

Metadata i e-pliktleveranser

Referens till det här dokumentet:

<http://www.kb.se/namespace/digark/metadaintro/v1/>

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 tillgänglighörandet, 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!

Leveranskanaler

Det kommer att finnas ett antal 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.

3. 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*.

4. 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 i kombination med Dublin Core samt METS i kombination med något av bibliografiska format som t.ex. MODS.

KB behöver komma fram till ett mindre antal modeller för leverans av metadata. För detta krävs fortsatta efterforskningar om vad som är möjligt för leverantörerna.

5. Tabeller över önskvärda metadata

Här nedan följer två tabeller som listar de informationsbitar (metadata) om resurserna som KB önskar få i samband med leverans. Den ena

tabellen behandlar information om själva resursen, och den andra tar upp sådant som är bra att veta om varje datafil.

Rubriker och beskrivningar i tabellen är medvetet uttryckta på en mer konceptuell än konkret nivå – exakt vilken information och hur denna uttrycks i form av ”metadatataggar” är beroende av leveransmetod och av vilken metadatastandard som används.

Förklaring av tabellen:

- **Rubrik.** Namn på den bit av information som beskrivs.
- **Hög prioritet.** Om ”Ja” betyder det att detta är information som är nödvändigt för att KB ska kunna hantera materialet på ett automatiserat sätt. Dessa är även markerade med en asterix (*).
- **Beskrivning.** En beskrivande förklaring (definition) av informationsbiten.

5.1. Metadata om resursen

Rubrik	001 *Identifikator
Hög prioritet	Ja
Beskrivning	En unik och beständig identifikator (identifierare) som givits resursen. Kan komma att användas vid återkoppling till leverantören och för att identifiera nya versioner av samma resurs. Exempel på identifikatorer: DOI, URN, Handle, GUID, ISRC, EAN, ett internt nummer, t.ex. utgivarens artikelnummer eller liknande. I vissa fall kan även resursens nätadress användas som identifikator. Se Nätadress .

Rubrik	002 *Nätadress
Hög prioritet	Ja
Beskrivning	Leverantörens nätadress till den elektroniska resursen vid publiceringstillfället. Kan komma att användas vid återkoppling till leverantören för att identifiera nya versioner av samma resurs. Om resursen inte har en unik nätadress krävs någon form av identifikator. Se Identifikator .

Rubrik	003 *Publiceringsdatum
Hög prioritet	Ja
Beskrivning	Tidpunkt för publicering på nätet angivet i något godkänt datumformat, t.ex. W3CDTF (profil av ISO 8601) eller RFC822.

Rubrik	004 *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.

Rubrik	005 *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. Kan även anges på objektnivå, se Objektets titel .

Rubrik	006 Beskrivning
Hög prioritet	Nej
Beskrivning	En kort redogörelse för resursens innehåll, t.ex. en sammanfattning från webbsida, baksidestext, ingress, katalogtext eller liknande.

Rubrik	007 *Tillgänglighet vid publicering
Hög prioritet	Ja
Beskrivning	Här anges värdet ' <i>gratis</i> ' 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.

Rubrik	008 *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 *Licens för objektet .

Rubrik	009 Upphovsrättsinnehavare
Hög prioritet	Nej
Beskrivning	Namn på den instans (person eller organisation) som äger rättigheterna till materialet. Kan även anges på objektnivå, se Upphovsrättsinnehavare för objektet

Rubrik	010 *Relaterad resurs
Hög prioritet	Ja
Beskrivning	En referens till en resurs som har en relation till, och anses

	viktig för förståelsen av, den beskrivna resursen. Exempel: förlagsserie som e-boken tillhör; webbtidning som artikeln är publicerad i; musikalbum som ljudspåret ingår i. Uttrycks med hjälp av 3 delar av information: <i>typ av relation; identifikator för den relaterade resursen; formatet för identifikatorn.</i>
--	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Rubrik	011 *Upphovsman
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.

Rubrik	012 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.). Kan även anges på objektnivå, se Objektets språk

Rubrik	013 Resurstyp/Genre
Hög prioritet	Nej
Beskrivning	En term som beskriver resursens huvudsakliga egenskap.

Rubrik	014 Ämne
Hög prioritet	Nej
Beskrivning	Resursens huvudsakliga innehåll uttryckt med hjälp av nyckelord eller termer från en ämnesordlista.

Rubrik	015 Geografisk täckning
Hög prioritet	Nej
Beskrivning	Ett geografiskt område som innehållet i resursen omfattar eller handlar om.

Rubrik	016 Täckning i tid
Hög prioritet	Nej
Beskrivning	En uppgift om tid som innehållet i resursen avser eller handlar om.

Rubrik	017 *Objekt (datafiler) som ingår i resursen
Hög prioritet	Ja.
Beskrivning	Information som klargör objektens (datafilernas) relation till resursen och till varandra. Kan i dess enklaste form vara en lista över de objekt (datafiler) som tillsammans utgör en resurs. Det är också värdefullt att få med vilken typ av relation objektet har till resursen, t.ex. om resursen innehåller 2 filer varav en innehåller text och den andra en bild.

5.2. Metadata om objekt som resursen består av

Rubrik	018 *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. Alla identifikatorer måste vara unika inom en och samma leverans. Se Objektets nätadress

Rubrik	019 *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 Objektets identifikator .

Rubrik	020 *Filformat
Hög prioritet	Ja
Beskrivning	Datafilens format uttryckt som MIME-typ.

Rubrik	021 Filstorlek
Hög prioritet	Nej
Beskrivning	Filens storlek uttryckt i antal bytes.

Rubrik	022 Checksumma
Hög prioritet	Nej
Beskrivning	En summa som räknats ut och som är unik för just denna fil. Räknas ut enligt MD5

Rubrik	023 *Kryptering eller lösenord
Hög prioritet	Ja
Beskrivning	Information om att datafilen är krypterad eller lösenordsskyddad. Krypteringsnyckel eller lösenord måste i så fall följa med.

Rubrik	024 *Licens för objektet
Hög prioritet	Ja
Beskrivning	Information om att någon form av licens gäller för objektet. Se även Licens .

Rubrik	025 Upphovsrättsinnehavare för objektet
Hög prioritet	Nej
Beskrivning	Namn på den instans (person eller organisation) som äger rättigheterna till objektets innehåll. Se även Upphovsrättsinnehavare .

Rubrik	026 Objektets språk
Hög prioritet	Nej
Beskrivning	Information om vilket (vilka) språk som dominerar i objektet. Se även Språk .

Rubrik	027 Objektets titel
Hög prioritet	Nej
Beskrivning	Ett namn som objektet (filen) är känd under om det skiljer sig från resursens titel. Se även Titel .