National Library of Sweden

Datum 2013-03-18 (DOS-anpassad 2025-04-07)

Version 2.0

Dnr

Introduktion till metadata i leveranser av elektroniska dokument till KB

Referens till detta dokument:
http://www.kb.se/namespace/digark/deliveryspecification/metadataintro/

National Library of Sweden

ÄNDRINGSHISTORIK

Version 2.0

En viss ändring i disposition av dokumentet har gjorts sedan version 1.0 och varje tabell har försetts med ett id-nr. Titeln har ändrats från det tidigare ”Metadata i e-pliktleveranser”.

National Library of Sweden

1. INLEDNING

Detta dokument vänder sig till leverantörer av e-plikt och ger en allmän introduktion till metadata som behöver medfölja i leveranserna. Med *metadata* menas information om och kring det elektroniska material som levereras.

2. BAKGRUND

I regeringens proposition 2011/12:121, Leveransplikt för elektroniskt material, hittar man följande stycke i inledningen till kapitel 8:

”Pliktexemplar av elektroniskt material ska vara åtföljt av information om tillgängliggörandet, det elektroniska materialets logiska format och samband med annat som omfattas av leveransplikt samt koder eller liknande som kan krävas för att man på avsett vis ska kunna ta del av innehållet i det elektroniska materialet.”

Lagtexten säger inte så mycket mer om vilka metadata som Kungliga biblioteket (KB) kan begära från leverantören. Utan tvivel är viss information helt nödvändig för att KB ska kunna identifiera och hantera materialet på ett automatiserat sätt. Annan information är kanske inte lika nödvändig men kan höja kvaliteten och därmed användbarheten väsentligt. Vissa typer av metadata kommer vi att med hjälp av program kunna extrahera vid leveransen, andra behöver komma från leverantören.

Metadata om resurserna förs även över till den nationella katalogen LIBRIS varmed information om resurserna kommer att bli sökbar och synlig direkt efter leverans.

Ju mer metadata som vi får tillgång till desto större blir möjligheterna för framtida bevarande, återsökning och visning av det inlevererade materialet!

National Library of Sweden

3. LEVERANSKANALER

KB erbjuder ett antal olika alternativ för att leverera. Dels kommer KB att kunna erbjuda olika tekniker, som t.ex. leverans via FTP, eller att KB med hjälp av robot hämtar hos leverantören.

Dessutom kommer vi att definiera ett antal ”leveranskanaler”, ibland kanske flera för en och samma leverantör. Vilka dessa är beror på leveransteknik och på vilket material som levereras.

4. METADATA OM LEVERANS, RESURS OCH OBJEKT

KB har i sin analys av vilken information som behövs analyserat tre olika nivåer för metadata – metadata om *leveransen*, om *resurserna* samt om de *objekt* som utgör resurserna.

En *leverans* kan innehålla en eller flera resurser. Med en *resurs* menar vi en ”publikation” eller ”dokument” - t.ex. en e-bok, ett nummer av en tidning, en artikel på webben, en film, ett foto, en text på webben, ett inspelat TV-inslag, en ljudinspelning på radion, etc. En resurs kan i sin tur bestå av en eller flera datafiler eller dataströmmar. Dessa kallar vi här för *objekt*.

5. STANDARDER OCH FORMAT FÖR METADATA

Metadata levereras i det format som KB och leverantören kommit överens om. KB föredrar leverans i något standardiserat, dvs. erkänt och ofta använt, format. Valet av metadata avgörs av sammanhanget. Det finns olika metadatastandarder för olika typer av resurser och för olika syften.

De standarder som KB för närvarande arbetar med är RSS, Dublin Core samt METS i kombination med något bibliografisk format som t.ex. MODS. Dessa beskrivs i särskilda tekniska specifikationer för e-pliktleverans.

National Library of Sweden

6. ÖNSKVÄRDA METADATA

Här nedan följer en lista i tabellform över den information (metadata) som Kungl. biblioteket (KB) önskar ska medfölja varje resurs som levereras enligt lagen om e-plikt. Varje tabell motsvarar ett metadata-element (”informations-bit”). Varje sådant metadata-element har ett ID-nummer och en tillhörande rubrik. ID-numret för varje element är detsamma som sedan används som referens i de tekniska specifikationerna för de olika metadataformaten (MODS, Dublin Core, RSS etc.). Tabeller med ID-nummer R101-R120 behandlar information om själva resursen, tabeller S201-203 beskriver strukturen för leveranspaketet, tabeller med ID-nummer F301-F311 beskriver de enskilda datafiler (objekt) som utgör en resurs.

Ett ”Ja” på raden för ”Hög prioritet” betyder att denna information är mycket viktig, ibland även nödvändig, att få med för att KB ska kunna hantera materialet på ett automatiserat sätt. Högprioriterade metadata är även markerade i rubrikraden med en asterix (*).

Varje tabell innehåller även en definition, en beskrivande förklaring, av metadata-elementet.

All information som ges i tabellerna är medvetet uttryckt på en mer konceptuell nivå. Vilka metadata som behöver skickas med, samt hur dessa uttrycks i form av ”metadatataggar”, är delvis beroende av leveranssätt samt valet av metadataformat. Mer exakta anvisningar kring detta ges i respektive teknisk specifikation.

National Library of Sweden

7. METADATA OM RESURSEN

En leverans kan innehålla en eller flera resurser. Med resurs menar vi den ”publikation” eller ”dokument” som givits ut på nätet som en ”enhet”, och ska levereras enligt lag - t.ex. en e-bok, en artikel på webben, en film, ett foto, ett inspelat TV-inslag, ett ljudspår, etc.

R101 *Identifikator	
Hög prioritet	Ja
Beskrivning	En unik och beständig identifikator (identifierare) som givits resursen. Identifikatorn används vid återkoppling till leverantören och för att identifiera nya versioner av samma resurs. Exempel på identifikatorer: DOI, EAN, GUID, Handle, ISRC, URN, ett internt nummer, t.ex. utgivarens artikelnummer eller liknande. I vissa fall kan även resursens nätadress användas som identifikator, se 102 *Nätadress.

R102 *Nätadress	
Hög prioritet	Ja
Beskrivning	Leverantörens nätadress (URL) vid publiceringstillfället. Kan komma att användas vid återkoppling till leverantören för att identifiera nya versioner av samma resurs.

R103 *Publiceringsdatum	
Hög prioritet	Ja
Beskrivning	Tidpunkt för publicering i något godkänt datumformat, t.ex. W3CDTF (profil av ISO 8601) eller RFC822.

R104 *Utgivare	
Hög prioritet	Ja
Beskrivning	Identifikator och/eller namn på den instans som gjort resursen tillgänglig i ett nätverk (förlag). Identifikatorn hämtas från KB:s leverantörsregister.

R105 *Titel	
Hög prioritet	Ja
Beskrivning	Det namn som resursen är känd under. Titeln finns vanligtvis på en framträdande plats, t.ex. rubriken till en artikel eller på förstasidan i en PDF-fil. Kan även anges på objektnivå, se F310 Objektets titel.

R106 Beskrivning	
Hög prioritet	Nej
Beskrivning	En kort redogörelse för resursens innehåll, t.ex. en sammanfattning från webbsida, abstract, baksidestext, ingress, katalogtext eller liknande. Kan även anges på objektnivå, se F311 Objektets beskrivning.

R107 *Tillgänglighet vid publicering	
Hög prioritet	Ja
Beskrivning	Här anges värdet *'gratis'* för att i förekommande fall indikera att resursen vid publiceringstillfället var fritt tillgänglig för var och en att läsa utan betalningskrav eller andra barriärer som t.ex. lösenordsskydd.

R108 *Licens	
Hög prioritet	Ja
Beskrivning	Information om att någon form av licens gäller för resursen. Anges i form av en identifikator/URL som refererar till en licens, t.ex. Creative Commons. Kan även anges på objektnivå, se F307 *Licens för objektet.

R109 Upphovsrättsinnehavare	
Hög prioritet	Nej
Beskrivning	Namn på den instans (person eller organisation) som äger rättigheterna till materialet. Kan även anges på objektnivå, se F308 Upphovsrättsinnehavare för objektet.

R110 Relaterad resurs	
Hög prioritet	Nej (men se R111 och R112)
Beskrivning	En referens till en resurs som har en relation till och anses viktig för förståelsen av den beskrivna resursen. De vanligast förekommande typerna av relationer beskrivs i R111-R114, men fler relationer skulle kunna förekomma. Uttrycks vanligen med hjälp av 3 delar av information: *typ av relation*; *identifikator för den relaterade resursen*; *formatet för identifikatorn*.

R111 *Serie (Relaterad resurs)	
Hög prioritet	Ja
Beskrivning	En referens till en serie som resursen ingår i. Typexempel är den förlagsserie som en e-bok tillhör. Bäst är att här ange en identifikator för serien (t.ex. ISSN). I vissa metadataformat finns även möjlighet att ange en bibliografisk referens innehållande seriens titel, identifikator och resursens eventuella nummerbeteckning i serien.

R112 *Värdpublikation (Relaterad resurs)	
Hög prioritet	Ja
Beskrivning	En referens till en värdpublikation som resursen ingår i. Typexempel på värdpublikation är den webbtidning som en artikel hämtas från, proceedings som ett konferensbidrag är utgiven i, musikalbum som ett ljudspår ingår i. Bäst är att här ange en identifikator för värdpublikationen (t.ex. DOI, EAN, GUID, URL, URN). I vissa metadataformat finns även möjlighet att ange en bibliografisk referens som kan innehålla såväl identifikator och titel på värdpublikation som andra uppgifter som kan vara viktiga för förståelsen av resursen.

R113 Annat format av (Relaterad resurs)	
Hög prioritet	Nej
Beskrivning	En referens till en resurs som är utgiven i ett annat fysiskt format men som är identiskt till innehållet. Typexempel är när en resurs ges ut både som PDF och EPUB. Bäst är att här ange formattyp och en identifikator för den relaterade resursen.

R114 Annan version av (Relaterad resurs)	
Hög prioritet	Nej
Beskrivning	En referens till en resurs som är i princip samma som den som beskrivs men där innehåll och/eller utförande har förändrats. Typexempel är en tidigare utgåva. Bäst är att ange en identifikator för den relaterade resursen.

R115 *Upphov	
Hög prioritet	Ja
Beskrivning	Namn på person eller institution som ansvarat för eller bidragit till resursens intellektuella eller konstnärliga innehåll, t.ex. författare, illustratör, fotograf, översättare, kompositör.

R116 Språk	
Hög prioritet	Nej
Beskrivning	Information om vilket (vilka) språk som dominerar i resursen. Anges gärna med hjälp av en textsträng eller kod från en känd standard, t.ex. ISO 639-2/B language codes (”swe”, ”eng”, etc.).

R117 Resurstyp	
Hög prioritet	Nej
Beskrivning	En term som beskriver resursens huvudsakliga egenskap (till form och innehåll). Några typexempel är artikel, bok, video, roman, ljud, bild, statistik, musik, etc.

R118 Ämne	
Hög prioritet	Nej
Beskrivning	Resursens huvudsakliga innehåll uttryckt med hjälp av nyckelord eller termer från en ämnesordlista. Upprepa elementet om flera ämnesord anges.

R119 Geografisk täckning	
Hög prioritet	Nej
Beskrivning	Ett geografiskt område som innehållet i resursen omfattar eller handlar om.

R120 Täckning i tid	
Hög prioritet	Nej
Beskrivning	En uppgift om tid som innehållet i resursen avser eller handlar om.

8. METADATA OM STRUKTUR

Varje leverans kan innehålla flera resurser som i sin tur kan bestå av en eller flera filer (objekt). Metadata som uttrycker vilka filer som hör till vilken resurs måste alltid finnas med.

S201 *Objekt (datafiler) som ingår i resursen	
Hög prioritet	Ja
Beskrivning	Information som klargör vilka objekt (datafiler) som ingår i resursen. Kan i dess enklaste form vara en lista med filnamn.

S202 Objektets roll i resursen	
Hög prioritet	Nej
Beskrivning	Information som klargör objektens (datafilernas) relation till resursen. Typexempel kan vara ”omslagsbild”, eller att en av filerna utgör resursens ”startpunkt”.

S203 Ordningsföljd mellan objekt	
Hög prioritet	Nej
Beskrivning	Information som klargör ordningsföljden mellan objekten

National Library of Sweden

9. METADATA OM OBJEKT

Varje resurs består av en eller flera filer (objekt). Vissa metadata om varje objekt behövs…

F301 *Objektets identifikator (Filnamn)	
Hög prioritet	Ja
Beskrivning	En alfanumerisk sträng som identifierar ett objekt (en fil, en dataström) som tillhör resursen. Kan vara ett filnamn, men i vissa fall kan även objektets nätadress användas som identifikator (se F302 *Objektets nätadress). Alla identifikatorer måste vara unika inom en och samma leverans.

F302 *Objektets nätadress	
Hög prioritet	Ja
Beskrivning	Leverantörens unika nätadress (URL) till objektet vid publiceringstillfället. Om objektet inte har en unik nätadress krävs någon form av identifikator. Se F301 *Objektets identifikator.

F303 *Filformat	
Hög prioritet	Ja
Beskrivning	Datafilens format uttryckt enligt någon konvention, t.ex. som *MIME-typ*, eller *formatnamn*, *ID* och *versionsbeteckning* hämtat ur ett formatregister.

F304 Filstorlek	
Hög prioritet	Nej
Beskrivning	Filens storlek uttryckt i antal bytes.

F305 Checksumma	
Hög prioritet	Nej
Beskrivning	En summa som räknats ut och som är unik för just denna fil. Räknas ut enligt MD5.

F306 *Kryptering eller lösenord	
Hög prioritet	Nej
Beskrivning	Information om att datafilen är krypterad eller lösenordsskyddad. Krypteringsnyckel eller lösenord måste i så fall följa med.

F307 *Licens för objektet	
Hög prioritet	Ja
Beskrivning	Information om att någon form av licens gäller för objektet. (Se även R108 *Licens).

F308 Upphovsrättsinnehavare för objektet	
Hög prioritet	Nej
Beskrivning	Namn på den instans (person eller organisation) som äger rättigheterna till objektets innehåll. (Se även R109 Upphovsrättsinnehavare).

F309 Objektets språk	
Hög prioritet	Nej
Beskrivning	Information om vilket (vilka) språk som dominerar i objektet. (Se även R116 Språk).

F310 Objektets titel	
Hög prioritet	Nej
Beskrivning	Ett namn som objektet (filen) är känd under om det skiljer sig från resursens titel (se även R105 *Titel).

F311 Objektets beskrivning	
Hög prioritet	Nej
Beskrivning	En kort redogörelse för objektets innehåll, t.ex. en bildtext. (Se även R106 Beskrivning).