

E-pliktleverans via RSS-feeds

Referens till detta dokument:

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

Ändringshistorik

Version 2.4.1

- Rättning av namnrymd för XML Schema instance

Version 2.4

- Ändrad hänvisning till formatet för angivning av publiceringsdatum (**R103**). RFC822, som har blivit ersatt, anges som format för publiceringsdatum i RSS 2.0 men med undantag att publiceringsår ska anges med fyra tecken. Följ rekommendationen om publiceringsåret i RSS 2.0 som överensstämmer med formatet RFC2822.

Version 2.3

- Elementet *dcterms:license* (**R108**) tillåter nu bara värden med format URI, inte längre vanlig text.
- Elementet *dcterms:isPartOf* (**R112**) med attribut *@xsi:type* som anger identifikatortyp, är obligatoriskt för de fall då värdpublikationen är leveranspliktig (tryckt eller elektronisk).
- Nytt element **Annat format**, *dcterms:isFormatOf* (**R113**) med attribut *@xsi:type* som anger identifikator för en annan utgåva i annat format (tryckt eller elektronisk) av levererad resurs (item).
- Elementet F303 *Filformat (för objekt utpekade av link) har bytt namn och referensnummer till **R117 *Resurstyp** för att särskilja det från det andra elementet *F303 Filformat (för objekt i media:content)*.
- Ny korrigerad specifikation av *media:license* (**F307**).
- Elementen **R118** och **F311** Kategori (Ämne) har lyfts ur specifikationen, eftersom KB inte använder dessa vid konvertering till arkivformat och bibliotekskatalog. (En RSS-feed med

motsvarande värden för */rss/channel/item/category* och */rss/channel/item/(media:group/)media:content/media:category* kommer fortfarande att validera, men dessa element blir då inte synliga utåt någons.)

Version 2.2

- Ett frivilligt element *dcterms:identifier* med attribut *@xsi:type* har lagts till för att möjliggöra angivande av **identifikortyp**, med syftet att förbättra beskrivningen av resursen och dubblettkontrollen.
- Elementen *dcterms:isPartOf* och *dcterms:references* kan nu också ha ett attribut *@xsi:type* som anger identifikortyp för referensen.

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 Upphovsrättsinnehavare 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 6).

- Prefix som *dc* eller *dct* används ibland också för namespace till DCMI Metadata Terms och det accepteras också, så länge man är konsekvent och det används endast för detta namespace. Oavsett prefix skall alltså alla Dublin Core-element i RSS-feeds 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 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/
XML Schema instance (används för type-attribut)	<i>xsi</i>	http://www.w3.org/2001/XMLSchema-instance	http://www.w3.org/TR/xmlschema-1/#no-xsi

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 finns två alternativ, som leverantören själv kan använda för validering. Enklast är att använda KB:s valideringstjänst, där man kopierar och klistrar in sin RSS-feed själv och sedan klickar ”Validera!”: <https://e-plikt.kb.se/validera/rss>.

Alternativt finns våra xml-scheman, byggda på XML Schema version 1.1, 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.

6. Format

Nedan följer tabeller med ett urval av element ur RSS, MediaRSS och dcterms där KB ställer särskilda krav på användningen för att leveransförandet 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 för resursen. Användas vid återkoppling till leverantören och för att identifiera nya versioner av samma resurser.	

R101a	Identifikator med typvärde (R)
/rss/channel/item/dcterms:identifier[@xsi:type]	
Obligatoriskt	Nej, men önskvärt för dubblettkontroll.
Format element	doi, ean, hdl, isan, isbn, ismn, isrc, issue-number, matrixnumber, upc, uri, urn
Format @attribut	Fasta värden: "dcterms:doi", "dcterms:ean", "dcterms:hdl", "dcterms:isan", "dcterms:isbn" etc.
Unik och beständig identifikator för resursen. Kan vara samma som i R101 <i>item/guid</i> . Men flera olika identifikatorer av olika typ kan anges (elementet kan upprepas). För att ange typvärde med @xsi:type, måste även namespace för xsi finnas med, antingen i rotelementet <rss> eller i det element där @xsi:type förekommer. Attributets typvärde måste ha samma prefix som det man använt för namespace http://purl.org/dc/terms/ i roten. Exempel:	
<pre><rss xmlns:dcterms=http://purl.org/dc/terms/ version="2.0"> <dcterms:identifier xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="dcterms:isbn">9783452679123</dcterms:identifier></pre>	

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 pekas ut av <link> kan även levereras i ett annat format, t.ex. utan annonser, navigering och liknande. Detta andra format pekas 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	Se pubDate på https://validator.w3.org/feed/docs/rss2.html
<p>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.</p> <p>OBS! Publiceringsåret ska alltid anges med fyra tecken.</p>	

R104	*Utgivare
/rss/channel/item/dcterms:publisher	
Obligatoriskt	Ja
Format	<a href="http://id.kb.se/organisations/SE<organisationsnummer>+ev.[-suffix]">http://id.kb.se/organisations/SE<organisationsnummer>+ev.[-suffix]
<p>Unik identifikator för den instans som gjort resursen tillgänglig i ett nätverk (utgivare). <i>Organisationsnummer</i> skrivs här <i>utan bindestreck</i>.</p> <p>Om flera olika utgivare (t.ex. tidningar) har samma organisationsnr som en gemensam leverantör (distributör, ägare) skall man efter överenskommelse med KB lägga till ett –suffix för att referera till den specifika utgivare (tidning) som publicerat resursen (artikeln). Suffixet skall föregås av bindestreck '-' och innehålla minst två tecken [Aa-Zz,0-9]</p> <p>Exempel:</p> <p><i>Dalademokraten</i>, levereras av Mittmedia AB, med organisationsnr. 556004-1815, som delas med flera andra tidningar. För <i>Dalademokraten</i> anges då under utgivare, med överenskommet suffix:</p> <pre><item> <dcterms:publisher>http://id.kb.se/organisations/SE5560041815-DD </dcterms:publisher> </item></pre>	

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, men önskvärt för rättighetsbehandling.
Format	URI
<p>Nätadress till ett licensavtal som omfattar alla ingående delar av dokumentet.</p> <p>Exempel: http://creativecommons.org/licenses/by/3.0/</p>	

R112	*-/Värdpublikation (R)
/rss/channel/item/dcterms:isPartOf[@xsi:type]	
Obligatoriskt	Ja, om leveransplikt gäller för värdpublikationen.
Format element	doi, ean, hdl, isan, isbn, ismn, isrc, issn, issue-number, matrix-number, upc, uri, urn
Format @attribut	Fasta värden: "dcterms:doi", "dcterms:ean", "dcterms:hdl", "dcterms:isan", "dcterms:isbn", "dcterms:issn" etc.
<p>En "värdpublikation" som en resurs (<i>item</i>) ingår i eller hör till. Om värdpublikationen också är leveranspliktig (tryckt eller elektronisk) är elementet obligatoriskt. Exempel är en tryckt lärobok med tillhörande elektroniskt publicerade övningsuppgifter. Vid e-pliktleverans av övningsuppgifterna anges då som minimum ISBN till huvudboken som "värdpublikation".</p> <p>Används även för att ange vilken värdpublikation, t.ex. en e-tidskrift, som en artikel ingår i, i de fall då artiklar från flera olika värdpublikationer förekommer i samma RSS-feed. Elementet kan upprepas om en artikel publicerats i flera olika värdpublikationer. Annars hämtas värdpublikation automatiskt från kanalen (channel/link). För formatering av elementet, se även under R101a Identifierator med typvärde.</p>	

R113	Annat format (R)
/rss/channel/item/dcterms:isFormatOf[@xsi:type]	
Obligatoriskt	Nej
Format element	doi, ean, hdl, isan, isbn, ismn, isrc, issn, issue-number, matrix-number, upc, uri, urn
Format @attribut	Fasta värden: "dcterms:doi", "dcterms:ean", "dcterms:hdl", "dcterms:isan", "dcterms:isbn", "dcterms:issn" etc.
<p>En referens till en resurs som är utgiven i ett annat format, men som är identiskt till innehållet. Det kan gälla t.ex. när en resurs ges ut både som pdf och epub. Ange en identifikator för den relaterade resursen. Elementet kan upprepas om flera utgåvor i olika format finns. För formatering av elementet, se även under R101a Identifikator med typvärde.</p> <p>Exempel (ISBN för tryckt utgåva av en elektronisk resurs):</p> <pre><dcterms:isFormatOf xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance xsi:type="dcterms:isbn">9789186197475</dcterms:isFormatOf></pre>	

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)
<p>Namn på person (eller institution) som <i>bidragit</i> till resursens intellektuella eller konstnärliga innehåll, t.ex. illustratör, översättare.</p>	

R117	*Resurstyp
/rss/channel/item/dcterms:format	
Obligatoriskt	Ja
Format	Media Type (mer känt som MIME-type)
<p>Media Type (MIME-type) enligt https://www.iana.org/assignments/media-types/media-types.xhtml för den fil som pekas ut av elementet R102 <link>.</p> <p>Består av: <i>förled / efterled</i> , där förledet anger mediatyp och efterledet filformat, t.ex. <i>application/pdf, audio/mpeg, image/tiff</i>.</p> <p>Exempel: <dcterms:format>text/html</dcterms:format></p>	

S201	*Objekt (datafiler) som ingår i resursen (R)
/rss/channel/item/(media:group/)media:content	
Obligatoriskt	Ja, om tillämpligt
Format	-
<p>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>.</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 (exempelvis med @type="text/html"; se även R102, F302).</p>	

F302	*Objektets nätadress (R) / Hämtningsadress för alternativ form
/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 alternativt format av resursen (se ovan S201).</p>	

-	Markör för alternativ form
/rss/channel/item/(media:group/)media:content/dcterms:isFormatOf	
Obligatoriskt	Nej
Format	URL (samma som i <link>)
Elementets värde är samma nätadress (URL) som anges i . Om detta element anges betyder det att överordnad (parent) media:content 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/channel/item/(media:group/)media:content/@type	
Obligatoriskt	Ja, om <i>media:content</i> finns
Format	Media Type (MIME-type)
Media Type (MIME-type) enligt https://www.iana.org/assignments/media-types/media-types.xhtml för den fil som pekas ut av denna media:content. Består av: <i>förled</i> / <i>efterled</i> , där förledet anger mediatyp och efterledet filformat. Exempel: <i>video/mp4, audio/mpeg, image/tiff</i> . 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).	

S201	*Objekt (mediafiler) som tillhör resursen, men som levereras separat (t.ex. via ftp) (R)
/rss/channel/item/dcterms:references[@xsi:type]	
Obligatoriskt	*Ja, om tillämpligt
Format element	doi, ean, hdl, isan, isbn, ismn, isrc, issn, issue-number, matrix-number, upc, uri, urn
Format @attribut	Fasta värden: "dcterms:doi", "dcterms:ean", "dcterms:hdl", "dcterms:isan", "dcterms:isbn" etc.
Alternativ för mediafiler som <i>inte</i> medföljer i <i>media:content</i> . För formatering av elementet, se exempel under R101a Identifier med typvärde .	

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 (R)
/rss/channel/item/(media:group/)media:content/media:license[@href]	
Obligatoriskt	Nej
Format element	Text
Format @attribut	URI
<p>Kan användas för att ange en licens som omfattar en viss del av dokumentet. Elementvärdet kan lämnas tomt, oavsett innehåll är det attributet @href som används av KB för rättighetshantering. Exempel:</p> <pre><media:license type="text/html" href="http://creativecommons.org/licenses/by/3.0/de/">Creative Commons tysk licens</media:license></pre>	

R118	Nyckelord (Ämne) (R)
/rss/channel/item/media:keywords	
Obligatoriskt	Nej
Format	Kommaseparerad lista med nyckelord
Relevanta nyckelord för <i>hela resursen</i> (item)	

F311	Nyckelord (Ämne) (R)
/rss/channel/item/(media:group/)media:content/media:keywords	
Obligatoriskt	Nej
Format	Kommaseparerad lista med nyckelord
Relevanta nyckelord för <i>en enskild fil</i> (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	Text. För person: Efternamn, Förnamn	
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).		