
ABM standard arbejdsgruppen

nedsat af Statens Arkiver, Styrelsen for Bibliotek og Medier og Kulturarvsstyrelsen

Titel : **Specifikationer for fælles præsentation af data fra arkiver, biblioteker og museer på internettet**

Dato : 2009-09-12

Status : Gældende ABM-specifikation

Redaktør: Leif Andresen
Email: LEA@bibliotekogmedier.dk

Publicering: <http://biblstandard.dk/abm/>

Specifikationer for fælles præsentation af data fra arkiver, biblioteker og museer på internettet

Indledning

Det digitale samarbejde blandt arkiver, biblioteker og museer (ABM) blomstrer som aldrig før. Der er grøde på det lokale og det regionale niveau. På landsplan kortlægges institutionernes samlinger af bøger, fotos, lyd, film, arkivalier, genstande m.v., og der gøres seriøse overvejelser omkring nationale strategier for digitalisering og for den efterfølgende bevaring og formidling af den digitaliserede kulturarv på tværs af sektorer og institutioner. Der er således spundet tråde til visioner om det fremtidige digitale samarbejde.

Når visionerne skal føres ud i livet er det vigtigt, at grundlaget er i orden. En lille men ikke uvæsentlig del af fundamentet beskrives i det foreliggende "specifikationssæt", der baserer sig på indholdet i rapporter publiceret i 2003 og 2006, og som er en færdiggørelse af et arbejde igangsat januar 2003.

Introduktion til specifikationssæt

Fælles ABM indholds- og transportformat

Der er defineret et XML schema, der beskriver de tilladte elementer i det fælles ABM-udvekslingsformat. Schemaet hedder dkabm_types. Dette schema er bygget op ved at tage dels internationale og anerkendte skemaer og dels schemas, der er konstrueret til danske forhold, og inkludere dem i et fælles schema.

Profil for udveksling af ABM-data

Dette dokument knytter sig til Fælles ABM indholds- og transportformat og beskriver den praktiske anvendelse af XML schemaet i det fælles format ved udveksling af data fra baser for arkiver, biblioteker og museer til fælles præsentation.

Transport af ABM data

Her bliver specificeret hvordan ABM data (i DKABM XML schemaet) kan indpakkes når de skal transporteres som filer, f.eks. på en DVD eller via FTP, eller indlejres i allerede eksisterende formater, såsom høstning med Open Archive Initiative (OAI) samt som svar på søgning med Search and Retrieval via URL (SRU).

Mapning ARKIBAS 4 til DKABM

Tabel med mapning fra lokalarkivernes ARKIBAS 4 til DKABM.

Mapning Daisy til DKABM

Tabel med mapning fra Statens Arkivers Daisy til DKABM.

Mapning danMARC2 til DKABM

Tabel med mapning fra bibliotekernes danMARC2 til DKABM.

Mapning Regin til DKABM

Tabel med mapning fra museernes Regin - Dansk Museums Dokumentations Standard (Museernes Samlinger) – til DKABM.

Aktuelle links til specifikationerne på <http://biblstandard.dk/abm/>

Baggrund

Kulturministeriet nedsatte i januar 2003 en styregruppe og arbejdsgruppe, der fik til opgave at komme med anbefalinger omkring det digitale samarbejde blandt arkiver, biblioteker og museer. Det skete dengang på baggrund af lokalt igangsatte projekter om udvikling af fælles ABM søgebaser. Arbejdsgruppen fik til opgave at vurdere fælles definitioner på dataindhold, dataformater og datatransport med henblik på at kunne etablere fælles præsentation af data på internettet. Det var således formålet at fastlægge og definere en fælles ramme således at de relevante informationer kunne præsenteres i sammenhæng for "den alment interesserede" borger.

Arbejdsgruppen tog udgangspunkt i de lokale forsøg for at kunne drage nytte af de allerede opnåede erfaringer og for at sikre, at kommende initiativer kunne bygge videre på dem. Formålet var også at minimere risikoen for, at der blev brugt unødvendige ressourcer ved efterfølgende lokale og regionale forsøg eller at disse udviklede sig på en sådan måde, at den mulige fremtidige integration blev vanskeliggjort.

Det var derfor arbejdsgruppens ambition at stille en "værktøjskasse" til rådighed, der både kunne tages i anvendelse ved fremtidige lokale og regionale ABM samarbejder og som også kunne skaleres til at blive brugt på nationalt plan.

De tre direktører for Statens Arkiver, Biblioteksstyrelsen og Kulturarvsstyrelsen, Johan Peter Noack, Jens Thorhauge og Steen Hvass udgør projektets styregruppe, mens arbejdsgruppen fik følgende sammensætning:

- Anders Sode-Pedersen, Statens Arkiver
- Jens Topholm, Aalborg Stadsarkiv
- Leif Andresen, Biblioteksstyrelsen
- Henning Midtgaard Hanssen, Statsbiblioteket
- Eske Wohlfahrt, Kulturarvsstyrelsen
- Henrik Jarl Hansen, Kulturarvsstyrelsen (formand)

Arbejdsgruppens første og foreløbige rapport fra 2003 indeholder forslag til rammer for standardisering af:

- Dataindhold
 - Her anbefales Dublin Core (DC)
- Dataformater
 - Her anbefales XML som udvekslingsformat
- Datatransport
 - Her anbefales Z.39.50, SRW og OAI-PMH
- Fælles præsentation (herunder selektion og emnesøgning)
 - Her fremlagdes overvejelser om en struktureret liste over emneord

Rapporten inkluderede også udkast til mapning mellem de sektorspecifikke databasers format til Dublin Cores 15 metadata-elementer: *Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights*.

Arbejdsgruppens første og foreløbige rapport fra 2003;
ABM Arbejdsgruppe. Afrapportering fra arbejdsgruppen om ABM-standarder til styregruppen, august 2003

<http://www.kulturarv.dk/tjenester/publikationer/emneopdelt/kulturarv-it/abm2003/index.jsp>

På baggrund af 2003 rapporten videreudviklede arbejdsgruppen anbefalingerne i en efterfølgende, afsluttende rapport fra 2006. Der blev således arbejdet videre med at raffinere håndteringen af metadata, dvs. konvertering af sektorspecifikke standarder til Dublin Core samt at overføre det resulterende Dublin Core univers til et sammenhængende XML-schema. Til at forstå denne del af arbejdet blev der indgået aftale med Danmarks Biblioteksskole og Index Data.

2006 rapporten fremlægger en samlet beskrivelse af dataindhold og dataformater i form af en overordnet model og en efterfølgende udfoldning af konverteringsbeskrivelser og et fælles XML-format. Endvidere indeholder rapporten nogle generelle overvejelser om datatransport og fælles præsentation.

Ved mapning fra de sektorspecifikke databaser til det valgte fælles format (Dublin Core) tages der udgangspunkt i formater, der er udviklet indenfor de enkelte ABM-sektorer, og som derfor er præget af forskellige faglige traditioner. Det giver sig bl.a. til udtryk i registreringens niveau og detaljeringsgrad. Mapningen er foretaget for:

- Lokalarkiver: *Arkibas 4*
- Statens Arkiver: *Daisy*
- Bibliotekerne: *DanMARC2*
- Museerne: *Dansk Museums Dokumentations Standard (Museernes Samlinger)*

Formålet med mapningen er at definere en fælles ramme, således at data fra arkiver, biblioteker og museer kan præsenteres i en sammenhæng. Mapningen til Dublin Cores 15 elementer giver i sagens natur en række begrænsninger i forhold til de originale, sektorspecifikke databaser. Det er således ikke alle informationer fra ABM institutionernes databaser, der vil kunne indgå i den fælles ramme, men så vidt muligt alle de data, der har betydning i forhold til borgervendt præsentation og søgning. Fordi målgruppen er bredere, er der heller ikke krav om at detailrigdommen skal ligge på samme niveau som i de faglige databaser. Men det skal dog afklares, hvordan man efter at have fundet en relevant post kan komme videre til de sektorspecifikke databaser, og dér få adgang til den fulde funktionalitet og alle informationer.

Med denne model har det været nødvendigt at acceptere et vist tab af informationer, da det som nævnt kun er udvalgte dele af de originale data, der kan repræsenteres i Dublin Core. Udfordringen har bestået i for hver database at definere hvor meget og til hvilket niveau data skulle konverteres. Udgangspunktet var:

- Det er kun informationer, der relevante for offentlig søgning, der konverteres til det fælles format.
- Der er ikke samme krav til at præsentere detaljer i den fælles præsentation som i de originale databaser.
- Det er vigtigt, at man får en henvisning til de originale data.

Når begrænsningerne er belyst, må det samtidig understreges, at mapningen af de sektorspecifikke databaser til Dublin Core åbner for helt nye sammenhænge på tværs af sektorerne og giver nye muligheder for samsøgning i de faglige databaser.

Med det rapportens forslag til en Dublin Core mapning af ABM-formater blev der givet et bud på et værktøj, der eksempelvis kan anvendes til en fortsat udvikling og implementering af lokale/regionale søgedatabaser. Men anbefalingerne kan også finde anvendelse på nationalt plan selvom udviklingen af de landsdækkende sektorspecifikke baser på nuværende tidspunkt ikke er rettet mod en fælles national ABM søgning. Den foreslåede mapning kan således både bruges i praktiske forsøg med lokale/regionale baser og i forbindelse med udvalgte data fra nationale baser. Dermed er der også åbnet for en afklaring af muligheder og perspektiver for ABM søgning på nationalt plan.

På baggrund af arbejdet med Dublin Core og XML schemas kommer arbejdsgruppen med følgende anbefalinger i 2006 rapporten:

Formater

Indholdsformat

- Dublin Core anbefales som et fælles, konkret indholdsformat, hvortil man kan mappe de udvalgte dele af de enkelte sektors specifikke databaseindhold, som set fra almenhedens synsvinkel er relevante for præsentation og søgning.
- Det må anses for påkrævet at præcisere eller kvalificere nogle af DCs 15 grundelementer med såkaldte "refinements". Der bør desuden inddrages visse

elementer (i første række "identifier" og "source") fra DC's Administrative Components.

Rammeformat

- Rammeformatet eller udvekslingsformatet bør være XML, som er en W3C standard. Der er i tilknytning til rapporten udarbejdet et XML schema, der kan rumme de fælles ABM-dataelementer.
- Dette schema kaldes DKABM og det inkluderer dels internationale og anerkendte schemas, dels schemas, der er konstrueret til danske forhold. Sidsnævnte omfatter dkdplus (danske udvidelser til DC bl.a. forskellige emneordslister til brug i elementet "subject").

Datatransport

- Der foreslås en central model med en fælles database, der til stadighed opdateres (harvesting) Den standard for høst af dokumenter og/eller metadata fra nettet, som det på nuværende tidspunkt er mest nærliggende at anbefale er OAI-PMH.
- Alternativt til den centrale model kan man forestille sig en web-service model baseret på den nye protokol SRW, der tillader søgninger på tværs af arkiver, biblioteker og museer. Det må derimod anses for at være for ressourcekrævende for arkiver og museer at implementere Z.39.50, som ellers er en gammel kending fra biblioteksverdenen.

Fælles præsentation

Selektion

- Selektion er en meget relevant problemstilling i arbejdet med ABM data og databaser. Der bør i alle tre institutionstypers materialer kunne foretages afgrænsninger. Det bør eksempelvis være muligt at afgrænse data af intern / forskningsmæssig / administrativ art, der ikke er beregnet for offentligheden.

Søgning

- Rapporten peger også på nytten af en struktureret ABM-emneordsliste og udvikling af fælles standarder for beskrivelse af tid og sted

Arbejdsgruppens afsluttende rapport:

Fælles præsentation af data fra arkiver, biblioteker og museer på internettet

Rapport fra arbejdsgruppen om ABM-standarder, februar 2006.

<http://www.kulturarv.dk/tjenester/publikationer/emneopdelt/kulturarv-it/abm2006/index.jsp>

Rapporten blev efterfølgende sendt i offentlig høring. På baggrund af de overvejende positive reaktioner og modtagne konstruktive forslag, har arbejdsgruppen om ABM standarder kunnet færdiggøre det arbejde, der blev påbegyndt i 2003 om at vurdere fælles definitioner på dataindhold, dataformater og datatransport med henblik på at kunne etablere fælles præsentation af data om materialer fra arkiver, biblioteker og museer.

Indkomne hørings svar:

Høringssvar på rapporten "Fælles præsentation af data fra arkiver, biblioteker og museer på internettet":

<http://www.kulturarv.dk/tjenester/publikationer/emneopdelt/kulturarv-it/abm2006/hoeringssvar.pdf>

Ved færdiggørelsen af specifikationssættet er der sket mange mindre justeringer af tabeller og tekst og der er sket en redaktionel ændring fra en samlet rapport til et antal specifikationer, der hver for sig enkelt kan opdateres, når det bliver relevant. Endvidere er tilføjet en specifikation af transport af data med beskrivelse af XML indpakning i forskellig sammenhæng.

Alle tabeller m.v. i 2006 rapporten skal således alene ses som udkast og anvendelse skal bygges på specifikationerne.

Set ud fra en helhedsbetragtning er det her fremlagte specifikationssæt blot en af mange brikker i ABM-puslespillet, men det er en af de vigtige at få på plads for at kunne gøre arbejdet færdigt.

Specifikationssættet giver et konkret bud på, hvordan vi på tværs af sektorerne kommer videre mod det samlede overblik over de materialer, der opbevares på ABM institutionerne, og som sådan er det et væsentligt, konkret bidrag til de aktuelle overvejelser omkring digitalisering, bevaring og formidling af kulturarven.

Vedligeholdelse af specifikationssættet

Styregruppen godkendte 15. oktober 2007 et sæt specifikationer.

Samtidig besluttede styregruppen at ABM standard arbejdsgruppen med uændret sammensætning blev gjort permanent med det løbende ansvar for at vedligeholde *Specifikationer for fælles præsentation af data fra arkiver, biblioteker og museer på internettet*.

En ny version af alle dele skete august 2009 og den er beskrevet her:

http://biblstandard.dk/abm/doc/Justering_ABM-dokumenter_august_2009.pdf

En ny version 2011 er alene sket med en 2011 udgave af *Fælles ABM indholds- og transportformat* og med nye schemaer med en 2011-dato. Se nærmere på forsiden

<http://biblstandard.dk/abm/>