



openTA-Newsdienst

Ziele, Verfahren, Nutzung

openTA-Dokument Nr. 2

Stand: 14.5.2014, Version 1.1

AutorInnen: Bettina Bauer, Ulrich Riehm (KIT-ITAS)

Dieses Dokument beschreibt die Ziele des openTA-Newsdienst sowie den Workflow von der Anmeldung bis zur Nutzung – ohne zu sehr in technische Details zu gehen. Es richtet sich in erster Linie an diejenigen, die gegebenenfalls zu entscheiden haben, ob die eigenen Institutsnachrichten, in den openTA-Newsdienst aufgenommen werden sollen, und an diejenigen, die sich allgemein über den openTA-Newsdienst informieren wollen. Der momentan verfügbare Prototyp des openTA-Newsdienst entspricht noch nicht den hier bereits skizzierten Ausbaustufen. Ein Handbuch für die technische Umsetzung ergänzt dieses Dokument (OpenTA-Dokument Nr. 3). Ein weiteres OpenTA-Dokument (Nr. 1) beschreibt die formale Klassifikation von Nachrichten und ihre Syntax im Rahmen von RSS- oder Atom-Feeds.

Kontakt: info@openta.net

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

Versionsgeschichte:

September 2013, erster Entwurf

Januar 2014 Version 1.0 (nur redaktionelle Änderungen)

Mai 2014 Version 1.1 (Linkaktualisierungen)

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	Das Wichtigste kurzgefasst.....	5
2	Ausgangslage und Zielsetzung	5
3	Anmeldung für openTA-Newsdienst, Teilnutzung	6
3.1	Wie erfolgt die Anmeldung?	6
3.2	Wenn nur ein Teil eines institutionellen Feeds für openTA zur Verfügung gestellt werden soll.....	7
3.3	Redaktionelle Betreuung, Auswahl.....	7
4	Elemente des openTA-Newsdienstes.....	7
4.1	Erstellung von RSS-/Atom-Feeds	7
4.2	Empfohlene Grundelemente der Nachricht	7
4.3	Empfohlene Anreicherung der News für erweiterte Nutzungsfunktionen.....	8
4.4	Sprachauszeichnung, Internationalisierung, Mehrsprachigkeit.....	9
4.5	Nutzungs- und Verwertungsrechte	10
4.6	Hierarchie der Merkmale	10
5	Übernahme und Speicherung in einer openTA-Newsdatenbank.....	11
6	Nutzung des openTA-Newsdienstes.....	11
6.1	Nutzungs- und Anzeigekanäle	11
6.2	Nutzergenerierte Anreicherungen.....	12
7	Open Source - Nachnutzung.....	12
8	Ansprechpartner.....	12

Tabellenverzeichnis

Tab. 1:	openTA-Kategorien für die formalen Nachrichtenkategorien	9
---------	--	---

1 Das Wichtigste kurzgefasst

Der openTA-Newsdienst führt die aktuellen Nachrichten der institutionellen Mitglieder des Netzwerks TA (NTA) zusammen und bietet einen Überblick über das Geschehen in der deutschsprachigen TA-Community. Das präferierte – wenn auch nicht einzige – Transportmedium für die Zusammenführung der verschiedenen Nachrichtenströme ist ein Atom- oder RSS-Feed. Die Absicht zur Teilnahme am openTA-Newsdienst sowie die URL des jeweiligen Feeds muss dem openTA-Team mitgeteilt werden.

Um eine selektive Nutzung des openTA-Newsdienstes zu gewährleisten, müssen bzw. sollen bestimmte Elemente innerhalb einer Nachricht enthalten sein. Dies sind u.a. der Titel der Nachricht, ein Kurztext, nach Möglichkeit der vollständige Nachrichtentext, der Link zur Nachricht, das Publikationsdatum, eine ID sowie der Urheber. Diese Elemente sind üblicherweise „automatisch“ in Feed-Nachrichten enthalten.

Es wird des Weiteren dringend empfohlen, ein von openTA entwickeltes Kategorienschema zur formalen Grobklassifikation der Nachrichten für jede Nachricht zu verwenden. Die wichtigsten Kategorien sind hier: Angebot, Institutionelles, Projekt, Publikation, Veranstaltung, Sonstiges. Das Kategorienschema kann auch für eigene Zwecke genutzt werden.

Der Prototyp des openTA-Newsdienst ist einerseits auf dem noch im Aufbau befindlichen openTA-Portal zu finden (<http://www.openta.net/news>). Darüber hinaus kann er nach den eigenen Wünschen konfiguriert und dann als Feed abonniert (<http://service.openta.net/feeds/atompub/news>) oder als Widget in eine eigene Webseite integriert werden (<http://www.openta.net/news#widget>).

2 Ausgangslage und Zielsetzung

Die meisten NTA-Institutionen informieren über ihre Forschungsaktivitäten, Publikationen, Veranstaltungen, gegebenenfalls auch über Ausschreibungen, Stellenangebote und Institutionelles. Diese Nachrichten werden typischerweise im Web-Angebot der jeweiligen Institution veröffentlicht oder über einen elektronischen Newsletter (als E-Mail-Abo oder als PDF) verbreitet. Aktuelle Meldungen erscheinen auch in elektronischen und gedruckten (Haus-)Zeitschriften oder als Pressemitteilungen. Als weiterer Distributionskanal für elektronische Nachrichten wird immer öfter ein RSS/Atom-Feed bereitgestellt, der von Interessierten abonniert werden kann.

Ziel des openTA-Newsdienst ist es, einen zentralen Anlaufpunkt für Nachrichten aus der deutschsprachigen TA-Community aufzubauen und dadurch die Sichtbarkeit von TA und der einzelnen Institute in der interessierten Öffentlichkeit zu erhöhen.

Der Zugang zum openTA-Newsdienst soll vielfältig sein (Web, Feed, E-Mail, Web-Widget, mobil) und eine über den Nachrichtenabruf erweiterte Funktionalität sowie Selektivität aufweisen.

Die erste Ausbaustufe des openTA-Newsdienstes umfasst die Übernahme von Atom- oder RSS-Feeds und die Ausgabe im openTA-Portal, als abonnierbarer Atom-Feed und als portierbares Web-Widget. Auch die Funktionalität und Selektierbarkeit ist in der ersten Ausbaustufe begrenzt.

Die folgende Beschreibung bezieht sich in erster Linie auf diese erste Ausbaustufe und gibt gegebenenfalls Hinweise auf geplante Erweiterungen. Die Darstellung orientiert sich an einem Workflow von der Anmeldung eines Feeds bei openTA bis zur Nutzung des openTA-Newsdienstes.

3 Anmeldung für openTA-Newsdienst, Teilnutzung

3.1 Wie erfolgt die Anmeldung?

Der openTA-Newsdienst umfasst die Nachrichten der institutionellen Mitglieder des NTA, soweit sie dies wünschen. Interessierte NTA-Mitgliedsinstitutionen wenden sich direkt über info@openTA.net an das openTA-Team.

Für eine spätere Ausbaustufe ist vorgesehen, dass die Anmeldung und die Pflege dieser Daten über das openTA-Mitgliederportal erfolgen kann.

Folgende Informationen werden hierfür benötigt:

- URLs für die einzelnen Nachrichtenquellen als RSS- oder Atom-Feed. Es können beliebig viele Feeds mit den entsprechenden Angaben angemeldet werden!

Zu jedem angemeldeten Feed können weitere Angaben gemacht werden:

- Angabe einer Kontakt-E-Mail-Adresse für Rückfragen zu den Newsangeboten. Wenn hier keine Angabe gemacht wird, dann gilt der im openTA-Mitgliederportal vermerkte Internetbeauftragte als Kontaktperson. Die E-Mail-Adresse wird in den Autoreneintrag (Institution) aufgenommen.
- Formale Kategorisierungen der Nachrichten (vgl. OpenTA-Newsdienst: Formale Kategorisierung, openTA-Dokument Nr. 1). Diese Angabe wird dringend empfohlen, wenn die Nachrichten eines Feeds durchgängig nur einer Kategorie zuzuordnen sind und diese Kategorisierung nicht bei den Einzelnachrichten selbst steht.
- Angaben zur Sprache der zur Verfügung gestellten Nachrichten (vgl. 4.4). Der Default-Wert ist „de“ (Deutsch).
- Angaben zur Nutzungslizenz (vgl. 4.5).
- Angaben dazu, ob der gesamte Nachrichtenstrom oder nur Teile für openTA übernommen werden sollen (vgl. 3.2).
- Angaben zum gewünschten Aktualisierungszyklus durch den openTA-News-Crawler. Der Default-Wert ist „täglich“.

In weiteren Ausbaustufen könnten zusätzliche Angaben ergänzt werden.

- Aus der Mitgliederverwaltung für die institutionellen Mitglieder des NTA werden für den Newsdienst die folgenden Informationen verwendet:
 - Adressinformationen, aus denen automatisch die Geoinformation für eventuelle Kartendarstellungen generiert werden.
 - Institutionen-Icon (bzw. stabiler Link auf das Institutionen-Icon), für die Kennzeichnung der Herkunft der jeweiligen Nachricht.
 - URL auf die das Icon verweisen soll. Dies wird in der Regel die Startseite des jeweiligen institutionellen Web-Angebots sein.

3.2 Wenn nur ein Teil eines institutionellen Feeds für openTA zur Verfügung gestellt werden soll

Einige institutionelle NTA-Mitglieder sind nur teilweise mit TA-Themen befasst. Sie wollen gegebenenfalls deshalb nur eine Auswahl aus ihren News für openTA zur Verfügung stellen. Dies sollte bei der Anmeldung des Feeds mitgeteilt werden (vgl. 3.1).

Die für den openTA-Newsdienst zu selektierenden Nachrichten müssen dann mindestens mit

```
<category term="openTA"/> (Atom) oder  
<category>openTA</category> (RSS)
```

gekennzeichnet sein (vgl. OpenTA-Newsdienst: Formale Kategorisierung, openTA-Dokument 1).

3.3 Redaktionelle Betreuung, Auswahl

Das openTA-Team behält sich vor, Nachrichten, die nicht in den thematischen Fokus „TA“ passen oder rein kommerzielle Interessen verfolgen, auszuschließen.

Eine weitere redaktionelle Bearbeitung der Nachrichten wird nicht vorgenommen. Zum Zwecke von Vorschauansichten kann jedoch der Inhalt automatisiert gekürzt (Beispiel: nur die ersten 100 Zeichen werden angezeigt) und die vollständige Version dann verlinkt werden.

Dubletten von Nachrichten werden nach Möglichkeit ausgeschlossen. Dazu wird die Identität der Nachrichten-ID (ID bzw. GUID), die Sprachangabe, das Publikationsdatum und die URL des anliefernden Feeds überprüft.

Ferner wird eine Nachricht aus derselben Quelle, deren Titel und Inhalt absolut identisch ist, als Dublette erkannt und nicht neu verarbeitet, selbst wenn ID oder Datumsangaben abweichen. Werden Nachrichten mit der gleichen ID, Sprachangabe und URL des Feeds aktualisiert und das Publikationsdatum der Nachricht unterscheidet sich von der vorherigen Version, wird die alte Version überspeichert.

4 Elemente des openTA-Newsdienstes

(Vgl. hierzu auch ausführlicher: Technisches Handbuch openTA-Newsdienst, openTA-Dokument Nr. 3)

4.1 Erstellung von RSS-/Atom-Feeds

Die meisten Content-Management-Systeme (CMS) erlauben Inhalte auch in Form von Feeds anzubieten. Diese Feeds können dann von openTA eingebunden werden. Die beiden Feed-Formate Atom und RSS 2.0 (andere RSS-Versionen nur eingeschränkt) werden von openTA voll unterstützt, wobei Atom zu bevorzugen ist, da es das modernere und besser standardisierte Format ist.

4.2 Empfohlene Grundelemente der Nachricht

Folgende Grundelemente sind üblicherweise in einer Nachricht enthalten:

- ID der Nachricht <id>, <guid>
- Titel <title>
- Teaser, Zusammenfassung <summary>, <description>
- Nachrichtentext <content>
- Link zur Nachricht <link>
- Publikationsdatum <published>, <pubDate>
- Aktualisierungsdatum <updated>
- Autor oder Urheber <author>

Als **Pflichtelemente** einer Nachricht werden Angaben zur **ID**, zum **Nachrichtentitel** und zum **Publikationsdatum** erwartet. Die herausgebende Institution des Feeds wird aus den Angaben, die bei der Anmeldung für den openTA-Newsdienst angegeben wurden, und aus den im Mitgliederportal hinterlegten Angaben ergänzt. Weitere im Ausgangsfeed aufgeführte Autoren werden als Mehrfacheintragung weitergegeben.

Es wird dringend empfohlen, sowohl einen Kurztext (Teaser, Zusammenfassung) als auch den gesamten Nachrichtentext (Volltext) zum Bestandteil des Feeds zu machen. Nur so können die Möglichkeiten der Volltextsuche im openTA-Nachrichtenarchiv vollständig ausgenutzt werden.

Wenn der öffentliche Feed der jeweiligen NTA-Institution aus bestimmten Gründen den Volltext nicht enthalten soll, wäre ein paralleler Feed mit dem Volltext, dessen Adresse nicht öffentlich gemacht wird, sondern nur für openTA verwendet wird, eine Lösungsmöglichkeit.

4.3 Empfohlene Anreicherung der News für erweiterte Nutzungsfunktionen

Zur besseren Suchbarkeit und um eine selektive Auswahl der aggregierten Nachrichten zu erlauben, sollen den Nachrichteneinträgen vom Nachrichtenproduzenten, also den NTA-Institution, weitere Merkmale angefügt werden. Diese von openTA ausgewerteten Kategorisierungen und Metainformationen stehen nicht in Konkurrenz zu schon vorhandenen Kategorisierungen der Nachrichtenerheber, sondern ergänzen diese.

Für eine Einteilung der Nachrichten in verschiedene formale Kategorien wird von openTA ein Auszeichnungsschema empfohlen (vgl. OpenTA-Newsdienst: Formale Kategorisierung, openTA-Dokument 1 und Tabelle 1). Diese Kategorien können sowohl in Deutsch als auch in Englisch verwendet werden (vgl. 4.4):

Gegebenenfalls können pro Nachricht auch mehrere Kategorien vergeben werden, wenn dies vom Inhalt her geboten erscheint.

Tab. 1: openTA-Kategorien für die formalen Nachrichtenkategorien

Deutsch	Englisch
Angebot	offer
Angebot/Stellenangebot	offer/job offer
Angebot/Kooperationsangebot	offer/cooperation offer
Angebot/Sonstiges	offer/other
Institutionelles	institutional matters
Institutionelles/Personalien	institutional matters/personnel matters
Institutionelles/Sonstiges	institutional matters/other
Projekt	project
Publikation	publication
Veranstaltung	event
Veranstaltung/Call for Papers	event/call for papers
Veranstaltung/Ankündigung	event/announcement
Veranstaltung/Bericht	event/report
Sonstige	other

Weitere ergänzende Informationen:

- Sprachangaben (vgl. 4.4)
- Lizenzinformationen (vgl. 4.5)
- Bilder oder andere nichttextliche Medientypen. Insbesondere Bilder werden als Ergänzung zum Text empfohlen. Die Nutzungsrechte müssen allerdings gewährleistet sein (vgl. 4.5).
- räumlicher Bezug der Nachricht

In weiteren Ausbaustufen könnten eventuell die folgenden Kategorien und Metainformationen ergänzt werden.

- *inhaltliche Klassifizierung*
- *Kategorisierung der Nachricht nach Adressatengruppen*
- *intendierte geographische Reichweite der Nachricht*

Weitere Kategorisierungen und Metainformationen der Urheber der Nachrichten im Feld <category> oder anderen Elementen der Nachricht werden mit übernommen, abgespeichert und ausgegeben und werden gegebenenfalls in weiteren Ausbaustufen für die Volltextsuche im openTA-Newsarchiv mit ausgewertet.

4.4 Sprachauszeichnung, Internationalisierung, Mehrsprachigkeit

OpenTA ist ein Fachportalprojekt, das sich auf die deutschsprachige Community der TA-Forscher und TA-Interessierten in Deutschland, Österreich und der Schweiz bezieht. Trotzdem stellt sich die Frage, wie mit nichtdeutschsprachigen Informationen umzugehen ist. Häufig werden News-Inhalte der NTA-Mitglieder auch in Englisch oder anderen Sprachen als Deutsch zur Verfügung gestellt. Die entsprechenden Feeds sollten dann mit einer Sprachkennung versehen werden und gemäß dieser Sprachkennung können Nutzer auch ihre Sprachpräferenz selektieren.

Die Festlegung der Sprache geschieht bei RSS nur auf Channel-Ebene im Kopfteil des Feeds mit dem Element <channel>. Enthält ein RSS-Feed Nachrichten in unterschiedlichen Sprachen, dann kann das Merkmal der Sprache in den Nachrichten nicht erkannt werden. Es wird dann empfohlen entweder sprachgetrennte RSS-Feeds anzubieten oder Atom als Format für den Feed zu verwenden.

Bei einem Atom-Feed wird auf Ebene der Nachricht als Attribut des Elements <entry> die Sprache angegeben.

Für die Sprachangabe wird ISO 639-1¹ verwendet. Für Deutsch muss also „de“ und für Englisch „en“ als Sprachangabe eingesetzt werden.

Die Sprache der Nachrichten kann aber auch bereits bei der Anmeldung für den openTA-Newsdienst festgelegt werden. Angaben auf Channel- oder Nachrichtenebene überschreiben diesen Default-Wert gegebenenfalls.

Die formalen Kategorien können in einer deutsch- oder englischsprachigen Variante genutzt werden. Die Nutzung dieser Sprachvarianten hat keinen Einfluss auf die erwartete Sprachversion der Nachricht, die allein über das Language-Tag bzw. die Angabe bei der Anmeldung des Feeds gesteuert wird.

Bei der Ausgabe der News werden durch openTA die Kategorien passend zur Sprache des jeweiligen Artikels ausgegeben, unabhängig davon, in welcher Sprachversion die Kategorien selbst übermittelt wurden.

Die Abbildung (Verlinkung) einzelner Nachrichten einer Sprache auf die entsprechende übersetzte anderssprachige Variante ist in openTA nicht vorgesehen.

4.5 Nutzungs- und Verwertungsrechte

Die Nutzungs- und Verwertungsrechte müssen über die ganze Wertschöpfungskette von der erstmaligen Bereitstellung durch die NTA-Institution, über die Speicherung und Bereitstellung durch openTA und die Ausgabe der Daten zur Anzeige, Weitergabe und Weiterverarbeitung beim Nutzer transparent und maschinenlesbar angegeben werden.

Bilder und andere nichttextliche Medien werden nicht in der openTA Newsdatenbank abgespeichert, sondern nur über einen Link auf die Ursprungsadresse verwendet.

Um diese aber im openTA-Newsdienst verwenden zu können, müssen die Mediennutzungsrechte (für Bilder etc.) den Nutzungsrechten der Textnachricht entsprechen.

Nähere Details zur korrekten Angabe von Creative-Commons-Lizenzen sind im Technischen Handbuch zum openTA-Newsdienst (openTA-Dokument 3) in Abschnitt 2.3 beschrieben.

4.6 Hierarchie der Merkmale

Es gilt, dass Angaben auf Item-Ebene, Angaben auf der Ebene des Channels und Angaben auf der Channel-Ebene, Angaben im Mitgliederportal überschreiben. Daneben sind aber auch Mehrfacheintragungen möglich. Dies gilt für Angaben in den Feldern <author> bzw. <contributor> und für die Eintragungen im Feld <category>. Für diese Angaben wird für jeden hinzukommenden Eintrag das Feld wiederholt.

¹ ISO 639-1 (Codes für die Darstellung der Namen von Sprachen - Teil 1 : alpha-2 Code). http://www.loc.gov/standards/iso639-2/php/code_list.php

5 Übernahme und Speicherung in einer openTA-Newsdatenbank

In der ersten Ausbaustufe werden die News-Feeds der NTA-Mitglieder über einen RSS/Atom-Nachrichtenstrom übernommen und in eine Datenbank gespeichert. Dieser Distributionsweg ist das präferierte aber nicht unbedingt einzige Modell für die Nachrichtenaggregation in openTA.

In weiteren Ausbaustufen sind weitere Übernahmemöglichkeiten vorstellbar. So etwa die Eingabe einzelner Nachrichten über ein von openTA angebotenes Webformular oder über Mikroformate auf Webseiten.

Alle News werden komplett – bei ergänzenden Medien (Bilder, Videos etc.) allerdings nur Links auf die Originaldateien – in der openTA-Newsdatenbank gespeichert und auf Dauer archiviert. So können auch ältere Nachrichten über ein Suchinterface im Portal durchsucht werden.

Aktualisierte Nachrichten (gleiche ID, gleiche Sprache, gleiche Feed-URL aber abweichendes Publikations- bzw. Aktualisierungsdatum) werden überspeichert.

6 Nutzung des openTA-Newsdienstes

6.1 Nutzungs- und Anzeigekanäle

Die verschiedenen Nutzungsmöglichkeiten des openTA-Newsdienstes sind:

- Nutzung des aktuellen News-Streams im openTA-Portal (<http://www.openta.net/news>). Derzeit umfasst der openTA-Newsdienst alle Nachrichten der letzten drei Monate. Diese können nach den folgenden Kategorien selektiert und gefiltert werden:
 - Sprache,
 - Urheberinstitution,
 - formale Kategorie,
 - einem freien Suchstring.

In weiteren Ausbaustufen könnte auch eine Sortierung und Selektion nach den folgenden Kriterien erfolgen:

- *Publikationsdatum,*
- *dem Ort und Land der Urheberinstitution*
- *inhaltlichen Kategorien oder*
- *Adressaten der Nachricht.*

Die institutionelle Herkunft der Nachricht ist bei von openTA erzeugten Darstellungen der Nachrichten bereits in der Vorschau erkennbar. Bei der Vollansicht der Nachricht wird immer ein Link zur Ursprungsnachricht ergänzend angegeben.

- Suche im Portal-Archiv, das alle bisher angelieferten Nachrichten umfasst. Die Funktionalität entspricht der des aktuellen Newsdienstes.
- Abonnement eines Atom-Feeds für den openTA-Newsdienst, der nach den oben beschriebenen Kategorien und Selektionsmöglichkeiten konfiguriert werden kann (<http://service.openta.net/feeds/atompub/news>).

- Nutzung des openTA-Newsdienstes als Widget, das in eine private oder institutionelle Website integriert werden kann.
Das Widget lässt sich auf dem openTA-Portal konfigurieren und dann portieren (<http://www.openta.net/news-feed-einbinden>).
- Für die mobile Nutzung des openTA-Newsdienstes werden vorhandene FeedReader für Smartphones oder Tablets empfohlen.

In einer weiteren Ausbaustufe könnte es auch

- *eine Abonnementmöglichkeit des konfigurierten Feeds per E-Mail und*
- *im Rahmen von my-openTA einen vorkonfigurierten openTA-Newsdienst unter my-openTA geben.*

6.2 Nutzergenerierte Anreicherungen

Nutzergenerierte Anreicherungen – Tagging, Kommentare, Bewertungen, Nutzungszählungen etc. – beim openTA-Newsdienst werfen im Gegensatz zu sonstigen Elementen des openTA-Portals (Blogs, Wikieinträge, Webseiten) besondere technische, aber auch datenschutzrechtliche Probleme auf. Mögliche Realisierungen müssen im Weiteren geprüft werden.

7 Open Source - Nachnutzung

Der openTA-Newsdienst basiert auf der JEE6 (Java Plattform, Enterprise Edition Version 6) und kann mit jedem JEE6-kompatiblen Applikationsserver genutzt werden. Eine Weitergabe und Nachnutzung ist prinzipiell möglich. Support für die Nachnutzung kann jedoch nicht geleistet werden.

8 Ansprechpartner

Für die Dauer des Projektes „Fachportal TA/openTA“ ist Ulrich Riehm (ITAS) für inhaltliche und organisatorische und Christian Schmitt für technische Belange Ansprechpartner. Diese sind über die Mailadresse info@openta.net erreichbar.