

**Statsbibliotekets
Metadatahåndbog
version 1
december 2016**

Indhold

1	Indledning	4
1.1	Formål, indhold og opbygning.....	4
2	Bibliografiske metadata	6
2.1	Definition.....	6
2.2	Formål	6
2.3	Krav til bibliografiske metadata	7
2.4	Nye katalogiseringsregler, RDA, fra 2018	8
2.5	Typer af bibliografiske metadata	8
2.5.1	Titeloplysning.....	8
2.5.2	Ophavsangivelse	9
2.5.3	Udgavebetegnelse	11
2.5.4	Serier.....	11
2.5.5	Publicering	11
2.5.6	Datering	12
2.5.7	Fysisk beskrivelse	12
2.5.8	Sprog.....	12
2.5.9	Emnedata.....	13
2.5.10	Anden indholdsbeskrivelse	15
2.5.11	Identifikatorer	15
2.5.12	Materialetype.....	15
2.5.13	Andre bibliografiske metadata (herunder kodede oplysninger)	16
3	Tekniske metadata.....	17
3.1	Definition.....	17
3.2	Formål	17
3.3	Typer af tekniske metadata	17
3.3.1	Tilblivelsesmetadata	17
3.3.2	Informationer om filen	17
4	Rettigheds- og licensmæssige metadata	19
4.1	Definition.....	19
4.2	Formål	19
4.3	Typer af rettigheds- og licensmæssige metadata	20
4.3.1	Tilgængelighed.....	20
4.3.2	Ressourcens tilladte leveringsformer	20
4.3.3	Autorisation og autentifikation.....	20
4.3.4	Ejer, licensudsteder og ophavsrethaver	21
5	Rettigheder til metadata	22
5.1	Definition.....	22
5.2	Formål	22
5.2.1	Publiceringskanaler	22
5.2.2	Ophavsret.....	23
5.2.3	Aftaleophør	23
5.2.4	Henvielse til metadataoprindelse	24
5.2.5	Dokumentation	24
6	Administrative metadata	25
6.1	Definition.....	25
6.2	Formål	25
6.3	Procesmodel for administrative metadata	25
7	Metadata for digitale forskningsdata	28
7.1	Definition.....	28
7.2	Formål	28
7.3	Krav til metadata	28

7.3.1	Registreringsformater	28
8	Standarder	30
8.1	Definition	30
8.2	Formål	30
8.3	Typer af standarder	30
Bilag "Administrative Metadata"		32
Administrative metadata for ressourcer i fysisk form		32
Administrative metadata for købte e-ressourcer		33
Administrative metadata for egne e-ressourcer		34
Bilag "Standarder"		35
Standarder ifm. katalogisering af fysiske materialer		35
Standarder ifm. katalogisering af fysiske materialer - SBCI		36
Standarder og formater ifm. købte E-ressourcer		37
Oversigt over datakilder og formater i Search/Corporate (ud over datakilder til købte e-ressourcer)		38

1 Indledning

Sidst opdateret 2016-11-16

1.1 Formål, indhold og opbygning

Metadatabasehåndbogen omfatter konkrete anvisninger og udpeger indsatsområder for metadataarbejdet på Statsbiblioteket (herefter kaldet SB). Med anvisninger menes krav til omfang, indhold og struktur i de metadata, der beskriver et fysisk eller elektronisk materiale. I arbejdet med metadata på SB skal disse anvisninger efterleves. I nær sammenhæng med metadatahåndbogen er SBs *Metadatatpolitik*, der fastlægger de overordnede rammer for det samlede arbejde med metadata.

Metadata betyder ”data om data”. Det er en fællesbetegnelse for mange forskellige typer af struktureret information, der bruges til at beskrive, administrere og genfinde samlinger. Metadata for et materiale eller en samling af materialer er vigtigt, fordi det giver en indgang til indholdet. Det er særlig vigtigt for de ressourcer af tekst, billeder, lyd og film, der ikke kan gøres søgbare som tekst. Uden metadata vil sådanne samlinger ikke, eller kun i meget begrænset omfang, kunne tilgængeliggøres. Metadata er også afgørende for bevaring af samlingerne og for adgangsstyring til samlingerne.

Formålet med metadatahåndbogen er at beskrive praksis, principper og prioriteringer for SBs metadataarbejde. Håndbogen skal således skabe et beslutningsgrundlag for håndtering, anvendelse og udvikling af metadata.

Det har i arbejdet med metadatahåndbogen ikke været muligt at give anvisninger på alle aspekter vedrørende metadata, enten pga. forskellig praksis for håndtering af det givne aspekt, eller fordi det ikke tidligere har været adresseret som en metadatamæssig udfordring. I metadatahåndbogen er de vigtigste af disse forhold fremhævet som indsatsområder.

SBs Metadatahåndbog består af definitioner, formål, krav og indsatsområder for følgende typer og aspekter af metadata:

- **Bibliografiske (beskrivende) metadata**
- **Tekniske metadata**
- **Rettigheds- og licensmæssige metadata**
- **Rettigheder til metadata**
- **Administrative metadata**
- **Metadata for forskningsdata**
- **Metadatastandarder.**

Ovenstående emner er inddelt i kapitler:

- Kapitel 2 vedrører de bibliografiske metadata. Bibliografiske metadata forstås her som data, der identificerer og beskriver indholdet af en ressource.
- Kapitel 3 vedrører tekniske metadata for filer som genereres eller modtages med henblik på digital bevaring. Tekniske metadata er informationer om filernes tilblivelse og om filernes egenskaber.
- Kapitel 4 vedrører de rettigheds- og licensmæssige metadata. Rettigheds- og licensmæssige metadata forstås her som data, der identificerer hvem der kan få adgang til en ressource eller en samling af ressourcer, og i hvilken form adgangen må ske. Rettighedsstyring har fået

stigende betydning, i og med at biblioteker ofte køber adgang til ressourcerne frem for at eje dem.

- Kapitel 5 vedrører rettigheder til metadata. Rettigheder til metadata forstås her som, hvordan metadata må tilgængeliggøres og udveksles.
- Kapitel 6 og bilaget ”Administrative Metadata” vedrører de metadata som anvendes til administration af bibliotekets ressourcer samt workflows der bør anvendes i metadataproduktionen. Administrative metadata forstås her som data, der beskriver en resources anskaffelse, indlemmelse, opbevaring og kassation.
- Kapitel 7 vedrører de metadata som beskriver indsamlede, observerede eller skabte forskningsdata.
- Kapitel 8 og bilaget ”Standarder” vedrører SBs anvendte standarder samt standarder metadata bør overholde.

Yderligere information fås ved henvendelse til metadatagruppen (jf. SBs *metadatapolitik*).

2 Bibliografiske metadata

Sidst opdateret 2016-11-10

2.1 Definition

Bibliografiske metadata er data, der identificerer og beskriver indholdet af en ressource (dette kaldes også deskriptive metadata). Det er data som f.eks. titler, ophavsangivelser, emneord, klassifikationer, noter, abstracts, publiceringsdata, fysisk beskrivelse og omfang.¹

Bibliografiske metadata kan tilvejebringes på forskellig vis:

1. Manuel registrering
2. Genbrug/konvertering af data fra andre kilder
3. Berigelse af data
4. Automatisk registrering.

Disse metoder kan også kombineres.

2.2 Formål

Katalogpolitik 1997:

Ifølge SBs katalogpolitik fra 1997² skulle de bibliografiske metadata kunne besvare følgende spørgsmål:

1. Har biblioteket et bestemt, navngivet dokument?
2. Har biblioteket et eller flere dokumenter om et bestemt emne?
3. Har biblioteket et eller flere dokumenter af et bestemt ophav?

Denne politik havde dog som forudsætning, at man kunne benytte bibliografier til mere udtømmende søgninger, og den var udformet på et tidspunkt, hvor store dele af samlingerne ikke var registreret.

Disse forudsætninger har imidlertid ændret sig i mellemtiden. Dels registrerer SB i dag materialer, som ikke findes i bibliografier (f.eks. småtryk), dels registrerer man også unikke materialer (f.eks. egne digitaliseringer og radio/tv-udsendelser). Hertil kommer, at SB i vidt omfang har adgang til registreringer på artikelniveau af samlingerne. Katalogen er således i dag langt mere omfattende, og da forventningen hos brugerne er, at katalogen afspejler størstedelen af samlingerne, kan der med rette stilles større krav til registreringen af visse materialetyper.

Katalogpolitik:

Formålet med de bibliografiske og beskrivende metadata er at kunne:

- 1) finde ressourcer
- 2) identificere ressourcer
- 3) vælge ressourcer³.

¹ Forskellige definitioner kan ses her:

Niso: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
Digitalbevaring.dk.: <http://digitalbevaring.dk/metadata/>

² KatPol-rapporten SBs katalogpolitik, udarbejdet af KatPol-udvalget, 1997

³ I overensstemmelse med FRBR (Functional Requirements for Bibliographic Records), se evt. <http://www.loc.gov/cds/downloads/FRBR.PDF> for en kortfattet beskrivelse.

Ad 1: Finde ressourcer

Brugeren skal være i stand til at finde bestemte navngivne ressourcer, finde ressourcer af et bestemt ophav, udgaver af et bestemt værk, ressourcer om et bestemt emne samt ressourcer med andre karakteristika.

For at opfylde det formål skal en søgning kunne rettes mod en bestemt kategori af data – f.eks. skal man kunne skelne mellem, om en person er ophav til en ressource eller er emnet for en ressource. Det er endvidere nødvendigt, at data har en form, der gør det muligt at udtrække standardiserede oplysninger til f.eks. afgrænsning af forfatter, materiale og sprog (= facetter).

Endelig kan data i bestemte kategorier (typisk emneord, klassifikationer, ophav, titler) benyttes til automatisk at finde andre ressourcer med ét eller flere fælles dataelementer og rankere dem efter graden af fælles data ('Lignende poster').

Ad 2: Identificere ressourcer

Brugeren skal være i stand til at afgøre, om de præsenterede ressourcer modsvarer det ønskede.

For eksempel: er det det korrekte tidsskrift eller et andet med en lignende titel? Hvad handler værket om? Hvem er ophavet til denne ressource? Svarer søgeresultatet til det forventede? For at opfylde dette formål skal søgeresultatet gerne kunne assistere, f.eks. ved at være rankeret med de mest relevante resultater først.

Ad 3: Vælge ressourcer

Brugeren skal kunne træffe et kvalificeret valg mellem flere lignende ressourcer.

Når brugeren skal vælge en bestemt ressource, sker det typisk ud fra oplysninger om ressourcens alder, format, sprog, indhold, omfang og tilgængelighed.

2.3 Krav til bibliografiske metadata

Når registreringsniveau og -praksis skal fastlægges, er der især tre aspekter, som bør indgå i vurderingen af metadata, uanset om man selv skal producere dem, eller de skal anskaffes udefra, og uanset hvilke type ressourcer, der skal behandles:

- Brugerens behov: Er beskrivelser og standardiserede data udformet med tanke på målgruppens behov? Er sproget i beskrivelser og standardiserede data det samme som benyttes af hovedparten af brugerne?
- Konsekvens og standardisering: Er beskrivelser og normative data beskrevet konsekvent og standardiseret, set i forhold til ønsket om størst mulig ensartethed og muligheder for udveksling af bibliografiske og normative data?
- Økonomi: Hvor og hvordan får man mest for pengene, hvis der er valgmuligheder?

I det efterfølgende beskrives de relevante typer af bibliografiske metadata samt de krav, SB stiller i forbindelse med disse. For hver type er angivet primære og sekundære krav:

- Primære krav beskriver de data, der som minimum skal være til stede, for at man sikkert kan identificere og genfinde en ressource, dvs. formålet er primært "lagerstyring".
- Sekundære krav beskriver de data, som derudover vedrører formidling af ressourcen til slutbrugere.

2.4 Nye katalogiseringsregler, RDA, fra 2018

I juni 2016 vedtog Slots- og Kulturstyrelsen at de nuværende anglo-amerikanske katalogiseringsregler skal erstattes af det internationale regelsæt RDA (Resource Description and Access), med virkning fra 1.1.2018.

Da hovedparten af SBs katalogdata tilvejebringes via import, skal SB forholde sig til de udfordringer, dette vil bringe.

Efter hver datatype er opregnet de krav, der i RDA blev anset for minimumskrav (*core elements*)⁴ på det tidspunkt, da metadatabogens forgænger, metadatatpolitikken vers. 1 og 2 blev til.

Da arbejdet med en dansk RDA-profil først er startet i efteråret 2016, har vi valgt ikke at revidere kapitlet på nuværende tidspunkt, men afvente implementeringen af RDA og dermed en forventet total revision af indholdet.

Det bærende element i RDA er værkvisning, altså idéen/ønsket om at kunne vise alle repræsentationer af et bestemt værk, uanset medium, på basis af den samme søgning. Dette er en betydelig ændring i forhold til de nuværende katalogiseringsregler, hvor man udelukkende registrerer den foreliggende udgave af værket.

2.5 Typer af bibliografiske metadata

2.5.1 Titeloplysning

Titeloplysninger er et vigtigt søgeelement, både som langordssøgning til at finde en præcis titel, og som ordsøgning til at finde 'noget om'.

Titeloplysninger benyttes til at identificere, at man har fundet det værk, man søgte, hvis man søger en bestemt titel, men også til at afgøre, om værket falder inden for det søgte fagområde i en emnesøgning.

Titlens identificerende elementer benyttes til at skelne mellem to værker med samme titel (fx udgivelsessted for et tidsskrift med uprægnant titel, eller dirigent af et musikalsk værk).

Titeloplysninger har også en vigtig funktion i forhold til præsentation. Hovedtitlen og første undertitel udtrækkes i SBs nuværende søgesystem til at præsentere værket i et oversigtsformat, som bruges til at lænke til det fulde format.

Titler benyttes ofte til opstilling af det fysiske materiale, men også til sortering af et søgesæt.

I SBs nuværende søgesystem Search og bibliotekssystem Aleph er der også mulighed for at lave registeropslag på titler, hvilket er vigtigt, ikke mindst i forhold til tidsskrifter (for at undgå uoverskuelige resultater på søgninger som fx Science og Philosophy).

Primære krav:

- Hovedtitel: som et absolut minimum må det kræves, at der er en hovedtitel, og at de undertitler, der er nødvendige for at identificere værket, gengives. Det er et krav, at man kan skelne mellem hoved- og undertitler, da man ellers ikke har mulighed for at separere

⁴ Centralt i RDA er, at fokus her er skiftet fra beskrivelsen af det enkelte dokument til i stedet at se dokumentet som en repræsentation af et værk. Vigtige begreber er Work (værk), Expression (repræsentation), Manifestation (dokument) og Item (eksemplar).

hovedtitlen til webfremvisning mv. Såfremt ingen titel forekommer, må der konstrueres en passende titel. Sprog og tegnsæt udgør en særlig problemstilling⁵.

Sekundære krav:

- Paralleltitler, originaltitler, værktitel ("uniform titel"), titler på tidligere og senere udgaver. Alle disse tjener til at muliggøre værkvisning, og derudover giver de vigtige informationer til søgeformål.
- Identifikation af indledende bestemt/ubestemt artikel, som muliggør udtræk til langordsregistre både med og uden indledende artikel.
- Alternative titler med variantformer, hvis titler indeholder tal, symboler o. lign. (skal der fx listes/søges på § eller Paragraf). Andre alternative titler, f.eks. titler som værket er kendt under, titler i andre skriftsprog, tilføjelser som muliggør skelnen af én titel fra en anden lignende titel, øvrige undertitler.

Core elements i RDA: Hovedtitel, værktitel, identificerende tilføjelser til værktitel.

2.5.2 Ophavsangivelse

Ophavsangivelser er nødvendige for at finde, hvad biblioteket stiller til rådighed af et bestemt ophav, ligesom de benyttes til at identificere ophavet til en funden ressource og til at vælge én bestemt ressource frem for en anden.

Ophavsangivelser kan deles op i de citerende ophavsangivelser (navnet som det står beskrevet i ressourcen) og ophavsangivelser som opslag (den standardiserede form af navnet).

2.5.2.1 Citerende ophavsangivelser

Primære krav:

- Såfremt den navneform, der står i materialet (fx Bob Greenblatt), afviger fra den standardiserede (Robert Greenblatt) bør den altid citeres, så der kan søges på begge former, og så ophavet kan genkendes⁶.

Sekundære krav:

- Citation af alle ophav (dog højst op til 3 med samme funktion).

Core elements i RDA: Elementet er et core element i RDA, dog er det kun et krav, at det første ophav til dokumentet skal citeres.

2.5.2.2 Ophavsangivelser som opslag (personer)

Ophavsangivelser som opslag benyttes dels til søgning, dels til afgrænsning. Endvidere benyttes ophavsangivelsen i standardiseret form til at identificere værket.

⁵ Det er nødvendigt at kende de standarder, der er anvendt til translitteration, såfremt der er foretaget en sådan. I modsat fald skal man også vide dette - det er en væsentlig oplysning for at kunne genfinde f.eks. kyrilliske og arabiske titler.

⁶ I den danske nationalbibliografi har man siden 2010 kun citeret, hvis der skønnes at være store afvigelser mellem de to navneformer. I ovennævnte tilfælde ville Bob Greenblatt ikke skønnes at være en sådan. De lande, vi i øvrigt importerer poster fra, fastholder citering, hvilket SB også gør ved grundkatalogisering.

Primære krav:

- At navnet er opdelt i efternavn, fornavn og identificerende tilføjelser (evt. blot ved at navnet er inverteret korrekt). Hvis der både er en standardiseret og en alternativ navneform for samme ophav i samme post, skal det være muligt at skelne mellem disse, da kun den standardiserede form skal anvendes som facet.

Sekundære krav:

- Valg af samme navneform hver gang, tilføjelser til navnet for at kunne skelne mellem flere personer med samme navn, tilføjelser af koder for funktion, sammenkobling af den standardiserede navneform med andre navneformer for samme person i form af henvisninger, (link til) poster til standardformer (=autoritetsposter) eller lignende⁷.

Core elements i RDA: Standardiseret navn samt alle de følgende oplysninger, når de er 'applicable and readily ascertainable'. Det kan f.eks. være adels- eller religiøs titel, dato for fødsel og død eller andre identificerende tilføjelser, samt hvis nødvendigt for at skelne: fuldt udskrevet navn, aktivitetsperiode, profession eller beskæftigelse.

2.5.2.3 Korporationer som ophav

Korporationer kan være ophav til publikationer som f.eks. virksomhedsregnskaber, udvalgsbetænkninger, lokalplaner og konferencerapporter, men de kan også være ophav til kunstværker, dvs. værker af mere unik karakter, såsom musikværker, teaterforestillinger mv. I sidstnævnte tilfælde er korporationen et vigtigt søgeelement, i førstnævnte er det ikke altid nødvendigvis et element, som slutbrugeren vil opfatte som ophav.

Primære krav:

- At et korporativt ophav altid citeres.
- At der i tilfælde af kunstværker laves opslag på det/de korporative ophav.

Sekundære krav:

- At der laves opslag på det/de korporative ophav.

Core elements i RDA: Foretrukne navn samt alle de følgende oplysninger, når de er 'applicable and readily ascertainable'. Det kan f.eks. være sted for conference, årstal associeret med en korporation, associeret institution, konferencenummer, samt hvis nødvendigt for at skelne: hovedkvarterets geografiske placering, associeret institution eller andre identificerende tilføjelser.

2.5.2.4 Valg af primært ophav

Valg af primært ophav udgør en særskilt problemstilling. Såfremt en ressource har flere ophav, skal ét eller flere af disse vælges til brug forskellige steder: i oversigtsformatet i brugergrænsefladen, til opstilling, til eksport til referenceværktøjer og øvrige citationer. Ikke mindst i forbindelse med

⁷ Der arbejdes p.t. internationalt på fælles initiativer som f.eks. **VIAF** (Virtual International Authority File), hvor man forsøger at sammenkoble standardiserede navneformer fra forskellige lande og som **ORCID**, hvor forskerne kan knytte publikationer til én identitet. Disse initiativer er vigtige at følge med i, da de også gradvis udvides til at omfatte værker og repræsentationer. Hvis SB skal have glæde af dette arbejde i de lokale kataloger, kræver det, at SBs egne data er så gode, at de kan indgå i udveksling. Hvis SB skal kunne benytte data til opslag i eksterne kilder (biografier, encyklopædier el. lign) er det ligeledes vigtigt, at navneformen optræder i standardiseret form.

forskningspublikationer, der ofte har adskillige bidragydere som alle har interesse i at stå som reference, udgør rækkefølgen en meget væsentlig problemstilling.

Primære krav:

- At det/de vigtigste (første) ophav fremgår enten som følge af formatering eller rækkefølge.

Sekundære krav:

- At der altid specifikt udpeges en 'værkansvarlig'.

Core elements i RDA: Første ophav citeret; som opslag: (første) primære ophav til værket, repræsentationen og dokumentet.

2.5.3 Udgavebetegnelse

Udgavebetegnelser benyttes primært til at vælge én given ressource frem for en anden.

Primære krav:

- Angivelse af egentlige udgavebetegnelser.

Sekundære krav:

- Citerende ophavsangivelser hørende til udgaven. SB bør desuden som pligtafleveringsbibliotek gengive oplagsdata, evt. blot i forbindelse med beholdningsangivelsen, således som det sker i dag.

Core elements i RDA: Udgavebetegnelse, betegnelse for en angivelse af revideret udgave.

2.5.4 Serier

Seriebetegnelser bruges som søgeelementer og som identifikation. Serietitlen er særligt vigtig, når en kontinuerlig udgivelsesrække ikke sikkert kan defineres som et tidsskrift, en række af enkeltstående monografier eller ét værk i flere bind.

Primære krav:

- Seriens titel, nummer, ISSN, underserietitel, underserienummer.

Sekundære krav:

- Parallel serietitel.

Core elements i RDA: Seriens titel, nummer, underserietitel, underserienummer.

2.5.5 Publicering

Publiceringsoplysninger for publiceret materiale er typisk sted og udgiver, og alternativt distributør. For upubliceret materiale kan det være fremstillingssted og producent. Publiceringsoplysninger hjælper med at identificere og vælge ressourcer. Derudover er oplysningerne også vigtige i forbindelse med rykning for afleveringspligtige materialer.

Primære krav:

- Mindst ét element af de følgende:
Udgivelsessted/distributionssted/fremstillingssted/forlag/udgiver/distributør/producent.

Sekundære krav:

- De øvrige af ovennævnte data.

Core elements i RDA: Mindst ét udgivelsessted/distributionssted/fremstillingssted og mindst én udgiver(forlag)/distributør/producent.

2.5.6 Datering

Datering benyttes til søgning, sortering, afgrænsning af søgeresultater, identifikation og til at vælge én ressource frem for en anden. Det er derfor vigtigt, at datering optræder i kodet/renset form. Derudover kan datering optræde i citerende version (hvor fx et arabisk årstal skrives, som det forekommer, sammen med en 'oversættelse' til et vestligt årstal, eller et usikkert årstal angives som usikkert).

Primære krav:

- For afsluttede/monografiske ressourcer: mindst ét årstal i søgbar form.

Core elements i RDA: Publiceringsdato/distributionsdato/produktionsdato/copyright-dato.

2.5.7 Fysisk beskrivelse

Den fysiske beskrivelse fortæller noget om omfang og fysisk medium (se også kap. 3: Tekniske metadata). Den er derfor vigtig af hensyn til identifikation og valg af én ressource frem for en anden (fx hovedrapporten på 300 sider eller resumérapporten af samme navn på 50 sider, i hhv. fysisk og digital form).

Da en del elektroniske ressourcer svarer til ressourcer i fysisk form, vil der ofte forekomme oplysninger om fx sidetal i e-bøger, eller spilletid på en tv-udsendelse der er tilgængelig online.

Primære krav:

- Fysisk medium for hovedmateriale og bilagsmateriale, omfang af hovedmateriale (når det kan fastslås).

Sekundære krav:

- Omfang af bilagsmateriale.

Core elements i RDA: Carrier type og omfang (hvis ressourcen er komplet).

2.5.8 Sprog

En betegnelse for ressourcens sprog benyttes til søgning, afgrænsning og valg af ressource. For at opfylde disse formål bør sprogbetegnelsen være i standardiseret form.

Primære krav:

- Angivelse af ressourcens hovedsprog, hvis ressourcen er i tekstlig form eller indeholder tale.

Sekundære krav:

- Angivelse af originalsprog, parallelsprog, synkroniserede sprog, sprog i undertekster og resuméer.

Core elements i RDA: Kun hvis nødvendigt til at skelne én repræsentation fra en anden.

2.5.9 Emnedata

Valg af emnedata afhænger i princippet ikke af, om materialerne udgives i fysisk eller elektronisk form, men mere af hvilke materialetyper der er tale om, da det som regel vil være bestemmende for, hvilke emnedata der vil være til rådighed.

I realiteten er der dog alligevel forskel. Der er fx p.t. flere valgmuligheder i forbindelse med den manuelle import af katalogdata til bøger i fysisk form end ved den maskinelle import af katalogdata til e-bøger, hvor data p.t. leveres og opdateres direkte i søgesystemet på basis af en generel aftale med dataleverandøren.

For andre typer af materiale indgår emnedata slet ikke som en mulighed i forbindelse med køb af metadata, og SB skal fra sag til sag tage stilling til, om der er behov/mulighed for at afsætte ressourcer til at berige med emnedata i en særskilt proces, evt. ved at påføre data manuelt.

For at sikre størst mulig nyttevirkning af emnedata, bør der anvendes så få klassifikationssystemer og forskelligsprogede emneord som muligt.

For at sikre størst mulig ensartethed i brugen af de valgte klassifikationssystemer og emneord, bør emnedata, der er oprettet som en del af nationalbibliografien eller en tilsvarende institution (nationalbibliotek/-arkiv), prioriteres højest.

Indsatsområde:

Det prioriteres højt, at metadata ved særlige opdateringskørsler beriges retrospektivt med såvel emneord som klassifikationer jf. nedenstående specifikationer.

Indsatsområde:

Det prioriteres højt at undersøge, hvordan anvendelse af "linked data" fra eksterne kilder kan berige katalogen. Eksempelvis VIAF <http://viaf.org/> for navne på ophav og muligheder for krydshenvisninger mellem forskellige klassifikationssystemer. I lyset af den hastigt stigende mængde af købte monografier i e-form⁸ (418.000 i 2015), bør det samtidig opprioriteres at forsøge at få de ønskede emnedata til at indgå i købte metadataleverancer.

2.5.9.1 Emneord

Emneord tjener til at beskrive ressourcens emne. Emneord kan være frie eller kontrollerede (dvs. optræde i en forhåndsdefineret liste over ord med fast betydning – en liste der kan være hierarkisk og indeholde relationer). Formålet med emneord er dels at øge muligheden for genfindning, og dels at give en beskrivelse af emnet, som kan forstås af slutbrugere.

⁸ E-bogsbestand opgivet til Biblioteksstyrelsen / Danmarks Statistik 2008-2012:

2008	53.395
2009	188.030
2010	210.286
2011	260.395
2012	283.576

Emneordene optræder som afgrænsningsmuligheder, men de benyttes også til at udvælge 'Lignende poster' i katalogen, ligesom de kan benyttes til videresøgning.

Primære krav:

- Såfremt der optræder emneord, bør de fortrinsvis være på dansk, engelsk eller på ressourcens sprog. Vi anbefaler følgende, i prioriteret rækkefølge:

For danske materialer:

- a. DBC's kontrollerede emneord
- b. Andre kontrollerede, dansksprogede emneord
- c. Kontrollerede, engelsksprogede emneord
- d. Andre dansk- eller engelsksprogede emneord.

For alle andre materialer:

- a. Kontrollerede emneord:
 - Library of Congress Subject Headings (LCSH)
 - Library of Congress Geographical Headings
 - Uniforme titler
 - Medical Subject Headings (MESH, fra National Library of Medicine).
- b. Andre engelsksprogede, kontrollerede emneord
- c. Andre engelsksprogede emneord
- d. Emneord på ressourcens eget sprog.

Indsatsområde:

Der bør lægges stor vægt på at øge mængden af registreringer med emneord, enten ved tildeling af emneord under registreringen eller ved automatisk berigelse i form af løbende eller specielle opdateringskørsler.

Core elements i RDA: Mindst ét 'subject relationship element'.

2.5.9.2 Klassifikation

Klassifikation består normalt i tildeling af en kode, der repræsenterer et emne i en hierarkisk emneoversigt, fx Dewey Decimal Classification, DK5 eller NLM Classification.

Klassifikation tjener til at udvide søgemulighederne og giver mulighed for at søge i andre poster inden for samme emne ('Lignende poster').

SB har ikke et krav om, at alle ressourcer tildeles en klassifikation inden for ét klassifikationssystem.

Sekundære krav:

- Klassifikationer er nyttige til at finde andre ressourcer, og er specielt gode til at finde ressourcer inden for overordnede eller underordnede begreber. Derfor bør der lægges stor vægt på genbrug af klassifikationer inden for alment kendte og benyttede klassifikationssystemer.

Såfremt der optræder klassifikationsdata, anbefaler vi nedenstående:

For danske materialer:

DK5

For alle andre materialer:

- a. Dewey Decimal Classification (DDC)
- b. National Library of Medicine (NLM).

Core elements i RDA: Mindst ét 'subject relationship element'.

2.5.10 Anden indholdsbeskrivelse

Anden indholdsbeskrivelse dækker over noter, abstracts, indholdsfortegnelser, forsider, fuldtekst, brugergenererede data m.m. Disse data kan øge genfinding, når de er søgbare, men kan dog også i visse tilfælde give støj i søgninger. Ofte vil søgbarhed dog være en fordel, specielt når søgeresultatet er relevansrankeret, så denne type information har lavere relevans end f.eks. titler og emneord. Endvidere kan det være en hjælp til at vælge én bestemt ressource frem for en anden.

Primære krav:

- Indholdsdata vægtes i brugerundersøgelser meget højt og bør derfor også vægtes, når SB vurderer datakilder.

Indsatsområde:

SB bør arbejde for at få gjort indholdsfortegnelser til trykte faglitterære publikationer søgbare.

Core elements i RDA: Intet.

2.5.11 Identifikatorer

Identifikatorer omfatter begreber som ISBN, ISSN, plademærker og -numre, FAUST-nr., URLer, URler, URNer, DOIer, lokale id'er, stregkoder m.m. Disse identifikatorer bruges dels lokalt, dels til at udveksle entydig information og derudover også til at berige poster m.m.

Primære krav:

- Mindst én identifikator, dog normalt alle internationale standardnumre.

Core elements i RDA: mindst én identifikator ("identifier") for dokumentet, mindst én for værket (kan være den samme, når der kun findes én manifestation af værket).

2.5.12 Materialetype

Materialetyper er meget væsentlige elementer i forhold til alle funktioner: søgning, afgrænsning, identifikation og valg af ressourcer. Materialetyper kan optræde på forskellige måder, enten i form af koder eller som definerede materialetyper. I SBs nuværende søgesystem er alle praksisreglernes

koder tilgængelige for den professionelle bruger.⁹ På tværs af datakilderne er der defineret et overskueligt antal fælles materialetyper, som præsenteres for slutbrugerne til afgrænsning af søgninger.

Primære krav:

- Data, der gør det muligt at dele søgeresultatet fornuftigt op.

Sekundære krav:

- Så mange data som muligt, der maskinelt kan oversættes til praksisreglernes afgrænsning og/eller benyttes til klartekstning i oversigtsformatet (fx koder, der muliggør visning af, at et materiale er en børnebog).

Core elements i RDA: Content type og carrier type.

2.5.13 Andre bibliografiske metadata (herunder kodede oplysninger)

Koder for festskrifter, bibliografier, skønlitteratur/faglitteratur, niveau og udgivelsesland er eksempler på andre bibliografiske metadata. I kodet/struktureret form kan oplysningerne gøres tilgængelige for søgning og/eller fremvisning. Der bør dog defineres et minimum af antal poster, der har disse oplysninger, før man udstiller dem.

Primære krav:

- Landekoden 'dk' i dansk udgivet materiale.

Sekundære krav:

- Koder for fagligt niveau, ligesom koder for hhv. skønlitteratur og faglitteratur er meget ønskelige at få bredt repræsenteret i katalogen.

Core elements i RDA: Intet.

⁹ For information om praksisregler for søgeveje, se:

http://www.kulturstyrelsen.dk/fileadmin/publikationer/andre_publicationer/praksisregler_soegeveje/index.htm

3 Tekniske metadata

Sidst opdateret 2016-11-10

3.1 Definition

Indtil videre omhandler dette kapitel de tekniske metadata for digitale filer, som genereres eller modtages med henblik på digital bevaring, enten midlertidigt (fx forskningsdata) eller permanent (pligtafl levering/dansk kulturarv). Tekniske metadata er informationer om filernes tilblivelse og egenskaber.

3.2 Formål

Tekniske metadata er afgørende for formidling og bevaring af de digitale samlinger. Tekniske metadata kan forklare, hvordan en samling er blevet til. De angiver forhold omkring samlingens fysiske tilstand, da den blev digitaliseret og dermed være udslagsgivende for samlingens bevaringsplan. De bør så vidt muligt angive hvilke krav der er til udstyr og/eller software for at kunne få adgang til data.

Indsatsområde:

Der er en række standarder for tekniske metadata, og det er en løbende opgave at afklare, hvilke standarder der evt. er egnede i forhold til en specifik samling.

3.3 Typer af tekniske metadata

3.3.1 Tilblivelsesmetadata

Metadata om tilblivelsen giver informationer om, hvorfor en fil ser ud, som den gør. I formidlingssammenhænge kan det være relevant at forstå filen i dens kontekst for at understøtte forståelse og troværdighed.

Primære krav:

- Dokumentation af digitaliseringsværktøjer brugt internt på SB: fx software, hardware, versionsnr., indstillinger m.m.

Sekundære krav:

- Man skal vide nok til at gøre det igen: hvis man har fyldestgørende metadata om digitaliseringsprocessen, så skal man i princippet kunne gentage processen og nå frem til en fil, der er identisk med filen fra den originale digitaliseringsproces. Relevante metadata kunne fx være data om:
 - omgivelser som luftfugtighed, temperatur, støj
 - værktøjer, herunder benyttet software, hardware, versionsnr. og indstillinger
 - personer/firmaer/institutioner involveret.

3.3.2 Informationer om filen

Metadata om filernes egenskaber kan benyttes til at genskabe og fortolke filen samt til at planlægge og foretage bevaringshandling.

Primære krav:

- Dokumentation af værktøjer brugt internt på SB. Dette kan fx være dokumentation af benyttet software, hardware, versionsnr., indstillinger m.m. i forbindelse med intern digitalisering.
- Checksum på filen. Dvs. resultatet af en algoritme kørt på filen, som giver et næsten unikt fingeraftryk.
- Filformat (dog ikke Netarkivet): informationer om filformat og versionsnummer med henvisning til, hvor man kan læse om formatet; dette kan fx være et link til en standard eller lignende.

Sekundære krav:

- Tilstrækkelige informationer til at kunne fortolke bits i filerne; dvs. en præcis og tilstrækkelig beskrivelse af formatet (fx en beskrivelse af, hvordan man vil kunne afspille MPEG videostream, selv om man ikke har softwaren til det).
- Egenskaber ved filen trækkes ud, fx:
 - Opløsning, farvedybde, størrelse etc. ved stillbilleder
 - Opløsning, framerate, længde, codecs¹⁰ etc. ved levende billeder
 - Omfang, opsætning etc. ved tekst
 - Frekvens, længde, codecs etc. ved lyd.
- Oplysninger om relevante indlejrede formater og dokumentation af værktøjer, også ved eksterne leverandører. Dette kan fx være dokumentation af benyttet software, hardware, versionsnr. og indstillinger, som er anvendt i forbindelse med digitaliseringsprojekter.

¹⁰ Det engelske ord **Codec** kommer fra **compressor-decompressor** eller **coder-decoder**. En codec er et program (en implementation) af hvordan man koder og afkoder mellem den originale datarepræsentation og en komprimeret datarepræsentation – og tilbage igen. Data til en codec er fx digitaliseret lyd, billede eller video.

4 Rettigheds- og licensmæssige metadata

Sidst opdateret 2016-10-25

4.1 Definition

Rettigheds- og licensmæssige metadata defineres i dette afsnit som informationer om, hvem der kan få adgang til en ressource, og i hvilken form adgangen må ske.

4.2 Formål

Formålet er primært dokumentation og konfiguration af manuelle og automatiserede adgangsstyringsprocesser:

- Manuelt, ved at en medarbejder slår op i metadata for at undersøge, om en ressource må udleveres eller kopieres.
- Automatiseret, ved at et system laver opslaget og på baggrund af dette afgør, om brugeren har adgang.

Rettigheds- og licensmæssige metadata relaterer sig som regel til større samlinger af materialer. Disse metadata kan enten være indlejret i den enkelte metadatapost (en trykt bog, der kun er til brug på læsesalen) eller relatere sig til en større samling af poster (alle artikler fra Cambridge University Press må uploades til Blackboard). Hyppigt er rettigheds- og licensmæssige metadata ikke som udgangspunkt digitaliseret, men fremgår af kontrakter eller lov på området. For ressourcer i elektronisk form er det essentielt, at rettigheds- og licensmæssige metadata digitaliseres. Dette sker i SBs ERMS (Electronic Ressource Management System), når der indgås nye aftaler om køb/leje af elektroniske ressourcer. Disse data er grundlaget for, at udvalgte licensbetingelser kan vises for bestemte brugergrupper, når de er logget ind.

En forudsætning for digitaliseringen er, at de rettigheds- og licensmæssige forhold udtrykkes med standardbegreber, og at der er fokus på dette fra aftaleindgåelse til implementering.

Der er tre basale forhold, som en slutbruger bør kunne få svar på i forbindelse med fremsøgning af en ressource:

1. Er ressourcen tilgængelig?
2. Hvor, hvornår og hvordan er ressourcen tilgængelig?
3. For hvem er ressourcen tilgængelig?

I forbindelse med ændring af tilgængelighedsvilkår kan bibliotekspersonalet endvidere have behov for at kontakte resourceejere, licensudsteder eller ophavsrethaver.

Rettigheds- og licensmæssige metadata kan derfor opdeles i følgende fire kategorier:

1. Tilgængelighed
2. Tilladte leveringsformer
3. Autorisation og autentifikation
4. Ejer, licensudsteder og ophavsret.

4.3 Typer af rettigheds- og licensmæssige metadata

4.3.1 Tilgængelighed

Tilgængelighed defineres som: Er ressourcen tilgængelig for de brugere, som har ret til at tilgå den. Brugerprofiler m.m. beskrives i afsnit 4.3.3 Autorisation og autentifikation.

En ressource kan være utilgængelig, hvis:

- den er undervejs i en indkøbs- eller indlemmelsesproces
- de tilgængelige fysiske eller elektroniske eksemplarer er udlånt
- en licens er ophørt eller først starter på et senere tidspunkt
- ressourcen er belagt med embargo (karenstid)
- en ophavsrethaver har nedlagt forbud mod, at ressourcen tilgængeliggøres.

Primære krav:

- Det bør fremgå, om en ressource er tilgængelig eller ej, samt hvorfor den eventuelt er utilgængelig.
- Det bør fremgå, hvor ressourcen er tilgængelig, fx via en bestemt URL eller fysisk lokation.
- Det bør fremgå, hvornår ressourcen er tilgængelig.

4.3.2 Ressourcens tilladte leveringsformer

En ressource kan stilles til rådighed via forskellige medier og i forskellige præsentationsformer. Metadata om ressourcens tilladte leveringsformer beskriver, hvordan en ressource må udleveres i henhold til aftale eller lov:

- Må den kopieres elektronisk?
- Må den printes?
- Kan brugeren selv hente ressourcen online hos udbyderen, eller skal det ske via biblioteket?
- Må ressourcen sendes til låneren elektronisk eller som papirkopi?
- Må ressourcen overføres til et e-lærings- eller pensumsystem?
- Må ressourcen indgå i et fjernlån til et andet bibliotek?
- Må ressourcen udlånes, kun udleveres til læsesal, eller slet ikke benyttes?
- Må ressourcen streames eller downloades?

I regi af Digital Library Federation er der lavet anvisninger for tilladte leveringsformer til købte e-ressourcer, som er implementeret i de gængse Electronic Resource Management systemer¹¹.

4.3.3 Autorisation og autentifikation

Metadata om autorisation og autentifikation giver informationer om brugerprofiler, som har adgang til ressourcen og om krav til brugervalidering. Standardprofiler er privatpersoner, studerende, ansatte og "walk-in" brugere (dvs. brugere som fysisk møder op på biblioteket). Desuden kan der differentieres mellem hvilken organisation eller enhed i organisationen, brugeren kommer fra.

Krav til brugervalidering vedrører, hvordan brugeren tilkendegiver sin rolle overfor adgangssystemet.

¹¹ Jf. <http://old.diglib.org/pubs/dlf102/ERMFINAL.pdf> (s. 152)

Indenfor biblioteksdomænet opereres primært med to forskellige valideringskoncepter:

- IP-adresse-baseret validering
- Login-baseret validering.

Ved IP-adresse-baseret validering gives adgang til brugere, som sidder på computere med bestemte IP-adresser, fx en computer på SB eller en computer på Aarhus Universitets campus.

Ved login-baseret validering gives adgang til brugere, som har en konto med godkendte brugerroller hos en godkendt institution, fx brugere som kan logge på Aarhus Universitets login-system, og som har rollen ”studerende”.

Primære krav:

- Adgangsgivende roller skal kunne udledes fra ressourcens metadata.

4.3.4 Ejer, licensudsteder og ophavsrethaver

Metadata om eksemplarejer, licensudsteder og ophavsrethaver beskriver de personer eller organisationer, som ejer, formidler eller har skabt ressourcen. Disse personer eller organisationer skal kontaktes, hvis en aftale om ressourcens anvendelse skal ændres. Det er også disse personer eller organisationer, som kan pålægge det formidlende bibliotek at ændre praksis omkring udstilling af ressourcen.

For fysiske ressourcer er eksemplarejeren det bibliotek, der har indkøbt materialet. Eksemplarejeren kan bestemme, om ressourcen må udveksles til andre biblioteker, og hvilke brugertyper der kan låne materialet, dog under hensyntagen til gældende ophavsret.

Ophavsrethaverne er de personer, som har skabt ressourcen, eller den organisation, som repræsenterer ophavsrethaveren (fx CopyDan). Ophavsrethaveren kan pålægge restriktioner mht. udstilling af ressourcen, fx at ressourcen ikke må være del af en service, eller at bestemte repræsentationer af ressourcen ikke må forekomme (fx forsidebilleder, lytteprøver m.m.).

For købte e-ressourcer er licensudsteder det forlag, som licensen er købt af.

Primære krav:

- Metadata skal indeholde referencer til aftalegrundlaget og ressource-ejeren, fx navnet eller et alias for aftalegrundlaget, som specificerer de rettmæssige aspekter.

5 Rettigheder til metadata

Sidst opdateret 2016-09-06

5.1 Definition

Rettigheder til metadata er afgørende for, hvordan metadata må tilgængeliggøres og udveksles.

5.2 Formål

Det kræver en del fremsynethed, når man skal indgå aftaler vedr. brug og tilgængeliggørelse af metadata.

Når metadata købes eller produceres, er det vigtigt at få klarhed over, hvordan metadata må tilgængeliggøres og udveksles via de services, de er tiltænkt. Det er vigtigt at vurdere brugen på langt sigt, da metadata sjældent udskiftes, når de først er taget i brug.

Rettigheder til metadata er et kompliceret anliggende, som kræver en kontraktvurdering af juridisk kompetente personer ved hver ny aftaleindgåelse. Af samme grund indeholder dette kapitel af metadatahåndbogen ikke anvisninger vedr. sikring af metadatarettigheder ved aftaleindgåelse, men skal være med til at sikre, at rettighedsaspektet er vurderet og indgår. Formålet med kapitlet er begrænset til at fremlægge en række anvendelsesorienterede problemstillinger, som bør vurderes i forbindelse med indgåelse af aftaler om anskaffelse af metadata.

5.2.1 Publiceringskanaler

Et af de primære formål med metadata er at opbygge søgeindekser og søgegrænseflader, som hjælper slutbrugeren med at fremsøge og vælge bibliotekets ressourcer. Det er i SBs interesse, at metadata er helt frie, og at de kan distribueres til download og fri brug af alle interesserede. Dette giver maksimale muligheder for at opbygge søgesystemer, som alle kan benytte, samt for at lade metadata indgå i tredjeparts søgesystemer, som fx Google, Danbib eller WorldCat. Helt frie metadata kræver som regel, at SB ejer metadata, at ophavsrettighederne til metadata er forældede, eller at metadata stilles til rådighed fra tredjepart uden restriktioner.

Nedenstående liste over SBs typiske publiceringskanaler bør indgå ved vurdering af metadataanvendelsen:

1. **SBs portaler:** Må SB lade metadata indgå i SBs indekser og portaler?
2. **API til ekstern høstning:** Må SB opbygge et API¹², hvorfra eksterne interessenter kan høste metadata?
3. **API til ekstern fremsøgning:** Må SB opbygge et API, hvorfra eksterne interessenter kan slå enkelte metadataposter op?
4. **Eksport til andre portaler:** Må SB eksportere metadata, fx til Danbib, hvorved andre biblioteker kan importere metadata, og metadata bliver udstillet via bibliotek.dk? Andre aktuelle eksempler på portaler, som SB eksporterer til, er: EUROPEANA, danmarkshistorien.dk og danskkulturarv.dk.
5. **Indeksering af Google og andre webcrawlere:** Skal der opstilles blokeringer i forhold til Google og andre webcrawlere, så metadata ikke hentes af disse?

¹² API: Application Programming Interface, forkortet API, er en softwaregrænseflade, der tillader et stykke software at interagere med andet software.

6. **Afgrænsede brugergrupper:** Er der restriktioner i forhold til, hvilke målgrupper metadata må udstilles til? De primære målgrupper, som hidtil har været i fokus på SB, er:
- a. Adgang på pligtafleveringsinstitutioner
 - b. Walk-in adgang på SB
 - c. Adgang for forskere
 - d. Adgang for studerende og ansatte ved en videregående uddannelsesinstitution
 - e. Adgang for studerende og ansatte på AU
 - f. Adgang for alle.

5.2.2 Ophavsret

Det er vigtigt at være opmærksom på, at der skal være en afklaring af de ophavsretlige spørgsmål, når man anvender metadata, på samme måde som der skal være en ophavsretlig afklaring i forhold til selve værket.

Et forhold, som ofte forhindrer SBs mulighed for at eksponere metadata, er netop ophavsretslike begrænsninger på indholdet af en ressource. Abstracts, thumbnails af forsider, snippets i form af videosekvenser eller tekstuddrag opfattes på SB ofte som metadata, da det er data, som er med til at beskrive en ressource, og som kan hjælpe brugeren med at vælge en ressource.

Der er også ofte ophavsmæssige begrænsninger til de deskriptive metadata (titel, forfatter, forlag og udgivelsesår m.m.), og her er det vigtigt at skelne mellem, om SB selv har produceret data, eller om data er købt/hentet fra en ekstern database.

Når der indgås aftaler med ophavsrethaver, eller den organisation som repræsenterer ophavsrethaver, er det vigtigt i kontraktvurderingen at gennemgå publiceringskanalerne fra afsnit 5.2.1.

Indsatsområde:

Traditionelt set har der ikke været gjort meget ud af at beskrive og håndhæve rettigheder til Marc-poster¹³ i biblioteksvæsenet, når der først var lavet en brugsretsaf tale, da brugerkredsen i praksis kun var biblioteksvæsenet selv. Men nu, hvor mange biblioteker begynder at overveje at stille data til rådighed for tredje part (Open Data-strategi), er det nødvendigt at undersøge, hvilke muligheder der i realiteten er for dette, bl.a. i forhold til virksomheders ønske om kommerciel anvendelse af metadata.

SB vil derfor, fx via Bibliografisk Råd, presse på for at genbrug af Marc-poster fra Danbib, OCLC¹⁴, LC (Library of Congress) og BL (British Library) bliver underlagt klare og gerne frie og åbne rettighedsaftaler. Eventuelle restriktioner skal tydeliggøres.

5.2.3 Aftaleophør

Ved køb af adgang til baser med metadata eller lignende services er det vigtigt at afklare retten til metadata ved aftaleophør. Har SB fortsat ret til at anvende metadata hentet inden for aftaleperioden, eller skal metadata fjernes fra SBs systemer?

¹³ Marc-poster: metadata struktureret efter Marc-formatet (MACHine-Readable-Cataloguing; se fx <http://www.loc.gov/marc/umb/um01to06.html>)

¹⁴ OCLC: <https://www.oclc.org/about.en.html>

Det er nemt at fjerne metadata, hvis de kun indgår som et berigende element på en søgeside. SB køber fx bogforsider fra en DBC-database med henblik på at berige brugergrænsefladen. Bogforsiderne kan uden den store serviceforringelse fjernes fra brugergrænsefladen, hvis aftalen ophører.

Det er langt sværere at fjerne metadata, hvis de indgår som et vigtigt element i SBs samlinger, fx programdata fra Ritzau eller Marc-poster på elektroniske tidsskrifter og bøger, som købes af den kommercielle leverandør ExLibris.

5.2.4 Henvisning til metadataoprindelse

Ved brug af metadata fra tredjepart stilles der ofte krav vedr. henvisning til metadataejer, fx via et sted på den portal som anvender metadata eller i forbindelse med hver visning af metadata. Krav som disse kan være svære at overholde, specielt når en portal anvender metadata fra mange forskellige metadataejere.

5.2.5 Dokumentation

Kontrakter, som indgås ved køb eller produktion af metadata, skal gemmes på forsvarlig vis, så de kan fremfindes og fortolkes, når nye services udvikles, eller hvis der stilles spørgsmål vedrørende den eksisterende brug.

6 Administrative metadata

Sidst opdateret 2016-09-06

6.1 Definition

Administrative metadata benyttes til at registrere lokale forhold vedrørende anskaffelse, indlemmelse, opbevaring og kassation angående bibliotekets samlinger. Data skal sikre, at man til enhver tid kan se, typisk:

- hvad ejer/lejer/opbevarer biblioteket?
- hvorfor har biblioteket et givet materiale?
- hvordan er det anskaffet?
- hvor befinder materialet sig?
- hvordan må det benyttes og af hvem?
- er det i givet fald udrangeret/afbestilt/bortkommet?

Disse data flyder til diverse fag- og administrative systemer.

De deskriptive metadata (se kap. 2) er også output fra ovenstående proces, men i biblioteksverdenen er der konsensus om at adskille dem fra de administrative metadata, da de deskriptive data angår materialet selv og i princippet ikke afspejler lokale forhold.

I takt med den stigende mængde af elektronisk formidlede ressourcer har de tekniske metadata (kap. 3) og de rettighedsmæssige metadata (kap. 4) imidlertid fået større betydning. Disse data betragtes også som administrative metadata.

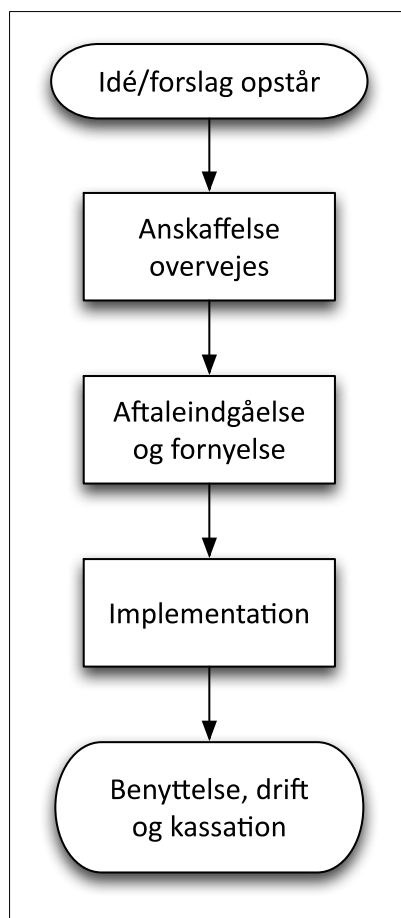
SB ønsker større begrebsmæssig kontrol af de forskellige typer af administrative metadata. I det følgende beskrives derfor de metadata, der falder uden for kategorierne deskriptive, tekniske og rettighedsmæssige metadata.

6.2 Formål

Administrative metadata forsvinder nemt i den enkelte medarbejders mailboks eller lokale drev, hvis der ikke gøres en indsats for at opsamle og systematisere dem. De er vigtige for at kunne lave statistik samt optimere administrative processer. De dokumenterer årsagen til, at et materiale er indkøbt, fravalgt eller kasseret. Det er god administrativ praksis af have styr på de administrative metadata, og de er ofte grundlag for revisionsmæssige rutiner.

6.3 Procesmodel for administrative metadata

Overordnet set anvender man de samme administrative processer i forbindelse med opbygning og tilgængeliggørelse af samlinger på tværs af SB.



Figur 1: Procesmodel for administrative metadata

Processen består af fem trin som illustreret på Figur 1. I bilag ”Administrative metadata” beskrives de enkelte trins dataelementer og den tilhørende systemunderstøttelse for henholdsvis ressourcer i fysisk form, købte e-ressourcer og egne e-ressourcer.

I det følgende beskrives de enkelte trin i modellen.

Første proces er ”Idé/forslag opstår”: Det kan fx være en bruger eller fagreferent, som ønsker, at SB abonnerer på et nyt e-tidsskrift eller indkøber en ny bog. Det kan også være en privatperson eller institution, der donerer en samling lyd- eller billedoptagelser til bevaring for eftertiden. Endelig kan det være igangsættelse af et projekt vedr. etablering, revidering eller berigelse af metadata. Eksempler på administrative metadata i denne proces er kontaktoplysninger, forslagsstiller/giver samt beskrivelse af produktet eller samlingen.

Anden proces er ”Anskaffelse overvejes”: Det skal undersøges, hvorvidt SB allerede har hele eller dele af produktet eller samlingen. Det skal undersøges, om produktet/samlingen falder indenfor SBs formål, om der er økonomi til at købe og indlemme produktet/samlingen, og i hvilken form produktet/samlingen skal indlemmes. For e-ressourcer skal der evt. etableres prøveadgang. Undersøgelsen kan være baseret på en evaluering fra flere medarbejdere/brugere. Eksempler på administrative metadata i denne proces er evalueringer, prøveadgangsopsætning og diverse analyser.

Tredje proces er ”Aftaleindgåelse og fornyelse”: Der skal forhandles priser, finansierungsgrundlag, rettigheder i forhold til anvendelse (licens). Der skal evt. skrives en kontrakt. Specielt for e-

tidsskrifter og e-bøger skal aftalerne fornyes efter et par år, og i den forbindelse skal pris, finansieringsgrundlag og licens genforhandles. Eksempler på administrative metadata i denne proces er priser, finansieringsgrundlag og kontrakter.

Fjerde proces er ”Implementation”: Produktet/samlingen skal leveres, betales, evt. restaureres, have strekkoder og opsættes eller konfigureres. Deskriptive metadata skal oprettes og tilgængeliggøres i søgesystemer, og adgangssystemer skal ajourføres i forhold til licensforhold. Eventuel funktionel bevaring skal opsættes via SBs metadata repository og bitbevaring via Bitmagasinet. Eksempler på administrative metadata i denne proces er faktura, beholdningsdata, opsætning og adgangsforhold.

Femte proces er ”Benyttelse, drift og kassation”: Transaktionsdata og statistik vedr. udlån/download og kassation af ikke bevaringsværdige materialer. Eksempler på administrative metadata i denne proces er cirkulationsdata, udestående gebyrer, brugsstatistik på elektroniske ressourcer, registrering af kasserede materialer og lagerbehov.

Primære krav:

- De administrative metadata skal løbende systemunderstøttes eller dokumenteres, således at kollegaer kan forstå, optimere eller migrere processerne.
- De administrative metadata skal kunne understøtte statistikbrug, således at relevant statistik kan udtrækkes.
- De administrative metadata skal være personuafhængige, således at nye medarbejdere kan indkøres i forbindelse med sygdom, medarbejderudskiftning eller opnormering af stab.
- Af de administrative metadata skal det fremgå, hvordan og under hvilke vilkår et materiale er anskaffet, samt årsagen til eventuelle mangler. Dvs. de administrative metadata danner grundlag for god administrativ praksis og dermed også revisionsmæssige rutiner.

Indsatsområde:

I projekter vedr. etablering, revidering eller berigelse af metadata, er der ofte lavet en stor indsats i form af metadata-analyser, -valg og -mapningsregler. Denne indsats går ofte tabt efter projektafslutning, og konsekvensen er, at nye projekter ikke kan bygge på tidligere valg og erfaringer. Der skal gøres en indsats for at sikre og langtidsbevare ovenstående dokumentation i videst muligt omfang, også retrospektivt. Samtidig skal der udarbejdes retningslinjer for, hvordan og hvor denne dokumentation sikres.

7 Metadata for digitale forskningsdata

Sidst opdateret 2016-11-29

7.1 Definition

Digitale forskningsdata er data der er indsamlet, observeret eller skabt (fx afledte data) med det formål at analysere dem og skabe originale forskningsresultater.

7.2 Formål

Metadata for forskningsdata skal så vidt muligt gøre forskningsdata FAIR (findable, accessible, interoperable and reusable)¹⁵.

7.3 Krav til metadata

Det er en stor fordel, at krav til metadata for forskningsdata bliver udarbejdet allerede ved forberedelsen af et forskningsprojekt, fx ved udformning af ansøgning om støtte.

SB bliver ofte inddraget i et projekt allerede fra starten, typisk fordi projektet skal anvende ressourcer, som biblioteket kan stille til rådighed for forskerne.

I andre tilfælde bliver SB først inddraget, når et forskningsprojekt er afsluttet og forskerne har brug for et sted at opbevare data i en given periode, typisk 5 år (jf. Horizon 2020 Programme).

7.3.1 Registreringsformater

SB anbefaler at man anvender Dublin Core (DC; <http://dublincore.org/>) eller DataCite Metadata Schema (schema.datacite.org) til generel beskrivelse.

Det vil dog ofte være tilfældet, at især repositories til langtidsbevaring eller deling forventer metadata i et bestemt format, og dette vil som regel være afgørende ved valg af format.

SB anbefaler, at man undersøger om der allerede eksisterer fag-specifikke metadata standarder, fx: DDI (www.ddialliance.org/), der anvendes indenfor samfundsvidenskab og humaniora. DDI bruges også af Rigsarkivet.

Uanset format bør nedenstående elementer så vidt muligt altid indgå (se også kap. 2 Bibliografiske metadata).

Primære krav:

- DOI (Digital Object Identifier) til identifikation af data
- Titeloplysning
- Ophavsangivelse
- Fulde navn, ORCID (Open Researcher and Contributor ID) og ansættelsessted/affiliation på alle ophav
- Udgiver (kan fx være et repository).
- Publiceringsdato: bør betegnes i ISO-standard for dato
- Materialetype/datatype
- Krav for at læse og/eller anvende data (se også kap. 3 Tekniske metadata)

¹⁵ H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Version 3.0. 26 July 2016.

Sekundære krav:

- Emneord: vi anbefaler emneord på fagets primærprog, suppleret af engelsk-sprogede emneord (se også afsnit 2.5.9 Emnedata).
- Datering: bør betegnes i ISO-standard for dato
 - Bevaringsperiode
 - Projektperiode
 - Indsamlingsperiode
 - Andre relevante datoer
- Relationer inkl. relationstype
 - Fx "isCitedBy" en publikation (DOI)
 - Fx "isMetadataFor" et andet datasæt
 - Relationer til andre datasæt
- Beskrivelse (uddybende beskrivelse af projektet): vi anbefaler at anvende fagets primærprog til beskrivelser, suppleret af en engelsk-sproget version.
- Rettigheder til data (se også kap. 4. Rettigheds- og licensmæssige metadata): vi anbefaler reference til Creative Commons-licenstype.
- Sprog: bør betegnes i ISO-standard for sprog.

8 Standarder

Sidst opdateret 2016-11-16

8.1 Definition

Den formelle definition på en standard lyder:

”Dokument til fælles og gentagen anvendelse, der angiver regler, vejledning eller karakteristiske træk ved aktiviteter eller ved resultaterne af disse. Dokumentet er fastlagt ved konsensus og vedtaget af et anerkendt organ. Hensigten er at opnå optimal orden i en given sammenhæng.”

(Fra <http://www.ds.dk/da/standardisering/ds-universitetet/hvorfor-laver-vi-standarder>)

8.2 Formål

Dette kapitel af SBs metadatahåndbog beskriver de metadatastandarder, som er de hyppigst benyttede på biblioteket lige nu.

For at lette overskueligheden, har vi forsøgt at give et billede af sammenhængen mellem proces, funktion og standard via flowdiagrammer (se bilag ”Standarder”).

Det er vigtigt at benytte anerkendte standarder og normer, både med hensyn til indhold (indholdsformat) og til struktur (dataformat), når man ønsker at udveksle data med andre brugere og systemer og/eller sikre langtidsbevaringen af data, herunder ikke mindst opfylde behovet for adgang, benyttelse og migrering til ændrede formater over tid.

Der er som regel flere standarder for samme opgave, og uanset hvilken standard man vælger, må man indgå et kompromis. Ingen standard kan dække alle materialer for alle brugere. Man må gøre sig klart, hvilken materialetype der skal beskrives, hvilken detaljeringsgrad man ønsker, hvilken type brugere man har, hvilke typer af standarder, der i forvejen er de hyppigst brugte til den pågældende opgave, osv.

8.3 Typer af standarder

Der findes forskellige typer af standarder, som groft kan grupperes således:

- a. Nationale og internationale standarder (fx ISO-standarder og DS-standarder), eksempelvis:
 - **ISO 2709:2008** Information and documentation - Format for information exchange
 - **ISO 9735:1988** Electronic data interchange for administration, commerce and transport (EDIFACT) - Application level syntax rules.
- b. Sektorspecifikke standarder (herunder NISO, som egentlig er en amerikansk national standard, men ofte anvendes internationalt), fx:
 - [ANSI/NISO Z39.50-2003 \(R2009\) Information Retrieval : Application Service Definition & Protocol Specification](#)
 - **DCMI** (Dublin Core Metadata Initiative)
 - **MODS** (Metadata Object Description Schema)
 - **MARC** (MAchine-Readable Cataloguing)

- MARC i lokal tilpasning:
DanMarc2, som er yderligere modificeret i **SBs bruttoformat**.
- c. Eksempler på standarder til specielle beskrivelser:
 - Radio- og tv-udsendelser: **PBCore** (Public Broadcast Core)
 - Visuelle objekter: **VRA Core** (Visual Resources Association)
 - Datasæt for de sociale videnskaber: **DDI** (Data Documentation Initiative)
- d. De facto standarder (fx leverandørspecifikke standarder, typisk for administrative metadata, katalogiseringsregler o. lign.):
 - **Eksemplarstatuskoder** (strukturen bestemmes af bibliotekssystemleverandøren, men fortolkes lokalt)
 - **Katalogiseringsregler og bibliografisk standard for danske biblioteker**
 - **"Role-based authorization"** (beskrivelse af regler for brugeres adgang til e-materialer)
 - **COUNTER** (statistik over brugen af e-ressourcer).

Primære krav:

- Brug af metadatastandarder og -formater på SB skal dokumenteres. Det er ikke nok at beskrive, hvilken standard der er anvendt i en given situation. Det er nødvendigt også at dokumentere, hvordan standarden er anvendt, altså de lokale fortolkninger.

Indsatsområde:

Det skal afdækkes, hvor der er behov for at forbedre dokumentation for SBs anvendelse af metadatastandarder og -formater.

Samtidig skal udarbejdes retningslinjer for, hvordan og hvor denne dokumentation sikres.

Bilag "Administrative Metadata"

Sidst opdateret 2016-09-06

Administrative metadata for ressourcer i fysisk form

Del af flowchart	Dataelementer	Hvor er data tilgængelige?
Idé/forslag opstår	Fagreferent Input fra AUbibliotek/slutbruger (inkl. reservationskø og bortkommet materiale) SBCI-indkøb	Leverandørbaser Aleph Diverse mails E-sag
Anskaffelse overvejes	Fjernlån eller køb? Pligtaflevering: Definition af afleveringspligt (fx print-on-demand) Sortering af pligtaflevering (hvad indgår) for eks. oplagsudtynding	Aleph Danbib E-sag Internettet
Aftaleindgåelse og fornyelse	Rabatter Kopieringsaftale Regler for brug af AV + radio/TV i fysisk form	Bogkøbsaftale? Copydan Pligtafleveringsloven
Implementering	Leverandør Faktura Bogbinder Adgang/ brug (eksemplar/processtatuskoder)	INDFAK Bogbindermodul/Aleph Aleph
Benyttelse og kassation	Transaktionsdata (lån, bestillinger, mellemværender, brugerregisteroplysninger) Brugsstatistik Betjening af uregistrerede brugere Kassation: pligtaflevering - nej andet – ja	Aleph E-sag Brugerregister Registrering af uregistrerede brugere

Administrative metadata for købte e-ressourcer

Del af flowchart	Dataelementer	Hvor er data tilgængelige?
Idé/forslag opstår	Leverandør Produkt Forslagsstiller	Intota Diverse mailbøxe
Overvejelse vedr. anskaffelse	Leverandør Produkt URL Prøveperiode Adgang: IP eller username/password Tilbagemeldinger	Intota Diverse mailbøxe
Aftaleindgåelse og fornyelse	Licens Platform Pris Intern finansiering Statistik	Intota
Implementering	Platform Leverandør Faktura Statistik URL og username/password Admin URL og username/password	Intota
Benyttelse, drift og kassation	Statistik Fornyelsespriser	Intota Assessment Diverse regneark Diverse mailbøxe DEFFs licensportal Libnet

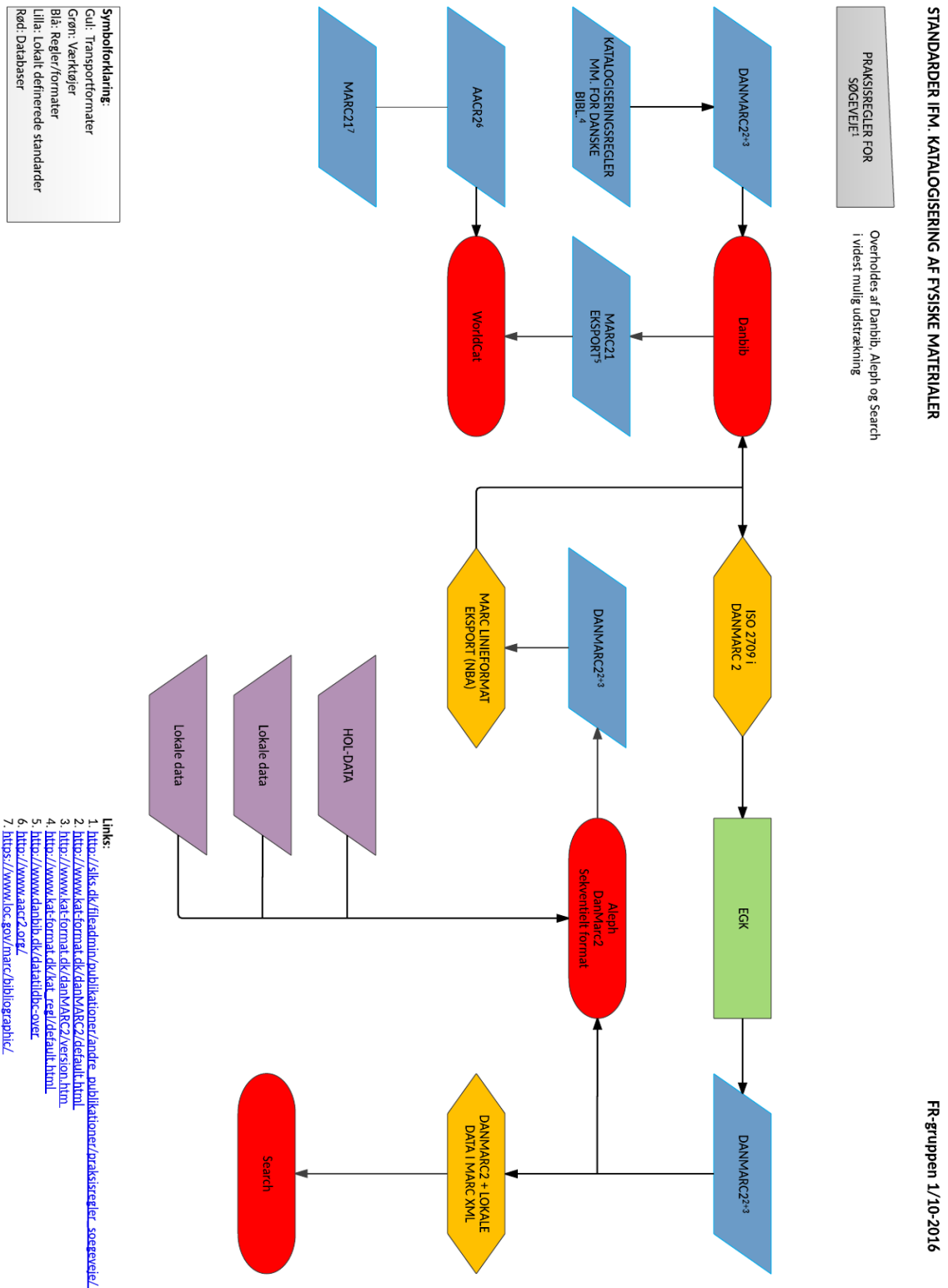
Administrative metadata for egne e-ressourcer

Del af flowchart	Dataelementer	Hvor er data tilgængelige?
Ide/forslag opstår	Kontaktperson/giver Kommisorium/mandat	Diverse mails Medier
Anskaffelse overvejes	Pligtaflevering? Gave? Er der overlap med eksisterende samlinger? Er det bevaringsværdigt? Samlingens størrelse? Samlingens format(er)? Metadata? Juridisk vurdering?	Internettet og andre medier Diverse mails og notater Pligtafleveringsloven
Aftaleindgåelse og fornyelse	Kilde/Giver inkl. kontaktinformation Bevaring Rettigheder Metadatarettigheder Tilgængeliggørelse	Aftaledokument Diverse mails Udbudsmateriale Kravspecifikation
Implementering	Samlingshistorik Indholdsår Digitaliseringsår Størrelse (antal filer, timer, programmer) Formater Kvalitet Placering Samlingsejer Metadata Risici Funktionel bevaring (DOMS) Bitbevaring (Bitmagasinet) Adgang Lovgivning - ifht. adgang Brugergrupper Kvalitetssikring	Modtagelsesblanket Wiki: digitale samlinger Wiki: metadata for digitale samlinger Samlingens bevaringsplan Checkliste for digitalisering Dokumenter med dokumentation af QA process
Benyttelse, drift og kassation	Statistik Antal kopier Antal checksummer Geografisk adskillelse Organisatorisk adskillelse Forskellige teknologiske platforme Checksum-check Kvalitetscheck Max indhentningstid (fra bestilling til data er online) Access Lagerbehov Ingest pr. år Midlertidigt online-lager	Mediestream statistik modul Skema for bitbevaringsniveau

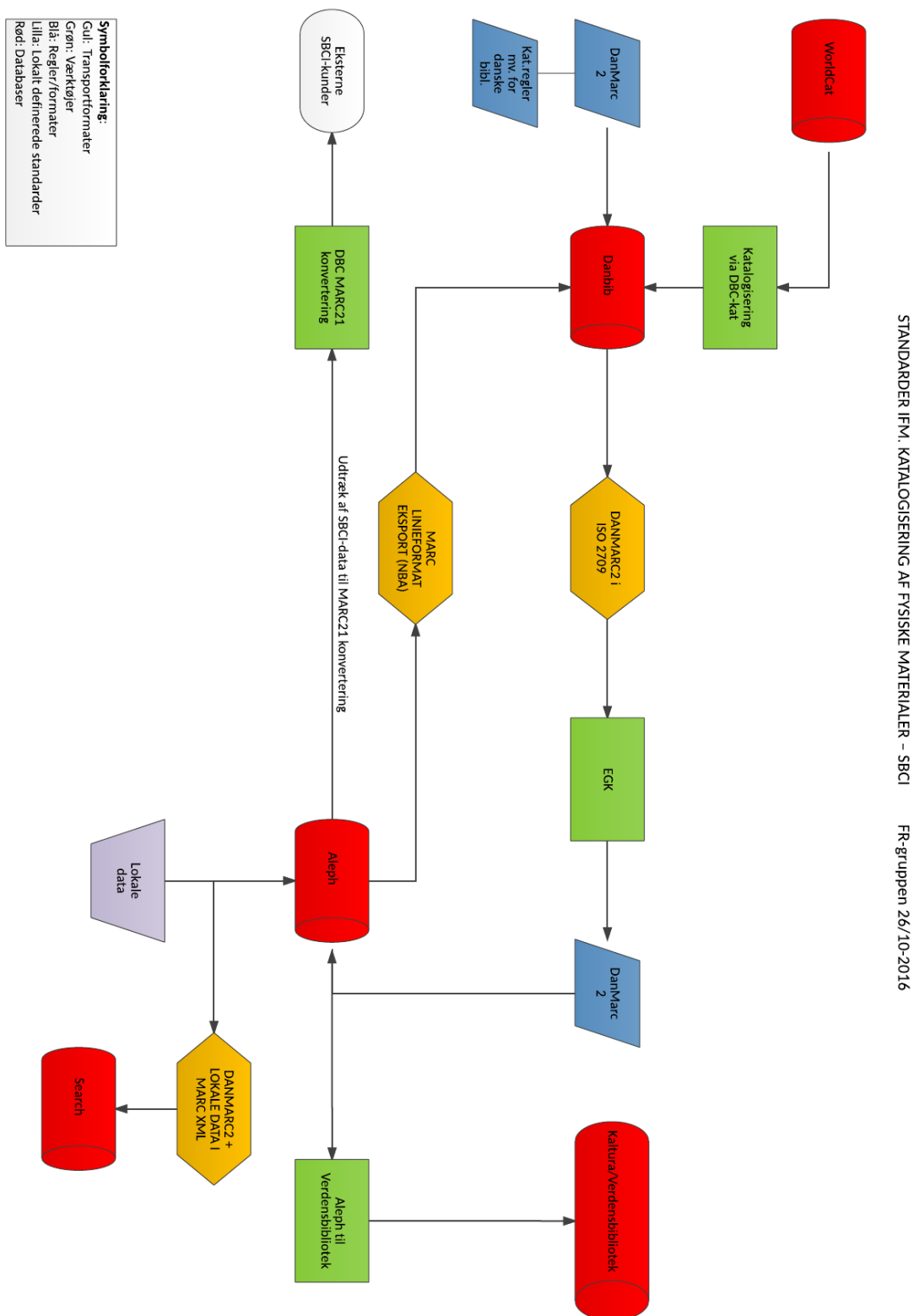
Bilag "Standarder"

Sidst opdateret 2016-11-09

Standarder ifm. katalogisering af fysiske materialer



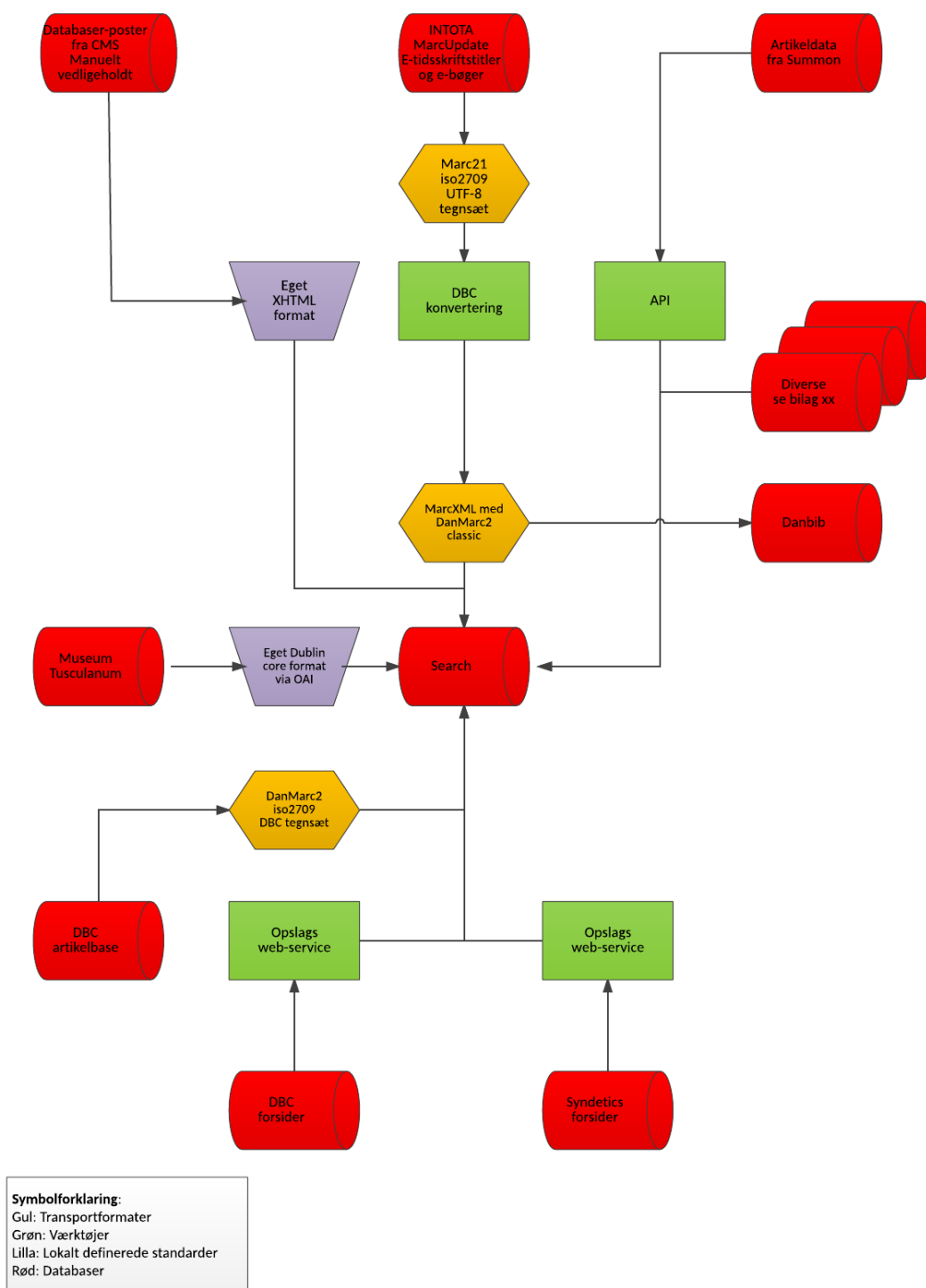
Standarder ifm. katalogisering af fysiske materialer - SBCI



Standarder og formater ifm. købte E-ressourcer

Standarder og formater ifm. købte E-ressourcer

26/10-2016



Oversigt over datakilder og formater i Search/Corporate (ud over datakilder til købte e-ressourcer)

1. OAI-poster fra diverse dataleverandører
Høstes (SB) ved hjælp af protokollen OAI-PMH <http://www.openarchives.org/pmh/> i Dublin Core format <http://dublincore.org/documents/dces/> i et XML-transportformat http://www.openarchives.org/OAI/2.0/oai_dc.xsd
2. PURE-poster fra Aarhus Universitet
Høstes (SB) i mxd-format <http://mx.forskningsdatabasen.dk/mxd/current/ddf-mxd.pdf>
3. De danske aviser – Jette Søllinge (mini-doms)
Indpakning FOXML
<http://fedorarepository.org/sites/fedorarepository.org/files/documentation/3.1/Introduction%20to%20FOXML.html>
4. Bil-årbogen og øvrige digitale årbøger (mini-doms)
oai_dc m.m. pakket ind i FOXML
5. Danske fagbibliografier 1 (ingen tilvækst) (Nordicom – ældre poster)
Proprietært Marc-lignende format pakket ind i MarcXML
6. Danske fagbibliografier 2 (ingen tilvækst) (Nordicom – nyere poster)
danMarc2-poster (med lokale bogstavfelter) pakket ind i MarcXML
7. Danske fagbibliografier 3 (ingen tilvækst) (MissNord)
Proprietært Marc-lignende format pakket ind i proprietært XML-format
8. Danske fagbibliografier 4 (ingen tilvækst) (DPP, DUB, DKB)
danMarc2-poster pakket ind i proprietært XML-format
9. Skæringer på rippede cd'er