

Standarder och metadata



Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 31(136)
Projekt eARD DP 4 & 6	Noteringar				

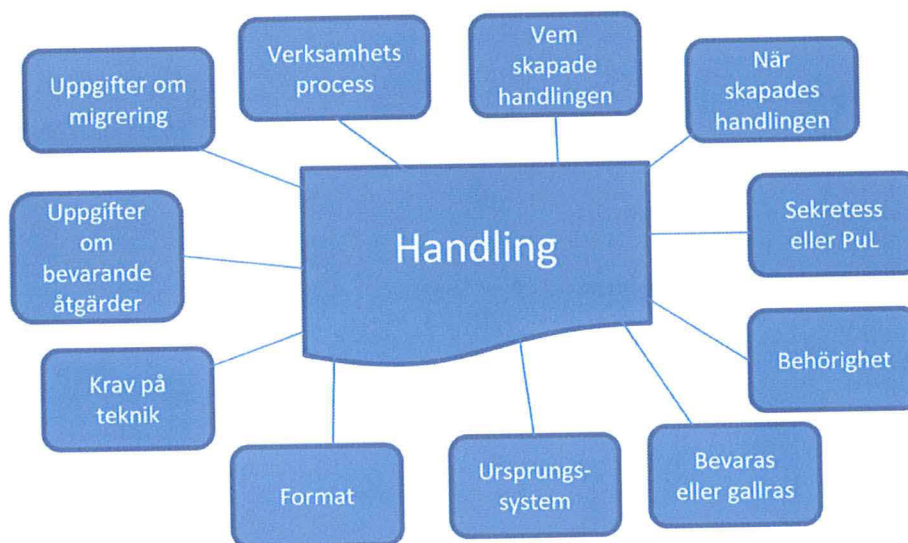
3. METADATA

FGS:erna anger hur information ska beskrivas, vilket görs genom metadata. Metadata har en nyckelroll i hanteringen av digital information, från det att den skapas till den bevaras (Evans, McKemmish & Bhoday, 2005). Det skapar förutsättningar för att kunna hitta, hantera, bevara och förstå informationen över tid (Cunningham, 2001).

Arkivarier har alltid hanterat metadata, t.ex. genom arkivredovisning och andra sökingångar till arkiven. Men i digital hantering ändras förutsättningarna. Till exempel måste vissa uppgifter som tidigare varit inbäddat i handlingarna och dess medium beskrivas separat. Informationen har också skapats under vissa tekniska förutsättningar och kräver kanske viss teknik för att kunna visas, vilket måste beskrivas för att kunna läsas över tid.

3.1 Vad är metadata

Metadata är data som beskriver handlingar och deras hantering. Det avser uppgifter som beskriver innehåll, sammanhang och struktur hos handlingar och deras hantering över tid (ISO 30300). Det gör det möjligt att registrera, söka och hitta, få förståelse för, använda, kontrollera och bevara informationen (International Organisation for Standardization (ISO), 2003). Det kan till exempel vara uppgifter om vem som skapat en handling, när den skapats och spårar användning och eventuella ändringar (QSA, 2012b). Man använder sig också av metadata för att styra åtkomst och behörighet till handlingar och det kan ge information om och när handlingar ska bevaras eller gallras. Sammanfattningsvis kan sägas att metadata är avgörande för möjligheterna att styra informationshanteringen och kontrollera informationen, och att upprätthålla dess sammanhang och bevisvärde (Wallace, 1993). I figur 3.1 nedan illustreras exempel på uppgifter (metadata) som kan behövas för att förstå och hantera en handling.



Figur 3.1: Exempel på metadata som kan tillhöra en handling.

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 32(136)
Projekt eARD DP 4 & 6	Noteringar				

3.2 Metadata och informationskvalitet

Genom att beskriva informationens sammanhang blir den meningsfull och kan användas och tolkas, även över tid. Metadata ger också uppgifter om informationen och dess hantering som krävs för att dess äkthet, tillförlitlighet och integritet ska kunna styrkas (QSA, 2012c, ISO, 2003). För att säkerställa handlingars äkthet måste man t.ex. veta vem som är upphovsman, när informationen skapats, vilken aktivitet handlingen dokumenterar och hur den relaterar till andra handlingar tillhörande samma aktivitet. Man måste också kunna påvisa att informationen inte ändrats otillåtet (International Research on Permanent Authentic Records in Electronic Systems (InterPARES)). Eftersom viss förändring kan ske med informationen, t.ex. vid migreringar, så måste också de aktiviteter som utförs med informationen dokumenteras. Det är också viktigt att upprätthålla sambandet mellan handling och dess metadata (QSA, 2012b).

3.3 Metadatahantering

Det bör finnas strategier, regelverk, rutiner och ansvar för metadatahanteringen på samma sätt som för dokumenthanteringen (ISO, 2003). En stor del av metadatan registreras automatiskt i systemen, varför det blir viktigt att säkerställa att systemen har kapacitet att dokumentera den metadata som behövs (QSA, 2012b). En standardisering av metadata kan bland annat säkerställa att nödvändig information finns med för att garantera informationens kvalitet (QSA, 2012c) och ger en gemensam struktur för återsökning.

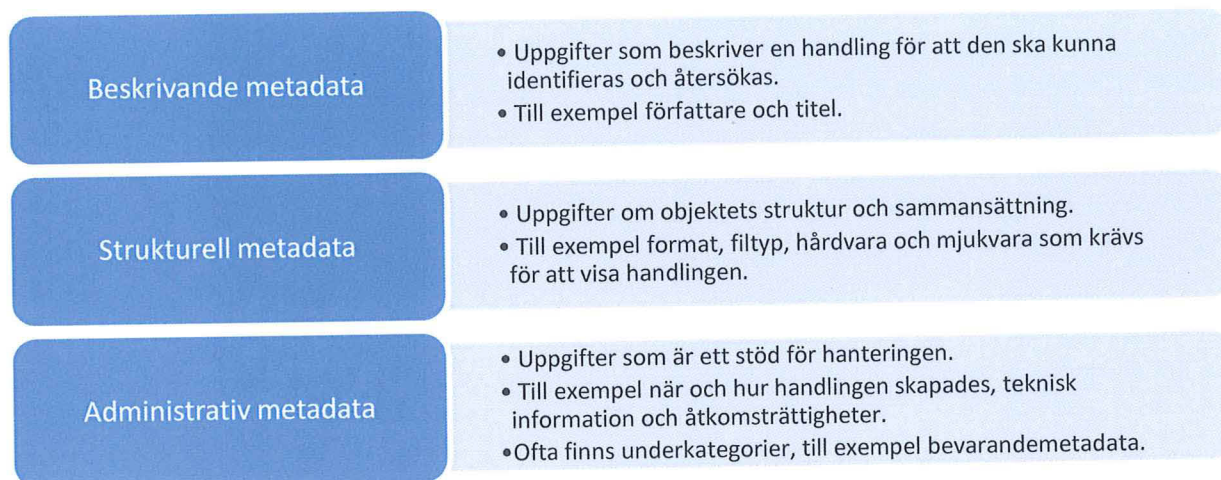
Metadata skapas och dokumenteras löpande i dokumenthanteringen från det att informationen skapas. Det skapas också specifikt vid leverans till e-arkiv och för fortsatt hantering i e-arkiv. Metadata används även i flera andra sammanhang och för olika syften, t.ex. vid publicering på webb och intranät.

3.4 Typer av metadata

För att hantera och bevara handlingar behövs olika typer av metadata. Det behövs t.ex. teknisk metadata som ger uppgifter om den tekniska miljön där handlingen skapats och hanterats, och vad som krävs för att visa handlingen. Det behövs information som berör handlingens struktur och lagring, som t.ex. format och viktiga tekniska beroenden. Det behövs också information om själva handlingen och dess relationer till andra handlingar. Slutligen behövs metadata för sökning och åtkomst och metadata för rättigheter för att definiera vem som har åtkomst och vad man har rätt att göra med handlingen (ISO, 2003). Metadata kan också styrka en handlingens äkthet, exempelvis teknisk metadata i en bild. Gallring eller annan förändring av metadata kan få konsekvenser och bör därför först utredas. En handlingens äkthet är också beroende av dess metadataas kvalitet, vilket tas upp mer under avsnitt om metadatakvalitet, 6.1.2 nedan.

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 33(136)
Projekt eARD DP 4 & 6	Noteringar				

Ofta kan metadata kategoriseras utifrån tre typer av metadata enligt figur 3.2 nedan.



Figur 3.2: kategorisering av olika typer av metadata (Franks & Kunde, 2006).

3.5 Föreskrifter för metadata

Föreskrifter om ändring i Riksarkivets föreskrifter och allmänna råd (RA-FS 1991:1) om arkiv hos statliga myndigheter; RA-FS 2008:4, föreskriver om statliga myndigheters arkivredovisning. Förutom att ange krav på arkivbeskrivningen, såsom att det ska finnas en arkivbeskrivning, klassificeringsstruktur med processbeskrivningar, och en arkivförteckning, så anges också grundkrav på uppgifter (metadata) om bland annat myndigheten, verksamheten, handlingar och handlingars förvar.

RA-FS:ar är tvingande endast för statliga myndigheter, men det finns ett också kommuner som arbetar med att implementera arkivredovisning i enlighet med denna RA-FS och som ställer krav på sådan redovisning i sin verksamhet, t.ex. Stockholms stad.

3.6 Standarder för metadata

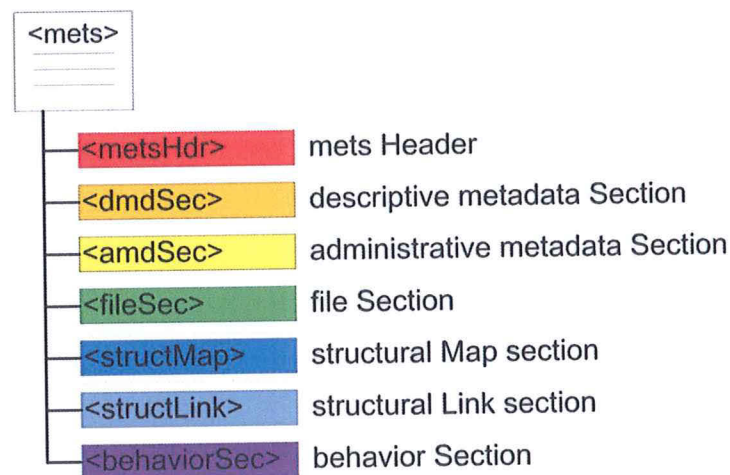
Det finns ett antal standarder för metadata och i arbetet med FGS:erna har några av dem använts som utgångspunkt. De finns beskrivna i egna dokument i projektdokumentationen inom eARD, men ges här en översiktlig beskrivning. METS har använts för FGS för paketstruktur för e-arkiv. PREMIS kan användas tillsammans med METS för bevarandemetadata. FGS för leveranstypen ärendehantering har inspirerats av MoReq (som även omfattar mer än metadata) men är inte en tillämpning rakt av.

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 34(136)
Projekt eARD DP 4 & 6		Noteringar			

3.6.1 METS (Metadata Encoding and Transmission Standard)

METS är en standard som ofta används vid paketering av arkivleveranser och ger en struktur för metadata. METS-filer kan också hanteras som sådana informationspaket som OAIS-modellen definierat (Semple, 2004). En s.k. METS-profil, som beskriver hur elementen i METS kommer att användas har tagits fram för FGS för paketstruktur för e-arkiv.

Strukturen i METS är sådan att det finns 7 sektioner som vardera placerar olika typer av metadata. I figur 3.3 nedan visas dess olika sektioner.



Figur 3.3: METS struktur för metadata. (eARD, 2013c, s. 3)

Mets Header: Här finns metadata om själva METS-dokumentet, såsom t.ex. id för METS-dokumentet och datum när det skapades.

Descriptive metadata: Här finns metadata om arkivbildare eller arkivet. Det kan vara metadata som ligger här eller hänvisa till extern metadata i t.ex. EAD eller EAC-CPF.

Administrative metadata: Här finns metadata som berör de digitala objekten. Det kan bland annat vara teknisk metadata som t.ex. filformat, metadata om rättigheter som t.ex. upphovsrätt, metadata om proveniens som t.ex. relationen mellan filer och metadata rörande bevarandeåtgärder.

File section: Här ges information om det digitala dataobjektet, t.ex. storlek och användning.

Structural Map section: Detta är den enda obligatoriska sektionen. Här beskrivs strukturen för de digitala objekten som dokumenterats i File section.

Structural Link section: Här kan man skapa pekare/hyperlänkar inom objekten i Structural Map section. Det kan t.ex. användas när en webbsida ska beskrivas med olika länkar.

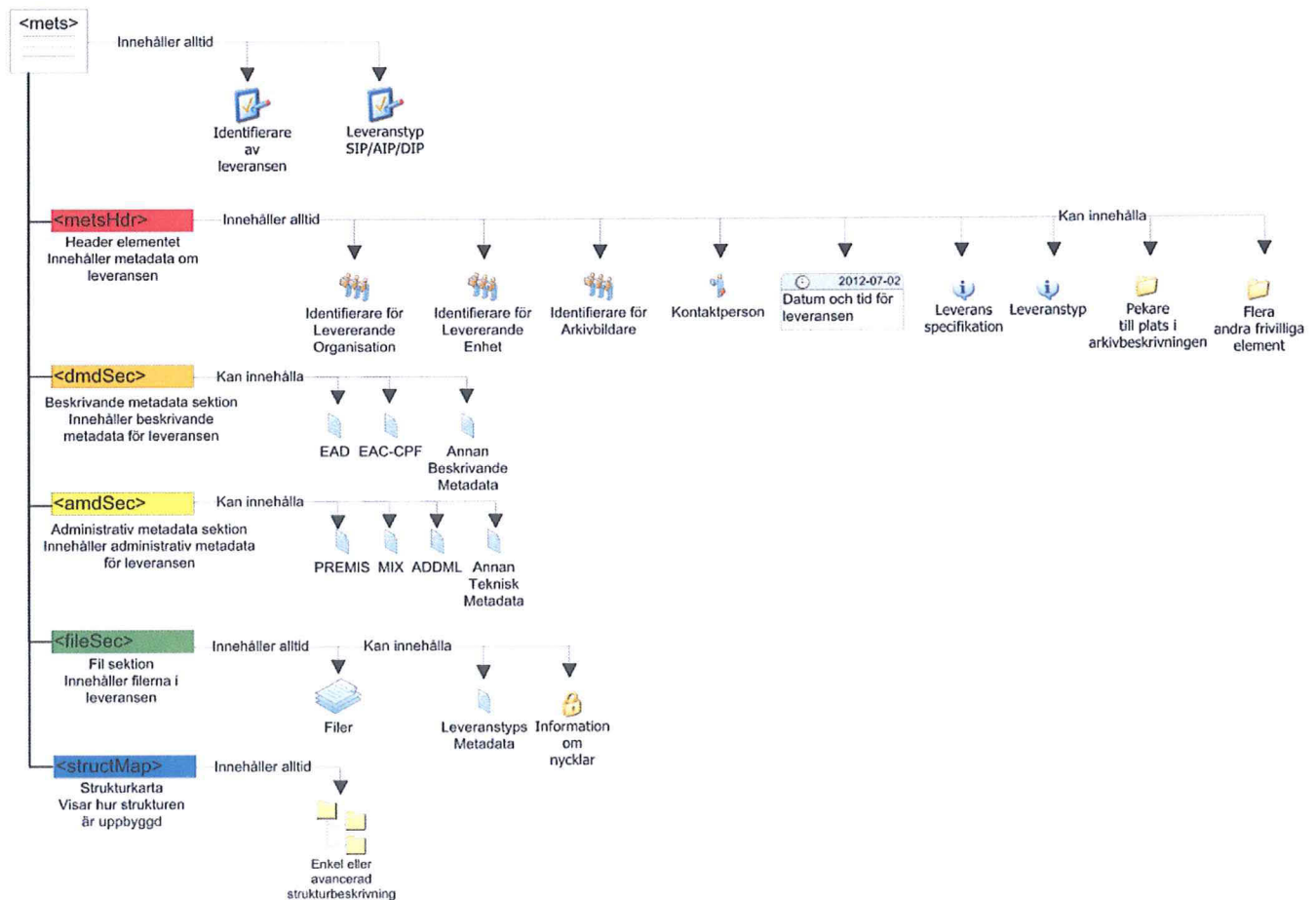
Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 35(136)
Projekt eARD DP 4 & 6		Noteringar			

Behavior Section: Här ges information om hur komponenter av informationen ska återges för användaren. T.ex. kan det omfatta information om vilka mjukvaruversioner som ska användas (eARD, 2013c).

METS använder XML för att koda strukturen. För att validera att det METS-schema som skapats följer framtagen METS-profil krävs sedan ett anpassat METS-schema (eARD, 2013c).

METS kan kombineras med metadata från andra standarder och i FGS:erna rekommenderas t.ex. att PREMIS används för bevarandemetadata.

I figur 3.4 nedan visas hur FGS:erna kan mappas mot METS struktur.



Figur 3.4: Hur FGS:erna skulle kunna mappas mot METS struktur. (eARD, 2013c, s. 7)

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 36(136)
Projekt eARD DP 4 & 6	Noteringar				

3.6.2 PREMIS (Preservation Metadata: Implementation Strategies)

Om man vill ha med information om bevarandeprocessen och t.ex. dokumentation av åtgärder som vidtagits för bevarandet av informationen i en SIP, rekommenderas det att PREMIS används.

PREMIS är en standard för att hantera, hitta och återskapa digital information. Den inkluderar proveniensinformation, aktiviteter kring digitala objekt och identifierar tekniska egenskaper. Metadata kan beskrivas i ett XML-dokument eller så kan informationen finnas i en databas, i fält som definierats för bevarandemetadata, eller kombinerat. Riksarkivet har tagit fram en svensk anpassning av PREMIS som används för behoven på Riksarkivet.

En datamodell har tagits fram där fem entiteter har identifierats; Intellectual entities, Objects (grunddata), Event (händelser), Agents (vem som gjort något) och Rights (eventuella restriktioner) (eARD, 2013d).

3.7 MoReq2010 (Modular Requirements for Records Systems)

Inom delprojekt 5 (metadata och e-tjänster för e-diarium), användes delar av MoReq2010 vid utarbetandet av grundläggande nivåer (t.ex. ett ärende och sedan handlingar som ingår i detta ärende) och metadata för ärendehantering (eARD, 2012). FGS för ärendehantering som sedan tagits fram går i MoReq 2010:s anda.

MoReq2010 är en uppsättning krav på ett informationshanteringssystem som kan användas vid kravställning och upphandling. Den beskriver grundläggande krav på system för registrering och hantering av information och innehåller även en specifikation för metadata. Tanken är att kraven ska vara anpassningsbara och kunna tillämpas för olika verksamheter och typer av organisationer. En gemensam uppsättning ”tjänster” definieras som kan passa olika typer av system och tanken är att de ska kunna införlivas i alla typer av system (eARD, 2013e; DLM Forum, 2011).

3.8 XML

I FGS:erna ingår s.k. XML-scheman, vilka finns publicerade i egna dokument.

XML, eller eXtensible Markup Language, används för att strukturera och klassificera information och används vid utbyte av information mellan olika system. En fördel med XML är att man kan skilja mellan innehåll och presentation. Ett XML-dokument är en textfil som struktureras med hjälp av s.k. taggar; $\langle \rangle$. T.ex. kan det se ut som följande:

```
<namn>Växjö</namn>
```

Den första delen är rotelementet och anger vad dokumentet beskriver. Snedstreckat markerar att elementet är slut.

Ett XML-schema tas fram för att strukturera informationen och för att kunna kontrollera att informationen är strukturerad och formaterad enligt specifikationen. Om ett dokument stämmer med



Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 37(136)
Projekt eARD DP 4 & 6	Noteringar				

reglerna i ett schema är det validerat. Det finns programvaror som maskinellt kan kontrollera om t.ex. en arkivleverans stämmer med den specifikation som överenskommits.

METS är ett exempel på ett standardiserat XML-schema (eARD, 2013f).
Med gemensamma scheman kan informationen valideras, lagras och sökas på ett enhetligt sätt (eARD, 2013g).

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 38(136)
Projekt eARD DP 4 & 6	Noteringar				

4. OAIS-MODELLEN

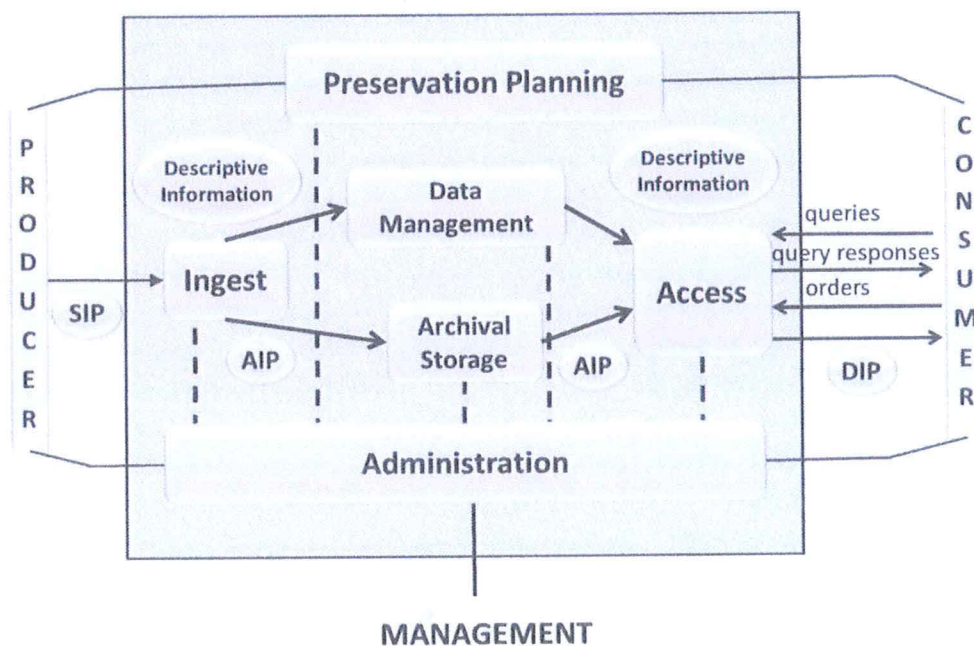
En standard som blivit etablerad i e-arkivsammanhang är den s.k. OAIS-modellen (Open Archival Information System). Den har även varit en utgångspunkt i arbetet med FGS:erna, varför en kort beskrivning ges här. Standarden finns att läsa i sin helhet på CCSDS hemsida, www.ccsds.org. FGS:erna går inte in på att definiera delar i ett OAIS-baserat arkiv, men utgår från att de e-arkiv som ska ta emot leveranser enligt FGS:erna följer OAIS-modellen och hanterar informationen i enlighet med standarden.

4.1 Om OAIS-modellen

OAIS-modellen är en ISO-standard som på en övergripande nivå definierar funktioner och roller som krävs i ett arkiv för att garantera informationens långsiktiga bevarande. Modellen definierar också tre typer av informationspaket som hanteras av arkivet; det som levereras till arkivet, det som lagras i arkivet och det som levereras ut från arkivet. Standarden har också definierat en terminologi som ofta används i e-arkivsammanhang (CCSDS, 2012).

4.2 Funktioner och roller i ett arkiv

Kortfattat kan man säga att det finns funktioner för att ta emot, lagra, hantera och lämna ut information samt planera för dess långsiktiga bevarande. I detta arbete är ett antal aktörer involverade och har olika ansvar. Att definiera funktioner och roller som krävs för ett långsiktigt bevarande av information är ett sätt att säkerställa att informationen hanteras på ett pålitligt sätt som tryggar dess äkthet och tillförlitlighet. I figur 4.1 nedan är en konceptuell modell av ett arkiv enligt OAIS-modellen.



Figur 4.1: Konceptuell modell över ett arkiv enligt OAIS-modellen, med ingående funktioner och roller (CCSDS, 2012, s. 4-1).

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 39(136)
Projekt eARD DP 4 & 6	Noteringar				

4.2.1 Roller i ett arkiv

För hanteringen av ett arkiv enligt OAIS-modellen har följande roller definierats:

- Producent som är de som levererar information till e-arkivet.
- Konsument som är de som begär ur information ur e-arkivet.
- Ledningen som äger informationen i e-arkivet, sätter ramarna för dess regelverk på en övergripande nivå och säkerställer finansieringen av arkivet.
- Administrationen som sköter den dagliga hanteringen av e-arkivet (CCSDS, 2012, s. 2-2 f.).

4.2.2 Funktioner i ett arkiv

OAIS-modellen definierar ett antal funktioner som bör finnas i ett arkiv för att säkerställa det långsiktiga bevarandet:

- Mottagande (ingest)
- Lagring (archival storage)
- Datahantering (data management)
- Administration (administration)
- Bevarandeplanering (preservation planning)
- Åtkomst (access)

4.2.2.1 Mottagande

Här tas information från producenter emot i form av SIP:ar (Submission Information Package; informationspaket för leverans¹), packas upp och kontrolleras. Därefter förbereds informationen för lagring och AIP:er (Archival Information Package; Arkivpaket²) skapas som formateras och dokumenteras enligt arkivets specifikationer. Metadata som ska användas för sökning överförs till arkivets databas och AIP:erna överförs till funktionen för lagring (CCSDS, 2012, s. 4-1).

4.2.2.2 Lagring

Här finns funktioner för att förvara, vårda och hämta upp AIP:er till åtkomstfunktionen vid återsökning och utlämnande av information. Lagringsfunktionen tar emot AIP:er och lägger dem för lagring i arkivsystemet och övervakar och kontrollerar att informationen sparats ned korrekt. Funktionen ska också säkerställa att lagringen sker på ett säkert sätt. Bland annat finns funktioner för lagringshantering, felsökningar och kontroller, samt hantering av säkerhetskopiering och möjlighet att återställa skadad information. Funktionen ska också säkerställa att informationen och dess metadata inte förändras vid eventuella byten av lagringsmedier (CCSDS, 2012, s. 4-2).

¹ Svensk översättning av Submission information package enligt eARDs begreppslista.

² Svensk översättning av Archival information package enligt eARDs begreppslista.

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 40(136)
Projekt eARD DP 4 & 6	Noteringar				

4.2.2.3 Datahantering

Denna funktion håller ordning på informationen i arkivet, hanterar arkivets databasfunktioner och möjliggör återsökning. Funktionen hanterar metadata som krävs för återsökning i arkivet och administrativ data som behövs för att administrera arkivet. Det möjliggör bland annat uttag av rapporter av olika slag. Datahanteringsfunktionen tar emot och bearbetar förfrågningar från funktionen för åtkomst och skickar svar på de frågor som ställs. Här sker också uppdateringar av olika typer av metadata för informationen i arkivet (CCSDS, 2012, s. 4-2).

4.2.2.4 Administration

Denna funktion sköter den praktiska hanteringen, förvaltningen och utvecklingen av arkivet som helhet och omfattar t.ex. förhandlingar med producenter om leveransöverenskommelser, tekniska frågor, migrering och att utarbeta och tillämpa regler och rutiner. Administrationen bedriver verksamheten utifrån direktiv från ledningen. Administrationen redovisar också rapporter, specifikationer och standarder till funktionen för bevarandeplanering som gör analyser och förslag till strategier. Till denna funktion hör också arbete med informationssäkerhet (CCSDS, 2012, s. 4-2).

4.2.2.5 Bevarandeplanering

Denna funktion ansvarar för omvärldsbevakning och ger rekommendationer för att säkerställa informationens långsiktiga tillgänglighet och användbarhet utifrån målgruppens behov. Det kan t.ex. omfatta förslag till migreringsplaner, bevarandestrategier och arkivstandarder och bevaka teknikförändringar och förändrade behov hos målgruppen. Förslagen överlämnas till administrationen som implementerar dem. Funktionen omfattar också att ta fram specifikationer för informationspaket och bidra med råd vid införande. Inom ramen för bevarandeplanering görs också risk- och konsekvensanalyser (CCSDS, 2012, s. 4-2).

4.2.2.6 Åtkomst

Här finns funktioner för återsökning, förfrågningar om och utlämnande av information från arkivet. Konsumenter kan skicka in förfrågningar och ta emot efterfrågad information i form av DIP:ar (Dissemination Information Package; informationspaket för distribution³). Här finns också kontroller för utlämnande av information så att inte känslig information kommer obehöriga tillhanda. Det kan både vara automatiska och manuella kontroller och prövningar (CCSDS, 2012, s. 4-2 f.).

³ Svensk översättning av Dissemination information package, enligt eARDs begreppslista.

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 41(136)
Projekt eARD DP 4 & 6	Noteringar				

4.3 Arkivens informationspaket

Information i ett arkiv som följer OAIS-modellen hanteras i informationspaket. Därigenom kan man hålla samman informationen med metadata som behövs för att kunna tolka, hitta och läsa informationen, oavsett vilket system den finns i. De tre informationspaketen som definierats i OAIS-modellen är

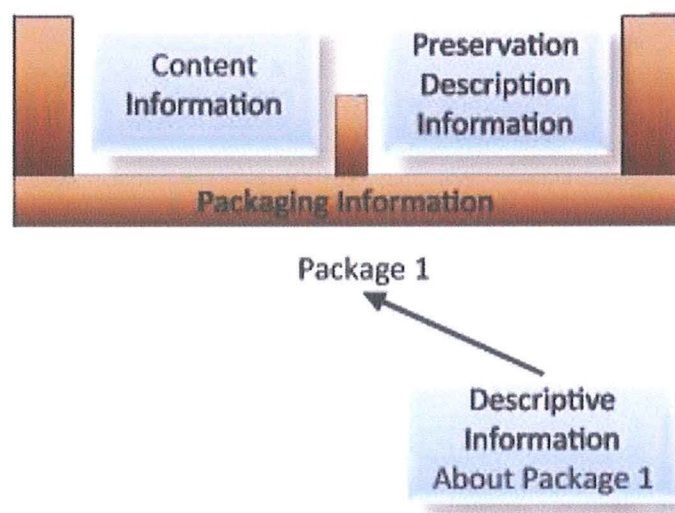
- Submission Information Package (SIP) som levereras till e-arkivet från producenterna.
- Archival Information Package (AIP) som är den information som lagras i e-arkivet.
- Dissemination Information Package (DIP) som är den information som lämnas ut till konsumenter från e-arkivet (CCSDS, 2012, s. 2-7 f.).

De olika informationspaketen kan ha lite olika uppsättning metadata. T.ex. tillförs ofta metadata för att hantera bevarandet av informationen när man skapar en AIP från en SIP. Det kan också vara vissa uppgifter hos en SIP som avser leveransen som kanske inte är relevant när informationen väl lagras i e-arkivet.

De FGS:er som tagits fram inom eARD avser leveranser till e-arkiv och berör därmed s.k. SIP.

4.4 Submission Information Package (SIP)

Information som levereras till e-arkiv ska paketeras som en s.k. Submission Information Package (SIP). En SIP består av information och olika typer av metadata och beskrivs nedan i figur 4.2.



Figur 4.2: Illustration av en SIP enligt OAIS-modellen (CCSDS, 2009, s.1.1 2-6).

Författare Tove Engvall	Avd	Telefon	Datum 13-05-31	Version 05	Sida 42(136)
Projekt eARD DP 4 & 6	Noteringar				

4.4.1 Innehållsinformation (content information)

Detta är den information som ursprungligen ska bevaras. Det består av ett dataobjekt (själva handlingen) och metadata kallad representationsinformation som krävs för att göra dataobjektet förståeligt (CCSDS, 2012, s. 4-26). Representationsinformationen kan t.ex. beskriva filstruktur och information om mjukvara och övrig teknik som krävs för att läsa informationen (Semple, 2004). Men annan dokumentation som beskriver dess historia och ursprung, hur den relaterar till andra observationer med mera ingår däremot inte i innehållsinformationen (eARD, 2013h).

4.4.2 Bevarandeinformation (Preservation Description Information)

Detta är metadata som är nödvändig för bevarandet och förståelsen av informationens innehåll och miljö där den skapades (CCSDS, s. 4-29). Det är information som stödjer trovärdigheten hos, åtkomsten till och sammanhanget hos innehållsinformationen över tid.

I figur 4.3 nedan visas vilka kategorier som bevarandeinformationen enligt OAIS-modellen kan delas in i:



Figur 4.3: Bevarandeinformation indelad i kategorier (CCSDS, 2012).