

E-pliktleverans via RSS-feeds

Referens till detta dokument:

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

Ändringshistorik

Version 2.1

- Ett frivilligt element *dcterms:isPartOf* har lagts till för att möjliggöra angivande av **värdepublikation** på artikelnivå, under *item*.

Version 2.0

- Obligatoriska element behöver nu **inte längre** komma i given ordning. Det är en anpassning baserad på XML Schema 1.1, som tillåter godtycklig ordning mellan obligatoriska element.
- XPath-uttryck för rss har korrigerats och anges nu utan prefix, eftersom RSS inte har något eget namespace.
- S201 *media:content* blir nu obligatoriskt, om tillämpligt. Det gäller speciellt bildfiler, ljudspår, videoklipp som hör till en resurs (item).
- S201 **Utpekad objekt (datafil) -**
//media:group/media:content/@isDefault i version 1.3 utelämnas. KB hämtar hem alla objekt i en *media:group*, oavsett innehåll.
- Ett element *dcterms:references* har lagts till för att möjliggöra utpekandet av en mediafil som levereras via annan kanal, t.ex. *ftp*.
- F308 **Upphovsrättsinnehavare/Upphov för objektet** har delats på två element: **(F308) Upphov för objektet – *media:credit***, och **F308 Upphovsrättshavare för objektet – *media:copyright***.
- Ett nytt avsnitt **5. Validering** har lagts till, med hänvisning till gällande xml-scheman.
- Information om vilka element som är *upprepningsbara* (R) har lagts till för varje element/attribut. (R) betyder här att elementet kan förekomma flera gånger *under ett item*.

Version 1.3

- Obligatoriska element måste nu först komma i given ordning (anpassning till xml-schema *rss_kbse.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. Oavsett vilket prefix man använder skall alltså alla Dublin Core-element i RSS-feeds levererade enligt denna specifikation ha namespace <http://purl.org/dc/terms/>. För tydlighets skull rekommenderar vi ändå att man använder prefixet *dcterms* för alla element i Dublin Core.

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> <i>dc</i>	http://purl.org/dc/terms/	http://dublincore.org/documents/dcmi-terms/

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

- 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 pekas 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 pekas 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.
- Observera att om ett metadataelement anges som icke obligatoriskt, betyder det inte automatiskt att den (de) fil(er) som beskrivs av metadataelementet är undantagna från leveransplikt.

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. Validering

För kontroll av att levererade RSS-feeds uppfyller kraven i denna specifikation tillhandahåller KB xml-scheman, som leverantören själv kan använda för validering. Dessa uppdaterade xml-scheman bygger på XMLSchemaversion 1.1. De finns att ladda ner här:

<http://www.kb.se/namespace/rss/index.html> .

För att valideringen skall fungera måste alla sex olika scheman (1. *rss-kbse.xsd*, 2. *dcterms_kbse.xsd*, 3. *media_kbse.xsd*, 4. *content_kbse.xsd*, 5. *dc.xsd*, 6. *dcmitype.xsd*) ligga i samma filkatalog som huvudschemat (nr. 1 ovan). I den programvara som används för valideringen, välj XML Schema **version 1.1** som default och någon av följande xml-parsers: **Xerces**, Relax NG, Schematron. (Saxon EE fungerar inte).

6. 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. Obligatoriska element har en * framför namnet och ett Ja i fältet för Obligatoriskt. Upprepningsbara element har ett (R) efter namnet. Det betyder att elementet kan förekomma flera gånger *under samma item*.

Även övriga delar av standarderna kan användas i en feed och vissa andra delar kommer användas av KB om de finns angivna.

R101	*Identifikator
/rss/channel/item/guid	
Obligatoriskt	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/channel/item/link	
Obligatoriskt	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/channel/item/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/channel/item/dcterms:publisher	
Obligatoriskt	Ja
Format	http://id.kb.se/organisations/SE<organisationsnummer>
Identifikator på den instans som gjort resursen tillgänglig i ett nätverk (utgivare). Organisationsnummer skrivs här utan bindestreck. Det går att efter överenskommelse med KB lägga till suffix för att referera till delar av en organisation.	

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

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

R107	*Tillgänglighet vid publicering
/rss/channel/item/dcterms:accessRights	
Obligatoriskt	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/channel/item/dcterms:license	
Obligatoriskt	Nej
Format	URI
<p>Kan användas för att ange en licens som omfattar alla ingående delar av dokumentet. Detta kan t.ex. vara Creative Commons.</p>	

R112	Värdpublikation
/rss/channel/item/dcterms:isPartOf	
Obligatoriskt	Nej
Format	URI
<p>Kan användas för att på artikelnivå ange en värdpublikation, t.ex. den e-tidskrift artikeln ingår i, i de fall då artiklar från flera olika värdpublikationer förekommer i samma RSS-feed. Annars hämtas värdpublikation från kanalen (<i>channel</i>).</p>	

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

R115	Upphov (bidragit) (R)
/rss/channel/item/dcterms:contributor	
Obligatoriskt	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/channel/item/dcterms:format	
Obligatoriskt	Ja
Format	MIME
Anger MIME-format på den fil som pekas ut av elementet <link>	

S201	*Objekt (datafiler) som ingår i resursen (R)
/rss/channel/item/(media:group/)media:content	
Obligatoriskt	Ja, om tillämpligt
Format	-
Pekar ut del av en resurs (item). Gäller särskilt för ljudspår, videoklipp, bildfiler som hör till resursen (med @type ="audio/...", "video/..."image/..." ; se nedan F303). Kan grupperas i <i>media:group</i> .	
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 (t.ex. med @type="text/html"; se även R102, F302).	

F302	*Objektets nätadress / Hämtningsadress för alternativ form (R)
/rss/channel/item/(media:group/)media:content/@url	
Obligatoriskt	Ja, om <i>media:content</i> finns
Format	URL
<p>Nätadress (URL) till objekt (filer) som hörde till den elektroniska resursen vid publicerings-tillfället.</p> <p>Kan även användas som hämtningsadress för alternativ form av resursen.</p>	

-	Markör för alternativ form
/rss/channel/item/(media:group/)media:content/dcterms:isFormatOf	
Obligatoriskt	Nej
Format	URL (samma som i <link>)
<p>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).</p>	

F303	*Filformat (för objekt i media:content)
/rss/channel/item/(media:group/)media:content/@type	
Obligatoriskt	Ja, om <i>media:content</i> finns
Format	MIME Media types: "video/...", "audio/...", "image/..."
<p>Anger MIME Media Type (http://www.freeformatter.com/mime-types-list.html) på den fil som pekas ut av denna <i>media:content</i>. Består av: <i>förled / efterled</i>, där förledet anger innehållstyp (t.ex. "video", "audio" eller "image"), och efterledet filformat. Exempel: <i>video/x-flv, audio/mpeg, image/jpeg</i>.</p> <p>För att ange samma resurs som i <i>link</i> men i alternativt format (utan annonser etc.) kan även exempelvis "<i>text/html</i>" förekomma som @type. (Se ovan S201).</p>	

[S201]	*Objekt (mediafiler) som tillhör resursen, men som levereras separat (t.ex. via ftp) (R)
/rss/channel/item/dcterms:references	
Obligatoriskt	*Ja, om tillämpligt
Format	uri
Alternativ för mediafiler som inte medföljer i media:content.	

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

F307	Licens för objektet
/rss/channel/item/(media:group/)media:content/media:license	
Obligatoriskt	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) (R)
/rss/channel/item/media:keywords	
Obligatoriskt	Nej
Format	Kommaseparerad lista med nyckelord
Relevanta nyckelord för en hel resurs	

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

(F308)	Upphov för objektet (R)	
/rss/channel/item/(media:group/)media:content/media:credit		
Obligatoriskt	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 objektets intellektuella eller konstnärliga innehåll.		

F308	Upphovsrättsinnehavare för objektet (R)	
/rss/channel/item/(media:group/)media:content/media:copyright		
Obligatoriskt	Nej	
Attribut	url	
Format	Text. För person: Efternamn, Förnamn	
Upphovsrättshavare till resursen (item) eller objektet (mediafilen).		

R118	Kategori (Ämne) (R)	
<i>/rss/channel/item/media:category</i>		
Obligatoriskt	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) (R)	
<i>/rss/channel/item/(media:group/)media:content/media:category</i>		
Obligatoriskt	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		