

Nora Sand Haugerud

Konseptuelle Modeller for Bibliografiske Data

Masteroppgave i Informatikk
Veileder: Trond Aalberg
Februar 2019

Nora Sand Haugerud

Konseptuelle Modeller for Bibliografiske Data

Masteroppgave i Informatikk
Veileder: Trond Aalberg
Februar 2019

Norges teknisk-naturvitenskapelige universitet
Fakultet for informasjonsteknologi og elektroteknikk
Institutt for datateknologi og informatikk

Sammendrag

Bibliotekskataloger har utviklet seg fra enkle kortkataloger til maskinlesbare og etterhvert mer kompliserte kataloger organisert etter diverse konseptuelle modeller. En del av disse konseptuelle modellene har det blitt jobbet med i mange år, men det er fortsatt ikke utviklet gode løsninger for å ta disse i bruk i praksis. En velkjent modell er FRBR, som kom allerede på 1990-tallet, og baserer seg på fire abstraksjonsnivåer. I 2011 ble BIBFRAME lansert, en modell som er inspirert av FRBR, og er en slags forenkling. Den deler også opp i ulike abstraksjonsnivåer, men fanger ikke opp helt det samme. Disse modellene er laget for å løse en del utfordringer knyttet til bibliotekstjenester, og vil kunne gi både mer og bedre informasjon og en helt annen brukeropplevelse.

Målet med denne masteroppgaven er å utforske noen av disse modellene, der de mest sentrale er de to nevnt ovenfor, FRBR og BIBFRAME. Oppgaven vil gå inn på forskjellene, og hvordan bruken av de ulike modellene vil påvirke søk og brukeropplevelsen. Den første delen av oppgaven ser på det rent teoretiske om hvordan modellene er bygget opp. Videre er det gjort både en kvalitativ og kvantitativ studie som ser på viktige forskjeller mellom FRBR og BIBFRAME. Den siste delen viser hvordan en resultatpresentasjon vil kunne se ut i et brukergrensesnitt der dataene er organisert etter BIBFRAME, og hvordan brukeropplevelsen av en slik presentasjon er sammenlignet med et resultat der hver enkel bok presenteres som en egen post.

Abstract

Library catalogues have moved from basic catalogue cards to machine readable and later more complex catalogues organized by various conceptual models. Some of these conceptual models have been worked on over decades, but there is still some work left to make them useful in real systems. A well-known model is FRBR, which was introduced already in the 1990's, and is based on four levels of abstraction. In 2011 BIBFRAME was released, a model inspired by FRBR, and is a kind of simplification. BIBFRAME is also divided in different levels of abstraction but doesn't capture the exact same as FRBR. These models are made to solve some of the challenges that exist in library services and will give more and better information and a totally different user experience.

The objective in this master thesis is to explore some of these models, where the most central is the two mentioned above, FRBR and BIBFRAME. The thesis will look at differences and how the use of the different models affect search and user experience. The first part of the thesis presents the background for different models and their structure, followed by a qualitative and quantitative study, exploring the differences between FRBR and BIBFRAME. The last part is about presenting search results in a user interface where the data is organized according to BIBFRAME, and how the user experience of this kind of presentation is compared to a result where each book is presented a single record.

Forord

Dette arbeidet er min masteroppgave som avslutning på masterstudiet i informatikk ved Norges teknisk-naturvitenskapelige universitet (NTNU). Det har vært et tema jeg ikke hadde så mye kunnskap om fra tidligere, og mye nytt å sette seg inn i. Det har vært en prosess som jeg har lært mye av, både når det gjelder det faglige og det å gjennomføre en større studie.

Jeg vil rette en stor takk til veilederen min, Trond Aalberg, for all hjelp og gode tilbakemeldinger hele veien. Jeg ønsker også å takke familie og venner for den støtte og motivasjon de har gitt meg underveis.

Innhold

Sammendrag	i
Abstract	iii
Forord	v
Forkortelser	xiii
1 Introduksjon	1
1.1 Motivasjon	1
1.2 Forskningsspørsmål	2
1.3 Metode	2
1.4 Struktur	2
2 Bakgrunn	5
2.1 FRBR/LRM	5
2.1.1 LRM	6
2.2 BIBFRAME	7
2.2.1 Klasser og relasjoner	8
2.3 MARC	8
2.4 Schema.org	9
2.5 Wikidata	10

2.6	Lenkede Data	11
2.7	Modellering	11
2.8	Mapping	12
2.9	Implementasjon	12
3	Modellering	15
3.1	Kvalitativ Studie	15
3.1.1	FRBR	15
3.1.2	BIBFRAME	17
3.1.3	Samleverk	18
3.2	Kvantitativ Studie	18
3.2.1	Abstraksjonsnivåene	18
4	Brukergrensesnitt	21
4.1	Verk	21
4.2	Instanser	23
4.3	Relaterte verk	23
5	Konklusjon	25
5.1	Diskusjon	25
5.2	Konklusjon	25
5.3	Fremtidig Arbeid	26
	Bibliography	27
	Referanser for Figurer	29

Figurer

2.1	FRBR-modell - Entiteter Gruppe 1	6
2.2	BIBFRAME 1.0 Modell	7
2.3	BIBFRAME 2.0 Modell	7
2.4	Schema.org - Hierarki	9
2.5	Wikibase Datamodell	10
2.6	Wikidata Books	11
3.1	Sult - FRBR	16
3.2	Sult - BIBFRAME	17
4.1	Søkeresultat BIBFRAME - kollapset	22
4.2	Søkeresultat BIBFRAME	22
4.3	Instanser - BIBFRAME	23
4.4	Relaterte verk - BIBFRAME	24

Tabeller

3.1	Totalt i samlingen	18
3.2	Antall uttrykk, manifestasjoner og instanser av verk	19

Forkortelser

BIBFRAME	Bibliographic Framework
BIBSURF	Discover Bibliographic Entities by Searching for Units of Interest, Ranking and Filtering
FRBR	Functional Requirements for Bibliographic Records
IFLA	International Federation of Library Associations and Institutions
LRM	Library Reference Model
MARC	The Machine-Readable Cataloging
RDA	Resource Description and Access
RDF	Resource Description Framework
SPARQL	Query Language for RDF
URI	Uniform Resource Identifier
XML	Extensible Markup Language

KAPITTEL 1

Introduksjon

Dette kapitlet gir en introduksjon til oppgaven og begynner med å forklarer litt av bakgrunnen for denne studien. Deretter er det beskrevet hvordan arbeidet skal utføres og problemstillinger som det da vil bli fokusert på. Til slutt er det gitt en oversikt over strukturen for resten av oppgaven.

1.1 Motivasjon

Bibliotekataloger har utviklet seg fra enkle kortkataloger til maskinlesbare og etterhvert mer kompliserte kataloger organisert etter diverse konseptuelle modeller. Biblioteker begynte på 1960-tallet å lagre data om sine samlinger i MARC-format, som kunne leses av maskiner. Dette formatet har blitt brukt i flere tiår, men man har etter hvert sett et større behov for å kunne beskrive innholdet på ulike abstraksjonsnivåer og relasjoner mellom forskjellige verk, ikke bare vise informasjon om hvert enkelt eksemplar for seg. Dette blant annet på grunn av internett, som har gjort denne type informasjon mer tilgjengelig for flere brukergrupper, ikke bare biblioteksansatte, og også måten man tilegner seg informasjon på.

Man ønsker å gjøre det lettest mulig for brukeren å finne det han leter etter, noe som kan være vanskelig ved bruk av MARC-poster. Et eksempel på dette kan være hvis noen har lest en bok som er en del av en serie, og ønsker å se hvilke andre bøker som finnes i denne serien. Da er det lite nyttig å få listet opp alle mulige utgivelser av den samme boka. For å løse dette har det blitt utviklet flere konseptuelle modeller. FRBR ble lansert på slutten av 1990-tallet, og i 2011 gikk Library of Congress¹ og Zepheira² ut med sine planer om BIBFRAME. Det er hovedsakelig disse to modellene det vil bli fokusert på her.

¹<https://www.loc.gov>

²<https://zepheira.com/>

1.2 Forskningsspørsmål

Denne oppgaven vil fokusere på søk i bibliografiske data og ta for seg ulike konseptuelle modeller knyttet til dette. Det er da to punkter vi gjennom denne studien skal prøve å finne ut av:

1. Hvilke strukturelle forskjeller er det på de ulike modellene?
2. Hvilken effekt vil slike modeller ha for implementasjon av søkesystem og resultatpresentasjon?

1.3 Metode

Til å begynne med vil det bli gått nærmere inn på noen forskjellige konseptuelle modeller, bakgrunnen for disse og hvordan de er bygget opp, i tillegg til lenkede data som er relevant for flere av disse modellene. Videre vil det bli en gjennomgang av studier gjort av andre for å se hva som har blitt jobbet med tidligere på dette området, der sentrale temaer er modellering, mapping og implementasjon.

Den neste delen vil ta utgangspunkt i en samling av verk, for å undersøke hvilke relasjoner som finnes mellom forskjellige verk og utgivelser, og hvordan det vil se ut for de to modellene FRBR og BIBFRAME. Det vil her gjøres både en kvalitativ studie, hvor det blir gått gjennom strukturelle forskjeller som gir utslag for hvordan blant annet ulike typer verk kommer til uttrykk, og kvantitativ studie der det er brukt en større samling for å kunne få noen tall som viser hvordan strukturen til modellene påvirker antallet av ulike entiteter og relasjoner.

Den siste delen vil dreie som om presentasjon av dataene, og hvordan en slik resultatpresentasjon med BIBFRAME vil kunne se ut, hvordan den skiller seg fra en presentasjon der resultatene ikke er organisert etter abstraksjonsnivå, og ikke minst hva dette har å si for brukeropplevelsen.

1.4 Struktur

Her er en oversikt over hvordan resten av oppgaven er utformet.

- **Kapittel 2 - Bakgrunnsteori** Dette kapitlet gi et innblikk i bakgrunnen for denne oppgaven og nærmere forklare noen formater og modeller.
- **Kapittel 3 - Modellering** I dette kapitlet vil det bli gjort en analyse av modellene FRBR og BIBFRAME.

- **Kapittel 4 - Bruergrensesnitt** Dette kapitlet vil se på hvordan en resultatpresentasjon vil kunne se ut der BIBFRAME er brukt .
- **Kapittel 5 - Konklusjon** Avslutningskapitlet vil ha en diskusjon rundt det som så langt har blitt gått gjennom i denne oppgaven for så å komme med en konklusjon. Til slutt vil se på hva det kan jobbes videre med.

KAPITTEL 2

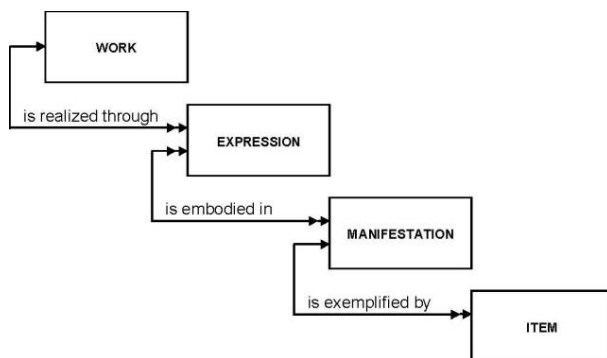
Bakgrunn

I dette kapitlet skal vi se nærmere på noen sentrale formater og modeller i biblioteksverdenen, deriblant BIBFRAME. Vi begynner med FRBR, som kan sees på som en forløper til den nyeste modellen, BIBFRAME, som er beskrevet i det neste avsnittet. Vi skal også litt innom MARC, som var det første formatet som ble brukt til katalogisering av bibliografiske data. Videre har vi Schema.org og Wikidata, og litt om lenkede data, som er en viktig del i disse modellene. Til slutt ser vi noe av det arbeidet andre har gjort, der sentrale temaer er modellering, mapping og implementasjon.

2.1 FRBR/LRM

FRBR, forkortelse for Functional Requirements for Bibliographic Records, er en konseptuell modell for å beskrive relasjoner mellom ulike entiteter, og ble utgitt av International Federation of Library Associations and Institutions¹ (IFLA) i 1998. Entitetene har blitt definert ut ifra hva som er av interesse for brukeren og er delt inn i tre forskjellige grupper (IFLA, 1998). Figur 2.1 viser den første entitetsgruppen som er de fire abstraksjonsnivåene i FRBR.

¹<https://www.ifla.org/>



Figur 2.1: FRBR-modell - Entiteter Gruppe 1

- **Verk:** Verket forteller kun om den abstrakte idéen, og hvem som står bak den. Et eksempel på et verk er Sult av Knut Hamsun. Skildringen om en person som forsøker å slå gjennom som forfatter, og hvordan opplevelsen av sult påvirker hans handlinger og tanker.
- **Uttrykk:** Verket kan være uttrykt på flere måter. Det kan være forskjellige språk og på ulike former, som for eksempel tekst eller lyd. Uttrykket til Sult var opprinnelig norsk tekst, men har senere blitt oversatt til andre språk og blitt utgitt blant annet i form av lydbok og film.
- **Manifestasjon:** Videre kan disse uttrykkene være utgitt i forskjellige formater, som innbundet eller elektronisk tekst, CD, DVD eller VHS. Den første manifestasjonen av Sult ble utgitt i København i 1890 av Philipsen i bokform.
- **Eksemplar:** Et eksemplar er en faktisk kopi. Altså kan et eksemplar være boka Sult av Knut Hamsun som du har i bokhylla di.

Gruppe 2 består av Person og Korporasjon, som er enten en individuell person eller en gruppe mennesker eller organisasjon. Disse kan ha ulike relasjoner, i form av roller, til et Verk, Uttrykk, Manifestasjon eller Eksemplar (Tillett, 2003). En person kan være forfatter av et verk eller en organisasjon kan være utgiver av en manifestasjon.

Gruppe 3 består av Begrep, Gjenstand, Hendelse og Sted, og beskriver emner for verk. Emnet til verket kan være relatert til alle de tre entitetsgruppene. For eksempel kan et verk ha et annet verk som emnet, eller det kan være en person, hendelse eller et sted (2003).

2.1.1 LRM

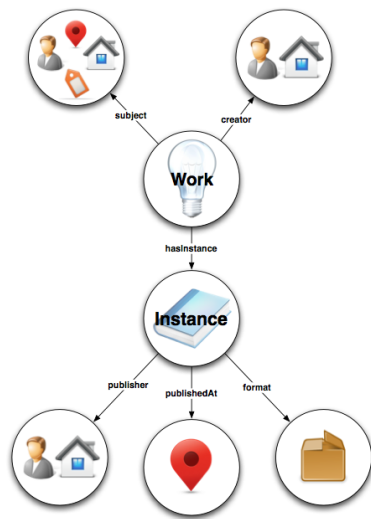
IFLA har ikke bare utviklet FRBR, men til sammen tre konseptuelle modeller, kjent som FR-familien av konseptuelle modeller. I tillegg til FRBR er det Functional Requirements

for Authority Data (FRAD) og Functional Requirements for Subject Authority Data (FRSAD). I 2016 ble det gitt ut en ny modell, IFLA Library Reference Model (LRM), som skulle samle de tre modellene i én. På grunn av inkonsistens mellom modellene var det ønskelig å lage en ny modell for å løse dette, og den nye modellen skulle også kunne brukes i et miljø med lenkede data (Pat Riva og Žumer, 2017).

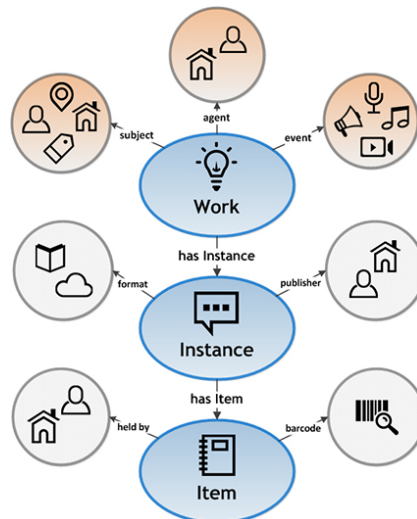
2.2 BIBFRAME

The Bibliographic Framework, kjent som BIBFRAME, er en datamodell for å beskrive bibliografiske data. Bibliographic Framework var et initiativ som for første gang ble introdusert i 2011 av Library of Congress, og var ment til å erstatte MARC-formatet som i lang tid hadde blitt sett på som utdatert. De ønsket å fokusere på web og ta i bruk prinsipper om lenkede data (Library of Congress, 2011). Året etter ble det første utkastet til dette rammeverket annonsert (Kroeger, 2013).

Figurene under viser Library of Congress sine illustrasjoner av modellen. Figur 2.2 viser den første versjonen av modellen, der kjerneklassene var "Verk", "Instans", "Autoritet" og "Kommentar". Figuren 2.3 til høyre viser den nyeste versjonen, BIBFRAME 2.0, som kom ut i 2016. Kjerneklassene her er "Verk", "Instans" og "Eksemplar" (Library of Congress, 2018).



Figur 2.2: BIBFRAME 1.0 Modell



Figur 2.3: BIBFRAME 2.0 Modell

På samme måte som i FRBR representerer disse kjerneklassene ulike abstraksjonsnivåer, der verk er det øverste nivået.

- **Verk:** I likehet med FRBR forteller Verk hva et bestemt arbeid handler om, den abstrakte idéen, og hvem som er ansvarlig for det. Men i tillegg til dette sier Verk i BIBFRAME hvordan det er formidlet. Om det for eksempel er tekst eller lyd, og på hvilket språk.
- **Instans:** En Instans av et Verk har en bestemt form. Et tekstlig verk kan være utgitt både som en fysisk bok og i elektronisk format. Instansen sier også hvem som er utgiver, og sted og dato for utgivelsen.
- **Eksemplar:** Eksemplar er det laveste abstraksjonsnivået og én enkel kopi av instansen. Eksemplaret har en strekkode og en lokasjon der man kan finne den.

Vi ser at både FRBR og BIBFRAME skiller på det abstrakte og konkrete, men de to modellenes måte å håndtere verk på er ikke helt det samme. I BIBFRAME ser vi at et verk har en instans som igjen har et eksemplar, altså er en instans bare en konkretisering av verket. FRBR-modellen derimot gjør det mulig å samle flere verk i én og samme manifestasjon. Ser vi på figur 2.1 igjen så er et verk realisert gjennom ett eller flere uttrykk, som er inneholdt i ett eller flere manifestasjoner, eksemplifisert ved ett eller flere eksemplarer. Et eksemplar er en eksemplifisering av én manifestasjon, men en manifestasjon kan inneholde opptil flere uttrykk som er en realisering av ett verk. Hvilken effekt denne forskjellen vil ha for ulike typer verk, for eksempel samleverk, skal vi se litt nærmere på i kapittel 3.

I tillegg til disse kjerneklassene har BIBFRAME tre nøkkelkonsepter: Agent, Emne og Hendelse (2018). Agent er personer eller organisasjoner og lignende på lik linje med entitetsgruppe 2 i FRBR-modellen. Emne og Hendelse beskriver mye av det samme som den tredje entitetsgruppen i FRBR.

2.2.1 Klasser og relasjoner

BIBFRAME-vokabularet består av klasser og egenskaper, der mange klasser er subklasser av de tre kjerneklassene beskrevet over. Egenskapene kan være relasjoner mellom klasser, eller attributter til klassene for å beskrive dem (2018). Vi har for eksempel egenskapene *hasInstance/instanceOf* og *hasItem/itemOf* som beskriver relasjonene mellom de tre kjerneklassene. Andre typer egenskaper er *hasExpression*, *partOf*, *responsibilityStatement* og *role*, som beskriver relasjoner mellom verk, instanser og agenter.

2.3 MARC

MARC, MACHine-Readable Cataloging, er formatet som blir brukt i det fleste bibliotekene i dag. Det ble utviklet på slutten av 1960-tallet av Henriette Avram, som jobbet for Library of Congress, for å gjøre bibliografiske dokumenter tilgjengelige for datamaskiner. Det nye MARC-formatet tok utgangspunkt i Dewey desimal-systemet som tidligere var brukt til katalogisering. Dette var et stort steg for bibliotekene fordi både tilgangen til

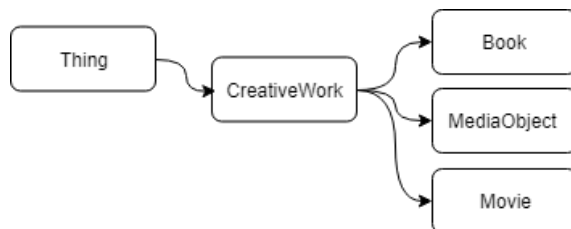
informasjon og deling bibliotekene seg imellom ble mye raskere og enklere (Schudel, 2006). MARC ble i løpet av kort tid en nasjonal standard i USA, og det tok ikke lang tid før det også ble en internasjonal standard. Etterhvert dukket det opp ulike versjoner av MARC-formatet, for eksempel NORMARC, som er den norske utgaven. På slutten av 1990-tallet kom MARC 21 ut og har blitt standarden som er mest brukt rundt om i verden.

Da internett dukket opp endret igjen behovet seg, og man trengte en ny standard tilpasset for web. Library of Congress kom opp med MARCXML, der MARC-postene ble gjort om til XML-skjema. Men det var mange som mente at hele MARC-formatet var utdatert og for begrenset for moderne behov, og at det var nødvendig med noe helt nytt. Problemet med MARC er blant annet at det er dårlig på å uttrykke relasjoner mellom ulike deler av informasjonen, og at det finnes en egen post for hver enkelt bok, noe som gjør at mange av postene inneholder samme informasjon. Noen som prøvde å løse dette var IFLA som utviklet FRBR.

2.4 Schema.org

Schema.org er et initiativ laget av Google, Yahoo, Bing og Yandex. Hensikten er at nettsider skal kunne bruke et felles vokabular til å strukturere dataene sine, for å hjelpe søkemotorer å forstå innholdet slik at søkeresultatet blir mer nøyaktig og brukerne lettere kan finne relevant informasjon (Schema.org, 2018a).

Datamodellen er utledet av RDF Schema og består av *Typer* og *Egenskaper*, der *Typene* er organisert i et hierarki, og er assosiert med et sett *Egenskaper* (2018b). I dette hierarkiet er alle entiteter en underkategori av *Ting*. Hovedklassene vi finner under *Ting* er Action, CreativeWork, Event, Intangible, MedicalEntity, Organization, Person, Place og Product, der CreativeWork er den klassen som kan brukes til bibliografisk data. Denne klassen inneholder informasjon om både tittel, forfatter, språk og format, og består ikke av flere abstraksjonsnivåer slik FRBR og BIBFRAME gjør. All informasjon er samlet under en og samme entitet.

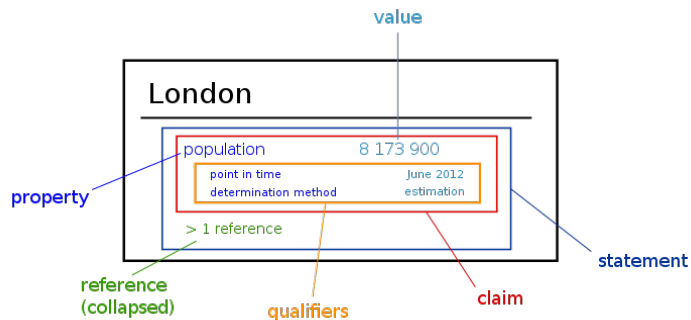


Figur 2.4: Schema.org - Hierarki

2.5 Wikidata

Wikidata er et Wikibase Repository, som er drevet av Wikimedia Foundation, og er en samling av *Entiteter* som det lagres forskjellig informasjon om. Disse dataene brukes både i Wikipedia, Wikiquote og andre Wikimedia-prosjekter, og er blant annet informasjonen man kan se i faktaboksene på Wikipedia.

Den konseptuelle modellen bak Wikibase baserer seg på semantiske tripler, noe vi kommer tilbake til i neste avsnitt 2.6 om lenkede data. Hvert element som beskrives er et *Item* og identifiseres ved hjelp av navn og en beskrivelse. Dataene som er knyttet til et *Item* representeres i form av *Statements*. Et *Statement* består av *Property*, *Value* og *Qualifiers* som tilsammen utgjør et *Claim*, i tillegg til en *Referanse* (MediaWiki, 2018).

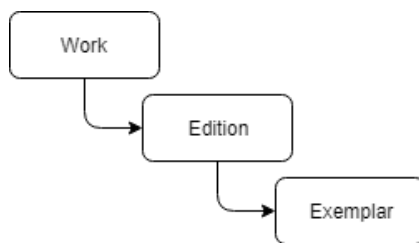


Figur 2.5: Wikibase Datamodel

Som vi kan se av figur 2.5 gir *Claimet* en beskrivelse av et *Item* ved at *Property* refererer til en *Value*. *Value* kan enten være en verdi eller et annet *Item*. Eksempel:

- Sult (Item) - utgitt (Property) - 1890 (Value)
- Sult (Item) - forfatter (Property) - Knut Hamsun (Value/Item)

For å beskrive bibliografisk informasjon, eller bøker, bruker Wikidata to entiteter, *Verk* og *Utgave*, der *Verk* er det abstrakte og *Utgave* er en utgivelse av et *Verk*. Verket inneholder informasjon som tittel, forfatter og språk, mens en *Utgave* gir informasjon blant annet om utgiver og oversetter. Wikidata beskriver også egenskaper for *Eksemplar*, bestemte kopier av bøker, som eier og lokasjon (Wikimedia, 2018).



Figur 2.6: Wikidata Books

2.6 Lenkede Data

BIBFRAME baserer seg på lenkede data for å kunne dele bibliografiske data og gjøre det tilgjengelig på nett. Lenkede data er en viktig del av den semantiske webben, og brukes for å lenke sammen data som det tidligere ikke har vært noen koblinger mellom slik at det også kan tolkes av maskiner. Vi har også lenkede åpne data. Åpne data vil si at dataene kan brukes og distribueres fritt av hvem som helst. (Ontotext, 2018).

Lenkede data bygger på fire prinsipper definert av Tim Berners-Lee i 2006.

1. Bruk URIer som navn for ting.
2. Bruk HTTP URIer slik at man kan slå opp disse navnene.
3. Når noen slår opp en URI, gi nyttig informasjon ved å bruke standardene (RDF*, SPARQL).
4. Inkluder lenker til andre URIer, slik at man kan oppdage flere ting.

All data som lagres får en URI, som er en unik identifikator. Denne URIen gjelder ikke bare for et enkelt datasett, men gjelder også globalt. På den måten har man mulighet til å lenke videre til andre datasett igjen og utnytte andres ressurser. Informasjonen lagres som RDF-tripler, bestående av subjekt, predikat og objekt. Subjektet er det som beskrives, predikatet forteller noe om subjektet og er en kobling mellom subjekt og objekt. Til slutt har vi objektet som er det det lenkes til. For eksempel:

- Sult (subjekt) - skrevetAv (predikat) - Knut Hamsun (objekt)

2.7 Modellering

En konseptuell modell har som formål å gi en forenklet beskrivelse av noe, som for eksempel FRBR som beskriver en måte å katalogisere dokumenter, uten å gå inn i detalj

på hvordan dette skal utføres i praksis. Utfordringen kan være at viktige punkter blir utelatt i det man forsøker å gjøre det enklest mulig å forstå teknologien eller prosessen som blir beskrevet. Det som er abstrakt kan være vanskelig å beskrive, noe jo både *Verk* og *Uttrykk* i FRBR-modellen er.

Alle de fire entitetene i FRBR-modellen, *Verk*, *Uttrykk*, *Manifestasjon* og *Eksemplar*, har vært brukt i katalogisering tidligere, men har ikke blitt beskrevet eksplisitt som fire entiteter slik FRBR gjør. Helt til å begynne med var det én-entitetsmodell der alt som beskrives er et enkelt eksemplar. Etterhvert som årene har gått og samlingene har blitt større har det blitt både to-, tre- og fire-entitetsmodeller, som beskriver ulike utgivelser, skiller på verk og til slutt har det blitt lagt til en entitet som er en form av det vi i FRBR kaller *Uttrykk* (Carlyle, 2006).

2.8 Mapping

Siden all bibliografisk data har vært lagret i MARC-format er det nødvendig å konvertere disse dataene slik at de kan brukes til nyere modeller som FRBR og BIBFRAME. Library of Congress har på sine sider lagt ut spesifikasjoner og verktøy for konvertering av MARC-poster til BIBFRAME. I 2017 var det en gruppe som gikk sammen for å gå gjennom disse og se hvor gode disse egentlig er og om det er noen problemer (Xu, Hess og Akerman, 2017).

Når det kommer til mapping mellom forskjellige konseptuelle modeller er det én viktig ting det stort sett handler om, og det er å bevare relasjoner (Zapounidou, Sfakakis og Papatheodorou, 2017).

BIBFRAME er som nevnt tidligere en modell som baserer seg på lenkede data. Forskere fra Online Computer Library Center (OCLC) har forsøkt å gjøre Schema.org egnet for å beskrive bibliografiske data, og mener det er vanskelig å komme utenom nettopp på grunn av at det er de største søkemotorene som står bak Schema.org, for å gjøre strukturerte data tilgjengelig på nett og gjenkjennbart i søk. Problemet med Schema.org er at det ikke beskriver de ulike abstraksjonsnivåene, noe som er en veldig viktig del både i FBRR og BIBFRAME. Det samme gjelder relasjoner, som ikke er definert godt nok (Godby, 2013).

2.9 Implementasjon

Hovedutfordringen når FRBR-modellen skal implementeres er å avgjøre hvilken gruppe et element skal tilhøre. Som nevnt tidligere er den konseptuelle modellen en forenklet beskrivelse. Når den skal implementeres derimot, er det nødvendig at disse beskrivelsene er mer presise, noe som kan tolkes ulikt og gi forskjellige implementasjoner av den samme modellen. For eksempel, hva skal regnes som et nytt verk? Da er det viktig å tenke på hva brukeren mest sannsynlig er ute etter og skape minst mulig forvirring. Carlyle, 2006 nevner blant annet en utgave med punktskrift, der denne notasjonen er den samme som

de alfabetiske bokstavene og kan derfor sees på som en manifestasjon av et eksisterende uttrykk. Men det vil nok være mest hensiktsmessig for brukeren å regne dette som et eget uttrykk, en ny realisering av verket.

Resource Description and Access, forkortet RDA, er en implementasjon som tar utgangspunkt i modellene FRBR OG FRAD. Noen deler stemmer veldig godt overens med modellene, mens andre deler avviker i større eller mindre grad (Riva og Oliver, 2012).

KAPITTEL 3

Modellering

Vi har nå sett på bakgrunnen for modellene FRBR og BIBFRAME og hvordan de er bygget opp. I dette kapittelet ønsker vi å ta for oss en samling av verk for å få et bedre innblikk i hvordan dette faktisk vil fungere. Den første delen er en kvalitativ studie, mens den andre delen er en kvantitativ studie.

3.1 Kvalitativ Studie

I denne kvalitative studien ønsker vi å undersøke hvordan de to modellene håndterer forskjellige typer verk, og hvilke konsekvenser det fører med seg. Vi har brukt verket “Sult” av Knut Hamsun som eksempel for modellene som er laget her. Det er da originalverket fra 1890, som har blitt utgitt på ulike formater og oversatt til flere andre språk, reviderte utgivelser, samleverk som “Sult” er en del av sammen med andre verk, filmadaptasjoner, og verk som handler om “Sult”.

3.1.1 FRBR

Figur 3.1 viser tre ulike verk, der to av dem har én uttrykksform hver og alle har en relasjon til verket “Sult” av Knut Hamsun. Dette verket er uttrykt på flere ulike språk både gjennom tekst, lyd og levende bilde.

De blå feltene viser *Verk*, de røde *Uttrykk* og de grønne er *Manifestasjoner*.



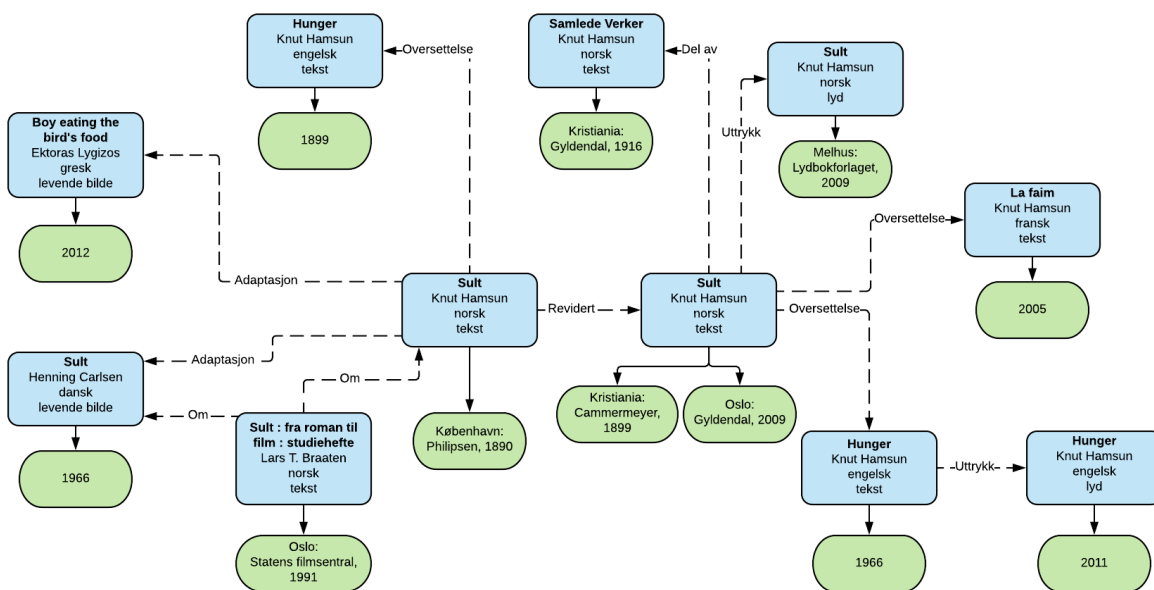
Figur 3.1: Sult - FRBR

Av modellen ser vi at det ikke er så veldig mye på verksnivå. Verket “Sult” av Knut Hamsun er veldig kjent og har eksistert i lang tid, noe som har gjort at det er blitt oversatt til en rekke ulike språk og blitt fortalt gjennom flere formater, og gir oss mange uttrykk i FRBR-modellen. Som vi ser er det en del forskjellig på uttrykksnivå for nettopp dette verket. De tre andre verkene som er lagt til i modellen derimot har bare én uttrykksform. “Sult : fra roman til film : studiehefte” er ikke et verk som prøver å formidle handlingen i “Sult”, men beskriver noe om det, og har derfor en relasjon til dette verket. Videre har vi to filmer, som jo er et slags uttrykk der den samme handlingen som er i det opprinnelige verket formidles. Men samtidig er det skapt noe nytt, og bør derfor defineres som egne verk. En film er bare basert på, ikke en direkte overføring, og de som lager den tar seg derfor gjerne en del friheter. Spesielt om vi ser på filmen “Boy eating the bird’s food” som til og med er lagt til moderne tid, så skiller den seg i ganske stor grad fra det originale verket av Hamsun.

3.1.2 BIBFRAME

Figur 3.2 under viser hvordan det samme eksempelet vil se ut med BIBFRAME. Man kan da se at det blir ganske mange flere verk, fordi det øverste nivået, altså verk, også sier noe om hvordan det er uttrykt. Til sammen er det her 11 verk, men hvert av disse har ikke så mange instanser.

De blå feltene er *Verk*, mens de grønne er *Instanser* av verket.



Figur 3.2: Sult - BIBFRAME

Ut ifra denne modellen er det ikke like lett å se hvor det hele egentlig starter, med mindre man ser på årstallet på instansnivå. Hamsun ga først ut "Sult" i 1890, som er det verket vi ser i midten til venstre i modellen. Rett til høyre ser vi også et verk "Sult" av Hamsun som norsk tekst. Senere gjorde han nemlig en del endringer i teksten, og det er stort sett den nyere utgivelsen lydinnspillinger og oversettelser har tatt utgangspunkt i. Siden det ofte blir tydelig markert om en utgivelse er fra den opprinnelige språkdrakten kan det være greit å skille på disse to verkene. Filmatiseringer er ikke en like direkte tolkning som oversatte verk, og kan derfor like gjerne være en adaptasjon av det første som det andre verket her, men de er her knyttet til det opprinnelige verket, der idéen først kom fra. Med BIBFRAME er det fort gjort å ende opp med flere verk som har samme tittel, forfatter, uttrykksform og språk, noe vi får et eksempel på her.

3.1.3 Samleverk

Noe man ofte kan komme borti når man har å gjøre med bibliografiske samlinger er *samleverk*, utgivelser bestående av flere verk. Om vi ser på den første modellen vår 3.1, er ikke dette definert som et eget verk i FRBR, nettopp fordi det ikke er noe nytt som ikke finnes fra før, bare eksakte kopier av allerede eksisterende verk. Derfor vil slike samleverk heller være manifestasjoner, som inneholder flere verk, med relasjoner til hvert av de verkene som den aktuelle utgivelsen måtte inneholde.

I motsetning til FRBR vil samleutgivelser i BIBFRAME regnes som et eget verk, i og med at verket i denne modellen også innebærer hvordan det er uttrykt. I stedet for å ha en instans som kan inneholde flere verk, slik en manifestasjon i FRBR gjør, blir det opprettet et nytt verk. Instansen er bare en konkretisering av verket, i form av en bok, elektronisk tekst og lignende.

3.2 Kvantitativ Studie

Denne andre delen av studien vil se på hvordan strukturen i de to modellene påvirker antallet av forskjellige entiteter og relasjoner. Det er her brukt et større sett, med 1500 MARC-poster, som er noe bearbeidet og konvertert til LRM og BIBFRAME. Denne samlingen inneholder verk av Hamsun, som er hentet fra fire forskjellige nasjonalbibliotek: det norsk (Nasjonalbiblioteket), tyske (Deutsche Nationalbibliothek), britiske (British Library) og spanske (Biblioteca Nacional de España). For å utføre dette arbeidet har vi brukt GraphDB, en semantisk grafdatabase, av Ontotext til å utforske dataene.

3.2.1 Abstraksjonsnivåene

Vi begynner med å se på abstraksjonsnivåene, som for FRBR er *verk*, *uttrykk* og *manifestasjoner*, og for BIBFRAME er *verk* og *instanser*. Tabell 3.1 viser de totale antallene vi har i datasettet. Vi ser at det er et betydelig mindre antall verk for dataene som er konvertert til LRM sammenlignet med BIBFRAME, og gjenspeiler det som vises i modellene fra den kvalitative studien der det også kom tydelig fram selv med en mye mindre samling.

	Verk	Uttrykk	Manifestasjoner	Instanser
LRM	169	368	1702	
BIBFRAME	2251			2690

Tabell 3.1: Totalt i samlingen

	Uttrykk	Manifestasjoner	Instanser
Sult	10	149	51
Victoria	12	100	30
Pan	11	116	31
Markens grøde	8	102	16
Mysterier	7	68	28
Paa gjengrodde stier	7	35	25
Børn av tiden	4	21	16
Sværmere	8	35	22
Den siste glæde	7	27	13
Brigantins saga	1	1	2

Tabell 3.2: Antall uttrykk, manifestasjoner og instanser av verk

For å undersøke sammenhengen mellom verk og de andre abstraksjonsnivåene har vi valgt ut ti forskjellige verk fra samlingen, og gjort noen SPARQL-spøringer i GraphDB på disse. I tabell 3.2 kan vi se hvor mange uttrykk, manifestasjoner og instanser som finnes for hvert av verkene.

Som nevnt tidligere er verk i FRBR kun den abstrakte idéen, mens det i BIBFRAME også forteller hvordan verket er uttrykt. Derfor har vi når vi her teller instanser for BIBFRAME, flere verk som er knyttet til det samme verket ved hjelp av egenskapene *hasPart* og *translationOf*. Tabellen viser det totale antallet instanser for verket, men dersom vi går inn på hvert enkelt så er stort sett alle verkene vi har sett på her relatert til mellom én og fire instanser hver.

Brukergrensesnitt

I dette kapitlet skal vi se på hvordan en presentasjon av søkeresultatet kan se ut i et brukergrensesnitt ved bruk av BIBFRAME. Eksempelene som blir vist her tar utgangspunkt i den samme samlingen som er brukt for modellene 3.1 og 3.2 i forrige kapittel, et lite utvalg av verk knyttet til “Sult” av Knut Hamsun. Designet er det samme som er brukt for BIBSURF¹, et grensesnitt utviklet for å eksperimentere med søk i bibliografiske data organisert etter FRBR/LRM.

4.1 Verk

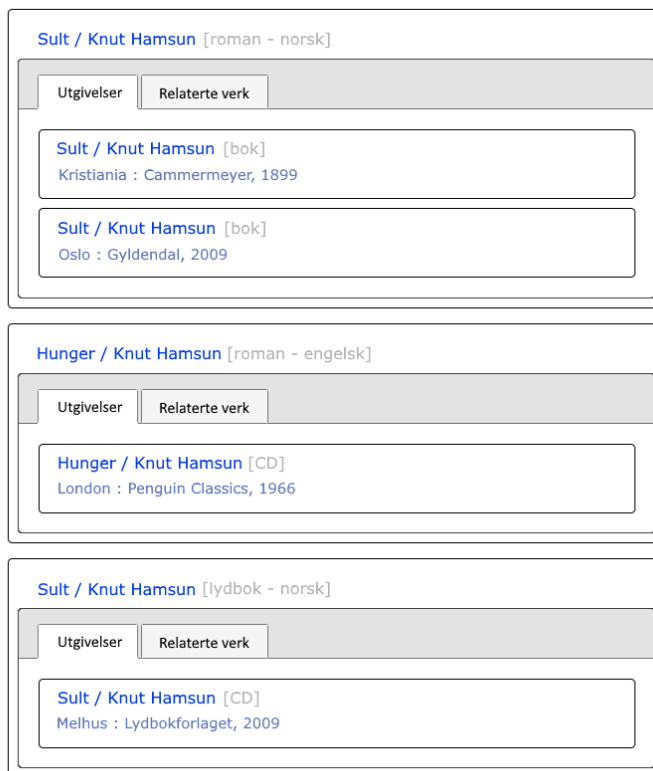
Figur 4.2 viser et eksempel på et søkeresultat der dataene er organisert etter BIBFRAME-modellen. Resultatet viser tre verk der en fane med instanser er åpen. Når disse er kol-lapset får vi et ganske kort og kompakt resultat sammenlignet med et mer tradisjonelt søk i data som ikke er lagret ut ifra abstraksjonsnivå. Som vi ser får vi oppgitt tittel og forfatter, eller den ansvarlige for verket, samt uttrykksform. Hver av resultatpostene har i tillegg to faner, der vi får listet opp instanser av verket og andre relaterte verk.

Brukeren får altså utgivelsene av det samme verket samlet på ett sted, og lettere oversikt over hvilke verk som faktisk finnes. I et søkeresultat der alle bøker er listet opp enkeltvis kan man ende opp med en ganske lang liste fordi et verk ofte har blitt utgitt flere ganger, og det kan være vanskelig for brukeren å orientere og oppleves litt overveldende.

¹<http://dijon.idi.ntnu.no/bibsurf/>



Figur 4.1: Søkeresultat BIBFRAME - kollapset



Figur 4.2: Søkeresultat BIBFRAME

4.2 Instanser

For hvert av verkene ser vi en liste med utgivelser. Dette er instanser av verket. Instansen oppgir tittel, forfatter og i hvilket format det er tilgjengelig, om det for eksempel er en bok, elektronisk fil eller DVD. I tillegg er det oppgitt informasjon om publisering av utgivelsen, og det skal være mulig å klikke på instansen for ytterligere informasjon.

Det å presentere disse utgivelsene under ett og samme verk gjør det mye enklere for brukeren å finne fram til det han er ute etter og utforske ulike alternativer. For lydbøker kan det være interessant å vite hvilke fortellere man har å velge blant, men veldig ofte spiller det ingen rolle hvilken utgivelse det er, og det å da få samlet dem slik gjør at brukeren lett ser hvilke utgivelser som faktisk er av det verket han er ute etter.



Figur 4.3: Instanser - BIBFRAME

4.3 Relaterte verk

I søkeresultatet ser vi at hvert verk har to faner, der den første viser en liste over utgivelser og den andre fanen viser andre verk som har en relasjon til dette verket. Dette gir brukeren mulighet til å oppdage nye verk og navigere seg videre. Ved å velge et verk kan man se hvilke utgivelser og relaterte verk som finnes for dette igjen. I et vanlig søkeresultat får man ikke opp hvilke verk som er relatert til hverandre, og man går glipp av verdifull informasjon. Eksempler på dette er bokserier der man kanskje har lest én bok, men vet ikke tittelen på de andre bøkene i serien. Da er det veldig nyttig å bare kunne finne fram til denne boka og få listet opp alle de andre bøkene i serien under relaterte verk. Et annet tilfelle er at man kanskje ikke en gang er klar over at det finnes flere bøker, og kan ved hjelp av slike relasjoner oppdage at det faktisk er en oppfølger. Noe flere kanskje har vært borti er å ha lest en bok man har likt veldig godt, og ønsker å finne flere av samme forfatteren. Dette er bare noen eksempler på hvordan det å utnytte relasjoner mellom

verk kan være til stor hjelp.



Figur 4.4: Relaterte verk - BIBFRAME

Konklusjon

Dette kapitlet vil diskutere noen av observasjonene som har blitt gjort gjennom denne studien, og se på hva vi kan trekke ut av dette. Helt til slutt ser vi på hva det vil kunne jobbes med videre fremover.

5.1 Diskusjon

Gjennom denne studien har vi gått gjennom noen konseptuelle modeller og ser at det er en del strukturelle forskjeller mellom dem. BIBFRAME er inspirert av FRBR, og de har derfor noen likeheter. Begge deler opp i ulike abstraksjonsnivåer slik at man kan skille mellom det abstrakte og konkrete, men måten dette håndteres på er ikke helt den samme. I BIBFRAME har man et verk som konkretiseres gjennom en instans, mens i FRBR er det manifestasjoner som har verk som innholde, og én manifestasjon kan inneholde ett eller flere verk. I tillegg er verksnivået i BIBFRAME mye større, og inneholder mer informasjon, noe som gjør at vi får mye flere å forholde oss til. I stedet for at det sees på som flere realiseringer av et verk blir det opprettet egne verk, som heller er knyttet sammen med ulike relasjoner. Vi har også vært innom Schema.org og Wikidata, der Schema.org er veldig hierarkisk, mens Wikidata igjen er delt opp i abstraksjonsnivåer.

5.2 Konklusjon

Vi har sett på en resultatpresentasjon for BIBFRAME, og denne viser hvordan det å dele opp i abstraksjonsnivåer gjør det lettere for brukeren å finne fram og skaper flere muligheter for å utforske det som finnes i samlingen. Likevel har denne modellen noen

svakheter som går på dette med verk. Det er foreløpig bare FRBR som ordentlig klarer å fange opp det, noe vi blant annet ser i forbindelse med samleverk. Uansett er det en klar fordel i det å skille konkrete utgivelser fra abstrakte idéer.

5.3 Fremtidig Arbeid

Videre vil det være nødvendig å jobbe med implementasjon og hvordan disse modellene på best mulig måte kan tas i bruk her.

Bibliografi

- Carlyle, A. (2006). Understanding FRBR as a conceptual model: FRBR and the bibliographic universe. *Library Resources Technical Services*, 50(4), 264–273. Date revised - 2007-05-10; Last updated - 2016-09-27; SubjectsTermNotLitGenreText - Cataloguing; Functional Requirements for Bibliographic Records. Hentet fra <https://search.proquest.com/docview/57629621?accountid=12870>
- Godby, C. J. (2013). The relationship between BIBFRAME and OCLC's linked-data model of bibliographic description: a working paper. OCLC Research.
- IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). Functional Requirements for Bibliographic Records : final report. *UBCIM publications*, 19.
- Kroeger, A. (2013). The Road to BIBFRAME: The Evolution of the Idea of Bibliographic Transition into a Post-MARC Future. *Cataloging & Classification Quarterly*, 51(8), 873–890. doi:10.1080/01639374.2013.823584
- Library of Congress. (2011). Library of Congress: Bibliographic Framework Initiative General Plan. <http://www.loc.gov/bibframe/news/framework-103111.html>. Aksessert 20. november 2018.
- Library of Congress. (2018). Overview of the BIBFRAME 2.0 Model. <https://www.loc.gov/bibframe>. Aksessert 16. september 2018.
- MediaWiki. (2018). Wikibase/DataModel. <https://www.mediawiki.org/w/index.php?title=Wikibase/DataModel&oldid=3006145>. Aksessert 16. januar 2019.
- Ontotext. (2018). What are Linked Data and Linked Open Data? <https://www.ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/>. Aksessert 10. januar 2019.
- Pat Riva, P. L. B. & Žumer, M. (2017). IFLA Library Reference Model - A Conceptual Model for Bibliographic Information.
- Riva, P. & Oliver, C. (2012). Evaluation of RDA as an Implementation of FRBR and FRAD. *Cataloging & Classification Quarterly*, 50(5-7), 564–586. doi:10.1080/01639374.2012.680848. eprint: <https://doi.org/10.1080/01639374.2012.680848>

- Schema.org. (2018a). About Schema.org. <https://schema.org/docs/faq.html>. Aksessert 14. januar 2019.
- Schema.org. (2018b). Data Model. <https://schema.org/docs/datamodel.html>. Aksessert 14. januar 2019.
- Schudel, M. (2006). Henriette Avram, 'Mother of MARC,' Dies. <https://www.loc.gov/loc/lcib/0605/avram.html>. Aksessert 20. november 2018.
- Tillett, B. (2003). What is FRBR? A Conceptual Model for the Bibliographic Universe. *Technicalities*, 25(5). Hentet fra <https://www.loc.gov/cds/downloads/FRBR.PDF>
- Wikimedia. (2018). Data Model. https://www.wikidata.org/wiki/Wikidata:WikiProject_Books. Aksessert 29. januar 2019.
- Xu, A., Hess, K. & Akerman, L. (2017). From MARC to BIBFRAME 2.0: Crosswalks. *Cataloging & Classification Quarterly*, 56(2-3), 224–250. doi:10.1080/01639374.2017.1388326. eprint: <https://doi.org/10.1080/01639374.2017.1388326>
- Zapounidou, S., Sfakakis, M. & Papatheodorou, C. (2017). Preserving Bibliographic Relationships in Mappings from FRBR to BIBFRAME 2.0, 15–26.

Referanser for Figurer

- **Figur 2.1:** https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- **Figur 2.2:** <http://www.loc.gov/bibframe/docs/images/bibframe.png>
- **Figur 2.3:** <https://www.loc.gov/bibframe/docs/bibframe2-model.html>
- **Figur 2.5:** https://commons.wikimedia.org/wiki/File:Wikidata_statement.svg

