Dirk Hommrich, Beate Pasucha, Matthias Razum und Ulrich Riehm

# Normdaten und Datenanreicherung beim Fachportal openTA

# Standards data and data accumulation in the specialist portal openTA

http://doi.org/10.1515/bd-2018-0031

**Zusammenfassung:** openTA ist ein webbasiertes Fachportal für das interdisziplinäre Forschungsfeld Technikfolgenabschätzung (TA). Der Beitrag geht zunächst auf die Vorgeschichte von openTA ein und stellt die wesentlichen Merkmale von openTA vor. Im Mittelpunkt steht die geplante Nutzung von Normdaten zur Anreicherung der Daten der openTA-Dienste und deren Verbreitung als Linked Open Data. Dabei sollen sowohl intellektuelle als auch (semi-)automatische Verfahren zum Einsatz kommen, um Entitäten wie Personen, Organisationen, Publikationen und Schlagworte eindeutig zu identifizieren.

Schlüsselwörter: Fachportal, Normdaten, Technikfolgenabschätzung

**Abstract:** openTA is a web-based specialist portal for the interdisciplinary research field technology assessment (TA). This article first describes the history of openTA and presents its main features. It focuses on the planned use of standards data for accumulating openTA service data and their circulation as Linked Open Data. Here, intellectual as well as (semi-) automatic methods are to be applied in order to clearly identify entities like persons, organisations, publications and keywords.

**Keywords:** specialist portal, standards data, technology assessment

Dirk Hommrich: dirk.hommrich@kit.edu

Beate Pasucha: beate.pasucha@fiz-karlsruhe.de
Matthias Razum: Matthias.razum@fiz-karlsruhe.de

Ulrich Riehm: ulrich@riehm.name

# 1 Einleitung

Das Fachportal Technikfolgenabschätzung, kurz openTA, ist ein webbasiertes Fachportal für das interdisziplinäre Forschungsfeld Technikfolgenabschätzung (TA), erreichbar unter https://www.openta.net/. Das Portal zielt darauf ab, Information, Kommunikation und Kooperation in der TA-Community zu unterstützen und zu verbessern und TA für die interessierte Öffentlichkeit leichter zugänglich zu machen. Es aggregiert Nachrichten, Termine und Publikationen aus dem Netzwerk Technikfolgenabschätzung (NTA), gibt einen Überblick über die Personen und Institutionen, die im Bereich der TA in Deutschland, Österreich und der Schweiz tätig sind, und ermöglicht die Erstellung personalisierter Dienste. openTA wird in Zusammenarbeit mit den Mitgliedsinstitutionen des Netzwerks Technikfolgenabschätzung (NTA) entwickelt und betrieben.<sup>1</sup>

Das Projekt wird durch die DFG gefördert. In einer ersten Förderphase von 2012 bis 2015 erarbeiteten das Institut für Technikfolgenabschätzung und Systemanalyse (ITAS), das Institut für angewandte Informatik (IAI) und die Bibliothek des Karlsruher Instituts für Technologie (KIT) die grundlegenden Funktionalitäten und Dienste des Portals. In einer zweiten Förderphase, die 2016 begann, führen ITAS und der neue Partner FIZ Karlsruhe die erfolgreiche Arbeit fort und bauen das Portal weiter aus. Ein wichtiger Aspekt bei der Weiterentwicklung ist der verstärkte Einsatz von Normdaten und die Anreicherung der angelieferten Daten.

In diesem Beitrag wird zunächst auf die Vorgeschichte von openTA eingegangen und die wesentlichen Merkmale von openTA vorgestellt. Im Mittelpunkt des Beitrags steht die Nutzung von Normdaten zur Anreicherung der Daten der openTA-Dienste.

<sup>1</sup> Der offizielle Titel des Vorhabens im Rahmen des Förderbereichs der DFG "Wissenschaftliche Literaturversorgungs- und Informationssysteme" ist "Kooperativer Aufbau eines Fachportals Technikfolgenabschätzung mit akteursspezifisch konfigurierbaren IuK-Diensten auf Basis dezentraler Informationsressourcen", die Förderkennzeichen GR 2897/7-1 und HE 5985/7-1. Siehe auch Riehm, U.; Nentwich, M. (2014): Was Sie schon immer über TA wissen wollten. Zwei neue Webportale, TAB-Brief Nr. 44, S. 49-50.

# 2 Zur Vorgeschichte von openTA

Die Vorgeschichte von openTA hat im Wesentlichen zwei Stränge: Die TA-Datenbank der 1990er Jahre und die Gründung und Entwicklung des Netzwerks TA ab 2004.

#### 2.1 TA-Datenbank

Der erste Strang reicht über 30 Jahre zurück. Die Vorläuferinstitution des seit 1995 bestehenden Instituts für Technikfolgenabschätzung und Systemanalyse (ITAS), die Abteilung für Angewandte Systemanalyse (AFAS), hatte zusammen mit FIZ Karlsruhe von 1986 bis ins Jahr 2000 eine "TA-Datenbank" erstellt und der Öffentlichkeit angeboten.² Per Post wurden Fragebogen an einige Hundert wissenschaftliche Institutionen, die mehr oder weniger mit TA zu tun hatten, verschickt und Daten über Institutionen, Projekte und Publikationen erfasst und – für damalige Verhältnisse durchaus innovativ – hypertextuell verlinkt sowie über das u. a. von FIZ Karlsruhe betriebene "Scientific & Technical Information Network" (STN) kostenpflichtig angeboten, später auch auf CD-ROM³ und gedruckt⁴ vertrieben. Im Jahr 2000 umfasste die TA-Datenbank Angaben zu 1.330 Institutionen, 3.376 Projekten und 6.964 Publikationen.

Mit dem Aufkommen des Internets war dieses Konzept nicht mehr aufrechtzuerhalten. Immerhin konnten die Daten von damals gerettet werden, so dass demnächst die fast 7.000 gut gepflegten und erschlossenen Publikationsdaten aus der Frühzeit der TA-Szene in Deutschland in den openTA-Publikationsdienst eingespeist werden können.

<sup>2</sup> Berg, Ingrid von; Bücker-Gärtner, Heinrich: Aufbau einer Datenbank über Institutionen, Projekte und Veröffentlichungen auf dem Gebiet der Technikfolgenabschätzung (TA). Karlsruhe: Kernforschungszentrum Karlsruhe 1988 (KfK-4448).

**<sup>3</sup>** Die CD-ROM kostete 1993 für "kommerzielle Nutzer" 1.500 DM, für wissenschaftliche Institutionen 1.000 DM und für Institutionen, die Informationen für die Datenbank geliefert hatten, 750 DM (TA-Datenbank-Nachrichten 2 (1993)4, S. 2–3).

**<sup>4</sup>** Coenen, Reinhard; Fürniß, Beate; Kupsch, Christel: Technikfolgenabschätzung im Spiegel ihrer Institutionen: Eine Dokumentation über deutsche Forschungseinrichtungen auf dem Gebiet der Technikfolgenabschätzung, Karlsruhe: Forschungszentrum Karlsruhe 2001.

#### 2.2 Netzwerk TA

Der zweite Strang beginnt mit der Gründung des Netzwerks TA (NTA) im Jahr 2004, ein lockerer Zusammenschluss von Personen und Institutionen auf dem Feld der Technikfolgenabschätzung in Deutschland, Österreich und der Schweiz. TA wird dabei nicht eng abgrenzend, sondern eher breit und offen verstanden, so dass sich dem Netzwerk auch wissenschaftliche Institutionen etwa aus der Nachhaltigkeits- und ökologischen Forschung oder auch der Zukunftsforschung angeschlossen haben. Das NTA hat sich insbesondere über die Durchführung von wissenschaftlichen Tagungen etabliert – im Jahr 2018 wird die 8. Tagung, die NTA8, in Karlsruhe stattfinden. Eine Arbeitsgruppe des NTA, die AG Information und Kommunikation, die sich gleich zu Beginn des NTA gebildet hatte, sorgte für den Aufbau einer kommunikativen und informativen digitalen Infrastruktur, betreute eine E-Mail-Liste<sup>5</sup> und eine Website.

Im April 2006 organisierte diese Arbeitsgruppe einen "Frühjahrsworkshop" unter dem Titel "Wissenschaftsvernetzung - Konzepte, Erfahrungen, Perspektiven", dessen Programm und Beiträge noch heute auf openTA zu finden sind6. In der Einladung hieß es u.a. "Insbesondere die über 'Portale' durchgeführte Vernetzung von Einzelwissenschaften und ihre dezentrale Organisation sind zu einer breiten Strömung geworden."

Der Workshop fand mit Unterstützung des Büros für Technikfolgenabschätzung beim Deutschen Bundestag (TAB) im Sitzungssaal des Bundestagsausschusses für Bildung, Forschung und Technikfolgenabschätzung im Paul-Löbe-Haus statt. Ein Ergebnis dieses Workshops war ein "9-Punkteplan" der AG IuK "zum Ausbau der informationstechnisch gestützten Vernetzung des Netzwerks TA", der im November 2006 auf der NTA2 in Berlin vorgelegt wurde. 7 Dieser sah u. a. vor:

- die Schaffung eines Netzwerks von "Internetbeauftragten" als Ansprechpartner bei den institutionellen Mitgliedern des NTA,
- der Betrieb einer NTA-Metasuchmaschine über alle Web-Angebote der institutionellen Mitglieder des NTA hinweg,
- ein NTA-Newsportal, u. a. auf Basis von RSS-Feeds,

<sup>5</sup> NTA-Mailingliste, https://www.openta.net/nta-mailingliste [Zugriff: 08.01.2018].

<sup>6</sup> Frühlingsworkshop der NTA AG IuK zum Thema "Wissenschaftsvernetzung – Konzepte, Erfahrungen, Perspektiven", https://www.openta.net/ag-iuk-fruhjahrsworkshop-2006 [Zugriff: 08.01.2018].

<sup>7</sup> Neun-Punkte-Plan der AG IuK des NTA zum Ausbau der informationstechnisch gestützten Vernetzung des Netzwerks TA, https://www.openta.net/documents/10179/36415/NTA-AG-IuK\_ Neun Punkte Plan [Zugriff: 08.01.2018].

- ein Open-Access-Report-Server im Sinne einer Vernetzung bestehender Repositorien und des Angebots eines eigenen,
- ein TA-Themenportal bei Wikipedia,
- eine TA-Linksammlung,
- eine (elektronische) TA-Review-Zeitschrift und
- die Erprobung von Web 2.0 Funktionalitäten.

Über einige dieser Ideen sind die Jahre hinweggegangen, wie etwa die Linksammlung, andere wurden weiter verfolgt und sind durch das Fachportal openTA mehr oder weniger umgesetzt (Newsdienst, Neuerscheinungsdienst, Publikationsdienst).

In diesem 9-Punkteplan wurde damals bereits festgehalten, dass die Umsetzung für einen Teil dieser Vorhaben durchaus als Netzwerkaktivität vorstellbar sei, andere aber eines größeren Ressourceneinsatzes bedürften. So wurde 2009 ein erster Antrag an die DFG gestellt, dem allerdings von den Gutachtern eine gründliche Überarbeitung anempfohlen wurde. Ein wichtiger Impuls für diese Überarbeitung erbrachte ein Workshop zum Thema "Innovative Konzepte für Wissenschaftsportale", der 2010 in Karlsruhe stattfand, und von ITAS mit weiteren Instituten des KIT organisiert wurde.<sup>8</sup> Auf diesem Workshop wurden u. a. die Folgerungen aus dem Scheitern des Großprojekts Vascoda<sup>9</sup> diskutiert, der zentralistische Ansatz von Fachportalen problematisiert ("Nicht der Wissenschaftler muss zum Fachportal gehen, sondern die Dienste des Fachportals müssen zum Wissenschaftler kommen") und die Konzepte des semantischen Netzes auf ihre Umsetzbarkeit ausgelotet.

Im gleichen Jahr wurde von der AG IuK des NTA zusammen mit den beteiligten KIT-Instituten eine Umfrage bei den Mitgliedsinstitutionen des NTA über deren Informationsdienste und Informationssysteme sowie die vorhandene technische Ausstattung und die unterstützten Standards durchgeführt.

Auch diese Ergebnisse flossen dann in den überarbeiteten Antrag (Februar 2011) ein, der schließlich im Januar 2012 (mit Kürzungsauflagen) bewilligt wurde. Das Projekt startete dann offiziell mit einem Auftaktworkshop am 31.10.2012 in Bern (im Kontext der NTA5).

<sup>8</sup> Innovative Konzepte für Wissenschaftsportale – im Hinblick auf ein Fachportal für die Technikfolgenabschätzung, https://www.openta.net/documents/10179/34933/Workshop+2010+TA-Portal/b692321e-79cb-4814-be47-0f3b06f01883 [Zugriff: 08.01.2018].

<sup>9</sup> Vascoda war der zentrale disziplinübergreifende, webbasierte Einstieg zu Fachangeboten wissenschaftlicher Bibliotheken und Informationseinrichtungen. Beginnend im Jahr 2001 wurde vascoda in mehreren Tranchen durch das BMBF bzw. die DFG gefördert, stellte dann allerdings 2011 seinen Betrieb wegen ausbleibender weiterer finanzieller Förderung ein.

Wenn openTA auch nicht im Rahmen des DFG-Förderschwerpunktes für die Fachinformationsdienste gefördert wird und keinen explizit bibliothekarischen, auf die Befriedigung des Spitzenbedarfs bei wissenschaftlichen Publikationen ausgerichteten Fokus aufweist, wurden doch eine der zentralen Anforderungen, die die DFG an Fachinformationsdienste stellt, von Anfang an erfüllt: Das Konzept wurde "aus der Wissenschaft für die Wissenschaft", in enger Kooperation mit der TA-Community und ihren Institutionen sowie auf Basis von Ist- und Bedarfserhebungen entwickelt.

Der Fortsetzungsantrag, nun in einer neuen Kooperation zwischen ITAS (federführend) und FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur, wurde im Juli 2016 (mit Kürzungsauflagen) bewilligt. Die zweite Projektphase begann offiziell mit dem 5. openTA-Workshop im November 2016<sup>10</sup> in Bonn im Kontext der Tagung NTA7<sup>11</sup>.

# 3 Was ist openTA?

openTA stellt Dienste zur Verfügung, die sich sowohl zentral über das Fachportal als auch dezentral über eine Integration in die Webangebote der jeweiligen TA-Institutionen (per Widget) nutzen lassen. Die openTA-Dienste umfassen den Newsdienst, den Kalenderdienst, den Publikations- mit Neuerscheinungsdienst und das Verzeichnis der persönlichen und institutionellen Mitglieder des NTA.

## 3.1 openTA-Newsdienst

Im Newsdienst werden die aktuellen Nachrichten der NTA-Mitgliedsorganisationen im Fachportal TA aggregiert und unter einer einheitlichen Oberfläche mit Such- und Filterfunktionen angeboten. Damit ist ein umfassender Nachrichtenüberblick über die Aktivitäten der TA-Institutionen in den D-A-CH-Ländern erstanden, der mittlerweile auch um Nachrichten des europäischen Netzwerks der parlamentarischen TA-Institutionen (EPTA) ergänzt wurde. Das Archiv der

**<sup>10</sup>** Fünfter Workshop 16. November 2016 in Bonn. Die Chance nutzen! Auftakt zur zweiten Förderphase für das Fachportal openTA, https://www.openta.net/workshops#openta-workshop5 [Zugriff: 08.01.2018].

<sup>11 &</sup>quot;Grand Challenges" meistern – 7. internationale Tagung des Netzwerks Technikfolgenabschätzung in Bonn, https://www.ea-aw.de/veranstaltungen/20-jahre-ea/7-nta-tagung.html [Zugriff: 08.01.2018].

Nachrichten umfasst über 4.300 Nachrichten zu Projekten, Publikationen, Veranstaltungen, Stellen- und Kooperationsangeboten sowie Nachrichten über die NTA-Institutionen (Stand: Januar 2018). 17 Institutionen liefern automatisiert Nachrichten über einen RSS- oder Atom-Feed an openTA. Eine redaktionelle Selektion oder Bearbeitung auf Seiten des openTA-Teams findet nicht statt. Die inhaltliche Verantwortung für die Inhalte auf openTA liegt beim Newsdienst wie bei den meisten anderen Diensten auch bei den Datengebern. Ein auf die eigenen Interessen individualisierbarer News-Feed, d. h. ein automatisierter Abruf der aktuellen Nachrichten, kann von den NutzerInnen, etwa im eigenen Mail-Programm, bezogen werden, so dass ein Besuch des Portals für die Nutzung des Newsdienstes dann nicht mehr nötig ist. 12

## 3.2 openTA-Kalenderdienst

Der openTA-Kalenderdienst bietet einen umfassenden Überblick über die öffentlichen Veranstaltungen des NTA und seiner Mitgliedsorganisationen sowie über Veranstaltungen mit TA-Relevanz. Er beruht im Wesentlichen auf den Kalenderdaten, die die NTA-Mitgliedsinstitutionen über ical-kompatible elektronische Kalender zur Verfügung stellen. Auch hierbei gibt es keine redaktionellen Eingriffe durch die openTA-Projektgruppe in Karlsruhe. Diese NTA-Termine werden ergänzt um "TA-relevante" Veranstaltungen außerhalb des NTA, die zusätzlich von Mitgliedern der AG IuK des NTA ausgewählt und erfasst werden.

## 3.3 openTA-Publikationsdienst

Der openTA-Publikationsdienst führt die Publikationen der kooperierenden NTA-Mitgliedsinstitutionen in einer Datenbank zusammen. Über den NTA-Fundus hinaus werden weitere Quellen etwa aus der DNB oder aus historischen Publikationsbeständen mit TA-Relevanz einbezogen, um eine vollständigere Abdeckung der TA-Literatur zu erreichen.

Ein wichtiger Bestandteil des Publikationsdienstes ist der monatlich erscheinende openTA-Neuerscheinungsdienst. Dieser gibt einen Überblick über TA-relevante Buchneuerscheinungen, die bei der Deutschen Nationalbibliothek (DNB) erfasst wurden. Im Neuerscheinungsdienst werden auch alle bei der DNB nach-

**<sup>12</sup>** Siehe auch Riehm, Ulrich; Hommrich, Dirk: Der openTA-Newsdienst. In: TATuP 26(2017)1–2, S. 89.

gewiesenen Publikationen von persönlichen Mitgliedern des NTA aufgenommen, von denen eine persönliche GND<sup>13</sup>-ID bekannt ist, auf deren Basis ein automatisierter Suchlauf monatlich erfolgt. Eine derzeit aus fünf Städten Deutschlands und Österreichs bestehende Redaktion mit sieben Personen wählt für jeden Monat etwa 20 bis 40 Titel aus einer automatisch selektierten Liste von etwa 200 bis 400 Titeln aus und veröffentlicht diese zusammen mit einem einführenden Editorial im openTA-Blog.

## 3.4 NTA-Mitgliederverzeichnis

Die öffentlich über openTA zugängliche Mitgliederdatenbank des Netzwerk TA (NTA) stellt für die TA-Community und die Öffentlichkeit eine zentrale Informationsressource über "TA-Experten" und "TA-Institutionen" und ihre Tätigkeitsfelder und Kompetenzen dar. Das NTA umfasst derzeit 50 institutionelle und mehr als 300 persönliche Mitglieder aus Deutschland, Österreich und der Schweiz.

Nähere Informationen zum Projekt finden sich unter https://www.openta.net/projekt-openta, zu den einzelnen Diensten unter https://www.openta.net/die-openta-dienste.

# 4 Vernetzung der openTA-Daten mit frei zugänglichen Normdaten

Die aktuell laufende, zweite Förderphase von openTA zielt unter anderem auf die Verbesserung der Metadatenqualität und auf die verstärkte Vernetzung des Systems mit anderen Quellen im Web, insbesondere Normdaten. Hierfür ist ein iteratives Vorgehen geplant. Im ersten Schritt steht die eindeutige Identifizierung zentraler Entitäten wie Personen, Organisationen, Publikationen und Schlagworten an. In einem zweiten Schritt können auf dieser Basis dann weitere Datenquellen erschlossen und die eigenen Daten angereichert werden.

Zunächst sind die Entitäten in den bei openTA vorliegenden Metadaten zu identifizieren, sie mit Normdatenquellen abzugleichen und dort vergebene Identifier in die eigenen Metadaten zu übernehmen. Am Beispiel einer Person könnte das aussehen wie in Abbildung 1 dargestellt.

<sup>13</sup> Gemeinsame Normdatei.

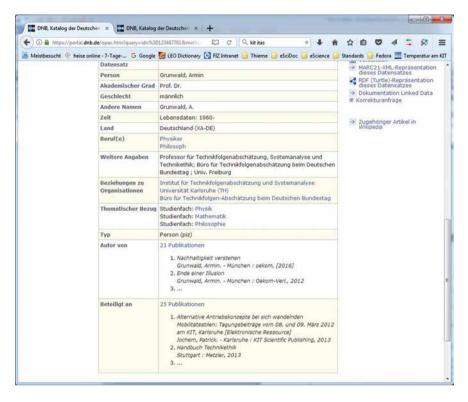


Abb. 1: Beispiel eines Personeneintrags in der Gemeinsamen Normdatei (GND).

Der Name einer Person, die etwa als Autor in einem der von openTA aggregierten Datensätze genannt wird, ist nicht immer eindeutig. So findet man im openTA-Publikationsdienst "Meier, Bernd" und "Meier, B." insgesamt 7-mal. Ob es sich dabei um eine, zwei oder mehr Personen handelt, muss anhand weiterer Merkmale erst ermittelt werden. Ideal wäre, wenn bereits die Daten bereitstellenden Institutionen ihre Daten mit eindeutigen Identifiern für Personen anreichern würden. Das geschieht bisher noch nicht, obwohl das openTA-Metadatenschema<sup>14</sup> dies vorsieht. Nur bei Daten, die von in der Regel großen bibliothekarischen Institutionen erstellt und gepflegt werden, wie von der DNB, sind solche Personen-IDs standardmäßig vorhanden und werden auch in den Datenbestand von openTA

<sup>14</sup> Riehm, Ulrich; Düpmeier, Clemens; Mönnich, Michael; Piltz, Len; Schmitt, Christian: open-TA-Publikationsdienst: Das open-TA-Metadatenschema und seine Umsetzung. (open-TA-Dokument Nr. 7, Version 2.0). Karlsruhe: 2015. https://www.openta.net/documents/10179/70804/open-TA Dok 7 Publikationsdienst-Metadatenschema/ [Zugriff: 09.01.2018].

übernommen. In den meisten Fällen werden aber intellektuelle oder (semi-)automatische Verfahren zum Einsatz kommen müssen, um zu einer Disambiguierung und eindeutigen Identifizierung insbesondere von Personen zu kommen.

# 5 Eindeutige Identifizierung und Disambiguierung

Um Entitäten eindeutig zu identifizieren, müssen zuerst geeignete Identifier festgelegt werden. Dabei muss man sich nicht notwendigerweise für einen einzelnen entscheiden, sondern kann pro Entität mehrere unterstützen. Tabelle 1 benennt für jede Entität relevante Identifier.

**Tab. 1:** Entitäten mit ihrem Vorkommen, geeigneten Identifiern und potenziellen Quellen für die Datenanreicherung.

Entität	Kommt vor	Identifier	Externe Quellen
Person	in allen openTA-Diensten	GND-Personen-ID	GND-Datensatz, Wikipedia,
	als Autor/Autorin;	VIAF-ID	Wikidata
	in der openTA Mitglieder-	ORCID	VIAF-Datensatz
	liste als persönliches		Öffentlich zugänglicher
	NTA-Mitglied		ORCID-Datensatz
Organisation	im openTA-Publikations-	GND-Organisations-ID	GND-Datensatz, Wikipedia,
	dienst als Affiliation;	ISNI	Wikidata
	im openTA-Kalenderdienst	VIAF-ID	ISNI-Datensatz
	als Veranstalter;		VIAF-Datensatz
	in der openTA Mitglieder-		
	liste als institutionelles		
	NTA-Mitglied		
Publikation	im openTA-Publikations-	DOI	CrossRef, Verlage
	und Neuerscheinungsdienst	ISBN	WorldCat, DNB, VlB, KVK,
		ISSN	ZDB, Verlage, ISSN-Portal
Schlagwort	im openTA-Publikations-	GND-Schlagwort-ID	GND-Datensatz
	und Neuerscheinungsdienst	DDC-Nummer	

#### 5.1 Personen

Die eindeutige Identifikation und Disambiguierung von Personen ist eine große Herausforderung. In Teilen kann dies durch intellektuelle Verfahren gewährleistet werden, etwa in dem die persönlichen Mitglieder aufgefordert werden, ihre eigenen Registrierungsdaten um entsprechende Identifier anzureichern. Der Mitgliederpersonendatensatz wird demnächst um entsprechende ID-Felder erweitert. Bei knapp über 300 persönlichen Mitgliedern erscheint dies, eventuell unterstützt durch redaktionelle Prozesse, als ein sinnvolles Vorgehen, wenn auch der Erfolg bisheriger Aktivitäten in diese Richtung begrenzt blieb. Bevorzugt als Identifier sind, sofern vorhanden, die GND-ID sowie die ORCID. Bereits seit längerer Zeit wird beim monatlichen Suchlauf bei der DNB für den openTA-Neuerscheinungsdienst nach allen bekannten GND-Personen-IDs der NTA-Mitglieder gesucht und die Treffer in den Neuerscheinungsdienst aufgenommen und so ein besonderer Nutzen oder "Mehrwert" für die Mitgliedschaft im NTA erzeugt.

Für die ca. 26.000 Publikationen in openTA-Publikationsdienst kann die Disambiguierung und Identifizierung von Autoren nur über automatische Verfahren erfolgen. Alleine auf Basis des Namens werden die Ergebnisse nur unbefriedigend sein. Entsprechend ist vorgesehen, zusätzliche Informationen wie etwa Co-Autoren, Zitationsnetzwerke, Ortsbezug, Affiliation und eine zeitliche Eingrenzung zu verwenden, um die Ergebnisse zu verbessern.

## 5.2 Organisationen

Organisationen tauchen in openTA einerseits als institutionelle Mitglieder, andererseits als Affiliation von Personen (z. B. Autoren) auf. Für die derzeit 50 institutionellen Mitglieder liegen bereits in den meisten Fällen im Mitgliederdatensatz die GND-Organisations-ID und die VIAF-ID vor.

Die Affiliationen von Autoren wird im Datenmodell des openTA-Publikationsdienstes dagegen aktuell überhaupt nicht berücksichtigt und liegt in den Ursprungsdaten nur in Ausnahmefällen vor, etwa bei den Daten, die openTA von der DNB übernimmt. Ein erster Schritt wäre die Anreicherung der vorhandenen Daten mit Organisationsangaben, was für einen Teil dieser Daten durch die Berücksichtigung der institutionellen Datenquellen sowie der Angaben zur

**<sup>15</sup>** Böhle, Knud: Wie Sie Ihre personenbezogenen Daten bei der Deutschen Nationalbibliothek überprüfen und anreichern können – und warum das wichtig ist. In: openTA-Blog 8.11.2016, https://www.openta.net/blog/-/blogs/wie-sie-ihre-personenbezogenen-daten-bei-der-deut schen-nationalbibliothek-uberprufen-und-anreichern-konnen-und-warum-das-wichtig-ist- [Zugriff: 09.01.2018].

**<sup>16</sup>** Böhle, Knud: Ein Schaufenster für neue Publikationen der NTA-Mitglieder! In: openTA-Blog 1.6.2016, https://www.openta.net/blog/-/blogs/ein-schaufenster-fur-neue-publikationen-dernta-mitglieder- [Zugriff: 09.01.2018].

Organisationszugehörigkeit in den Daten der persönlichen NTA-Mitglieder, die in der Regel ja auch Autorinnen oder Autoren sind, als relativ leicht realisierbar erscheint.

#### 5.3 Publikationen

Bei Publikationen stehen mit der ISSN bzw. ISBN in vielen Fällen bereits Identifier zur Verfügung. Generell wäre eine Verwendung eines Digital Object Identifiers (DOI) wünschenswert. Immer mehr Publikationen aus Periodika besitzen auch einen DOI. Während in der Datenbank des openTA-Publikationsdienstes für das Jahr 2010 nur bei 23 % der Zeitschriftenartikel Angaben zu einem DOI vorhanden sind, ist dieser Anteil für das Jahr 2016 bereits auf 46 % und für das Jahr 2017 auf 67 % angestiegen. Für die anderen Publikationen können DOIs über Suchen in CrossRef (Periodika) und DNB-Katalog bzw. WorldCat erfolgen, wobei sich schon gezeigt hat, dass die aggregierten Metadaten in openTA nicht durchgehend mit den Metadaten in CrossRef bzw. DNB-Katalog übereinstimmen und ein solcher Abgleich mit unscharfen Suchen erfolgen muss.

Für einige Publikationstypen, die eher grauer Literatur zuzuordnen sind (wie z. B. Vorträge), gibt es selten Identifier. Für diese Publikationen steht auch nicht zu erwarten, dass sie sich umfassend eindeutig und persistent identifizieren lassen.

## 5.4 Schlagworte

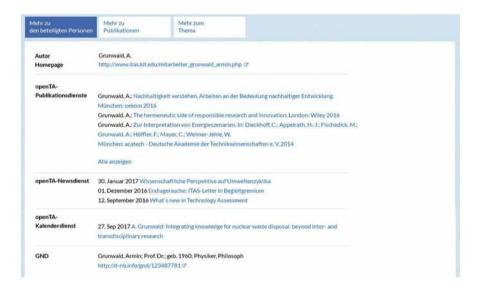
Schlagworte liegen aktuell nur für einen kleinen Ausschnitt der in openTA angebotenen Daten vor. Die vorliegenden Publikationsdaten mit Schlagworten stammen fast alle von der DNB und sind mit GND-Identifiern versehen. Die Herausforderung liegt hier weniger in der eindeutigen Identifizierung als vielmehr in der gezielten Anreicherung von Datensätzen um Schlagworte.

## 6 Datenanreicherung

Sobald die Entitäten eindeutig identifiziert sind, stehen vielfältige Möglichkeiten der Erschließung weiterer Datenquellen und der Datenanreicherung offen. Zum einen können Entitäten innerhalb einzelner Dienste und über Dienstgrenzen hinweg untereinander verlinkt werden, auch und gerade bei Daten aus unter-

schiedlichen Quellen. So könnte man z.B. den Angaben der persönlichen Mitglieder des NTA deren bei openTA verfügbaren Publikationsnachweise zuordnen und so für die Mitglieder einen besonderen Service und für die Nutzer von openTA einen informativen Mehrwert bieten.

Weitaus interessanter ist es aber, mit dem für die Entität vergebenen Identifier in externen Quellen wie der GND, ORCID oder Wikidata zu suchen, um ergänzende Informationen zu finden, die über die von den Partnern gelieferten Daten hinausgehen. Abbildung 2 zeigt am Beispiel des openTA-Publikationsdienstes, wie eine solche Anreicherung sowohl auf Basis einer internen Verlinkung als auch aus externen Quellen erfolgen könnte.



**Abb. 2:** Beispiel für eine mögliche Anreicherung von Informationen zu einem Autor aus interner Verlinkung und externen Quellen.

Die Anreicherung kann auf zwei unterschiedlichen Wegen erfolgen. Einerseits können die Daten aus externen Quellen in den Datenbestand von openTA statisch übernommen, zum anderen aus openTA heraus dynamisch auf externe Datenquellen verlinkt werden. Welche Variante jeweils gewählt wird, ist dann im Einzelfall zu entscheiden und hängt u.a. von der erwarteten Änderungsrate der jeweiligen Information und von Performanz- und Wartungsgesichtspunkten ab.

#### 6.1 Personen

Zu Personen, insbesondere den Mitgliedern des NTA, lassen sich Informationen über (weitere) Veröffentlichungen, biographische Informationen und Arbeitsschwerpunkte anreichern. Liegt eine GND-ID vor, können Daten aus dem GND-Record der Person entnommen werden. Darüber hinaus finden sich interessante ergänzende Daten z. B. in Wikidata. Sofern die Person über eine ORCID-ID verfügt und in dem dortigen Profil öffentlich einsehbare Daten hinterlegt hat, können auch diese zur Anreicherung herangezogen werden. Bei der Aggregation aus mehreren Quellen müssen allerdings Präzedenzregeln definiert werden, da gegebenenfalls widersprüchliche Informationen bei den unterschiedlichen Quellen vorliegen. Generell soll die jeweilige Quelle der angereicherten Daten immer mit angegeben werden. In der letzten Zeile der Abb. 2 wird dieses Prinzip deutlich: Einerseits werden biographische Daten (Geburtsjahr, Beruf) aus dem Personendatensatz der GND direkt angezeigt, andererseits kann der Nutzer über den Link auf den GND-Datensatz sich die gesamte dort verfügbare Information erschließen und die Quelle "überprüfen".

### 6.2 Organisationen

Über eindeutig identifizierte Organisationen lassen sich weitere Veröffentlichungen dieser Einrichtungen einsammeln, aber auch weitere Informationen wie Anschrift, Georeferenz oder Auftrag und Arbeitsschwerpunkte. Wichtige Quellen hierfür sind wiederum die GND, der VIAF- und der ISNI-Datensatz, eventuell ergänzt um Informationen aus Wikidata.

### 6.3 Publikationen

Bei Publikationen ist das Ziel, die von den NTA-Partnern gelieferten Metadaten zu vervollständigen und zu ergänzen, etwa um Abstracts, Inhaltsverzeichnisse und, sofern verfügbar, Links auf Volltexte. Wichtige Quellen hierfür sind der Katalog der DNB sowie Verlagsserver, die über CrossRef angesprochen werden können. Über die ISSN können teilweise ähnliche Informationen angezogen werden, wobei hier zusätzlich Titel, Autoren und Erscheinungsdatum zur Eingrenzung der Suchergebnisse herangezogen werden müssen. Für Monographien erlaubt die ISBN, über den Katalog der DNB, dem WorldCat von OCLC oder das Verzeichnis lieferbarer Bücher des Börsenvereins des deutschen Buchhandels vergleichbare Informationen zu beziehen – z. B. auch das Buchcover.

Darüber hinaus kann aber auch die Verfügbarkeit in Bibliotheken (KVK, ZDB, WorldCat) angezeigt werden. Weitere Ideen für Anreicherungen betreffen Referenzen, Zitationen, Rezensionen und von Dritten vergebene Schlagworte (etwa von der DNB).

### 6.4 Schlagworte

Neben der Anreicherung mit von Dritten vergebenen Schlagwörtern ist geplant, auch die (semi-) automatische Vergabe von Schlagworten zu erproben. Es existieren zwei erfolgversprechende Verfahren, aus Abstracts zu Publikationen Kandidaten für Schlagworte zu extrahieren. Häufig kommen Verfahren der maschinellen Verarbeitung natürlichen Sprache (Natural Language Processing, NLP) zum Einsatz, die gerade im Englischen gute Ergebnisse erzielen und über Tools wie KEA<sup>17</sup> einfach einsetzbar sind. Über OpenNLP<sup>18</sup> kann man auch deutsche Sprachmodelle verwenden. Daneben kann über die Berechnung von sog. "N-Gramms" eine Identifizierung von Kandidaten aus dem Korpus aller Publikationen in openTA erfolgen. Dabei wird der Text in Fragmente zerlegt, hier meist in Worte. Die Kombination von jeweils N Fragmenten wird dabei als N-Gramm bezeichnet. Die N-Gramme werden üblicherweise anschließend gerankt, um ungeeignete Terme auszuschließen. Im nächsten Schritt werden dann die verbliebenen Terme gegen einen Thesaurus abgeglichen, um die freien Terme auf ein kontrolliertes Vokabular abzubilden - idealerweise auf die GND, über die man dann eindeutig identifizierte Schlagworte erhält. Auch für dieses Verfahren kann z. B. KEA zum Einsatz kommen.

<sup>17</sup> KEA (Keyphrase Extraction Algorithm) ist ein Algorithmus zur Extraktion von Schlagwörtern aus Textdokumenten. Er unterstützt sowohl die freie Indexierung als auch die Indexierung anhand eines kontrollierten Vokabulars, http://community.nzdl.org/kea/ [Zugriff: 12.01.2018]; Witten, Ian H.; Paynter, Gordon W.; Frank, Eibe; Gutwin, Carl; Nevill-Manning, Crag G: Kea: Practical automatic keyphrase extraction. In: Proceedings of the 4th ACM Conference on Digital Libraries (1999), S. 254–255. DOI: 10.1145/313238.313437.

**<sup>18</sup>** OpenNLP ist ein auf Machine Learning aufsetzendes Open-Source-Werkzeug für die Verarbeitung von natürlichsprachlichem Text, https://opennlp.apache.org/ [Zugriff: 12.01.2018].

# Dissemination der Daten als Linked Open Data

Neben dem Anreichern der Metadaten aus externen Quellen will openTA zukünftig auch die eigenen Metadaten in Form von Linked Open Data<sup>19</sup> zur Nachnutzung durch Dritte bereitstellen. Zentrale Aufgabe dabei ist das Mapping der Daten, die die zentralen Entitäten beschreiben, auf geeignete LOD-Vokabulare. Tabelle 2 zeigt für die zu berücksichtigenden Entitäten die vorgesehenen Vokabulare, auf die die Informationen gemappt werden sollen.

Entität	aus openTA-Dienst	Vorliegendes Format	LOD-Vokabular	
RSS Feeds	News, Blog	RSS, ATOM; Blogposts	AtomOWL	
	http://bblfish.net/work/atom-owl/2006-06-06/AtomOwl.html			
Termine	Kalender	iCalendar	iCalendar als RDF	
	https://www.w3.org/TR/rdfcal/			
Bibliographische Angaben	Neuerscheinungen Publikationen	MARC21, BibTeX, RIS	BIBO, Dublin Terms	
	http://bibliontology.com/ http://dublincore.org/documents/dcmi-terms/			
Organisationen	Mitglieder Publikationen	intern	Core Organization Ontology	
	https://www.w3.org/TR/vocab-org/			
Personen	Mitglieder Publikationen Blog	intern	FOAF	
	http://www.foaf-project.org/			

Die Bereitstellung von LOD soll in Form von Daten-Dumps erfolgen. Das Angebot einer SPARQL-Schnittstelle zur dynamischen Abfrage wird aktuell noch untersucht, um Aufwand und Nutzen abzuwägen.<sup>20</sup> Darüber hinaus ist eine Dissemi-

<sup>19</sup> Berners-Lee, Tim: Linked Data. 18.06.2009, https://www.w3.org/DesignIssues/LinkedData. html [Zugriff: 12.01.2018].

<sup>20</sup> Warren, Robert: Why you need a SPARQL server. In: MUNINN Project Blog, 12.07.2013, https:// blog.muninn-project.org/node/68 [Zugriff: 15.01.2018]; Rog, Dave: The Enduring Myth of the

nation der Daten in den Webseiten von openTA geplant. Dazu sollen alle in der jeweiligen Webseite erwähnten Entitäten (also Personen, Organisationen, Publikationen und Schlagworte) mitsamt ihren Identifiern über die Vokabulare von schema.org<sup>21</sup> ausgezeichnet und so maschinell suchbar gemacht werden.

## 8 Ausblick

Das Fachportal openTA basiert auf zwei "Säulen", einer technisch-informatischen und einer organisatorisch-sozialen:

Zum einen kann aus der Aggregation vorhandener Informationen eine neue Qualität für die interessierte Community entstehen, nicht nur durch die umfassendere Datengrundlage, sondern vor allem durch die erweiterten Nutzungsfunktionalitäten, die bei den Datenbereitstellern oft nicht gegeben sind. Die Aggregation, die prinzipiell natürlich auch "händisch" über eine "Zentralredaktion" vorstellbar ist, wie dies bei einer Reihe anderer Fachportale und Fachinformationsdienste der Fall ist, erfolgt bei openTA dank der Verwendung weit verbreiteter Standards automatisiert. Anders lässt sich der Betrieb eines solchen Portals für eine verhältnismäßig kleine Community, die nicht nur interdisziplinär zusammengesetzt, sondern auch in ihren Organisationsstrukturen sehr heterogen ist, kaum auf Dauer sicherstellen.

Zum anderen müssen aus dem Netzwerk der auf dem Gebiet der TA arbeitenden Institutionen und Personen diejenigen zeitlichen, finanziellen, personellen und motivationalen Ressourcen bereitgestellt werden, die trotz eines hohen Automatisierungsgrades erforderlich sind, um den Betrieb des Portals auf Dauer zu gewährleisten. Dies betrifft in erster Linie die Überwachung der Dienste inkl. einer gewissen Qualitätskontrolle, die Akquise weiterer Quellen und die Unterstützung der Nutzung, um nicht zu sagen des "Marketings" für die Nutzung. Aus der Initiative der Personen im Netzwerk können sich darüber hinaus weitere Angebote entwickeln, wie bereits mit dem redaktionell betreuten und oben beschriebenen openTA-Neuerscheinungsdienst und dem openTA-Blog geschehen. Mit der Initiative von Personen im Netzwerk TA ist nicht zuletzt auch die zwischenmenschliche und wissenschaftsbetriebliche Herausforderung ange-

SPARQL Endpoint. In: Dave's Blog, 04.06.2013, https://daverog.wordpress.com/2013/06/04/the-enduring-myth-of-the-sparql-endpoint/ [Zugriff: 12.01.2018].

<sup>21</sup> http://schema.org/docs/schemas.html [Zugriff: 12.01.2018].

sprochen, das häufig unentgeltliche Engagement vieler Einzelpersonen mit Wertschätzung anzuerkennen.

Beide Seiten der "soziotechnischen" Plattform openTA werden weiterentwickelt. Die technischen Lösungen sollen weiter verbessert, leichter wartbar und nutzbar gemacht werden. Die personelle Trägerschaft für das Portal soll erweitert, Kooperationen ausgebaut und unter Mitwirkung von NTA-Mitgliedsorganisationen dezentralisiert werden. Die Förderung durch die DFG ermöglicht auf diesem keinesfalls einfachen Weg einen großen Schritt voranzukommen und damit openTA als integralen Bestandteil der "TA-Szene" und ihres professionellen Selbstverständnisses zu etablieren.



Dirk Hommrich
Karlsruher Institut für Technologie (KIT)
Institut für Technikfolgenabschätzung und Systemanalyse (ITAS)
Postfach 3640
76021 Karlsruhe
Deutschland
E-Mail: dirk.hommrich@kit.edu

Beate Pasucha

FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen Deutschland

E-Mail: beate.pasucha@fiz-karlsruhe.de

ORCID: orcid.org/0000-0003-0844-571X



Matthias Razum

FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen Deutschland

E-Mail: **Matthias.razum@fiz-karlsruhe.de** ORCID: orcid.org/0000-0002-5139-5511



Ulrich Riehm

Karlsruher Institut für Technologie (KIT)
Institut für Technikfolgenabschätzung und Systemanalyse (ITAS)
Postfach 3640
76021 Karlsruhe

Deutschland

E-Mail: ulrich@riehm.name

ORCID: orcid.org/0000-0002-5107-8305