



openTA-Publikationsdienst:
Das openTA-Metadatenschema und
seine Umsetzung

openTA-Dokument Nr. 7

Stand: 1.4.2015, Version 2.0

Autoren: Ulrich Riehm, Clemens Döpmeier, Michael Mönnich, Len Piltz, Christian Schmitt

Dieses Dokument beschreibt das openTA Metadatenmodell für bibliographische Informationen und seine Abbildung („mapping“) auf andere gängige bibliographische Formate. Das Dokument richtet sich in erster Linie an die verantwortlichen Personen, die in den jeweiligen Institutionen, insbesondere den NTA-Mitgliedsinstitutionen, mit der Betreuung der Publikationsdaten zu tun haben, und den openTA-Publikationsdienst für die Verbreitung ihrer Publikationen nutzen wollen.

Kontakt: Ulrich Riehm über info@openta.net

Die openTA-Dokumente sind abrufbar unter der Adresse
<https://www.openta.net/veroeffentlichungen-und-openta-dokumente>

Versionsgeschichte:

06.12.2013, erster Entwurf

18.12.2013, redaktionelle Änderungen nach Teamdiskussion

29.01.2014, redaktionelle Änderungen nach Feedback aus der Community

28.02.2014, redaktionelle Änderungen nach Teamdiskussion

22.09.2014, redaktionelle Änderungen (überarbeitete Angaben zu geplantem Repositorium und semantischem Crawler, weitere Erläuterungen zu schema.org und Mapping neuer schema.org-Elemente, Korrektur in Tab. 7 openTA > BibTeX im Feld „Seite von-bis“)

Die Versionen 1.x wurden federführend von Maike Abel verantwortet.

1.4.2015, Version 2, stellt eine grundlegende Überarbeitung des Dokuments dar, in der Fehler korrigiert, insbesondere aber Erfahrungen aus dem Aufbau des openTA-Publikationsdienstes und daraus abgeleitete Veränderungen am Metadatenschema und den Mappingtabellen eingeflossen sind



Rechtehinweis: CC BY-SA 4.0

Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International



Im Rahmen des Forschungsvorhabens „Kooperativer

Aufbau eines Fachportals Technikfolgenabschätzung“, gefördert durch die DFG, durchgeführt von den KIT-Instituten ITAS, IAI und der Bibliothek.

Inhaltsverzeichnis

1	Vorwort zur Version 2	5
2	Ziele und Bestandteile des openTA-Publikationsdienstes	6
3	Datenakquise.....	7
3.1	Methoden der Datenakquise	7
3.2	Ausgangslage auf Seiten des NTA.....	7
4	Das openTA-Metadatenchema für bibliographische Informationen	9
4.1	Mindestanforderungen.....	9
4.2	Publikationstypen.....	9
4.3	Das openTA-Metadatenchema.....	10
4.4	Umgang mit Dubletten.....	14
5	Abbildung (Mapping) auf bibliographische Formate	15
5.1	Import und Export.....	15
5.2	Nutzung der Software Metafactory und Metamorph	15
5.3	Anmerkungen zu den bisher genutzten Formaten	15
5.3.1	BibTeX.....	15
5.3.2	MARC21	19
5.3.3	RIS.....	22
5.3.4	XML.....	28
5.4	Weiter in Frage kommende bibliographische Formate	29
5.4.1	BIBFRAME	29
5.4.2	DINI-Empfehlungen zur RDF-Repräsentation bibliographischer Daten.....	31
5.4.3	Dublin Core	31
5.4.4	PACITA	32
5.4.5	schema.org	32
	Anhang: Auszug aus bisher genutzten Mappingtabellen	33

Tabellenverzeichnis

Tab. 1: Elemente des openTA-Metadatenschemas	11
Tab. 2: Elemente des Metadatenschemas für die verschiedenen Publikationstypen.....	13
Tab. 3: openTA > BibTeX.....	17
Tab. 4: BibTeX > openTA.....	18
Tab. 5: BibTeX-Entry Type > openTA-Publikationstyp > BibTeX-Entry Type.....	19
Tab. 6: openTA > RIS.....	25
Tab. 7: RIS-Fields > openTA-Metadatenfeld	26
Tab. 8: RIS-Reference Types > openTA-Publikationstyp.....	28
Tab. 9: Mapping angelieferter BibTeX-Files.....	33
Tab. 10: Mapping angelieferter RIS-Files.....	34

Abbildungsverzeichnis

Abb. 1: Beispiel eines BibTeX-Publikationseintrags.....	16
Abb. 2: MARC21-XML-Repräsentation der DNB einer Monographie von Paschen et al. 2004	19
Abb. 3: Beispiel eines RIS-Publikationseintrags.....	23
Abb. 4: BIBFRAME-Repräsentation einer Monographie von Nentwich/König 2012 durch die DNB	30

1 Vorwort zur Version 2

Nach einem Prototyp vom Juni 2014 wurde im Dezember 2014 eine erste produktiv nutzbare Version des openTA-Publikationsdienstes freigeschaltet. Die Übernahme von Publikationsdaten aus heterogenen Quellen innerhalb und außerhalb des Netzwerks TA (NTA) basierte auf Überlegungen und Konzepten für ein openTA-Metadatenschema, die Ende 2013 entstanden und in verschiedenen Versionen veröffentlicht sowie im Jahr 2014 mit verschiedenen Publikationsdatensätzen umgesetzt wurden. Insbesondere mit den weit verbreiteten Publikationsformaten BibTeX und RIS konnten dabei Erfahrungen gesammelt werden, die zu einigen Korrekturen und Ergänzungen führten, die in diese Version 2 des openTA-Dokuments Nr. 7 einfließen.

2 Ziele und Bestandteile des openTA-Publikationsdienstes

Vor Beginn des openTA-Projektes gab es keinen systematischen Überblick über die im breit verstandenen Themenfeld der Technikfolgenabschätzung (TA) erscheinende Literatur. Doch schon allein bei den Institutionen des NTA werden jährlich einige Tausend Publikationen veröffentlicht. Der Nachweis dieser Literatur auf den Websites der Mitgliedsinstitutionen – soweit überhaupt vorhanden – entspricht nicht immer heutigen Standards, was die Aktualität, Vollständigkeit, Aufbereitung, Zugänglichkeit und Durchsuchbarkeit angeht. Außerdem stehen diese Quellen nur dezentral und nicht über einen zentralen Anlaufpunkt zur Verfügung.

Die openTA- Publikationsdienst hat daher zum Ziel, den Nachweis von und den Zugang zu wissenschaftlicher Literatur aus der TA-Community zu verbessern. Es soll die von den Mitgliedern des NTA publizierte Literatur (NTA-Publikationen) gebündelt und für eine breitere öffentliche Nutzung sichtbar gemacht werden. Weiterhin soll der openTA-Publikationsdienst die NTA-Institutionen darin unterstützen, literaturbezogene Metainformationen strukturiert und nach gängigen Schemata abzubilden und anderen wissenschaftlichen Publikationsdiensten zur Verfügung zu stellen (etwa BASE oder Google Scholar).

Für die openTA-Publikationsdatenbank wird ein möglichst vollständiger Nachweis der Publikationen der NTA-Mitglieder angestrebt. Darüber hinaus wird dieser Bestand erweitert um TA-relevante Publikationsquellen, etwa aus der Deutschen Nationalbibliothek (DNB) oder dem Sozialwissenschaftlichen Fachportal sowiport. Die Publikationssprache ist nicht auf Deutsch eingeschränkt.

Es wird angestrebt, möglichst viele Publikationen als frei zugängliche Volltextpublikationen nachzuweisen, ohne dass der Dienst auf solche „Open Access-Volltexte“ beschränkt wurde. Es werden auch Volltexte nachgewiesen, die nur restriktiv (z.B. gegen Bezahlung oder über Rahmenverträge) zugänglich sind.

Der openTA-Publikationsdienst ist die Basis für weitere anpassbare und portierbare Dienste. Die selektive Übernahme (Export) in Literaturverwaltungssysteme ist bereits realisiert. Angeboten wird ein Export in die Formate BibTeX oder RIS und in ein einfaches Textformat. Ein TA-Neuerscheinungsdienst, der sich aus dem Publikationsdienst speist, wurde schon testweise publiziert und soll demnächst in den Routinebetrieb überführt werden. Geplant sind des Weiteren u.a. ein abonnierbarer Profildienst als Feed oder ein Widget für die Integration des Publikationsdienstes in eine eigene Website.

Seit Dezember 2014 steht eine erste Version des openTA-Publikationsdienstes zur Verfügung. Zum Zeitpunkt der Veröffentlichung dieses Dokuments umfasst der Publikationsdienst zehn Quellen, darunter sechs NTA-Mitgliedsinstitutionen, mit insgesamt über 20.000 Literaturnachweisen. Die Erfahrungen mit dem Import dieser sehr heterogener Publikationsdatenquellen fließen in die vorliegende Überarbeitung des openTA-Dokuments Nr. 7 mit ein.

3 Datenakquise

3.1 Methoden der Datenakquise

Die openTA-Publikationsdatenbank kann prinzipiell über verschiedene Wege Publikationsmetadaten importieren. Die eleganteste Methode ist ein Abruf von Datensätzen über eine standardisierte Schnittstelle an der Quelle. In der Regel verfügen Bibliothekskataloge über solche Schnittstellen (etwa SRU, OAI-PMH, Z39.50, SPARQL Endpoint). Bei den NTA-Institutionen sind nach den bisherigen Erfahrungen diese Schnittstellen eher selten anzutreffen, liefern nur eine unzureichende Datenqualität oder die Daten sind unvollständig.

Die momentan gängigste Methode ist deshalb, dass die NTA-Institutionen ihre Publikationsdaten als File unter einem bestimmten Namen und einer vereinbarten Adresse auf ihrem Web-Server zur regelmäßigen Abholung durch openTA ablegen. Dabei werden insbesondere die verbreiteten bibliographischen Formate BibTeX (Kap. 5.3.1) oder RIS (Kap. 5.3.3) verwendet.

Ein weiterer Weg der Datenakquise ist das Crawlen (Harvesting) von Webseiten nach Publikationsinformationen mit Hilfe moderner Suchmaschinentechnologien. Dies wird gegenwärtig nicht unterstützt. Die Bedingungen für das Crawlen von Webseiten nach Publikationsdaten würden erheblich verbessert, wenn entsprechende Mikroformate oder Mikrodaten für Publikationen, die in den HTML-Code der Seite eingebettet sind, vorhanden wären (Kap. 5.4.5).

Schließlich ist vorgesehen, dass Publikationsdaten direkt über ein Web-Formular in die openTA-Publikationsdatenbank von dafür legitimierten Personen und Institutionen eingegeben werden können.

3.2 Ausgangslage auf Seiten des NTA

Die Voraussetzungen für das Harvesten von Publikationsinformationen bei den NTA-Mitgliedern sind sehr unterschiedlich.

Die wenigstens NTA-Mitgliedsinstitutionen haben ein eigenes OA-Repository oder sind an eine Institution angeschlossen, die ein OA-Repository betreibt, auf das openTA zugreifen könnte.

Einige NTA-Mitgliedsinstitutionen haben Publikationsdatenbanken für interne oder auch externe Zwecke, die eine günstige Voraussetzung für die Datenübernahme oder den Datenexport bieten. Solche Publikationsdatenbanken können ebenfalls den übergeordneten Einheiten (etwa der Universität) zugeordnet sein.

Andere NTA-Mitgliedsinstitutionen weisen ihre Publikationen nur auf statischen HTML-Seiten ohne Datenbankunterstützung, zumeist aus einem CMS heraus, nach.

Zusätzlich gibt es Mitgliedsinstitutionen, die nur die „Hauspublikationen“ (etwa eigene Berichts- und Buchreihen oder Zeitschriften) nachweisen, aber nicht die Publikationen, die „extern“ in Zeitschriften oder anderen Verlagspublikationen erscheinen.

4 Das openTA-Metadatenchema für bibliographische Informationen

Grundlage für den openTA-Publikationsdienst ist ein Metadatenchema für bibliographische Daten. Das Metadatenchema weist eine geringe Anzahl an Pflichtfeldern auf, die als Mindestanforderung von Datenlieferanten bedient werden müssen (Kap. 4.1). Diese Einschränkung auf wenige Pflichtfelder trägt der Tatsache Rechnung, dass die datenanliefernden NTA-Institutionen bisher oft über keine umfassende Publikationserfassung verfügen (Kap. 3.2). Aber erst durch ergänzende Angaben über die wenigen Pflichtfelder hinaus wird ein aussagekräftiger bibliographischer Nachweis erzielt. Und über die reinen bibliographischen „Metadaten“ zu einer Publikation hinaus werden möglichst vielfältige ergänzende Angaben zur jeweiligen Publikation angestrebt. Dies können etwa Abstracts, Schlagworte, Klassifikationen, Inhaltsverzeichnisse, Titelblätter und natürlich auch das gesamte Dokument im Volltext sein.

Darüber hinaus sollen nach Möglichkeit standardisierte Identifikationssysteme unterstützt werden, so etwa für die Publikation (etwa ISBN, ISSN, DOI) oder für Autoren (etwa ORCID, GND), an die Nutzungsfunktionen angeschlossen werden können. Solche Identifikationssysteme sind auch wichtig, um Dubletten, auf die man unweigerlich stoßen wird, auszuschließen (Kap. 4.4).

4.1 Mindestanforderungen

Für die Aufnahme von Publikationsnachweisen in die openTA-Publikationsdatenbank sind nur wenige Angaben verpflichtend. Diese Pflichtangaben unterscheiden sich je nach Publikationstyp. Es handelt sich um Angaben zu

1. Autor(en) oder Herausgeber(n),
2. Titel oder Bandtitel und das
3. Erscheinungsjahr.

Durch die Reduktion auf ein Minimalset an Pflichtangaben soll sichergestellt werden, dass die anliefernden Institutionen keine aufwändigen Änderungen an ihrer Publikationsverwaltung vornehmen müssen und auch bei einer relativ schlechten Datenqualität Publikationsdaten zur Verfügung gestellt werden können. Diese Mindestanforderungen orientieren sich im Übrigen an denen von Google Scholar.

Mindestanforderungen heißt aber auch, dass openTA – und die zuliefernden Institute – keineswegs mit diesen drei minimalen Angaben zufrieden sein sollten. Vielmehr wird eine möglichst umfassende Beschreibung der betreffenden Publikation angestrebt. Diese sind im openTA-Metadatenchema niedergelegt (Tab. 1).

4.2 Publikationstypen

openTA unterscheidet im Metadatenchema für bibliographische Informationen die folgenden Publikationstypen:

1. Monographie: Unter einer Monographie wird eine selbständige, in der Regel Verlagspublikation eines Autors oder eines Autorenteam zu einem Thema verstanden.

2. Sammelband: Unter dem Publikationstyp „Sammelband“ wird ein Buch (selbständige Literatur) verstanden, das weiter untergliedert ist in Aufsätze oder sonstige Beiträge, deren Autoren jeweils gesondert ausgewiesen sind (unselbständige Literatur). Ein Sammelband kann z.B. ein Tagungsband, eine Festschrift oder ein Proceedings-Band sein.
3. Aufsatz aus Sammelband
4. Periodikum: Ein Periodikum kann sowohl eine bei einem Verlag veröffentlichte (wissenschaftliche) Zeitschrift als auch ein regelmäßig erscheinender Newsletter eines Instituts sein. Mit dem Typ „Periodikum“ soll insbesondere der Nachweis von Periodika, die von NTA-Institutionen herausgegeben werden, ermöglicht werden.
5. Aufsatz in Periodikum
6. Bericht: Bei dem Publikationstyp „Bericht“ kann es sich um Jahresberichte, Projektabschlussberichte, Forschungsberichte u.ä. handeln, die typischerweise nicht in einem Verlag erschienen sind. Dieser Publikationstyp wird deshalb oft zum Genre der „grauen Literatur“ gezählt. Die Grenzen zwischen Verlagsbuch- und Berichtspublikationen sind allerdings fließend.
7. (unveröffentlichter) Vortrag: Während veröffentlichte Vorträge, etwa in Proceedings, dem Publikationstyp „Aufsatz aus Sammelband“ zugeordnet werden, stellen unveröffentlichte Vorträge einen eigenen Publikationstyp dar. Angegeben werden möglichst die oder der Vortragende, der Vortragstitel, Ort, Anlass und Datum des Vortrags. Verweise auf Abstracts oder Folien zum Vortrag sind möglich.
8. Sonstiges: Unter „Sonstiges“ fallen alle Publikationen, die keiner der obigen Kategorien (eindeutig) zugeordnet werden können. Dabei kann es sich z.B. um einen Blog-Eintrag, ein Interview (in einem Audio-Format) oder ein Vorlesungsskript handeln.

Diese Publikationstypen decken alle für die TA-Community relevanten wissenschaftsüblichen Veröffentlichungsformen ab. Zu jedem Publikationstyp gibt es spezifische Metadatenfelder (Tab. 2).

4.3 Das openTA-Metadatenchema

Tabelle 1 listet alle Elemente des openTA-Metadatenchemas für bibliographische Informationen auf. Die Hinweise zu den Inhalten der einzelnen Elemente („Ansetzungsregeln“) sind als Empfehlungen für das Erstellen von Datensätzen zu verstehen, die openTA zur Verfügung gestellt werden. Die Tabelle enthält außerdem Hinweise darauf, ob Datenfelder mehrfach verwendet werden können oder nicht.

- Als Zeichensatz sollte UTF-8 verwendet werden.
- In den Freitext-Datenfeldern kann grundsätzlich eine unbegrenzte Anzahl an Zeichen eingegeben werden.
- openTA empfiehlt, sich bei der Eingabe von Informationen an etablierten Standards zu orientieren, z.B. bei der Angabe der Publikationssprache.

Tab. 1: Elemente des openTA-Metadatenchemas

openTA-Metadatenelement	„Ansetzungsregel“, erlaubte Feldwerte, Ausfüllhinweis	Mehrfachverwendung
Publikationstyp	Auswahlliste mit folgenden Werten: <ul style="list-style-type: none"> ▪ Monographie ▪ Sammelband ▪ Aufsatz aus Sammelband ▪ Periodikum ▪ Aufsatz in Periodikum ▪ Bericht ▪ (unveröffentlichter) Vortrag ▪ Sonstiges 	nein
Autor(en)	Nachname, Vorname	ja
Herausgeber	Nachname, Vorname	ja
Titel	Freitextfeld	ja
Bandtitel	Freitextfeld	ja
Erscheinungsjahr	numerisch, YYYY	ja
Verleger/institutioneller Herausgeber	Freitextfeld	ja
Verlagsort/Ort der Herausgeber	Freitextfeld	ja
Angaben zur Reihe	Freitextfeld	nein
Auflage	Freitextfeld	nein
Titel des Periodikums	Freitextfeld	nein
Jahrgang	alpha-numerisch	nein
Heft	Freitextfeld	nein
Artikelnummer	Freitextfeld	nein
Seite von-bis	Freitextfeld	nein
Angaben zur Veranstaltung	Freitextfeld	nein
Abstract/Beschreibung	Freitextfeld	ja
Schlagnote/Klassifikation	Freitextfeld; Empfehlung: Verwendung eines kontrollierten Vokabulars (z.B. DDC, GND) in der Form: DDC: 323.48	ja
Sprache	Angabe nach RFC1766 und ISO 639-1 (zweistellig) oder ISO 639-2 (dreistellig)	ja
Anmerkungen	Freitextfeld	nein
Link auf Volltext	URL	ja
Link auf sonstige die Publikation beschreibende Materialien	URL mit Freitext; Empfehlung: URL mit Bezeichner angeben, z.B. Abstract: URL, Inhaltsverzeichnis: URL	ja
Link auf Originaldaten in der Quelle	URL	nein
DOI	Angabe in der Form: doi:10.ORGANISATION/ID	nein
Autorenidentifikation	Angabe je nach Form des Identifiers; Empfehlung: zweiteilige Angabe, d.h. Typ und ID, etwa: LCCN: no00088944 (dies orientiert sich an Wikipedia)	ja
ISBN	10- oder 13-stellig; Empfehlung: 13-stellig, Angabe mit Bindestrichen	nein
ISSN	alpha-numerisch; Empfehlung: mit Bindestrich	nein

Eindeutige Kennzeichnung des Datensatzes bei der Quelle (ID)	Freitext	nein
Rechtehinweis	Freitextfeld; Empfehlung: maschinenlesbare, normierte Angabe wie z.B. Creative Commons; es kann sich um Lizenzangaben für die Publikation oder für den Metadatenatz handeln	ja
Medientyp	Auswahlliste mit folgenden Werten: <ul style="list-style-type: none"> ▪ Druckwerk ▪ Netzwerkressource ▪ Datenträger 	nein
Inhaltstyp	Auswahlliste mit folgenden Werten: <ul style="list-style-type: none"> ▪ Rezension ▪ Editorial ▪ begutachtet ▪ Hochschulschrift 	ja
Erstellungs-/Übernahmedatum	Systemfeld, das von openTA automatisch gefüllt wird bzw. von einem Bearbeiter manuell gefüllt werden kann.	nein
Datum der letzten Änderung	Systemfeld, das von openTA automatisch gefüllt wird bzw. von einem Bearbeiter manuell gefüllt werden kann.	nein
Ursprüngliche Quelle (Person/Institution)	Systemfeld, das von openTA automatisch gefüllt wird.	nein
Zuletzt geändert von	Systemfeld, das von openTA automatisch gefüllt wird.	nein
Sonstige Angaben	Systemfeld, das von openTA automatisch gefüllt wird bzw. von einem Bearbeiter manuell gefüllt werden kann.	nein

Die folgende Tabelle 2 verdeutlicht, dass nicht jedes Metadatenelement für jeden Publikationstyp sinnvollerweise verwendet werden kann. Außerdem wird zwischen verpflichtenden und optionalen Angaben unterschieden. Auch wenn nur wenige Angaben als verpflichtend gekennzeichnet sind, heißt dies nicht, dass diese Minimalangaben für ausreichend gehalten werden. Angestrebt wird immer die Übernahme der reichhaltigsten, aktuellsten und besten verfügbaren Daten.

Legende:

P	= Pflicht
O	= Optional
-	= trifft nicht zu

Tab. 2: Elemente des Metadatenchemas für die verschiedenen Publikationstypen

Publikationstyp Element	Mono- graphie	Sam- mel- band	Aufsatz aus Sam- mel- band	Perio- dikum	Aufsatz in Perio- dikum	Bericht	(un- veröf- fent- lichter) Vortrag	Sonsti- ges
Autor(en)	P	-	P	-	P	P	P	P
Herausgeber		P	O	O	O	O	-	O
Titel	P		P	O	P	P	P	P
Bandtitel		P	O	-	-	O	-	O
Erscheinungsjahr	P	P	P	-	P	P	P	P
Verleger/institutioneller Herausgeber	O	O	O	O	O	O	-	O
Verlagsort/Ort der Heraus- geber	O	O	O	O	O	O	-	O
Angaben zur Reihe	O	O	O	-	O	O	-	O
Auflage	O	O	O	-	O	O	O	O
Titel des Periodikums	-	-	-	P	O	-	-	-
Jahrgang	O	-	-	O	O	-	-	-
Heft	-	-	-	O	O	-	-	-
Artikelnummer	-	-	-	O	O	-	-	-
Seite von-bis	O	O	O	-	O	O	O	O
Angaben zur Veranstaltung	O	O	O	O	O	O	O	O
Abstract/Beschreibung	O	O	O	O	O	O	O	O
Schlagworte/Klassifikation	O	O	O	O	O	O	O	O
Sprache	O	O	O	O	O	O	O	O
Anmerkungen	O	O	O	O	O	O	O	O
Link auf Volltext	O	O	O	O	O	O	O	O
Link auf sonstige die Publikation beschreibende Materialien	O	O	O	O	O	O	O	O
Link auf Originaldaten in der Quelle	O	O	O	O	O	O	O	O
DOI	O	O	O	O	O	O	O	O
Autorentifikation	O	O	O		O	O	O	O
ISBN	O	O	O	-	-	O	-	O
ISSN	O	O	O	O	O	O	-	O
Eindeutige Kennzeichnung des Datensatzes bei der Quelle (ID)	O	O	O	O	O	O	O	O
Rechtehinweis	O	O	O	O	O	O	O	O
Medientyp:	-	-	-	-	-	-	-	-
<i>Druckwerk</i>	O	O	O	O	O	O	O	O
<i>Netzwerkressource</i>	O	O	O	O	O	O	O	O
<i>Datenträger</i>	O	O	O	O	O	O	O	O
Inhaltstyp:	-	-	-	-	-	-	-	-
<i>Rezension</i>	-	-	O	O	O	O	-	O
<i>Editorial</i>	-	-	O	-	O	O	-	O
<i>begutachtet</i>	O	O	O	O	O	O	-	O
<i>Hochschulschrift</i>	O	O	O	O	O	O	-	O
Erstellungs- /Übernahmedatum	O	O	O	O	O	O	O	O
Datum der letzten Ände- rung	O	O	O	O	O	O	O	O
Ursprüngliche Quelle (Person/Institution)	O	O	O	O	O	O	O	O
Zuletzt geändert von	O	O	O	O	O	O	O	O
Sonstige Angaben	O	O	O	O	O	O	O	O

4.4 Umgang mit Dubletten

Wenn Publikationen in verschiedenen Quellen nachgewiesen werden, kommt es zu „Dubletten“ von bibliographischen Metainformationen zu einer Publikation. Es ist dabei nicht ungewöhnlich, dass sich die Metainformationen für die gleiche Publikation mehr oder weniger deutlich unterscheiden, da die Sichtweisen auf eine Publikation unterschiedlich sein können. Die Dublette bezieht sich dann nicht auf die Metadaten, sondern auf das von den Metadaten beschriebene Objekt. Es ist nicht einfach zu entscheiden, wie solche (unechten) „Dubletten“ verschiedener Metadatenbeschreibungen zu einer Publikation, obwohl sie z.B. anhand der ISBN oder des DOI – wenn verfügbar – einfach identifiziert werden könnte, behandelt werden sollen. Der – noch nicht implementierte, aber derzeit angedachte – Lösungsansatz von openTA sieht vor, die verschiedenen Metadatenätze zu einer Publikation in einer Gruppe zusammenzufassen, so dass dem Nutzer bei Interesse alle Metadatenätze zur Verfügung stehen.

Ein etwas anders gelagertes „Dublettenproblem“ liegt vor, wenn ein Datenlieferant einen Datensatz bei sich ändert (aktualisiert) und den überarbeiteten Datensatz zu dem gleichen Titel erneut an openTA liefert. Sollte es keine eindeutige ID des entsprechenden Datensatzes mit einem Updatedatum geben, muss von openTA die Veränderung des Datensatzes automatisch erkannt werden, was gegenwärtig über den Vergleich von Hashwerten geschieht.

5 Abbildung (Mapping) auf bibliographische Formate

5.1 Import und Export

Das openTA-Metadatenchema ist eine im Projektkontext operative Zwischenschicht für den Im- und Export aus bzw. in bibliographische Standardformate und sollte nicht verstanden werden, als ein Vorschlag für einen neuen bibliographischen Standard. Die im Folgenden aufgeführten Tabellen enthalten für den Import aus und den Export in bibliographische Standardformate die Abbildungsrelationen („mapping“) zwischen den Elementen des jeweiligen Formats. Datensätze, die in die openTA Datenbank importiert werden, werden komplett mit allen Feldinhalten bei openTA abgespeichert. Feldinhalte, die nicht eindeutig auf das openTA-Schema gemappt werden können, bleiben bestehen. Sie können wieder komplett und unverändert exportiert werden.

5.2 Nutzung der Software Metafacture und Metamorph

Im Rahmen des bei der DNB angesiedelten Projektes Culturegraph entstand Metafacture, ein Baukasten für die Metadatenverarbeitung. Ein Bestandteil ist die Datentransformationssprache Metamorph.¹ Metafacture und Metamorph wird im Rahmen von openTA für das Mapping von bibliographischen Formaten auf das openTA-Metadatenchema erfolgreich genutzt.

5.3 Anmerkungen zu den bisher genutzten Formaten

Im openTA-Projekt wurden bisher Publikationsquellen für den openTA-Publikationsdienst in den folgende Formate verarbeitet:

- BibTeX
- MARC21
- RIS
- XML

Während MARC21 ein ausdifferenzierter bibliothekarischer Standard darstellt, sind BibTeX und RIS im Kontext des Umgangs mit wissenschaftlicher Literatur zwar weit verbreitet, aber nicht mehr als eine sich ständig verändernde, relativ variable Konvention. XML ist eine allgemeine Auszeichnungssprache, die auch für bibliographische Daten genutzt werden kann, dafür aber keine spezifischen Elemente und Vorgaben kennt.

5.3.1 BibTeX

BibTeX ist ein Programm zur Erstellung von Literaturangaben und -verzeichnissen in TeX- oder LaTeX-Dokumenten, das in den 1980er Jahren von Oren Patashnik entwickelt wurde.² Das zugrundeliegende BibTeX-Format hat eine einfache Syntax und umfasst ein Set von Publikationstypen („entry types“) und zugehörigen beschreibenden Feldern („fields“).

¹ Siehe u.a. Geipel (2012): Metamorph User Guide.
[https://github.com/fsteeg/culturegraph-documentation/blob/master/src/user-guide.tex](https://github.com/fsteeg/culturegraph-documentation/blob/master/src/user-guide.tex;);
<https://github.com/culturegraph/metafacture-core/wiki>, <http://culturegraph.github.io/>,
<https://wiki.dnb.de/display/CULTUREGRAPH/Software#Software-Metamorph>, 12.1.2015.

² Siehe z.B. <http://www.ctan.org/tex-archive/biblio/bibtex/base> und
<http://de.wikipedia.org/wiki/BibTeX>

Das Format hat sich durch die einfache Struktur und seine relativ große Verbreitung von dem ursprünglichen Zweck, Erstellung von Literaturverzeichnissen in wissenschaftlichen Publikationen, gelöst und ist (auch) zu einem Format für den Austausch bibliographischer Informationen geworden. Es wird von vielen Literaturverwaltungsprogrammen, wie etwa Endnote oder Citavi, für den Im- und Export unterstützt. Viele Bibliothekskataloge bieten ebenfalls einen BibTeX-Export an.

Abb. 1: Beispiel eines BibTeX-Publikationseintrags

```
@article{e4dc3ea5-9476-3664-a7be-7eb24a0d4245,  
  author = "Wang, Yufei and Höller, Samuel and Viebahn, Peter and Hao, Zhengping",  
  title = "Integrated assessment of CO2 reduction technologies in China's cement industry",  
  journal = "International journal of greenhouse gas control",  
  year = "2014",  
  abstract = "The main objective of this article is to evaluate CO2 mitigation potential and to  
  calculate costs avoided by the use of different CO2 mitigation technologies in China's cement  
  sector, namely energy efficiency improvements, use of alternative fuels, clinker substitution  
  and carbon capture and storage (CCS). Three scenarios are designed based on the projection of  
  cement output and technology development over the next 40 years (2010-2050). 2.5, 4.7 and 4.3  
  Gt tonnes of CO2 will be saved totally in basic scenario and two low carbon scenarios up to  
  2050. By comparing these technologies along the scenarios, it can be concluded that CO2  
  emissions can mainly be reduced by energy efficiency improvements and use of alternative fuels.  
  Clinker substitution, which reduces the clinker-to-cement ratio as well as energy intensity, results  
  in significant cost advantages. CCS, including post-combustion capture and oxy-fuel combustion capture,  
  could play an important role in the capture of CO2 in the cement industry, and is expected to be in  
  commercial use by 2030.",  
  language = "EN",  
  doi = "10.1016/j.ijggc.2013.10.004",  
  pages = "27 - 36",  
  volume = "20"  
}
```

Quelle: OpenTA-Publikationsdienst, 8.1.2015

<https://www.openta.net/publikation?id=e4dc3ea5-9476-3664-a7be-7eb24a0d4245>

BibTeX ist weniger ein bibliographischer Standard als eine über viele Jahre im Gebrauch und in Veränderung befindliche Konvention. Es gibt kein „Normierungsgremium“, das das aus den 1980er Jahren stammende Schema aktualisieren würde. Als maßgebliche Referenz kann Oren Patashniks Dokument „BibTeXing“ aus dem Jahr 1988 gelten. Seit dem sind viele neue „BibTeX-fields“ in Verwendung, die von openTA berücksichtigt werden, soweit sie für die Abbildung des openTA-Metadatenschemas als wichtig und in mehreren Quellen als gebräuchlich erscheinen. Diese Anpassungsfähigkeit ist Vor- und Nachteil zugleich.

Nachteile des BibTeX-Formats sind vor dem Hintergrund der bisherigen Erfahrungen im openTA-Projekt insbesondere, dass BibTeX zwei Publikationstypen („entry types“) nicht kennt, die das openTA-Metadatenschema vorsieht, und auch von den uns zuliefernden Quellen bedient werden: Sammelband und (nicht publizierter) Vortrag. Sammel- oder Herausgeberbände werden bei BibTeX typischerweise als „book“ oder als „proceedings“ gekennzeichnet. Aber nicht jeder Sammelband ist ein Konferenzband und bei Büchern wird typischerweise zwischen Monographien eines Autors oder eines Autorentams und Büchern mit Beiträgen verschiedener Autoren, die von einem oder mehreren Herausgebern betreut wurden, unterschieden. Vorträge können innerhalb der Standardvorgaben nur als entry type „misc“ gekennzeichnet werden, was unbefriedigend ist.

In einzelnen Fällen wurden deshalb zusätzliche openTA-spezifische Felder in das BibTeX-Format aufgenommen, um die beiden Publikationstypen kenntlich zu machen. Dazu muss die zuliefernde Institution in der Lage sein, das BibTeX-File zu verändern und zu ergänzen.

Sollte ein solcher Bedarf und eine solche Möglichkeit bestehen, dann sollte direkt mit dem openTA-Projekt-Team über info@openTA.net Kontakt aufgenommen werden.

Für die Publikationstypen „Vortrag“ und „Sammelband“ wird vorgeschlagen, ein openTA-spezifisches Feld gemäß der BibTeX-Syntax zusätzlich in die BibTeX-Beschreibung aufzunehmen („ota_publtyp = „talk“,“ und „ota_publtyp = „collection“,“), aber vom „standardisierten“ BibTeX-Entry Type (für den Sammelband „@book“, für den Vortrag „@misc“) nicht abzuweichen.

Tab. 3: openTA > BibTeX

openTA Element	BibTeX Element
Autor(en)	author
Herausgeber	editor
Publikationstyp	Entry Type (Tab. 5)
Titel	title
Bandtitel	booktitle
Erscheinungsjahr	year
Verleger/institutioneller Herausgeber	publisher
Verlagsort/Ort der Herausgeber	address
Auflage	edition
Angaben zur Reihe	series
Titel des Periodikums	journal
Jahrgang	volume
Heft	number
Artikelnummer	note
Seite von-bis	pages
Angaben zur Veranstaltung	organization
Abstract/Beschreibung	abstract
Schlagworte/Klassifikation	keywords
Sprache	language
Anmerkungen	note
Link auf Volltext	url
Link auf sonstige die Publikation beschreibende Materialien	url
Link auf Originaldaten in der Quelle	url
DOI	doi
Autorenidentifikation	orcid
ISBN	isbn
ISSN	issn
Eindeutige Kennzeichnung des Datensatzes bei der Quelle (ID)	note
Rechteinweis	copyright
Medientyp	howpublished
<i>Druckwerk</i>	-
<i>Netzwerkressource</i>	-
<i>Datenträger</i>	-
Inhaltstyp	note
<i>Rezension</i>	note
<i>Editorial</i>	note
<i>begutachtet</i>	note
<i>Hochschulschrift</i>	s. Tab. 10 Entry Type Masterthesis oder Phdthesis
Erstellungs-/Übernahmedatum	<i>kein Mapping, da es sich um Angaben zum Datensatz und nicht zu der Ressource handelt; wird von openTA gefüllt</i>
Datum der letzten Änderung	<i>kein Mapping, da es sich um Angaben zum Datensatz und nicht zu der Ressource handelt; wird von openTA gefüllt</i>
Ursprüngliche Quelle (Person/Institution)	<i>kein Mapping, da es sich um Angaben zum Datensatz und nicht zu der Ressource handelt; wird von openTA gefüllt</i>
Zuletzt geändert von	<i>kein Mapping, da es sich um Angaben zum Datensatz und nicht zu der Ressource handelt; wird von openTA gefüllt</i>
Sonstige Angaben	note

Tab. 4: BibTeX > openTA

BibTeX-Element	openTA Element
abstract	Abstract/Beschreibung
address	Verlagsort/Ort der Herausgeber
affiliation	Anmerkungen
annotate	Anmerkungen
author	Autor
booktitle	Titel
chapter	Anmerkungen
contents	Abstract/Beschreibung
copyright	Rechteinweis
crossref	Anmerkungen
doi	DOI
edition	Auflage
editor	Herausgeber
Entry Types	Publikationstyp (Tab. 5)
eprint	Netzwerkressource
howpublished	Medientyp
institution	Verleger/ institutioneller Herausgeber
isbn	ISBN
issn	ISSN
journal	Titel des Periodikums
key	Anmerkungen
keywords	Schlagworte/Klassifikation
language	Sprache
LCCN	Schlagworte/Klassifikation
month	Anmerkungen
location	Verlagsort/Ort der Herausgeber
note	Anmerkungen
number	Heft
orcid	Autorenidentifizier
organization	Angaben zur Veranstaltung
pages	Seite von-bis (bei Aufsätzen oder Anzahl der Seiten bei einem Bericht, Monographie etc.)
price	Anmerkungen
publisher	Verlag/institutioneller Herausgeber
school	Anmerkungen
series	Reihe
size	Anmerkungen
title	Titel
type	Publikationstyp
ota_publtyp	Publikationstyp (ota_publtyp für die präzisere Abbildung der Publikationstypen über die vorhandenen Entry Types von BibTeX hinaus)
url	Link auf sonstige die Publikation beschreibende Materialien
volume	Jahrgang
year	Erscheinungsjahr

Tab. 5: BibTeX-Entry Type > openTA-Publikationstyp > BibTeX-Entry Type

BibTeX-Entry Type	openTA-Publikationstyp	BibTeX
@article	Aufsatz in Periodikum	@article
@book	Monographie	@book
@book and ota_publtyp = „collection“	Sammelband	@book
@booklet	Bericht	@techreport
@conference	Aufsatz aus Sammelband	@inbook
@inbook	Aufsatz aus Sammelband	@inbook
@incollection	Aufsatz aus Sammelband	@incollection
@inproceedings	Aufsatz aus Sammelband	@inproceedings
@manual	Bericht	@techreport
@mastersthesis	Publikationstyp = Sonstiges (soweit nicht anders erkennbar) Inhaltstyp = Hochschulschrift	@masterthesis (keine eindeutige Zuordnung ob master- oder phdthesis möglich, wenn keine weitere Information verfügbar)
@misc	Sonstiges	@misc
@misc and „ota_publtyp = „talk“	Vortrag	@misc
@phdthesis	Publikationstyp = Sonstiges (soweit nicht anders erkennbar) Inhaltstyp = Hochschulschrift	@phdthesis(keine eindeutige Zuordnung ob master- oder phdthesis möglich, wenn keine weitere Information verfügbar)
@proceedings	Sammelband	@book
@techreport	Bericht	@techreport
@unpublished	Sonstiges	@misc

Die Tabelle enthält das Mapping für die Publikationstypen von BibTeX nach openTA und von openTA nach BibTeX. Für den Publikationstyp „Vortrag“ und „Sammelband“ sind openTA-spezifische BibTeX-Felder eingeführt, da die vorgegebenen BibTeX-Entry Types diese beiden Publikationstypen nicht kennen. Bei der Rückumwandlung von openTA nach BibTeX geht diese spezifische Information verloren. Dies ist auch beim Mapping anderer Entry Types auf openTA-Publikationstypen (etwa Bericht und Inhaltstyp = Hochschulschrift) der Fall.

5.3.2 MARC21

MARC (MACHINE-Readable Cataloging) ist ein weitverbreitetes bibliographisches Datenformat, das vor allem zwischen Bibliotheken zum Austausch von Katalogdaten verwendet wird. Seien Ursprünge gehen in die 1960er Jahre zurück. MARC21 ist die aktuelle Version, die seit 1999 eine Reihe von MARC-Dialekte vereinheitlicht hat. Auch die Deutsche Nationalbibliothek (DNB) liefert ihre Daten seit 2009 im MARC21-Format aus.

MARC21 ist ein sehr stark ausdifferenziertes Format, das für die Beschreibung und Übertragung von „einfachen“ Publikationsdaten zu komplex ist. Die „Feldbeschreibung der Titeldaten der Deutschen Nationalbibliothek und der Zeitschriftendatenbank im Format MARC 21“ umfasst allein 379 Seiten.³

Im Folgenden – in Ausschnitten – die MARC21-XML-Repräsentation des Buches: Paschen, H.; Coenen, Chr.; Fleischer, T.; Grünwald, R.; Oertel, D.; Revermann, Chr.: Nanotechnologie. Forschung, Entwicklung, Anwendung. Berlin, Heidelberg, New York: Springer 2004.

Abb. 2: MARC21-XML-Repräsentation der DNB einer Monographie von Paschen et al. 2004

```
<record xmlns="http://www.loc.gov/MARC21/slim" type="Bibliographic">
  <leader>02973pam a2200433 c 4500</leader>
  <controlfield tag="001">970163886</controlfield>
  <controlfield tag="003">DE-101</controlfield>
  <controlfield tag="005">20130723093228.0</controlfield>
  <controlfield tag="007">tu</controlfield>
```

³ <http://nbn-resolving.de/urn:nbn:de:101-2014070106>

```

<controlfield tag="008">040212s2004          gw  |||||  ||||
00||||ger</controlfield>
<datafield tag="015" ind1=" " ind2=" ">
  <subfield code="a">04,A33,1193</subfield>
  <subfield code="z">04,N11,0716</subfield>
  <subfield code="2">dnb</subfield>
</datafield>
<datafield tag="016" ind1="7" ind2=" ">
  <subfield code="2">DE-101</subfield>
  <subfield code="a">970163886</subfield>
</datafield>
<datafield tag="020" ind1=" " ind2=" ">
  <subfield code="a">3540210687</subfield>
  <subfield code="c">Pp. : EUR 99.95, sfr 158.00</subfield>
  <subfield code="9">3-540-21068-7</subfield>
</datafield>
<datafield tag="024" ind1="3" ind2=" ">
  <subfield code="a">9783540210689</subfield>
</datafield>
<datafield tag="035" ind1=" " ind2=" ">
  <subfield code="a">(DE-599)DNB970163886</subfield>
</datafield>
<datafield tag="035" ind1=" " ind2=" ">
  <subfield code="a">(OCoLC)249520110</subfield>
</datafield>
<datafield tag="040" ind1=" " ind2=" ">
  <subfield code="a">1145</subfield>
  <subfield code="b">ger</subfield>
  <subfield code="c">DE-101</subfield>
  <subfield code="d">1210</subfield>
  <subfield code="e">rakwb</subfield>
</datafield>
<datafield tag="041" ind1=" " ind2=" ">
  <subfield code="a">ger</subfield>
</datafield>
<datafield tag="044" ind1=" " ind2=" ">
  <subfield code="c">XA-DE-BE</subfield>
</datafield>
<datafield tag="082" ind1="7" ind2="4">
  <subfield code="a">620</subfield>
  <subfield code="q">DE-101</subfield>
  <subfield code="2">22sdb</subfield>
</datafield>
<datafield tag="084" ind1=" " ind2=" ">
  <subfield code="a">620</subfield>
  <subfield code="q">DE-101</subfield>
  <subfield code="2">sdb</subfield>
</datafield>
<datafield tag="090" ind1=" " ind2=" ">
  <subfield code="a">b</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Nanotechnologie</subfield>
  <subfield code="b">Forschung, Entwicklung, Anwendung</subfield>
  <subfield code="c">H. Paschen ...</subfield>
</datafield>
<datafield tag="260" ind1="3" ind2=" ">
  <subfield code="a">Berlin</subfield>
  <subfield code="a">Heidelberg</subfield>
  <subfield code="a">New York</subfield>
  <subfield code="b">Springer</subfield>

```

```

    <subfield code="c">2004</subfield>
  </datafield>
  <datafield tag="300" ind1=" " ind2=" ">
    <subfield code="a">XV, 366 S.</subfield>
    <subfield code="b">Ill., graph. Darst.</subfield>
    <subfield code="c">24 cm</subfield>
  </datafield>
  <datafield tag="500" ind1=" " ind2=" ">
    <subfield code="a">Literaturverz. S. 327 - 359</subfield>
  </datafield>
  <datafield tag="520" ind1=" " ind2=" ">
    <subfield code="a">Wegen ihres Potenzials zur grundlegenden Veränderung
... Nanotechnologie.- Handlungsbedarf.- Literatur.- Anhänge.</subfield>
  </datafield>
  <datafield tag="650" ind1=" " ind2="7">
    <subfield code="0">(DE-588)4327470-5</subfield>
    <subfield code="0">(DE-101)043274706</subfield>
    <subfield code="a">Nanotechnologie</subfield>
    <subfield code="2">gnd</subfield>
  </datafield>
  <datafield tag="650" ind1=" " ind2="7">
    <subfield code="0">(DE-588)4078176-8</subfield>
    <subfield code="0">(DE-101)040781763</subfield>
    <subfield code="a">Technikbewertung</subfield>
    <subfield code="2">gnd</subfield>
  </datafield>
  <datafield tag="653" ind1=" " ind2=" ">
    <subfield code="a">(VLB-FS)Nanotechnologie</subfield>
  </datafield>
  <datafield tag="653" ind1=" " ind2=" ">
    <subfield code="a">(VLB-FS)Nanotechnik</subfield>
  </datafield>
  <datafield tag="653" ind1=" " ind2=" ">
    <subfield code="a">(VLB-FS)Forschungsaktivitäten</subfield>
  </datafield>
  <datafield tag="653" ind1=" " ind2=" ">
    <subfield code="a">(VLB-PF)BB: Gebunden</subfield>
  </datafield>
  <datafield tag="689" ind1="0" ind2="0">
    <subfield code="0">(DE-588)4327470-5</subfield>
    <subfield code="0">(DE-101)043274706</subfield>
    <subfield code="D">s</subfield>
    <subfield code="a">Nanotechnologie</subfield>
  </datafield>
  <datafield tag="689" ind1="0" ind2="1">
    <subfield code="0">(DE-588)4078176-8</subfield>
    <subfield code="0">(DE-101)040781763</subfield>
    <subfield code="D">s</subfield>
    <subfield code="a">Technikbewertung</subfield>
  </datafield>
  <datafield tag="689" ind1="0" ind2=" ">
    <subfield code="5">DE-101</subfield>
    <subfield code="5">DE-101</subfield>
  </datafield>
  <datafield tag="700" ind1="1" ind2=" ">
    <subfield code="0">(DE-588)170786110</subfield>
    <subfield code="0">(DE-101)170786110</subfield>
    <subfield code="a">Paschen, Herbert</subfield>
    <subfield code="d">1933-</subfield>
    <subfield code="4">ctb</subfield>
  </datafield>

```

```
<datafield tag="856" ind1="4" ind2="2">
  <subfield code="m">V:DE-576</subfield>
  <subfield code="q">application/pdf</subfield>
  <subfield code="u">http://d-nb.info/970163886/04</subfield>
  <subfield code="3">Inhaltsverzeichnis</subfield>
</datafield>
<datafield tag="925" ind1="r" ind2=" " >
  <subfield code="a">ra</subfield>
</datafield>
</record>
```

Quelle: DNBt, <http://d-nb.info/970163886/about/marcxml>, 23.3.2015

Das Mapping von MARC21-Daten auf das openTA-Metadatenschema ist nicht nur wegen der Komplexität der verwendeten Datenfelder aufwändig, sondern auch deshalb problematisch, weil die Vielfalt der Datenfelder auch eine Vielfalt von Verwendungsformen erlaubt, die erkannt und „richtig“ behandelt werden müssen. Zudem weichen die Verwendungsformen unter den „MARC21-Bibliotheken“ voneinander ab. Trotzdem gelten MARC21-Daten von (National-)Bibliotheken als der „Goldstandard“.

Da der openTA-Publikationsdienst, neben Publikationsdaten aus den NTA-Mitgliedsinstitutionen, auch sonstige TA-Literatur aufnimmt, kommen dafür als repräsentative und hochwertige Importquelle diese Bibliotheken in Frage. Ein Export aus dem openTA-Publikationsdienst in MARC21 ist jedoch nicht vorgesehen. Eine openTA-MARC21-Mapping-Tabelle wird wegen der speziellen Problematik hier nicht angeboten.

5.3.3 RIS

Das RIS-Format ist weit verbreitet, um Titelinformationen von bibliographischen Online-Datenbanken sowie Bibliothekskatalogen in Literaturverwaltungsprogramme zu importieren. Das Format ist – ähnlich wie das BibTeX-Format – relativ einfach aufgebaut: nach einem Bezeichner, der aus zwei Großbuchstaben besteht, folgen zwei Leerzeichen, dann ein Bindestrich und ein weiteres Leerzeichen (Abb. 3). Dann folgen die eigentlichen Feldinhalte. Das RIS-Format wird von dem Unternehmen Thomson Reuters gepflegt.⁴ Die im Folgenden verwendeten Quellen für die RIS-Reference Types und RIS-Fields sind Thomson Reuters: Reference Manager 12. User's Guide. 2008, (<http://www.refman.com/support/docs/ReferenceManager12.pdf>, 14.1.2015) und die RIS Format Specifications (http://www.refman.com/support/risformat_intro.asp, mit einem Excel-File „EN RIS Documentation.xls“ zuletzt geändert am 9.5.2011, <http://refman.com/support/direct%20export.zip>, 14.1.2015). Letztere weist deutlich mehr Reference Types und Fields auf als der User's Guide. Ähnlich wie BibTeX wird RIS von vielen Literaturverwaltungsprogrammen für den Im- und Export unterstützt sowie von Bibliothekskatalogen als Ausgabeformat angeboten.

⁴ Siehe z.B. http://www.refman.com/support/risformat_intro.asp, http://de.wikipedia.org/wiki/RIS_%28Dateiformat%29, <http://refman.com/support/direct%20export.zip> sowie [Appendix C und D im Handbuch Reference Manager 12](http://www.refman.com/support/docs/ReferenceManager12.pdf) <http://www.refman.com/support/docs/ReferenceManager12.pdf>, 8.1.2015.

Abb. 3: Beispiel eines RIS-Publikationseintrags

```

TY - BOOK
AU - Caviezel, Claudio
TI - Climate Engineering
TI - Kann und soll man die Erderwärmung technisch eindämmen?
Y1 - 2014
U1 - U2
ED - edition sigma
CY - Berlin
LA - DE
UR - http://deposit.d-nb.de/cgi-bin/dokserv?id=4805933&prov=M&dok_var=1&dok_ext=htm
SN - 978-3-8360-8141-2
M1 - Studien des Büros für Technikfolgen-Abschätzung beim Deutschen Bundestag
SP - 336 S.
Y2 - 41
ER -

```

Quelle: OpenTA-Publikationsdienst, 9.1.2015
<https://www.openta.net/publikation?id=7b606d6f-0007-36c8-9a9a-d282284e63f6>

RIS unterscheidet im Reference Manager 12 User's Guide aus dem Jahr 2008 35 Publikationstypen, deutlich mehr als das BibTeX-Format mit nur 14 „entry types“. Allerdings findet man unter diesen „reference types“ auch solche, die für den openTA-Publikationsdienst nicht von Bedeutung und entsprechend nicht im openTA-Metadatenschema vorgesehen sind, etwa „audiovisual material“, „bill“, „hearing“, „motion picture“, „music“, „sound“, „unenacted bill“.

(Nicht publizierte) Vorträge könnten als PCOMM (personal communication) oder UNPB (unpublished work) gekennzeichnet werden.

Innerhalb der 35 Reference Types des User's Guide fehlt ein Publikationstyp für Herausgeber- oder Sammelbände. Allerdings ist dieser in den „RIS Format Specifications“, die 56 Reference Types umfassen, als „Edited Book“ (EDBOOK) enthalten.⁵

Auch in Bezug auf die Datenfelder („tags“) weist RIS eine deutliche größere Komplexität als BibTeX auf. Dies erlaubt zwar präzisere Beschreibungen, führt aber auch – ähnlich wie bei MARC21 – dazu, dass die Variabilität der Nutzung der Datenfelder zwischen den unterschiedlichen RIS-Quellen – gegebenenfalls auch innerhalb einer Quelle – steigt. Der Herausgeber eines Buches, um nur ein Beispiel zu nennen, kann im Feld „ED“ oder auch „A2“ stehen (Tab. 10).

Ob BibTeX oder RIS als „besseres“ Format zu bevorzugen wäre, ist schwierig zu entscheiden. Dies kann letztlich nur vor dem Hintergrund der konkreten Publikationsdaten einer bestimmten Quelle beurteilt werden.

Im Folgenden noch einige Erfahrungen aus den bisher importierten RIS-Formaten (vgl. hierzu auch Tab. 10):

⁵ Thomson Reuters (2008): „RIS“ format documentation. Adding a „direct export“ button to your web page or web application, S. 8 f., <http://refman.com/support/direct%20export.zip>, 14.1.2015

- Die Erzeuger des RIS-Formats an der Publikationsdatenquelle benutzen nach unseren bisherigen Erfahrungen, den eigentlich im erweiterten Satz der RIS-Reference Type vorhandenen Reference Type EDBOOK für den Sammelband (Herausgeberband, Sammelwerk) nicht. Beim Import von RIS-Daten muss dieser deshalb über andere Merkmale erschlossen werden.
- Die RIS-Fields werden in Abhängigkeit eines Reference Types unterschiedlich genutzt. So kann etwa T2 bei einem Aufsatz aus einem Sammelband den Titel des Sammelbandes enthalten, bei einem Buch oder Sammelband den Reihentitel einer Buchreihe.
- Die Vielfalt der verfügbaren RIS-Reference Types und RIS-Fields führt dazu, dass unterschiedliche Quellen ihre Metadaten unterschiedlichen RIS-Reference Types oder RIS-Fields zuordnen.

Tab. 6: openTA > RIS

openTA Element	RIS Element
Publikationstyp	TY
<i>Monographie</i>	BOOK
<i>Sammelband</i>	EDEBOOK
<i>Aufsatz aus Sammelband</i>	CHAP
<i>Periodikum</i>	JFULL
<i>Aufsatz in Periodikum</i>	JOUR
<i>Bericht</i>	RPRT
<i>(unveröffentlichter) Vortrag</i>	PCOMM
<i>Sonstiges</i>	GEN
Autor(en)	AU
Herausgeber	ED
Titel	TI
Bandtitel	BT
Erscheinungsjahr	Y1
Verleger/institutioneller Herausgeber	PB
Verlagsort/Ort der Herausgeber	CY
Angaben zur Reihe	U2 ⁶
Auflage	ET
Titel des Periodikums	JF
Jahrgang	VL
Heft	IS
Artikelnummer	C7
Seite von-bis	SP und EP
Angaben zur Veranstaltung	U3
Abstract/Beschreibung	AB
Schlagworte/Klassifikation	KW
Sprache	LA
Anmerkungen	N1
Link auf Volltext	UR
Link auf sonstige die Publikation beschreibende Materialien	UR
Link auf Originaldaten in der Quelle	UR
DOI	DO
Autorenidentifikation	U1
ISBN	SN
ISSN	SN
Eindeutige Kennzeichnung des Datensatzes bei der Quelle (ID)	AN oder ID
Rechtehinweis	AV
Medientyp	U1
<i>Druckwerk</i>	U2
<i>Netzwerkressource</i>	ELEC
<i>Datenträger</i>	DATA
Inhaltstyp	U5
<i>Rezension</i>	U5
<i>Editorial</i>	U5
<i>begutachtet</i>	U5
<i>Hochschulschrift</i>	TY - THES
Erstellungs-/Übernahmedatum	<i>kein Mapping, da es sich um Angaben zum Datensatz und nicht zu der Ressource handelt; wird von openTA gefüllt</i>
Datum der letzten Änderung	<i>kein Mapping, da es sich um Angaben zum Datensatz und nicht zu der Ressource handelt; wird von openTA gefüllt</i>
Ursprüngliche Quelle (Person/Institution)	<i>kein Mapping, da es sich um Angaben zum Datensatz und nicht zu der Ressource handelt; wird von openTA gefüllt</i>
Zuletzt geändert von	<i>kein Mapping, da es sich um Angaben zum Datensatz und nicht zu der Ressource handelt; wird von openTA gefüllt</i>
Sonstige Angaben	N1

⁶ Die Literaturverwaltungssoftware CITAVI z.B. erkennt beim Import U2 nicht als Angabe zur Reihe. Citavi exportiert den Sammelband als Book und schreibt die Reihenangabe in T2.

Tab. 7: RIS-Fields > openTA-Metadatenfeld

RIS Element	openTA Element
A1	Autor
A2	Herausgeber (bei Aufsatz aus Sammelband CHAP und Sammelband BOOK oder CONF)
A3	Angaben zur Reihe (Reihenherausgeber bei Aufsatz aus Sammelband CHAP und Sammelband BOOK oder CONF)
AB	Abstract/Beschreibung
AD	Anmerkungen (AD enthält die Adresse des Autors)
AN	Eindeutige Kennzeichnung des Datensatzes bei der Quelle (ID)
AU	Autor (bei EDBOOK der Herausgeber)
AV	Rechteinweis
BT	Bandtitel
C1	Anmerkungen (C1 wird je nach Reference Type sehr unterschiedlich gebraucht)
C2	Anmerkungen (C2 wird je nach Reference Type sehr unterschiedlich gebraucht)
C3	Anmerkungen (C3 wird je nach Reference Type sehr unterschiedlich gebraucht)
C4	Anmerkungen (C4 wird je nach Reference Type sehr unterschiedlich gebraucht)
C5	Anmerkungen (C5 wird je nach Reference Type sehr unterschiedlich gebraucht)
C6	Anmerkungen (C6 wird je nach Reference Type sehr unterschiedlich gebraucht)
C7	Artikelnummer bei Aufsatz in Periodikum oder Anmerkungen
C8	Anmerkungen (C8 wird je nach Reference Type sehr unterschiedlich gebraucht)
CA	Anmerkungen oder Titel (Caption)
CN	Anmerkungen(Call Number)
CP	Verlagsort/Ort des Herausgebers [nur aufgeführt im Reference Manager 12 User's Guide]
CT	Anmerkungen („caption“, Bildunterschrift oder Titel)
CY	Verlagsort/Ort des Herausgebers
DA	Anmerkungen (meist für „dated accessed“ oder andere Datumsangaben)
DB	Anmerkungen („name of database“)
DO	DOI
DP	Anmerkungen („database provider“)
ED	Herausgeber
EP	Seite von-bis (hier nur die „ending page number“)
ER	ignorieren (ein Publikationsformat abschließender „tag“)
ET	Auflage
ID	Eindeutige Kennzeichnung des Datensatzes bei der Quelle (ID) [nur aufgeführt im Reference Manager 12 User's Guide]
IS	Heft (bei Aufsatz in Periodikum), Anzahl Bände in einer Reihe (bei Aufsatz aus Sammelband)
J1	Titel des Periodikums (J1 soll Kürzel enthalten)
J2	Titel des Periodikums (J2 soll Kürzel enthalten)
JA	Titel des Periodikums (JA soll Kürzel enthalten)
JF	Titel des Periodikums (JF soll den vollen Titel des Periodikums enthalten)
JO	Titel des Periodikums (JO soll Kürzel enthalten)
KW	Schlagworte/Klassifikation
L1	Anmerkungen (L1 enthält Angaben zu einem „file attachment“)
L2	Link auf Volltext
L3	Link auf sonstige die Publikation beschreibende Materialien
L4	Anmerkungen (L4 enthält Angaben zu „figures“.)
LA	Sprache
LB	Anmerkungen („label“)
M1	Anmerkungen (je nach „reference type“ eine „number“ unterschiedlicher Art)
M2	Seite von-bis (hier nur die „start page“ beim „reference type“ „journal article“)
M3	Anmerkungen („type of article“ beim „reference type“ „journal article“)
N1	Anmerkungen
N2	Abstract/Beschreibung
NV	Anmerkungen (meist für „number of volumes“)
OP	Anmerkungen (meist für „original publication“)
PB	Verleger/institutioneller Herausgeber
PY	Erscheinungsjahr
RI	Inhaltstyp: begutachtet (meist für „reviewed item“)
RN	Anmerkungen („research notes“)
RP	Auflage („reprint edition“)
SE	Anmerkungen (SE wird je nach „reference type“ sehr unterschiedlich gebraucht.)
SN	ISBN oder ISSN (je nach „reference type“ oder auch eine andere ID-Nummer)
SP	Seite von-bis (in der Regel „start page“)

ST	Titel („short title“)
T1	Titel [nur aufgeführt im Reference Manager 12 User's Guide]
T2	Bandtitel (bei Aufsatz aus Sammelband), Reihenangabe (bei Monographie)
T3	Reihenangabe („series title“)
TA	Anmerkungen („translated author“)
TI	Titel
TT	Anmerkungen („translated title“)
TY	Publikationstyp (s.u.)
U1	Anmerkungen („user definable“) [nur aufgeführt im Reference Manager 12 User's Guide]
U2	Anmerkungen („user definable“) [nur aufgeführt im Reference Manager 12 User's Guide]
U3	Anmerkungen („user definable“) [nur aufgeführt im Reference Manager 12 User's Guide]
U4	Anmerkungen („user definable“) [nur aufgeführt im Reference Manager 12 User's Guide]
U5	Anmerkungen („user definable“) [nur aufgeführt im Reference Manager 12 User's Guide]
UR	Link auf sonstige die Publikation beschreibende Materialien
VL	Jahrgang („volume“)
Y1	Erscheinungsjahr [nur aufgeführt im Reference Manager 12 User's Guide]
Y2	Anmerkungen („access date“)

Die Angaben zum Gebrauch der RIS-Elemente orientieren sich an „EN RIS Documentation.xls“ zuletzt geändert am 9.5.2011, (<http://refman.com/support/direct%20export.zip>, 14.1.2015). Die erhaltenen RIS-Files für openTA weichen davon teilweise deutlich ab (Tab. 10).

Tab. 8: RIS-Reference Types > openTA-Publikationstyp

ABST	Abstract	Sonstiges
ADVS	Audiovisual Material	Sonstiges
ART	Artwork	Sonstiges
BILL	Bill/Resolution	Sonstiges
BOOK	Book, Whole	Monographie
CASE	Case	Sonstiges
CHAP	Book chapter / Book Section	Aufsatz aus Sammelband
CTLG	Catalog	Sonstiges
COMP	Computer program	Sonstiges
CONF	Conference Proceedings	Sammelband
CPAPER	Conference Paper	Aufsatz aus Sammelband
DATA	Data file	Publikationstyp = Sonstiges Medientyp = Datenträger
EBOOK	Electronic book	Publikationstyp = Monographie Medientyp = Netzwerkressource
ECHAP	Electronic Book Section	Publikationstyp = Aufsatz aus Sammelband Medientyp = Netzwerkressource
EDBOOK	Edited book	Sammelband
EJOUR	Electronic Article	Publikationstyp = Aufsatz in Periodikum Medientyp = Netzwerkressource
ELEC	Electronic Citation	Sonstiges
GEN	Generic	Sonstiges
HEAR	Hearing	Sonstiges
ICOMM	Personal Communication / Internet Communication	Publikationstyp = Aufsatz in Periodikum Medientyp = Netzwerkressource
INPR	In press	Aufsatz in Periodikum
JFULL	Journal (full)	Periodikum
JOUR	Journal article	Aufsatz in Periodikum
MANSCPT	Manuscript	Sonstiges
MAP	Map	Sonstiges
MGZN	Magazine article	Aufsatz in Periodikum
MPCT	Motion picture	Sonstiges
MUSIC	Music score	Sonstiges
NEWS	Newspaper Article	Aufsatz in Periodikum
PAMP	Pamphlet	Sonstiges
PAT	Patent	Sonstiges
PCOMM	Personal communication	Sonstiges oder Vortrag
RPRT	Report	Bericht
SER	Serial (Book, Monograph)	Periodikum
SLIDE	Slide	Sonstiges
SOUND	Sound recording	Sonstiges
STAT	Statute	Sonstiges
THES	Thesis/Dissertation	Publikationstyp = Sonstiges (soweit nicht anders erkennbar) Inhaltstyp = Hochschulschrift
UNBILL	Unenacted bill/resolution	Sonstiges
UNPB	Unpublished work	Sonstiges
VIDEO	Video recording	Sonstiges

Die Angaben zu den Reference Types sind entnommen aus Reference Manager 12 User's Guide und ergänzt um einige relevante Reference Types aus „EN RIS Documentation.xls“ zuletzt geändert am 9.5.2011 (<http://refman.com/support/direct%20export.zip>, 14.1.2015).

5.3.4 XML

XML ist kein bibliographisches Format, sondern eine allgemein einsetzbare Auszeichnungssprache für die Strukturierung von Daten, deren Syntax durch das World Wide Web Consortium (W3C) spezifiziert wurde. Im Abschnitt über MARC21 wurde schon gezeigt, dass der MARC21-Export der DNB als XML-File ausgegeben wird. XML wird als Ausgabesprache auch von speziellen Datenbanken genutzt, deren bibliographische Formate zwar

nicht weit verbreitet, aber deren Inhalte für openTA von inhaltlicher großer Relevanz sind. Entsprechende spezielle XML-Import-Routinen wurden deshalb für die sozialwissenschaftliche TA-Literatur von Gesis-Sowiport⁷ erstellt als auch für die sogenannte TA-Datenbank, die von 1988 bis 2000 vom ITAS in Kooperation mit dem FIZ Karlsruhe aufgebaut wurde.⁸ Für diese Spezialfälle werden keine Mapping-Tabellen aufgeführt.

5.4 Weiter in Frage kommende bibliographische Formate

Die genannten Formate sind bibliothekarische Standards und/oder finden häufig bei den NTA-Institutionen Verwendung. Eine Reihe weiterer bibliographischer Datenformate müssen auf Grund ihrer Verbreitung oder wegen ihrer besonderen Relevanz für TA-Publikationen über die bereits eingesetzten hinaus in Betracht gezogen werden. Dazu zählen:

- BIBFRAME
- DINI-Empfehlungen zur RDF-Repräsentation bibliographischer Daten
- Dublin Core
- PACITA
- schema.org

5.4.1 BIBFRAME

BIBFRAME ist eine Initiative der Library of Congress aus dem Jahr 2011. Es geht dabei um einen neuen Ansatz für bibliographische Informationen und den Austausch bibliographischer Daten, der die MARC21-Welt ablösen und bibliographische Daten über den engeren Kontext der Bibliothekswelt für das Internet bereitstellen soll. Kernelemente des BIBFRAME Modells sind die Unterscheidung zwischen einem (geistigen) Werk und seiner Realisierung sowie die Orientierung auf die Verlinkung über kontrollierte Identifier im Sinne des Semantic Web. BIBFRAME schlägt ein eigenes Vokabular vor⁹, das in RDF definiert ist, und folgt damit nicht dem Ansatz, bestehende Vokabulare für Linked Open Data (LOD) für neue Anwendungsfälle in spezifischen Kombinationen zu nutzen und nur gegebenenfalls zu ergänzen.¹⁰

BIBFRAME wird auch von der Deutschen Nationalbibliothek (DNB) im Rahmen der international besetzten Early Experimenters Group unterstützt und begleitet.¹¹ Die DNB bietet bereits prototypisch BIBFRAME-Repräsentationen für ihre Katalogdaten an (Abb. 4). Für openTA sind diese sicherlich noch weit in die Zukunft reichenden Entwicklungen auch deshalb interessant, da Linked Open Data-Ansätze auf im Rahmen von openTA auf ihre Anwendungsrelevanz hin überprüft werden sollen.

⁷ Böhle, Knud (2014): Technikfolgenabschätzung als Gegenstand sozialwissenschaftlicher Literatur. Eine annotierte Bibliografie deutschsprachiger Quellen von 1978 bis 2013. Köln: GESIS-Leibniz Institut für Sozialwissenschaften, https://www.openta.net/documents/10179/34933/Recherche-Spezial_TA_Druckversion/.

⁸ TA-Datenbank nicht mehr online verfügbar. In: TA-Datenbank-Nachrichten, Nr. 1, 10. Jg., März 2001, S. 160 f., https://www.tatup-journal.de/tadn011_news05.php, 9.1.2015.

⁹ <http://bibframe.org/vocab-list/>, 13.1.2015.

¹⁰ Grundlegend: Library of Congress (2012): Bibliographic framework as a web of data: Linked data model an supporting services. Washington, <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>, 13.1.2015; weitere, aktuelle Informationen: <http://www.loc.gov/bibframe/> <http://bibframe.org/>.

¹¹ <http://www.dnb.de/DE/Wir/Projekte/Laufend/bibframe.html>, 13.1.2015.

Es ist nicht überraschend, dass der BIBFRAME-Ansatz auch seine Kritiker hat. So kritisiert das US-amerikanische National Information Standards Organization (NISO) die zu starke Orientierung an den speziellen Interessen und Praktiken der Library of Congress, wodurch Communities wie die Dublin Core Metadata Initiative, die IFLA oder schema.org zu wenig berücksichtigt würden.¹²

Abb. 4: BIBFRAME-Repräsentation einer Monographie von Nentwich/König 2012 durch die DNB

```
<rdf:RDF xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:bf="http://bibframe.org/vocab/"
xmlns:bfpr="http://bibframe.org/bfpr/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:dcterms="http://purl.org/dc/terms/">
  <bf:Instance rdf:about="http://d-nb.info/1011915952">
    <bf:modeOfIssuance>Einbändiges Werk</bf:modeOfIssuance>
    <bf:instanceOf rdf:resource="http://d-nb.info/bf_temp/work_1011915952" />
    <bf:isbn13>9783593395180</bf:isbn13>
    <bf:isbn10>3593395185</bf:isbn10>
    <bf:ean>9783593395180</bf:ean>
    <bf:nbn>11,N22</bf:nbn>
    <bf:nbn>12,A38</bf:nbn>
    <bf:local>
      <bf:identifierScheme>VLB / Netzpublikationen</bf:identifierScheme>
      <bf:identifierValue>3819524</bf:identifierValue>
    </bf:local>
    <bf:responsibilityStatement>Michael Nentwich ; René
König</bf:responsibilityStatement>
    <bf:extent>X, 237 S.</bf:extent>
    <bf:dimensions>22 cm</bf:dimensions>
    <bf:illustrativeContentNote>Ill., graph. Darst.</bf:illustrativeContentNote>
    <bf:note>Literaturangaben</bf:note>
    <bf:title>Cyberscience 2.0</bf:title>
    <bf:providerEntity>
      <bf:providerName>Campus-Verl.</bf:providerName>
      <bf:providerPlace>Frankfurt, M., New York, NY</bf:providerPlace>
      <bf:providerDate>2012</bf:providerDate>
    </bf:providerEntity>
    <bf:frequency rdf:resource="http://marc21rdf.info/terms/continuingfre%23/u" />
  </bf:Instance>
  <bf:Work rdf:about="http://d-nb.info/bf_temp/work_1011915952">
    <bf:hasInstance rdf:resource="http://d-nb.info/1011915952" />
    <bf:associatedAgent>
      <bf:Person>
        <bf:hasGNDLink rdf:resource="http://d-nb.info/gnd/171839854" />
        <bf:label>Nentwich, Michael</bf:label>
        <bf:resourceRole rdf:resource="http://id.loc.gov/vocabulary/relators/aut" />
      </bf:Person>
    </bf:associatedAgent>
    <bf:subject rdf:resource="http://d-nb.info/gnd/7548364-6" />
    <bf:subject rdf:resource="http://d-nb.info/gnd/4255573-5" />
    <bf:class-ddc>001.2</bf:class-ddc>
```

¹² Carpenter (2012): A Study to Determine the Needs and Requirements for Extending the Usability of the New Bibliographic Framework into the Global Networked Information Environment, http://www.niso.org/apps/group_public/document.php?document_id=9975&wg_abbrev=ccm, 13.1.2015. Kritik im Detail an der Nutzung von Identifier etwa bei Pohl (2014): Name authority files & linked data. In: Übertext: Blog, 10.7.2014, <http://www.uebertext.org/2014/07/name-authority-files-linked-data.html>, 13.1.2015.

```

    <bf:class-ddc>001.2</bf:class-ddc>
    <bf:title>Cyberscience 2.0</bf:title>
  </bf:Work>
</rdf:RDF>

```

Quelle: DNB-Katalog, <http://d-nb.info/1011915952/about/lds>, 29.1.2015

5.4.2 DINI-Empfehlungen zur RDF-Repräsentation bibliographischer Daten

Die Arbeitsgruppe Titeldaten des Kompetenzzentrums Interoperable Metadaten der Deutschen Initiative für Netzwerkinformation (DINI-AG KIM) hat im September 2013 eine Empfehlung zur RDF-Repräsentation bibliographischer Daten vorgelegt. Damit strebt sie eine Harmonisierung der RDF-Repräsentationen von Titeldaten im deutschsprachigen Raum an und will darüber möglichst einen Quasi-Standard etablieren. Im Vergleich zu BIBFRAME unterscheidet sich der DINI-Vorschlag u.a. dahingehend, dass bereits vorliegende „namespaces“ berücksichtigt werden und kein neuer – wie bei BIBFRAME – vorgeschlagen wird. Bei der Auswahl vorliegender Ontologien richtet man sich nach deren Verbreitung. Die folgende Liste stellt die berücksichtigten „namespaces“ nach der (vermuteten) Häufigkeit der Nutzung in absteigender Reihenfolge dar:

1. Dublin Core terms
2. Dublin Core elements 1.1
3. Bibliographic Ontology (BIBO)
4. Resource Description and Access Vocabularies (RDA)
5. International Standard Bibliographic Description (ISBD)
6. RDF Schema
7. Upper Mapping and Binding Exchange Layer (UMBEL)
8. schema.org: Library extension terms
9. OWL Web Ontology Language
10. ISBD Mediatype

Die DINI- und die BIBFRAME-Initiative verfolgen das gemeinsame Ziel, bibliographische Daten für die Nutzung im (semantischen) Web als Linked Open Data bereitzustellen. Dieses Ziel wird auch von openTA geteilt.

5.4.3 Dublin Core

Von der Dublin Core Metadata Initiative (DCMI) wurde ein Set an Elementen entwickelt, um Dokumente und andere Objekte im Internet zu beschreiben. Die „DCMI Metadata Terms“ empfehlen über die 15 „core elements“ hinaus zusätzliche Felder sowie detaillierende Felder (element refinements), die eine auf speziellere Bedürfnisse zugeschnittene Beschreibung bzw. Kategorisierung erlauben. Alle Felder sind optional, können mehrfach auftauchen und in beliebiger Reihenfolge stehen.¹³ Dublin Core ist als Metadatenformat für bibliographische Informationen zwar in einigen Aspekten nicht ausreichend detailliert, allerdings ist es insbesondere im Kontext von Open Access Publikationsrepositorien weit verbreitet.

¹³ Siehe z.B. <http://dublincore.org/documents/dces/>, <http://dublincore.org/documents/dcmi-terms/> und http://de.wikipedia.org/wiki/Dublin_Core

Die bisherigen Erfahrungen im Projekt openTA waren allerdings dahingehend, dass Dublin Core-Metadaten für Publikationen aus Repositorien weniger feingliedrig und inhaltsreich waren als verfügbare Alternativen. Da bisher immer inhaltsreichere Format-Alternativen aus Bibliothekskatalogen zur Verfügung standen, wurde Dublin Core bisher nicht eingesetzt.

5.4.4 PACITA

PACITA (Parliaments and Civil Society in Technology Assessment) war von 2011 bis 2015 EU-gefördertes Projekt. Es hatte zum Ziel, die Möglichkeiten von wissenschaftlicher Politik zu Themen aus Wissenschaft, Technologie und Innovation und deren institutionelle Basis zu stärken. Am Projekt waren 15 regionale bzw. nationale parlamentarische Stellen für Technikfolgenabschätzung beteiligt.¹⁴ Die deutschen, österreichischen und Schweizer PACITA-Partner sind auch Mitglied im NTA. Im Rahmen des PACITA-Projekts wurde ein Portal für die Technikfolgenabschätzung entwickelt.¹⁵ Das Portal weist u.a. Publikationen von europäischen TA-Institutionen nach¹⁶ und hat ein eigenes Format für den Austausch von bibliographischen Daten entwickelt, das relativ „grobkörnig“ ist und die Erzeugung eines JSON-Files voraussetzt. Wegen dieser „Grobkörnigkeit“ wird das Pacita-Format nicht von openTA unterstützt. Bei Bedarf könnte es allerdings für den Austausch zwischen dem PACITA-TA-Portal und openTA verwendet werden.

5.4.5 schema.org

schema.org ist eine Initiative von Google, Microsoft und Yahoo, die 2011 ins Leben gerufen wurde. Ziel ist es, semantische HTML-Tags zu entwickeln, die maschinenlesbar sind. Durch Einbinden von schema.org-Tags in Webseiten können Inhalte durch Suchmaschinen besser (semantisch) ausgelesen und interpretiert werden. Webseiten werden mit einem Mark-Up versehen, wobei die Formate RDFa, JSON-LD oder Mikrodata verwendet werden können. Dadurch wird einem definierten Teil einer Webseite eine bestimmte Bedeutung zugemessen. schema.org ist im Entwicklungsstadium.¹⁷ Es findet zurzeit kein Export von openTA nach schema.org statt. Für openTA wurden die Eigenschaften (properties) der schema.org-Elemente „Thing“, „Creative Work“, „Book“, „Article“, „Scholarly Article“, „TechArticle“, „NewsArticle“ und „Review“ gelistet. Es ist davon auszugehen, dass die weiteren Klassen von „Creative Work“ (z.B. „Movie“, „Painting“, „TVSeries“) für openTA keine Rolle spielen. Es gibt Initiativen, schema.org für bibliographische Zwecke zu erweitern, z.B. SchemaBibEx.¹⁸

¹⁴ Siehe <http://www.pacitaproject.eu/>

¹⁵ Siehe <http://www.technology-assessment.info>

¹⁶ Siehe <http://www.technology-assessment.info/index.php/publications>

¹⁷ Siehe z.B. <http://schema.org/> und <http://de.wikipedia.org/wiki/Schema.org>

¹⁸ Siehe z.B. <http://www.w3.org/community/schemabibex/> und <http://www.niso.org/publications/isq/2013/v25no4/wallis/>

Anhang: Auszug aus bisher genutzten Mappingtabellen

Die folgenden Tabellen geben einen Einblick in die vielfältige Nutzung der Formate BibTeX und RIS, wie sie bisher von den Kooperationspartnern für den openTA-Publikationsdienst bereitgestellt wurden.

Tab. 9: Mapping angelieferter BibTeX-Files

	KIT-ITAS	Wuppertal Institut	Öko-Institut
Primärautor(en)		author,	author
Autor(en)	author (mehrere Autoren getrennt durch „“)	author (mehrere Autoren getrennt durch "and")	author (mehrere Autoren getrennt durch "and")
Publikationstyp	@xxx und openTA-Spezialformat für collection „ota_publtyp = „collection“,“ und talk: „ota_publtyp = „talk“,“	"@xxx" (wenn Entrytyp @book und keine Angabe für "author" aber Angaben für "editor", dann Publikationstyp „Sammelband“)	"@xxx" und „ota_publtyp = „xxx“,“
Herausgeber	editor	editor	(editor)
Titel	title	title	title
Bandtitel	booktitle	booktitle, wird verwendet bei EntryTyp @incollection oder @inproceedings für den Titel des Sammelbands/Proceedings	(booktitle)
Erscheinungsjahr	year	year	year
Verleger/institutioneller Herausgeber	publisher	publisher	publisher
Verlagsort/Ort der Herausgeber	address	address	(address)
Angaben zur Reihe	series	series	(series)
Auflage	(edition)	edition	(edition)
Angaben zur Veranstaltung	organization	(organization)	(organization)
Anmerkungen	(note)	(note)	(note)
Titel des Periodikums	journal	journal	(journal)
Jahrgang	volume	volume	(volume)
Heft	number	number	(number)
Nummer des Artikels	-	-	-
Seite von-bis	pages, in der Variante "8-10", nur für Aufsätze aus Periodika und Sammelbänden	pages, in der Variante {14 - 17} für Aufsätze, oder in der Variante {253} für Bücher, Berichte	(pages)
Abstract/Beschreibung	abstract	abstract	abstract
Schlagnote/Klassifikation	(keywords)	(keywords)	keywords
Sprache	language	language, zweistellig nach ISO 639-1	language, zweistellig nach ISO 639-1
Link auf Volltext	url	url	url
Link auf sonstige die Publikation beschreibende Materialien	url	url	(url)
Link auf Quelle	url	url	(url)
DOI	doi	doi	(doi)
Autorenidentifikation	-	-	-
ISBN	isbn	isbn	isbn
ISSN	issn	(issn)	(issn)
Quellen-ID	-	-	@publtyp{xxxxx,
Rechtehinweis	copyright	(copyright)	(copyright)
Medientyp	-	-	-
Inhaltstyp	-	-	ota_inhaltstyp
Erstellungs-/Übernahmedatum	automatische Systeminformation	automatische Systeminformation	
Datum der letzten	automatische Sys-	automatische Systemin-	

	KIT-ITAS	Wuppertal Institut	Öko-Institut
Änderung	teminformation	formation	
Ursprüngliche Quelle (Person/Institution)	automatische Systeminformation	automatische Systeminformation	
Zuletzt geändert von	automatische Systeminformation	automatische Systeminformation	
Sonstige Angaben	-	-	

() Angaben in Klammern werden momentan nicht geliefert.
 - kein entsprechendes BibTeX-Feld vorhanden.

Tab. 10: Mapping angelieferter RIS-Files

	FhG-ISI	ITA Wien	IZT Berlin
Publikationstyp	TY (Da Sammelband nicht als eigener Typ kodiert ist, muss er erschlossen werden. Vortrag ist „TY - PCOMM“)	TY (Da Sammelband nicht als eigener Typ kodiert ist, muss er erschlossen werden: if ((authors is empty) and (editors is not empty)) then pubType = COLLECTIVE_VOLUME)	TY (TY - PCOMM ist Vortrag; NEWS ist Aufsatz in Periodikum)
Autor(en)	AU	AU	AU
Herausgeber	ED	Wenn Sammelband, dann Hrsg = A3, wenn TY CHAP oder CPAPER dann Herausgeber = A2, ED wird nicht verwendet.	Wenn Sammelband, dann Hrsg = A2, wenn CHAP dann Hrsg = ED
Titel	TI	TI (des Buches, des Aufsatzes)	T1
Bandtitel	TI oder BT	Wenn Sammelband, dann Bandtitel=TI; wenn CHAP oder CPAPER, dann Bandtitel T2; T2 aber auch Titel der Zeitschrift bei TY JOUR oder TY MGZN.	Bei Sammelband T1, bei Aufsatz aus Sammelband (CHAP) T2
Erscheinungsjahr	PY	PY (DA liefert zusätzlich ein Tagesdatum, wird ignoriert)	PY oder Y2
Verleger/ institutioneller Herausgeber	PB	PB (bei TY RPRT steht in PB das Gleiche wie in A4, deshalb A4 ignorieren)	PB
Verlagsort/Ort der Herausgeber	CY	CY (bei mehreren Orten mit / getrennt)	CY
Angaben zur Reihe	T3	T3, bei TY RPTR kommt auch die Kombination N1 + SN oder nur SN vor	Bei BOOK (Monographie oder Sammelband) T2, bei CHAP T3
Auflage	(ET)	(ET) Angaben zur Auflage finden sich auch in TI, PB, T2, AB	ET (aber auch mit anderen Angaben belegt)
Titel des Periodikums	JO oder BT	Wenn TY JOUR oder MGZN dann T2, T2 aber auch bei CPAPER und CHAP der Bandtitel!	Wenn TY JOUR, dann T2; wenn TY NEWS dann JF
Jahrgang	VL	VL	VL (enthält teilweise auch eine Bandzählung von Reihentitel)
Heft	IS	IS	IS
Nummer des Artikels	-	-	-
Seite von-bis	bei Artikel SP und EP (von-	SP (EP wird nicht	SP und EP

	FhG-ISI	ITA Wien	IZT Berlin
	bis), bei Buch, Bericht nur SP (Anzahl Seiten)	verwendet)	
Angaben zur Veranstaltung	U1	In C1 gibt es Angaben zum Ort der Veranstaltung beim TY CPAPER	-
Anmerkungen	-	-	TS (mit Angaben zur ursprünglichen Quelle)
Abstract/ Beschreibung	AB	AB	AB (aber selten)
Schlagnworte/Klassifikation	KW	(KW)	KW (mehrere Schlagworte mit Komma abgetrennt)
Sprache	LA, zweistellig nach ISO 639-1	(LA)	LA (in der Form english, deutsch, arabisch ...)
Link auf Volltext	UR	UR	UR und/oder L1 (Linkadresse wird in L1 in der Regel wiederholt, kann dann ignoriert werden, es gibt auch den Fall, dass nur L1 vorhanden ist, dann übernehmen.)
Link auf sonstige die Publikation beschreibende Materialien	UR	-	-
Quellen-URL	UR bzw. Kombination aus ID und feststehender URL-Adresse	-	-
DOI	DO	DO	(DO)
Autorenidentifikation	-	-	-
ISBN	SN, zur Unterscheidung von ISBN und ISSN wird ISBN bzw. ISSN der Nummer vorangestellt. Bsp.: SN - ISBN 9783658044671	keine Angaben zur ISBN, SN wird gegen die Intention des Standards, der ISBN/ISSN-Nutzung vorschreibt, anderweitig genutzt	SN = ISSN oder ISBN (ISSN ist immer 8-stellig, numerisch "18606229"; ISBN gibt es in unterschiedlichen Varianten: 13-stellig "ISBN 978-3-929173-92-5", "9783929173864", 10-stellig "3929173778")
ISSN	SN, zur Unterscheidung von ISBN und ISSN wird ISBN bzw. ISSN der Nummer vorangestellt. Bsp.: SN - ISSN 0958-305X	keine Angaben zur ISSN, SN wird anderweitig genutzt	SN, s.o.
Quellen-ID	ID	-	U2 (ein interner Zähler)
Rechtheinweis	AV	(AV)	(AV)
Medientyp	-	-	-
Inhaltstyp	Wenn TY - THES, dann Medientyp = Hochschulschrift	Wenn TY - THES, dann Medientyp = Hochschulschrift. Bei diesem Typ kommt auch M3 und N1 vor. N1 doppelt teilweise AB oder steht auch anstatt AB. M3 hat den Inhalt "Doktorarbeit" und kommt nur einmal vor. Im Standard stehen die Mn-Felder für Miscellaneous. AB, N1, N2 ist auch für Abstract vorgesehen.	-
Erstellungs-/Übernahmedatum	automatische Systeminformation	automatische Systeminformation	automatische Systeminformation

	FhG-ISI	ITA Wien	IZT Berlin
Datum der letzten Änderung	automatische Systeminformation	automatische Systeminformation	automatische Systeminformation
Ursprüngliche Quelle (Person/Institution)	automatische Systeminformation	automatische Systeminformation	automatische Systeminformation
Zuletzt geändert von	automatische Systeminformation	automatische Systeminformation	automatische Systeminformation
Sonstige Angaben	-	-	-

() Angaben in Klammern werden momentan nicht geliefert.

- kein entsprechendes BibTeX-Feld vorhanden.