

NOTAT

9. september 2005
J.nr.: 331-6
LEA

Bilag H til danZIG-møde 21. september 2005

XML Schema for dansk metadata udveksling

OPDATERET udgave af Bilag C til danZIG-møde 21. september 2005

Indledning

Det fælles projekt Kulturarvsstyrelsen, Biblioteksstyrelsen og Statens Arkiver om ABM-standarder har fået lavet en analyse af anvendelse af Dublin Core til de fælles data. I analysen foreslås en anvendelse af Dublin Core kombineret med Administrative Components – suppleret med nogle ekstra refinements til Dublin Core. Analysen er udarbejdet af konsulenter fra Danmarks Biblioteksskole og de tilhørende XML Schemaer af Index Data. ABM-gruppen der er nedsat af Kulturarvsstyrelsen, Biblioteksstyrelsen og Statens Arkiver vil efter planen i løbet af efteråret udgive en rapport fælles format for arkiver, biblioteker og museer til præsentation for publikum.

Som en del af dette projekt er et XML schema udarbejdet til at kunne rumme data fra de tre sektorer på dette generelle niveau. Det fælles XML Schema kaldet *dkabm* er d.d. endelig teknisk færdigt.

Da skemaet *dkabm* bl.a. rummer Dublin Core modsvarende en mapning fra danMARC2 til Dublin Core, er det oplagt også at benytte dette arbejde som udgangspunkt for en fælles standard for XML formaterede metadata records i den danske bibliotekssektor. Dette spørgsmål lægges ud til offentlig høring ved distribution på danZIG-listen og metadata-dk listerne og drøftes på danZIG-mødet 21. september 2005.

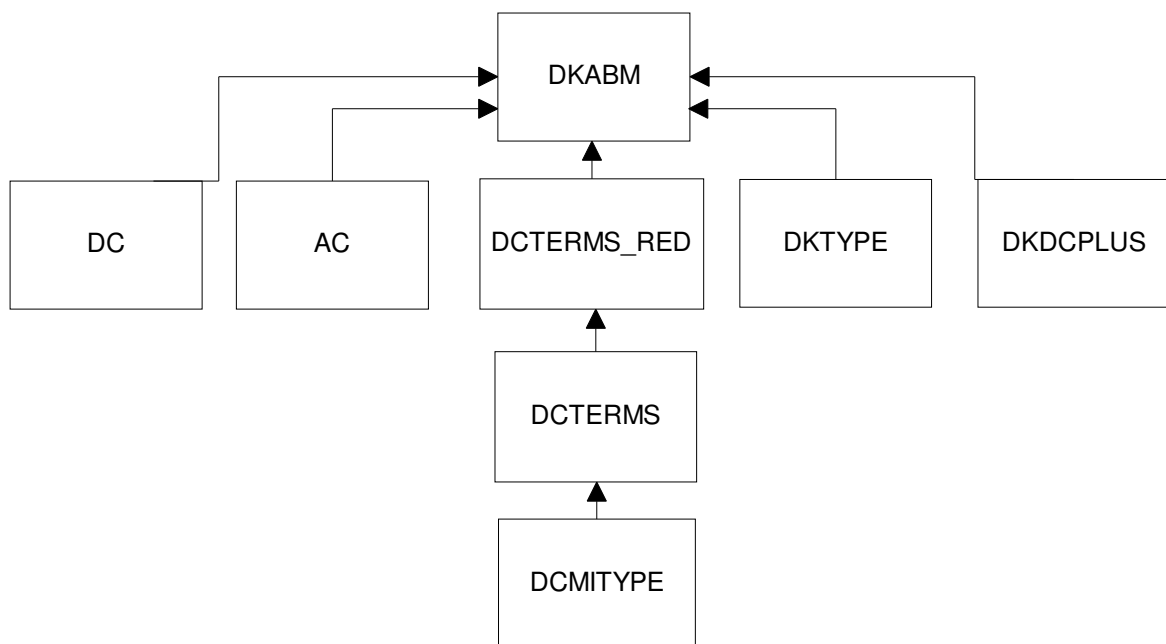
Formålet er at få lokaliseret evt. behov for yderligere tilføjelser (formentlig primært til *dkdcplus*) inden ABM-standard arbejdet er endeligt afsluttet, således at XML Schemaet fra dette projekt rummer alle nødvendige felter.

ABM skemaet

Der er defineret et XML-skema, der beskriver det fælles ABM-udvekslingsformat som hedder *dkabm* og kan findes på (http://www.bs.dk/standards/dkabm_2005-09-09.xsd). Dette skema er bygget op ved at tage dels internationale og anerkendte skemaer og dels skemaer, der er konstrueret til danske forhold, og inkludere dem i et fælles skema. Det drejer sig om følgende skemaer:

- **dc.xsd** (<http://purl.org/dc/elements/1.1/dc.xsd>) - definerer de 15 standard Dublin Core elementer som er gennemgået ovenfor.
- **ac.xsd** (http://www.bs.dk/standards/ac_2005-09-01.xsd) - definerer Dublin Core Administrative Componentets, som er blevet gennemgået ovenfor
- **dcterms.xsd** (<http://purl.org/dc/terms/dcterms.xsd>) - indeholder ekstra Dublin Core elementer, refinements og typer. F.eks. er det her at "temporal" og "spatial" defineres for dc:coverage.
- **ISO639-2.xsd** (<http://www.bs.dk/standards/ISO639-2.xsd>) – er en såkaldt redefinition af schemaet dcterms.xsd fra DCMI, hvor det eneste der er ændret er typen ISO639-2 der definerer gyldige sprogkoder
- **dcmitype.xsd** (<http://purl.org/dc/dcmitype/dcmitype.xsd>) - beskriver en Dublin Core type indeholdende liste over ressource typer
- **dktype.xsd** - indeholder den danske pendant til dcmitype.xsd (se bilagene), der oplister forskellige ressource typer som for eksempel 'Tekst', 'Lyd' eller 'Samling'.
- **dkdcplus.xsd** – indeholder danske udvidelser til dc felterne.

Udover elementerne fra ovennævnte schemer indeholder dkabm også nogle danske udvidelser til Dublin Core schemaerne. Disse elementer benævnes *dkdcplus*, men er ikke et særskilt schema. Der er sket visse justeringer lige før deadline, så denne illustration viser tankegangen omend ikke alle detaljer passer:



Strukturen af skemaerne er som følger: dkabm er hovedskemaet og heri inkluderes dc, ac, ISO639-2, dktype og dkdcplus.

Det skal bemærkes at dkabm schemaet er bygget op ved at dc-elementer er basis og ac elementer ligger i en "indre container" ac, idet der er sammenfald mellem elementnavne i ac og dc. Bemærk også at ac elementer ligger først i schemaet.

Referencer til indhold:

DC, dcterms m.v. - Dublin Core er dokumenteret i *DCMI Metadata Terms*, se: <http://dublincore.org/documents/dcmi-terms/>

Administrative Components er dokumenteret i dokumenteret i AC - *Administrative Components*,

se: <http://www.bs.dk/standards/AdministrativeComponents.htm>

Dansk udvekslingsformat for metadata. DC/AC formatet,

se: <http://www.bs.dk/publikationer/andre/dcac/index.htm>

dkdcplus

Udvidelsen til Dublin Core *dkdcplus* for danske forhold indeholder for nærværende:

DC.Creator	PreferredName	Anvendes til angivelse af vedtagen navneform for ophav
DC.Creator	AlternativeName	Anvendes til angivelse af andre navnformer for ophav
DC.Creator	Temporal med scheme Period	Anvendes til angivelse af tid for et ophavs virke. Der angives startår og slutår.
DC.Descripti on	Version	Version, udgave m.v.

Til DC.Coverage Temporal defineres et scheme **Location** svarende til Box i namespace dcterms.

Location indeholder to components:

type med flg. tilknyttede liste:

Place	stednavn under sogn (fra ADS/ARENA)
Parish	sogn (fra ADS/ARENA)
Unitary	herred (fra ADS/ARENA)
County	amt, gammel amt (fra ADS/ARENA)
Country	Land, geografisk område (fra ADS/ARENA)

name med navnet på enheden

Eksempel: location=County name=Sorø Amt

Til DC.Subject knyttes liste SubjectType over mulige værdier:

DK5	DK5-klassemærke
DBCS	DBC's emneordssystem: skønlitterært emneord
DBCF	DBC's emneordssystem: faglitterært emneord
DBCM	DBC's emneordssystem: musik emneord

Denne liste skal suppleres med klassifikationssystemer fra arkiver og museer inden arbejdet med ABM schemaet er afsluttet.

Felter i ABM skemaet

De lovligte felter i ABM skemaet er listet i nedestående skema. I forslaget er to felter defineret som obligatoriske: DC.Title og AC.Identifier.

Felt	Refinement	Namespace
DC.Title		dc
	alternative	dcterms
DC.Creator		dc
	preferredName	dkdcplus
	alternativeName	dkdcplus
	temporalPeriod med scheme Period (Period fra dcterms)	dkdcplus
DC.Subject	Lister over values SubjectType med værdierne DK5, DBCS, DBCF og DBCM i dkdcplus.	dc
DC.Description		dc
	tableofContents	dcterms
	abstract	dcterms
	version	dkdcplus
DC.Publisher		dc
DC.Contributor		dc
DC.Date		dc
	created	dcterms
	valid	dcterms
	available	dcterms
	issued	dcterms
	modified	dcterms
	dateAccepted	dcterms
	dateCopyrighted	dcterms
	dateSubmitted	dcterms
DC.Type	Lister over values fra dcmi-type og dansk version i dkdcplus	dc
DC.Format		dc
	Extent	dcterms
	Medium	dcterms
DC.Identifier		dc
DC.Source		dc
DC.Language		dc
DC.Relation		dc
	isVersionOf	dcterms
	hasVersion	dcterms
	isReplacedBy	dcterms
	replaces	dcterms
	isRequiredBy	dcterms
	requires	dcterms
	isPartOf	dcterms
	hasPart	dcterms

	isReferencedBy	dcterms
	references	dcterms
	isFormatOf	dcterms
	hasFormat	dcterms
	conformsTo	dcterms
DC.Coverage		dc
	Spatial med Scheme Location	determs (scheme Location fra dkdcplus)
	Temporal med scheme Period	dcterms (scheme Period fra dcterms)
DC.Rights		dc
AC.Identifier		ac
AC.Source		ac
AC.Scope		ac
AC.Comment		ac
AC.Location		ac
AC.Language		ac
AC.Rights		ac
AC.DateRange		ac
AC.Handling	Action Activity Name Email Contact Date Affiliation	ac
AC.Database		ac
AC.Transmitter		ac
AC.Filename		ac
AC.Technicalformat		ac
AC.characterSet		ac
AC.bibliographicFormat		ac
AC.resultFile		ac

Informativt annex

Kommentar / beskrivelse i ABM rapporten om anvendelsen af de enkelte felter. Derfor er nogle af kommentarer tæt knyttet til konvertering fra enten arkiv, bibliotek eller museum. Bemærk at ikke alle elementer i dkabm er med her, da dette i realiteten fungerer som en profil i forhold til det samlede schema – som har hele dc, dcterms og ac inkluderet og hvor ikke alle elementer i dcterms og ac anvendes.

Felt	Refinement	Beskrivelse
DC.Title		Anvendes til angivelse af ressourcens titel. En titel kan være givet til en ressource hvilket ofte vil være tilfældet i arkiverne og museerne ved deres registrering af sager / genstande / arkivalier. En titel kan også være uddraget fra ressourcen hvilket er tilfældet ved bibliotekerne.
	Alternative	Anvendes til angivelse af alternativ titel.
DC.Creator		Anvendes til angivelse af ophav til dokumenter (bibliografisk ophav) og til angivelse af sagoprettende / sagsansvarlige institutioner / personer.
	PreferredName	Anvendes til angivelse af vedtagen navneform for ophav
	AlternativeName	Anvendes til angivelse af andre navnformer for ophav (DAISY)
	TemporalPeriod med scheme Period	Anvendes til angivelse af tid for et ophavs virke. Der angives startår og slutår.
DC.Subject		Angivelse af emnedata. Hvis der er anvendt et velbeskrevet system skal dette angives som "Scheme". I de tilfælde hvor der er anvendt et Scheme som ikke er beskrevet i DC skal der etableres et fast navn for dette.
DC.Description		Anvendes til beskrivelse af ressourcen. Dette kan være i form af abstract, gengivelse af indholdsfortegnelse m.m. Ved registreringer fra Regin vil den tekstuelle beskrivelse af en sag / genstand være angivet her.
DC.Publisher		Udgiver som er ansvarlig for at stille ressourcen til rådighed, som f.eks. forlag. Anvendes kun ved bibliografiske ressourcer.

DC.Contributor		Bidragydere, udover primære ophav angivet i DC.Creator, som f.eks. illustratører eller oversættere. Anvendes kun ved bibliografiske ressourcer.
DC.Date		Anvendes til angivelse af "udgivelsesdato". F.eks. ved genstande i museer kan der være identificeret et bestemt år / dato for skabelse af genstanden.
DC.Type		Anvendes til angivelse af ressourcens genre. I de tilfælde hvor der ikke findes data i den oprindelige registrering angives dette jvf. DCMI Type Vocabulary 2 . Værdien Collection anvendes for alle poster på samlende niveau. Dette af hensyn til præsentation af poster således at den oprindelige sammenhæng i posterne kan bevares.
	Anvendte værdier i scheme dcmi-type: Collection PhysicalObject Text	Se iøvrigt http://dublincore.org/documents/dcmi-type-vocabulary/ for flere typer
DC.Format		Anvendes til angivelse af den fysiske udstrækning eller varighed af ressourcen.
	Extent	
	Medium	
DC.Identifier		Identifikation af den enkelte ressource.
DC.Source		Anvendes til angivelse af værk som ressourcen evt. er baseret på. Anvendes kun ved bibliografiske ressourcer.
DC.Language		Anvendes til angivelse af sproget for det intellektuelle indhold af dokumentet. Anvendes kun ved bibliografiske ressourcer.
DC.Relation		hasPart / isPartOf Anvendes til angivelse af relationer til dele i f.eks. en sag eller ved hoved/bind-poster i bibliografiske ressourcer. Der henvises ved hjælp af identifikation angivet i AC Identifier.
	isVersionOf	
	hasVersion	
	isReplacedBy	
	replaces	
	isRequiredBy	
	requires	
	isPartOf	
	hasPart	

	isReferencedBy	
	references	
	isFormatOf	
	hasFormat	
	conformsTo	
DC.Coverage		Anvendes til angivelse af materialets emnemæssige dækning mht. til tidsperiode og sted. Disse er hentet fra ADS / ARENA med egne tilføjelser (markeret med *) for periode.
	Spatial med Scheme Location	Anvendes til angivelse af materialets emnemæssige dækning mht. til tidsperiode og sted. Disse er hentet fra ADS / ARENA.
	Temporal med scheme Period	tekstlig beskrivelse for periode med mulighed for start og slut
DC.Rights		Anvendes til oplysninger om eventuelle klausuler vedrørende adgang til materialet.
AC.Identifier		Post-identifikator
AC.Source		Registrerende entitet
AC.Scope		Ukvalificeret tænkes elementet anvendt til angivelse af det bredere formål med at registrere den pågældende resource, fx. Nationalbibliografi, Netkatsamarbejde mv.
AC.Comment		Skal også kunne anvendes i tilknytning til AC.activity
AC.Location		Bruges kun hvis metadata ligger et andet sted
AC.Language		Materialets sprog
AC.Rights		Rettigheder
AC.DateRange		Gyldighed
AC.Handling		Anvendes til beskeder om fremtidige handlinger vedr. posten
AC.Activity		Anvendes til oplysninger om hvad der er sket med posten
AC.Name		Person eller korporation
AC.Email		Email-adresse på person eller korporation
AC.Contact		Fx. Postadresse, tlf.nr. eller URL
AC.Date		Dato for handlingen
AC.Affiliation		Ansvarlig institution
AC.Database		Navn/kode for base som batchfil skal sendes til
AC.Transmitter		Navn/kode for person/korporation
AC.Filename		Batchfilens navn
AC.Technicalformat		Fx ISO2709, XML, HTML