

National strategi for autoritetsdata

Indhold

Hvorfor en national autoritetsdata-strategi?	3
Visionen	4
Mål	4
Indsatsområder	5
Samarbejde om entydig identifikation	5
VIAF - genbrug og internationalt samarbejde om autoritetsdata	6
Nationalt autoritetsdata repositorium med sammenhæng til VIAF	7
Sammenhæng mellem autoritetsdata og bibliografiske data i bibliotekernes infrastruktur	7
Linked open data	7
RDA - international dataudveksling i fremtiden	8
Funktionelle krav til autoritetsdata	8
Typer af autoritetsdata	9
Implementering af strategi	11
Det nationale autoritetsdatarepositorium (NAR)	11
Forklaringer til figuren	12
URI som ID for autoritetsentiteter	12
Referencer til andre identifikatorer	13
Autoritetsdata og RDA. Typer af autoritetsdata	13
Genbrug	13
Regler	14
Dansk hyldeopstilling og standardtitler	14
Formatet	14
Praktisk brug af autoritetsdata	14
Prioritering af typerne – rækkefølge for implementering mv.	15
Implementeringsprojekter	16
Regler og best practice	16

National strategi for autoritetsdata - Version 2.0 (2016)

Det nationale repositorium - tekniske løsninger	16
Øvrige tiltag	16
Bilagsfortegnelse	17

Hvorfor en national autoritetsdata-strategi?

Bibliografisk Råd har igangsat arbejdet med udarbejdelsen af en national strategi for autoritetsdata. Hensigten er dels at få fokus på de udfordringer i det bibliografiske arbejde med autoritetsdata, som vi står overfor i dag, og som vi mener, er ved at ændre sig radikalt, dels at få anbefalet et antal indsatsområder, som kan gøre os i stand til at imødegå disse udfordringer.

Biblioteksverdenen arbejder i dag med udgangspunkt i FRBR¹, hvor hensigten med autoritetsdata er, at man skal sikre autoriserede søgeindgange på de entiteter, der bliver identificeret i katalogiseringsprocessen, således at man samler eller relaterer data. Dette er principper, som er rigtig gode for de brugere, som skal søge og identificere personer, materialer osv. i en katalog eller database, og som vi ved mange års erfaringer med at producere bibliografiske data har kunnet se et stort potentiale i at lave.

Autoritetsdata er dog ikke udelukkende et begreb, som vedrører valg og udformning af en autoritativ form i et bibliografisk univers, men også et begreb, som dækker identificering af identiteter og deres indbyrdes fælles tilhørsforhold og relationer i mange andre sammenhænge som f.eks. forskere, geografiske lokaliteter etc.

Nationale digitaliseringsprojekter og webkataloger har givet brugerne mange flere muligheder på nettet. Det betyder, at søgeadfærden for vores brugere er i en digital kontekst og behovet for at få understøttet søgninger i denne sammenhæng er en udfordring for bibliotekerne. Set i en autoritativ kontekst vil opgaven være at sikre adgang til globale autoriteter og skabe relationer, som kan give merværdi og støtte navigation. Brugerne forventer, at bibliotekerne er en autoritativ kilde.

Den internationale biblioteksverden står overfor at skulle implementere et nyt regelsæt, RDA², som er afløseren for det anglo-amerikanske regelsæt, AACR³, som de nuværende danske katalogiseringsregler bygger på. Grundprincipperne for RDA følger principperne i FRBR, hvilket indebærer, at der i RDA er fokus på, at der skal beskrives og tildeles "access points" til forskellige elementer (entiteter) vedrørende det registrerede materiale, samt på at angive entiteternes indbyrdes relationer, fremfor, som de hidtidige regler foreskriver, at opbygge en fuldstændig bibliografisk poststruktur for hvert enkelt biblioteksmateriale. I den sammenhæng får autoritetsdata en vigtig rolle, da man for at navigere i en sådan datastruktur, mange gange bør have adgang til en autoritativ indgang for at komme videre. For bibliotekerne er det desuden væsentligt, at der skal være mulighed for at opbygge et sådant hjælpemiddel på en rationel måde, hvilket RDA gerne skulle kunne understøtte.

Andre internationale initiativer hvor autoritetsdata spiller en væsentlig rolle er BIBFRAME⁴ og Schema.org⁵.

¹ FRBR (Functional Requirements for Bibliographic Records): <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

² RDA (Resource Description and Access: <http://www.rda-jsc.org/rda.html>

³ AACR2 (Anglo-American Cataloguing Rules, 2. edition): http://en.wikipedia.org/wiki/Anglo-American_Cataloguing_Rules

⁴ BIBFRAME (Bibliographic Framework Initiative): <http://bibframe.org/>

⁵ Schema.org: <http://schema.org/>

National strategi for autoritetsdata - Version 2.0 (2016)

For BIBFRAMEs vedkommende er hensigten at imødekomme udviklingen indenfor web-teknologien/arkitekturen ved at give fleksible metoder til håndtering og formatering af fremtidens katalogisering, sikre linked data og sikre fælles metoder til overførsel af bibliografiske informationer etc. I denne model indgår autoriteter også som en hoveddel af universet, og de udgøres her af et antal veldefinerede entiteter med relationer til et værk og dens manifestation. Entiteterne kan være personer, steder, emner, organisationer, med relation til værket/manifestationen. Disse autoriteter skal sikre effektive og troværdige pegepinde, som kan hjælpe til med at navigere og frembringe relaterede entiteter. BIBFRAME-autoriteter er dog ikke tiltænkt at skulle erstatte allerede eksisterende autoriteter men tænkes indskudt som et ekstra "abstraktionslag", der kan samle indgang til flere forskellige web-baserede autoriteter.

Schema.org er ikke udsprunget af biblioteksdomænet, men af et samarbejde mellem nogle af de største søgemaskineaktører, bl.a. Bing, Google, Yahoo! og Yandex, og det har til formål at skabe fælles rammer for udnyttelse af mulighederne i det semantiske web. For bibliotekerne er berigelse af data med opmærkning i schema.org en oplagt mulighed for at bringe bibliotekernes velkurerede data - herunder ikke mindst autoritetsdata – i spil i en større kontekst på nettet.

Der er således mange områder, hvor autoritetsdata er i stigende fokus, og det er vores vurdering, at det er nødvendigt med en national strategi for området, hvis vi som samlet dansk biblioteksvæsen skal kunne møde de udfordringer vi står over for. Autoritetsdata bør være en del af det metadatasamarbejde, som der er en lang og god tradition for i de danske biblioteker.

Det er håbet, at vi med fastlæggelse af en national strategi kan bidrage til at få identificeret de konkrete udfordringer og til at få påpeget de væsentligste indsatsområder for de kommende år – samtidig med at vi undervejs skal være åbne for at rette ind efter udviklingen, da der til stadighed ellers kan være fare for, at teknologien vil overhale os.

Visionen

Visionen er, at man med etablering af et nationalt autoritetsdatasamarbejde kan optimere resource discovery for bibliotekernes slutbrugere og samtidig understøtte såvel internationalt som nationalt samarbejde om skabelse og genbrug af autoritetsdata indenfor bibliotekssektoren – på kortere og længere sigt.

Visionen er også, at et sådant samarbejde kan realiseres således, at bibliotekernes autoritetsdata fremadrettet har en karakter og en tilgængelighed, der også kan gøre dem attraktive for aktører udenfor biblioteksdomænet.

Mål

Til opfyldelse af visionen kan der formuleres en række konkrete mål samt peges på, hvordan disse mål kan realiseres.

Følgende er de vigtigste overordnede formål med autoritetsdata:

Øge brugernes mulighed for resource discovery i bibliotekernes data ved at integrere yderligere metadata, der relaterer sig til de beskrevne bibliografiske objekter. Eksempler på sådanne supplerende data kan være biografiske oplysninger om en forfatter, koordinaterne på en lokalitet, emners semantiske relationer etc.

Give åben adgang til data, dvs. gøre data tilgængelige, såvel teknisk som rettighedsmæssigt, på en måde der vil gøre det muligt såvel for biblioteker som for aktører uden for biblioteksdomænet at bidrage til og udnytte nationale autoritetsdata.

Understøtte éntydig identifikation af enkelte entiteter, der indgår i de bibliografiske data, det være sig personer, korporationer, værker, lokaliteter, emner eller andet.

Dette er helt centralt i forhold til den måde, hvorpå man fremover vil skulle strukturere bibliografiske data, hvad enten der er tale om RDA eller BIBFRAME. Brugs-mæssigt vil dette også være af stor betydning i forhold til f.eks. rettighedsstyring (hvem skal krediteres for hvad?) eller søgning (find alle repræsentationer af den samme entitet – uanset det konkrete tekstlige udtryk i metadata)

Fremme samarbejde om metadataskabelse - internationalt såvel som nationalt - ved at sammenstille variantformer af benævnelser for den samme entitet på tværs af forskellige katalogiseringsregler/-praksis og sprog, vil det være muligt at genbruge andre institutioners autoritetsdata uden nødvendigvis at gå på kompromis med egne regler/egen praksis

Rationalisere metadataskabelse bl.a. ved at give mulighed for automatisk at tilføje relationer til alternative former når opslagsform er valgt.

Indsatsområder

Ud fra disse overordnede mål kan der peges på en række strategiske satsningsområder for et nationalt samarbejde om autoritetsdata i Danmark.

Samarbejde om éntydig identifikation

En væsentlig funktion for autoritetsdata er at understøtte den éntydige identifikation af en entitet – det være sig en person, en organisation, en lokalitet eller andet.

Der findes i dag 2 væsentlige internationale tiltag, der understøtter en éntydig identifikation af personer: ISNI⁶ og ORCID⁷. Begge systemer har som formål at tildele et persistent ID til personer, men hvor ISNI overvejende har fokus på skabere af kreative værker, har ORCID primært sit sigte mod forskere og forskningsregistreringens behov. ISNI og ORCID har i januar 2014 underskrevet et "Memorandum of Understanding" der har til formål at sikre sammenhængen – herunder den tekniske interoperabilitet mellem de 2 systemer.

⁶ ISNI (International Standard Name Identifier). For en nærmere beskrivelse se: <http://www.isni.org/>

⁷ ORCID (Open Researcher and Contributor ID). For en nærmere beskrivelse se: <http://orcid.org/>

National strategi for autoritetsdata - Version 2.0 (2016)

Der er i DEFF-regi givet tilskud til et projekt, der skal få alle forskere i Danmark ført ind i den globale forskerdatabase ORCID⁸. Det er DTU (Danmarks Tekniske Universitet), der er koordinator på projektet, der tæller 6 universiteter, Kulturministeriets forskningsinstitutioner ved Det Kongelige Bibliotek samt alle UC'erne (University Colleges) via UC-Viden.

Der er derimod, så vidt vides, pt. ingen konkrete biblioteksrelaterede ISNI-projekter på nationalt niveau.

Den nationale autoritetsdataindsats skal omfatte allerede eksisterende initiativer vedr. tildeling af persistente ID'er til personer som ISNI og ORCID, herunder fastlægge hvorledes ISNI skal indgå i danske nationalbibliografiske autoritetsdata.

I det omfang, der internationalt findes andre veletablerede og udbredte systemer for persistente og unikke ID'er for andre entiteter end personer, bør disse også anvendes i danske autoritetsdata, og en nærmere analyse af området er derfor også et vigtigt indsatsområde.

VIAF - genbrug og internationalt samarbejde om autoritetsdata

Et vigtigt potentiale ved autoritetsdata er mulighederne for genbrug og internationalt samarbejde om data.

Det danske nationale samarbejde om autoritetsdata bør derfor foregå i tæt sammenhæng til det internationale samarbejde på området.

I den forbindelse spiller **VIAF**⁹ en helt central rolle.

Udover, at man her har etableret en fælles platform for koordinering og publicering af en lang række landes nationale autoritetsdata, kan man også gennem VIAF-samarbejdet tilføje ekstra værdi til de nationale autoritetsdata: VIAFs match af autoritetsdata fra de deltagende lande/organisationer gør det muligt for den enkelte deltager efterfølgende at hente og genbruge data fra andres autoritetsposter. Som en væsentlig faktor skal det nævnes, at et samarbejde mellem VIAF og ISNI sikrer adgang til ISNI via VIAFs autoritetsdata. Også andre berigelser, der foretages direkte i VIAF, fx links til Wikipedia, vil kunne genbruges og indgå i de nationale data.

Gennem deltagelse i VIAF-samarbejdet gives der således muligheder for at berige og udbygge data til gavn for bibliotekernes brugere.

Danmark deltager allerede nu i VIAF-samarbejdet gennem eksport af danske nationalbibliografiske autoritetsposter fra DBC's autoritetsbase.

Fortsat dansk deltagelse i VIAF-samarbejdet skal være en del af det nationale samarbejde om autoritetsdata.

⁸ Afrapportering af forprojekt, der dannede baggrund for det igangværende [Deff-projekt](#). En kort præsentation af projektet kan ses i Perspektiv, 2014,6:

<http://bf.dk/FagmagasinetPerspektiv/Bladet/2014/Perspektiv6/EtNummerTilAlleForskere>

⁹ VIAF (Virtual International Authority File): <http://viaf.org>

National strategi for autoritetsdata - Version 2.0 (2016)

Nationalt autoritetsdata repositorium med sammenhæng til VIAF

Det nationale samarbejde om skabelse og genbrug af metadata bør understøttes af et nationalt repositorium for autoritetsdata. De nationalbibliografiske data vil være et oplagt udgangspunkt for dette.

Til realisering af dette skal der etableres et fælles dansk autoritetsdatarepositorium der - som minimum - rummer autoritetsdata fra den danske nationalbibliografi.

Som nævnt er sammenhængen til internationale autoritetsdata vigtig, og det forudsættes derfor, at et dansk autoritetsdatarepositorium skal kunne interagere med tilsvarende internationale kilder.

Inden realiseringen af et nationalt autoritetsdatarepositorium er der imidlertid en række forhold, der skal klarlægges nærmere. Blandt disse er de vigtigste: fastlæggelse af hvilke autoritetsdatatyper, der skal være omfattet af det nationale samarbejde, samspillet mellem de nationalbibliografiske autoritetsformer og bibliotekernes data, understøttelse af åben adgang til data samt sammenhængen til det internationale autoritetsdataarbejde som fx VIAF.

Ovenstående spørgsmål vil blive uddybet i afsnittet om implementering af autoritetsdatastrategien.

Sammenhæng mellem autoritetsdata og bibliografiske data i bibliotekernes infrastruktur

Et særligt indsatsområde vil være at sikre optimal sammenhæng mellem autoritetsdata og bibliotekernes metadata, det være sig bibliotekernes eksisterende bibliografiske data eller metadata, der fremadrettet skabes eller genbruges af danske biblioteker.

Der skal derfor foretages en nærmere analyse af de konkrete behov på dette område og herudfra opstilles mulige modeller for, hvorledes autoritetsentiteter kan implementeres i bibliotekernes infrastruktur.

Linked open data

Det er et mål, at der skal være åben adgang til de danske autoritetsdata. Dette gælder såvel i forhold til den forretningsmodel, der vælges, som for den tekniske tilgængeliggørelse af data.

Linked open data vil være et nøglebegreb i denne sammenhæng, og der findes forskellige internationale initiativer, der bl.a. arbejder med at opstille datamodeller, der strukturerer og tilgængeliggør metadata – herunder autoritetsdata - som linked data.

De væsentligste er i denne sammenhæng BIBFRAME, schema.org og Europeanas datamodel: EDM¹⁰.

Et grundlæggende træk ved disse og andre datamodeller, der understøtter LOD, er, at autoritetsdata indgår som selvstændige entiteter på linje med andre entiteter fra det bibliografiske univers.

En anden ting, som det er væsentligt at nævne, er, at autoritetsdatabegrebet i disse modeller rækker ud over en traditionel opfattelse af autoritetsdata, idet det foruden navne på personer og korporationer, steder, emner

¹⁰ EDM (Europeana Data Model): <http://pro.europeana.eu/edm-documentation>

etc. også åbner for, at andre datatyper, som fx materialetyper og udgiveroplysninger kan betragtes som autoritetsdata.

Danske autoritetsdata skal tænkes ind i denne kontekst og der skal udarbejdes konceptuelle modeller, der understøtter deres funktion i et linked data univers.

RDA - international dataudveksling i fremtiden

Ideen med linked data bygger på en datamodel, hvor metadata i form af ”lukkede” bibliografiske poster vil blive afløst af selvstændige entiteter der linkes sammen via relationer.

Denne model går igen i det nye katalogiseringsparadigme, der er under implementering internationalt i form af RDA. I RDA spiller autoritetsdata en meget væsentlig rolle i de relationer, som skabes mellem entiteterne.

Samtidigt med forslaget om en national strategi for autoritetsdata har Bibliografisk Råd afleveret en indstilling til Slots- og Kulturstyrelsen med anbefaling om, at vi i DK skifter vores katalogiseringsregler til RDA.

Funktionelle krav til autoritetsdata

Ud fra en række brugsscenarier for autoritetsdata i danske biblioteker (se i Bilag 3), samt på baggrund af de ovenfor beskrevne aktuelle udviklingstendenser inden for det bibliografiske område, kan der opstilles følgende overordnede funktionelle krav til autoritetsdata, der således skal kunne:

- understøtte éntydig identifikation af en given entitet gennem brug af unikke og persistente identifikatorer (URI'er¹¹)
- sammenholde tekstlige variantformer for samme entitet og give mulighed for at vælge en foretrukken form
- understøtte semantiske netværk, der muliggør meningsfuld navigation mellem relaterede entiteter
- fremme interoperabilitet ved at være struktureret og formateret ud fra fælles rammer

Ikke mindst under inspiration af den udvikling, der foregår inden for det semantiske web og linked data, bør det fremadrettet være et unikt ID (en URI) der er den nøgle, der sammenholder forskellige repræsentationer af den samme entitet – det være sig en person, en organisation, et emne, en lokalitet etc.

Unikke ID'er kan i denne sammenhæng være allerede eksisterende URI'er, som fx ISNI eller ORCID eller der kan anvendes nationalt eller lokalt fastlagte URI'er. De nærmere retningslinjer herfor skal fastlægges.

Brug af URI vil bl.a. betyde, at der fremadrettet ikke vil være behov for en éntydig, autoritativt vedtaget navneform som kollokativt element.

Hermed dog ikke være sagt, at det ikke i mange sammenhænge, fx i nationalbibliografien, vil være både praktisk og hensigtsmæssigt at opretholde kravet om en vedtagen navneform/foretrukken term – således som

¹¹ URI (Uniform Resource Identifier): http://en.wikipedia.org/wiki/Uniform_resource_identifier

det da også er angivet i de nuværende katalogiseringsregler. Som eksempel kan nævnes, at autoritetsdata ofte også fungerer som fælles opstillingsdata, f.eks. i folkebibliotekerne. Her vil kravet om fortsat at kunne angive en vedtagen/foretrukken form til entiteten være ufravigeligt.

Den nationale indsats må derfor finde en passende balance mellem på den ene side at understøtte entydig identifikation gennem tildeling af URI'er og, på den anden side, hvor det er hensigtsmæssigt, at give mulighed for, at forskellige varianter af navnet på samme entitet kan eksistere parallelt, med mulighed for at kunne angive, at der i blandt dem er en form, der er den foretrukne inden for en gruppe af brugere.

I forhold til kravet om at understøtte semantiske sammenhænge vil det tillige være væsentligt, at relationer mellem entiteter indgår som en del af autoritetsdata. Som eksempler herpå kan nævnes oplysninger om evt. navneskift (især for personer og korporationer) eller relationerne mellem emner som fx over- eller underordnede begreber.

Samarbejde om autoritetsdata indgår som en væsentlig del af strategien, og en forudsætning for interoperabilitet vil være, at der er fælles rammer for strukturering af autoritetsentiteter. Set i dette perspektiv vil fastlæggelse af fælles rammer for strukturering og formatering af autoritetsdata være et væsentligt middel til at opnå den ønskede funktionalitet.

I forbindelse med virkeliggørelse af strategien vil der derfor ligge en vigtig opgave i at pege på holdbare, fælles formatløsninger, der på samme tid både skal være optimale i forhold til fremtidige løsninger baseret på linked data og desuden skal kunne understøtte samspillet mellem autoritetsentiteter og bibliotekernes eksisterende data indenfor de etablerede rammer.

I princippet gælder ovenstående funktionelle krav for alle typer af autoritetsdata, men i praksis vil man skulle forholde sig forskelligt, alt efter hvilken type af autoritetsdata, der er tale om. I de følgende afsnit vil dette blive uddybet nærmere i forhold til de enkelte typer af autoritetsdata.

Typer af autoritetsdata

Med udgangspunkt i en traditionel opfattelse af autoritetsdata, er følgende datatyper relevante som autoriteter:

- personer
- korporationer
- (titler)¹²
- emner ('topics')

Set i lyset af RDA og BIBFRAME, vil også andre dataelementer skulle tænkes ind som autoritetsdata i fremtiden. Det drejer sig først og fremmest om FRBR-entiteterne

¹² Titler som selvstændige autoritets-entiteter er ikke relevante i RDA, hvor titel indgår som attribut til works og expressions

National strategi for autoritetsdata - Version 2.0 (2016)

- work
- expression

Det skal bemærkes, at ovenstående omfatter autoritetsdatatyper, der i dag må anses for de mest relevante. Udviklingen kan ændre ved dette og andre typer kan komme til. Området skal derfor løbende vurderes.

Herudover vil andre dataelementer med stor grad af genbrug, og hvor ensartethed er ønskelig kunne overvejes som selvstændige entiteter. Blandt disse kan nævnes:

- lokaliteter
- sprog
- materialetyper

Vi har dog på nuværende tidspunkt valgt ikke at inddrage typer af autoritative data, der har karakter af indhold fra lukkede lister/vokabularer (sprog og materialetyper).

Udelukkelsen skyldes primært, at denne type data mere hensigtsmæssigt kan håndteres ved at koble sig på internationale initiativer såsom RDA registry for materialetyper <http://www.rdaregistry.info/> og eventuelt tilsvarende dokumentation for sprog- og landekoder.

Som nævnt gælder strategiens overordnede visioner og mål naturligvis for alle autoritetsdatatyper, men den konkrete målsætning kan variere fra datatype til datatype, alt efter relevans på nationalt/fællesskabs niveau eller kompleksitet.

Implementering af strategien vil derfor skulle differentiere i forhold til typerne og afspejle de konkrete problemstillinger, der karakteriserer hver enkelt datatype. Endvidere skal der udarbejdes anbefaling af prioritering af eller evt. rækkefølge for implementering af de forskellige typer.

De funktionelle krav vil tillige være afgørende for hvilke attributter, der knyttes til de forskellige typer af autoritetsdataentiteter. Til illustration af dette vil de *væsentligste* attributter, der bør knyttes til personautoriteter være:

- vedtaget navn
- variantformer – herunder varianter i forskellige transskriptioner og alfabeter
- identifikatorer knyttet til personen (ISNI, ORCID)
- relationer til biografisk og historisk information mv. Fx oplysning om navneskift, fødsels- og dødsår, nationalitet, køn, arbejdsmæssig tilknytning

I Bilag 5 ses tilsvarende oversigt for andre autoritetsdatatyper. Udarbejdelse af en detaljeret beskrivelse for alle relevante autoritetsdatatyper inklusiv angivelse af et anbefalet minimumsniveau skal indgå som en del af implementeringen af autoritetsdatastrategien. I samme forbindelse skal også krav i forlængelse af overgang til RDA indtænkes.

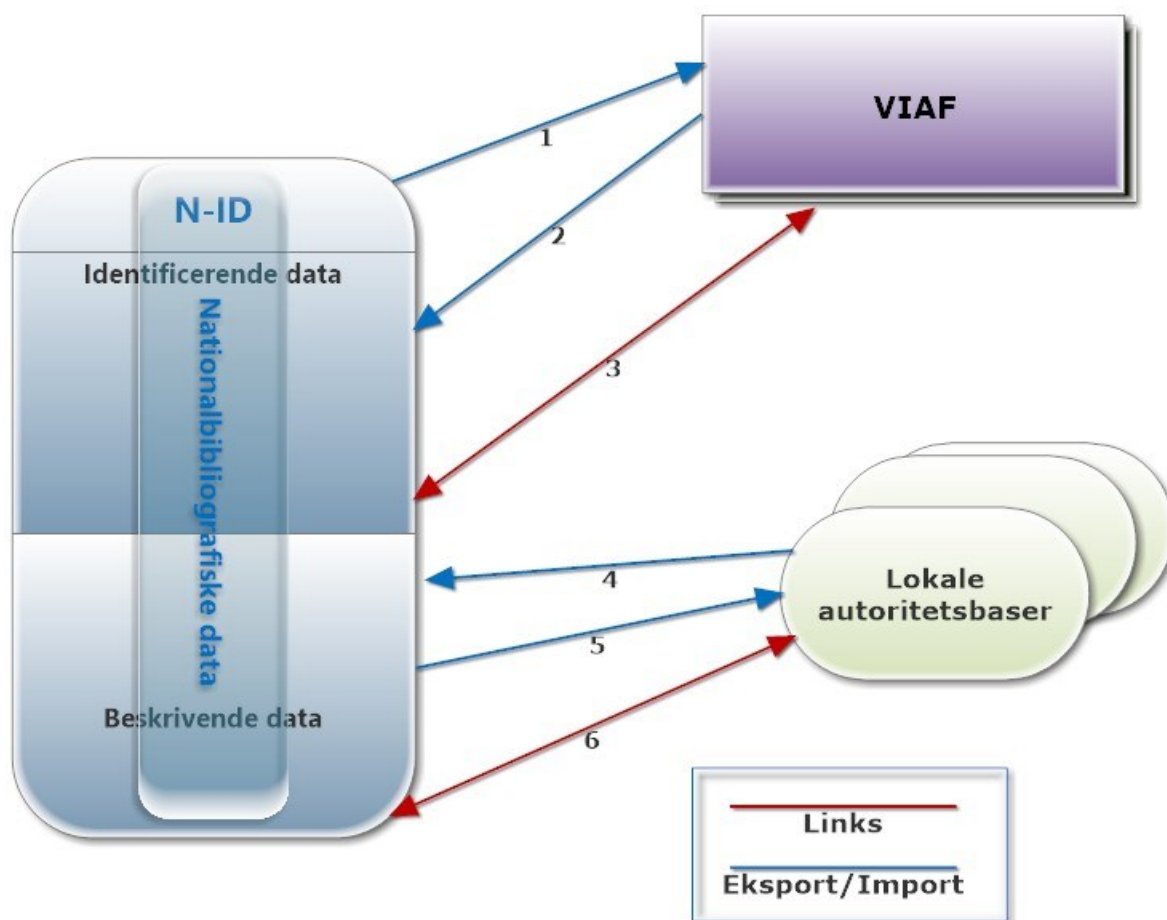
Implementering af strategi

Det nationale autoritetsdatarepositorium (NAR)¹³

Som en del af implementering af autoritetsdata i bibliotekernes nationale infrastruktur, foreslås det som tidligere nævnt, at der etableres et nationalt autoritetsdatarepositorium, NAR, som omdrejningspunkt for det nationale og internationale samarbejde omkring autoritetsdata i Danmark.

Der er udarbejdet en konceptuel datamodel herfor, herunder hvordan man forestiller sig sammenhængen til den internationale autoritetsdatabase, VIAF.

En dyberegående beskrivelse af den foreslåede datamodel fremgår af Bilag 1, men overordnet set foreslås en model som skitseret i denne figur:



¹³ Navnet "Det nationale autoritetsdatarepositorium (NAR)" er en foreløbig arbejdstitel. Den endelige fastlæggelse af navnet indgår i implementeringsprocessen.

Forklaringer til figuren

En entitet i det nationale autoritetsdatarepositorium (NAR) indeholder 3 elementer eller grupper af elementer: Et unikt ID for entiteten (N-ID), et eller flere identificerende dataelementer, fx VIAF-ID, ISNI-nummer eller ORCID. Herudover et eller flere beskrivende dataelementer, fx entitetens navn og variantformer heraf eller oplysninger om nationale tilhørsforhold.

En delmængde af entiteterne i det nationale repositorium udgøres af autoritetsdata fra Nationalbibliografien.

Flow'et mellem repositorier er angivet med pile:

- Entiteter fra det nationale repositorium uploades til VIAF (1)
- Der vil kunne importeres visse dataelementer fra VIAF til det nationale repositorium (2)
- Der oprettes links mellem det nationale repositorium og VIAF (3)
- Lokale systemer uploader entiteter til det nationale repositorium (4)
- Lokale systemer downloader entiteter fra det nationale repositorium (5)
- Hvis match, beriges den nationale entitet med link til den lokale entitet og URI returneres til den lokale base (6)

Modellens gyldighed er vurderet i forhold til de use cases, der anses for de væsentligste:

- Oprettelse af ny fællesressource
- Oprettelse af fællesressource for entitet, der allerede findes som lokal autoritetsentitet.
- Flere biblioteker ønsker at benytte samme data om en entitet
- Tilføjelse af supplerende data til en fællesressource
- Præsentation af udvalgte dele af en ressource
- Tilføjelse relation mellem en fællesressource og en lokalressource
- Indberetning af bibliografisk post, der indeholder VIAF-URI som eneste identifikation af personnavn

En mere uddybende præsentation af de valgte use cases fremgår af Bilag 1. Det skal dog understreges, at der ikke har været intentionen at udarbejde et fuldstændigt sæt af use cases, idet det må forudsættes at listen udvides, i forbindelse med et egentligt implementeringsprojekt.

URI som ID for autoritetsentiteter

Som ID for entiteter i autoritetsdatarepositoriet anbefales det, at man anvender URI'er i henhold til anbefalinger fra [ISA: D7.1.3 - Study on persistent URIs, with identification of best practices and recommendations on the topic for the MSs and the EC](#).

URI'er udformes med udgangspunkt i denne skabelonen:

[skema://domæne:port/type/kategori/reference]

I overensstemmelse med anbefalinger vedr. linked data vælges som **skema** protokollen http eller https. For at sikre et unikt ID, angives **referencen** som et nummer trukket fra en fælles, central nummerrulle for det nationale repositorium.

Under forudsætning af, at referencen er unik på tværs af begrebsmæssig kontekst, synes derimod ikke, at være afgørende tekniske grunde for at anvende en underdeling med kategori, og det er derfor anbefalingen, at man undlader dette.

En nærmere redegørelse for indstilling vil fremgå af bilag 1.

Referencer til andre identifikatorer

Som det vil fremgå af den foreslåede datamodel anses relationen til VIAF for helt essentiel for det nationale repositorium.

Herudover anbefales det at autoritetsentiteter for personer så vidt muligt relateres til de tilsvarende entiteter i ISNI og ORCID's systemer.

Analyser vedr. mulige identifikatorer for andre autoritetstyper som fx korporationer, værker, lokaliteter mv. bør indgå i forarbejdet til implementering af disse, i den takt, det besluttes – jf. indstilling fra Gruppe 2 (Bilag 2, side 36).

Autoritetsdata og RDA. Typer af autoritetsdata

Uanset om en beslutning om at indføre RDA som fælles danske katalogiseringsregler kommer sammen med eller efter implementering af autoritetsdata, vil det være vigtigt at danske autoritetsentiteter kan matche autoritetsentiteter der er udformet i overensstemmelse RDA's behov.

Der er derfor foretaget en nærmere analyse af i hvilke sammenhænge RDA må forventes at påvirke beslutninger vedrørende implementering af autoritetsdata i Danmark.

I sammenhæng med denne analyse gives også et bud på hvilke dataelementer der ideelt set bør indgå i fremtidige autoritetsdata, herunder muligheder for genbrug af udenlandske data, samt i hvilke rækkefølge, set i forhold til autoritetsdatatyper, man vil anbefale autoritetsdata implementeret.

Hertil kommer nogle overvejelser vedr. de formatmæssige krav, som følger af implementering af autoritetsdata.

Analysen kan i sin helhed ses som Bilag 2. Dens hovedkonklusioner og anbefalinger er som følger:

Genbrug

Analysen viser, at direkte genbrug af udenlandske autoritetsdata som danske autoritative data i mange tilfælde ikke vil være muligt:

Årsagerne hertil er bl.a.:

- Som genbrugskilde er VIAF ikke en entydig kilde men svarer til en international fælleskatalog. I nogle tilfælde kan man kompensere herfor ved at prioritere genbrugskilder, men nedenstående punkter komplicerer alligevel genbruget.
- Skriftssprog og translitteration
- Udbredt udenlandsk praksis omkring udeladelse af artikler. Om dette skal være dansk praksis skal endeligt fastlægges i den danske profil, men da det ikke er specielt brugervenligt skulle argumentet herfor udelukkende være af genbrugshensyn.
- RDA's preference mht. at foretrække tilføjelser mv. på den katalogiserende institutions sprog

MEN: udenlandske autoritetsdata vil i mange tilfælde kunne bruges som udgangspunkt for og inspiration til berigelse af danske data.

Regler

Arbejdet med autoritetsdata i et nationalt repositorium er ikke bundet til et valg af RDA men kan isoleret set være ønskeligt for tydeligere at illustrere relationer. Man kan godt implementere et dansk autoritetsrepositorium uden RDA, men man skal i givet fald så selv fastlægge regler for, hvad der skal indgå i beskrivelsen af de enkelte entiteter. Her vil det være naturligt at se på, hvilke elementer RDA arbejder med bl.a. for at kunne genbruge oplysninger.

Dansk hyldeopstilling og standardtitler

Dansk hyldeopstilling bør betragtes som en del af danske autoritetsdata via en særlig markering. Den titel der vælges til opstilling vil i langt de fleste tilfælde alligevel indgå som varianttitel i autoritetsposten.

De hidtidige danske standardtitler på musik bør ligeledes fremgå af autoritetsdata og derved bygge bro imellem hidtidig og fremtidig praksis.

Formatet

Hvad enten implementering vælges inden for et MARC-miljø eller ej, skal der tages højde for at alle relevante elementer, der hidtidig er fremgået af hjælpepostformatet eller som noget nyt fremgår af RDA, kan angives. I analysen er anført samtlige konstaterede mangler indenfor det nuværende format.

Hvis implementeringen sker inden for MARC, skal den hidtidige logik omkring navne helt for sig og titler helt for sig (enten felt 100 eller felt 130 i en post – aldrig begge) løses, så der kan ske den nødvendige kobling af ophav og titler som authorized access points.

Praktisk brug af autoritetsdata

Mange værker udkommer kun en enkelt gang og oversættes formentlig heller aldrig. Der bør tages stilling til, om vi i dansk sammenhæng vil (og kan) oprette værkposter på alt fra starten, hvor vi i hvert fald for de bibliografiske posters vedkommende, befinder os i et MARC-miljø.

Dertil kommer at afgøre præcis, hvordan værk- og expression posters sammenhæng skal illustreres i en dansk kontekst. Skal vi have en værkpost, der linker til sit ophav, en- eller flere expressionposter, der linker til

værkposten og ikke gentager oplysninger herfra eller skal etableres flade strukturer med gentagelser af oplysninger.

Svaret på ovenstående hænger meget sammen med de muligheder for rationelle arbejdsgange og overblik, genbrug mv., som de fremtidige værktøjer vil kunne stille til rådighed.

Brugsscenerier omkring samspil mellem bibliografiske poster og entitetsposter og imellem entitetsposter bør beskrives nærmere både helt generelt og også specifikt. (Et eksempel på sidstnævnte: Hvad skal der ske fx når man i en bibliografisk musikpost har en henvisning til en autoritetspost og ønsker standardtitlen fremvist, fordi materialet er af ældre dato og er opstillet herunder?)

Der skal endvidere tages stilling til, hvad der vil være hensigtsmæssigt at udføre retrospektive rettelser på når datamodel og format kendes.

Prioritering af typerne – rækkefølge for implementering mv.

Arbejdsgruppen anbefaler følgende prioritering af typerne:

1) Personnavne

Man vil med fordel kunne starte med dette område, da DBC allerede har etableret en pt. intern autoritetsbase dækkende en væsentlig del af det nationalbibliografiske område. En opgave vil bl.a. være at finde en løsning på håndtering af de ikke-nationalbibliografiske navne samt opgradering af et i første omgang udenlandsk navn til dansk nationalbibliografi. Disse spørgsmål forventes besvaret via datamodellen.

2) Værktitler på musik.

Så snart vi går i gang med at bruge RDA, vil dette være et presserende problem, så opbygning af repositorium inkl. løsning på dansk opstilling og relation til gamle standardtitler på dette område bør prioriteres højt.

Det bør undersøges, hvor mange danske standardtitler vi egentligt har haft anvendt på musik. Det bør efterfølgende vurderes, om vi skal lave en gennemgang af disse eller som en anden mulighed: starte med at lægge standardtitlerne op som autoriteten og rette til RDA efterhånden, som titlerne genbruges, når der kommer nye materialer. I vurderingen bør også indgå om noget i større omfang kan klares maskinelt fx via match med VIAF og oversættelse til dansk af udvalgte termer.

3) Korporationsnavne

Gerne sideløbende med punkt 2 vil det være relevant at opbygge autoritetsdata på nationalbibliografiske korporationsnavne på linje med personnavne.

4) Øvrige værk- og expressionstitler

Implementeringsprojekter

Implementering af den foreslåede strategi vil, udover etablering af de tekniske løsninger, også indebære en yderligere en række opgaver af specificerende karakter. De væsentligste af disse vil være fastlæggelse af regler og best practice for det fremtidige samarbejde om autoritetsdata.

Regler og best practice

Et samarbejde mellem danske biblioteker omkring autoritetsdata vil indebære et behov for yderligere standardiseringstiltag på området, i form af regler eller vejledning om best practice på en række områder. De vigtigste er i den forbindelse:

- Alt afhængig af beslutning vedr. RDA som danske katalogiseringsregler skal der udarbejdes beskrivelse af ønskede attributter for alle relevante autoritetsdatatyper inklusiv angivelse af et anbefalet minimumsniveau. Hvis der træffes beslutning om RDA inden implementering af autoritetsdata vil dette kunne indgå som en naturlig del af implementeringen heraf.
- Det skal vurderes, om nuværende varierende praksis for valg af autoritative former bunder i historiske forhold, eller om der stadig er et aktuelt behov. I de tilfælde, hvor der fortsat vil være behov for at bruge flere autoritetsformer til samme entitet, skal dette understøttes af såvel regler som format på en sådan måde, at der etableres korrekte relationer mellem variantformerne. Træffes beslutning om RDA vil denne vurdering kunne indgå som en naturlig del af udarbejdelsen af den danske profil.
- Eksisterende og nye fælles formatløsninger skal ajourføres eller udarbejdes. De konkrete opgave i denne sammenhæng vil dels være afhængig af de formatmæssige beslutninger, der træffes i forlængelse af en overgang til RDA, dels bero på, hvilken fysisk implementering af det nationale repositorium, der besluttes.
- Eventuelle juridiske begrænsninger vedr. persondata i autoritetsrepositorier skal analyseres og beskrives.

Det nationale repositorium - tekniske løsninger

Omsætning af den foreslåede logiske datamodel til en fysisk implementering bør ske indenfor rammerne af - og i sammenhæng med – den øvrige udvikling af bibliotekernes fælles infrastruktur.

Det vil være at væsentligt afprøve modellen gennem et "proof of concept". Det anbefales derfor, at lade dette indgå i 2 af de projekter, der er medtaget på den nationale biblioteksinfrastrukturens releaseplan 2016: R15.19: "Personautoriteter i Databrønden" og R16.44: "Teknologisk afklaring af RDA, Bibframe og autoritetsdata".

Øvrige tiltag

- En forudsætning for en vellykket implementering af strategien vil være, at den har bred opbakning i biblioteksvæsenet. Dette vil bl.a. være betinget af, at bibliotekernes faglige specialister og beslutningstagere får mulighed for at drøfte sagen. Slots- og Kulturstyrelsen bør derfor sikre rammerne for en bred drøftelse blandt bibliotekerne.

Bilagsfortegnelse

Bilag 1: Et nationalt repositorium for autoritetsdata

Bilag 2: Autoritetsdata – Datatyper og RDA-sammenhænge

Bilag 3: Brugsscenarier

Bilag 4: Autoritetsdata i dag – status

Bilag 5: Oversigt over væsentlige attributter til forskellige autoritetsdatatyper

Bilag 6: Ordliste og definitioner