



Date: 2013-05-03
Version: 1.3
Ref:

Legal deposit of electronic documents by means of RSS feeds

Change history

Version 1.3

- Mandatory elements must now come first under each item in a given sequence (adaptation to the corresponding xml-schema *rss_kb.xsd* now published).
- The element R105 *<title>* is now mandatory, while R106 *<description>* is optional.
- The element R107 *<dcterms:accessRights>* now gets one of two possible values: 'gratis' or 'restricted'.
- The elements R118 *media:keywords* and R118 *media:category* have been split into two tables each (R118 and F311); one is for the description of the resource (*item*) as a whole, the other table is for constituent files (objects), described as *media:content*.
- Support for retrieval via *https* is now implemented by KB.

Version 1.2

- Under the heading "Publishing": Namespace addresses for MediaRSS and Dublin Core is now correct.

Version 1.1

- The element *link* is now mandatory
- New element *dcterms:isFormatOf* (optional) added

Editorial change:

- References (ID-no and headings) to the document "Introduktion till metadata i leveranser av elektroniska dokument till KB" (in Swedish) at the top of most tables.

1. Introduction

This document describes how legal deposit suppliers of electronic documents can deliver material by means of RSS feeds.

2. Publishing

A supplier wishing to deliver via feeds publishes an RSS feed at a fixed URL. The frequency of retrieval is settled in an agreement with Kungl. biblioteket (KB) – The National Library of Sweden. The RSS feed needs to fulfil the following conditions:

- The format should be RSS 2.0 with the addition of some elements from MediaRSS 1.5.0 (<http://search.yahoo.com/mrss/>) and Dublin Core DCMI Metadata Terms (<http://purl.org/dc/terms/>).

For MediaRSS the best information about the format is presently found here: <http://www.rssboard.org/media-rss/>. Please observe, however, that this is still only a draft document, as stated in the editor's note, while the link to the official specification at yahoo.com is currently inactive.

For Dublin Core, please note that KB uses the more comprehensive namespace for DCMI Metadata Terms, *not* the more restricted Dublin Core Metadata Element Set (<http://purl.org/dc/elements/1.1/>), encompassing only 15 elements, being only a small subset of DCMI Metadata Terms. Some of the Dublin Core elements in this specification are exclusively part of the DCMI Metadata Terms. This concerns for example *accessRights* and *isFormatOf* in the format tables below (section 5).

The prefix *dc* is sometimes used also for the namespace of the more comprehensive DCMI Metadata Terms (<http://purl.org/dc/terms/>). This is accepted, as long as it is consistently used exclusively for this namespace. For the sake of clarity, however, we recommend the use of the prefix *dcterms* for all elements in Dublin Core.

Summary of recommended formats, prefixes and namespaces:

Format	Commonly used prefix	Namespace URI	Documentation
RSS 2.0	-	-	http://www.rssboard.org/rss-specification
MediaRSS 1.5.0	<i>media</i>	http://search.yahoo.com/mrss/	http://www.rssboard.org/media-rss
DCMI Metadata Terms	<i>dcterms</i>	http://purl.org/dc/terms/	http://dublincore.org/documents/dcmi-terms/

- Each item in the RSS feed represents an electronic document that is to be delivered. Metadata about the document is expressed in the formats RSS, MediaRSS and *dcterms* for each item. The files that make up the electronic document that is represented by an item are designated using the elements `<link>` from RSS [and], `<media:content>` from MediaRSS and `dcterms:format`. These designations contain information about where the files are published (URL) and their logical format (MIME-type). The files designated in this way should be accessible for retrieval by KB, at the URLs given. All URLs should be of type *http* or *https*.
- *Items* in the feeds should be sorted by the value in the element `<pubDate>` in descending order, i.e. with the latest published item first.
- The element `<pubDate>` should have as its value the time when the latest version of the document was published. If a new version of the document is published, the value of `<pubDate>` should be updated and the *item* put in the correct sequence further up in the feed.
- All *items* that are published or have new versions published during a time period corresponding to the interval of retrieval (settled in the agreement between KB and the supplier) must be shown in the feed.
- All 7 mandatory elements (R101-R105, R107, F303 *dcterms:format*) must appear first under each item and in this sequential order for validation with the xml-schema now published: *rss_kb.xsd*.

3. Limitations

Delivery by means of feeds has certain limitations:

- Delivery of files with DRM-protection is not supported.

4. Security

It is anticipated that suppliers may wish to deliver material that is not meant to be publically accessible and thus needs to be protected from unauthorized retrieval by a third party. This can be solved using one of the following methods:

- IP-filter – the supplier configures its network infrastructure so as to allow retrieval only from certain IP-addresses, used by KB only.
- Log in with Basic Authentication – the supplier protects the resources by username/password to be used by KB.

Support for retrieval via *https* is now implemented by KB.

5. Format

This is a selection of elements from RSS, MediaRSS and *dcterms*, for which KB specifies certain requirements in order to make delivery via feeds work properly. Other parts of the given standards (for RSS, MediaRSS and *dcterms*) may also be used in a feed and, if supplied, some of those will also be used by KB (as supplementary metadata). If any element from RSS 2.0 under <item> is repeated, only the first instance will be used by KB.

R101	Identifier
/rss:rss/rss:channel/rss:item/rss:guid	
Mandatory	Yes
Format	Optional
Unique and permanent identifier of the resource. Will be used for feedback to the supplier and to identify new versions of the same resource.	

R102	Internet address
/rss:rss/rss:channel/rss:item/rss:link	
Mandatory	Yes
Format	URL
Internet address (URL) of the electronic resource at the time of publishing, that is where the resource was first made publicly available. May be used for feedback to the supplier and to identify new versions of the same resource. This URL will be used to retrieve one of the files that make up the document. The same information content that is designated by <link> may also be delivered in another format, without ads, navigation etc. This other format will then be given by an element <media:content>, having as a child-element <dcterms:isFormatOf> with the same URL value as <link>. (See below F302 and following element).	

R103	Publishing date
/rss:rss/rss:channel/rss:item/rss:pubDate	
Mandatory	Yes
Format	RFC822
Time of publishing. If the resource is republished in a new version, this date should be updated. All items in a channel should be sorted by this field in descending order, with newest item first.	

R104	Publisher
/rss:rss/rss:channel/rss:item/dcterms:publisher	
Mandatory	Yes
Format	http://id.kb.se/organisations/SE<organisation number>
Identifier of institution (publisher) that made the resource available on the internet. Swedish organisation number, written without hyphen. In agreement with KB a suffix might be added for reference to particular divisions within an organisation.	

R105	Title
/rss:rss/rss:channel/rss:item/rss:title	
Mandatory	Yes
Format	Text
Name of resource.	

R106	Description
/rss:rss/channel/rss:item/rss:description	
Mandatory	No
Format	Text
Short description of content, e.g. introduction, abstract or summary of resource.	

R107	Accessibility at the time of publishing
/rss:rss/channel/rss:item/dcterms:accessRights	
Mandatory	Yes
Format	Controlled values (either of two): <i>gratis</i> <i>restricted</i>
Indicate if the resource is freely accessible (' <i>gratis</i> ') at the time of publication. In other case, if there are particular conditions for access to the resource (such as payment, subscription, encryption), select ' <i>restricted</i> '.	

R108	License
/rss:rss/channel/rss:item/dcterms:license	
Mandatory	No
Format	URI
Can be used to state a license encompassing all constituent parts of the document. This could be for example Creative Commons.	

R115	Statement of responsibility (Creator)
/rss:rss/channel/rss:item/dcterms:creator	
Mandatory	No
Format	For names of persons: Family name, Given name (role)
Name(s) of person(s) or institution(s) that are responsible for the intellectual or artistic content, for example author, photographer, composer.	

R115	Statement of responsibility (Contributor)
/rss:rss/channel/rss:item/dcterms:contributor	
Mandatory	No
Format	For names of persons: Family name, Given name (role)
Name(s) of person(s) or institution(s) that contributed to the intellectual or artistic content of the resource, for example illustrator, translator.	

F303	File format (for object designated by <i>link</i>)
/rss:rss/channel/rss:item/dcterms:format	
Mandatory	Yes
Format	MIME
MIME-type of the file designated by the <rss:link>-element	



S201	Constituent objects (files) that make up the resource
	/rss:rss/rss:channel/rss:item/(media:group/)media:content
Mandatory	No
Format	-
Points to a part of the document. Can be grouped in <i>media:group</i> , but only one of the representations in a <i>media:group</i> will be retrieved by KB.	

S201	Premium object (file) in a <i>media:group</i>
	/rss:rss/rss:channel/rss:item/(media:group/)media:content/@isDefault
Mandatory	Yes, if <i>media:group</i> is provided
Format	true false
Decides which <i>media:content</i> from one <i>media:group</i> , that will be retrieved.	

F302	Internet address of constituent object / Retrieval address for alternative format
	/rss:rss/rss:channel/rss:item/(media:group/)media:content/@url
Mandatory	Yes, if <i>media:content</i> is provided.
Format	URL
Internet address (URL) to this part of the electronic resource at the time of publication. This URL will be used for retrieval of one of the files that make up the document.	

-	Indicator of alternative format
	/rss:rss/rss:channel/rss:item/(media:group/)media:content/dcterms:isFormatOf
Mandatory	No
Format	URL (same value as in <link>)
Same URL as in <link>. The presence of this element indicates that the parent element <i>media:content</i> describes the same content as in <link>, but in another format. (See above F302 and R102).	

F303	File format (for object in <i>media:content</i>)
	/rss:rss/rss:channel/rss:item/(media:group/)media:content/@type
Mandatory	Yes, if <i>media:content</i> is provided
Format	MIME
MIME-type of the file designated by this <i>media:content</i>	

F305	Checksum
	/rss:rss/rss:channel/rss:item/(media:group/)media:content/media:hash
Mandatory	No
Format	MD5
An MD5 checksum for the designated file.	

F307	License for a constituent object (file)
/rss:rss/rss:channel/rss:item/(media:group/)media:content/media:license	
Mandatory	No
Format	URI
Can be used to indicate a license that is valid for a certain part of the document, for example Creative Commons.	

R118	Keywords (subject) for item
/rss:rss/rss:channel/rss:item/media:keywords	
Mandatory	No
Format	Comma separated list of keywords
Keywords relevant for the resource (<i>item</i>) as a whole.	

F311	Keywords (subject) for constituent file (object)
/rss:rss/rss:channel/rss:item/(media:group/)media:content/media:keywords	
Mandatory	No
Format	Comma separated list of keywords
Keywords relevant for a constituent file (object) described as <i>media:content</i> .	

F308	Holder of rights pertaining to object (file) / contributor	
/rss:rss/rss:channel/rss:item/media:credit /rss:rss/rss:channel/rss:item/(media:group/)media:content/media:credit		
Mandatory	No	
Attribute	role	Role according to scheme
	scheme	urn:ebu urn:yvs
Format	For names of persons: Family name, Given name (role)	
Name and role for the person (or institution) that has contributed to the artistic or intellectual content of the resource.		

R118	Category (subject) of item	
/rss:rss/rss:channel/rss:item/media:category		
Mandatory	No	
Attribute	scheme	URI
	label	Category name in clear text
Format	Value according to selected scheme	
Can be used for categorizing the resource (<i>item</i>) as a whole		

F311	Category (subject) of constituent file (object)	
/rss:rss/rss:channel/rss:item/(media:group/)media:content/media:category		
Mandatory	No	
Attribute	scheme	URI
	label	Category name in clear text
Format	Value according to selected scheme	