

**Are D. Gulbrandsen**

---

# **Autoritetsregister og ID for verk**

**– Fra Kallimakhos til realisering av FRBR og Lenkede data**

**Masteroppgave i bibliotek- og informasjonsvitenskap  
Høgskolen i Oslo og Akershus, Fakultet for samfunnsfag**

## **Sammendrag**

Forskningsområdet for masteroppgaven er verkets og forfatterens rolle i en bibliografisk verden der man beveger seg fra informasjonssiloer, og MARC-koding av katalogposter, til distribuerte systemer og Lenkede åpne data (Linked Open Data). Verksbegrepet analyseres fra et historisk perspektiv, og sentralt i dette er selve formålet med bibliotek katalogen.

Oppgaven tar utgangspunkt i PODE-prosjektets ønske om et nasjonalt autoritetsregister med åpne, nasjonale unike ID-er for verk.

Designkriterier for et autoritetsregister for verk basert på FRBR og Lenkede data drøftes. Disse er basis for å lage en konseptuell modell av et forenklet autoritetsregister som etablerer unike identifikatorer for verk og bruker semantisk teknologi.

Modellen ble realisert med RDF og OWL, og en enkel prototype av autoritetsregisteret ble implementert med Ontologi-rammeverket Protégé. Det er laget enkle skjema i prototypen for innlegging av personer, verk og manifestasjoner. Til slutt ble de viktigste designkriteriene testet på prototypen med SPARQL-spørringer.

## **Abstract**

The master thesis explores the role of the work in a bibliographic universe moving from information silos and MARC encoding to distributed systems and Linked Open Data.

The bibliographic work concept is analyzed from a historical perspective, and central in this are the main objectives for the bibliographic catalog.

The starting point of the thesis is the closing request from the PODE-project, for national authority records establishing open unique identifiers for bibliographical works and authors.

Design criteria for a new type of authority system based on FRBR and semantic technology are discussed, forming the basis for a simplified conceptual model, - an ontology.

The ontology is realized using RDF and OWL, and a simplified prototype of the authority system is implemented using the Protégé ontology editor and knowledge-base framework. Finally the most important design criteria are tested using SPARQL queries.

## Forord

En stor takk til veilederne Unni Knutsen og Nils Pharo som har holdt ut med meg som deltidsstudent i en lengre periode.

Hjertelig takk til Robert Engels for hjelp med å komme i gang med ontologieditoren Protégé.

Oslo, juni 2012

Are D. Gulbrandsen

*Entities are a state of mind. No two people agree on what the real world view is.*

Metaxides(Kent & Hoberman, 2012)

# Innholdsfortegnelse

<b>Forord</b> .....	<b>3</b>
<b>Innledning</b> .....	<b>7</b>
Problemstilling .....	8
<i>Avgrensinger</i> .....	8
Oppgavens struktur .....	9
<b>Metode</b> .....	<b>10</b>
Data, metadata og semantikk .....	11
<i>Metadata</i> .....	12
<i>Strukturerte data</i> .....	13
<i>Semantikk</i> .....	15
RDF og Semantisk web .....	16
<i>Lenkede data</i> .....	18
<i>RDF kodet med Turtle syntaks</i> .....	20
<i>SPARQL</i> .....	21
<i>Ontologirammeverket Protégé</i> .....	23
<b>Verkbegrepets historie</b> .....	<b>24</b>
I begynnelsen var ordet, og verket .....	24
<i>Sivilisasjonen og skriftspråket oppstår</i> .....	25
Det første biblioteket? .....	26
Opprinnelsen til verksbegrepet i bibliografiske systemer .....	28
<i>En autoritativ versjon av Homers verker</i> .....	28
<i>Kallimakhos og biblioteket i Alexandria</i> .....	30
Gjenoppgdagelsen av verket .....	31
<b>Verk i bibliografiske systemer</b> .....	<b>37</b>
Deichmanske Biblioteks webkatalog .....	37
Pode-prosjektet .....	39
Det norske biblioteklandskapet .....	41
Internasjonale utviklingstrekk .....	41
<i>FRBR og RDA</i> .....	41
<i>RDF og Lenkede data</i> .....	41
<b>Behovet for identifikatorer</b> .....	<b>43</b>
<i>Navn og tittel er dårlig egnet som identifikator</i> .....	43

<i>Meningsbærende identifikatorer</i> .....	45
Ikke-meningsbærende identifikatorer.....	45
Et ontologisk problem .....	46
<b>FRBR</b> .....	<b>48</b>
Verket i FRBR-modellen .....	49
<i>Tittel</i> .....	50
<i>Litterær form</i> .....	51
<i>Årstall</i> .....	51
<i>Skillende karakteristika</i> .....	52
<i>Beregnet målgruppe</i> .....	52
<i>Kontekst</i> .....	52
<i>Attributt- og relasjonsverdi ved gjenfinning og identifisering av verk</i> .....	52
Manifestasjonen i FRBR-modellen.....	53
Etablerte identifikator typer .....	54
FRBRisering og MARC .....	55
<b>Autoritetsregistre</b> .....	<b>56</b>
<b>Oppsummering og rammer for et autoritetsregister</b> .....	<b>65</b>
Bruk av FRBR og lenkede data .....	65
Hvordan bør en unik identifikator for verk se ut? .....	66
Diskusjon av kriterier for et nasjonalt autoritetsregister .....	66
<i>SPARQL-søk etter verks-id</i> .....	67
<i>ISBN som input</i> .....	67
<i>Identifiserende data for manifestasjon</i> .....	68
<i>ID for forfatter</i> .....	69
<i>Trenger vi ID for uttrykk?</i> .....	69
Attributter for verk.....	69
Verkets rolle i katalogen og autoritetsregisteret.....	71
<i>Ett navn, flere navn og visningsnavn på ulike språk</i> .....	71
Brukerbehovet og verksavgrensning .....	72
<b>Et autoritetsregister for verk basert på FRBR og RDF</b> .....	<b>77</b>
ID for verk.....	78
Oppsummering og drøfting av designkriterier for autoritetsregisteret .....	79
<i>Søkegrensesnitt</i> .....	79
<i>Attributter og relasjoner for verk</i> .....	80

<i>Attributter og relasjoner for manifestasjon</i> .....	80
<i>Attributter og relasjoner for person</i> .....	80
Ontologidesign.....	80
<i>Person</i> .....	81
<i>Verk</i> .....	82
<i>Manifestasjon</i> .....	82
<i>Konseptuell modell av Ontologi</i> .....	83
Implementasjon av en forenklet prototype i Protégé .....	84
<i>Realisering av ontologi i Protégé</i> .....	85
Test av designkriterier på prototypen.....	87
<i>SPARQL-spørring med ISBN som input</i> .....	87
<i>SPARQL-spørring med tittel og år som input</i> .....	87
<i>SPARQL-spørring som returnerer både ID, tittel og år for verk</i> .....	88
<b>Konklusjon</b> .....	<b>89</b>
Videre arbeid.....	91
<i>Automatisk generering av data i autoritetsregisteret</i> .....	91
<i>Utnytte en rik bibliografisk kilde</i> .....	91
<i>Etablering av identitet uten identifikatorer</i> .....	91
<i>Generere unike verks-ID ved hjelp av NOID</i> .....	92
<i>Publisering som Lenkede åpne data</i> .....	92
<i>Tilrettelegge og oppfordre til gjenbruk av data</i> .....	93
<i>Visningsnavn for verk i ulike kontekster</i> .....	93
<b>Litteraturliste</b> .....	<b>93</b>
<b>Vedlegg 1: RDF og OWL-kode for ontologien</b> .....	<b>103</b>
<b>Vedlegg 2: RDF eksempel-kode for forfatter, verk og manifestasjoner</b> .....	<b>105</b>
<b>Vedlegg 3: Protégé prosjektfil og datafil</b> .....	<b>107</b>

## Innledning

Forskningsområdet for denne oppgaven er verkets (og forfatterens) rolle i bibliografiske systemer. Oppgaven drøfter verkets rolle i en bibliografisk verden der man beveger seg fra informasjonssiloer, og MARC-koding av katalogposter, til distribuerte systemer og Lenkede åpne data (Linked Open Data). Sentralt i denne drøftingen er selve formålet med bibliotek katalogen.

I kapittelet om verkbegrepets historie argumenteres det for at verket er den opprinnelige bibliografiske entiteten. Ved biblioteket i Alexandria klarte Kallimakhos allerede ca. 250 år f.Kr. å realisere en katalog over verk.

Verksbegrepet historie kan følges videre fra Sir Anthony Panizzi, som kjempet for verkets posisjon i bibliografiske systemer i 1849, via Seymour Lubetzky's Paris-prinsipper som var utgangspunktet for AACR (Anglo-American Cataloguing Rules), til FRBR (Functional Requirements for Bibliographic Records) som igjen er basis for RDA (Resource Description and Access).

*English-speaking countries have collaborated on cataloging for over a century, and with IFLA's support there is a worldwide movement for universal bibliographic control. Underpinning all of this is the idea of the "work". (Denton, 2007)*

Moderne bibliografiske systemer har imidlertid ikke klart å realisere verkets hovedrolle, mer enn 150 år etter at Panizzi formulerte de moderne grunnprinsippene og behovet, 60 år etter Lubetzky's formulering av dagens hovedprinsipper og formål for bibliotek katalogen, og snart 15 år etter at FRBR ble publisert.

Verket har tilsynelatende fått en underordnet rolle i dagens bibliografiske systemer.  
– Verket vises ikke i biblioteksystemene jeg bruker.

Et av de viktigste bibliotekprosjektene i Norge de siste årene var PODE-prosjektet ved Deichmanske bibliotek (2007-2011), som utforsket mulighetene for å gjøre noe med dette. Prosjektet hadde brukeren og brukertjenester som hovedfokus, og utforsket FRBRisering av folkebibliotek katalogen og RDF (Resource Description Framework) som alternativ til

MARC-formatet. Det ble brukt semantiske teknologier (RDF, OWL og SPARQL) til å publisere bibliografiske informasjon som Lenkede data.

Pode-prosjektet avsluttet sin siste bloggpost med en trepunkts ønskeliste basert på arbeidet i prosjektet (Podeprosjektet , 2011a):

1. Åpen dokumentert tilgang til alle katalogdata
2. Et moderne metadataformat som er åpent og fleksibelt
3. Nasjonalt autoritetsregister og åpne, nasjonale unike ID-er for verk og forfattere

Ønskene var utformet som en utfordring til Nasjonalbiblioteket, katalogfaglige komiteer på nasjonalt (og internasjonalt) plan og biblioteksystemleverandørene.

## *Problemstilling*

Denne oppgaven drøfter Pode-prosjektets ønske om et nasjonalt autoritetsregister med åpne, nasjonale unike ID-er for verk.

Hovedspørsmålet er hva som skal til for å tilfredsstille Pode-prosjektets ønske?

- Pode-prosjektet utforsket FRBR, semantiske teknologier og Lenkede data. Er dette naturlige teknologivalg for et autoritetsregister, gitt dagens utvikling hos BIBSYS, Nasjonalbiblioteket og viktige internasjonale bibliotekaktører?
- Hva ligger i begrepet verk, og hvilken rolle spiller verket i bibliotek katalogen?
- Hva ligger i begrepet autoritetsregister, og hvordan kan et autoritetsregister løse Pode-prosjektets hovedutfordringer?
- Hvordan bør en unik ID for verk se ut?

Oppgaven tester en hypotese om at det lar seg gjøre å modellere og prototype et forenklet autoritetsregister som løser Podes utfordring ved hjelp av FRBR, RDF, OWL og Lenkede data, innenfor rammene av masteroppgaven.

## **Avgrensinger**

Arbeidet avgrenses til monografier (vanlige bøker, ikke flerbindsverk), da det blir for omfattende å ta høyde for alle materialtyper (tidsskrifter, musikk, video, kart, noter mm).



Oppgaven avgrenses også når det gjelder skapere av verk, fra personer og korporasjoner, til kun personlige forfattere.

## *Oppgavens struktur*

Metodekapittelet redegjør for teori og metode som brukes for å svare på forskningsspørsmålene, og hvordan jeg går frem for å teste hypotesen i problemstillingen. Fordi RDF og Lenkede data brukes som metode for realisere den konseptuelle modellen inkluderes en gjennomgang av kunnskapsstatus innen temaer som metadata, semantisk web og lenkede data.

Kapittelet om verkbegrepets historie er en gjennomgang av opprinnelse og bruk av begrep og idé. Sentralt i denne analysen er også formålet med bibliotek katalogen.

Kapittelet Verk i bibliografiske systemer gir en oversikt over status for utvikling av bibliografiske systemer i Norge og internasjonalt. Det vises også eksempler på at forfatter og verk ikke eksisterer som objekter i katalogmodellen.

Kapittelet om behovet for identifikatorer gir en bakgrunn for arbeidet med identifikator for verk og kapittelet om FRBR gir en oversikt over FRBR-modellen og noen av utfordringene ved forsøk på å FRBRisere MARC-baserte kataloger.

Kapittelet om autoritetsregistre forsøker å gi en oversikt over tradisjonelle autoritetsregistre og den konseptuelle modellen i FRAD-rapporten.

I diskusjonskapittelet drøftes problemstillinger som er viktige for å kunne bestemme designkriterier for autoritetsregisteret.

I kapittelet *"Et autoritetsregister for verk basert på FRBR og RDF"* defineres først designkriterier for et nasjonalt autoritetsregister basert på FRBR, RDF og OWL, som realiserer identifikatorer for verk. Designkriteriene er basis for å lage en konseptuell modell av et forenklet autoritetsregister, - en ontologi.

Ontologien realiseres med RDF og OWL, og en begrenset prototype av systemet implementeres i ontologirammeverket Protégé. Til slutt blir de viktigste designkriteriene testet på prototypen med SPARQL-spørringer.

## Metode

Teoretisk grunnlag for oppgaven er bl.a. dagens bibliotekmodeller og bibliotekteori, bl.a. FRBR- og FRAD-rapporten, identifikatorstandarder, bøker, rapporter, artikler og blogger.

Litteraturstudier og kvalitativ analyse av etablerte modeller, standarder og teori er brukt som metode i første del av oppgaven. Metoden er delvis en historisk gjennomgang av feltet for å tolke hva som ligger i begrepet verk, og hvilken rolle verket har i bibliotek katalogen? Det er her gjort en historisk analyse basert på foreliggende litteratur.

I tillegg brukes plandokumenter, nyhetsmeldinger og policy-dokumenter fra ulike bibliotekaktører som et grunnlag for å vurdere trender for teknologivalg i Norge og internasjonalt.

I andre del av oppgaven brukes en mer praktisk orientert metode. Konseptuell analyse brukes som metode for å komme frem til designkriterier for et autoritetsregister.

Designkriteriene er basis for å lage en konseptuell modell av et forenklet autoritetsregister, en ontologi.

Ontologien realiseres med RDF og OWL, og en begrenset prototype av systemet implementeres i ontologiverktøyet Protégé.

Modellering og prototyping av et minimalt autoritetsregister for verk brukes som metode for å evaluere om det er mulig å lage et system ut ifra designkriteriene, og for å vurdere om det foreslåtte designet kan løse hovedutfordringene som er kartlagt.

Som test av prototypen kjøres SPARQL-spørringer mot ontologien og eksempel-instanser i Protégé for å se om designkriteriene er tilfredsstillt. Siden Protégé og SPARQL brukes, kan eksperimentene og prototypen etterprøves av andre. Det er bare å installere Protégé, laste opp prosjektet som er brukt i oppgaven og teste med SPARQL-spørringene.

RDF og Lenkede data brukes som metode for realisere modellen. Dette er teknologier som faller inn under merkelappene web-arkitektur og semantisk web hos W3C, så det er naturlig å gi et overblikk over grunnleggende filosofi og teori her i metodedel av oppgaven. Derfor

inkluderes en gjennomgang av kunnskapsstatus innen temaer som metadata, semantisk web og lenkede data.

## *Data, metadata og semantikk*

Det kan som en innledning være nyttig å redegjøre hva som er forskjellen på data og metadata, hva som menes med semantisk teknologi, og se på hvilken sammenheng det er mellom semantikk og data.

*Data* er en term som er brukt og brukes i så mange ulike sammenhenger at begrepet kan tolkes svært vidt. Det interessante i denne sammenhengen er å se begrepet i forhold til begrepene *informasjon*, *metadata* og *mening* (semantikk).

Følgende definisjoner har et brukbart abstraksjonsnivå i denne konteksten (Informasjon, udatert):

*Data er symboler organisert i bestemte mønstre som representerer fakta, observasjoner og/eller ideer, og som det er mulig å kommunisere, fortolke og manipulere ved hjelp av menneskelige eller automatiske prosesser.*

*Informasjon er den forståelsen eller fortolkningen et menneske trekker ut av data på grunnlag av kjente konvensjoner for den benyttede representasjonsform. En slik tolkning av data vil være betinget av: Evnen til å oppfatte data, kulturell kontekst, konvensjoner for tolkning, forståelse av protokoll for innkoding av data samt evne til å forstå meningsinnholdet i data.*

I denne sammenhengen, og for å forenkle litt, kan *representasjonsform* erstattes med *syntaks* (koding). *Forståelse* kan erstattes av *semantikk* (mening).

Hvis syntaksen (kodingen) og semantikken (meningen) for dataene er kjent kan dataene tolkes som informasjon.

Dette betyr at vi kan stå overfor data som vi ikke greier å tolke, og som derfor ikke er informasjon for oss fordi de ikke gir mening.

Data som vi ikke klarer å tolke uten mye bearbeiding eller prosessering er en egen sjanger. Det er gjerne dette som for mange er begrepet bak termen data. Begrepene rådata, måledata, grunndata og datasett brukes ofte som en presisering av at det er dette som menes.

Tolkning av data som informasjon kan også avhenge av kontekst (sammenheng) og evne til å forstå meningsinnholdet i informasjonen. Vi kan forstå at noe er informasjon for andre uten å forstå meningen i informasjonen selv. (F.eks. vil mange dra kjensel på formelen  $E = mc^2$  og knytte den til Albert Einstein uten at de nødvendigvis kan redegjøre for den dypere meningen med denne informasjonen.)

I praksis vil begrepene data og informasjon derfor ofte brukes synonymt. Det er vanlig å bruke begrepet informasjon eller informasjonsressurs om utolkede data hvis man er relativt sikker på at det er informasjon for andre.

## Metadata

Historien om metadata kan sies å ha startet i bibliotekene, der bøkernes metadata var i katalogen, som kan sies å være en forkortet representasjon av informasjonen i bøkene utfylt med bibliotekarenes kunnskap om boken.

Tenk deg at du leter etter informasjon om noe, og har funnet en informasjonsressurs som kanskje er interessant. I tillegg til selve innholdet i informasjonsressursen er det ofte en del annen informasjon man kan bruke til å vurdere innholdet:

- Hva handler det om? (Tittel, sammendrag og emneord)
- Hva slags type dokument er det? Hvilket format? (Tekstlig dokument (pdf, odf eller word?), - en podkast eller en video?)
- Hvor aktuelt er innholdet i dokumentet? Når ble dokument opprettet, hva er sist endret dato?
- Hvem er forfatter? - Eller hvordan ble datasettet/informasjonen laget? (Hvis det ikke er en bestemt forfatter)
- Hvem har ansvaret for informasjonen, og hvordan kan man kontakte vedkommende?

Alt dette kan karakteriseres som metadata, og man kan godt tenke på det som ulike typer merkelapper på informasjonsressursene.

Metadata startet faktisk som merkelapper (*sillyboi/sillybos*) som hang ut av papyrus- eller pergamentrullene og kort beskrev innholdet, slik at man kunne finne frem til riktig dokument uten å ta frem og rulle ut hele dokumentet. (Blum, 1991, s. 63, 93)

Hvis det knyttes metadata til et dokument har vi bedre muligheter for å finne frem til og vurdere relevans for enkeltdokumenter innenfor mengden av alle dokumenter som er tilgjengelig.

En vanlig definisjon av metadata er at det er data om data. Konteksten i dette kapittelet er tolkingen av elektronisk tekst publisert på WWW. Med tolkning menes både menneskelig tolkning og tolkning ved hjelp av dataprogrammer.

Metadata kan defineres som strukturerte data som beskriver en informasjonsressurs.

## **Strukturerte data**

Strukturerte data vil si at syntaksen og semantikken for dataformatet er kjent, og at formatet er maskinlesbart.

Termen maskinlesbare data brukes vanligvis om formater som er standardiserte og kodet på en strukturert måte slik at programmer kan tolke innholdselementene.

I såkalt ustrukturerte dokumenttyper (f.eks. tekst-, pdf- og word-dokumenter) må innholdet tolkes som naturlige språk. Datamaskiner får lite eller ingen hjelp til å identifisere innholdet fra den tekniske kodingen av teksten. Kodingen er vanligvis laget for å vise frem dokumentet for en bruker, dvs. man koder utseendet av dokumentet.

Hvis tekst i et HTML-dokument (*Hyper Text Markup Language*) er kodet med elementet `h1` vet vi kun at teksten inne i elementet er en overskrift. Elementnavn kan i HTML sies å være merkelapper som beskriver innholdet.

```
<h1>INF2820 - Datalingvistik</h1>
```

Dette kan være bra nok til at man innenfor en bestemt kontekst (i en bestemt sammenheng) kan gjenbruke eller flette informasjon fra ulike kilder, såkalt *mash-up*.

Hvis utgangspunktet er HTML kalles dette vanligvis for skraping av websiden (*data scraping* eller *web scraping*). Ved skraping kan man utnytte at det brukes faste maler for informasjonen. Det kan lykkes å hente ut de ulike informasjonselementene hvis informasjonen er skrevet inn i dokumentene på en måte som er lik fra dokument til dokument. Logikken og konteksten som ligger til grunn for tolkningen ligger ofte i dokumentets URL (*Uniform Resource Locator*) dvs. web-adresse, som i dette tilfellet gir noe mer informasjon:

<http://www.uio.no/studier/emner/matnat/ifi/INF2820/>

I formatet XML (*Extensible Markup Language*) kan man definere en mer logisk struktur, og dokumentet kan da regnes som strukturert tekst. I XML-eksempelet under er det tre elementer. Tittel-elementet er satt sammen av to andre elementer. Først emnenavn og deretter emnekode.

```
<tittel>
  <emnenavn>Datalingvistikk</emnenavn>  <emnekode>INF2820</emnekode>
</tittel>
```

Grensen mellom data og metadata har alltid vært uklar. Tittelen for en bok er f.eks. både data (informasjon i boken) og metadata (informasjon om boken).

Hvis vi har strukturert dokumentkoding, som i XML, vil grensene mellom data og metadata gli over i hverandre. I et XML-dokument kan i prinsippet alt som står i dokumentet være både data og metadata.

Ofte vil man ikke klare å fortolke data uten en beskrivelse av hva slags data det er snakk om og hvordan de er kodet. Metadata er vanligvis svært nyttig også for datasett som er strukturert kodet og maskinlesbare.

Det kan her være nyttig å definere ulike typer metadata. En vanlig kategori er administrative metadata, som for eksempel opprettet-dato og sist-endret-dato for et dokument. Denne typen metadata er uunnværlige i et publiseringssystem. Informasjonen finnes ofte ikke i dokumentet, men i publiseringssystemet, som administrerer metadataene. Metadataene vil ofte vises i signaturen for dokumentet. Dette kan være uunnværlig informasjon for et datasett. I et bibliotek er tradisjonelt dette uinteressante metadata. Man forholder seg

isteden til utgivelses-dato (eller kun utgivelses-år). Hva som er interessante metadata er altså avhengig av kontekst.

## Semantikk

I XML er elementnavnene kun svært korte tekstlige beskrivelser, som må tolkes av et menneske innenfor en domenespesifikk kontekst. Muligheten for å definere elementnavn som meningsbærende merkelapper og muligheten for lage et XML-skjema som definerer en konsistent dokumentstruktur gjør at en mash-up vanligvis blir mer pålitelig enn hvis HTML er utgangspunktet.

Elementnavnene er imidlertid vanligvis ikke veldefinert, I den forstand at man kan bestemme om samme elementnavn i to dokumenter med ulikt skjema faktisk har samme betydning. Elementnavnet er vanligvis en kort term som kan ha ulik betydning i ulike kontekster. XML-standarden beskriver ingen standard måte å definere hva som er semantikken, dvs. betydningen eller meningen, hvis tekst kodes med et bestemt element.

Det finnes mange eksempler på ting som har flere navn, - synonymer. Her er det ofte en utfordring å vite om betydningen er den samme eller om det er en viktig nyanseforskjell. *Tittel* og *Overskrift* kan f.eks. være synonymer i noen sammenhenger. En liknende problemstillinger finnes i oversettelse mellom ulike språk. *Title* eller *heading* som elementnavn kan også bety det samme. Synonymer kan gi god prosa, men vanskelig maskinkommunikasjon.

Tittel i et XML-dokument kan i noen sammenhenger bety en persons tittel, f.eks. Overingeniør, mens i andre sammenhenger kan det være tittelen på en bok. Hvis to eller flere forskjellige ting deler samme navn kaller vi det homonymer. Dette kan noen ganger være opphav til god humor, men gir ofte vanskelig maskinkommunikasjon.

Det gjemmer seg ofte store utfordringer i elektronisk samhandling og kobling av informasjon hvis man ser nærmere på mening og definisjon for begreper som brukes som termer.

I den såkalte Kungsbacka-saken i Sverige ble ca. 1000 personer politianmeldt for mistanke om trygdemisbruk etter at myndighetene sammenlignet informasjon i ulike offentlige systemer. Under 2 % av de anmeldte ble domfelt. Et av forholdene som kom frem under

etterforskningen var at det fantes mer enn 25 rettslige definisjoner av inntekt innenfor ulike deler av offentlig forvaltning. (Bing, udatert), (Freese, 1987), (Bygrave, 1996)

Innen lingvistikk (språkvitenskap) er semantikk studiet av mening i språk, og semantisk teknologi dreier seg om å bruke datateknologi til å tolke mening i informasjonsressurser og data. Som det vises av eksempelet fra Kungsbacka er det selv for mennesker en kompleks utfordring å bestemme om begreper som brukes innenfor to domener er identiske. Hva det menes med at to begreper er identiske er det skrevet filosofiske artikler om.

Både RDF og Emnekart regnes som semantiske teknologier. Forenklet kan vi si at for begge bestemmes likhet av om den samme identifikatoren er brukt for begge forekomstene av begrepet. Som identifikatorer brukes URI. Mer om dette under.

Dublin Core (DC) er et eksempel på en metadatastandard med godt definert semantikk.

## *RDF og Semantisk web*

RDF (*Resource Description Framework*) ble opprinnelig laget av W3C (*World Wide Web Consortium*) som en datamodell for metadata, og standarden fikk status *recommendation*<sup>1</sup> i 1999 (*Resource Description Framework (RDF) Model and Syntax Specification*, udatert). 1999-versjonen var en beskrivelse av konseptuell datamodell med en XML serialiseringssyntaks.

Ha i bakhodet at det er en glidende overgang fra metadata til data. Metadataaktiviteten i W3C ble for flere år siden flyttet over i gruppen som arbeider med semantisk web.

*The Semantic Web is a web of data.*

*It is about common formats for integration and combination of data drawn from diverse sources (where the original Web mainly concentrated on the interchange of documents)*

---

<sup>1</sup> W3Cs hovedoppgave er å jobbe for tekniske standarder relatert til webteknologi. Hvis en standard blir ratifisert av W3C får den status *recommendation*.



*It is also about language for recording how the data relates to real world objects.*

Det var Tim Berners Lee, oppfinneren av WWW, som lanserte uttrykket *Semantisk Web* i artikkelen *The Semantic Web* i tidsskriftet *Scientific American* i mai 2001, sammen med Ora Lassila som stod bak RDF-standarden. (Berners-Lee, Hendler, & Lassila, 2001)

*The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. (Berners-Lee et al., 2001)*

Det var denne artikkelen som på mange måter fanget folks interesse for semantisk web. Samtidig er dette en artikkel som har fått skylden for overdrevne forventninger,



misforståelser og kritikk mot semantisk web som teknologi.

I hovedeksempelet i artikkelen får Pete en telefon fra søsteren om at moren trenger behandling hos en medisinsk spesialist. Ved hjelp av såkalte agenter (programvare) reserverer de time for moren hos et legesenter mens de kjører bil. Eksempelet innebærer bl.a. samordning av legejournal, forsikringsselskapets system, kalendere,

geografisk avstand og reisetid fra hjem og kontor (inkludert beregnet rushtidstrafikk), behandlingstilbud og vurdering av kvaliteten på behandlingstilbudet hos potensielle legesentre.

Selv mer enn 10 år etter er vi langt unna å realisere en enkel versjon av det som ble beskrevet her. Vi er midlertid mye nærmere en enklere semantisk web der datamaskiner kan koble informasjon fra ulike informasjonskilder automatisk – *lenkede data*.



## Lenkede data

Begrepet *Linked Data* (Lenkede data) ble lansert av Tim Berners Lee i et design notat til arbeidsgruppen for semantisk web. (Berners-Lee, 2006)

Erkjennelsen var at lenkene i stor grad var det som gjorde webben fruktbar, det var lenkene som skapte et nettverk (eng: web) av informasjon.

Lenkede data utnytter kjernen i Semantisk web-teknologi, og fokuserer på de viktigste prinsippene i forhold til å bruke det som en web-teknologi.

Det er fire hovedregler:

1. Bruk URI som navn på ting
2. Bruk http-URI så det er mulig å slå opp på referansen
3. Send nyttige data tilbake hvis noen aksesserer en URI, og bruk standardene til dette (RDF-familien og SPARQL)
4. Lenk data til andre data med http-URI, så det er mulig å følge lenkene videre og finne mer data

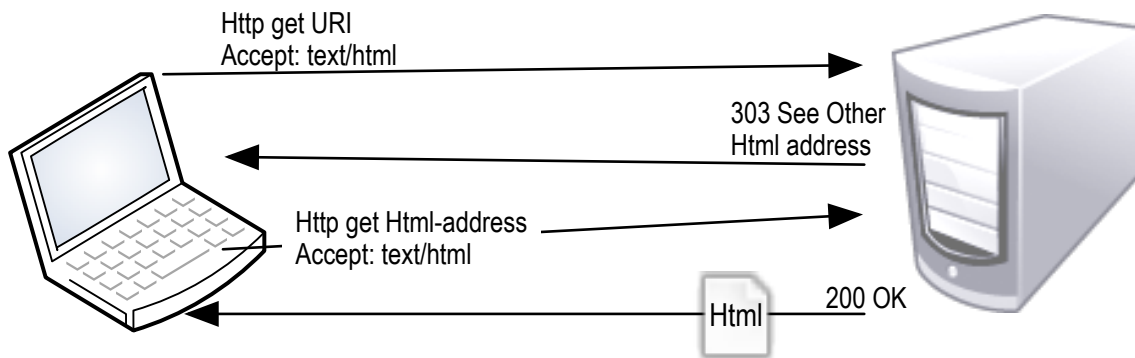
Under en presentasjon på TED-konferansen i 2009 fikk Berners-Lee publikum til å rope ut slagordet sammen med ham (Berners-Lee, 2009):

*Raw Data Now!*

En utfordring er hvordan man skal klare å gi informasjon til mennesker om hva en URI identifiserer (kodet med html), og samtidig sende data hvis en maskin sender en forespørsel. En løsning på dette er såkalte http 303-redirects. "303 See Other" er en av statuskodene i http-standarden<sup>2</sup>.

---

<sup>2</sup> Et eksempel på en annen statuskode, som de fleste nok kjenner bedre, er statuskoden "404 Not Found" som returneres hvis ressursen ikke finnes.



**Figur 1: Http 303 content negotiation**

Figuren illustrerer forhandlingen som skjer hvis en klient spør etter en vanlig html-fil.

1. Klienten spør etter html på en URI
2. Men får en 303-melding fra serveren og en annen adresse som svar
3. Klienten prøver å hente html på den nye adressen
4. Serveren sender html-fil og 200 statuskode som betyr OK

Tilsvarende ville skjedd hvis en klient spurte etter "application/rdf+xml", som er RDF med XML-syntaks (eller et annet format som serveren støtter), men serveren ville da videresendt til en annen adresse der en fil kodet med det ønskede formatet fantes.

Berners-Lee lanserte også en rangering av ulike tilnærminger til å legge datasett åpent tilgjengelig ut på web, *Linked Open Data*. Det viktigste er å få ut dataene åpent tilgjengelig, men det er lettvinde og bedre måter å gjøre det på (Berners-Lee, 2006):

- ★ Available on the web (whatever format) *but with an open licence, to be Open Data*
- ★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)
- ★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)
- ★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
- ★★★★★ All the above, plus: Link your data to other people's data to provide context

En ting som er verdt å merke seg er følgende:

*Large datasets provide a SPARQL query service, but the basic linked data should be provided as well. (Berners-Lee, 2006)*

Dette er et hint til en av de store utfordringene som fortsatt er uløst. Man må i prinsippet velge om dataene skal være navigerbare eller søkbare. Det er derfor relativt vanlig å både legge ut datasettet slik at det kan nåes via http og tilby søkbarhet via et SPARQL endpoint (som er forklart under), i praksis ofte som to forskjellige kopier av dataene. (Giese, 2012)

## RDF kodet med Turtle syntaks

Det finnes flere syntakser for RDF, og Turtle er regnet for å være den enkleste å lese, derfor er det denne som brukes i oppgaven. All RDF-kode skrevet med Turtle kan dessuten brukes i spørrespråket SPARQL. ("Turtle - Terse RDF Triple Language," 2011)

RDF kan sies å bestå av enkle setninger som alltid har tre komponenter, såkalte tripler:

```
subjekt predikat objekt
```

Alle objekter og begreper må gis entydige navn i form av unike identifikatorer, som må være en URI. En URL er en URI, så i eksemplene her bruker vi det. Objekt kan også være en tekststreng.

Hvis vi skal representere setningen "Sult er skrevet av Knut Hamsun", må vi gjenbruke eller definere en URI for hver av delene.

Det enkleste er ID for predikatet, relasjonen mellom subjekt og objekt, her kan vi gjenbruke *creator* fra Dublin Core:

```
http://purl.org/dc/terms/creator
```

Vi kunne ha hentet en identifikator for Hamsun fra BIBSYS' autoritetsregister for personer, <http://data.bibsys.no/data/>, men her kreves både kjennskap til spørrespråket SPARQL og kjennskap til ontologien de har brukt.

Det er per mai 2012 mye enklere å søke i google etter "viaf hamsun". Viaf (Virtual International Authority File) er et internasjonalt samarbeid som kombinerer mange lands autoritetsfiler til én autoritetstjeneste for navn på personer og organisasjoner. Første treff er <http://viaf.org/viaf/59083692/> som er siden for Knut Hamsun. Url for siden er ID for Hamsun, og de ulike lands autoritetsinformasjon er oversiktlig lenket opp fra siden.

Dette er også en bedre ID å bruke hvis målet er å bruke samme ID som flest mulig andre på lang sikt.

Når det gjelder ID for verket Sult er dette et hovedtema i oppgaven, men foreløpig gjenbruger vi ID fra Pode-prosjektet:

```
http://www.bibpode.no/work/Hamsun_Knut_sult
```

For å gjøre setningen mer leselig kan vi definere prefikser som gjør at vi kan forkorte ID-ene, dette fungerer rett og slett som forkortelser slik at koden skal bli mer leselig:

```
@prefix dcterms: <http://purl.org/dc/terms/>.
```

```
@prefix pode: <http://www.bibpode.no/work/>.
```

```
pode:Hamsun_Knut_sult dcterms:creator <http://viaf.org/viaf/59083692/> .
```

Hvis man har flere utsagn om samme subjekt finnes en forkortelse der man kun skriver subjekt-id én gang og bruker semikolon. Her er et eksempel på dette hentet fra vedlegget i oppgaven med flere utsagn om verket sult:

```
local-ontology:id/work/00000002
  a      frbr:Work ;
  dcterms:title "Sult"^^xsd:string ;
  fabio:hasPublicationYear "1890"^^xsd:int ;
  fabio:hasManifestation
    local-ontology:id/manifestation/hamsun-sult-1944 ,
    local-ontology:id/manifestation/9788205394810 ,
    local-ontology:id/manifestation/8205277486 ,
    local-ontology:id/manifestation/8205248931 ;
  frbr:creator <http://viaf.org/viaf/97006051/> .
```

## SPARQL

SPARQL (SPARQL Query Language for RDF) er spørrespråket som brukes for RDF ("SPARQL Query Language for RDF," 2008). Det minner litt om SQL for RDF. Det er ikke rom for skrive en innføring i SPARQL i denne oppgaven, men fordi spørrespråket vil brukes for å teste ontologien vises prinsippet med et par eksempler.

Hvis man har en RDF-database, som ofte kalles triple-store, har man vanligvis implementert SPARQL. Hvis det er implementert som en webtjeneste som aksepterer SPARQL-uttrykk som input og returnerer resultatene er det vanlig å si at man har et SPARQL endpoint.

Vi kan ta utgangspunkt i et utsagn fra forrige kode-eksempel. Verket Sult har id

```
local-ontology:id/work/00000002
```

Det er lagt inn relasjon til fire ulike manifestasjoner med predikatet

```
fabio:hasManifestation
```

Hvis vi ønsker å spørre etter alle manifestasjoner i et SPARQL endpoint kan vi skrive følgende (har her droppet definisjon av prefix som i en ekte situasjon må med):

```
local-ontology:id/work/00000002 fabio:hasManifestation ?manifestasjoner
```

Vi kan altså erstatte deler av utsagnet med en variabel som starter med "?". SPARQL vil da finne alle utsagn der dette mønsteret passer.

Vi må også fylle ut med spørre-syntax som ligner litt på SQL:

```
select *
where

{local-ontology:id/work/00000002 fabio:hasManifestation ?manifestasjoner
.}
```

Resultatet ved spørring mot ontologien i vedlegg 2 ville vært registrerte manifestasjoner relatert til dette verket.

Et annet eksempel er å finne manifestasjonen som har et spesielt ISBN-nummer. Vi bruker predikatet prism:isbn for å angi en relasjon mellom manifestasjonen og ISBN-nummeret, så her spør vi etter subjekter som er relatert til et spesifikt ISBN-nummer.

```
select ?manifestation
where
{?manifestation prism:isbn "82-05-21006-3" .}
```

Svaret er Sværmere (1993) av Knut Hamsun.

## Ontologirammeverket Protégé

Jeg velger å bruke verktøyet Protégé både til modellering av ontologi og registrering av instanser. Dette er sannsynligvis den mest brukte ontologieditoren for RDF og OWL.

En ferdig utviklet løsning vil måtte bruke et triplestore til å lagre modell og informasjon, og bruke dette som et SPARQL endpoint som andre kan søke i. Både PODE-prosjektet, BIBSYS (rådata nå) og British Library (nasjonalbibliografi) har gjort dette. Oppsett av dette åpent tilgjengelig på et webdomene er dessverre ikke gjennomførbart innenfor denne masteroppgaven.

Det er imidlertid mulig å teste aktuelle spørringer med SPARQL direkte i Protégé. Som test av prototypen kjøres SPARQL-spørringer mot ontologien og eksempel-instanser i Protégé for å se om designkriteriene er tilfredsstillende.

## Verkbegrepets historie

Bibliografiske systemer har hatt grunnleggende problemer med å forholde seg til verket som abstrakt bibliografisk entitet i bibliografiske modeller i mer enn 170 år etter at Panizzi formulerte de moderne bibliografiske grunnprinsippene og behov i 1841.

Jeg vil argumentere for at behovet for å forholde seg til verket har vært tilstede siden tidenes morgen.

### *I begynnelsen var ordet, og verket*

Før bibliotekene, før skriftspråket, fantes muntlig litteratur. Myter, sagn, eventyr, sanger, lover og religiøse fortellinger ble overlevert muntlig fra generasjon til generasjon. Muntlig tradisjon var grunnlaget for overlevering av kultur og kunnskap i gamle samfunn før skriftspråket oppstod, og også i mange år etter. Mange norske folkeeventyr og fortellinger ble ikke skrevet ned før på 1800-tallet.

I en tradisjon med muntlig overlevering er verket det essensielle. Kilden for en slik muntlig overlevering er stort sett ukjent eller anonym, og hver fremføring av en historie er en unik versjon av den. Samme historie vil etterhvert kunne få flere ulike variasjoner, men folk forholder seg likevel til *historien*, dvs. det vi kan kalle *verksbegrepet*.

Det vil uunngåelig oppstå variasjoner, og det er også kjent at poeter i muntlig tradisjon gjerne tilpasset innholdet til mottagerne. For endel typer litteratur, som f.eks. sagn, myter, lover og religiøse fortellinger, var det imidlertid svært viktig at historiene forandret seg minst mulig etterhvert som de ble gjenfortalt, og tilhørerne førte streng kontroll med at overleveringen var riktig. Verseform ble innført som hjelp til å holde formen og innholdet mest mulig fast. Verse målet (metrikken) gjorde at innholdet kunne holdes mest mulig statisk ved hjelp av rytme og rimmønster. (Filliozat, 2005) (Muntlig tradisjon, udatert) (Muntlig litteratur, udatert)

Jeg mener vi kan her snakke om verkets opprinnelse. Historien kom før boken, verksbegrepet oppstod før skriftspråket, før bibliotekene. Verket er den opprinnelige bibliografiske entiteten.



## Sivilisasjonen og skriftspråket oppstår

De første sivilisasjonene oppstod i den vestlige kulturens vugge i Midtøsten. Yngre steinalder kalles også bondesteinalderen eller neolittisk tid, og var en overgang fra nomadiske jeger-samfunn til bofaste jordbrukssamfunn, som i Midtøsten startet for ca. 12000 år siden. Dette sammenfaller med at den siste store istiden sluttet brått for 11700 år siden. (global warming (Earth science) : Climatic variation since the last glaciation, udatert)

Verdens første bysamfunn begynte å utvikle seg fra ca. 8000-6000 år f.Kr (Jeriko og Çatalhöyük) (Soja, 2011), og den såkalte keramiske epoken startet når keramikk og leirvarer oppstod ca. 7000 år f.Kr. (Den fruktbare halvmåne, udatert)

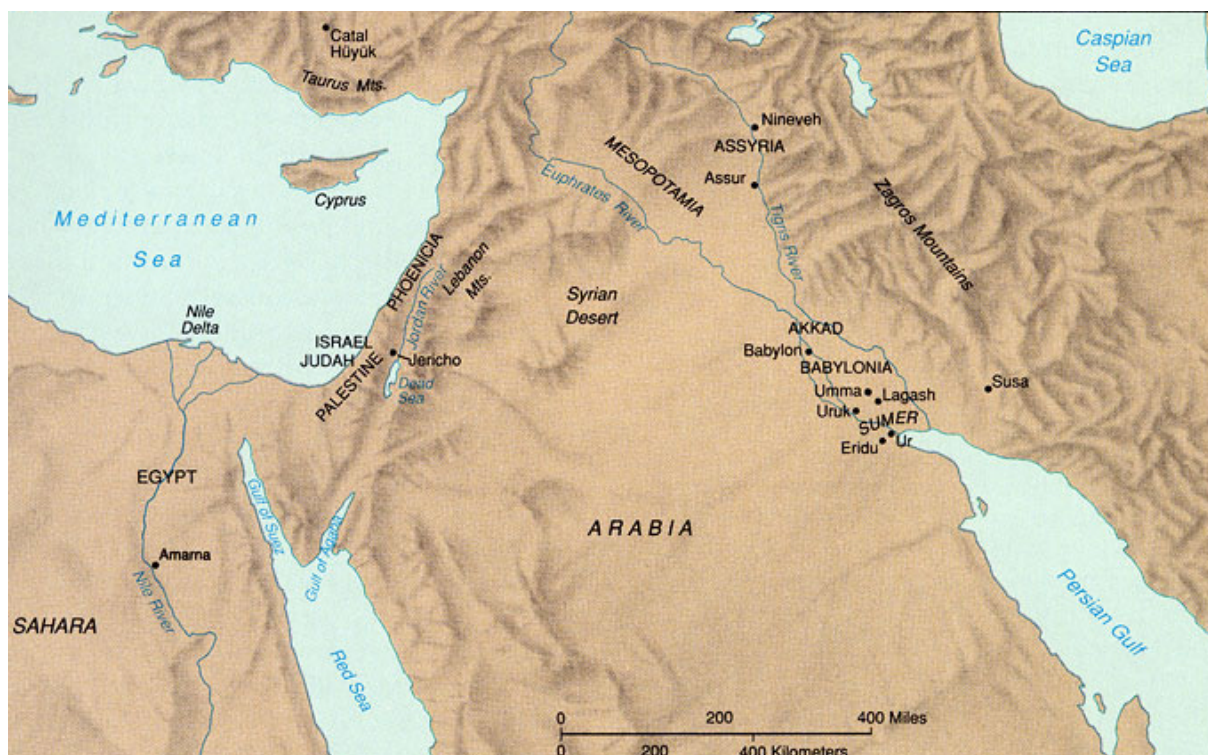
Den sumeriske oldtidsbyen Uruk var en av verdens første bystater, dannet ca. 4000 år f.Kr(Lawler, 2012), og utviklet seg etterhvert til verdens største by med ca. 80 000 innbyggere (Taylor, 2012). Dette var hovedstaden til kong Gilgamesh ca. år 2500 f.Kr.

Skriftspråket oppstod som et administrativt verktøy etterhvert som organiserte bysamfunn grodde frem. De eldste eksemplene på skriftspråk er leirtavler fra ca. 3300 f.Kr, som er funnet i Uruk. Det er antatt at Uruk hadde ca. 20 000 innbyggere da skriftspråket oppstod i Mesopotamia i overgangen mellom yngre steinalder og bronsealderen, ca. 3400 år f.Kr.(Glassner, 2003, s. 1).

De tidligste leirtavlene med skrift hadde såkalt logografisk skrift, der hvert tegn var piktogrammer, dvs. symboler som representerte et ord. Dette utviklet seg til kileskrift over de neste tusen årene. (Glassner, 2003, s. 122)



Figur 2: Utvikling av kileskrifttegnet SAG, som betyr hode, fra ca. 3000-1000 år f.Kr.(SAG, 1990)



Figur 3: Sumerian Civilization (Sumerian Civilization, 2008)

### *Det første biblioteket?*

Skriftspråk og en samling nedskrevne tekster er en forutsetning for et bibliotek, men det er et definisjonsspørsmål når en samling blir et arkiv, og når et arkiv går over til å bli et bibliotek.

De fleste har i våre dager en god del bøker hjemme, men de færreste føler noe stort behov for å konstruere en egen katalog for å holde orden. Tilsvarende måtte arkiver bli av en viss størrelse før det var nødvendig å bruke andre hjelpemidler enn en systematisk lagring av tekster.

Et bibliotek er ikke kun en haug med tekster. Man kan argumentere for at både utvalg av tekster, sjangere, og en viss støtte for gjenfinning av tekstene må til før man kan karakterisere en tekstsamling for et bibliotek. Et bibliotek av en viss størrelse forutsetter en form for bibliotekskatalog, som inneholder bibliografisk informasjon, metadata om tekstene, for at gjenfinning skal støttes tilfredsstillende.

Man må imidlertid ikke se seg blind på at katalogen er forutsetningen for at man skal kalle noe et bibliotek. Katalogen er både et hjelpemiddel for administrasjon av samlingen og

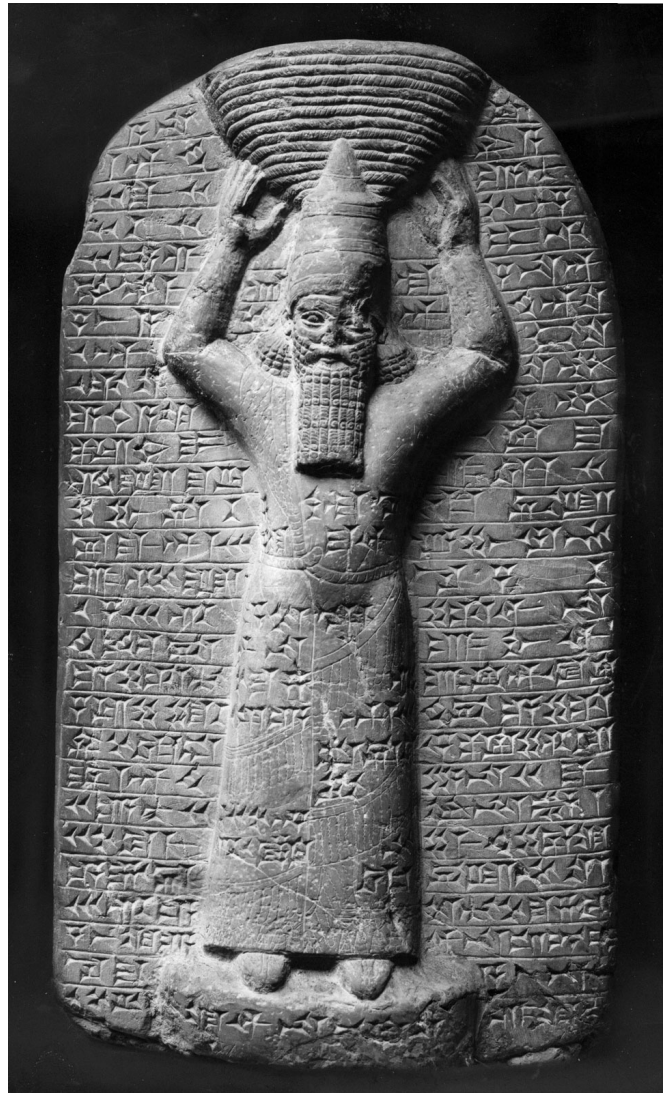
gjenfinning, og behovet for hjelp er avhengig av størrelsen på samlingen. Helt frem til slutten av 1700-tallet var det relativt vanlig å tenke at man ikke hadde behov for en katalog til å holde orden på bøkene i et bibliotek, det var tilstrekkelig med en systematisk oppsetting av bøkene i hyllene.(Blum, 1991, s. 64).

Verdens antageligvis eldste bibliotek/arkiv er funnet i den sumeriske oldtidsbyen Nippur, i Mesopotamia, ca. 15 mil sør for Bagdad. Ca. 40 000 leirtavler der de eldste kan dateres tilbake til ca. 3000 år f.Kr. I Nippur fant man også det som sannsynligvis er verdens eldste bibliotekskatalog: To leirtavler med usorterte lister over hhv. 68 og 62 litterære tekster datert til ca. 2000 år f.Kr. (Weitemeyer, 2009), (Casson, 2002, s. 4), (Peters, 1905)

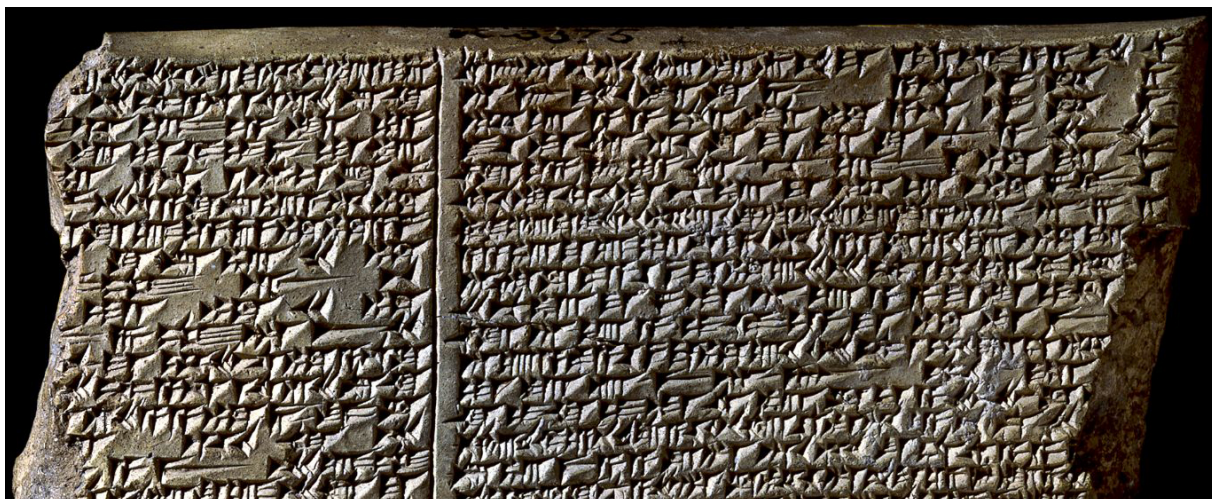
For å finne mer utviklede former for bibliotek må vi sannsynligvis hoppe frem til den assyriske kongen Ashurbanipal, som grunnla det kongelige biblioteket i byen Nineveh i Babylon i perioden 668-630 f.Kr.

Dette er regnet som verdens første systematisk oppbygde bibliotek.

Hovedorganiseringen av samlingen var etter emne. (Bales, 2009, s. 3) Man har i dag funnet mer enn 30 000 leirtavler fra dette biblioteket.(The Library of Ashurbanipal, udatert)



Figur 4: Ashurbanipal ca. 660 f.Kr. (Foto: Public domain/Wikimedia)



Figur 5: Gilgamesh-eposet, funnet i Ashurbanipals bibliotek. AN32864001  
© Trustees of the British Museum

Sagn forteller at Alexander den Store så biblioteket, og ble inspirert til å samle alle verk i områdene han erobret og oversette dem til gresk. (Philips, 2010, s. 1-2) At kunnskap var makt var en svært viktig pådriver for de første bibliotekene, som definitivt ikke var folkebiblioteker.

Det som regnes som historiens første virkelige systematiske bibliotekskatalog og nasjonalbibliografi oppstod imidlertid allerede ca. 250 år f.Kr. i Alexandria.

Byen Alexandria ble grunnlagt av Alexander den store i ca. år 331 f.Kr. Ifølge mytene skal han samtidig ha angitt plasseringen av biblioteket.

### *Opprinnelsen til verksbegrepet i bibliografiske systemer*

De fleste diskusjoner av den historiske opprinnelsen til verksbegrepet i bibliografiske systemer går ca. 150 år tilbake, til Panizzi. (Yee, 1994), (Denton, 2007), (Coyle, 2010)

Jeg mener imidlertid at man må gå mye lenger tilbake.

### **En autoritativ versjon av Homers verker**

Homer er en av verdenshistoriens mest kjente forfattere, men man vet ironisk nok ikke hvor han kom fra og om det er ham som har skrevet både Odysseen og Iliaden. Det er bevist at Både Iliaden og Odysseen har røtter i muntlig tradisjon (Staikos & Cullen, 2004, s. 63-64). Iliaden og Odysseen er kodet på versmålet hexameter, og har hhv. 15680 og 12100 linjer.

Det er usikkert når den første versjonen ble skrevet ned, men én versjon av historien er at Peisistratos, Athens hersker ca. 560 f.Kr, beordret en nedskrevet versjon av Homers verker etter en dispuTT med Megara over hva som var den riktige eller autoritative versjonen av Illiaden (Staikos & Cullen, 2004, s. 69). Man laget her en skriftlig versjon for å få en autoritativ versjon av verket.

Et verk kan sies å være en sosial konstruksjon. Hva som er en autoritativ eller populær oppfatning av grensene for et verk vil kunne være kulturelt avhengig, dvs. det kan variere både i forhold til historisk tidspunkt og være avhengig av brukerens kontekst og subjektive oppfatning.

*Attempts to define the concept of a work and how it relates to texts and documents have been undertaken by many scholars in various disciplines. There is lively debate among the critical, literary, and textual theorists Barthes, Foucault, Tanselle, McGann, Greetham, and Shillingburg, to name just a few. Within the library and information studies community, Patrick Wilson has been a major theoretical voice. Many of these definitions of a work contradict each other and these contradictions often stem from specific social and ideological agendas. In fact, it could be said, that the idea of a work is one that is socially constructed. (Genz, 2002)*

*The concept of a work is an important and controversial concept for many others in addition to cataloging theorists: textual scholars, literary critics, philosophers of art, and copyright lawyers, among others. Michel Foucault remarks that "a theory of the work does not exist". It is true that no generally accepted theory exists; we could collect dozens of different ones. (Wilson, 1989)*

Det er utenfor rekkevidde av denne oppgaven å gå inn i en dypere vitenskapsfilosofisk diskusjon av dette. Det viktige i denne sammenhengen er å være klar over at det er avvikende oppfatninger av hva et verk er, hva som regnes innenfor og utenfor grensene for ett verk, og hvor grensene går mellom forskjellige verk. Regler og avgrensinger av verk i katalogen er definert i de bibliografiske standardene som brukes i biblioteket, f.eks. i FRBR, AACR2 og RDA, men det vil sannsynligvis være avvikende oppfatninger blant brukerne av systemene.

## Kallimakhos og biblioteket i Alexandria

Det er usikkert hvor mange skriftruller som fantes i biblioteket i Alexandria på bibliotekaren Kallimakhos tid, ca. 300 år f.Kr. Antagelsene varierer mellom 400 000 og 700 000 skriftruller. (MacLeod, 2000, s. 5), (Philips, 2010, s. 3)

Biblioteket katalogiserte skriftrullene etter emner, og etter forfatter innen hvert emne. Dette var også en oversikt over kopier, dvs. den tids ulike utgaver av verkene.

Biblioteket ble kjent for å ta originalene og gi tilbake kopier til eierne:

*That books are a passion, and collecting them, a pathology, should be an Alexandrian saying. The Library had little regard for intellectual property - or even for property rights per se. It is said that Ptolemy III (246-221 BC) wrote to all the world's sovereigns, asking to borrow their books for copying. When Athens lent him texts of Euripides, Aeschylus and Sophocles, he had them copied, but kept the originals, cheerfully forfeiting the fortune of fifteen talents he had deposited as bond.*  
(MacLeod, 2000, s. 4)

Alle dokumenter på denne tiden var kopiert for hånd, tegn for tegn. Skriftspråket hadde på denne tiden allerede eksistert i mer enn 2000 år, og skrivefeil og avvikende versjoner var et problem.

Datidens kopiering førte til ekstra store utfordringer i forhold til å holde oversikt over originaltekst, senere kopier og hvilke kopier disse igjen var basert på. Det man kan kalle tekstkritisk lesning var derfor en viktig oppgave, og man jobbet systematisk med å finne frem til en opprinnelig versjon av de viktigste verkene. (Blum, 1991, s. 64)

Bibliotekaren Kallimakhos implementerte katalogen pinakes, der tema, forfatter og verk var de viktigste bibliografiske entitetene, ca. 300 år f.Kr:

*Even the members of the Museion were better served by a catalog of works than by a catalog of copies, which was unwieldy because it listed all extant specimens. These were of interest only for the philologists who were concerned with textual criticism.*  
(Blum, 1991, s. 238)

## *Gjenoppdagelsen av verket*

Etter biblioteket i Alexandria gikk det nærmere 1500 år før behovet for å holde styr på verk og kopier i det bibliografiske systemet ser ut til å dukke opp igjen.

Sir Thomas Bodley tok i 1598 på seg oppgaven med å gjenoppbygge universitetsbiblioteket i Oxford. Under arbeidet med å kjøpe inn bøker til biblioteket fant han ut at katalogen ofte ikke kunne gi svar på om samlingen inneholdt et verk fra før, bl.a. hvis verket inngikk som en del av et annet verk og ikke var ført inn under verkets forfatter. (Lubetzky, 1953) På denne tiden var katalogen i bokform. Katalogen fra 1620 var den første kjente moderne katalog som var ordnet primært på forfatter, og dermed samlet alle verkene til en forfatter i katalogen. (Yee, 1994, s. 3)

Innledningen til 1674-versjonen av katalogen kan sies å være det første eksempelet på katalogiseringsregler i den anglo-amerikanske tradisjonen. Her beskrives det at alle verk av en forfatter skal samles under en enhetlig skriveform av forfatterens navn. Implisitt i dette lå det også at alle utgaver av verket også samles. Sorteringen av utgaver på språk og publiseringsdato kan sees i 1674-versjonen av katalogen. (Yee, 1994, s. 3) (Smiraglia, 2002, s. 333)

I 1698 publiserte Frederic Rostgaard en ny metode for å bygge opp en bibliotekskatalog der alle utgaver av samme verk alltid ville gjenfinnes samlet i katalogen. (Strout, 1956, s. 266)

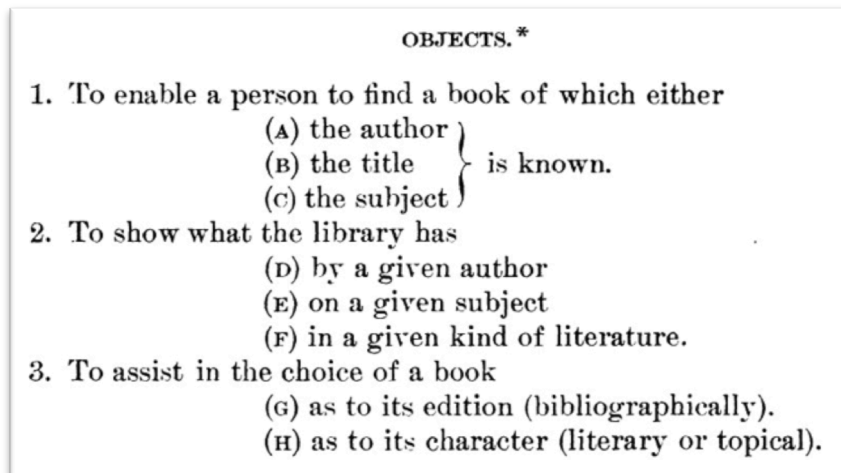
Italieneren Antonio Panizzi (senere sir Anthony Panizzi) satte problemstillingen på kartet i arbeidet sitt med å realisere et nytt bibliografisk system for British Museum fra 1836.

Senere bibliografiske systemer har strevet med noen grunnleggende forutsetninger i mer enn 150 år etter at Panizzi formulerte de moderne grunnprinsippene i *Rules for the Compilation of the Catalogue*, etter hvert kjent som *Panizzi's 91 rules*.

*A reader may know the work he requires; he cannot be expected to know all the peculiarities of different editions, and this information he has a right to expect from the catalog.* Antonio Panizzi ca. 1850 (Svenonius, 2000, s. 11)

Det samme grunnleggende problemet er dessverre fortsatt tilstede både om man handler hos Amazon eller forsøker å låne en bok på biblioteket. Jeg er vanligvis interessert i å lese et bestemt verk, men må gjerne rote en del rundt for å få oversikt over pris og/eller tilgjengelighet for ulike utgaver.

Charles A. Cutter skrev i 1876 rapporten *Rules for a dictionary catalog*, der han formulerte formålet med bibliotekskatalogen, her hentet fra 4. Utgave fra 1904 (Cutter, 1904):



Legg merke til at Cutter ikke beskriver verket eksplisitt. "what the library has by a given author" er i beste fall vagt.

De første anglo-amerikanske katalogiseringsreglene fra 1908 inkluderte ikke regler for ordning av kortene i kortkatalogen. Dette ble overlatt til separate veiledninger. Hvert bibliotek stod også fritt i å bestemme sine egne ordningsregler for kortene. (Yee, 1994, s. 5)

Library of Congress hadde f.eks. strikte regler for oppsetting av kortene som beskrev at en forfatters verk skulle plasseres sammen. Ulike utgaver av et verk ble gruppert sammen, uavhengig av tittel-enderinger. Ordningen av kortene var imidlertid ikke avmerket på kortene, så andre biblioteker kunne ikke gjenbruke ordningen. (Yee, 1994, s. 5)

Seymour Lubetzky, som av mange er regnet som det 20. århundrets store bibliotek-teoretiker, kritiserte senere Cutter og etterfølgende standarder for at reglene ikke skiller klart og konsistent mellom bok og verk.

Lubetzky formulerte katalogens hovedformål:



*The first objective is to enable the user of the catalog to determine readily whether or not the library has the book he wants.*

...

*The second objective is to reveal to the user of the catalog, under one form of the author's name, what works the library has by a given author and what editions or translations of a given work.*

(Lubetzky 1953a@36}

Det er katalogens andre hovedformål (*second objective*) som er mest interessant i denne sammenhengen.

Ifølge Lubetzky var et klart skille mellom bok og verk en fundamental forskjell på de etablerte katalogiseringsreglene og AACR (Anglo-American Cataloging Rules) som ble publisert i 1967 (Lubetzky, 1969, s. 15). Lubetzkys definisjon av formålet med katalogen ble imidlertid ikke en del av AACR, og ble aldri en formell del av katalogiseringsreglene. (Yee, 1994, s. 8)

*The need for the second objective arises from the fact that the works of an author may be issued under different names as a result of a change, translation, transliteration, or even misprint of the author's name, and the editions of a work may be issued under different titles for similar reasons, and could, therefore, be separated in the catalog.*

*The cataloger is thus required, in recording a work, to establish the identity of the author and the relationships of the work.*

(Lubetzky & American Library Association, 1953, s. 37)

AACR var basert på de såkalte Paris-prinsippene, resultatet fra den internasjonale konferansen om katalogiseringsprinsipper i 1961. Disse prinsippene var i stor grad basert på Lubetzkys arbeid, og reformulerte hovedformålet for bibliotekskatalogen (Denton, 2007):

*The catalogue should be an efficient instrument for ascertaining*  
*2.1 whether the library contains a particular book specified by*

*(a) its author and title, or*

*(b) if the author is not named in the book, its title alone, or*

*(c) if author and title are inappropriate or insufficient for identification, a suitable substitute for the title; and*

- 2.2 (a) *which works by a particular author and*  
(b) *which editions of a particular work are in the library*  
(Svenonius, 2000, s. 16)

Dette var en tid da kortkatalogen fortsatt var i bruk. MARC (MAchine-Readable Cataloging) ble utviklet av Library of Congress på 60-tallet og ble en internasjonal standard i 1973. MARC-formatet kan sies å være en modell av de gamle papirbaserte bibliotekskortene. Formatet begynte som et prosjekt der man så på mulighetene for å konvertere informasjonen på papirbaserte bibliotekskort til elektronisk form med tanke på å trykke opp kort. (Avram, 1976, s. 1712, 1722, 1724) Strukturen er bestemt av ANSI-standarden (American National Standards Institute) Z39.2 (senere ISO 2709), som ble utviklet som en standard for koding av informasjon på trykte bibliotekskort. Dette var en av de første IT-standardene som ble laget.

Det er ca. 160 år frem i tid fra Panizzi til rapporten *Functional Requirements for Bibliographic Records* (FRBR), som ble sluppet av IFLA i 1998. (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1997/2009) Rapporten bygget videre på Paris-prinsippene og lanserte en ny konseptuell modell for bibliografiske data, som man mente var nødvendige for å kunne løse brukernes grunnleggende behov. Vyene var store, men det videre arbeidet har gått tregt fremover. Det er naturlig å spørre seg selv hvorfor ikke FRBR har fått mer gjennomslag som modell i biblioteksystemer de siste 10 årene.

Et svar på dette kan ironisk nok være utfordringene som oppstod da de elektroniske katalogene med MARC-kodede poster ble innført som en erstatning for kortkatalogen.

At gamle løsninger skaper store problemer i overgangen til nyere teknologi er ikke noe nytt:

*The mere lapse of time aggravates old problems and creates new ones – and some of the most serious new problems are the old solutions, which seemed good to us a generation or two ago but which now embarrass us because we should like to change them if we knew how, if it were not too difficult, and if we could afford to pay the necessary cost.*

(Hoesen, 1944, s. 100)

Med en kortkatalog hadde man store muligheter for å legge arbeid i sammenstilling av kortene, det vi kan kalle sammenstilling av informasjon i katalogen. Dette arbeidet var imidlertid som nevnt tidligere ikke en del av katalogiseringsreglene.

Ironisk nok tapte man en god del på digitaliseringen av papirbaserte kortkataloger og overgang til MARC-formatet (Hegna & Murtomaa, 2003).

Det viser seg at tapet av struktur i oppstilling av postene (papirkortene vs. MARC-postene) i katalogen ikke har latt seg erstatte av fritekstsøk. Ref. søkeeksemplene fra Deichman i begynnelsen av neste kapittel.

James Duff Brown ga i 1916 ut boken *Library classification and cataloguing* (Brown, 1916), og den er på mange måter like relevant i dag. Det er fortsatt hit vi ønsker å komme per 2012, nesten hundre år senere.

Han foreslår følgende oppstilling av kort i kortkatalogen for verkene til en forfatter som har mange verk og utgivelser:

5. Single works in chronological order of publication
  - Originals, including manuscripts
  - Reprints
  - Translations in foreign languages
  - Parodies
  - Criticism, etc., of single works
6. Collected works in chronological order of publication
  - Author's editions
  - Editor's editions
  - Selections
  - Paraphrases and condensed version
  - Dramatic versions of works
  - Musical settings of works

I kortkatalogen ville dette materialisert seg ved at man ved oppslag på forfatteren ville møte en skuff med forfatterens samlede produksjon, delt opp kronologisk i verk (Hegna, 2007).

Dette gjorde det mye enklere å få oversikten over forfatterens bibliografi enn i dag.

Man kan si at både forfatteren, verkene og uttrykkene fantes som objekter i denne katalogmodellen, selv om det i første rekke var i form ordningen av kortene og såkalte fanekort, som hadde som hovedfunksjon å vise grupperingen de vanlige bibliotekskortene.

Selv om informasjon om forfatter, verk og uttrykk er implisitt i en MARC-post eksisterer ikke disse informasjonsobjektene i dagens bibliografiske modell, som kun baserer seg på poster som er katalogiserte manifestasjoner.

Det intellektuelle arbeidet med oppstilling av katalogkort forsvant ved overgangen til elektronisk informasjon, og ble ikke erstattet. Resultatet ble en stor mengde av separate poster med flat struktur slik vi har i dag (Hegna, 2007).

Hvis man søker i en elektronisk katalog i dag (OPAC, - Online Public Access Catalog) vil resultatlisten ofte være uten struktur, selv om MARC-postene er basert på samme katalogiseringspraksis som kortene i kortkatalogen var.

## Verk i bibliografiske systemer

Hva er status for utviklingen av bibliografiske systemer i Norge og internasjonalt? Hvilke teknologivalg gjøres ved utvikling av bibliografiske systemer i 2012? Er FRBR og Lenkede data naturlige teknologivalg, gitt utviklingen hos BIBSYS, Nasjonalbiblioteket og viktige internasjonale biblioteksaktører?

### *Deichmanske Biblioteks webkatalog*

I oppstarten av arbeidet med oppgaven gjorde jeg et søk etter *Ibsen* hos Deichmanske bibliotek, Norges største folkebibliotek. Av de 25 første treffene hadde ingen Ibsen som forfatter, og representerte altså ingen av forfatterens verk.

Trefflisten så ut som en relativt tilfeldig oppstilling av ulike utgaver av bøker og andre medier på ulike språk. Det var ikke lett å se hvilken sammenheng det var mellom søkeordet og en del av treffene:

**Deichmanske bibliotek** Søk i hele bibliotek  
 Søk i flere bibliotek

Nytt søk   Søkemetoder   Databaser   MappaMi   Hjelp   ( 0 )

Du søkte på **Ibsen** som Fritekst Side 1 / 25 ->> (vis alle)

**Søket ga 1250 treff i følgende kategorier:**

<b>Medier:</b>	Bok (428) DVD (16) VHS (12) Musikknoter (12) CD Musikk (7)	mer... (3)
<b>Språk:</b>	Bokmål (307) Engelsk (61) Flerspråklig (45) Dansk (43) Norsk (nynorsk) (18)	mer... (23)
<b>Sjangre:</b>	Biografier (47) Spillefilmer (21) Norske essays (9) Festskrifter (7) Hørespill (7)	mer... (10)
<b>Emner:</b>	Ibsen, Henrik (207) Biografier (49) Biografier Forfattere (48) Norsk litteratur (37) Historie og kritikk (35)	mer... (671)

**Litteraturtyper:** Faglitteratur (346) Skjønnlit. (139) Skuespill (79) Antologi (13) Dikt (10)  
Tilpass søket

Bilde	#	Forfatter	Tittel	Medietype	År	Status	Utv.liste
	1		<b>100 år etter : om det litterære livet i Norge i 1890-åra</b>	Bok	1993	På hylla	
	2		<b>Anglo-Scandinavian cross-currents</b>	Bok	1999	På hylla	
	3		<b>Anna Stouslands minnebok</b>	Bok	1997	På hylla	
	4		<b>Bilder fra Henrik Ibsens liv</b>	VHS	1996	På hylla	
	5		<b>Boken om bøker : aarskrift for bokvenner</b>	Bok	1926-1945	På hylla	
	6		<b>Boken om mor : en litterær antologi fra Henrik Wergeland til vår egen tids forfattere</b>	Bok	1988	På hylla	
	7		<b>Bokspor : norske bøker gjennom 350 år</b>	Bok	1993	På hylla	
	8		<b>"Bunden af en taksyld uden lige" : om svenskspråklig Ibsen-formidling 1857-1906</b>	Bok	2005	På hylla	
	9		<b>Cogito . 10</b>	Bok	2005	Utlånt	
	10		<b>Con amore : festskrift til Arne Hannevik på 70-årsdagen 15.12. 1994</b>	Bok	1994	På hylla	

Figur 6: Eksempel på søk i katalogen til Deichmanske bibliotek 7. Oktober 2009

Hovedproblemet, slik jeg ser det, er at hverken forfatteren eller verkene hans har en selvstendig plass i katalogen. De eksisterer ikke som objekter i modellen som systemet er basert på. Det er kun utgavene, manifestasjonene av verkene, som er objekter i systemet i form av MARC-poster.

Det har vært en liten utvikling i brukergrensesnittet fra 2009 til 2012, men fortsatt har verken forfatteren eller verkene hans noen selvstendig plass i katalogen:

The screenshot shows the Deichmanske Bibliotek website interface. At the top, there is a search bar with the text 'Ibsen' entered. To the right of the search bar are fields for 'Lånernummer' and 'PIN-kode', and buttons for 'Forbli pålogget', 'Glemt PIN?', and 'Logg inn'. Below the search bar, it says 'Søket ga 1262 treff' and 'Sortering Relevans'. The search results are listed in a table-like format, with each entry showing a book cover, the title 'Ibsen', the author, and the publication year. Each entry has 'Bestill' and 'Husk denne' buttons. On the right side, there are several filter menus: 'Huskeliste', 'Du søkte', 'Avgrens søket', 'Medier', 'Språk', 'Tekstspråk', 'Sjangre', 'Emner', and 'Litteraturtype'. The 'Du søkte' menu shows 'Ibsen' as the search term. The 'Avgrens søket' menu shows various categories and their counts, such as 'Bok (1142)', 'CD Lydbok (30)', 'DVD (23)', 'Musikknoter (19)', 'CD Musikk (18)', 'mer... (3)'. The 'Språk' menu shows 'Bokmål (817)', 'Engelsk (151)', 'Dansk (63)', 'Tysk (44)', 'Fransk (37)', 'mer.. (38)'. The 'Tekstspråk' menu shows 'Engelsk (19)', 'Bokmål (9)', 'Italiensk (2)', 'Finsk (1)', 'Portugisisk (1)', 'mer.. (2)'. The 'Sjangre' menu shows 'Biografier (98)', 'Biografier Forfattere (96)', 'Hørespill (26)', 'Filmer (19)', 'Filmer Drama (18)', 'mer... (44)'. The 'Emner' menu shows 'Ibsen, Henrik (422)', 'Historie og kritikk (70)', 'Norsk litteratur (57)', '839.8209 (56)', 'Peer Gynt (40)', 'mer... (1278)'. The 'Litteraturtype' menu shows 'Faglitteratur (732)', 'Skjønnlit. (516)', 'Skuespill (377)', 'Dikt (38)', 'Roman (19)', 'mer... (7)'. The search results list includes:
 

- Ibsen** [redaksjon: Espen Hammer ... [et al.]] på Bokmål | utgitt 1993. Ledig (Avd.: Voksenavdelingen, Hovedbiblioteket - Plassering: m 839.82 Ibsen,H. )
- Ibsen** av Michael Meyer ( abridged by the author ) på Engelsk | utgitt 1992. Ledig (Avd.: Voksenavdelingen, Hovedbiblioteket 839.82 Ibsen, H. )
- Ibsen** av Willy Dahl ( Billedred.: Else Gullvåg ) på Bokmål | utgitt 1974. Ledig (Avd.: Voksenavdelingen, Hovedbiblioteket 839.82 Ibsen,H. )
- Ibsen** av Willy Dahl på Bokmål | utgitt 1974. Ledig (Avd.: Voksenavdelingen, Hovedbiblioteket u 839.82 Ibs )
- Ibsen** av George Wilson Knight på Engelsk | utgitt 1962. Ledig (Avd.: Voksenavdelingen, Hovedbiblioteket - Plassering: m 839.82 Ibsen, H. )
- Ibsen** av Aurelien Marie Lugué-Poë på Fransk | utgitt 1936. Ledig (Avd.: Voksenavdelingen, Hovedbiblioteket - Plassering: m 839.82 Ibsen, H. )
- Ibsen** av Edmund Gosse utgitt 1907. Ledig (Avd.: Voksenavdelingen, Hovedbiblioteket - Plassering: m BI 839.82 Ibsen, H. )

Figur 7: Eksempel på søk i katalogen til Deichmanske bibliotek 5. mai 2012

Det kan innvendes at det er gjort et vanlig søk i katalogen, og at et avansert søk kunne ha gitt bedre resultater. Jeg antar imidlertid at mange brukere benytter søkefeltet på forsiden,

siden det er dette som er standard søkefelt. Jeg tar også utgangspunkt i at et forfattersøk er et av de aller vanligste søk i en katalog, og at dette er et av hovedformålene med katalogen.

## *Pode-prosjektet*

Pode-prosjektet tok utgangspunkt i akkurat denne utfordringen. Prosjektet testet FRBR, semantiske teknologier og Lenkede data for å se om dette kunne bedre brukeropplevelsen (Podeprosjektet , 2011a).

Podeprosjektet sprang ut av biblioteklaboratoriets mashup-konkurranse i 2007, og var finansiert av ABM-utvikling. Prosjektorganisasjonen var plassert hos Deichmanske bibliotek.

Prosjektet hadde brukeren og brukertjenester som hovedfokus, og utforsket FRBRisering av folkebibliotekskatalogen og RDF (Resource Description Framework) som alternativ til MARC-formatet. Det ble brukt semantiske teknologier (RDF/SPARQL) til å publisere bibliografiske informasjon som Lenkede data.

Hovedfokus i FRBRiseringsprosessen var nettopp på brukervennlighet i trefflistene (Podeprosjektet , 2011). Prototypen *Linked Authors* var for søk etter forfatter, og trefflisten lister opp forfatterens verk kronologisk, akkurat som anbefalt i *Library classification and cataloguing* (Brown, 1916). Webkatalogen ser endelig ut til å nå opp til kortkatalogens brukervennlighet.

- Den Gaadefulde (1877)
- Bjørger (1878)
- Et gjensyn (1878)
- Fra det moderne Amerikas Aandsliv (1889)
- Lars Oftedal (1889)
- Sult (1890)
- Mysterier (1892)
- Redaktør Lynge (1893)
- Ny Jord (1893)
- Pan (1894)
- Ved Rigets Port (1895)
- Livets Spil (1896)
- Siesta (1897)
- Aftenrøde (1898)
- Victoria (1898)

- Munken Vendt (1902)
- I Æventyrland (1903)
- Kratskog (1903)
- Dronning Tamara (1903)
- Det vilde Kor (1904)
- Sværmere (1904)
  - ◆ Bok på dansk
  - ◆ Bok på engelsk
  - ◆ Bok på fransk
  - ◆ Bok på italiensk
  - ◆ Bok på norsk bokmål
    - \* Samlede romaner. 1932
    - \* Samlede verker. 1934
    - \* Samlede verker. 1954
    - \* Samlede verker. 1963. 395 s.
    - \* Samlede verker. 1976. 395 s.
    - \* Samlede verker. 1992. 395 s.
    - \* Samlede verker / Knut Hamsun. 2007. 209 s.
    - \* Sværmere. 1904. 133 s.
    - \* Sværmere. 1972. 88 s. 1 bl.
    - \* Sværmere / Knut Hamsun. 1993. 88 s.
    - \* Sværmere / Knut Hamsun ; etterord av Nils M. Knutsen. 1996. 130 s.
  - ◆ CD-plate, Lydbok på norsk bokmål
  - ◆ Lydbok, Lydkassett på norsk bokmål
  - ◆ Bok på polsk
  - ◆ Bok på rumensk
  - ◆ ...

I eksempelet er det klikket på *Sværmere*, og en liste over uttrykk vises sortert på språk. Av disse er *Bok på norsk bokmål* valgt, og en liste over manifestasjonene vises i kronologisk rekkefølge i forhold til utgivelse.

Prosjektet søkte om penger til videreføring av arbeidet i 2011, men fikk avslag fra Nasjonalbiblioteket. PODE-prosjektet avsluttet sin siste bloggpost med en trepunkts ønskeliste basert på arbeidet i prosjektet (PODEprosjektet , 2011b):

1. Åpen dokumentert tilgang til alle katalogdata
2. Et moderne metadataformat som er åpent og fleksibelt
3. Nasjonalt autoritetsregister og åpne, nasjonale unike ID-er for verk og forfattere

Ønskene var utformet som en utfordring til Nasjonalbiblioteket, katalogfaglige komiteer på nasjonalt (og internasjonalt) plan og biblioteksystemleverandørene.



## *Det norske biblioteklandskapet*

BIBSYS har siden PODE-prosjektet ble avsluttet publisert sitt autoritetsregister for navn som lenkede data under navnet *Rådata nå!* Dette vil si at det nå finnes unike id-er for forfattere. Prosjektet ble valgt til årets bibliotekprosjekt for fagbiblioteker 2011 for å publisere sitt autoritetsregister for navn åpent tilgjengelig som lenkede åpne data. (Grønnestad, 2012)

BIBSYS er også i gang med å realisere et nasjonalt autoritetsregister for personer og korporasjoner på oppdrag fra Nasjonalbiblioteket.

BIBSYS planlegger å erstatte sitt egenutviklede system med OCLC WorldShare Management Services i 2013. Dette innebærer en overgang til MARC21. Dette innebærer at brukerne skal bruke Worldcat local som system, og blir tett knyttet opp mot OCLC og Worldcat. (Røyneberg, 2010)

BIBSYS og nasjonalbiblioteket har ingen planer om å lage et autoritetsregister med verks-ID<sup>3</sup>.

## *Internasjonale utviklingstrekk*

### **FRBR og RDA**

RDA er en ny registreringsstandard som er ment å ta over etter AACR2. Den bygger på FRBR og gjenbruker entitetene definert i FRBR og FRAD (Oliver, 2010a).

Det er etterhvert mange land som har besluttet å ta i bruk RDA, bl.a. USA og Storbritannia. Både *Library of Congress*, *OCLC* og *British Library* støtter dermed RDA. (Library of Congress, 2011), (OCLC, 2011)

### **RDF og Lenkede data**

*Bibliographic Framework Transition Initiative* er lansert av *Library of Congress*:

*The new bibliographic framework project will be focused on the Web environment, Linked Data principles and mechanisms, and the Resource Description Framework (RDF) as a basic data model. (Library of Congress, 2011)*

---

<sup>3</sup> Muntlig svar fra Jonny Edvardsen på spørsmål fra Nils Pharo ved Bibliotekmøtet 23. Mars 2012.

OCLC og Worldcat satser stort på RDF og Lenkede data i 2012. OCLC leder arbeidet med å bruke RDF og Lenkede data til publisering av Viaf-identifikatorer (Virtual International Authority File) for personer og korporasjoner. De har et tilsvarende prosjekt i gang rundt å publisere Dewey data som Lenkede data. OCLC har også ansatt Richard Wallis, en kjent person innenfor fagfeltet Semantisk web, i stillingen som teknologievangelist. (Zaino, 2012)

*The British Library* utvikler en versjon av den britiske nasjonalbibliografien som skal tilgjengeliggjøres som Lenkede åpne data. ("Free Data Services," udatert)

## Behovet for identifikatorer

Vi bruker identifikatorer alle sammen, hver dag. Å identifisere ting er nært knyttet til språk og navn. Navn er en slags identifikator som angir hva vi snakker om. Det er imidlertid ikke en unik identifikator, men mennesker er gode til å disambiguere – kunsten å tolke hva den andre parten snakker om ut ifra hva som kommuniseres.

Selv mellom mennesker er tydelig kommunikasjon vanskelig. Den finske Professoren Osmo A. Wiio, som forsket på menneskelig kommunikasjon formulerte *Wiios lover*, der første lov er: (Korpela, udatert)

*Communication usually fails, except by accident*

En alternativ tolkning er at mennesker er usedvanlig flinke til å takle uklar kommunikasjon, i motsetning til datamaskiner.

(Piantadosi, Tily, & Gibson, 2011) argumenterer for at språk er et effektivt verktøy for menneskelig kommunikasjon nettopp fordi det er vagt og flertydig og avhengig av kontekst for å forstås.

Bibliotekbrukere bruker navn på forfattere og titler på verk for å finne det de leter etter i bibliotek katalogen. Tidligere slo de opp på forfatter i den alfabetiserte kortkatalogen, mens de i dag bruker de samme data i søk. Hvorfor er det problematisk å bruke navn og tittel for å identifisere det riktige verket og den riktige personen?

Titler kan forandre seg over tid, og man ender over tid lett opp med flere former av navn og tittel i katalogen. Titler kan lett bli stavet feil og noen verk er kjent under flere titler. Vi lever også i en internasjonal verden og ulike språk bruker ulike titler på verk.

FRBR forsøker å løse dette ved å definere én standardtittel og tillate flere varianttitler.

Man skulle tro at det var enklere med navn, men det motsatte er tilfellet.

### Navn og tittel er dårlig egnet som identifikator

Å bruke navn som identifikator kan føre med seg uoverkommelige språklige utfordringer og feilstavelser, se eksempelet ved Tsjajkovskij under.

Standardtittelen er originaltittelen, dvs. den originale tittelen for boken med den originale skrivemåten.

Satt sammen med forfatterens navn og fødselsår og eventuelt dødsår for forfatteren blir dette brukt som identifiserende tekst i autoritetsregistre.

Forskjellige språk og alfabeter vil her skape uoverkommelige problemer i forhold til å etablere internasjonale identifikatorer basert på standardtittel (kyrillisk, kinesisk, japansk osv.).

Den russiske komponisten Tsjajkovskij har et notorisk vanskelig navn som skrives på utallige måter, og er et godt eksempel på utfordringene som kan oppstå. Noen navnevariasjoner er uunngåelige ved transkribering mellom ulike alfabeter, men andre skyldes rene feil. Her følger et rikt utvalg reelle varianter som er i bruk:

- Pyotr Ilyich Tchaikovsky (en.wikipedia.org)
- Pjotr Iljitsj Tsjajkovskij, Pjotr Tsjajkovskij og Peter Tsjajkovskij (no.wikipedia.org)
- Pyotr Ilyich Chaykovsky (en.wikipedia.org)
- Пётр Ильич Чайковский (en.wikipedia.org)
- Пётр Ильич Чайковский (no.wikipedia.org)
- Петр Ильич Чайковский (Tchaikovsky forum)
- Pyotr Ilyich Tchaikovsky (Tchaikovsky forum)
- Tchaikovsky, Tschaikovsky og Tschaikowsky (Varianter komponisten selv brukte)
- Chaykovskiy (Tchaikovsky forum)

I tillegg til disse er det funnet mer enn 40 ulike navneformer for komponisten Tsjajkovskij i autoritetsregisteret til Library of Congress (authorities.loc.gov). (Tchaikovsky Forum: Translation into English of Tchaikovsky's Name, udatert), (Pyotr Ilyich Tchaikovsky, udatert) og (Pjotr Tsjajkovskij , udatert)

## Meningsbærende identifikatorer

*When you have a large pile of stuff, you need a way to identify it. The more meaningful the names, the worse they scale. (Weinberger, 2005)*

Meningsbærende identifikatorer har endel felles med navn på ting. Det er mulig å lese og tolke hva de står for. Det kan være mulig å gjette seg til hva de skal være, og her ligger en fallgrube. For at de skal være brukbare for maskiner må de være identiske, det kan ikke være skrivefeil.

Det er mange eksempler på feilstavinger i katalogen, og at titler på verk endres over tid, så man kan stille spørsmål ved hvor nyttig en meningsbærende identifikator med f.eks. en innebygget originaltittel ville være.

Ulike titler og navn er dessuten av svært ulik lengde og med forskjellig antall deler. Hvis dette skal brukes i praksis i en identifikator må man lage mange regler for bl.a. forkortelse av navn eller tittel.

Tegnsettet som kan brukes i en URI er begrenset, man kan bl.a. ikke bruke mellomrom, så man måtte kode om endel tegn til andre tegn i det gyldige tegnsett.

Lengde og tegnsettproblematikk er ikke utfordringer for en ikke-meningsbærende ID.

Ved bruk av en identifikator vil verket ha én id og det kan i tillegg registreres flere navn. Det er identifikatoren som er maskinlesbar og enkel å bruke til å sjekke likhet mellom ressurser.

### *Ikke-meningsbærende identifikatorer*

Alternativet er en ikke-meningsbærende id. Viaf-id for komponisten Tsjajkovskij er et eksempel:

`http://viaf.org/viaf/99258155`

Vi er vant til ikke-meningsbærende identifikatorer i dagliglivet vårt:

- Telefonnummer
- Bankkontonummer
- Kredittkortnummer

- Personnummer

Med RDF og Lenkede data må man bruke standardiserte URI identifikatorer. Man bør i tillegg bruke en gyldig http-URI. (Berners-Lee, 2006)

Domenenavnet er avhengig av hvilken organisasjon som gjør det. Hva som kommer etter domenenavnet avhenger bl.a. av arkitekturen på løsningen. Følgende er eksempler på ikke meningsbærende identifikatorer som er i bruk:

- [http://www.bibpode.no/work/Hamsun\\_Knut\\_sult](http://www.bibpode.no/work/Hamsun_Knut_sult)
- <http://viaf.org/viaf/99258155/>
- <http://data.bibsys.no/data/notrbib/authorityentry/x90061718>

### *Et ontologisk problem*

Identifikatorer og identifikatormodeller, og standarder, kan sies å være sosiale konstruksjoner. De er nært knyttet til modeller, som gir ett syn på hva som er viktig og hvordan ting henger sammen.

Definisjon av identifikatorer er også en ontologisk problem, du må vite hvilket abstraksjonsnivå du opererer på. (Weinberger, 2005)

*It's a complex field. That's because providing unique IDs is an ontological problem. You need to know the level of abstraction at which you're dealing*

*and, even then, "When are two copies of Hamlet the same?" is way too difficult a question to throw at a computer without a whole lot of human sorting-through. For example, even if a system had access to the full text (which these systems do not), two heavily annotated student editions might look quite different.*

*So far, at least in the realm of books, the successful sorting-throughs have been motivated by crass commercial needs. That's why they work. But the commercial sense of ontology — two books are the same if the accounting entries are the same — isn't the only one that matters to readers. This is not an issue we're ever going to get perfectly right because there isn't one right answer. The rare book collectors are always going to have a different sense of what needs a unique ID than are the public school teachers.*

*But that won't stop us from slapping numbers on things using schemes that slice up the world in ways that work at least pretty well for us.  
(Weinberger, 2005)*

## FRBR

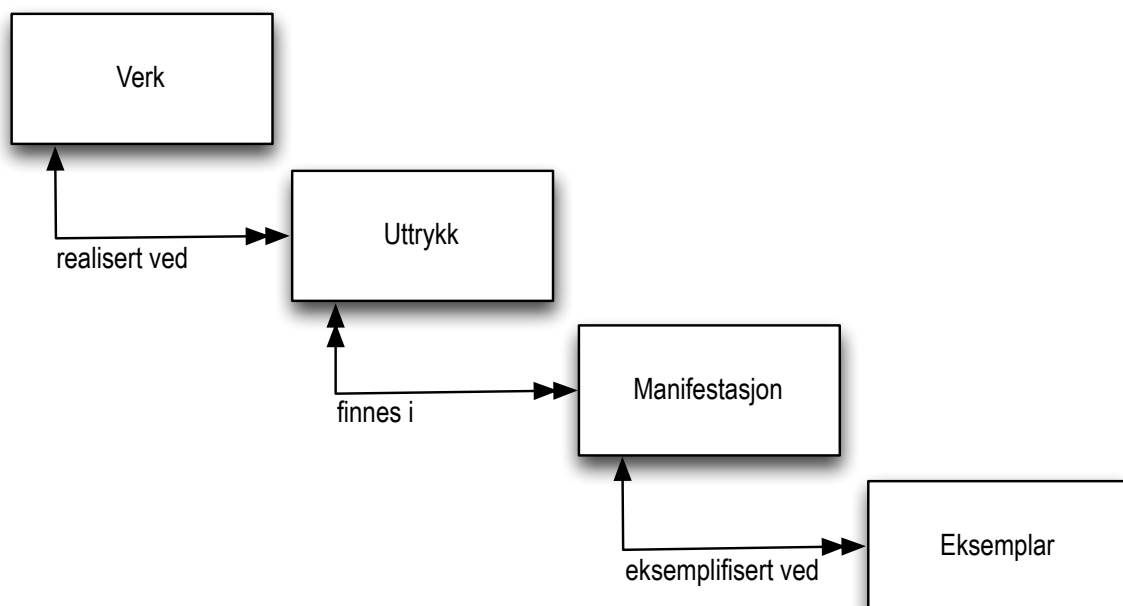
Functional Requirements for Bibliographic Records (FRBR) er en konseptuell bibliografisk datamodell.

FRBR-modellen beskriver tre hovedgrupper av entitetstyper for bibliografiske data:

1. Verk, uttrykk, manifestasjon og eksemplar
2. Person, korporasjon
3. Begrep, gjenstand, hendelse og sted

Den opprinnelige FRBR-rapporten tok kun for seg entitetsgruppe I, og det er denne som er hovedfokus i denne oppgaven. I tillegg er rapporten Functional Requirements for Authority Data (FRAD) (Patton, (FRANAR), International, & International, 2009) relevant i denne sammenhengen siden den bygger videre på og tildels avviker fra FRBR. FRAD beskriver en konseptuell modell for autoritetsdata.

Dette kapittelet er en gjennomgang av entitetene som er relevante i denne oppgaven, dvs. gruppe én og person-entiteten.

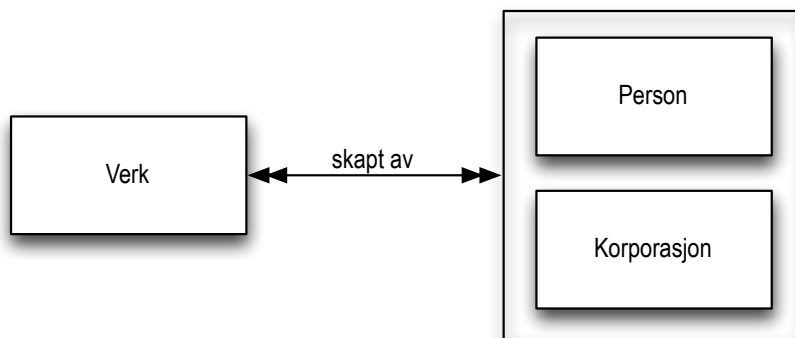


Figur 8: FRBR Entitetsgruppe 1 (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1997/2009, s. 13), (Holm, IFLA Study Group on the Functional Requirements for Bibliographic Records, & Nasjonalbiblioteket, 2001, s. 15)



De to nederste lagene i gruppe I av modellen, manifestasjon og eksemplar, er godt representert i dagens bibliotekskataloger. Det er de to øverste lagene, verk og uttrykk som er dårlig støttet.

Resultatet oppdager man raskt hvis man prøver å søke på noen av disse entitetene i en bibliotekskatalog, slik eksempelet fra Deichman tidligere viser.



**Figur 9: Relasjoner mellom verk og person/korporasjon**

### *Verket i FRBR-modellen*

Verk har en svært enkel definisjon i FRBR:

*Verk er et selvstendig intellektuelt eller kunstnerisk arbeid*

I FRAD spesifiseres det at verket er det intellektuelle eller kunstneriske innholdet.

Verk er en abstrakt bibliografisk entitet. Vi kan snakke om og skrive om et verk, men ikke ta på det fysiske. Hvis vi snakker om en fysisk versjon av verket, er det i henhold til modellen eksemplaret vi snakker om.

Rapporten er åpen på at dette er en abstrakt entitet som det kan være vanskelig å definere klare grenser for, og at grensene er kulturelt betinget. (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1997/2009)

I FRAD (Patton et al., 2009, s. 27) spesifiseres det også nærmere hvorfor grensene for verket kan være avhengig av kultur og kontekst:

*Note: The concept of what constitutes a work and where the line of demarcation lies between one work and another may differ from one culture to another, from one discipline to another or from one cataloguing agency to another. Consequently, bibliographic conventions established by various cultures, national groups, or cataloguing agencies may differ with respect to the criteria they use for determining the boundaries between one work and another. The specifics of cataloguing rules may also have a bearing on what is recognized as a work.*

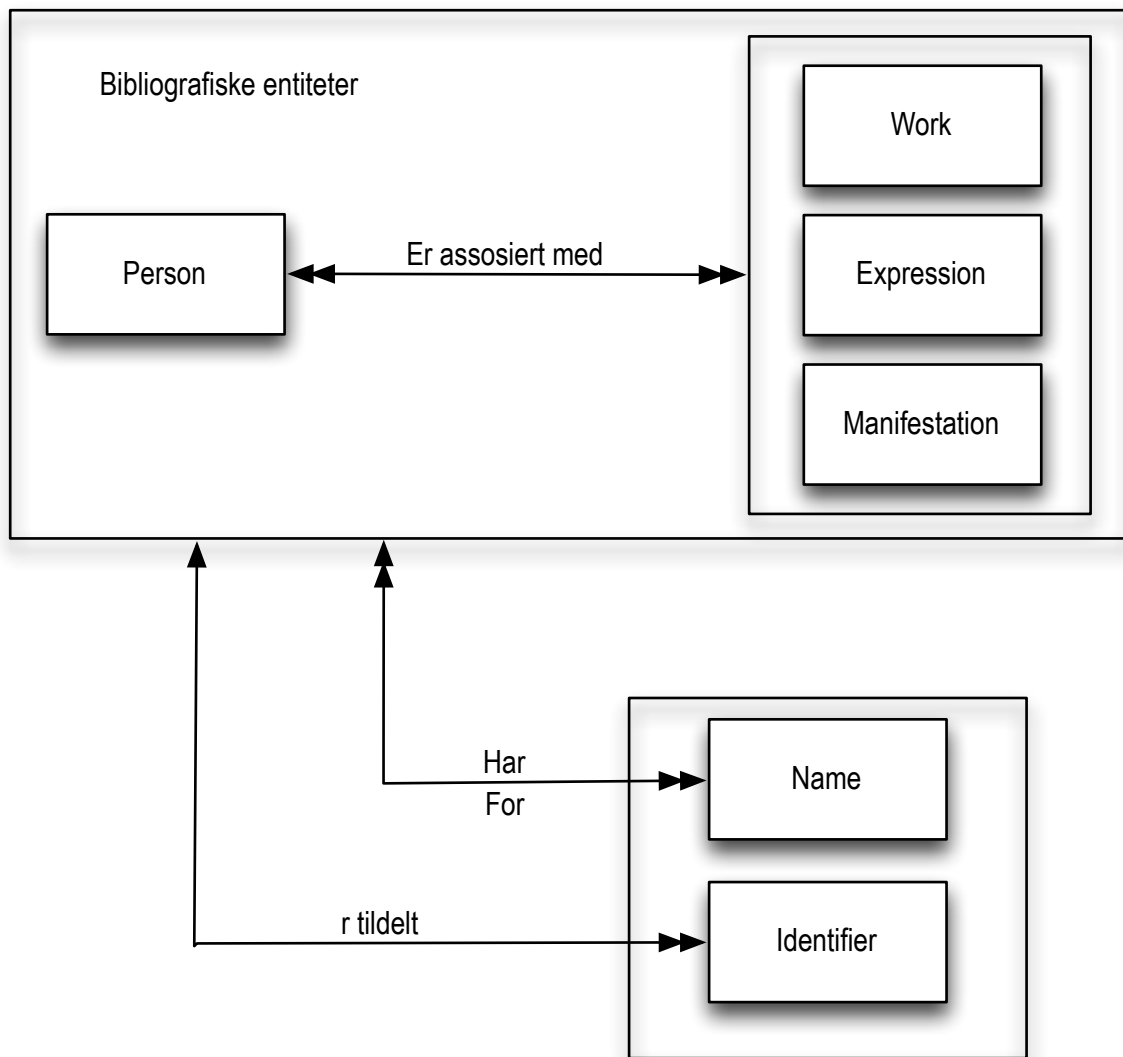
Det er definert tolv attributter for verk, men kun følgende verks-attributter er relevante for monografier (det finnes noen flere bl.a. for musikk og kart).

### **Tittel**

Her skal *standardtittel* brukes, dvs. én av utgavenes titler som er valgt ut til å representere verket. Ifølge norske katalogiseringsregler er dette originaltittelen.

Verket kan ha flere varianttitler. Det er også åpnet for at verket kan ha to parallelle standardtitler.

I FRAD er ikke tittel et attributt knyttet til verk. I stedet er entiteten *navn* definert. Navn kan knyttes til alle bibliografiske entiteter, og brukes bl.a. til navn for personer, titler for verk og tittel for manifestasjoner.



Figur 10: Oversikt over relasjoner i FRAD

### Litterær form

Dette er litterær sjanger, dvs. f.eks. Roman, Novelle, Dikt, Skuespill, Tegneserie m.fl.

### Årstall

Rapporten spesifiserer at førstevalget er datoen (vanligvis året) da verket ble skapt. Hvis det ikke finnes en sikker dato for skapelsen skal dato (vanligvis år) for førsteutgivelse brukes.

## Skillende karakteristika

Her spesifiseres det at attributtet er for *enhver karakteristikk* som kan hjelpe til å skille verket fra andre verk. Dette tyder på at det er beregnet på menneskelig tolkning. Som et eksempel er nevnt opphavssted for verket.

## Beregnet målgruppe

Gruppen som verket er beregnet for, basert på alder, utdanningsnivå eller andre kategorier.

## Kontekst

Konteksten verket ble skapt i. "Det 17. århundrets gjeninnsettelse av monarkiet i England" og "Den estetiske bevegelse på slutten av det 19 århundre" er nevnt som eksempler i FRBR. Dette attributtet er med andre ord beregnet på menneskelig tolkning.

## Attributt- og relasjonsverdi ved gjenfinning og identifisering av verk

FRBR-rapporten gjør en vurdering av verdien av de ulike attributtene og relasjonene i forhold til de ulike brukeropp gavene som er beskrevet. Her er oversikten forenklet til kun gjenfinning og identifisering av verk, siden det er det som er hovedfokus i denne masteroppgaven. Merk at utgangspunktet for disse vurderingene kun er sett fra brukerens side ved søk i katalogen på ett attributt. I vår sammenheng er i mer interessert i sannsynligheten for at en kombinasjon av attributter entydig kan bestemme et verk.

Attributter og relasjoner	Identifisering	Gjenfinning
Tittel	Høy	Høy
Litterær form	Middels	Middels
Årstall	Lav	Lav
Skillende karakteristika	Lav	Lav
Beregnet målgruppe	Lav	Lav

Kontekst		
Relasjon til ansvarlig person (forfatter)	Høy	Høy

### *Manifestasjonen i FRBR-modellen*

Denne entiteten er godt representert i form av katalogposter i dagens kataloger. FRBR definerer hele 38 attributter som kan brukes.

I denne sammenhengen tas kun med tabellen med vurdering av verdien av de ulike attributtene og relasjonene i forhold til gjenfinning og identifisering av verk og manifestasjon.

Her tas kun med de viktigste attributter og relasjoner der verdien for å finne verket er vurdert som høy eller middels. Samme reservasjon til verdien i vår kontekst gjelder som for verks-tabellen foran. Dette er vurdert ut ifra en bruker som søker i katalogen på ett attributt.

<b>Attributter og relasjoner</b>	<b>Identifisere verk</b>	<b>Identifisere manifestasjon</b>
Tittel	Høy	Høy
Ansvarsangivelse	Middels	Høy
Utgave-/heftebetegnelse		Middels
Forlegger/distributør		Høy
Utgivelsesår/distribusjonsår	Middels	Høy
Materialtype		Høy
Manifestasjonsidentifikator		Høy

Ansvarsangivelse betyr her ”én eller flere personer eller grupper som er ansvarlige for å skape eller frembringe det intellektuelle eller artistiske innholdet i manifestasjonen”.

### *Etablerte identifikator typer*

Når det gjelder manifestasjon er det lett å tenke at ISBN dekker behovet, det er imidlertid en stor andel av postene som ikke har ISBN. (Gatenby, 2010) viser til en analyse av Worldcat der kun 26% av postene har ISBN.

LCCN (Library of Congress Control Number) har vært i bruk side 1898, og kan være et alternativ for bøker publisert i USA. Nummeret hadde sin opprinnelse på katalogkortene, som ble trykket opp og distribuert. Hvert kort hadde et LCCN. Fra 2008 har det vært mulig å bruke http-uri for LCCN ved hjelp av LCCN Permalink service. 70575609 er f.eks. ID for *Mysterier* (1970) av Knut Hamsun, og <http://lccn.loc.gov/70575609> kan brukes som http-uri ID.

OCN (OCLC record identifiers) er et annet alternativ hvis manifestasjonen er registrert i Worldcat. 751954942 er f.eks. ID for *Sult* (1890) av Knut Hamsun.

(Gatenby, 2010) stiller spørsmålet ”trenger vi ID for uttrykk?”. For denne entiteten er identifikatorer nesten fraværende. ISTC (International Standard Text Code) vurderes til å ha plassert seg et sted mellom uttrykk og verk, og kan kanskje brukes på uttrykksnivå. (, udatert) ISTC har 16 alfanumeriske tegn, f.eks.

ISTC 0A9-2002-12B4A105-7 Dette er en sammensatt nøkkel der f.eks. de tre første tegnene representerer en registrator, mens de fire neste er år.

For verk er det heller ikke mange valgmuligheter. I tillegg til ISTC nevnt over publiserer OCLC OWI (OCLC Work identifiers).

VIAF (Virtual International Authority File) er implementert og driftet av OCLC for mer enn 20 nasjonalbiblioteker og endel andre bibliotekorganisasjoner. Denne tjenesten kombinerer de ulike autoritetsregisterne for navn til én tjeneste. Et eksempel på en ID er <http://viaf.org/viaf/71378383/> som er for forfatteren Henrik Ibsen.

## FRBRisering og MARC

Dagens bibliografiske systemer er i stor grad basert på MARC-formatet, så FRBRisering vil i de fleste tilfeller bety å hente ut FRBR-informasjon fra MARC-poster.

Det har vært gjort flere forsøk på å trekke ut informasjon fra MARC-baserte kataloger for å kunne generere en FRBR-modell. De første erfaringene var imidlertid ikke gode. I artikkelen *Datamining MARC to find: FRBR?* (Hegna & Murtomaa, 2003, s. 1) er en av hovedkonklusjonene:

*Even though the information in the MARC records holds attributes relevant for identifying the work, expression and manifestation entities, the accuracy and formal syntax are too simple to be properly handled by programs.*

Hegna og Murtomaa forsøkte her å trekke FRBR-relasjoner ut av MARC-postene til de finske og norske nasjonalbibliografiene. De tok utgangspunkt i et forfattersøk.

*When you search for a distinct person, the hit list should either consist of his or her works in some order, alphabetical perhaps or chronological or by a list of the different functions he plays related to the entities (author, illustrator, translator).*

Trond Aalberg hadde noe bedre erfaringer fra forsøk på å FRBRisere BIBSYS' bibliografiske database (Aalberg, Haugen, & Husby, 2006, s. 455), men erfarte også her at hovedutfordringen var inkonsistente data, dvs. dårlig datakvalitet i postene:

*The results from our conversion demonstrate that a perfect set of FRBR entities and relationships can be produced if the initial records contain sufficient and consistent information. For other sets of records the conversion tool creates duplicate entities and erroneous relationships, ...*

Både datakvalitet, katalogiseringsregler og Marc-formatet i seg selv representerer en stor utfordring i forhold til FRBRisering av bibliotekatalogene.

Pode-prosjektet, nevnt tidligere i oppgaven, valgte i første omgang seg ut fire forfattere: J.R.R Tolkien, William Shakespeare, Per Petterson og Knut Hamsun, men resultatene fra første runde var ikke bra.

Hamsun skrev 40 bøker, totalt 72 verk (inkludert enkeltnoveller). ("Sammenligning Frbr Versjon 1 Og 2," 2010) I utgangspunktet fikk man 585 treff ved avansert søk på forfatterfeltet etter "Hamsun, Knut". 63 av disse var ulike varianter av romanen Sult.

Målet var, som presentert i kapittelet om PODE-prosjektet, å presentere en oversiktlig og strukturert liste over verk i kronologisk rekkefølge.

Den første runden med konvertering var ikke vellykket. Hovedsakelig pga. manglende informasjon i Marc-postene og inkonsistent katalogiseringspraksis. I denne prosessen ble det identifisert 149 verk av Hamsun. (Westrum, Rekkavik, & Tallerås, 2012, s. 3)

Prosjektet besluttet å rette opp postene før konvertering, og prioriterte derfor Hamsun og Petterson. For monografier bestod rettingen først og fremst i å legge til originaltitler i poster som manglet dette. Det var også mange tilfeller av Det største problemet var manglende originaltittel i oversatte verker, så for Petterson var det lite å rette opp. Hamsun har både en større produksjon, flere oversettelser og utgavene spenner over flere år, så her var det mer arbeid. Flere av bøkene har endret tittel opp gjennom årene. ("Prinsipper for Rettinger Av Hamsun-poster," 2010) I andre runde, etter retting av postene, reduserte man antall identifiserte Hamsun-verk til 84. (Westrum et al., 2012)

## **Autoritetsregistre**

Autoritetsregistre har tradisjonelt vært brukt i katalogiseringsarbeid for å standardisere navneformer, emneord og titler, fordi dette vil gjøre gjenfinning av ressurser lettere. Standardiserte navneformer og standardtitler gjorde det mulig å koble sammen informasjon på tvers av ulike poster i katalogen.

Et eksempel fra kortkatalogen kan illustrere dette. Autoritetsregisteret var separat fra publikumskatalogen, men informasjonen kunne inkluderes i publikumskatalogen som f.eks. se-henvisninger.



Standardtittel ble utformet når et verk var kjent under flere titler<sup>4</sup>, f.eks. Hamlet. Kort-eksempler er hentet fra ("Authority Control in the Card Environment," udatert). Uf står for "used for":

```
Hamlet.  
  uf Tragedy of Hamlet  
  uf Prince of Denmark  
  
  ( )
```

**Figur 11: Eksempel på autoritetsregisterkort med standardtittel**

```
Prince of Denmark  
  see  
  
  Hamlet.  
  
  ( )
```

**Figur 12: Eksempel på kryssreferanse i publikumskatalogen**

```
822.3 Shakespeare, William, 1564-1616.  
  [Hamlet]  
  The Prince of Denmark .....  
  ( )
```

**Figur 13: Eksempel på katalogkort for boken**

---

<sup>4</sup> Vi har ikke hatt denne praksisen med standardtittel på skjønnlitteratur og faglitteratur i Norge.

*Authority control of names, work identifiers, and subjects is a special kind of standardization that is much more likely to be necessary in large cataloging and union cataloging projects than in descriptive bibliography projects. When authority control is applied effectively, users are enabled to find the works they seek under any variant of author name (Mark Twain vs. Samuel Clemens), variant of title (Tom Sawyer vs. Adventures of Tom Sawyer), or subject synonym they might use in their search (e.g. hypnosis vs. hypnotism). In addition, homonyms are differentiated (e.g. cold the disease vs. cold the temperature), people with the same name are differentiated (e.g. George Bush), and works with the same title are differentiated (e.g., 5th symphony). And finally, a structure is created that allows a user to broaden and narrow a search as desired. (Yee, 2007, s. 314)*

Dårlig koordinering av navneformer fører fortsatt til problemer med flere poster for én manifestasjon, såkalte dubletter, når katalogposter aggregeres. Manglende standardtittel fører også til at det blir vanskeligere eller umulig å automatisk koble sammen manifestasjoner til verk ved FRBRisering.

*I en bibliotekskatalog eller bibliografisk database kan det utøves autoritetskontroll på mange forskjellige informasjonslementer. De mest aktuelle er i tillegg til personnavn og korporasjonsnavn: stedsnavn, tidsperioder, emneord, klassifikasjon, serietitler og standardtitler. (Den norske katalogkomité, 2004, s. 6)*

*Norske bibliotek har tradisjonelt brukt betydelige ressurser på katalogisering av sine samlinger. Ved siden av klassifikasjon og emneindeksering representerer valg av navneformer en av de mest ressurs- og kompetansekrevende arbeidsoppgavene. Ved at det bygges opp et sentralt autoritetsregister over etablerte navneformer, vil en oppnå store fordeler ved at den enkelte katalogiseringsenhet kan unngå å måtte gjenta et nitidig kontrollarbeid som allerede vil være utført ved etableringen av korrekt navneform i det sentrale registeret. (Den norske katalogkomité, 2004, s. 5)*

Nasjonalt autoritetsregister for bøker finnes ikke i Norge, og det finnes knapt eksempler internasjonalt. (Den norske katalogkomité, 2004), (Knutsen, 2009), (Smiraglia, 2004)

(Den norske katalogkomité, 2004) nevner imidlertid "at det finnes en liste over standardtitler for norske middelalderverk". Denne er utgitt i bokform (Rindal & katalogkomité, 1994). Det er også nevnt at BIBBI, Biblioteksentralens bibliografiske database, har autoritetsregister for standardtitler.

(Knutsen, 2009) nevner at det er behov for autoritetsregistre over standartitler for ulike komponisters verker.

Det meste som er skrevet om autoritetsregistre bærer preg av å være fra en annen tid. Det dreier seg om standardisering av tekstlige titler brukt i katalogen for å sikre konsistent gjenfinning ved søk.

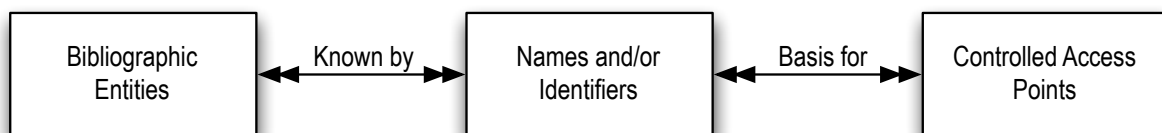
BIBSYS' nye autoritetsregister for navn basert på Lenkede data kan være en modell for hvordan dette bør gjøres. Følgende er et eksempel på en RDF-kodet autoritetsfil for William Shakespeare (Bertheussen, Greenall, & Husby, 2010):

```
<x90052737> a foaf:Person ;
  radatana:catalogueName "Shakespeare, William" ;
  dcterms:modified "19910422" ;
  foaf:name "William Shakespeare" ;
  rdfs:label "William Shakespeare" ;
  skos:altLabel "Uil'jam Sekspir";
  skos:altLabel "William Shakspeare";
  skos:altLabel "William Shakspear";
  skos:altLabel "William Shakspere";
  skos:altLabel "William Shakespear";
  dcterms:identifier "x90052737";
  owl:sameAs <http://dbpedia.org/page/William_Shakespeare> ;
  owl:sameAs <http://libris.kb.se/resource/auth/198702> ;
  owl:sameAs <link.to.NL_Hungary> ;
  owl:sameAs <http://d-nb.info/gnd/118613723/about> ;
  owl:sameAs <http://viaf.org/viaf/96994048/rdf.xml> ;
  foaf:page
  <http://ask.bibsys.no/ask/action/result?cmd=&kilde=biblio&fid=forfatter&term=Shakespeare%2C+William&op=and&fid=bd&term=&bibliografi=ingen&arstall=&sortering=sortdate-&treffPrSide=10> ;
  foaf:page <bibsys_authority_william_shakespeare> ;
  foaf:page <bibsys_william_shakespeare> ;
  foaf:made <links.to.the.bibsys.objektid.RDF.representations> .
```

Her er identifiserende informasjon navn og navnevarianter kodet som attributter for personentiteten.

owl:sameas er brukt til å si at denne identifikatoren identifiserer samme person som den andre identifikatoren. Dette gjør at personnodene vil bli slått sammen, eller flettet sammen hvis en rdf-graf inneholder opplysninger knyttet til en av de andre ID-ene.

Rapporten *Functional Requirements for Authority Data (FRAD)* (Patton et al., 2009), *Funksjonskrav til autoritetsdata*, er et skritt videre i en annen retning, da det lages en konseptuell modell for autoritetsdata. Arbeidet ble startet i 1999 og rapporten kom 10 år senere. Rapporten bygger på og tildels utvider FRBR-modellen. Rapporten beskriver autoritetsdata på et helt annet abstraksjonsnivå. Det kan være vanskelig å trekke linjene fra beskrivelsene i rapporten til BIBSYS sin løsning basert på lenkede data. Følgende figur danner basis for den konseptuelle modellen.



**Figur 14: Fundamental Basis for the Conceptual Model (Patton et al., 2009, s. 19)**

*Controlled access point* er en av hovedentitetene i modellen og brukes gjennomgående i FRAD. Begrepet er abstrakt og brukes på flere ulike måter.

I den norske versjonen av FRBR-rapporten er termen *Access point* oversatt med *søkeelement*. I *Erklæring om internasjonale katalogiseringsprinsipper* (IFLA, 2008) brukes termen *søkeinnganger*. I *Bibliotekstermer på svenska och engelska* (Dodd & Uppsala universitetsbibliotek, 2003) er oversettelsen *sökingång*. Ingen av disse definisjonene er helt dekkende.

#### *Access point*

*A unit of information in a bibliographic record under which a person may search for and identify items listed in the library catalog or bibliographic database. (ODLIS — Online Dictionary for Library and Information Science, udatert)*

Hvis vi trekker linjene tilbake til kortkatalogeksempelet og til et søkegrensesnitt i en katalog, dreier det seg tilsynelatende om attributtene som er indeksert og kan slås opp på i kortkatalogen og attributtene som tilbys som søkefelt i en OPAC (Online Public Access Catalog).

(Oliver, 2010b, s. 15) oversetter begrepet *Controlled Access Point* med "authorized forms, variant forms of names/references, etc."

FRAD-rapporten har følgende definisjon av *Controlled Access Point*:

*A name, term, code, etc., under which a bibliographic or authority record or reference will be found.*

*Includes access points designated as authorized (or preferred) forms of names (i.e., authorized access points) as well as those designated as variant forms of name (i.e., variant access points).*

*Includes access points based on personal, family, and corporate names.*

*Includes access points based on titles (i.e., names) for works, expressions, manifestations, and items.*

*Includes access points consisting of a combination of two access points, as in the case of a creator/title access point for a work which consists of an authorized access point for the name of the creator combined with an authorized access point for the name (i.e., the title) of the work.*

*Includes access points based on identifiers, such as standard numbers, classification indicia.*

Min tolkning av dette er at begrepet kan oversettes med "identifiserende autoritetsdata", noe som skulle dekke både standardiserte former av ulike attributter og unike identifikatorer.

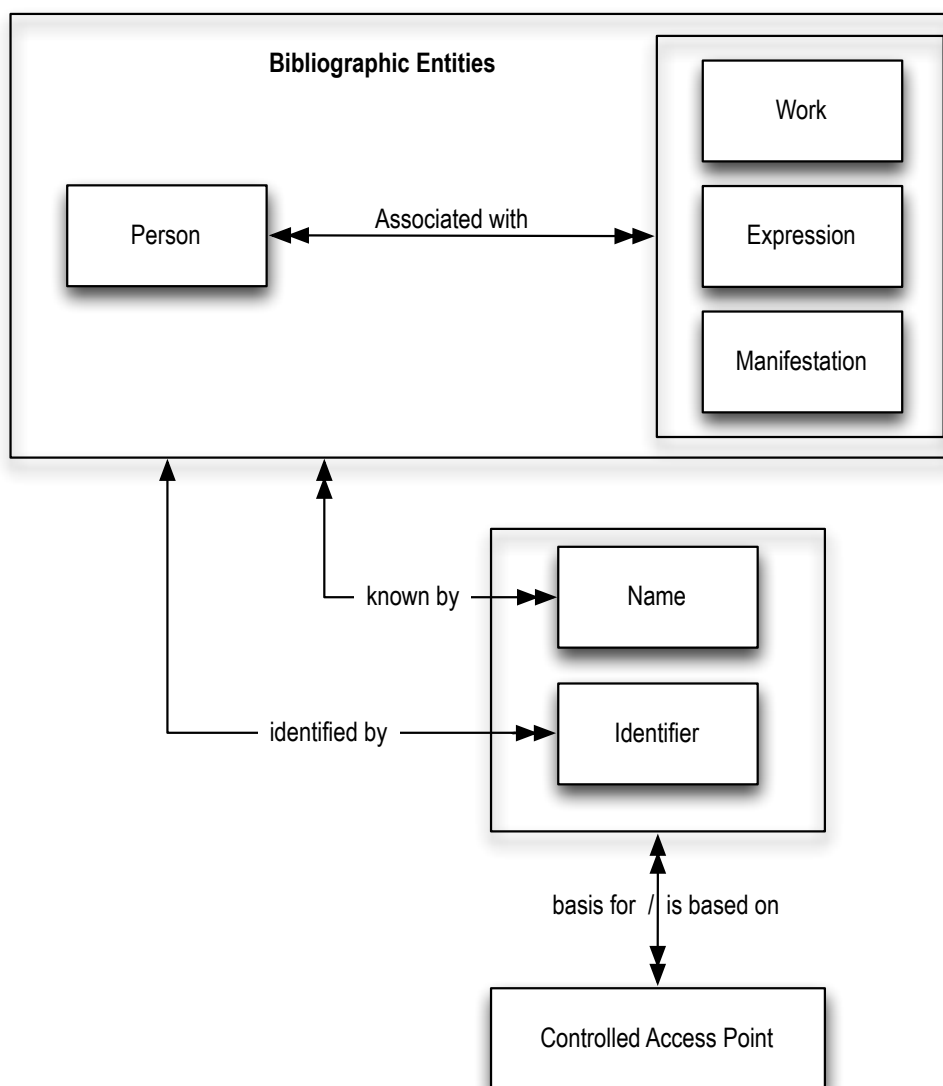
I første kapittel beskrives formålet med autoritetsdata som følger:

*Authority data represents the controlled access points and other information that institutions use to collocate works by a specific person, family, or corporate body, or the various editions of a title.*

*Controlled access points include authorized forms and variant forms of name assembled by cataloguers to identify an entity. For the purposes of this study, only name and title entities are addressed fully;*

*Authority control, which means both the identification of entities represented by controlled access points and the ongoing management of them, is integral to the functioning of a catalogue.*

*Authority control is beneficial to cataloguers able to identify and distinguish between the controlled access points within a catalogue. More importantly, authority control benefits end users, enabling them to search any controlled form of an author's name or of a title to retrieve bibliographic resources within catalogues.*



**Figur 15: Conceptual Model for Authority Data**

I denne figuren har jeg kun tatt med de mest interessante entitetene i denne konteksten og har endret relasjonsnavnene noe i forhold til opprinnelig figur i (Patton et al., 2009, s. 23).

Legg merke til at navn og identifikator er samlet i en gruppe, men at det går forskjellige relasjoner til hver av dem. I Figur 14: Fundamental Basis for the Conceptual Model ble relasjonen *known-by* brukt. I den detaljerte figuren har Patton to forskjellige relasjoner til hhv. navn og identifikator: *has-appellation* og *is-assigned*.

I Merriam-Webster er definisjonen for Appellation:

1. : an identifying name or title : designation
2. archaic : the act of calling by a name

Det er en mange-til-mange-relasjon mellom bibliografiske entiteter og navn. Dette skulle bety at en bibliografisk entitet kan ha flere navn, og at navnene ikke nødvendigvis er entydige, men kan deles av flere bibliografiske entiteter. Det er ikke en identifiserende relasjon. Dette er som diskutert tidligere en av utfordringene med navn og titler. Flere personer kan hete det samme og flere bøker kan ha samme tittel.

Derfor har jeg valgt å beholde relasjonstermen brukt i grunnmodellfiguren, *known-by*.

Det er en én-til-mange-relasjon mellom bibliografiske entiteter og identifikator. Dette skulle bety at en bibliografisk entitet kan ha flere identifikatorer, men at en identifikator kun kan brukes av én bibliografisk entitet. Dette er en identifiserende relasjon, en nøkkel, en unik identifikator. Jeg har derfor valgt å bruke relasjonsnavnet *identified-by* i figuren.

Begrepet *Controlled Access Point* står på flere måter litt i veien for å se universet som beskrives i denne konseptuelle modellen i forhold til konseptuelle modeller for metadata (som f.eks. Dublin Core) og Semantisk web.

Når konteksten er design av et autoritetsregister basert på semantisk teknologi, kan følgende være en forenklet tolkning av modellen. Det er for bibliografiske entiteter behov for kontrollerte former av navn, titler mm i et autoritetsregister. I tillegg er det behov for identifiserende data, nøkler, i form av unike identifikatorer for de ulike bibliografiske entitetene. Både kontrollerte navneformer, titler og identifikatorer er grunnlag for autoritetsregisteret.



## Oppsummering og rammer for et autoritetsregister

Utgangspunktet i denne oppgaven er Podedprosjektets ønske om et nasjonalt autoritetsregister og åpne, nasjonale unike ID-er for verk.

Hovedspørsmålet er hva som skal til for å tilfredsstillere Podedprosjektets ønske.

Det ble stilt flere spørsmål i problemstillingen som det nå er mulig å drøfte:

- Podedprosjektet utforsket FRBR, semantiske teknologier og Lenkede data. Er dette naturlige teknologivalg for et autoritetsregister, gitt dagens utvikling hos BIBSYS, Nasjonalbiblioteket og viktige internasjonale bibliotekaktører?
- Hvordan bør en unik ID for verk se ut?
- Hva ligger i begrepet autoritetsregister, og hvordan kan et autoritetsregister løse Podedprosjektets hovedutfordringer?
- Hva ligger i begrepet verk, og hvilken rolle spiller verket i bibliotek katalogen?

### *Bruk av FRBR og lenkede data*

Det ble innledningsvis stilt spørsmål om teknologien Podedprosjektet brukte virket som naturlige teknologivalg for et autoritetsregister, gitt dagens utvikling hos BIBSYS, Nasjonalbiblioteket og viktige internasjonale bibliotekaktører.

Podedprosjektet utforsket FRBR, semantiske teknologier og Lenkede data. - Er dette fornuftige teknologivalg i denne sammenhengen?

Ut ifra gjennomgangen i kapitlene om Det norske biblioteklandskapet og Internasjonale utviklingstrekk ser det ut til å være en klar trend at bibliotekorganisasjonene ønsker å dele bibliografiske data via WWW med RDF og Lenkede åpne data.

Dette ligger til grunn for beslutningene BIBSYS har tatt om autoritetsregistre. Dette er også Implisitt i arbeidet i Library of Congress og Worldcat de siste årene. Det er flere eksempler på autoritetsregistre basert på Lenkede data.

FRBR ser også ut til å ha fått mer vind i seilene det siste året. En av årsakene er at RDA er akseptert av store deler av det internasjonale biblioteklandskapet, og at denne standarden bygger på entitetene i FRBR.

Ut ifra teknologivalg andre aktører gjør kan det konkluderes med at PODE-prosjektets valg var gode.

### *Hvordan bør en unik identifikator for verk se ut?*

Ut ifra gjennomgangen i kapittelet om Lenkede data er det lett å konkludere med at en identifikatorer for verk er en forutsetning hvis autoritetsregisteret skal baseres på FRBR og realiseres som Lenkede data. Identifikatoren bør også være en http-URI.

Ut ifra gjennomgangen i kapittelet om Behovet for identifikatorer, virker det fornuftig å legge opp til en ikke-meningsbærende identifikator. At VIAF er et vellykket eksempel på en identifikator av den samme typen som brukes i et internasjonalt samarbeid om autoritetsregister for personer og korporasjoner støtter denne vurderingen.

### *Diskusjon av kriterier for et nasjonalt autoritetsregister*

I problemstillingen ble det stilt spørsmål om hva ligger i begrepet autoritetsregister, og hvordan et autoritetsregister kan løse PODE-prosjektets hovedutfordringer.

Kapittelet om Autoritetsregistre ga en oversikt over tradisjonell bruk av autoritetsregistre. Den konseptuelle modellen for autoritetsregistre i FRAD-rapporten ble også presentert.

Fra gjennomgangen av PODE-prosjektet vet vi at mangel på originaltittel gjorde at FRBRiseringsprosessen i første runde ikke ble vellykket. Originaltittel er som nevnt i FRBR-kapittelet standardtittel i henhold til norske tradisjoner, og brukes som navn på verket i FRBR-modellen. Innlegging av originaltittel i postene gjorde det mulig å koble sammen postene som tilhørte samme verk.

Sett fra PODE-prosjektets perspektiv ønsker vi først og fremst å tilfredsstillere behovet for identifikasjon av verk ut ifra informasjonen i en katalogpost (manifestasjon). Dette kan også skje ved at data fra en katalogpost er søke-input til autoritetsregisteret, som returnerer id for tilhørende verk.

Erfaringen i Pode-prosjektet var imidlertid at informasjonen i katalogposten i utgangspunktet ikke var god nok til å identifisere hvilket verk manifestasjonen tilhørte. Man måtte redigere postene og legge inn originaltittel for å få til dette.

En alternativ, og komplementerende tilnærming, er at autoritetsregisteret har registrert data for manifestasjonene, og at det for hver manifestasjon er registrert en relasjon til riktig verk. Verksentiteten må da ha en tilordnet verksidentifikator som kan returneres.

Denne modellen gjør at Pode-prosjektet ikke trenger å rydde postene for å få koblet manifestasjonene til riktig verk.

### **SPARQL-søk etter verks-id**

Vi ønsker altså å gi andre systemer muligheten til å søke og finne ID for verk ut ifra opplysningene de har i katalogposten. SPARQL kan tilby dette.

Bruken av SPARQL endpoint som autoritetsregister kan sammenlignes med søk mot en relasjonsdatabase som bruker SQL der input er en nøkkel for manifestasjonen og output er en verks-id.

Pode-prosjektet brukte et SPARQL endpoint, så dette teknologivalget ville vært uproblematisk sett i deres perspektiv. Autoritetsregisteret *Rådata nå!* fra BIBSYS bruker også et SPARQL endpoint, så dette støtter vurderingen.

### **ISBN som input**

Det enkleste scenariet er at bibliotekposten har et ISBN-nr. ISBN er en unik nøkkel for manifestasjonen<sup>5</sup> Som nevnt i kapitlet om Etablerte identifikatortyper har imidlertid kun ca. 26% av postene i Worldcat ISBN.

Det første designkriteriet er at ontologien må ha manifestasjon som entitet, som må ha ISBN som informasjonsattributt. Det må også være en relasjon fra manifestasjon til verk, og verket må ha verks-id. På denne måten kan et SPARQL-søk med ISBN som input returnere en verks-id.

---

<sup>5</sup> Unik i prinsippet, men det finnes tilfeller av gjenbruk som kan skape problemer.

Siden vi planlegger å bruke Lenkede data vil alle entiteter i utgangspunktet ha en http-URI id.

Et annet scenario er at det ikke finnes ISBN for en manifestasjon. Da må det være mulig å bruke en kombinasjon av andre data som input for å identifisere riktig manifestasjon.

I en annen sammenheng kunne det vært aktuelt å bruke en alternativ unik ID, som f.eks. LCCN eller OCN, men det er lite hjelp i dette i vår kontekst, siden Deichmans katalog i utgangspunktet ikke inneholder disse identifikatorene.

Siden konteksten er et nasjonalt autoritetsregister kunne en intern BIBSYS-id vært en alternativ nøkkel for manifestasjon, men for Pode-prosjektet ville ikke dette ha hjulpet.

Alternativt må man gjøre søk med en kombinasjon av andre data.

## **Identifiserende data for manifestasjon**

Tabellen i kapittelet om Manifestasjonen i FRBR-modellen viser verdien av de ulike attributtene og relasjonene for en manifestasjon i forhold til gjenfinning og identifisering av manifestasjonen.

Her er det i utgangspunktet flere attributter og relasjoner å velge mellom.

I tillegg til manifestasjonsidentifikator, som kan være ISBN, velger jeg å ta med tittel, forfatter og utgivelsesår.

Det kan finnes eksempler på at tittel, forfatter og år ikke er nok informasjon til å bestemme en manifestasjon entydig, men her vil vi akseptere at disse tilfellene vil føre til treff på flere manifestasjoner. Hvis disse manifestasjonene er relatert til forskjellige verk vil flere verks-id returneres fra søket og resultatet må vurderes av et menneske.

Hvis denne løsningen skal implementeres i større skala er det flere alternativer man kan bruke for å etablere identitet, se (Linking data without common identifiers, 2011) og (Takhirov, Duchateau, & Aalberg, 2011) for en beskrivelse av mer robuste strategier for hvordan man kan lenke sammen FRBR-entiteter og Lenkede data uten delte identifikatorer. Dette fagfeltet har flere navn, bl.a. record linkage, entity resolution, identity resolution og deduplication. Det er dessverre utenfor rekkevidde for denne masteroppgaven å gå nærmere inn på disse løsningene, utover å nevne dette som mulig videre arbeid.

Ontologien må altså gjøre det mulig å bruke en kombinasjon av tittel, forfatter-ID og år for å finne riktig manifestasjon i ontologien.

Et annet tilfelle som bør støttes er at input av både forfatter og tittel er tekststrenger. I tilfeller der forfatter ikke er kjent må ontologien støtte en kombinasjon av tittel og år.

## **ID for forfatter**

Et annet spørsmål er hva slags forfatter-id som bør velges. Vi kan velge å støtte ID brukt i BIBSYS sitt autoritetsregister og/eller VIAF-ID. Jeg velger her å bruke VIAF fordi dette er et datasett som også inkluderer BIBSYS sine autoritetsdata for personer.

## **Trenger vi ID for uttrykk?**

Vi ønsker å gi andre systemer muligheten til å søke og finne ID for verk ut ifra opplysningene de har i katalogposten. Input er identifiserende data for manifestasjon.

FRBR-modellen er basert på at relasjonen fra manifestasjon til verk går via uttrykk, det er ikke definert noen relasjon direkte fra verk til manifestasjon.

FaBiO (the FRBR-aligned Bibliographic Ontology) har sett dette behovet og har utvidet FRBR-modellen med en relasjon direkte mellom verk og manifestasjon: (Shotton & Peroni, 2012)

```
fabio:hasManifestation og fabio:isManifestationOf
```

Relasjonen kan f.eks. brukes slik:

```
@prefix fabio: http://purl.org/spar/fabio/
```

```
<Verks-ID> fabio:hasManifestation <manifestasjons-ID>
```

```
<Manifestasjons-ID> fabio:isManifestationOf <verks-ID>
```

## *Attributter for verk*

En av kjerneforskjellene på autoritetsregisteret beskrevet i denne oppgaven og et tradisjonelt autoritetsregister går ved bruk av standardtittel.

*The idea that an entity with more than one name should be referred to in a bibliographic database by one form only is familiar to most in the library community, although not, perhaps to most library users. (Maxwell, 2008, s. 67)*

Hovedgrunnen til å bruke én navneform i autoritetsregisteret er at man ønsker at alle bibliografiske poster som refererer til samme entitet bruker samme navneform, slik at alle gjenfinnes når en bruker søker. Som Maxwell påpeker er det ikke gitt at dette er nødvendig hvis det brukes andre modeller og annen teknologi. (Maxwell, 2008, s. 67)

Ved å bruke en identifikator til å relatere entitetene til hverandre vil det ikke være behov for å bruke standardtittelen til dette.

I utgangspunktet kan det virke som det ikke er behov for å gi annen output fra systemet enn en verks-id.

Men det vil være en fordel at standardtittel og år inkluderes. Dette vil gjøre at systemet kan fungere som et tradisjonelt autoritetsregister. Det er to grunner til å ta med standardtittel og årstall for verk i ontologien:

1. Verks-entiteten i FRBR-modellen har tittel og årstall
2. Brukerperspektivet og behovet for å presentere verket med én tittel.

Katalogens andre hovedformål er å presentere brukeren for en liste over en forfatters verk, og vi har diskutert at et forfattersøk bør resultere i en kronologisk liste over forfatterens verk.

Hvis et verk skal ha en standardtittel er det i forhold til punkt 2 fordi vi ønsker en tittel som kan representere alle variantene av verket ved presentasjon av verket i katalogen for brukeren. Sett fra brukerperspektivet er det også behov for å støtte muligheten for opplisting av verk kronologisk.

Behovet for standardtittel har altså endret seg fra bruk som en identifiserende tekststreng til et behov for å avklare hvilken tittel som kan representere alle verkene. Identifikatoren har overtatt standardtittelens gamle rolle.

Årstallet bør være med som attributt både fordi det er med i FRBR-modellen og fordi man ønsker å presentere verkslisten kronologisk.

Et moderne autoritetsregister for verk bør derfor tilby en unik identifikator for verket, standardtittel for verket og årstall.

Da fyller løsningen også oppgaven som tradisjonelt autoritetsregister ved å levere standardtittel.

### *Verkets rolle i katalogen og autoritetsregisteret*

Et av spørsmålene i problemstillingen var hva som ligger i begrepet verk, og hvilken rolle verket spiller i bibliotek katalogen.

Den historiske gjennomgangen og analysen av verksbegrepet ga noen svar, og viste at Poded-prosjektets strukturering av trefflisten ved forfattersøk, er veldefinert og bygger på lange tradisjoner. Gjennomgangen av FRBR- og FRAD-modellene ga også en innsikt.

Her følger drøfting av noen uavklarte sider ved verket.

### **Ett navn, flere navn og visningsnavn på ulike språk**

Noen mener verk ikke burde ha en tittel som attributt, men at verkets tittel er en union av alle titler for manifestasjonene. (What is a (FRBR) Work?, 2009)

Dette er en filosofisk diskusjon, men vi kan som nevnt over bringe diskusjonen tilbake til verkets historie og til brukerperspektivet. Vi ønsker en tittel som kan representere alle varianter av verket.

Spørsmålet er om vi kanskje ønsker flere navn. FRBR-modellen åpner for mer enn én standardtittel og flere variantnavn. FRAD har objektifisert tittelen og kalt den *navn*, slik at det har blitt en egen entitet som er relatert til bl.a. verk. Fordelen med dette er at vi kan hekte metadata på tittelen, og f.eks. angi flere typer titler.

I FRAD har navn flere attributter knyttet til seg, bl.a. type navn og språk.

(What is a (FRBR) Work?, 2009) nevner utfordringer rundt visning av verk, spesielt i flerspråklige systemer:

*When you follow the concept of uniform titles, the Work title should be the title of the original. This means that we would be showing our users *Война и мир* as the title for the Work that most of them will know as *War and Peace*. We could show them the English language title, but what if your catalog users are global? What if some of them will only understand the title if you display it in French or Turkish or Chinese? If a Work has an identifier (which is only useful for machine processing, not for display to humans), then you can let users choose what language they prefer in Work displays. (Obviously having some default for the case where the user's preferred language isn't available.)*

Her går det an å hente inspirasjon fra Emnekartmodellen ("TMDM," 2008). Dette er også en semantisk metamodell i likhet med RDF. En av forskjellene er at navn og variantnavn er en del av grunnmodellen.

I et emnekart er det vanlig å definere variantnavn av type *display name*, visningsnavn. I Emnekartmodellen finnes også scope som kan oversettes med kontekst. Alle deler av modellen kan tilordnes scope, som representerer konteksten hvor utsagnet er gyldig.

Et verk kunne i en slik modell ha hatt mange visningsnavn i ulike kontekster, der kontekstene var språk. Det ville ikke vært nødvendig for autoritetsregisteret å ha registrert alle variantene. Norsk og engelsk ville vært naturlig. Andre kunne ha lagt til andre språk etter behov.

Det ser ut til at rammeverket for å få til dette ligger i FRAD. Dette kan være en modell for hvordan det er mulig å angripe utfordringen med å vise riktig navn hvis man har flere navn å velge mellom.

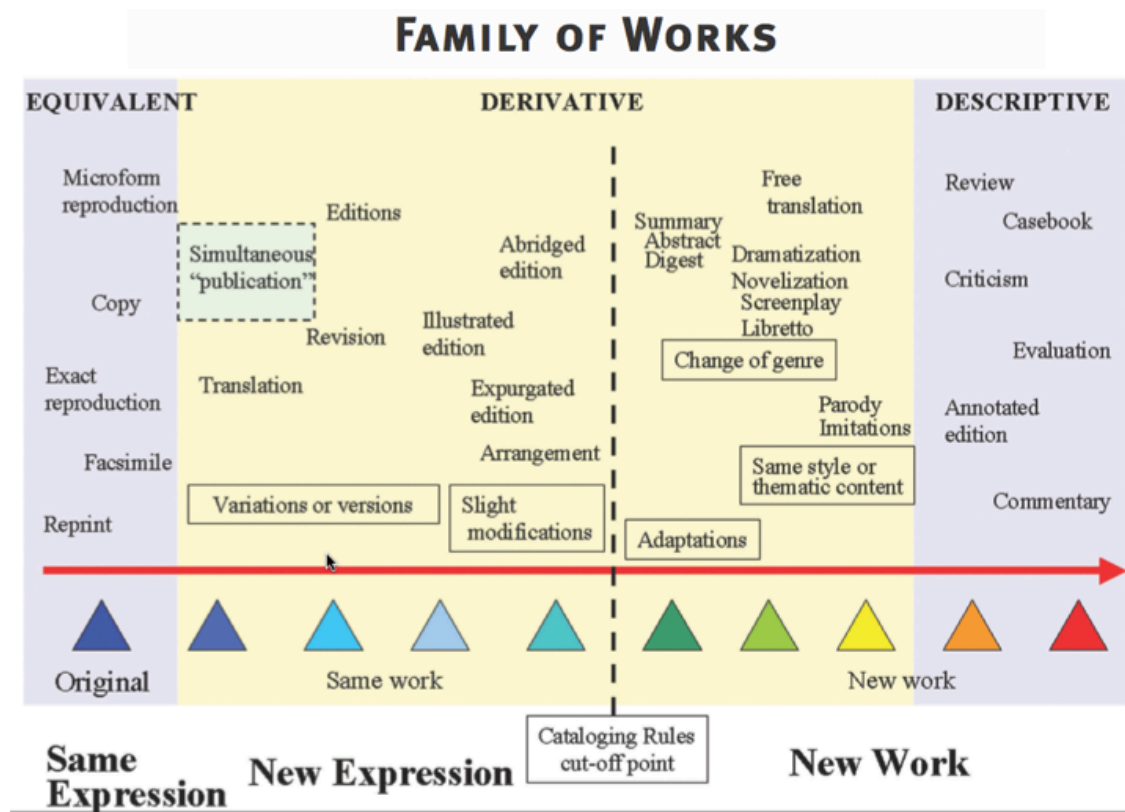
### *Brukerbehovet og verksavgrensing*

Det ligger en motsetning i definisjonen av verk i FRBR.

På den ene siden kan man si at det er et aristoteliske paradigme for kategorisering som ligger til grunn for FRBR. Entitetenes attributter er grunnlaget for medlemskap i en kategori,



og katalogiseringsreglene bruker dette til å forsøke å definere klare grenser mellom ett verk, uttrykk og et avledet verk.



Figur 16: Grafisk fremstilling av verksgransene (Tillett, 2001)

På den andre siden sier FRBR-rapporten klart fra om at grensene mellom ulike verk ikke er klare, og kan være kulturelt betinget. (IFLA Study Group on the Functional Requirements for Bibliographic Records, 1997/2009, s. 16)

Verk en sosial konstruksjon. Dette gjør at det er umulig å sette klare universelle grenser for når en oppfatter to manifestasjoner som samme verk eller to forskjellige verk. (Patton et al., 2009, s. 27), (Genz, 2002), (Wilson, 1989)

I enda større grad enn for bibliotekaren som skal klassifisere gjelder dette brukeren. Det er ikke klare universelle grenser for når en bruker oppfatter to manifestasjoner som samme verk eller to forskjellige verk.

En alternativ forklaringsmodell for klassifikasjon er prototypingsteori, der medlemmer i en kategori har ulik grad av medlemskap og grensene mellom ulike kategorier er mer diffuse.

Graden av medlemskap vil både være subjektiv (avhengig av bibliotekbrukeren) og avhengig av kontekst (hvilken situasjon bibliotekbrukeren er i). (Lakoff, 1994)

Katalogiseringsreglene setter en grense, og det oppstår et nytt verk ved bearbeiding av et verk, hvis grensene mellom litterære former eller kunstformer krysses. Et eksempel er filmatisering.

Her kan det se ut til at katalogens andre formål og det som hittil har blitt sett på som en ideell opplisting kommer til kort. Vi trenger noe mer. Fordi grensene mellom verk kan være uklare er det behov for også å presentere relaterte verk i forhold til konteksten man er i.

Det er sannsynlig at brukerbehovet variere mye med hensyn til hvor grensene for akseptable alternative varianter av verket går. En andel av brukerne vil synes at filmen representerer samme verk og synes den er et godt alternativ til boken i endel sammenhenger.

Et eksempel kan være et søk etter *Sven Nordquist*, som bl.a. har skrevet bøkene om *Gubben og katten*. Hvis vi antar at katalogen er basert på FRBR, og trefflisten er en kronologisk oversikt over verk, vil *Gubben og katten og nissemaskinen (1995)* vises i listen. Det som imidlertid ikke vil vises i trefflisten er animasjonsfilmen *Gubben og katten og nissemaskinen*. Dette er forskjellige verk, hhv. Boken og filmen, og Nordquist er ikke hovedansvarlig for filmen.



Figur 17: To forskjellige verk. Boken til venstre og filmen til høyre.

Den tyske eller russiske boken, eller en eventuell gresk lydbok, representerer imidlertid samme verk som den norske boken, men ulike uttrykk. Modellmessig er de dermed nærmere den norske boken enn filmen i eksempelet over. I en treffliste basert på katalogens andre formål og Browns retningslinjer for presentasjon (Brown, 1916) vil disse vises, mens filmatiseringen ikke vil være tatt med.



Figur 18: Den tyske og greske oversettelsen i bokform og den greske lydboken er samme verk som den norske boken.

Informasjon kodet som RDF og Lenkede data danner en graf av noder som er lenket sammen. En bivirkning av dette er at det også legger til rette for å lage brukergrensesnitt der brukeren kan navigere gjennom grafen. Dette trenger ikke å være basert på visualisering som en graf, men kan se ut som fasettert navigasjon via websider med lenker. Dette gjør at biblioteket ikke tvinges til å lage én hovedsortering av treffliste for søk, men har et stort potensiale i å lage brukergrensesnitt som i større grad kan la brukeren utforske samlingen.

Dette kan selvfølgelig gjerne kombineres med søk, og en av hovedutfordringene blir å presentere potensielle navigasjonsakser for brukeren i kontekst på en måte som ikke er forvirrende. *Filmatisering* er navigasjonsakse som er en god kandidat.

## Et autoritetsregister for verk basert på FRBR og RDF

Ut ifra diskusjonen foran trenger man følgende tjenester og logiske deler for å etablere et autoritetsregister basert på FRBR og Lenkede Data:

- En metode for å la andre systemer finne ID for verk ut ifra opplysningene de har i katalogposten. Dette gjøres ved et søk for å finne riktig manifestasjon og så følge relasjonen til tilhørende verk.
- En enkel datamodell, ontologi, basert på FRBR og RDF
- Publisering som Lenkede Data og med SPARQL endpoint
- Et system for generering og tilordning av unik ID for verk
- Et grensesnitt for å registrere og rette opp informasjon om forfatter, verk og tilhørende manifestasjoner i autoritetsregisteret.

Modellen bør være så enkel som mulig. Både fordi det er begrenset hva som kan gjøres innenfor rammene av en masteroppgave, men også fordi det er en fordel med en enklest mulig modell.

Det er et mål å maksimere gjenbruk av modeller laget av andre. Det ville være en teknisk fordel å kun legge inn en id som peker til mer informasjon hos andre hvis vi ikke er avhengige av å vise informasjonen i dette systemet. Dette har imidlertid en praktisk avveining.

Som et eksempel kunne vi i prinsippet klart oss med å registrere kun en ID for forfatter, uten navn og andre data. Dette kalles ofte en stubb (eng: stub), ufullstendig informasjon der vi må bruke en ID for å hente mer informasjon fra andre systemer ved f.eks. visning.

Et menneske som skal registrere informasjon i autoritetsregisteret vil ikke klare å tolke en person-id, og forfatternavn kan være nødvendig kontekst for å kunne redigere informasjon om et verk.

Ut ifra hensyn til den som redigerer må vi enten ha forfatternavnet i systemet vårt eller må koble oss opp mot andre systemer og hente det hver gang vi skal vise forfatternavn. Siden et forfatternavn i ontologien også gir mulighet for å gi svar på søk som inneholder forfatternavn er det fornuftig at dette er en del av ontologien.

Det er enkelt å utvide ontologien på et senere tidspunkt hvis det viser seg at vi har behov for flere informasjonsattributter for entitetene eller en mer kompleks modell.

## *ID for verk*

ID for verk må være en gyldig http-URI, ref. retningslinjene for Lenkede data:

<http://www.w3.org/DesignIssues/LinkedData.html>

En god kandidat som utgangspunkt for generering av identifikatorer er programpakken og rammeverket NOID: Nice Opaque Identifier (Minter and Name Resolver),

<https://wiki.ucop.edu/display/Curation/NOID>

Dette er et rammeverk med en java-implementasjon som er fri programvare med åpen kildekode og en fri lisens. Det er også lagt til rette for å kombinere generering av identifikatorer med mer avanserte løsninger for å legge til rette for persistente identifikatorer.

Som nevnt tidligere finnes det allerede noen sammenlignbare eksempler på identifikatorer brukt i lenkede data:

- [http://www.bibpode.no/work/Hamsun\\_Knut\\_den\\_sidste\\_glaede](http://www.bibpode.no/work/Hamsun_Knut_den_sidste_glaede)
- <http://viaf.org/viaf/99258155/>
- <http://data.bibsys.no/data/notrbib/authorityentry/x90061718>

Mitt forslag til identifikatorstruktur er:

```
http://<domenenavn>/<arkitektur-og-kode-for-verksid>/<8-sifret-tall>
```

Domenenavnet er avhengig av hvilken organisasjon som implementerer løsningen. I podeprosjektet ble **www.bibpode.no** brukt. Det viktigste her er at man velger et domenenavn som man har muligheten til å beholde kontrollen over i minst 100 år fremover.

Organisasjoner kommer, går, blir slått sammen og skifter navn. Persistente (vedvarende) identifikatorer er derfor en utfordring. Av samme årsak er det derfor en fordel at identifikatorene publiseres av en etablert organisasjon som har evne til å forvalte en slik løsning på svært lang sikt.

Når det gjelder neste komponent velger jeg å gjøre det enkelt og foreslår **id/work**. Før implementasjon i en organisasjon må dette uansett avklares nærmere. Hva som kommer etter domenenavnet avhenger både av IT-arkitektur i organisasjonen, vurderinger av stabilitet over tid og IT-politikk i den aktuelle organisasjonen.

Den siste komponenten er rett og slett et 8-sifret tall, som et telefonnummer, men uten mellomrom mellom sifferne. Dette er en god balanse mellom et antall siffer som mennesker klarer å forholde seg til og en mengde potensielle unike identifikatorer som er stor nok. Både Viaf og BIBSYS bruker 8-sifrede tall i person-id.

To eksempler på hvordan en id kunne se ut kan da være:

```
http://oda.hio.no/id/work/12345678
```

```
http://nb.no/id/work/34567890
```

## *Oppsummering og drøfting av designkriterier for autoritetsregisteret*

Autoritetsregisteret skal baseres på RDF , OWL, SPARQL og Lenkede data.

### **Søkegrensesnitt**

Skal støtte distribuert søk via et SPARQL endpoint.

#### *Input*

Søkegrensesnittet skal baseres på SPARQL, og ontologien skal støtte søk mot en kombinasjon av følgende attributter og relasjoner for manifestasjonsentiteten:

- ISBN
- Tittel
- År
- Forfatter

#### *Output*

Ontologien skal kunne levere kombinasjoner av følgende informasjon om verket

- Verks-id

- Tittel
- År
- Forfatter

### **Attributter og relasjoner for verk**

- Tittel (attributt)
- År (attributt)
- Relasjon til forfatter
- ID på formen <http://are.com/id/work/12345678>

### **Attributter og relasjoner for manifestasjon**

- ISBN (attributt)
- Tittel (attributt)
- År (attributt)
- Relasjon til forfatter
- ID

### **Attributter og relasjoner for person**

- Navn
- Relasjon til verk
- VIAF ID

### *Ontologidesign*

Ut ifra diskusjonen av behov for gjenfinning foran ser det ut til at vi i utgangspunktet kan slippe unna med en svært enkel ontologi.

Det er ikke behov for å modellere uttrykk siden det lages en relasjon direkte mellom verk og manifestasjon. Denne relasjonen er som tidligere nevnt en utvidelse av FRBR-modellen som er definert i FaBiO (the FRBR-aligned Bibliographic Ontology).



Det ikke er noen fasitsvar i forhold til valg av predikater i ontologien. Det er flere vokabularer å velge mellom for de aktuelle entitetene som er tilnærmet identiske. Man må søke litt og se hva som er i bruk hos andre og ta en sjanse. I forhold til funksjonaliteten i denne prototypen er det ikke så nøye om man velger et predikat fra dcterms- eller frbr-vokabularet.

Hvis systemet skal realiseres i større skala er dette noe som må vurderes nærmere, og det kan være basert på praktiske forhold som hva nære samarbeidspartnere bruker eller IT-politiske beslutninger.

## Person

Som diskutert tidligere trenger vi i kun informasjonsattributtene ID og navn.

Vi har valgt å bruke VIAF-identifikator.

Det er også et spørsmål om hva slags navneform vi skal bruke. Vi kan eventuelt bruke en hovedform og flere varianter, men ønsker å gjøre det enkelt. Dette drøftes nærmere under.

I forhold til hovedmålet for ontologien er det ikke nødvendig å registrere en relasjon fra forfatter til verk når vi registrerer en forfatter. Det er tilstrekkelig å legge inn en relasjon fra verk til forfatter når vi registrerer verk.

British Library definerer en egen relasjon *hasCreated* som er definert som owl:inverseOf dct:creator: (Tim.Hodson@talis.com, 2011)

```
hasCreated owl:inverseOf dcterms:creator
```

Jeg velger imidlertid å bruke frbr:creatorOf og definere denne som owl:inverseOf frbr:creator, se vedlegg I for detaljer.

### *Navnevarianter*

Når det gjelder navn har jeg tvilt meg frem til å støtte flere navnevarianter. Dette blir litt avhengig av bruken og mulighetene for å manipulere input-strenger. Invertert form er standard i katalogpostene, og vi ønsker antageligvis dette for å få en alfabetisk oversikt sortert på etternavn. Vi ønsker imidlertid fornavn-etternavn ved visning for publikum.

I foaf-vokabularet er ikke invertert form vanlig, så jeg har definert en lokal navnevariant `invertedName` som er en subtype av `foaf:name`. I tillegg har jeg brukt `foaf:givenName` og `foaf:familyName`, som er anbefalt fremfor `foaf:name`. Se vedlegg I eller vedlagt elektronisk ontologifil for detaljer. I en virkelig verden ville man ha manipulert tekststrengene og kunne populert alle disse variantene fra én kilde.

## Verk

Vi ønsker å ta med ID og attributtene tittel og år.

Relasjon til forfatter:

```
dcterms:creator
```

Det må defineres en relasjon til manifestasjoner:

```
fabio:hasManifestation
```

Det vil være fornuftig å registrere verk før registrering av manifestasjoner. Vi behøver ikke registrere manifestasjoner når vi registrerer verk. Vi kan modellere denne relasjonen som refleksiv:

```
fabio:hasManifestation owl:inverseOf fabio:isManifestationOf
```

Dette vil gjøre at det holder å registrere relatert verk for alle manifestasjoner, da vil vi automatisk kunne få en relasjonen den andre veien.

Det viser seg at fabio-vokabularet har predikatet `fabio:hasPublicationYear`, så vi bruker dette for både verk om manifestasjon.

## Manifestasjon

Som diskutert tidligere er det behov for informasjonsattributtene ISBN og tittel og år.

Det finnes flere muligheter for å velge predikat som knytter ISBN til manifestasjon. Valget faller på `prism:isbn`, som er brukt av mange andre. (*PRISM Basic Namespace (prism:)*, udatert)

```
http://prismstandard.org/namespaces/1.2/basic/isbn
```

Når det gjelder tittel er predikatet `dcterms:title` mye brukt, så dette predikatet velges.

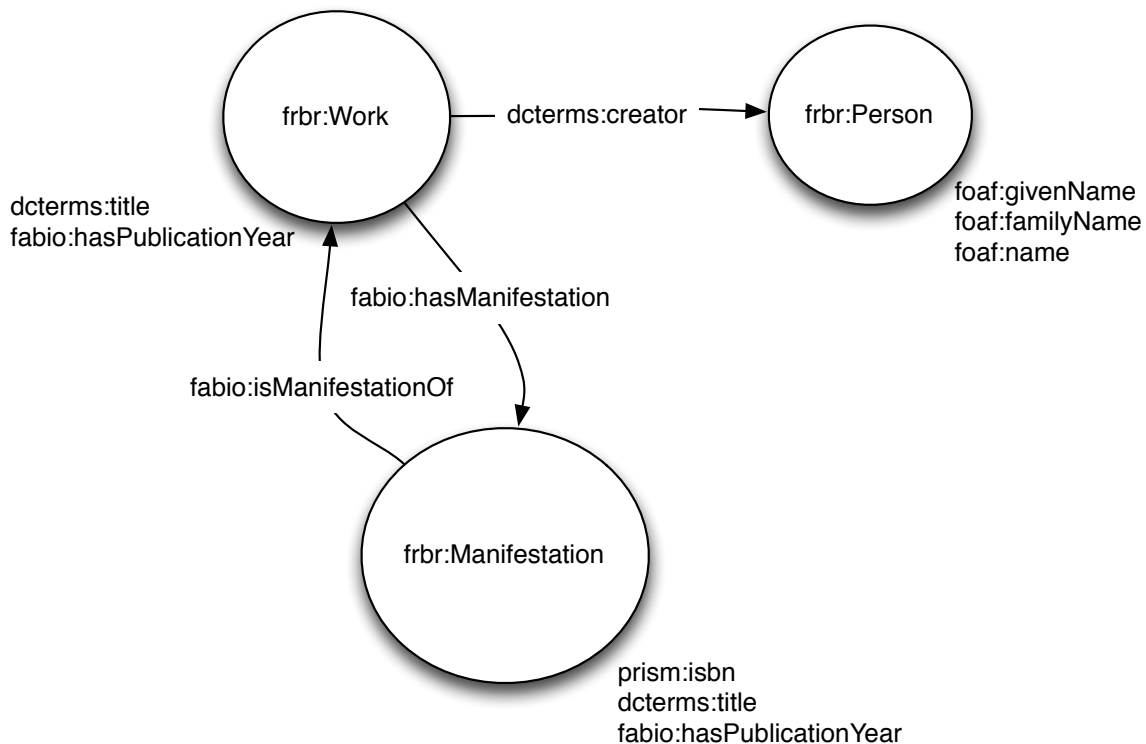
`http://purl.org/dc/terms/title`

Som for verk bruker vi predikatet `fabio:hasPublicationYear` fra `fabio`-vokabularet.

Som drøftet tidligere bruker vi predikatet `fabio:isManifestationOf` fra manifestasjon til verk:

`fabio:isManifestationOf`

## Konseptuell modell av Ontologi



**Figur 19: Konseptuell modell for autoritetsregisteret**

*Manglende relasjon fra manifestasjon til forfatter?*

Det kan virke som om det mangler en relasjon fra manifestasjon til forfatter i ontologien.

Rent praktisk vil alltid forfatteren registreres først. Deretter registreres verkene, som alltid har en relasjon til forfatteren (unntatt hvis forfatter er ukjent).

Alle manifestasjoner relateres til et verk.

Relasjonen direkte fra manifestasjon til forfatter er dermed implisitt i ontologien. Man kan finne forfatter for manifestasjonen ved å finne verkets forfatter. Det er dermed unødvendig å kreve at dette registreres for manifestasjon, det holder å relatere manifestasjon til verk.

Det er mulig å lage en regel i ontologien som utleder denne relasjonen automatisk ved hjelp av en regel i ontologien som bruker såkalt deduksjon (inference). Dette vil imidlertid kreve en kraftigere resoneringsmotor enn ontologien bruker per i dag, så dette vurderes å være unødvendig.

Vi kan enkelt finne forfatter(e) for en manifestasjon ved hjelp av spørring med SPARQL, så vurderingen er at det er bedre å bruke SPARQL enn resonering til dette.

### *Minimalistisk modell med gode utvidelsesmuligheter*

Et av designkriteriene her er å lage dette så minimalistisk som praktisk mulig.

En av de store fordelene med semantisk teknologi i forhold til f.eks. relasjonsdatabaser er at det er svært enkelt å utvide ontologien etterhvert som behov og ønsker utvikler seg. Man kan legge opp til en evolusjonær utviklingsprosess der ontologien gror frem dynamisk.

### *Implementasjon av en forenklet prototype i Protégé*

Denne oppgaven har en relativt bred problemstilling, så det gjaldt å finne en enkel måte å realisere ontologien på innenfor rammene av en masteroppgave. Det er en stor fordel om prototypen som lages kan kjøres av andre, men det er ikke rom for å lage en full prototype som en webtjeneste.

Protégé er kanskje standardverktøyet innen modellering med RDF og OWL. Det er åpen og fri programvare utviklet ved Stanford Center for Biomedical Informatics Research. Det er en Ontologieditor og et kunnskapsbase-rammeverk.

Målet i dette kapitlet er å vise at jeg har implementert ontologien som en prototype i Protégé og har testet prototypen ved å kjøre SPARQL-spørringer.

Protégé kan også tilpasses bruk som editor ved å lage skjemaer for å tilrettelegge for registrering av informasjon om verk, forfattere og manifestasjoner. På denne måten kan Protégé brukes som verktøy for eventuelle bibliotekansatte som skal legge inn informasjon i autoritetsregisteret. En tilsvarende løsning er allerede i bruk i Pop-senteret, der IT-arkitekturen også baserer seg på RDF/OWL og SPARQL. (Engels, 2011)

## Realisering av ontologi i Protégé

Ontologien ble kodet med RDF og OWL og lastet opp i verktøyet Protégé. Se vedlegg I for RDF- og OWL-kode.

Det ble laget skjemaer (forms) i Protégé for å kunne registrere instanser av forfatter, verk og manifestasjoner, se figur 20, 21 og 22.

Forfatteren Knut Hamsun og verkene Sult og Sværmere med åtte tilhørende manifestasjoner ble registrert ved å skrive inn informasjon via skjemaene i Protégé, se vedlagt RDF-kode kode i vedlegg I eller prosjektfilene som er elektroniske vedlegg, beskrevet i vedlegg 3.

The screenshot shows the 'INDIVIDUAL EDITOR' for the URI `http://viaf.org/viaf/97006051/`, which is an instance of the class `frbr:Person`. The editor displays four property forms:

- foaf:givenName:** A table with columns 'Value' and 'Lang'. The value 'Knut' is entered in the 'Value' column.
- foaf:familyName:** A table with columns 'Value' and 'Lang'. The value 'Hamsun' is entered in the 'Value' column.
- invertedName:** A table with columns 'Value' and 'Type'. The value 'Hamsun, Knut' is entered in the 'Value' column, and the type 'string' is entered in the 'Type' column.
- foaf:name:** A table with columns 'Value' and 'Lang'. Two values are entered in the 'Value' column: 'Knut Hamsun' and 'Hamsun, Knut'.

Figur 20: Definisjon av input-skjema for forfatter i Protégé

Skjemaene er av enkleste sort og må finpusses hvis de skal brukes i praksis. Det kommer f.eks. automatisk et språkattributt på flere felt i Protégé, noe som ikke burde være med i denne konteksten.

**INDIVIDUAL EDITOR for <http://are.com/id/work/00000001> (instance of frbr:Work)**

For Individual: <http://are.com/id/work/00000001>

**dcterms:title** 🔍 + ✕

Value	Type
Sværmere	string

**frbr:hasPublicationYear** 🔍 + ✕

Value	Type
1904	int

**frbr:creator** ✕ + ↵

- ◆ <http://viaf.org/viaf/97006051/>

**frbr:hasManifestation** ✕ + ↵

- ◆ <http://are.com/id/manifestation/8205210063>
- ◆ <http://are.com/id/manifestation/8205257094>
- ◆ <http://are.com/id/manifestation/8252534686>
- ◆ <http://are.com/id/manifestation/hamsun-sværmere-1904>

Figur 21: Definisjon av input-skjema for verk i Protégé

**INDIVIDUAL EDITOR for <http://are.com/id/manifestation/8205210063> (instance of frbr:Manifestation)**

For Individual: <http://are.com/id/manifestation/8205210063>

**dcterms:title** 🔍 + ✕

Value	Type
Sværmere	string

**frbr:hasPublicationYear** 🔍 + ✕

Value	Type
1993	int

**prism:isbn** 🔍 + ✕

Value	Type
82-05-21006-3	string

Figur 22: Definisjon av input-skjema for manifestasjon i Protégé

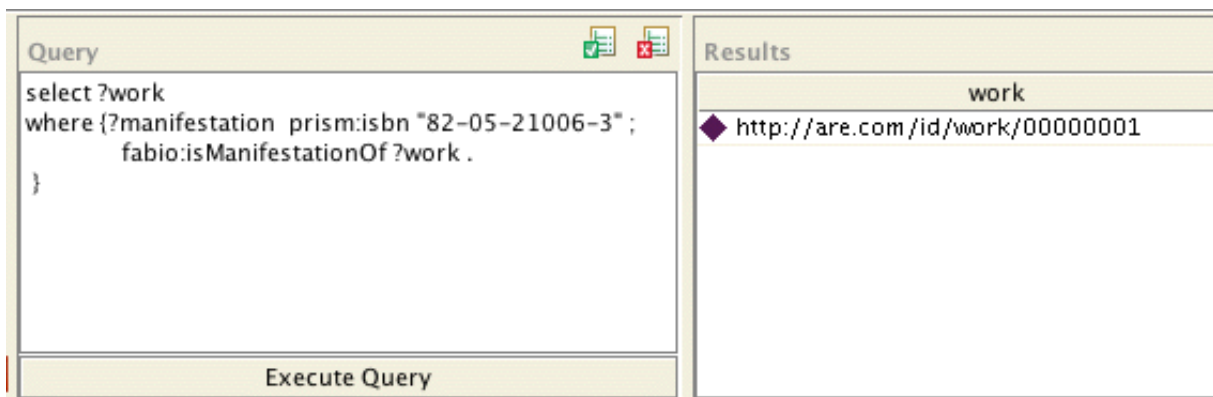
## Test av designkriterier på prototypen

Målet i dette kapittelet er å vise at den realiserte modellen lar seg bruke i henhold til designkriteriene.

Protégé har innebygget støtte for SPARQL-spørringer. Dette ble utnyttet til å lage søk som testet designkriteriene, og returnerte verks-id ved input av data for manifestasjon.

Se figur 23, 24 for skjermbilder og SPARQL-koden under.

### SPARQL-spørring med ISBN som input



The screenshot shows a SPARQL query interface. On the left, a text area contains the following query:

```
select ?work
where {?manifestation prism:isbn "82-05-21006-3";
      fabio:isManifestationOf ?work .
}
```

Below the query area is a button labeled "Execute Query". On the right, a "Results" table is displayed with one row:

work
<a href="http://are.com/id/work/00000001">http://are.com/id/work/00000001</a>

Figur 23: SPARQL-uttrykk med ISBN som input og ID for verk som output

```
select ?work
where

{?manifestation prism:isbn "82-05-21006-3";fabio:isManifestationOf ?work
.}
```

Resultat: <http://are.com/id/work/00000001>

### SPARQL-spørring med tittel og år som input

```
select ?work
where

{?manifestation dcterms:title "Sult";fabio:hasPublicationYear
"1944"^^xsd:int;fabio:isManifestationOf ?work .}
```

Resultat: <http://are.com/id/work/00000002>

## SPARQL-spørring som returnerer både ID, tittel og år for verk

Query	Results						
<pre>select ?work ?worktitle ?workyear where {?manifestation dcterms:title "Sult";       fabio:hasPublicationYear "1944"^^xsd:int;       fabio:isManifestationOf ?work .       ?work dcterms:title ?worktitle;       fabio:hasPublicationYear ?workyear. }</pre>	<table border="1"><thead><tr><th>work</th><th>worktitle</th><th>workyear</th></tr></thead><tbody><tr><td>◆ <a href="http://are.com/id/work/00000002">http://are.com/id/work/00000002</a></td><td>Sult</td><td>1890</td></tr></tbody></table>	work	worktitle	workyear	◆ <a href="http://are.com/id/work/00000002">http://are.com/id/work/00000002</a>	Sult	1890
work	worktitle	workyear					
◆ <a href="http://are.com/id/work/00000002">http://are.com/id/work/00000002</a>	Sult	1890					

Figur 24: SPARQL-uttrykk som returnerer ID, tittel og år for tilhørende verk.

```
select ?work ?worktitle ?workyear
where
```

```
{?manifestation dcterms:title "Sult";fabio:hasPublicationYear
"1944"^^xsd:int;fabio:isManifestationOf ?work .      ?work
dcterms:title ?worktitle;fabio:hasPublicationYear ?workyear.}
```

Resultat: <http://are.com/id/work/00000002> Sult 1890



## Konklusjon

Masteroppgaven har drøftet PODE-prosjektets ønske om et nasjonalt autoritetsregister med åpne, nasjonale unike ID-er for verk.

PODE-prosjektet utforsket FRBR, semantiske teknologier og Lenkede data. Oppgaven drøftet om dette var fornuftige teknologivalg for et autoritetsregister. Ut ifra en gjennomgang av utviklingstrender i det norske biblioteklandskapet og Internasjonale utviklingstrekk ble det konkludert med at PODE-prosjektets valg var gode.

Et av spørsmålene var hvordan en unik ID for verk bør se ut. Oppgaven har drøftet og designet en ID for verk, som tilfredsstiller designkriteriene som ble etablert. Det er foreslått at rammeverket NOID kan brukes til å generere identifikatorer, men dette er ikke testet i prototypen.

Det er gjort en historisk og konseptuell analyse av hva som ligger i begrepet verk, og hvilken rolle verket har i bibliotek katalogen, basert på foreliggende litteratur.

Oppgaven har også drøftet hva som ligger i begrepet autoritetsregister, og hvordan et autoritetsregister kan løse PODE-prosjektets hovedutfordringer.

Hovedspørsmålet var hva som skulle til for å tilfredsstille PODE-prosjektets ønske?

Hypotesen var at det lot seg gjøre å modellere og prototype et forenklet autoritetsregister som løser PODEs utfordring ved hjelp av FRBR, RDF, OWL og Lenkede data, innenfor rammene av masteroppgaven.

Konseptuell analyse og drøfting ble brukt for å komme frem til designkriterier for et autoritetsregister. Designkriteriene var basis for å lage en konseptuell modell av et forenklet autoritetsregister, en ontologi.

Ontologien ble realisert med RDF og OWL, og en begrenset prototype av systemet ble implementert i ontologiverktøyet Protégé.

Det ble definert skjemaer for innlegging av instansinformasjon. Dette er en enkel prototype med tanke på at bibliotekarer kan gjøre det samme. Arkivarer ved Popsenteret bruker per

juni 2012 Protégé på denne måten til innlegging av informasjon i samlingen, så det fungerer i praksis.

Forfatteren Knut Hamsun, verkene Sult og Sværmere og åtte tilhørende manifestasjoner ble registrert ved å skrive inn informasjon via skjemaene i Protégé.

Som test av prototypen ble det kjørt SPARQL-spørringer mot ontologien og eksempelinstanser i Protégé, for å se om designkriteriene var tilfredsstillende. Det fungerte som planlagt.

Konklusjonen er at arbeidet har forsterket hypotesen om at det er mulig å lage et autoritetsregister basert på FRBR og Lenkede data som løser PODE-prosjektets utfordringer.

Basert på arbeidet kan det se ut til at en relativt enkel ontologi løser utfordringene.

Dette er en komplementerende tilnærming i forhold til FRBRiseringstilnærmingen brukt i PODE-prosjektet, og metodene kan gjerne kombineres. Denne tilnærmingen gjør at PODE-prosjektet ikke trenger å rydde postene for å få koblet manifestasjonene til riktig verk.

Det må imidlertid gjøres et autoritetsarbeid i å legge inn data inn i systemet. Som nevnt i forslaget til videre arbeid er det potensiale for å automatisere store deler av denne prosessen.

Dette burde være et godt utgangspunkt for å teste en utvidet prototype, som ikke er like forenklet, og nærmere det praktiske behovet.

Jeg er fornøyd med det forskningsmessige designet av prototypen. Det var en stor fordel å bruke Protégé. Siden Protégé og SPARQL brukes, kan eksperimentene og prototypen etterprøves av andre.

Jeg har lagt ved to Protégé-filer ved innlevering, som inneholder prosjektinfo og data, se nærmere beskrivelse i vedlegg 3. Interesserte kan installere Protégé på vanlig måte, og laste inn hele prosjektet slik at de kan teste det jeg har gjort.

Vitenskapelig sett er jo etterprøvbare eksperimenter et poeng, og da bør man jo kunne laste opp datagrunnlaget, som i mitt tilfelle er to filer på til sammen 200 KB.

## *Videre arbeid*

Prototypen kan utvikles videre i nær sagt alle retninger. Både med mer data og rikere funksjonalitet.

### **Automatisk generering av data i autoritetsregisteret**

I den forenklete løsningen i oppgaven må data om forfatter, verk og manifestasjoner legges manuelt inn i autoritetsregisteret. Det burde være svært gode muligheter for å automatisere ihvertfall deler av denne prosessen.

Registrering av data for attributter knyttet til manifestasjoner burde i det minste kunne gjøres automatisk. Det vil i så fall forenkle arbeidet med manifestasjonsregistrering til kun lenking av manifestasjon til riktig verk.

### **Utnytte en rik bibliografisk kilde**

Et nasjonalt autoritetsregister bør hente data fra en rik katalog for å ha nok informasjon om ulike utgaver og verkstittel, f.eks. fra en nasjonalbibliografi.

En nasjonalbibliografi kan også fungere som et autoritetsregister, dvs. være autoritativ med hensyn til identifikatorer, tittel, forfatter mm. for innholdet i nasjonalbibliografien. Det er dette British Library jobber mot når de har valgt å publisere nasjonalbibliografien som Lenkede åpne data.

### **Etablering av identitet uten identifikatorer**

Det kan være aktuelt å legge mer arbeid i å kunne etablere verks- og manifestasjonsidentitet automatisk når man ikke har identifikatorer i katalogposten. Se (Linking data without common identifiers, 2011) og (Takhirov et al., 2011) for en beskrivelse av mer robuste strategier for hvordan man kan lenke sammen FRBR-entiteter og Lenkede data uten delte identifikatorer. Dette fagfeltet har flere navn, bl.a. record linkage, entity resolution, identity resolution og deduplication. Det var dessverre utenfor rekkevidde for denne masteroppgaven å gå nærmere inn på disse løsningene, utover å nevne dette som mulig videre arbeid.

## Generere unike verks-ID ved hjelp av NOID

Det er utenfor rekkevidde av oppgaven å implementere NOID, både siden jeg baserer meg på å bruke Protégé og fordi det blir for stort omfang innenfor masteroppgaven. Dette må imidlertid gjøres hvis løsningen skal realiseres.

## Publisering som Lenkede åpne data

I metodedelen ble det referert til Tim Berners-Lees regler og anbefalinger for Lenkede data. Han graderer her ulike implementasjoner av Lenkede data med stjerner. Autoritetsregisteret kan få stjerner for hver av disse punktene:

1. Publisere data med en åpen lisens
2. RDF er maskin-lesbare data
3. RDF er et standardisert format
4. Publisering ved hjelp av åpne standarder fra W3C (RDF og SPARQL)
5. Lenke til andre data

De tre første kravene er tilfredsstillt forutsatt at det publiseres med en åpen lisens.

Løsningen vil gjenbruke, dvs. lenke til, person-id for forfattere. Enten det brukes id fra BIBSYS' autoritetsregister eller VIAF-id vil dette gi den fjerde stjernen.

De femte stjernen, for punkt 4, er litt mer usikker, og avhengig av hvordan man tolker kravet. Noen mener her det er nok med et SPARQL endpoint, men i tillegg bør det være et system som resolver identifikatorene for å få lenkede data, hvis ikke blir det som nevnt i metodekapittelet søkbare data. Det er derfor vi bruker http-URI som id. Hvis noen prøver å bruke identifikatoren bør de få svar. Ideelt bør maskiner få RDF-kode som svar og personer som putter identifikatoren inn i en nettleser bør få en tekstlig beskrivelse av hva det er en identifikator for.

En relativt vanlig måte å løse dette er med en såkalt 303-redirect, som diskutert i metodekapittelet.

## Tilrettelegge og oppfordre til gjenbruk av data

Siden innholdet publiseres som lenkede data ligger det til rette for gjenbruk.

Interesserte kan laste ned de essensielle RDF-data for å presentere en bibliografi for forfatterne som er registrert i autoritetsregisteret. Dette er informasjon som er interessant langt utover bibliotekene, f.eks. for:

- Alle som selger bøker
- Alle som lager webtjenester relatert til bøker

Dette er også interessant informasjon å bruke sammen med informasjon om forfatteren gjennom en mash-up.

## Visningsnavn for verk i ulike kontekster

Hvis man kun har én tittel for verk må denne brukes til visning i alle kontekster. En viktig kontekst er språk, og her har typisk et verk forskjellig navn på ulike språk.

Som diskutert tidligere kan inspirasjon hentes fra emnekartmodellen og display-name og scope. Et verk kan i denne modellen ha hatt mange visningsnavn i ulike kontekster, der kontekstene bl.a. kan være språk. Det burde være mulig å modellere noe tilsvarende i RDF og OWL.

Det ville ikke være nødvendig for autoritetsregisteret å registrere alle variantene. Norsk og engelsk ville vært naturlig. Andre kan utvide informasjonen ved å legge til andre språk etter behov.

## Litteraturliste

Aalberg, T., Haugen, F., & Husby, O. (2006). A Tool for Converting from MARC to FRBR. In J. Gonzalo, C. Thanos, M. Verdejo, & R. Carrasco (Eds.), *Lecture Notes in Computer Science: Vol. 4172. Research and advanced technology for digital libraries* (pp. 453-6). Springer Berlin / Heidelberg. Hentet fra [http://dx.doi.org/10.1007/11863878\\_41](http://dx.doi.org/10.1007/11863878_41)

Authority Control in the Card Environment. (udatert). Authority Control in the Card Environment.. Hentet fra <http://libraries.vermont.gov/sites/libraries/files/tsu/lesson2card.htm>

- Avram, H. D. (1976). International Standards for Interchange of Bibliographic Records in Machine-Readable Form. *Library Resources and Technical Services*, 20(1), 25-35.
- Bales, S. E. (2009). Tracing the Archetypal Academic Librarian. *Library Philosophy and Practice*, 2009(November). Hentet fra <http://digitalcommons.unl.edu/libphilprac/309/>
- Berners-Lee. (2009). The next Web of open, linked data. *TED conference*. Hentet fra [http://blog.ted.com/2009/03/13/tim\\_berners\\_lee\\_web/](http://blog.ted.com/2009/03/13/tim_berners_lee_web/)
- Berners-Lee, T. B. (2006). Linked Data - Design Issues. *W3.Org*. Hentet fra <http://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T. B., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34-43.
- Bertheussen, L. E., Greenall, R., & Husby, O. (2010). *BIBSYS authority to linked data*. Hentet fra <https://docs.google.com/document/pub?id=IrxzLvyVgf5aDVtiKdZ4qV6HLdlvelLF4cH5Ey9cQNK0>
- Bing, J. (udatert). Kunnskapsbaserte system for representasjon av rettslige normer. *Presentasjon på konferansen emnekart 2009*. Hentet fra <http://forum.dataforeningen.no/attachment.php?attachmentid=3158>
- Blum, R. (1991). *Kallimachos: The alexandrian library and the origins of bibliography*. University of Wisconsin Press .
- Brown, J. D. (1916). *Library classification and cataloguing*. London: Grafton.
- Bygrave, L. (1996). Ensuring Right Information on the Right Person(s): Legal Controls of the Quality of Personal Information. *Manuscript Series on Information Technology and Administrative Systems*, 4(4/96). Hentet fra [http://www.jus.uio.no/ifp/om/organisasjon/afin/forskning/notatserien/1996/4\\_96.html](http://www.jus.uio.no/ifp/om/organisasjon/afin/forskning/notatserien/1996/4_96.html)
- Casson, L. (2002). *Libraries in the ancient world* (reprint, illustrated ed.). New Haven: Yale University Press. Hentet fra <http://www.scribd.com/doc/29547947/Libraries-in-the-ancient-world>

Coyle, K.. *Coyle's information* (2009). What is a (FRBR) Work? Hentet 4 Juni. 2012 fra <http://kcoyle.blogspot.no/2009/08/what-is-frbr-work.html>

Coyle, K. (2010). *Understanding the semantic web : Bibliographic data and metadata*. Chicago, IL: ALA TechSource.

Cutter, C. A. (1904). *Rules for a dictionary catalog*. United States: U.S. Bureau of Education. Hentet fra <http://digital.library.unt.edu/ark:/67531/metadc1048/m1/1/>

Den fruktbare halvmåne. (udatert). I *Wikipedia*. Hentet 6 Mar. 2012 fra [http://no.wikipedia.org/wiki/Den\\_fruktbare\\_halvm%25ne](http://no.wikipedia.org/wiki/Den_fruktbare_halvm%25ne)

Den norske katalogkomité. (2004). Nasjonalt autoritetsregister for person- og korporasjonsnavn. [Utredning] (Utredning). Nasjonalbiblioteket. Hentet fra <http://www.nb.no/content/download/2797/20723/version/1/file/96autregister-sluttrapport-okt2004.pdf>

Denton, W. (2007). FRBR and the History of Cataloging. *Understanding FRBR: What It Is and How It Will Affect Our Retrieval*. Ed. Arlene G. Taylor. Westport, Connecticut: Libraries Unlimited.

Denton, W. (2007). FRBR and the History of Cataloging. *Understanding FRBR: What It Is and How It Will Affect Our Retrieval*. Ed. Arlene G. Taylor. Westport, Connecticut: Libraries Unlimited.

Dodd, S., & Uppsala universitetsbibliotek. (2003). *Bibliotekstermer på svenska och engelska (continually updated resource)* [worldcat.Org]. Uppsala: Uppsala universitet. Hentet fra [http://www.worldcat.org/title/bibliotekstermer-pa-svenska-och-engelska/oclc/789839531&referer=brief\\_results](http://www.worldcat.org/title/bibliotekstermer-pa-svenska-och-engelska/oclc/789839531&referer=brief_results)

Engels, R. (2011). *Where is the music in Semantics? Europeana tech conference*. Wien. Hentet fra [http://www.europeanaconnect.eu/europeanatech/documents/Engels\\_Robert.pdf](http://www.europeanaconnect.eu/europeanatech/documents/Engels_Robert.pdf)

Filliozat, P. -S. (2005). Ancient Sanskrit Mathematics: An Oral Tradition and a Written Literature. In K. Chemla (Ed.), *Boston Studies in the Philosophy of Science: Vol. 238. History of science, history of text* (pp. 137-57). Springer Netherlands. Hentet fra [http://dx.doi.org/10.1007/1-4020-2321-9\\_7](http://dx.doi.org/10.1007/1-4020-2321-9_7)

FRBRisering. (2011). FRBRisering. *Podes prosjektsider*. Hentet 15 Mai. 2012 fra <http://www.bibpode.no/?q=node/11>

Free Data Services. (udatert). Free Data Services.. Hentet fra <http://www.bl.uk/bibliographic/datafree.html>

Freese, J. (1987). *Den maktfullkomliga oförmågan*. Stockholm: Wahlström & Widstrand.

Garshol, L. M.. *Slideshare: Linking data without common identifiers*(2011). Linking data without common identifiers. Hentet 29 Mai. 2012 fra <http://www.slideshare.net/larsga/linking-data-without-common-identifiers>

Gatenby. (2010). Workshop on FRBR and Identifiers. *ELAG 2010*. Hentet fra <http://elag2010.nationallibrary.fi/files/2010/06/ELAG-2010-workshop-on-FRBR-and-identifiers.pdf>

Genz, M. (2002). Review: The Nature of "A Work": Implications for the Organization of Knowledge by Richard P. Smiraglia. . *JSTOR: Journal of Education for Library and Information Science*, 43(1 (Winter, 2002)), 87-89. Hentet fra <http://www.jstor.org/stable/40323993>

Giese, D. M. (2012). Lenket data og spørbare data - er fem stjerner bra nok? *Communities in action 2012*. Oslo. Hentet fra <https://wiki.cantara.no/pages/viewpage.action?pageId=27788038>

Glassner, J. -J. (2003). *The invention of cuneiform* (illustrated ed.) Baltimore: JHU Press.

global warming (Earth science) : Climatic variation since the last glaciation. (udatert). I *Britannica online encyclopedia*. Hentet 15 Mar. 2012 fra <http://www.britannica.com/EBchecked/topic/235402/global-warming/274818/Climatic-variation-since-the-last-glaciation?anchor=ref979219>

Grønnestad. (2012).(2012, February 11). Årets bibliotekprosjekt 2011. *Bok Og Bibliotek*, ( nr 1/2012), 20-22. Hentet fra [http://www.bokogbibliotek.no/images/stories/pdf\\_2012/bob-nr1-2012-web-s20-26.pdf](http://www.bokogbibliotek.no/images/stories/pdf_2012/bob-nr1-2012-web-s20-26.pdf)

Hegna, K. (2007). The objectives of cataloguing, the technology and the user interface.



Hegna, K., & Murtooma, E. (2003). Data mining MARC to find: FRBR? *International Cataloguing and Bibliographic Control*, 32(3), 52-55.

Hoesen, H. B. V. (1944). Perspective in Cataloging, with Some Applications. *The Library Quarterly*, 14(2), 100-107. Hentet fra <http://www.jstor.org/stable/4303213>

Holm, L. A., IFLA Study Group on the Functional Requirements for Bibliographic Records, & Nasjonalbiblioteket. (2001). *Funksjonskrav til bibliografiske poster : Sluttrapport*. Oslo: Nasjonalbiblioteket.

IFLA. (2008). *Erklæring om internasjonale katalogiseringsprinsipper*. Hentet fra <http://www.nb.no/content/download/7894/56208/version/1/file/Katalogprinsipper.pdf>

IFLA Study Group on the Functional Requirements for Bibliographic Records. (2009). *UBCIM publications, new series: Vol. 19. Functional requirements for bibliographic records : Final report*. München: K.G. Saur Verlag GmbH & Co. (Original work published 1997) Hentet fra <http://www.ifla.org/files/cataloguing/frbr/frbr.pdf>

Informasjon. (udatert). I *Wikipedia*. Hentet 5 Jun. 2012 fra <http://no.wikipedia.org/wiki/Informasjon>

ISTC - *International Standard Text Code*. *ISTC - international standard text code*(udatert).Hentet 5 Juni. 2012 fra <http://www.istc-international.org/html/>

Kent, W., & Hoberman, S. (2012). *Data & reality* (3 ed.). Westfield, NJ: Technics Publications, LLC.

Knutsen, U. (2009). Fragmentering eller fellesløsning? : organisering av norsk bibliografisk produksjon. *ABM Skrift*, (60). Hentet fra <http://www.abm-utvikling.no/publisert/abm-skrift/abm-skrift-60>

Korpela, J. (udatert). *A commentary of wiio's laws*. Hentet fra <http://www.cs.tut.fi/~jkorpela/wiio.html>

Lakoff, G. (1994). *Women, fire, and dangerous things : What categories reveal about the mind*. Chicago: University of Chicago Press.

Lawler, A. (2012). Uncovering Civilization's Roots. *Science*, 335(6070), 790-793.

The Library of Ashurbanipal. *British museum* (udatert). The Library of Ashurbanipal. Hentet 21 Mai. 2012 fra

[http://www.britishmuseum.org/research/research\\_projects/all\\_projects/ashurbanipal\\_library\\_phase\\_1.aspx](http://www.britishmuseum.org/research/research_projects/all_projects/ashurbanipal_library_phase_1.aspx)

Lubetzky, S. (1953). Development of Cataloging Rules.

Lubetzky, S. (1969). Principles of Cataloging. Final Report. Phase I: Descriptive Cataloging.

Lubetzky, S., & American Library Association, B. o. C. P. a. R. (1953). *Cataloging rules and principles a critique of the A.L.A. Rules for entry and a proposed design for their revision. Prepared for the board on cataloging policy and research of the A.L.A. Division of cataloging and classification.* Washington: Processing Dept., Library of Congress. Hentet fra

<http://catalog.hathitrust.org/api/volumes/oclc/1373808.html>

MacLeod, R. (2000). *The library of alexandria: Centre of learning in the ancient world.* London: I.B. Tauris .

Marcum, D.. (2011). A Bibliographic Framework for the Digital Age. *Bibliographic framework transition initiative (library of congress).* Hentet 1 Mai. 2012 fra

<http://www.loc.gov/marc/transition/news/framework-103111.html>

Maxwell, R. L. (2008). *FRBR : A guide for the perplexed.* Chicago: American Library Association.

Muntlig litteratur. (udatert). I *Wikipedia*. Hentet 6 Mar. 2012 fra

[http://no.wikipedia.org/wiki/Muntlig\\_litteratur](http://no.wikipedia.org/wiki/Muntlig_litteratur)

Muntlig tradisjon. (udatert). I *Wikipedia*. Hentet 6 Mar. 2012 fra

[http://no.wikipedia.org/wiki/Muntlig\\_tradisjon](http://no.wikipedia.org/wiki/Muntlig_tradisjon)

OCLC Policy Statement on RDA Cataloging in WorldCat for the U.S. Testing Period and Beyond. (2011). OCLC policy statement on RDA Cataloging in WorldCat for the U.S.

testing period and beyond. *Oclc.Org*. Hentet 1 Mai. 2012 fra  
<http://www.oclc.org/us/en/rda/policy.htm>

ODLIS — *Online Dictionary for Library and Information Science*. (udatert).. Hentet fra  
[http://www.abc-clio.com/ODLIS/odlis\\_A.aspx](http://www.abc-clio.com/ODLIS/odlis_A.aspx)

Oliver, C. (2010a). *Introducing RDA : A guide to the basics*. Chicago: ALA Editions.

Oliver, C. (2010b). *Introducing RDA : A guide to the basics*. Chicago: ALA Editions.

Patton, G. E., (FRANAR), I. W. G. O. F. R. A. N. O. A. R., International, F. F. O. L. A. A. I. C. S. S., & International, F. F. O. L. A. A. I. S. O. C. A. I. S. (2009). *Functional requirements for authority data : A conceptual model*. München: K.G. Saur.

Peters, J. P. (1905). The Nippur Library. *Journal of the American Oriental Society*, 26, 145-164.  
doi:10.2307/592887

Philips, H. (2010). The Great Library of Alexandria? *Library Philosophy and Practice*, 2010(August). Hentet fra <http://digitalcommons.unl.edu/libphilprac/417/>

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). The communicative function of ambiguity in language. *Cognition*.

Pjotr Tsjajkovskij . (udatert). I *Wikipedia*. Hentet 29 May 2012 fra  
<http://no.wikipedia.org/wiki/Tchaikovsky>

Prinsipper for Rettinger Av Hamsun-poster. (2010). Prinsipper for rettinger av Hamsun-poster.. Hentet fra <http://www.bibpode.no/blogg/?p=1271>

PRISM Basic Namespace (*prism:*). (udatert).. Hentet fra  
<http://prismstandard.org/namespaces/1.2/basic/>

Pyotr Ilyich Tchaikovsky. (udatert). I *Wikipedia, the free encyclopedia*. Hentet 29 May 2012 fra  
<http://en.wikipedia.org/wiki/Tchaikovsky>

Resource Description Framework (RDF) Model and Syntax Specification. W3C(udatert).  
Resource Description Framework (RDF) Model and Syntax Specification. Hentet 6 Juni. 2012  
fra <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>

Rindal, M., & katalogkomité, D. N. (1994). *Anonyme norske mellomalderverk : Ei liste over standardtitlar*. Oslo: Norsk bibliotekforening.

Røyneberg, E. (2010). Nytt biblioteksystem. *BIBSYS brukermøte 2011*. Hentet fra  
[http://www.bibsys.no/norsk/bibliotekar/brukermoter/brukermote\\_2011/streaming/](http://www.bibsys.no/norsk/bibliotekar/brukermoter/brukermote_2011/streaming/)

SAG. (1990). *I Wikipedia*. Hentet fra <http://en.wikipedia.org/wiki/File:SAG.svg>

Sammenligning Frbr Versjon 1 Og 2. (2010). Sammenligning frbr versjon 1 og 2.. Hentet fra  
<http://www.bibpode.no/blogg/?p=1284>

Shotton, D., & Peroni, S. (2012). FaBiO, the FRBR-aligned Bibliographic Ontology. Hentet fra  
<http://purl.org/spar/fabio>

Smiraglia, R. P. (2002). The Progress of Theory in Knowledge Organization. *Library Trends*,  
50(3), 330-49.

Smiraglia, R. P. (2004). Authority control of works: Cataloging's chimera? *Cataloging &  
Classification Quarterly*, 38(3-4), 291-308.

Soja, E. W. (2011). Cities and states in geohistory. *Contention and Trust in Cities and States*,  
211-226.

SPARQL Query Language for RDF. (2008). SPARQL Query Language for RDF.. Hentet fra  
<http://www.w3.org/TR/rdf-sparql-query/>

Staikos, K., & Cullen, T. (2004). *The history of the library in western civilization. From minos to  
cleopatra* (1 ed., Vols. 5). New Castle, Del. USA; Goy-Houten, Netherlands; Athens: Oak  
Knoll Press ; HES & De Graaf Publishers ; Kotinos Publications.

Strout, R. F. (1956). The Development of the Catalog and Cataloging Codes. *The Library  
Quarterly*, 26(4), 254-275. doi:10.2307/4304573

Sumerian Civilization. (2008, Mai 7). I *New world encyclopedia*. Hentet 15 Apr. 2011 fra [http://www.newworldencyclopedia.org/entry/Sumerian\\_Civilization?oldid=703366](http://www.newworldencyclopedia.org/entry/Sumerian_Civilization?oldid=703366)

Svenonius, E. (2000). *The intellectual foundation of information organization*. Cambridge, Mass.: MIT Press.

Takhirov, N., Duchateau, F., & Aalberg, T. (2011). Linking FRBR Entities to LOD through Semantic Matching. In S. Gradmann, F. Borri, C. Meghini, & H. Schuldt (Eds.), *Lecture Notes in Computer Science: Vol. 6966. Research and advanced technology for digital libraries* (pp. 284-95). Springer Berlin / Heidelberg. Hentet fra [http://dx.doi.org/10.1007/978-3-642-24469-8\\_30](http://dx.doi.org/10.1007/978-3-642-24469-8_30)

Taylor, P. J. (2012). Extraordinary Cities: Early City-ness and the Origins of Agriculture and States. *International Journal of Urban and Regional Research*, 36(3), 415-447.  
doi:10.1111/j.1468-2427.2011.01101.x

Tchaikovsky Forum: Translation Into English of Tchaikovsky's Name. *Tchaikovsky forum*(udatert). Tchaikovsky Forum: Translation into English of Tchaikovsky's Name. Hentet 29 Mai. 2012 fra <http://www.tchaikovsky-research.net/en/forum/forum0059.html>

Tillett, B. (2001). Bibliographic Relationships. In C. A. Bean & R. Green (Eds.), *Relationships in the organization of knowledge* (pp. 19-35). Dordrecht; Boston; Norwell, MA: Kluwer Academic Publishers.

Tim.Hodson@talīs.com. (2011). British Library Terms RDF schema. Hentet fra <http://www.bl.uk/schemas/bibliographic/blterms-v1-2.rdf>

Topic Maps — Data Model. (2008). Topic Maps — Data Model.. Hentet fra <http://www.isotopicmaps.org/sam/sam-model/>

Turtle - Terse RDF Triple Language. (2011). Turtle - Terse RDF Triple Language.. Hentet fra <http://www.w3.org/TeamSubmission/turtle/>

Weinberger, D. (2005). The year of unique IDs. *JOHO - journal of the hyperlinked organization*. Hentet fra <http://www.hyperorg.com/backissues/joho-dec05-05.html#id>

Weitemeyer, M. (2009). Archive and Library Technique in Ancient Mesopotamia. *Libri*, 6(1-4), 217-238.

Westrum, A.. (2011a). Tre år med PODE. *Podens prosjektblogg*. Hentet 7 Mars. 2012 fra <http://www.bibpode.no/blogg/?p=1787>

Westrum, A.. (2011b). Tre år med PODE. *Podens prosjektblogg*. Hentet 7 Mars. 2012 fra <http://www.bibpode.no/blogg/?p=1787>

Westrum, A. L., Rekkavik, A., & Tallerås, K. (2012). Improving the presentation of library data using FRBR and Linked data. *Code4Lib Journal*, (16). Hentet fra <http://journal.code4lib.org/articles/6424>

Wilson, P. (1989). Interpreting the second objective of the catalog. *The Library Quarterly*, 339-353.

Yee, M. (2007). Cataloging Compared to Descriptive Bibliography, Abstracting and Indexing Services, and Metadata. *Cataloging & Classification Quarterly*, 44(3/4), 307-327.

Yee, M. M. (1994). What is a work? Part I, The User and the Objects of the Catalog. *Cataloging & Classification Quarterly*, 19(1).

Zaino, J. (2012). Richard Wallis Talks About New Role At OCLC. *Semanticweb.Com*. Hentet fra [http://semanticweb.com/richard-wallis-talks-about-new-role-at-oclc\\_b28358](http://semanticweb.com/richard-wallis-talks-about-new-role-at-oclc_b28358)

## Vedlegg 1: RDF og OWL-kode for ontologien

```
@prefix local-ontology: <http://are.com/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix frbr: <http://purl.org/vocab/frbr/core#> .
@prefix fabio: <http://purl.org/spar/fabio/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix prism: <http://prismstandard.org/namespaces/1.2/basic/> .
```

```
#### Work-class and related attributes and relations
#### -----
```

```
frbr:Work
    a          owl:Class .
```

```
## Work Attributes ##
```

```
dcterms:title
    a          owl:DatatypeProperty ;
    rdfs:domain
        [ a          owl:Class ;
          owl:unionOf (frbr:Work frbr:Manifestation)
        ] .
```

```
fabio:hasPublicationYear
    a          owl:DatatypeProperty ;
    rdfs:domain
        [ a          owl:Class ;
          owl:unionOf (frbr:Work frbr:Manifestation)
        ] ;
    rdfs:range xsd:int .
```

```
## Work relations ##
```

```
frbr:creator
    a          owl:ObjectProperty ;
    rdfs:domain frbr:Work ;
    rdfs:range  frbr:Person ;
    owl:inverseOf frbr:creatorOf .
```

```
fabio:hasManifestation
    a          owl:ObjectProperty ;
    rdfs:domain frbr:Work ;
    rdfs:range  frbr:Manifestation ;
    owl:inverseOf fabio:isManifestationOf .
```

```
##### Manifestation-class and related attributes and relations
##### -----
```

```
frbr:Manifestation
  a      owl:Class .
```

```
## dcterms:title shared with work
## fabio:hasPublicationYear shared with work
```

```
prism:isbn
  a      owl:DatatypeProperty ;
  rdfs:domain frbr:Manifestation .
```

```
fabio:isManifestationOf
  a      owl:ObjectProperty ;
  rdfs:domain frbr:Manifestation ;
  rdfs:range frbr:Work ;
  owl:inverseOf fabio:hasManifestation .
```

```
##### Person-class and related attributes and relations
##### -----
```

```
frbr:Person
  a      owl:Class .
```

```
foaf:givenName
  a      owl:DatatypeProperty ;
  rdfs:domain frbr:Person ;
  rdfs:range xsd:string .
```

```
foaf:familyName
  a      owl:DatatypeProperty ;
  rdfs:domain frbr:Person ;
  rdfs:range xsd:string .
```

```
foaf:name
  a      owl:DatatypeProperty ;
  rdfs:domain frbr:Person ;
  rdfs:range xsd:string .
```

```
local-ontology:invertedName
  a      owl:DatatypeProperty ;
  rdfs:subPropertyOf foaf:name .
```

```
frbr:creatorOf
  a      owl:ObjectProperty ;
  rdfs:domain frbr:Person ;
  rdfs:range frbr:Work ;
  owl:inverseOf frbr:creator .
```



## Vedlegg 2: RDF eksempel-kode for forfatter, verk og manifestasjoner

```
@prefix local-ontology: <http://are.com/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix frbr: <http://purl.org/vocab/frbr/core#> .
@prefix fabio: <http://purl.org/spar/fabio/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix prism: <http://prismstandard.org/namespaces/1.2/basic/> .

#### Person: Knut Hamsun ####

<http://viaf.org/viaf/97006051/>
  a frbr:Person ;
  foaf:givenName "Knut"^^xsd:string ;
  foaf:familyName "Hamsun"^^xsd:string ;
  foaf:name "Knut Hamsun"^^xsd:string .
  local-ontology:invertedName "Hamsun, Knut"^^xsd:string ;
  frbr:creatorOf local-ontology:id/work/00000001 ,
                local-ontology:id/work/00000002 ;

#### Verk: Sult ####

local-ontology:id/work/00000002
  a frbr:Work ;
  dcterms:title "Sult"^^xsd:string ;
  fabio:hasPublicationYear "1890"^^xsd:int ;
  fabio:hasManifestation
    local-ontology:id/manifestation/hamsun-sult-1944 ,
    local-ontology:id/manifestation/9788205394810 ,
    local-ontology:id/manifestation/8205277486 ,
    local-ontology:id/manifestation/8205248931 ;
  frbr:creator <http://viaf.org/viaf/97006051/> .

local-ontology:id/manifestation/hamsun-sult-1944
  a frbr:Manifestation ;
  dcterms:title "Sult"^^xsd:string ;
  fabio:hasPublicationYear "1944"^^xsd:int ;
  fabio:isManifestationOf local-ontology:id/work/00000002 .

local-ontology:id/manifestation/8205248931
  a frbr:Manifestation ;
  dcterms:title "Sult"^^xsd:string ;
  fabio:hasPublicationYear "1999"^^xsd:int ;
  prism:isbn "82-05-24893-1"^^xsd:string ;
  fabio:isManifestationOf local-ontology:id/work/00000002 .

local-ontology:id/manifestation/8205277486
  a frbr:Manifestation ;
  dcterms:title "Sult"^^xsd:string ;
  fabio:hasPublicationYear "2001"^^xsd:int ;
  prism:isbn "82-05-27748-6"^^xsd:string ;
  fabio:isManifestationOf local-ontology:id/work/00000002 .
```

```
local-ontology:id/manifestation/9788205394810
a      frbr:Manifestation ;
dcterms:title "Sult"^^xsd:string ;
fabio:hasPublicationYear "2009"^^xsd:int ;
prism:isbn "978-82-05-39481-0"^^xsd:string ;
fabio:isManifestationOf local-ontology:id/work/00000002 .
```

```
#### Verk: Sværmere ####
```

```
local-ontology:id/work/00000001
a      frbr:Work ;
dcterms:title "Sværmere"^^xsd:string ;
fabio:hasPublicationYear "1904"^^xsd:int ;
fabio:hasManifestation
    local-ontology:id/manifestation/8205210063 ,
    local-ontology:id/manifestation/hamsun-sværmere-1904 ,
    local-ontology:id/manifestation/8252534686 ,
    local-ontology:id/manifestation/8205257094 ;
frbr:creator <http://viaf.org/viaf/97006051/> .
```

```
local-ontology:id/manifestation/hamsun-sværmere-1904
a      frbr:Manifestation ;
dcterms:title "Sværmere"^^xsd:string ;
fabio:hasPublicationYear "1904"^^xsd:int ;
fabio:isManifestationOf local-ontology:id/work/00000001 .
```

```
local-ontology:manifestation/8205210063
a      frbr:Manifestation ;
dcterms:title "Sværmere"^^xsd:string ;
fabio:hasPublicationYear "1993"^^xsd:int ;
prism:isbn "82-05-21006-3"^^xsd:string ;
fabio:isManifestationOf local-ontology:id/work/00000001 .
```

```
local-ontology:id/manifestation/8252534686
a      frbr:Manifestation ;
dcterms:title "Sværmere"^^xsd:string ;
fabio:hasPublicationYear "1996"^^xsd:int ;
prism:isbn "82-525-3468-6"^^xsd:string ;
fabio:isManifestationOf local-ontology:id/work/00000001 .
```

```
local-ontology:id/manifestation/8205257094
a      frbr:Manifestation ;
dcterms:title "Sværmere"^^xsd:string ;
fabio:hasPublicationYear "2006"^^xsd:int ;
prism:isbn "82-05-25709-4"^^xsd:string ;
fabio:isManifestationOf local-ontology:id/work/00000001
```

## Vedlegg 3: Protégé prosjektfil og datafil

En zip-fil med to Protégé-filer er lastet opp elektronisk ved innlevering,

are.pprj er kodet i Protégés eget prosjektformat, her ligger bl.a. skjemadesignet.

are.owl inneholder dataene, dvs. ontologien og instansene.

Interesserte kan installere Protégé på vanlig måte, og laste inn hele prosjektet slik at de kan teste det som er gjort, f.eks. SPARQL-spørringene selv.

Jeg har brukt Protégé versjon 3.4.8 i prosjektet.

Vitenskapelig sett er etterprøvbarhet i eksperimenter et poeng, og da bør man jo kunne laste opp datagrunnlaget, som i mitt tilfelle er to filer på til sammen 200 KB.