



openTA News Service Formal Categorization

openTA Document No. 1

Last Update: Apr 22, 2014, Version 1.0

English translation of: „Formale Kategorisierung openTA Newsdienst“. openTA Dok. Nr. 1
English Translation: Maike Abel (KIT Library)

Author: Christian Schmitt (KIT-IAI)

Contact: info@openta.net

The openTA documents can be retrieved from
<http://www.openta.net/veroeffentlichungen-und-openta-dokumente>

Version history:

March 2013 first draft

January 2014 Version 1.0 (editorial changes only)

April 2014 English translation

Licence Notice: CC BY-SA 4.0

Attribution-ShareAlike 4.0 International

Summary and target audience: This documentation describes how entries in the news feeds of NTA institutions can be categorized. The information is targeted at the web administrators of the institutional members of the NTA. The openTA News Service is also described in three further documents:

- openTA News Service: Ziele, Verfahren, Nutzung (Objectives, Procedure, Use) (openTA Dokument No. 2) (only in German)

- openTA News Service: Technisches Handbuch (Technical Handbook) (openTA Dokument No. 3) (only in German)
- openTA News Service: FAQs. (openTA-Dokument No. 8) (only in German)

Part of the DFG research project “Kooperativer Aufbau eines Fachportals Technikfolgenabschätzung“, funded by the DFG, conducted by the KIT institutes ITAS, IAI and the KIT Library.

Table of Contents

1	Preface.....	5
2	openTA Formal Categories.....	5
3	Use in Web Feeds.....	5
3.1	Use in Atom feeds.....	6
	Examples.....	7
3.2	Use in RSS feeds.....	8
	Examples.....	8
	Appendix A: APP-Category Document for the openTA Formal Categories.....	11
	German.....	11
	English.....	11

List of Tables

Table 1: openTA Formal News Categories (de/en).....	5
Table 2: openTA Formal Categories – Hierarchical and Flat Structure (de/en)	6
Table 3: openTA Formal Categories (de/en) with Term Attribute.....	7

1 Preface

As part of the DFG project openTA a service is developed by means of which news items from the TA-community are collected, processed and made available in a unified way. In the following it is described how entries from the news feeds of NTA institutions can be categorized in order to bundle them by specific topics and across several institutions in the openTA News Service.

2 openTA Formal Categories

openTA has defined the following category hierarchies for the classification of news (and other content):

Table 1: openTA Formal News Categories (de/en)

German	English
1. Angebot	1. offer
a) Stellenangebot	a) job offer
b) Kooperationsangebot	b) cooperation offer
c) Sonstiges	c) other
2. Institutionelles	2. institutional matters
a) Personalia	a) personel matters
b) Sonstiges	b) other
3. Projekt	3. project
4. Publikation	4. publication
5. Veranstaltung	5. event
a) Call for Papers	a) call for papers
b) Ankündigung	b) announcement
c) Bericht	c) report
6. Sonstige	6. other

3 Use in Web Feeds

There are two relevant standards for the dissemination of news items in the form of web feeds, the family of RSS formats and the Atom format. Both standards allow the specification of categories in order to classify each entry semantically.

It is however not possible to structure these categories hierarchically. The hierarchy therefore needs to be mapped into a flat structure. For openTA this was achieved by placing the parent category first and separating the following hierarchy levels by a slash (/). This results in the following mapping of openTA categories to the according RSS / Atom format category declarations as they should be specified in feeds consumed by the openTA feed aggregation system:

Table 2: openTA Formal Categories – Hierarchical and Flat Structure (de/en)

	Flat Structure (de)	Flat Structure (en)
1. Angebot	Angebot	offer
a) Stellenangebot	Angebot/Stellenangebot	offer/job offer
b) Kooperationsangebot	Angebot/Kooperationsangebot	offer/cooperation offer
c) Sonstiges	Angebot/Sonstiges	offer/other
2. Institutionelles	Institutionelles	institutional matters
a) Personalia	Institutionelles/Personalia	institutional matters/personnel matters
b) Sonstiges	Institutionelles/Sonstiges	institutional matters/other
3. Projekt	Projekt	project
4. Publikation	Publikation	publication
5. Veranstaltung	Veranstaltung	event
a) Call for Papers	Veranstaltung/Call for Papers	event/call for papers
b) Ankündigung	Veranstaltung/Ankündigung	event/announcement
c) Bericht	Veranstaltung/Bericht	event/report
6. Sonstige	Sonstige	other

If a content management system (CMS) is employed for creating feeds and the chosen feed component of the CMS allows the use of categories, the generated category description in the feed entry should be one of those entries listed in the middle (if the language is German) or right column (if the language is English) of the above table (flat structure, de or en). These category identifiers can often be filled in by the user itself, or an administrator has to prefill the category selection with one of the above lists. In the latter case, the system administrator in charge may be able to help.

If the used feed component allows the administrator to additionally specify a classification scheme for the configured category selection list, the correct reference is:

<http://service.openta.net/feeds/atompub/scheme/formal-category>

The openTA categories should be clearly separated from any other categories which are configured in the news feed category configuration. If it is not possible to associate a scheme reference with categories the openTA feed aggregation allows it to add the prefix “openTA/” to each category to clearly distinguish openTA categories from any other categories configured for the news feed.

3.1 Use in Atom feeds

Two standards are relevant concerning Atom feeds. For the description of feed content the Atom Syndication Format (ASF) is used. For the standardized dissemination and processing of ASF-based contents the Atom Publishing Protocol (AtomPub/APP) is used.

An unlimited number of categories can be included for each entry in an ASF feed according to the following format (Example for the English version¹):

```
<category term="angebot-misc"
scheme="http://service.openta.net/feeds/atompub/scheme/formal-
category?lang=en"
label="offer/other"/>
```

¹ The German equivalent:

```
<category term="angebot-misc"
scheme="http://service.openta.net/feeds/atompub/scheme/formal-category"
label="Angebot/Sonstiges" />
```

The value of the *term* attribute defines the machine-readable category key, the optional value of the *scheme* attribute defines the reference on a category document which describes all possible categories. The optional *label* attribute indicates the human-readable name of the respective category. The values for the term attribute are as follows:

Table 3: openTA Formal Categories (de/en) with Term Attribute

German	English	Term
1. Angebot	1. offer	angebot
a) Stellenangebot	a) job offer	angebot-stelle
b) Kooperationsangebot	b) cooperation offer	angebot-koop
c) Sonstiges	c) other	angebot-misc
2. Institutionelles	2. institutional matters	inst
a) Personalia	a) personel matters	inst-pers
b) Sonstiges	b) other	inst-misc
3. Projekt	3. project	projekt
4. Publikation	4. publication	pub
5. Veranstaltung	5. event	veranst
a) Call for Papers	a) call for papers	veranst-call
b) Ankündigung	b) announcement	veranst-ankuend
c) Bericht	c) report	veranst-bericht
6. Sonstige	6. other	misc

The document format is standardized according to the APP (see Appendix A). The value of the *scheme* attribute should be the URL at which the document can be retrieved. For the formal categories of openTA this is:

de: <http://service.openta.net/feeds/atompub/scheme/formal-category>

en: <http://service.openta.net/feeds/atompub/scheme/formal-category?lang=en>

Examples

The examples below each contain two category entries, one for the openTA category „job offer“, which is a sub-category of „offer“ (1.1), and another for the category “project” (3)

Example 1: Recommended notation

```
<entry xmlns=http://www.w3.org/2005/Atom xml:lang="en">
<title>Example</title>
<id>Example ID</id>
<updated>2013-02-25T12:30:02Z</updated>
<category
  scheme="http://service.openta.net/feeds/atompub/formal-category?lang=en"
  term="angebot-stelle" />
  label="offer/job offer" />
<category
  scheme="http://service.openta.net/feeds/atompub/formal-category?lang=en"
  term="projekt"
  label="project" />
</entry>
```

Example 2: Notation if the scheme and / or label reference cannot be included

In many CMS systems a category declaration of an article can only be specified by choosing either a language specific category label from a predefined list or by directly typing in the category name in a text field. And these category declarations will then be (“incorrectly”) mapped to an Atom feed category definition as term attribute (and not as label attribute as it should). To respect this misbehavior the openTA feed aggregation system also recognizes “label” values as terms if label and scheme attributes are omitted. Thus, the following

example which uses English labels as term values will be correctly interpreted. Thus, English or German labels can be used as category names (aka term values) in the case that the CMS feed component cannot create a feed with correct scheme, term and label attributes.

```
<entry xmlns="http://www.w3.org/2005/Atom">
  <title>Example</title>
  <id>Example ID</id>
  <updated>2013-02-25T12:30:02Z</updated>
  <category term="offer/job offer"/>
  <category term="project"/>
</entry>
```

Example 3: Notation if scheme reference cannot be included and a clear separation of openTA categories from other categories is intended:

If the formal openTA categories shall be indicated in order to differentiate them from other schemes used but if the scheme entry is missing this can be solved by adding the prefix "openTA/".

```
<entry xmlns="http://www.w3.org/2005/Atom">
  <title>Example</title>
  <id>Example ID</id>
  <updated>2013-02-25T12:30:02Z</updated>
  <category term="openTA/offer/job offer"/>
  <category term="openTA/project"/>
</entry>
```

3.2 Use in RSS feeds

RSS is a family of data formats based on XML. Different functionalities are supported depending on the format version. The indication of categories is identical in each version. For each entry in an RSS feed a random number of categories can be assigned in the following format:

```
<category domain="http://example.org/bar">foo</category>
```

The name of the category is defined by the value of the *category* tag and the taxonomy (i.e. set of given category names) is defined by the optional indication of the *domain* attribute. No standard exists for the indication of the taxonomy but use of an URL is recommended. Since there is also no standard in RSS for the description of categories which a URL would reference a reference to the category documents of the APP should be included in the domain attribute:

```
de: http://service.openta.net/feeds/atompub/scheme/formal-category
en: http://service.openta.net/feeds/atompub/scheme/formal-category?lang=en
```

Examples

The examples below each contain two category entries, one for the openTA category „job offer“, which is a sub-category of „offer“ (1.1), and another for the category “project” (3)

Example 1: Recommended notation

```
<item>
  <title>Example</title>
  <guid> Example ID</guid>
```



```

<pubDate>2013-02-25T12:30:02Z</pubDate>
<category
domain="http://service.openta.net/feeds/atompub/formal-category?lang=en">
offer/job offer</category>
<category
domain="http://service.openta.net/feeds/atompub/formal-category">
project</category>
</item>

```

Note, that in RSS category declarations language specific labels and not term keys will be used as value for the category declaration. Both, German or English labels can be used. The domain attribute can even be omitted.

Example 2: Notation if indication of domain reference is not possible

If it is not possible to indicate the domain attribute this can be left out. The entry will still be recognized correctly by openTA. However, the entry will lose context since a neutral reader will not be able any more to interpret which categories stem from the same scheme.

```

<item>
<title>Example</title>
<guid>Example ID</guid>
<pubDate>2013-02-25T12:30:02Z</pubDate>
<category>offer/job offer</category>
<category>project</category>
</item>

```

Example 3: Notation if domain reference cannot be included and a clear separation of openTA categories from other categories is intended

If the formal openTA categories shall be indicated in order to differentiate them from other schemes used but if the *domain* entry is missing this can be solved by adding the prefix "openTA/".

```

<item>
<title>Example</title>
<guid>Example ID</guid>
<pubDate>2013-02-25T12:30:02Z</pubDate>
<category>openTA/offer/job offer</category>
<category>openTA/project</category>
</item>

```


Appendix A: APP-Category Document for the openTA Formal Categories

German

URL:

<http://service.openta.net/feeds/atompub/scheme/formal-category>

Content:

```
<categories xmlns="http://www.w3.org/2007/app"
xmlns:atom="http://www.w3.org/2005/Atom">
  <atom:category term="angebot" label="Angebot" />
  <atom:category term="angebot-stelle" label="Angebot/Stellenangebot" />
  <atom:category term="angebot-koop" label="Angebot/Kooperationsangebot" />
  <atom:category term="angebot-misc" label="Angebot/Sonstiges" />
  <atom:category term="inst" label="Institutionelles" />
  <atom:category term="inst-pers" label="Institutionelles/Personalia" />
  <atom:category term="inst-misc" label="Institutionelles/Sonstiges" />
  <atom:category term="projekt" label="Projekt" />
  <atom:category term="pub" label="Publikation" />
  <atom:category term="veranst" label="Veranstaltung" />
  <atom:category term="veranst-call" label="Veranstaltung/Call for Papers" />
  <atom:category term="veranst-ankuend" label="Veranstaltung/Ankündigung" />
  <atom:category term="veranst-bericht" label="Veranstaltung/Bericht" />
  <atom:category term="misc" label="Sonstige" />
</categories>
```

English

URL:

<http://service.openta.net/feeds/atompub/scheme/formal-category?lang=en>

Content:

```
<categories xmlns="http://www.w3.org/2007/app"
xmlns:atom="http://www.w3.org/2005/Atom">
  <atom:category term="angebot" label="offer" />
  <atom:category term="angebot-stelle" label="offer/job offer" />
  <atom:category term="angebot-koop" label="offer/cooperation offer" />
  <atom:category term="angebot-misc" label="offer/other" />
  <atom:category term="inst" label="institutional matters" />
  <atom:category term="inst-pers" label="institutional matters/personnel mat-
ters" />
  <atom:category term="inst-misc" label="institutional matters/other" />
  <atom:category term="projekt" label="project" />
  <atom:category term="pub" label="publication" />
  <atom:category term="veranst" label="event" />
  <atom:category term="veranst-call" label="event/call for papers" />
  <atom:category term="veranst-ankuend" label="event/announcement" />
  <atom:category term="veranst-bericht" label="event/report" />
  <atom:category term="misc" label="other" />
</categories>
```