

E-pliktleverans via RSS-feeds

Referens till detta dokument:

<http://www.kb.se/namespace/digark/deliveryspecification/deposit/rssfeeds/>

Ändringshistorik

Version 1.3

- Obligatoriska element måste nu först komma i given ordning (anpassning till xml-schema *rss_kb.xsd* som publiceras samtidigt)
- Elementet R105 Titel *<title>* blir nu obligatoriskt, medan R106 Beskrivning *<description>* blir frivilligt. (Tidigare fanns möjligheten att välja vilken som helst av dessa två.)
- Elementet R107 Tillgänglighet vid publicering *dcterms:accessRights* får nu ett av två möjliga värden: *gratis* eller *restricted*.
- Elementen R118 Nyckelord *media:keywords* och R118 Kategori *media:category* har delats upp på två olika tabeller vardera (R118 och F311); den ena för beskrivning av resursen som helhet direkt under *item*, den andra tabellen för beskrivning av medföljande filer under *media:content*.
- Stöd för hämtning via *https* är nu implementerat.

Version 1.2

Under rubriken ”Publicering”: Adresser till namespace för MediaRSS respektive Dublin Core är nu korrekta. (**korr. 2013-04-05**). Lagt till en sammanfattning i tabellform över format och namespaces.

Version 1.1

- Elementet *link* är nu obligatoriskt
- Nytt element *dcterms:isFormatOf* tillagt

Redaktionella ändringar: Titeln är något ändrad från version 1.0 (E-pliktleverans via feeds). Referenser (id-nr och rubrik) till dokumentet ”Introduktion till metadata i leveranser av elektroniska dokument till KB” inleder de flesta tabeller.

1. Inledning

Detta dokument beskriver hur en leverantör kan leverera material som omfattas av e-plikt med hjälp av feeds.

2. Publicering

En leverantör som vill leverera material via en feed publicerar en RSS-feed på en fix URL. Hämtningsfrekvensen överenskommes med KB. Denna feed behöver uppfylla följande:

- Formatet skall vara RSS 2.0 med tillägg av vissa element från MediaRSS 1.5.0 (<http://search.yahoo.com/mrss/>) och Dublin Core DCMI Metadata Terms (<http://purl.org/dc/terms/>).

För MediaRSS hittar man för närvarande bäst information om formatet här: <http://www.rssboard.org/media-rss>. Men observera att detta (ännu så länge) bara är ett utkast (draft) till en ny version 1.5.1 av formatet, och att tills vidare är det version 1.5.0 ovan som gäller, fastän länken dit är inaktiv.

För Dublin Core vill vi särskilt göra uppmärksam på att KB använder den mer omfattande namnrymden för DCMI Metadata Terms, *inte* det mer inskränka Dublin Core Metadata Element Set (<http://purl.org/dc/elements/1.1/>), som omfattar endast 15 termer och är en liten delmängd av DCMI Metadata Terms. Några av Dublin Core elementen i specifikationen finns bara i DCMI Metadata Terms, det gäller till exempel R107 *accessRights* och ”Markör för alternativ form” *isFormatOf* i formattabellerna nedan (avsnitt 5).

Det förekommer att man använder bara *dc* som prefix för namespace till DCMI Metadata Terms och det accepteras också, så länge man är konsekvent och det används endast för detta namespace. Men för tydlighets skull rekommenderar vi ändå att man använder prefixet *dcterms* för element i Dublin Core.

Sammanfattning av format och namnrymder (namespaces) med vanligt förekommande prefix:

Formatnamn	Vanligt förekommande prefix	Namespace URI	Dokumentation
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/

- Varje *item* i RSS-flödet representerar ett elektroniskt dokument som skall levereras. Metadata om dokumentet uttrycks med hjälp av RSS, MediaRSS och dcterms i *itemet*. De datafiler som tillsammans utgör det elektroniska dokumentet pekats ut med hjälp av elementen <link> från RSS och <media:content> från MediaRSS. Dessa utpekningar innehåller information om var filen är tillgängliggjord (URL) och dess logiska format (MIME-typ). De filer som pekats ut måste vara tillgängliga för hämtning av KB på de URL:er som anges. Alla URL:er skall vara av typen *http* eller *https*.
- *Items* skall vara sorterade på värdet i elementet <pubDate> med senast publicerade *item* först.
- Elementet <pubDate> skall innehålla tidpunkten då den senaste versionen av dokumentet publicerats. Om en ny version av dokumentet publiceras skall värdet på <pubDate> uppdateras och *itemet* skall sorteras in på rätt plats i listan.
- Alla de *items* som publiceras eller får nya versioner under en tidsperiod som motsvarar hämtningsintervallet måste synas i feeden.
- Alla 7 obligatoriska element (R101-R105, R107, F303 *dcterms:format*) måste komma först och i den ordningen under varje *item* för validering mot xml-schemat som nu publiceras: *rss_kb.xsd*.

3. Begränsningar

Leverans via feeds har vissa begränsningar:

- Leveranser av filer som har DRM-skydd stöds ej

4. Säkerhet

Det förutses att leverantörer kan vilja leverera material som ej skall vara publikt tillgängligt och som därför behöver skyddas från obehörig hämtning av tredje part. Detta kan lösas med någon eller några av följande metoder:

- IP-filtrering – Leverantören konfigurerar sin nätverksinfrastruktur att endast tillåta hämtningar från vissa IP-adresser som används av KB
- Inloggning med Basic Authentication – Leverantören skyddar resurserna med användarnamn/lösenord och kommunicerar dessa till KB

KB erbjuder nu stöd för leverantörer som önskar skydda resurser genom att endast tillåta hämtning via *https*.

5. Format

Detta är ett urval av element ur RSS, MediaRSS och *dcterms* där KB ställer särskilda krav på användningen för att leveransförfarandet via feeds skall fungera. Även övriga delar av standarderna kan användas i ett feed och vissa andra delar kommer användas av KB om de finns angivna.

Om något element från RSS 2.0 under item repeteras så används endast den första förekomsten.

R101	Identifikator
/rss:rss/rss:channel/rss:item/rss:guid	
Obligatorisk	Ja
Format	Valfritt
Unik och beständig identifikator som givits resursen. Denna identifikator kommer att användas vid återkoppling till leverantören och för att identifiera nya versioner av samma resurser.	

R102	*Nätadress
/rss:rss/rss:channel/rss:item/rss:link	
Obligatorisk	Ja
Format	URL
<p>Nätadress (URL) till den elektroniska resursen vid publiceringstillfället, dvs. till den internetadress där resursen först tillgängliggjordes för allmänheten. Kan komma att användas vid återkoppling till leverantören och för att identifiera nya versioner av samma resurser.</p> <p>URLen i <link> kommer att användas för att hämta en av filerna som utgör resursen.</p> <p>Samma information som pekats ut av <link> kan även levereras i ett annat format, t.ex. utan annonser, navigering och liknande. Detta andra format pekats då ut av en <media:content> innehållande ett <dcterms:isFormatOf> som pekar på länken i <link>.</p>	

R103	*Publiceringsdatum
/rss:rss/rss:channel/rss:item/rss:pubDate	
Obligatoriskt	Ja
Format	RFC822
Tidpunkt för publicering. Om resursen publiceras i en ny version skall detta datum uppdateras. Alla items i en channel skall vara sorterade på detta fält (i fallande datumordning) med nyast item först.	

R104	Utgivare
/rss:rss/rss:channel/rss:item/dcterms:publisher	
Obligatorisk	Ja
Format	http://id.kb.se/organisations/SE<organisationsnummer>
<p>Identifikator på den instans som gjort resursen tillgänglig i ett nätverk (utgivare).</p> <p>Organisationsnummer skrivs här utan bindestreck. Det går att efter överenskommelse med KB lägga till suffix för att referera till delar av en organisation.</p>	

R105	Titel
/rss:rss/rss:channel/rss:item/rss:title	
Obligatoriskt	Ja
Format	Text
Det namn som resursen är känt under.	

R106	Beskrivning
/rss:rss/rss:channel/rss:item/rss:description	
Obligatorisk	Nej
Format	Text
Kort redogörelse för resursens innehåll, t.ex. ingress, sammanfattning eller abstract.	

R107	Tillgänglighet vid publicering
/rss:rss/rss:channel/rss:item/dcterms:accessRights	
Obligatorisk	Ja
Format	Kontrollerad värdemängd: <i>gratis</i> <i>restricted</i>
<p>Ange om resursen är fritt tillgänglig (<i>gratis</i>) vid publiceringstillfället.</p> <p>I annat fall, om det för tillgång till resursen gäller särskilda villkor (betalning, abonnemang, kryptering) anges <i>restricted</i>.</p>	

R108	Licens
/rss:rss/rss:channel/rss:item/dcterms:license	
Obligatorisk	Nej
Format	URI
Kan användas för att ange en licens som omfattar alla ingående delar av dokumentet. Detta kan t.ex. vara Creative Commons.	

R115	Upphov (intellektuellt / konstnärligt ansvarig)
/rss:rss/rss:channel/rss:item/dcterms:creator	
Obligatorisk	Nej
Format	För person: Efternamn, Förnamn (roll)
Namn på person (eller institution) som är <i>ansvarig</i> för resursens intellektuella eller konstnärliga innehåll, t.ex. författare, fotograf, kompositör	

R115	Upphov (bidragit)
/rss:rss/rss:channel/rss:item/dcterms:contributor	
Obligatorisk	Nej
Format	För person: Efternamn, Förnamn (roll)
Namn på person (eller institution) som <i>bidragit</i> till resursens intellektuella eller konstnärliga innehåll, t.ex. illustratör, översättare.	

F303	Filformat (för objekt utpekad av link)
/rss:rss/rss:channel/rss:item/dcterms:format	
Obligatorisk	Ja
Format	MIME
Anger MIME-format på den fil som pekats ut av elementet <rss:link>	

S201	Objekt (datafiler) som ingår i resursen
/rss:rss/rss:channel/rss:item/(media:group)media:content	
Obligatorisk	Nej
Format	-
<p>Pekar ut del av dokumentet. Kan grupperas i <i>media:group</i> men endast en av representationerna i en <i>media:group</i> kommer att hämtas av KB.</p> <p>Kan även användas för att ange en fil med samma innehåll som det som pekas ut av <i>link</i>, men i annat format, t.ex. utan annonser, navigering och liknande. (Se R102, F302).</p>	

S201	Utpekat objekt (datafil) i en grupp om flera
/rss:rss/rss:channel/rss:item/(media:group)media:content/@isDefault	
Obligatorisk	Ja, om <i>media:group</i> är angivet
Format	true false
Anger vilken <i>media:content</i> ur en <i>media:group</i> som skall hämtas.	

F302	Objektets nätadress / Hämtningsadress för alternativ form
/rss:rss/rss:channel/rss:item/(media:group)media:content/@url	
Obligatorisk	Ja, om <i>media:content</i> finns
Format	URL
Nätadress (URL) till denna del av den elektroniska resursen vid publiceringstillfället. Denna URL används för att hämta en av filerna som utgör dokumentet.	

-	Markör för alternativ form
/rss:rss/rss:channel/rss:item/(media:group/)media:content/dcterms:isFormatOf	
Obligatorisk	Nej
Format	URL (samma som i <link>)
Elementets värde är samma nätadress (URL) som anges i <link>. Om detta element anges betyder det att överordnad (parent) <i>media:content</i> beskriver samma information som det som pekas på av <link>, men i en annan form (se ovan R102, S201 och F302).	

F303	Filformat (för objekt i media:content)
/rss:rss/rss:channel/rss:item/(media:group/)media:content/@type	
Obligatorisk	Ja, om <i>media:content</i> finns
Format	MIME
Anger MIME-format på den fil som pekas ut av denna <i>media:content</i> .	

F305	Checksumma
/rss:rss/rss:channel/rss:item/(media:group/)media:content/media:hash	
Obligatorisk	Nej
Format	MD5
Anger en checksumma för den utpekade filen beräknad enligt algoritmen MD5.	

F307	Licens för objektet
/rss:rss/rss:channel/rss:item/(media:group/)media:content/media:license	
Obligatorisk	Nej
Format	URI
Kan användas för att ange en licens som omfattar en viss del av dokumentet. Detta kan t.ex. vara Creative Commons.	

R118	Nyckelord (Ämne)
/rss:rss/rss:channel/rss:item/media:keywords	
Obligatorisk	Nej
Format	Kommaseparererad lista med nyckelord
Relevanta nyckelord för en hel resurs	

F311	Nyckelord (Ämne)
/rss:rss/rss:channel/rss:item/(media:group/)media:content/media:keywords	
Obligatorisk	Nej
Format	Kommaseparererad lista med nyckelord
Relevanta nyckelord för en enskild fil (ett objekt) som hör till resursen	

F308	Upphovsrättsinnehavare/Upphov för objektet	
/rss:rss/rss:channel/rss:item/media:credit		
/rss:rss/rss:channel/rss:item/(media:group/)media:content/media:credit		
Obligatorisk	Nej	
Attribut	role	Roll enligt tillåtna värden i scheme
	scheme	urn:ebu urn:yvs
Format	För person: Efternamn, Förnamn (roll)	
Namn och roll för person (eller institution) som bidragit till resursens intellektuella eller konstnärliga innehåll.		

R118	Kategori (Ämne)	
	/rss:rss/channel/rss:item/media:category	
Obligatorisk	Nej	
Attribut	scheme	URI
	label	Klartextnamn på kategorin
Format	Värde enligt valt scheme	
Kan användas för att ange en kategorisering av resursen.		

F311	Kategori (Ämne)	
	/rss:rss/channel/rss:item/(media:group)media:content/media:category	
Obligatorisk	Nej	
Attribut	scheme	URI
	label	Klartextnamn på kategorin
Format	Värde enligt valt scheme	
Kan användas för att ange en kategorisering av en enskild fil (ett objekt) som hör till resursen		