWW zugänglich und werden in einer XML-Struktur ausgeliefert. Die Form der Dateneinträge ist allerdings stellenweise nicht exakt normiert, was zu Problemen führen kann. Es ist nicht unerheblich, ob der erste Eintrag in ein Namensfeld als Vor- oder Nachname zu interpretieren ist

2 Eine Übersicht zur Terminologie und Verwendung von OAI-PMH gibt Warner, S. (2001). Die Präsentation von Naravaran, N. (2010) enthält eine hilfreiche Visualisierung der Zusammenhänge. Siehe die vergleichbaren Folien von Brungai, L. (2011). Gaudinat, A.; Beausire, J.; Fuss, M. et al. (2017) haben eine Studie zur Reichweite und Überschneidung der gebräuchlichsten OAI-PMH konformen Meta-Kataloge vorgelegt.

3 <https://www.loc.gov/marc/bibliographic/>

4 <http://www.loc.gov/standards/mets/>

5 <http://xml.coverpages.org/mpeg21-didl.html>

6 <https://dublincore.org/>

(„Berthold Ludwig“, „Elisabeth Anne“).<sup>7</sup> Die Festschreibung eines Minimalkonsenses zur Formalisierung von Informationen im bibliographischen Datentransfer war wohl entscheidend für den eindrucksvollen Erfolg von OAI-PMH. Er wird durch zahlreiche informationstechnische, praxisorientierte und medientheoretische Publikationen belegt, die sich die Umsetzung der Deklarationen von Budapest und Berlin zum Ziel gesetzt haben.<sup>8</sup>

Die offizielle Registrierungsstelle<sup>9</sup> meldet zum unten angeführten Datum 5.318 protokoll-konforme Repositorien („data providers“), der „Bielefeld Academic Research Engine“ (BASE)<sup>10</sup> nennt 9.101 „Datenlieferanten“. Die Zahlen alleine geben allerdings nur ein ungefähres Bild. Das Volumen der erfassten Archive variiert, manche stagnieren oder sind verwaist. Dennoch ist mit OAI-PMH ein Instrument geschaffen worden, das die Vorstellung einer transparenten Organisation der verfügbaren Kenntnisse über wissenschaftliche Forschungsergebnisse exemplarisch zu realisieren hilft. Die Bedeutung dieses Umstands ist im Vergleich mit den in weiterer Folge entwickelten proprietären Angeboten des „social web“ zu ermessen, auf die im dritten Abschnitt dieses Beitrags zurückgekommen wird. Ebenso wichtig wie die Archive „vor Ort“ sind die im Protokoll vorgesehenen „Harvesters“, denen die Aufgabe zufällt, das Datenmaterial gesammelt in einer konsistenten, benutzer:innen-freundlichen Form zur Verfügung zu stellen (BASE, CORE<sup>11</sup>, OpenAIRE<sup>12</sup>, Zenodo<sup>13</sup>). Die erwähnte unvollständige Operationalisierung von Dublin Core erfordert z. B. Korrekturarbeiten seitens der Aggregatoren<sup>14</sup>. Die Auslieferung großer Datenmengen läuft außerdem nicht immer störungsfrei, doch solche Schwierigkeiten scheinen die Verbreitung des Protokolls nicht behindert zu haben. Als Beispiel der anhaltenden Attraktivität der beschriebenen Infrastruktur kann eine kürzlich unternommene Recherche dienen. Seit dem Jahr 2000 wurden weltweit etwa 50 neue Open Access e-Journals aus dem Bereich der Philosophie gegründet<sup>15</sup>. Die

---

7 Vgl. zum Problem <https://de.comm.infosystems.www.authoring.misc.narkive.com/P8ihO9lz/public-core-autoren>.

8 Siehe die Budapest und Berliner Open-Access-Deklarationen: <https://www.budapestopenaccess-initiative.org/read/german-translation/> und <https://openaccess.mpg.de/Berliner-Erklarung>. Zur Entwicklung der Open-Access-Bewegung: Suber, P. (2016). Die Deutsche Initiative für Netzwerkinformation hat eine Informationsbroschüre herausgegeben: Deutsche Initiative für Netzwerkinformation e. V. (2005). Zur deutschen Diskussion der Open-Archives-Initiativen: Herb, U. (2012).

9 <https://www.openarchives.org/Register/BrowseSites>

10 <https://www.base-search.net/>

11 <https://core.ac.uk/>

12 <https://www.openaire.eu/>

13 <https://zenodo.org/>

14 Müller, S. (2020)

15 Hrachovec, H. (2021)

Mehrzahl ihrer Inhalte stehen über eine OAI-PMH-Schnittstelle zum Abruf zur Verfügung, sodass die Journalseiten gleichzeitig als frei zugängliche Archive fungieren. Über die genannten „Harvesters“ sind die Beiträge dem interessierten Publikum zugänglich.

### 3. Geschichte

Als OAI-PMH entwickelt wurde<sup>16</sup>, war das Protokoll darauf ausgelegt, „Service Providers“ mit den Grunddaten zur bibliometrischen und bibliothekarischen Aufbereitung zu versorgen. Zu ihren Aufgaben sollten nicht bloß die Erfassung wissenschaftlicher Publikationen, sondern auch die Dokumentation ihrer Verbreitung durch Abfragen und Zitationen gehören. Die direkte Zugänglichkeit der faktischen Dokumente fehlte allerdings, wie auch im traditionellen Katalogeintrag die verzeichneten Bücher selbst nicht enthalten waren. Dementsprechend bezeichnet das Ensemble von Informationen, welches im Rahmen von OAI-PMH transportiert wird, die betreffenden Ressourcen nicht mit einer URI. „ListIdentifiers“ gibt eine archivinterne Adresse an, Weblinks sind fakultativ. Zur Jahrtausendwende war noch nicht absehbar, dass in Zukunft hocheffektive Suchmaschinen wissenschaftlich relevante Texte mit Hilfe ausgefeilter Algorithmen (ziemlich) präzise aus einer Unmenge von Webseiten herausfiltern könnten. Dieses Defizit von OAI-PMH motivierte verschiedene Vorschläge zu Protokollerweiterungen. Im Rahmen der OAI-Initiative selbst entwickelten Carl Lagoze und Herbert Van de Sompel „Open Archives Initiative – Open Reuse and Exchange“ (OAI-ORE) zur protokollkonformen Erfassung medial komplexer digitaler Objekte (2007)<sup>17</sup>. Die Verteilung und Wiederverwertung der Einträge soll danach durch ihre Definition als Aggregate gewährleistet werden. Sie bestehen aus sachlich zusammengehörigen Komponenten, die im Unterschied zu OAI-PMH durch eine „Resource Map“ erschlossen werden. Durch sie wird ein Zugang zum semantischen Web geschaffen, also zur Möglichkeit einer Charakteristik solcher Aggregation in gängigen Datenformaten (z. B. RDF/XML, RDFa oder AtomXML).<sup>18</sup> Andere Ansätze zur Ausweitung der Kapazitäten von OAI-PMH sind im Rahmen externer Projekte verfolgt worden. „Linked Data“ bezeichnet eine Anzahl von Konventionen zur maschinen-lesbaren Verknüpfung von Informationsbeständen des WWW. Bezogen auf frei zugängliche Daten bilden sie einen Teil

<sup>16</sup> Den Ausgangspunkt bildete die Einladung zu einem Treffen in Santa Fe (Juli 1999), bei dem die Kompatibilität der zu diesem Zeitpunkt aktuellen e-prints-Formate zur Debatte stand. Federführend waren Paul Ginsparg, Rick Luce und Herbert Van de Sompel. Zur Geschichte vgl. das informative Tutorium: <http://www.ukoln.ac.uk/metadata/oa-forum/tutorial/page2.htm>

<sup>17</sup> <http://www.openarchives.org/ore/>

<sup>18</sup> Siehe [https://www.w3schools.com/xml/xml\\_rdf.asp](https://www.w3schools.com/xml/xml_rdf.asp); <https://rdfa.info/>; <http://www.atom-enabled.org/developers/syndication/>

des „semantischen Webs“, das sich aus derartigen heterogenen Inhalten knüpfen lässt. OAI-PMH Datensätze können systematisch durch Bestände der „Linked Open Data Cloud“ (LOD-Cloud) ergänzt werden<sup>19</sup>. Dabei werden die OAI-PMH „Identifiers“ durch gängige URLs ergänzt, die auf semantisch angereicherte Ressourcenbeschreibungen verweisen. Sie binden die Metadaten in eine LOD-Cloud ein<sup>20</sup>.

Anders als OAI-PMH zeigten dessen Ergänzungen durch semantische Zusätze relativ wenige Folgewirkungen. Einer Umfrage aus dem Jahr 2018 zufolge publizieren 65 % von 142 erfassten Institutionen „linked data“ zu experimentellen Zwecken „to demonstrate what could be done with datasets as linked data“; 45 % geben an, dass sie das Konzept an hauseigenen Daten ausprobieren wollten<sup>21</sup>. Eine Breitenwirkung des Konzepts ist nicht festzustellen, wohl aber einige Versuche, in diese Richtung weiter zu forschen. So konvertiert das OntoOAI-Modell Ergebnisse von OAI-PMH-Sammlungen nach RDF, reichert sie mit LOD-Daten an und produziert nach einer Validierung in LOD-Tripeln semantische Graphen, aus denen neue Wissensbestände extrahiert werden können<sup>22</sup>. Einen weiteren eigenständigen Ansatz zur Datenverbreitung verfolgt das „ResourceSync Framework“<sup>23</sup>. Die Idee dahinter ist, dass die Synchronisation über den Abgleich der Seitenverzeichnisse geeigneter Portale („sitemaps“) die Bereitstellung gemeinsamer Informationen und ihrer Aktualisierungen mit einem Schlag auf elegante Art lösen soll. Julien A. Raemy analysierte in einer Studie 2020 mit Blick auf die Zielvorstellungen des Europeana-Projektes das Angebot an Alternativen zu OAI-PMH. 71,2 % (von 37 Teilnehmer:innen an der Umfrage) verwenden dieses Format zur Aggregation von Metadaten<sup>24</sup>. Angeichts des Aufwands, der mit der Einrichtung dieser Informationsarchitektur verbunden war, äußerten sich einige Befragte skeptisch gegenüber einer Veränderung des status quo<sup>25</sup>. Raemy rechnet nicht damit, dass die Sammlung von Metadaten in der Betreuung des kulturellen Erbes in absehbarer Zeit auf alternative Formate wechselt wird<sup>26</sup>. Seine Untersuchung steuert (ungewollt) eine wichtige zusätzliche Erklärung dafür bei. Unter den angeführten Optionen ist kein Vorschlag, der auch nur in die Nähe der Standardisierung, Akzeptanz und Verbreitung von OAI-PMH käme. Weder die Schwachstellen des 20 Jahre alten Protokolls noch die Aussichten

19 Linked Open Data. Europeana Pro: <https://pro.europeana.eu/page/linked-open-data>

20 Haslhofer, B.; Schandl, B. (2008). Der 2008 vorgestellte Ansatz wurde nicht weiter verfolgt.

21 Raemy, J. A. (2020), S. 96.

22 Becerril-Garcia, A.; Aguado-López, E. (2018)

23 ResourceSync Framework Specification – Table of Contents: <http://www.openarchives.org/rs/toc>

24 Raemy, J. A. (2020), S. 34.

25 Raemy, J. A. (2020), S. 39.

26 Raemy, J. A. (2020), S. 54.

auf erweiterte Funktionalitäten scheinen seine Hauptrolle in der Branche gefährden zu können.

Debatten über OAI-PMH sind eng mit der Operationsweise institutionell verankerter Archive („institutional repositories“) verbunden. Richard Poynder hat deren, aus der Sicht mancher Skeptiker enttäuschende, Entwicklung mit Unzulänglichkeiten in Verbindung gebracht, die sich bis zum programmatischen Beginn der Initiative in der richtungsweisenden Santa-Fe-Tagung (1999) zurückverfolgen lassen<sup>27</sup>. Die halbherzige Einbeziehung des HTML-Linkmechanismus und die mangelnde Koordination der Zusatzdienste sind durch die Eigenschaften der kontextarmen, teilstandardisierten Konvention zum Metadaten-Austausch mitverursacht. Die „Confederation of Open Access Repositories“ (COAR), eine 2009 gegründete Interessengemeinschaft von 156 Partnern aus allen Bereichen der wissenschaftlichen Infrastruktur, setzt mit der Initiative „Next Generation Repositories“ bei dieser Diagnose an<sup>28</sup>. Sie greift die Ziele des Santa-Fe-Treffens in verändertem Kontext auf. Repositorien sollen die Grundlage einer global vernetzten Infrastruktur wissenschaftlicher Kommunikation werden. Das System soll stärker an der Forschung orientiert und innovationsoffen von der „scholarly community“ selbst verwaltet werden. Die programmatische Beschreibung eines umfassenden Organisationsrahmens – von der Ebene wissenschaftlicher Inhalte über deren Publikation bis hin zur Dissemination – aus dem Jahr 2022 liegt vor<sup>29</sup>. Sie enthält allerdings keine technischen oder organisatorischen Details, an denen sich die Umsetzung des Entwurfs orientieren könnte.

## 4. Politik

Der beschriebene status quo von OAI-PMH<sup>30</sup> in der Aufbereitung und Verbreitung wissenschaftlicher Fachliteratur ist durch Hinweise auf parallele Entwicklungen der Informations- und Kommunikationstechnologie in dieser Sache zu ergänzen. Technische Fortschritte beruhen auf sozialen Vorgaben, die ihre Richtung bestimmen und umgekehrt von ihnen affiziert werden. Herbert Van de Sompel, ein prominenter Akteur innerhalb der meisten hier dargestellten Initiativen, hat die Voreingenommenheit für den Fall von OAI-PMH retrospektiv und selbtkritisch formuliert:

---

27 [https://poynder.blogspot.com/2016/09/q-with-cnis-clifford-lynch-time-to-re\\_22.html](https://poynder.blogspot.com/2016/09/q-with-cnis-clifford-lynch-time-to-re_22.html)

28 <https://www.coar-repositories.org/news-updates/what-we-do/next-generation-repositories/>

29 <https://www.coar-repositories.org/news-updates/pubfair-version-2-now-available/>

30 Den Versuch einer Einschätzung unternimmt Poynder, R. (2019). Siehe auch Hrachovec, H. (2018).

It is a perspective in which a repository resembles a brick and mortar library, a library that one can go visit and that allows such visits subject to policies – the protocol – that may simultaneously be well intended and idiosyncratic. This kind of repository, although it resides on the web, hinders seamless access to its content because it does not fully embrace the ways of the web.<sup>31</sup>

Institutionen (Universitäten, Forschungsinstitute, Museen) administrieren ihre Ressourcen nach biblio-bürokratischen Richtlinien. Sie schaffen Hierarchien und garantieren dadurch weitgehende Konsistenz. Dazu benötigen sie normierte Erhebungsbögen und ein vormodelliertes Datenmanagement. Im Gegensatz dazu ist das WWW eine unübersichtliche Mega-Kollektion unterschiedlichster Inhalte, die mit solchen Mitteln nicht zu bändigen sind. Der erfolgreichen Weiterentwicklung institutioneller Archive steht im Weg, dass sich die Orientierungsmöglichkeiten im Angebot der Wissensgesellschaft verlagert haben.

Die traditionserprobten bibliographischen Metadaten, die OAI-PMH erfasst und transportiert, sind ein Instrument, Nadeln im Heuhaufen zu finden, den die Datenexpansion fortwährend vergrößert. Die mächtigen Suchmaschinen, allen voran Google, sind auf diese Situation eingestellt. Sie operieren, soweit man weiß, nicht auf der Grundlage konventioneller Datenbanken, sondern mit Big-Data-Algorithmen, die sich dem Untersuchungsfeld kontinuierlich anpassen<sup>32</sup>. Anders als OAI-PMH sind sie in der Lage, einem allenfalls vorhandenen Link auf eine Ressource automatisch zu folgen. Google hat, wie Stevan Harnad treffend bemerkt, genügend Zeit und Geld, um daraus eine hocheffektive Wissenschaftsplattform zu entwickeln, während die betroffenen Institutionen im ständigen Konkurrenzkampf um knappe öffentliche Mittel und Sponsorengelder stehen<sup>33</sup>. Peter Suber, ein Open-Access-Verfechter der ersten Stunde, kommt schon 2004 im Vergleich zwischen der Archivierung in kuratierten digitalen Archiven und der Indizierung durch Suchmaschinen zum Ergebnis, „that putting an eprint on your personal web site won't always be worse, or won't be much worse, than depositing it in an OA-OAI archive.“<sup>34</sup> Das Hauptargument für seine ernüchterte Einschätzung ist wie bei Harnad, dass

31 Van de Sompel, H.; Nelson, M. L. (2015), S. 3.

32 Stephens, O. (2012)

33 „The biggest Quasitory of all is the Virtual Quasitory called Google Scholar (GS). GS has mooted most of the fuss about interoperability because it full-text-inverts all content. It's a nuclear weapon, but it is in no hurry. Unlike institutions and funders, GS is under no financial pressure. And unlike publishers, it does not have the ambition or the need to capture and preserve publishers' obsolete, parasitic functions (even though, unlike publishers, GS is in an incomparably better position to maximise functionality on the web). GS is waiting patiently for the research community to get its act together.“ Comment von Stevan Harnad: <https://poynder.blogspot.com/2016/10/institutional-repositories-response-to.html>.

34 Suber, P. (2016), S. 6.

kontinuierlich verbesserte Lernalgorithmen den riesigen Datenmengen besser gewachsen sein könnten als die in OAI-PMH implementierten, von der Erfahrung von Bibliothekar:innen ausgehenden, in Gremien standardisierten Protokolle. Einem Weltkonzern steht die heterogene, hochspezialisierte, administrativ weitgehend autonome Vielfalt kultureller und wissenschaftlicher Sammlungen gegenüber. Der künftige Verlauf dieser Konstellation kann an dieser Stelle nicht prognostiziert werden.

Metadaten sind gegenüber den Inhalten, die sie kennzeichnen, neutral – keineswegs aber im sozio-ökonomischen Gebrauch. Sie sind ein Mittel, um Übersicht zu bewahren und Abläufe zu lenken. In der Praxis ist OAI-PMH ein Instrument zum standardisierten Datenaustausch zwischen Institutionen, überwiegend in extrakommerziellem Interesse – ein transparenter, unveräußerlicher Regelsatz zur Aufbereitung geteilter Informationen. Bekanntlich gelten diese Prinzipien weder für die global agierenden Suchmaschinen noch für die Serviceanbieter in privater Hand, die sich freier (und freiwillig bereitgestellter) Inhalte bedienen, um verkäuflichen Mehrwert zu gewinnen. Ihre Produkte basieren auf einer Infrastruktur, die im Prinzip für alle zugänglich, im Effekt jedoch von einschränkenden proprietären Interessen überlagert ist. Die Herkunft von OAI-PMH aus der selbstverwalteten Organisation kultureller und wissenschaftlicher Produkte bedingt auch seine Grenzen. Gary Hall hat darauf aufmerksam gemacht, dass die gegenwärtige mediale Umgebung tendenziell nicht mehr durch den „Besitz“ von Inhalten, sondern zunehmend durch die Ausübung von Datenkontrolle gekennzeichnet ist:

In this world who gate-keeps access to (and so can extract maximum value from) content is less important, because that access is already free, than who gate-keeps (and so can extract maximum value from) the data generated around the use of that content, which is used more because access to it is free.<sup>35</sup>

Die von G. Hall angesprochene institutionelle Verwaltung von Inhalten mit Hilfe von OAI-PMH ist eine Reminiszenz aus der Gelehrtenrepublik, die ihre Arbeitsergebnisse an Universitäten, Akademien und Archiven in eigener Machtvollkommenheit pflegte. Das Protokoll ist, wie diese Aufbewahrungsstätten selbst, im Datenuniversum ernsthafter Konkurrenz ausgesetzt. Die Transparenz und Abwesenheit von privatwirtschaftlichen Eingrenzungen, die OAI-PMH motiviert und maßgeblich ermöglicht, steht auf dem Prüfstand. Stabile Ordnungsmuster von branchenspezifischer Bedeutung finden sich, davon ist auszugehen, eingelagert in eine Unübersichtlichkeit, an der eine solche Orientierung sich zu bewähren hat.

---

35 Hall, G. (2012)

## Orientierungshilfen

### Registrierung, Validierung

Directory of Open Access Repositories: <https://v2.sherpa.ac.uk/opendoar/>

OAI-PMH Registered Data Providers: <https://openarchives.library.cornell.edu/Register/BrowseSites>

OAI-PMH Validator & Data Extractor: <https://validator.oaipmh.com/>

Open Archives Initiative - Repository Explorer: <http://oai.clarin-pl.eu/>

## Tutorials

OAI for Beginners - the Open Archives Forum Online Tutorial: <http://www.ukoln.ac.uk/metadata/oa-forum/tutorial/>

OAI-PMH Implementation: <http://eprints.rclis.org/4586/1/tutorial3muller.pdf>

OSTI OAI Repository Manual: [https://www.osti.gov/sites/www.osti.gov/files/public/OSTI\\_OAI\\_Repository\\_Manual\\_1\\_1\\_0.pdf](https://www.osti.gov/sites/www.osti.gov/files/public/OSTI_OAI_Repository_Manual_1_1_0.pdf)

Das Open Archives Initiative Protocol for Metadata Harvesting: Zielsetzung, Funktionalität, Einsatzgebiete: [https://zenodo.org/record/1253735/files/Diederichs\\_Wuttke\\_OAI\\_PMH\\_Term\\_Paper\\_2018.pdf](https://zenodo.org/record/1253735/files/Diederichs_Wuttke_OAI_PMH_Term_Paper_2018.pdf)

## Bibliografie

Becerril-García, Arianna; Aguado-López, Eduardo (2018): A Semantic Model for Selective Knowledge Discovery over OAI-PMH Structured Resources. In: *Information* 9 (6), p. 144. <https://doi.org/10.3390/info9060144>

Brungai, Lena (2011): OAI and OAI-PMH. <https://www.slideshare.net/LenaBruncaj/oai-and-oaipmh> (abgerufen am 12.01.2022)

Gaudinat, Arnaud; Beausire, Jonas; Fuss, Megan et al. (2017): Global Picture of OAI-PMH Repositories through the Analysis of 6 Key Open Archive Meta-Catalogs. <https://arxiv.org/abs/1708.08669> (abgerufen am 12.01.2022)

Deutsche Initiative für Netzwerkinformation e.V. (Hg.) (2005): Elektronisches Publizieren an Hochschulen. Inhaltliche Gestaltung der OAI-Schnittstelle – Empfehlungen. <https://edoc.hu-berlin.de/bitstream/handle/18452/2131/2-de2.pdf> (abgerufen am 14.01.2022)

Hall, Gary (2012): Has Critical Theory Run Out of Time for Data-Driven Scholarship? In: *Debates in the Digital Humanities*. <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbd1e/section/1a9b138c-eb51-4f48-bcb8-039505f88ff8> (abgerufen am 14.01.2022)

Haslhofer, Bernhard; Schandl, Bernhard (2008): The OAI2LOD Server. Exposing OAI-PMH Metadata as Linked Data. <http://events.linkeddata.org/ldow2008/papers/03-haslhofer-schandl-oai2lod-server.pdf> (abgerufen am 14.01.2022)

Herb, Ulrich (Hg.) (2012): Open Initiatives. Offenheit in der digitalen Welt und Wissenschaft – Onlineversion. <https://publikationen.sulb.uni-saarland.de/bitstream/20.500.11880/30534/1/OnlineversionOpenInitiativesUlrichHerb.pdf> (abgerufen am 14.01.2022)

Hrachovec, Herbert (2018): Zugang für alle? Rhetorik und Realität der Open Access-Initiativen. In: *Information Wissenschaft & Praxis* 69 (4), S. 161-170. <https://doi.org/10.1515/iwp-2018-0022>

Hrachovec, Herbert (2021): Jüngere Open Access Journale in Philosophie. <https://doi.org/10.5281/zenodo.5140827>

Müller, Stefan (2020): Datenaustausch schwer gemacht. Das Beispiel OAI-PMH-Schnittstellen. <https://dhmuc.hypotheses.org/2827> (abgerufen am 14.01.2022)

Naravaran, Nikesh (2010): Open Archives Initiatives for Metadata Harvesting. <https://www.slideshare.net/nikeshn/open-archives-initiatives-for-metadata-harvesting> (abgerufen am 12.01.2022).

Poynder, Richard (2019): Open Access. Could Defeat Be Snatched from the Jaws of Victory? <https://richardpoynder.co.uk/Jaws.pdf> (abgerufen am 06.04.2023)

Raemy, Julien Antoine (2020): Enabling Better Aggregation and Discovery of Cultural Heritage Content for Europeana and Its Partner Institutions. [https://julsraemy.github.io/assets/doc/Mastersthesis\\_europeana\\_raemyjulien\\_FV.pdf](https://julsraemy.github.io/assets/doc/Mastersthesis_europeana_raemyjulien_FV.pdf) (abgerufen am 14.01.2022)

Stephens, Owen (2012): What Does Google Do? CORE Blog. <https://blog.core.ac.uk/2012/03/> (abgerufen am 14.01.2022)

Suber, Peter (2016): Knowledge Unbound. Selected Writings on Open Access, 2002–2011. Cambridge, MA: MIT Press. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:26246071>

Van de Sompel, Herbert; Nelson, Michael L. (2015): Reminiscing About 15 Years of Interoperability Efforts. In: *D-Lib Magazine* 21 (11/12). <https://doi.org/10.1045/november2015-vandesompel>

Warner, Simeon (2001): Exposing and Harvesting Metadata Using the OAI Metadata Harvesting Protocol. A Tutorial. In: *High Energy Physics Libraries Webzine* 4. <https://webzine.web.cern.ch/4/papers/3/> (abgerufen am 12.01.2022)

**Herbert Hrachovec** ist ao. Univ.-Prof. i.R. an der Universität Wien. Lehraufträge für Technik- und Medienphilosophie. Publikationen aus Analytischer Philosophie, Ästhetik, Medienphilosophie und Metaphysik. <http://hrachovec.philo.at>