

**Rapport från ENSAM, E-arkivsamverkansgruppens
metadatagrupp**

2010-05-12

Uppdrag

I den nationella E-arkivsamverkansgruppen (ENSAM) har man sett behovet av att ta fram en gemensam profil över hur ett paket i OAIS-modellen kan se ut. Man tillsatte då en arbetsgrupp för metadata som fick i uppdrag att ta fram en sådan profil. Profilen ska kunna användas oavsett var i OAIS-modellen paketet finns samt oberoende av vem användaren är. Metadatagruppen kom fram till att man kan enas om hur basprofilen ska byggas upp. Därefter får respektive användare skapa en utökningsprofil med sina speciella regler.

Dessa profiler är ej fastslagna att användas utan det är varje organisation och brukare av profilerna som fastställer användning.

Gruppen har haft följande medlemmar:

Renata Arovelius, Sveriges lantbruksuniversitet (SLU)
Steve Olsson, Stockholms läns landsting (SLL)
Martina Engsjö-Lindgren, Stockholms läns landsting (SLL)
Stefan Fogelvik, Stockholms stadsarkiv (SSA)
Catrine Bergstrand, REgionalt Digitalt Arkiv (REDA) och Riksarkivet (RA)
Caspar Gielissen, Sambruk, kommunala e-tjänster
Stina Degerstedt, Kungliga Biblioteket (KB)
Elisabet Bengtsson, Kungliga Biblioteket (KB), Vid ett tillfälle
Eli Hjorth Reksten, Linköpings Universitet (LiU)
Anna Ekström, Landstingsnätverket R7
Göran Lindqvist, LDB Centrum
Mats Berggren, Riksarkivet (RA)
Karin Bredenberg, Riksarkivet (RA), sammankallande

Rapport från ENSAM, E-arkivsamverkansgruppens metadatagrupp	1
Uppdrag.....	2
Arbetets gång och ställningstaganden	4
Beslut.....	4
Namngivning av bland annat id	4
METS grundprofil	4
Att använda ENSAM's basprofiler	5
Arbetsgång vid användandet av ENSAM's basprofiler	5
METS	6
Introduktion	6
Dokumentets struktur	6
METS-profiler	7
Basprofiler	7
Yttre och inre METS	8
Namngivning av profilerna	8
Användandet av element i basprofilerna	8
Agenter	16
Övriga standarder.	18
Beskrivande metadata	18
Administrativt metadata	19
Kombinationsformat.....	20
Länkar.....	23

Arbetets gång och ställningstaganden

Vi inledde gruppens arbete med en genomgång av standarder som kan användas för denna typ av information. Vi enades om att för paketbeskrivningen använda METS-standarderna. Detta då det är den standard som har kommit längst. Det finns även andra standarder för paketstruktur. En är DIDL men den hanterar en paketering baserad på formatet MPEG. NASA har tagit fram XDFU som liknar METS. Den är inte fullt lika använd och inte lika lätt att finna information om som METS. Så valet att använda METS grundar sig på att det redan finns ett antal profiler att titta på och standarderna uppfyller de krav på dokumentation om filerna som gruppen vill ha uppfylla.

Beslut

Gruppen fattade att antal principiella beslut.

Vi beslutade att:

- Med den tillgängliga tiden för gruppens arbete kan vi inte specificera användandet av övriga metadatastandarder. Dessa metadatastandarder kan utöka informationsmängden om filerna. Vi kan däremot tala om vilka som finns.
- Det inte är möjligt att göra en profil som passar alla. Vi tar därför fram en basprofil. Denna får sedan respektive användare utöka för att passa vald användning.
- Skapa en basprofil där samtliga beskrivna filer skall finnas med i paketet.
- Skapa en basprofil speciellt för bibliotek. Denna profil behöver tillåta att information om de paketerade objekten kan finnas på annan plats. Vid leveranstillfället kan man behöva hämta metadata från andra system, t.ex. LIBRIS. I övrigt skall samtliga filer som beskrivs finnas i paketet.
- Starkt rekommendera att samtliga XML-dokument som finns i ett paket skall åtföljas av de XML-scheman som använts.
- Starkt rekommendera att det är obligatoriskt med checksumma på samtliga filer.
- Starkt rekommendera att checksumme-typen ska vara MD5 eller SHA-x.
- Starkt rekommendera att nivån av Structmap skall bestämmas på lokal nivå. Structmap är det enda obligatoriska elementet i standarden.
- Starkt rekommendera att i den mån bevarande-metadata finns ska PREMIS användas.

Namngivning av bland annat id

Vi har valt att inte definiera exakt hur id och liknande ska namnges då detta kan variera för respektive användare. Vi rekommenderar därför starkt att man i sin utökningsprofil definierar detta.

Man bör använda standardformer för namn.

För ett id som gäller identifiering av en organisation rekommenderas användande av standarden ISO 15511 Information and documentation - International Standard Identifier for Libraries and Related Organizations (ISIL)

METS grundprofil

Element samt attribut som vi valt att använda i basprofilen är en gemensamt bestämd lägsta nivå för att ett paket ska vara validerbart och autentiskt enligt de krav vi i gruppen ställer. En

basprofil gör det möjligt för att alla som använder samma basprofil, att assistera varandra vid eventuella problem.

Valda element och attribut gås igenom i stycket ”Användandet av element i basprofilerna”.

Att använda ENSAM:s basprofiler

När man väljer att använda ENSAM's basprofiler måste man göra ett antal ställningstaganden:

- Kommer ett paket se likadant ut oavsett om det är en SIP, AIP eller DIP?
- Behöver basprofilen utökas? Se stycket ”Basprofiler”.
- Hur ska namngivningen av id-begreppen vara?
- Hur ser ordlistor för till exempel USE-attributet ut?
- Hur namnges paketet?
- Hur ser agenternas information ut? Se stycket ”Agenter”.
- Hur ser eventuell uppackningsinformation ut?
- Hur ska uppbyggnaden av StructMap vara utformad?
- Användande av yttre och inre METS för kvalitetssäkring? Se stycket ”Yttre och inre METS”.
- Behövs det mer information om filerna än det som är obligatoriskt i basprofilerna?
- Vilka andra regler kommer att gälla för paketen?

Arbetsgång vid användandet av ENSAM:s basprofiler

För att göra en egen anpassning.

1. Definiera egna regler.
2. Skapa en utökningsprofil.
3. Skapa ett anpassat schema.
4. Publicera schemat på en beständig adress.
5. Skapa METS-dokument

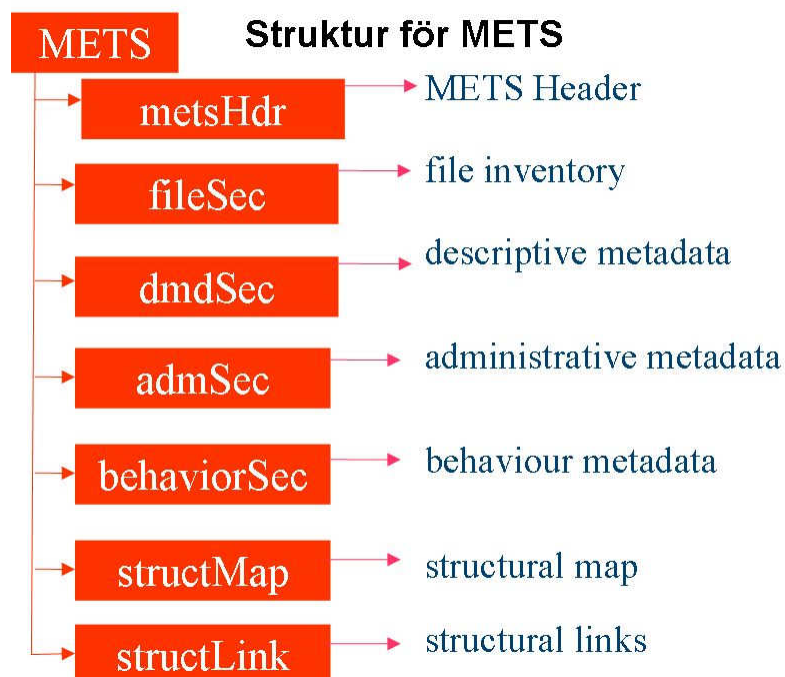
METS

Introduktion

Metadata Encoding and Transmission Standard (METS) är en struktur för att koda och packa ihop metadata för ett digitalt objekt. METS är en flexibel standard som använder XML för att koda upp strukturen. Detta gör att standarden är oberoende av ett visst programmeringsspråk eller operativsystem. XML är läsbart av de flesta program vilket gör att man även för läsning är oberoende. Man måste däremot ha en programvara som kan hantera XML om man vill redigera dessa filer. Exempel på dessa programvaror är XML Spy, oXygen.

Dokumentets struktur

Ett METS-dokument består av sju sektioner.



METS-Header: I denna sektion sparas information om själva METS-filen. Man kan spara ett id för METS-filen, datum för skapande, datum för modifiering, status på METS-filen. Man kan spara information om agenter det vill säga information om aktörer/roller som har varit delaktiga i skapandet av filen, samt att man kan spara ett alternativt objektid för METS-filen som är skilt ifrån det id som är sparad i rotelementet.

File Section: Denna sektion innehåller alla elektroniska versioner som bildar det digitala biblioteks objektet. I sektionen har man element av typen fileGrp som grupperar element av typen file eller separata fileobjekt. Man kan antingen ha en pekare på en extern fil eller bädda in filen i elementet.

Descriptive Metadata: Denna sektion innehåller metadata som gäller beskrivning av arkivbildare eller arkivet. Sektionen kan peka på metadata som finns externt (exempelvis en EAD) eller ha informationen inbäddad i sektionen. Alternativt kan man ha båda varianterna. Man kan ha flera stycken beskrivande metadata som sparas varsitt eget element dmdSec där varje element får ett eget id så att man kan referera till dem i andra delar av METS-filen.

Administrative Metadata: Denna sektion hanterar de tekniska och administrativa metadata som avser de digitala objekten (filer) samt källmaterialet för att skapa objekten. I sektionen finns det fyra delar av administrativt metadata:

Teknisk Metadata. Information om skapandet av filerna deras format och en beskrivning av deras användning.

Intellectual Property Rights (IPR) Metadata. Information om copyright och licens.

Source Metadata. Beskrivande och administrativt metadata om den analoga källan som de digitala objekten härstammar ifrån.

Digital Provenance Metadata. Information gällande källa/mål förhållanden mellan filer, inkluderande ursprungs-/härledningsförhållanden mellan filer och information gällande migrering/transformering som skett på filer från ursprungsartefakt till den nuvarande formen av ett digitalt objekt.

Sektionen är uppbyggd precis som Descriptive Metadata och använder externa eller interna metadata och man kan ha flera stycken administrativa metadata som sparas ett eget element amdSec där varje element får ett eget id så att man kan referera till dem i andra delar av METS-filen.

Behaviour Section: Denna sektion används för att ge information om hur komponenter av det digitala objektet ska renderas för användaren. Man kan spara information om vilka specifika mjukvaruversioner som ska användas eller om det finns speciella parametrar som ska användas vid exekvering av komponenten.

Structural Map: Detta är METS-dokumentets hjärta. Denna sektion går mest att jämföra med hur en bok i pappersformat ser ut. Man bygger upp kapitel med hjälp av div-element och i dessa skapar man referenserna till filer, filgrupper, beskrivande metadata och administrativ metadata så att man får en hierarkisk struktur som kan presenteras för användare av det digitala arkivpaketet.

Structural Links: Denna sektion används till att skapa pekare/hyperlänkar inom objekt i Structural Map sektion. Är användbart till exempel när man ska arkivera en webbsida.

METS-profiler

Vad är då det som kallas för en METS-profil? Jo, det man skapar är en beskrivning i en XML-fil som talar om hur man kommer att använda originalschemat för METS och det är detta som kallas för profil. Man talar om vilken XML-fil med begränsningar och utökningar av ett METS-schema som man använder när man säger att man använder en METS-profil. Att skapa en egen METS-profil innebär att man gör en XML-fil som bygger på originalprofilen och i denna talar man om vilka begränsningar och utökningar som man ger sin profil. För att uppnå detta sätt att hantera METS så har man skapat två scheman. Dels ett schema som visar hur METS-profilen ser ut och ett schema som beskriver hur ett METS-dokument ser ut. I schemat som beskriver METS-dokumentet har man i headern lagt till ett attribut PROFILE och där anger man vilken XML-fil med begränsningar som man hänvisar till för att göra de undantag/utökningar i METS-schemat som man gör i sitt METS-dokument.

Basprofiler

Man kan använda basprofilerna utan att göra några förändringar. Vill man komplettera med fler regler så gör man en egen tilläggsprofil. Denna nya profil hänvisar till basprofilen och innehåller dessutom de egna regler man lägger till. Det handlar om att man till exempel kompletterar ordlistor eller bestämmer hur id ska namnges. En regel som finns i basprofilen behöver inte upprepas i tilläggsprofilen. Till basprofilerna har det anpassats scheman som

validerar enligt de regler som finns i dem. För sin egen tilläggsprofil kopierar man dessa scheman och kompletterar med de regler som man vill skall valideras.

Yttre och inre METS

För att underlätta lagring och kvalitetssäkra transporter är det önskvärt att paketera samman filer med hjälp av till exempel tar- eller zipformat. I paketet ingår en METS-fil (inre) innehållande referenser till filer och deras checksummor. Även paketen ska förses med checksummor. För att åstadkomma detta används en yttre METS-fil innehållande checksumma för paketet som helhet samt en checksumma för den i paketet ingående inre METS-filen. Båda dessa METS-dokument följer samma profil och samma schema.

Namngivning av profilerna

Vi har valt att kalla profilerna för Swedish Information Package (SWEIP) respektive Swedish Information Package B (SWEIPB).

Användandet av element i basprofilerna

Detta är en översikt i tabellform av alla de attribut och element som förekommer i METS-profilerna SWEIP och SWEIPB. Med kursiv text med inledning ”Riksarkivet:” beskrivs i tabellen de diskussioner som förs vid Riksarkivet.

<METS> element grupp:

Element/attribut	Förekomster	Profilregler
<METS>	1 och endast 1 (M)	Måste innehålla attributen OBJID, PROFILE och TYPE.
LABEL	0 eller 1	Innehåller en beskrivning av paketet. Attributet är inte obligatoriskt men det rekommenderas att attributet används för att ge en beskrivning i klartext av paketets innehåll.
OBJID	1 och endast 1 (M)	Identitetskod för paketet. Sätts av den organisation som skapar paketet.
PROFILE	1 och endast 1 (M)	URL till den plats där profilen finns Ex: http://xml.ra.se/METS/SWEIP.xml
TYPE	1 och endast 1 (M)	Detta attribut har ingen värdelista enligt standarden METS 1.9. När denna profil följs skall dock följande fasta värdelista användas: SIP, AIP, DIP

<metsHdr> element grupp:

Element/attribut	Förekomster	Profilregler
<metsHdr>	1 och endast 1 (M)	Måste innehålla attributet CREATEDATE, elementet <metsDocumentID> samt ett eller flera <agent>-element.
CREATEDATE	1 och endast 1 (M)	Tid när METS-filen skapats.
<metsDocumentID>	1 och endast 1 (M)	Här rekommenderas att XML-filens filnamn används. Filnamnet får inte innehålla svenska tecken, blanktecken eller specialtecken.
<agent>	2 eller flera. (M)	Se definition av <agent>-elementet nedan.

Element/attribut	Förekomster	Profilregler
		<i>Två eller flera <agent>-element skall användas för att ge information om den organisation som skapat paketet.</i>

<agent> element grupp:

Element/attribut	Förekomster	Profilregler
<agent>	2 eller flera (M)	Måste innehålla attributen ROLE och TYPE samt elementet <name>. <i>För kompletterande information se stycket om agenter.</i>
ROLE	1 och endast 1 (M)	Värden i detta attribut anges enligt den värdelista som anges i standarden METS 1.9. Se nedan: ARCHIVIST CREATOR CUSTODIAN DISSEMINATOR EDITOR IPOWNER PRESERVATION OTHER
TYPE	1 och endast 1 (M)	Värden i detta attribut anges enligt den värdelista som anges i standarden METS 1.9. Se nedan: INDIVIDUAL ORGANIZATION OTHER
OTHERTYPE	0 eller 1	Om attributet TYPE har värdet "OTHER" så skall attributet OTHERTYPE finnas med. Detta attribut har ingen värdelista enligt METS 1.9. När denna profil följs skall dock värdelistan nedan användas. Listan har endast ett värde: SOFTWARE Detta värde används för att ange information om de mjukvaror som genererat paketet.
<name>	1 och endast 1 (M)	Namn på agenten.
<note>	0 eller flera.	Information om agenten.

<dmdSec> element grupp:

Element/attribut	Förekomster	Profilregler
<dmdSec>	0 eller flera.	Elementet <dmdSec> är inte obligatoriskt. <i>Riksarkivet: Om en statlig myndighet skapar SIP:ar eller AIP:er för långtidsbevarande så vore det möjligt att använda ett <dmdSec>-element för EAD-information och ett <dmdSec>-element för EAC-information.</i>

ID	1 och endast 1 (M)	Unik identitet för elementet inom METS-filen.
<mdRef> eller <mdWrap>		Metadata kan hanteras på två sätt i METS. Antingen ligger metadata i egna filer. Dessa refereras då via elementet <mdRef>. Metadata i XML-form kan också inkluderas i själva METS-filen. Då kapslas metadata in av elementet <mdWrap>. Se nedan för definition av elementen <mdRef> och <mdWrap>

<amdSec> element grupp:

Element/attribut	Förekomster	Profilregler
<amdSec>	0 eller flera.	Elementet <amdSec> är inte obligatoriskt. <i>Riksarkivet: Om en statlig myndighet skapar SIP:ar för leverans till Riksarkivet så kan ett eller flera <amdSec>-element användas för att referera till erforderlig metadata.</i>
ID	1 och endast 1 (M)	Unik identitet för elementet inom METS-filen.
<techMD>	0 eller flera.	Elementet <techMD> är inte obligatoriskt. <i>Riksarkivet: Om en statlig myndighet skapar SIP:ar för leverans till Riksarkivet så kan ett <amdSec>-element innehålla ett <techMD>-element som refererar till metadata enligt ADDML.</i>
ID i <techMD>	1 och endast 1 (M)	Unik identitet för elementet inom METS-filen.
<mdRef> eller <mdWrap> i elementet <techMD>		Metadata kan hanteras på två sätt i METS. Antingen ligger metadata i egna filer. Dessa refereras då via elementet <mdRef>. Metadata i XML-form kan också inkluderas i själva METS-filen. Då kapslas metadata in av elementet <mdWrap>. Se nedan för definition av elementen <mdRef> och <mdWrap>
<digiprovMD>	0 eller flera.	Elementet <digiprovMD> är inte obligatoriskt. Om bevarandeinformation (digital provenance information) skall registreras så rekommenderas att PREMIS-standardens används för detta syfte. <i>Riksarkivet: Vid leverans till Riksarkivet bör det vara tillåtet men inte obligatoriskt att leverera en <digiprovMD> som refererar till metadata i PREMIS-format.</i>
ID i <digiprovMD>	1 och endast 1 (M)	Unik identitet för elementet inom METS-filen.
<mdRef> eller <mdWrap> i elementet <digiprovMD>		Metadata kan hanteras på två sätt i METS. Antingen ligger metadata i egna filer. Dessa refereras då via elementet <mdRef>. Metadata i XML-form kan också inkluderas i själva METS-filen. Då kapslas metadata in av elementet <mdWrap>. Se nedan för definition av elementen <mdRef> och <mdWrap>

<mdRef> element grupp:

Element/attribut	Förekomster	Profilregler
<mdRef>	0 eller 1	Elementet är inte obligatoriskt (se ovan) men om elementet <mdRef> skall användas så måste det innehålla attributen CREATED, MDTYPE, ID, SIZE, MIMETYPE, CHECKSUMTYPE, CHECKSUM, LOCTYPE, xlink:href och xlink:type.
CREATED	1 och endast 1 (M)	Tid när den refererade filen skapats.
MDTYPE	1 och endast 1 (M)	Värden i detta attribut anges enligt den värdelista som anges i standarden METS 1.9. Se nedan: MARC , MODS, EAD, DC, NISOIMG, LC-AV, VRA, TEIHDR, DDI, FGDC, LOM, PREMIS, PREMIS:OBJECT, PREMIS:AGENT, PREMIS:RIGHTS, PREMIS:EVENT, TEXTMD, METSRIGHTS, ISO 19115:2003 NAP, OTHER.
OTHERMDTYPE	0 eller 1	Om attributet MDTYPE innehåller värdet "OTHER" så måste det finnas ett attribut OTHERMDTYPE. Detta attribut har ingen värdelista enligt METS 1.9. När denna profil följs skall dock värdelistan nedan användas: ADDML, EAC-F, EAC-CPF, EAG, METS
ID	1 och endast 1 (M)	Unik identitet för elementet inom METS-filen.
SIZE	1 och endast 1 (M)	Filstorlek för den refererade metadatafilen
MIMETYPE	1 och endast 1 (M)	MIME-typ för den refererade metadatafilen
CHECKSUMTYPE	1 och endast 1 (M)	Typ av checksummealgorithm. Anges enligt attributets värdelista i METS 1.9. Se nedan: Adler-32, CRC32, HAVAL,

Element/attribut	Förekomster	Profilregler
		MD5, MNP, SHA-1, SHA-256, SHA-384, SHA-512, TIGER, WHIRLPOOL <i>Checksummor av typerna MD5 eller SHA-X skall användas.</i>
CHECKSUM	1 och endast 1 (M)	Checksumman för den angivna algoritmen.
LOCTYPE	1 och endast 1 (M)	I METS tillåts flera värden i detta attribut men i denna SIP-profil tillåts endast värdet URL. Det förutsätts också att xlink-referensen är av typen "file:" följt av ett filnamn. Alla filer som refereras skall finnas fysiskt i SIP:en.
Xlink:href	1 och endast 1 (M)	Skall innehålla prefix "file:" följt av filnamnet.
Xlink:type	1 och endast 1 (M)	Skall innehålla värdet "simple".

<mdRef> element grupp SWEIPB i dmdSec:

Denna grupp användes när referens sker till exempelvis LIBRIS. Om det inte är en sådan referens skall elementet följa den ordinarie <mdRef>-strukturen.

Element/attribut	Förekomster	Profilregler
<mdRef>	0 eller 1	Elementet är inte obligatoriskt (se ovan) men om elementet <mdRef> skall användas så måste det innehålla attributen MDTYPE, ID, MIMETYPE, LOCTYPE, xlink:href och xlink:type.
MDTYPE	1 och endast 1 (M)	Värden i detta attribut anges enligt den värdelista som anges i standarden METS 1.9. Se nedan: MARC , MODS, EAD, DC, NISOIMG, LC-AV, VRA, TEIHDR, DDI, FGDC, LOM, PREMIS, PREMIS:OBJECT, PREMIS:AGENT, PREMIS:RIGHTS, PREMIS:EVENT, TEXTMD,

Element/attribut	Förekomster	Profilregler
		METSRIGHTS, ISO 19115:2003 NAP, OTHER.
OTHERMDTYPE	0 eller 1	Om attributet MDTYPE innehåller värdet "OTHER" så måste det finnas ett attribut OTHERMDTYPE. Detta attribut har ingen värdelista enligt METS 1.9. När denna profil följs skall dock värdelistan nedan användas: ADDML, EAC-F, EAC-CPF, EAG, METS
ID	1 och endast 1 (M)	Unik identitet för elementet inom METS-filen.
MIMETYPE	1 och endast 1 (M)	MIME-typ för den refererade metadatafilen
LOCTYPE	1 och endast 1 (M)	I METS tillåts flera värden i detta attribut men i denna SIP-profil tillåts endast värdet URL.
Xlink:href	1 och endast 1 (M)	Skall innehålla prefix "file:" följt av filnamnet när en fil refereras. När det är en länk används http:
Xlink:type	1 och endast 1 (M)	Skall innehålla värdet "simple".

<mdWrap> element grupp:

Element/attribut	Förekomster	Profilregler
<mdWrap>	0 eller 1	Elementet är inte obligatoriskt (se ovan) men om elementet <mdWrap> skall användas så måste det innehålla attributen MDTYPE samt antingen elementet <xmlData> eller <binData>.
MDTYPE	1 och endast 1 (M)	Värden i detta attribut anges enligt den värdelista som anges i standarden METS 1.9. Se nedan: MARC, MODS, EAD, DC, NISOIMG, LC-AV, VRA, TEIHDR, DDI, FGDC, LOM, PREMIS, PREMIS:OBJECT, PREMIS:AGENT, PREMIS:RIGHTS, PREMIS:EVENT, TEXTMD, METSRIGHTS, ISO 19115:2003 NAP,

Element/attribut	Förekomster	Profilregler
		OTHER.
OTHERMDTYPE	0 eller 1	Om attributet MDTYPE innehåller värdet "OTHER" så måste det finnas ett attribut OTHERMDTYPE. Detta attribut har ingen värdelista enligt METS 1.9. När denna profil följs skall dock värdelistan nedan användas: ADDML, EAC-F, EAC-CPF, EAG, METS
<binData>	0 eller 1	Detta element används för att kapsla in metadata i binär form. Profilen tillåter detta men det rekommenderas inte.
<xmlData>	0 eller 1	Detta element används för att kapsla in metadata i XML-form. Det rekommenderas att metadata anges i XML-form.

<fileSec> element grupp:

Element/attribut	Förekomster	Profilregler
<fileSec>	1 och endast 1 (M)	METS-filen måste alltid innehålla ett <fileSec>-element. Detta måste innehålla minst ett <fileGrp>-element. Dessa <fileGrp>-element innehåller i sin tur <file>-elementen. Se nedan.
<fileGrp>	1 eller flera (M)	All datafiler måste underordnas en filgrupp.
USE	0 eller 1	Attributet USE är inte obligatoriskt i denna profil men det rekommenderas starkt att attributet används. Det är upp till varje organisation att ange värden och definiera sitt användande av detta attribut.
<file>		Se <file> elementet nedan

<file> element grupp:

Element/attribut	Förekomster	Profilregler
<file>	0 eller flera	Måste innehålla attributen ID, MIMETYPE, CREATED, CHECKSUM, CHECKSUMTYPE och SIZE samt elementet <Flocat> vilket innehåller attributen LOCTYPE, xlink:href och xlink:type.
ID	1 och endast 1 (M)	Unik identitet för filelementet inom METS-filen.
MIMETYPE	1 och endast 1 (M)	MIME-typ för den refererade filen
CREATED	1 och endast 1 (M)	Tid när den refererade filen skapats.
CHECKSUM	1 och endast 1 (M)	Checksumma enligt angiven algoritm.
CHECKSUMTYPE	1 och endast 1 (M)	Typ av checksummealgoritm. Anges enligt attributets värdelista i METS 1.9. Se nedan: Adler-32, CRC32,

Element/attribut	Förekomster	Profilregler
		HAVAL, MD5, MNP, SHA-1, SHA-256, SHA-384, SHA-512, TIGER, WHIRLPOOL <i>Checksummor av typerna MD5 eller SHA-X skall användas.</i>
USE	0 eller 1	Attributet USE är inte obligatoriskt i denna profil men det rekommenderas starkt att attributet används. Det är upp till varje organisation att ange värden och definiera sitt användande av detta attribut.
SIZE	1 och endast 1 (M)	Filstorlek är obligatorisk
ADMID	0 eller 1	Refererar till ett eller flera metadataelement under <amdSec>.
DMDID	0 eller 1	Refererar till ett eller flera metadataelement av typen <dmdSec>.
<Flocat>	1 och endast 1 (M)	Måste innehålla attributen LOCTYPE, xlink:href och xlink:type.
LOCTYPE	1 och endast 1 (M)	I METS tillåts flera värden i detta attribut men i denna SIP-profil tillåts endast värdet URL. Det förutsätts också att xlink-referensen är av typen "file:" följt av ett filnamn. Alla filer som refereras skall finnas fysiskt i SIP:en.
Xlink:href	1 och endast 1 (M)	Innehåller prefix "file:" följt av filnamnet.
Xlink:type	1 och endast 1 (M)	Skall innehålla värdet "simple".

<structMap> element grupp:

Element/attribut	Förekomster	Profilregler
<structMap>	1 eller flera	En METS-fil måste innehålla minst ett <structMap>-element. Detta måste innehålla minst ett <div>-element. Det är upp till varje organisation att definiera innehållet i <structMap> och underordnade element.
LABEL	0 eller 1	Innehåller en beskrivning i klartext om <structMap>-elementet. Attributet är inte obligatoriskt men det rekommenderas att attributet används.
<div>	1 eller flera	Se elementet <div> nedan

<div> element grupp:

Element/attribut	Förekomster	Profilregler
<div>	1 eller flera	Det är upp till varje organisation att definiera innehåll i attribut och element under <div>.
LABEL	0 eller 1	Innehåller en beskrivning i klartext om <div>-elementet. Attributet är inte obligatoriskt men det rekommenderas att attributet används.
ADMID	0 eller 1	Refererar till ett eller flera metadataelement under <amdSec>.
DMDID	0 eller 1	Refererar till ett eller flera metadataelement av typen <dmdSec>.
<fptr>	0 eller flera	Filpekare till en fil i <fileSec>
FILEID	1 och endast 1 (M)	Om elementet <div> innehåller ett element av typen <fptr> så måste <fptr>-elementet innehålla en referens till ett existerande ID för en fil i <fileSec>
<div>	0 eller flera	Ett <div> element kan underindelas i godtyckligt antal <div>-element.

Agenter

Detta element ger information om vilka som är ansvariga för METS-dokumentet. En agent kan vara både en människa, en organisation eller ett program. I grundprofilerna finns det två obligatoriska agenter. Se nedan för ett exempel. Ytterligare agenter kan komma att behövas. Men, respektive användare av profilerna behöver skapa regler för dessa ytterligare agenter. Agent elementet innehåller två under element. I name-elementet skrivs namnet på agenten. I note-elementet (som är upprepningsbart) kan kompletterande information ges så som till exempel identifieringskod.

Vår tanke är att den första obligatoriska agenten med ROLE="ARCHIVIST" och TYPE="ORGANIZATION" beskriver arkivbildaren. Den andra obligatoriska agenten med ROLE="CREATOR" och TYPE="ORGANIZATION" beskriver den enhet eller organisation som skapat och levererat METS-dokumentet.

Vi förslår även ett ytterligare antal agenter enligt nedanstående.

- En agent med ROLE="CREATOR" och TYPE="INDIVIDUAL" som är den man kan kontakta vid problem.
- En agent med ROLE="CREATOR", TYPE="OTHER" och OTHERTYPE="SOFTWARE" som beskriver en programvara. Antingen programmet som skapat materialet eller programmet som deltagit i paketeringen.
- En agent med = "PRESERVATION" och TYPE="ORGANIZATION" som beskriver det arkiv eller den arkivfunktion som är ansvarig för bevarandet.
- En agent med ROLE="PRESERVATION", TYPE="OTHER" och OTHERTYPE="SOFTWARE" som beskriver ett program som skall utföra något med paketet, exempelvis ett mottagningsprogram. I note-elementen ges styrvariabler som programmet behöver.

Obligatoriska agenter:

```

<mets:agent ROLE="ARCHIVIST" TYPE="ORGANIZATION">
  <mets:name>Arkivbildare</mets:name>
</mets:agent>
<mets:agent ROLE="CREATOR" TYPE="ORGANIZATION">
  <mets:name>Leverande enhet eller organisation</mets:name>
</mets:agent>

```

Förslag på ytterligare agenter:

```

<mets:agent ROLE="CREATOR" TYPE="INDIVIDUAL">
  <mets:name>Kontaktperson</mets:name>
  <mets:note>Telefonnummer och övriga kontaktuppgifter</mets:note>
</mets:agent>
<mets:agent ROLE="CREATOR" TYPE="OTHER" OTHERTYPE="SOFTWARE">
  <mets:name>Paketeringsapplikationen X</mets:name>
</mets:agent>
<mets:agent ROLE="PRESERVATION" TYPE="ORGANIZATION">
  <mets:name>Mottagande arkiv</mets:name>
</mets:agent>
<mets:agent ROLE="PRESERVATION" TYPE="OTHER" OTHERTYPE="SOFTWARE">
  <mets:name>Arkivets uppkningsprogram</mets:name>
  <mets:note>Eventuella styrvariabler för uppkningsprogrammet</mets:note>
  <mets:note>Eventuella ytterligare styrvariabler</mets:note>
</mets:agent>

```

Övriga standarder.

I METS sektioner för beskrivande- och administrativt metadata kan man infoga eller peka på metadata som beskrivs med andra standarder. Vi har inte haft möjlighet att detalj granska och specificera dessa men ger här en beskrivning av vad som kan användas.

Beskrivande metadata

EAC-CPF

Vid utbyte av information om arkivbildare, som kan vara en myndighet, en organisation, ett företag eller en person, använder man XML-standarderna EAC-CPF (Encoded Archival Context - Corporate bodies, Persons, Families). Standarderna upprätthålls av Society of American Archivists tillsammans med Berlin State Library. EAC-CPF bygger på den allmänna standarden ISAAR(CPF).

EAD

Vid utbyte av information om arkivets innehåll, i princip innehållet i en arkivförteckning, använder man XML-standarderna EAD (Encoded Archival Description). Standarderna upprätthålls av Library of Congress tillsammans med Society of American Archivists. EAD bygger på den allmänna standarden ISAD(G).

EAG

EAG är en förkortning av 'Encoded Archival Guide'. Den tillhandahåller ett elektroniskt format för lagring, publicering och utbyte av information om arkivens samlingar. EAG upprättades av det spanska ministeriet för kultur för att märka upp alla 42000 spanska och iberisk-amerikanska arkivsamlingar. Numera är den en de facto-standard.

DC

DC (Dublin Core) från Dublin Core Metadata Initiative (DCMI) är tänkt att vara enkelt att använda, ska kunna appliceras på alla typer av media med syfte att förbättra åtkomst av digitala resurser på Internet. DC har kommit att bli den standard som ofta bestämmer miniminivån för utbyte av metadata mellan olika domäner. Det finns två uppsättningar DC metadata: *Dublin Core Element Set* som också är internationell standard (ISO 15836:2009), samt *DCMI Metadata Terms* som inkluderar DC Element Set men är utbyggd med fler termer.

MARCXML

Acronymen MARC står för MACHine-Readable Cataloging och har sitt ursprung i ett format som utvecklades för 40 år sedan. Syftet var att möjliggöra för datorer att utbyta, använda och tolka bibliografisk information. Dess dataelement har sedan utgjort grunden för bibliotekens kataloginformation ända fram till idag. MARC21 är det format som idag används av de flesta bibliotekssystem. MARCXML är helt enkelt MARC21 överfört till xml. Underhålls av Network Development and MARC Standards Office vid Library of Congress.

MODS

MODS (Metadata Object Description Schema) är ett xml-schema för bibliografiska metadata som kan användas för att beskriva många olika typer av objekt. Det är framförallt inom biblioteken man träffar på MODS som där ses som "enklare än MARC" men "rikare än DC".

Det används ofta som ett ”utbytesformat” för valda delar av katalogposter i MARC21, men även för att skapa nya beskrivningar av objekt i bibliotekens samlingar. Network Development and MARC Standards Office vid Library of Congress ansvarar för underhållet.

Administrativt metadata

TextMD

TextMD (Technical Metadata for Text) är ett xml-schema i vilket man kan ange tekniska metadata för textbaserade digitala objekt, till exempel OCR-lästa bilder från en digitaliserad bok. Används ofta som tillägg till administrativa metadata inom standarden METS, men kan även användas inom PREMIS. TextMD kan till exempel innehålla information om vilken teknisk plattform eller mjukvara som använts vid OCR-läsningen, om vilken teckenuppsättning som förekommer, språk, uppmärkningsspråk, sidföljd och mycket mer. Network Development and MARC Standards Office vid Library of Congress ansvarar för underhållet.

PREMIS

PREMIS står för Preservation Metadata: Implementation Strategies, senaste version är 2.0. Det är en standard framtagen för att effektivt hantera, finna och återskapa den digitala informationen. PREMIS bevarandemetadata innehåller proveniensinformation, fångar aktiviteter kring digitala objektet samt identifierar tekniska egenskaper. Den ger även autenticitet till objektet.

PREMIS datamodell är uppbyggd av fem entiteter, Intellectual entites, Objects, Event, Agents samt Rights. Varje digital objekts grunddata fångas i Object, en händelse (Event) som har påverkat objektet samt vem (Agents) och eventuella restriktioner (Rights) som har satts på objektet fångas och nertecknas. Library of Congress tillhandahåller scheman i XML för detta, det finns även en svensk version som Riksarkivet ansvarar för.

MIX

NISO MIX = NISO metadata for Images in XML

Mix är en standardisering av teknisk metadata för att hantera samlingar av digitala (raster)bilder. Det är ett XML schema som utvecklats för att ge instruktioner om hur man kan/får/skall bevara, bearbeta och skicka digitala bilder i och mellan olika system över tid.

ADDML

Archives Data Description Markup Language (ADDML) är ett XML-schema avsett att beskriva elektroniska arkiv. Schemat är utarbetat av Norska Riksarkivet och används i dag av riksarkiven i Norge och Sverige. Aktuell version är version 8.2 från januari 2009. En ADDML-fil kan beskriva en grupp av filer, ett ”datasett” med norsk terminologi. Vid svenska riksarkivet kommer en ADDML-fil att innehålla referenser till samtliga filer ingående i ett arkivinformativpaket (AIP). ADDML kommer att användas för born-digital-AIP:er, dvs AIP:er som innehåller filer levererade till Riksarkivet från andra myndigheter. ADDML kommer inte att användas för de AIP:er som skapas som ett resultat av Riksarkivets egen skanningsverksamhet.

ADDML kan användas för teknisk, strukturell och allmänt beskrivande metadata för levererade filer. ADDML innehåller en rad element och attribut som kan användas för att beskriva filer med namn, filtyp, filstorlek etc. Det finns också en rad beskrivande element både för enskilda filer och för paketet (datasettet) som helhet. Ett av syftena med ADDML har

också varit att kunna registrera och bevara postbeskrivningar för flata filer som exporterats ur databaser. ADDML kan beskriva tabeller, poster och fält samt primärnycklar och främmande nycklar. Det innebär att ADDML kan användas för att fullständigt beskriva filer som exporterats från databaser och att ADDML också kan användas för att återskapa databaser från bevarade flata filer. Förutom en rad fasta element och attribut innehåller ADDML också ett par generiska element vilket möjliggör att schemat kan anpassas för att härbärgera ytterligare information specifik för en viss arkivmyndighet.

Audio och video

Det finns idag inga färdiga standarder som kan hantera metadata kring audio och video. Library of Congress har två enkla scheman som inte är publicerade. Audio Engineering Society (AES) håller på och tar fram standarder.

Kombinationsformat

Nedanstående format som beskrivs innehåller både administrativt data och metadata. Varje organisation måste bestämma var dessa ska placeras i METS-dokumentet.

Ekonomi

För ekonomi finns det idag de scheman som eXtensible Business Reporting Language (XBRL) tagit fram. Dessa innehåller dock ingen information om anläggningstillgångar. Svenska SIE-gruppen har tagit fram ett format för utbyte av information mellan ekonomisystem och det stöds av de flesta systemen på den svenska marknaden. Det är ett flatfilsuttag. Under 2005 experimenterade man med ett utbytesformat i XML men detta arbete har tyvärr avstannat. Riksarkivet och LDB-projektet gjorde ett test av formatet och fann att med ett antal justeringar och komplimenteringar så skulle det kunna uppfylla kraven för leverans av myndigheters ekonomidata till Riksarkivet. Försöket har dock inte blivit helt utvärderat men materialet finns tillgängligt som utkast.

MoReq

MoReq är en förkortning av 'Model Requirements for the management of electronic records'. Den är en modellspecifikation och innehåller allmänna krav för system för hantering av elektroniska register. Specifikationen har utvecklats av Europeiska kommissionen på begäran av DLM-forum, en oberoende europeisk sammanslutning av arkiv och andra offentliga och privata organisationer som är intresserade av arkivering och förvaltning av arkiv och information.

MoReq2 är avsedd för användare och leverantörer av elektroniska ärendehanterings- eller diariesystem och tjänster. Dess huvudsakliga syfte är att tjäna som underlag för anbudsförfarande (även känt som en begäran om förslag) och som ett regelverk för att styra utvecklingen av nya elektroniska ärendehanterings- eller diariesystem. Den kan också användas för att ge en grund för revision av befintliga ärendehanterings- eller diariesystem eller tjänster, och som en resurs för akademiska eller kommersiella utbildare.

MoReq2 publicerades år 2008 och är en uppdaterad och utökad version av de tidigare standardkraven. Den innehåller en basmodul med de grundläggande kraven som är nödvändiga för en trovärdig hantering av elektroniska register. Valfria moduler kan läggas till om ytterligare funktioner önskas.

Genom MoReq2 infördes också det s.k. 'kapitel noll'. Syftet med detta är att varje land ska kunna lägga till sina nationella krav och därigenom ta hänsyn till olika språk, lagstiftning, föreskrifter och traditioner för arkivering.

NOARK

Är en norsk standard för elektroniska ärendehanteringssystem. Sedan 1984 har NOARK (Norsk Arkivsystem) varit de facto standard för norska statliga och kommunala myndigheter. Medan NOARK 1-3 var en standard för register (diarier) var NOARK 4 även standard för digital arkivering av dokument. Med NOARK 5 är processen med att få kompatibiliteten med Moreq 2 startat.

NOARK 5 består av 3 skikt: 1 Inre kärna, 2 Yttre kärna, och 3 NOARK komplett. Den inre kärna innehåller krav på grundläggande funktionalitet för ärendehantering medan den yttre kärnan innehåller krav på externa moduler.

SIARD

SIARD (Software Independent Archival of Relational Databases) är en icke kommersiell, öppen standard för långtidsbevarande av relationsdatabaser. Standarden är utvecklad av det Schweiziska Federala Arkivet (SFA) inom ramen för "ARELDA"-projektet (Archiving of Electronic Data).

SIARD baseras på ISO-standarder: Unicode, XML, SQL 1999 och ZIP. 2008 antogs SIARD som ett gällande arkivformat för arkivering av relationsdatabaser av det gemensamma europeiska initiativet "PLANETS"- Preservation and Long term Access through Networked Services.

SIARD hanterar en databas som ett enskilt dokument (handling) och sparar den som en fil. Filen finns som ett icke komprimerat ZIP 6 arkiv. Arkivet innehåller en uppsättning av XML-filer över databasens innehåll och, om så påkallat, binära filer från BLOB (Binary Large Objects) och CLOB (Character Large Objects) men med referenser i XML; allt i en speciell mappstruktur.

Alla data presenteras, i princip, i Unicode 7.

En relationsdatabas består av en mängd information (objekt) som är strukturerad på ett visst sätt, dvs. av olika scheman, översikter ("standard queries"), regler ("procedures") och tabeller. En databas i SIARD-formatet kommer att bestå av metadata och data (databasinnehåll). Metadata beskriver databasstruktur och data representerar tabellernas innehåll. Genom metadata är det också möjligt att ta reda på var data finns arkiverade.

Metadata och data är sparade tillsammans som ett icke komprimerat ZIP-arkiv med filnamnet *.siard*.

Metadata finns i mappen "header" och innehållet/data i mappen "content".

Siard-filens struktur ser ut på följande sätt:

header

metadata.xsd (XML-Schema-Definition)
metadata.xml

content

schema1
table1
table.xsd
table.xml
lob1 (Large Object)

```

        record1.txt / record1.bin
    lob2
        record1.txt / record1.bin
    ...
table2
    table.xsd
    table.xml
    ...
schema2
    ...

```

SIARD är utformad oberoende av de informationspaket som finns i OAIS-modellen och gäller enbart för relationsdatabaser. Formatet kan ingå i en AIP som en del av denna.

Geografisk information

Geografisk information är ett samlingsbegrepp för all data som kan ges en rumslig adressering och därtill hörande geografiska referenssystem (koordinatsystem). Man brukar använda begreppet GIS (geografiskt informationssystem) för att därmed också inkludera metoder, verktyg för insamling, bearbetning, analys och presentation av geografisk information. Det finns en mängd platser på nätet inom detta område men två som kan vara utgångspunkter för vidare utforskande är: www.geodata.se och www.opengeospatial.org.

Det finns mycket samverkan kring frågor rörande geografisk information och GIS. INSPIRE-direktivet (Infrastructure for Spatial Information in Europe) bindande bestämmelser för att inrätta en infrastruktur för geodata medlemsländerna i EU.

Ett av de områden som ingår i direktivet avser metadata.

Geodata.se är en informationssida och gemensam webbplats för [EG-direktivet Inspire](#), den [nationella geodatastrategin](#) och [Geodataprojektet](#). Här hittar du allmän information och dokument kring geodataområdet.

OpenGis är en open sourcesamverkan kring geografisk information och GIS som även hanterar standarder för metadata.

<http://www.opengeospatial.org/>

Här finns metadata för bl a GML (Geographic Markup Language):

“The OpenGIS® Geography Markup Language Encoding Standard (GML) The Geography Markup Language (GML) is an XML grammar for expressing geographical features. GML serves as a modeling language for geographic systems as well as an open interchange format for geographic transactions on the Internet. As with most XML based grammars, there are two parts to the grammar – the schema that describes the document and the instance document that contains the actual data. A GML document is described using a GML Schema”

KML (Keyhole Markup Language):

“Google submitted KML (formerly Keyhole Markup Language) to the Open Geospatial Consortium (OGC) to be evolved within the OGC consensus process with the following goal: KML Version 2.2 has been adopted as an OGC implementation standard. Future versions may be harmonized with relevant OGC standards that comprise the OGC standards baseline.”

m fl.

Arbetet inom OpenGis täcker in hela kedjan från specificering av metadata till visualisering/publicering som e-tjänster. Det mesta går att ladda ner.

Länkar

ADDML: <http://www.arkivverket.no/standarder/addml>
Basprofil: <http://xml.ra.se/METS>
Dublin Core DC: <http://dublincore.org/>
EAC-CPF: <http://eac.staatsbibliothek-berlin.de/>
EAD: <http://www.archivists.org/saagroups/ead/>
EAG DTD: <http://censoarchivos.mcu.es/CensoGuia/doc/EAG.DTD.txt>
EAG Spansk kulturministeriet: <http://censoarchivos.mcu.es/CensoGuia/proyecto.htm>
EAG tagbeskrivning:
http://censoarchivos.mcu.es/CensoGuia/doc/Repertorio_de_etiquetas_EAG_Alfa_0.2.doc
Ekonomischema, utkast: <http://xml.ra.se/economics/>
Geografiskt data bla: <http://geodata.se>
ISAAR(CPF): <http://www.ica.org/sites/default/files/ISAAR2EN.pdf>
ISAD(G): http://www.ica.org/sites/default/files/isad_g_2e.pdf
ISIL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=52666
MARCXML: <http://www.loc.gov/standards/marcxml/>
MIX: <http://www.loc.gov/standards/mix>
METS: <http://www.loc.gov/standards/mets>
MODS: <http://www.loc.gov/standards/mods/>
MoReq Download: <http://www.moreq2.eu/moreq2>
MoReq Infoblad i olika språk: <http://moreq2.eu/other/leaflets>
MoReq2 webbsida: <http://www.moreq2.eu> och <http://www.moreq.info>
NOARK: <http://www.arkivverket.no/arkivverket/Offentlig-forvaltning/Noark>
Open GIS: <http://www.opengeospatial.org>
PREMIS: <http://www.loc.gov/standards/premis/> och <http://xml.ra.se/PREMIS>
SIARD: <http://www.bar.admin.ch/themen/00772/00811/index.html?lang=de>
SIE gruppen: <http://www.sie.se/default.asp>
TextMD: <http://www.loc.gov/standards/textMD/>
XBRL, scheman: <http://www.xbrl.org/SpecRecommendations/>
XBRL, Sverige: <http://www.xbrl.se>