

Integrasjon og bruk av gazetteers og tesauri i digitale bibliotek
Søk og gjenfinning via geografisk referert informasjon

Hovedfagsoppgave ved
Institutt for Datateknikk og Informasjonsvitenskap
Norges teknisk-naturvitenskapelige universitet

Marit Olsen
Juni 2004

Forord

Denne avhandlingen gjenspeiler mitt hovedfagsarbeid ved Institutt for Datateknikk og Informasjonsvitenskap, NTNU.

Fagområdet informasjonsforvaltning i digitale bibliotek har forgreininger til mange komplekse problemområder og problemstillingen i denne oppgaven berører flere av disse.

Hovedtema for avhandlingen er gazetteers og tesauri og hvordan strukturer som disse kan integreres og brukes for å støtte gjenfinning av dokumenter i digitale bibliotek. Gazetteers og navnetypesaurus er begreper som brukes til stadighet i denne avhandlingen. Digitale gazetteers var helt nytt for meg, og navnetyper var for meg et ukjent begrep da jeg startet dette arbeidet.

Et uttrykk sier at *veien blir til mens man går* og det kan man trygt si om veien hit også. Det har tidvis vært vanskelig å velge ståsted og sette grenser for veien videre. Min hang til å gruble over detaljer har ikke akkurat gjort det enklere.

Jeg har lært utrolig mye underveis, både faglig og ikke minst av selve prosessen som har ført meg hit.

Professor Ingeborg T. Sølvberg skal ha mange slags takk. Hun har vært min faglige veileder og har gitt meg uvurderlig input gjennom denne prosessen. Hun har et stort hjerte både for faget og for sine studenter. Tusen takk!

Flere har bidratt til at jeg har kommet i mål. Familie og venner har vært tålmodige og mine nærmeste medstudenter har vært gode inspirasjonskilder. Takk skal dere ha!

Jeg må også sende en takk til Jeanine Lilleng. Da jeg slet som verst støttet hun meg til å ta et veivalg som viste seg å være både riktig og viktig. Takk for det!

Trondheim 25. Juni, 2004

Marit Olsen

Sammendrag

Geografiske navn og navnetyper er begreper og termer som de fleste av oss kan forholde oss til og forstå. En god del av den informasjon som produseres i verden er georeferert på en eller annen måte.

Å benytte geografiske navn som søketermer ved søk etter informasjonsobjekter er ikke uvanlig, men gjenfinningsmuligheter begrenses ofte av at en er avhengig av eksakt match mellom søketermer og registrerte søkbare metadata i databaser.

Konseptbaserte motivbeskrivelser av fotografier er komplisert og arbeidskrevende. Mye av arbeidet gjøres ofte manuelt og innebærer en fare for at indekseringen påvirkes av indekserers subjektive oppfatning av motivet som skal indekseres. Manuell indeksering er tidkrevende og flere personer kan være involvert i indeksering for en og samme samling. Detaljeringsgraden i indeksering kan variere og i tillegg kan mangel på vokabularkontrollerende virkemidler medføre inkonsistens i indekseringen. Å søke etter og finne relevante fotografier kan, for mange samlinger, være svært vanskelig dersom bruker ikke vet noe om bakgrunnen for indekseringen og hvilke begreper indekserer forholder seg til. Vi kan ha optimale databaseløsninger og søkesystemer, men oppleve å allikevel komme til kort i søkeprosessen som følge av inkonsistens i indeksering.

Med tilgang til enorme mengder av informasjon samt variasjon i indekseringspraksis for ulike systemer og samlinger står vi overfor store utfordringer på dette området.

Beskrivende metadata er kritisk for å skape gode forhold for konseptbasert gjenfinning, men beskrivende metadata er ikke nok. Vokabularkontroll er et nøkkelord.

I denne avhandlingen foreslår jeg et system som integrerer en gazetteer i et gjenfinningssystem med det mål å produsere søkeresultater som tilsynelatende bærer preg av at ressursene er beskrevet enhetlig og strukturert selv om dette ikke er tilfelle.

Viktige prinsipper er testet ved å implementere et system med en konkret søkedel. En gazetteer som inkluderer en navnetypesaurus er utviklet og implementert integrert i et system med bilder fra Galleri Nor. Galleri Nor er et digitalt fotogalleri som administreres av Nasjonalbiblioteket Rana.

Gazetteeren som er utviklet for dette prosjektet inkluderer en del av en navnetypesaurus. Gazetteeren er basert på ADL Gazetteer Content Standard (ADL GCS). ADL GCS er utviklet ved University of California, Santa Barbara (UCSB). Kjerneelementer for en forekomst i gazetteeren er geografisk navn, navnetype og lokasjonsinformasjon. En navnetype sier noe om et steds natur. Jeg foreslår et navnetypeskjema som benyttes for tilordning av navnetype til en forekomst. Lokasjonsinformasjon er basert på relasjoner mellom forekomster.

Ved spørring etter stedsnavn og /eller navnetype gjøres automatisk oppslag i gazetteeren, og hierarkiske strukturer utnyttes for automatiske queryutvidelser. En spørring etter *Trondheim by* vil gi treff også på ressurser som er beskrevet med *Ila* eller med *Bispegata* som spesifiserte stedsnavn.

For å kontrollere vokabularet ved spørring etter navnetyper må bruker velge navnetype fra en hierarkisk indeks i brukergrensesnittet. Som støtte ved valg av navnetype er en navnetypesaurus tilgjengelig via en lenke fra søkesiden.

Implementasjonen er basert på Java-teknologi og MySQL relasjonsdatabaser.

Implementert funksjonalitet for systemet er begrenset til å kunne søke på stedsnavn, og

navnetyper. Systemet er testet og sammenlignet med Galleri Nor. Testresultatene er oppmuntrende med tanke både på recall og presisjon.

Avhandlingen beskriver noen etablerte metadataformater og flere systemer og samlinger. Det er stor variasjon i indekseringspraksis.

Som et ledd i tilnærmingen til Galleri Nor vurderer jeg muligheten for å bruke Dublin Core (DC) som utvekslingsformat for beskrivende metadata for Galleri Nor. Med tanke på bildene i Galleri Nor diskuterer jeg elementene i DC. Overraskende mange problemstillinger dukket opp underveis i diskusjonen selv om jeg gjorde et forsøk på å begrense denne. Konklusjonen på denne diskusjonen er at DC kan brukes som utvekslingsformat selv om det ikke er uproblematisk. Jeg legger fram forslag til metadataskjema på samlingsnivå og på ressursnivå for Galleri Nor, basert på DC.

Selv om man bestreber seg på å løse problemer forbundet med inkonsistens i beskrivende metadata kan jeg vanskelig se for meg en strategi som løser alle problemene.

En konklusjon på mitt arbeid er at strukturer som gazetteers og tesauri i tillegg til å kunne fungere som selvstendige systemer for kunnskapsorganisering, har egenskaper som kan utnyttes til å støtte andre arbeidsprosesser i digitale bibliotek. Integret i et gjenfinningssystem kan selv forholdsvis enkle utgaver av slike strukturer utnyttes til å forbedre gjenfinningsmulighetene.

Et forslag til videreføring av dette prosjektet er å videreutvikle de ulike komponentene som systemet består av slik at de kan fungere som selvstendige enheter i tillegg til at de kan være dypt integrert i et gjenfinningssystem. Komponentene bør hver for seg ha klart definerte grensesnitt mot andre komponenter. Et godt utgangspunkt kan være å definere tjenesteprotokoller for de ulike komponentene. En naturlig utvidelse vil være å inkludere flere og større samlinger. Det må også være mulig å inkludere andre typer samlinger som tekstdokumenter og lydsamlinger. I et operasjonelt system kan det være nødvendig å revurdere noen av de beslutninger som er tatt for prototypen som er beskrevet i avhandlingen.

Innholdsoversikt

FORORD	3
SAMMENDRAG	5
INNHALDSOVERSIKT	7
TABELLOVERSIKT	11
FIGUROVERSIKT	12
1. INTRODUKSJON TIL OPPGAVEN	13
1.1. OPPGAVEFORMULERING.....	13
1.2. PROBLEMBESKRIVELSE.....	13
1.3. MÅL FOR OPPGAVEN.....	14
1.4. DELMÅL.....	14
1.5. AVGRENSING AV OPPGAVEN.....	14
2. DIGITALE BIBLIOTEK	15
GEOGRAFISK REFERERT INFORMASJON I DIGITALE BIBLIOTEK.....	19
3. SØK OG GJENFINNING	21
3.1. INFORMASJONSSØKING.....	21
3.2. SØKEPROSESSEN.....	21
3.3. GJENFINNING.....	24
3.3.1. <i>Recall</i>	24
3.3.2. <i>Presisjon</i>	24
3.3.3. <i>Relevans</i>	24
3.4. BRUKERUNDERSØKELSER.....	24
3.5. SØKESYSTEMER.....	25
3.6. BILDEGJENFINNING.....	26
3.6.1. <i>Automatisk innholdsbasert gjenfinning (Content-Based Image Retrieval, CBIR)</i>	26
3.6.2. <i>Konseptbasert gjenfinning</i>	26
3.6.3. <i>Motivindeksering</i>	26
4. VIKTIGE STRUKTURER	29
4.1. METADATA.....	29
4.1.1. <i>Tilnærminger til metadata</i>	31
4.1.2. <i>Interoperabilitet mellom metadata</i>	32
4.1.3. <i>Metadata for bilder</i>	33
4.2. TESAURUS.....	41
4.2.1. <i>Vokabularkontroll gjennom tesaurus</i>	42
4.2.2. <i>Relasjoner mellom termer i tesaurus</i>	43
4.2.3. <i>Display av termer i tesaurus</i>	44
4.2.4. <i>Prinsipper for valg av foretrukne termer</i>	45
4.3. GAZETTEERS.....	46
5. SYSTEMER OG SAMLINGER	53
5.1. FROM DIGITAL LIBRARY TO MULTIMEDIA APPLICATION.....	53
5.2. SCHOOLNET.....	53
5.3. ADL OG ADEPT.....	56
5.3.1. <i>ADL arkitektur</i>	56
5.4. GEOXWALK.....	57
5.5. TEXTUAL – GEOSPATIAL INTEGRATION SERVICES FOR THE NATIONAL SMETE DIGITAL LIBRARY.....	58
5.6. SAMLINGER I NASJONALBIBLIOTEKET.....	58
6. MITT SYSTEM	61

6.1.	GALLERI NOR I DAG	61
6.2.	ADL GAZETTEER	66
6.2.1.	ADL Gazetteer Content Standard (ADL GCS)	67
6.2.2.	ADL Feature Type Tesauros (ADL FTT).....	68
6.2.3.	Identifiserte kriterier for å integrere gazetteers i digitale bibliotek	68
6.3.	SYSTEMFORSLAG	70
6.3.1.	Brukere – hvem er de og hvilke ønsker/behov har de.....	70
6.3.2.	Bildekategorier	70
6.3.3.	Hva skal vi kunne søke etter	71
6.3.4.	En norsk gazetteer for Galleri Nor.....	71
6.3.5.	Elementer fra ADL GCS som inngår i forslaget.....	73
6.3.6.	Navnetypesaurus for Galleri Nor.....	75
6.3.7.	Interoperabilitet med andre samlinger	77
6.3.8.	Dublin Core for Galleri Nor.....	77
6.3.9.	Diskusjon og forslag til DC metadata for Galleri Nor.....	77
6.3.10.	Oversikt over DC metadataforslag til Galleri Nor	82
7.	PROTOTYPING.....	85
7.1.	IMPLEMENTASJON I FORHOLD TIL IDENTIFISERTE KRITERIER	86
7.1.1.	Andre forhold.....	86
7.2.	FUNKSJONELLE KRAV TIL PROTOTYPEN	87
7.2.1.	Søk via stedsnavn.....	87
7.2.2.	Søk via navnetype	87
7.2.3.	Søk via stedsnavn og navnetype.....	87
7.2.4.	Navnevarianter	88
7.3.	IKKE FUNKSJONELLE KRAV	88
7.4.	TEKNOLOGI FOR IMPLEMENTASJON.....	88
7.5.	ARKITEKTUR FOR PROTOTYPEN	88
7.6.	GRENSESNITT MOT BRUKER	91
7.6.1.	Søkeform for forespørsel fra bruker	92
7.6.2.	Presentasjon av søkeresultat	95
7.7.	JAVA-KLASSER.....	96
7.7.1.	Kommunikasjon mellom JSP og Java-klasser	96
7.7.2.	Kommunikasjon mellom Java-klassene	97
7.8.	DATABASER	98
7.8.1.	Gazetteer.....	98
7.8.2.	Samling.....	99
7.8.3.	Data i dtabasene.....	99
7.8.4.	Navnetypesaurus.....	99
7.8.5.	Bildene i testsamlinga.....	99
7.8.6.	Metadata om bildene	100
7.8.7.	Stedsnavn i Gazetteeren.....	101
7.8.8.	Navnetyper i Gazetteeren	102
7.8.9.	Relasjoner mellom forekomster	102
7.8.10.	Relasjoner mellom typer	103
8.	TESTING OG EVALUERING.....	104
8.1.	TESTING	104
8.1.1.	Testing av prototypen	104
8.1.2.	Prototypen vs. Galleri Nor	105
8.2.	DISKUSJON OG EVALUERING	115
8.2.1.	Testresultater	115
8.2.2.	Prototypen som helhet	115
8.2.3.	Gazetteeren.....	116
8.2.4.	Navnetypesaurus.....	117
8.2.5.	Hierarkisk indeks.....	117
8.2.6.	Automatisk queryutvidelse	117
8.2.7.	Dublin Core for bildebeskrivelser	117
8.2.8.	Oppgaven som helhet.....	118
8.2.9.	Veien videre.....	119

9. REFERANSELISTE.....	122
VEDLEGG.....	130
A. DUBLIN CORE	132
I. DUBLIN CORE FOR GALLERI NOR - DISKUSJON	132
II. DCMI METADATA TERMER PR. 04.03.2003.....	140
B. ADL GAZETTEER CONTENT STANDARD.....	148
C. NASJONALBIBLIOTEKETS DATABASEMODELL FOR GALLERI NOR.....	152
D. TESTSAMLINGA – BILDER OG METADATA.....	154
E. TABELLER FOR TESAURUS.....	168
I. NAVNETYPEHIERARKI.....	168
II. FORETRUKNE TERMER USES FOR IKKE FORETRUKNE TERMER.....	169
III. IKKE FORETRUKKET TERM USE FORETRUKKET TERM.....	171
F. SQL RESULTAT GAZETTEER	172
G. DOKUMENTASJON AV JAVA-KODE	176
I. KLASSEDIAGRAM.....	177
II. KLASSER OG METODER	178
MARO.GEOFINNER CLASS SPORRING.....	178
<i>Sporring</i>	178
<i>getStedsNavn</i>	179
<i>getNavneType</i>	179
<i>getPord</i>	179
MARO.GEOFINNER CLASS DATABASESOKER.....	180
<i>Databasesoker</i>	180
<i>gjorQuery</i>	180
MARO.GEOFINNER CLASS GAZETTEERSOKER	181
<i>Gazetteersoker</i>	182
<i>gjorOppslagGazetteer</i>	182
<i>gjorQuery</i>	182
<i>lagreSted</i>	183
<i>leggTilSted</i>	183
<i>harSted</i>	183
<i>finnSteder</i>	183
<i>finnSteder</i>	183
<i>finnTyper</i>	184
<i>finnTyper</i>	184
<i>lagreType</i>	184
<i>leggTilType</i>	184
<i>harType</i>	184
<i>finnStedOgType</i>	184
<i>tomSteder</i>	185
<i>finnBildeSteder</i>	185
<i>finnTypeSteder</i>	185
MARO.GEOFINNER CLASS OVERSETTER.....	186
<i>Oversetter</i>	187
<i>finnSted</i>	187
<i>finnVarNavn</i>	187
<i>finnSteder</i>	187
<i>erDelAvSted</i>	187
<i>harSteder</i>	187
<i>finnType</i>	187
<i>finnSted</i>	187
<i>erDelAvType</i>	187
<i>harTyper</i>	187
MARO.GEOFINNER CLASS STED	188
<i>Sted</i>	188

<i>getStedId</i>	189
<i>getStedNavn</i>	189
<i>getNavnType</i>	189
MARO.GEOFINNER CLASS BILDESOKER	190
<i>Bildesoker</i>	191
<i>gjørOppslagBilder</i>	191
<i>gjørQuery</i>	191
<i>gjørQuery</i>	192
<i>finnBilder</i>	192
<i>lagBildeSporring</i>	192
<i>finnBildeSteder</i>	192
<i>finnBildeSteder</i>	192
<i>finnSted</i>	193
<i>finnBilderFraSted</i>	193
MARO.GEOFINNER CLASS BILDE	194
<i>Bilde</i>	194
<i>getBildeId</i>	195
<i>getBildeSteder</i>	195
<i>getSted</i>	195
<i>leggTilSted</i>	195
H. KILDEKODE	196
I. DATABASESCRIPT FOR GAZETTEEREN	196
II. DATABASESCRIPT FOR BILDEMETADATA	197
III. SCRIPT FOR BRUKERGRENSESNIFF	198
<i>form.jsp</i>	198
<i>resultat.jsp</i>	200
<i>inc1.jsp</i>	202
<i>inc2.jsp</i>	202
<i>gftesaurus.html</i>	203
IV. JAVAKODE	206
<i>Klasse Sporring</i>	206
<i>Klasse Databasesoker</i>	207
<i>Klasse Gazetteersoker</i>	208
<i>Klasse Oversetter</i>	217
<i>Klasse Sted</i>	218
<i>Klasse Bildesoker</i>	219
<i>Klasse Bilde</i>	224
<i>Klasse SokeProgram</i>	226

Tabelloversikt

Tabell 4-1 Retningslinjer for valg av foretrukne termer (uf = use for)	45
Tabell 6-1 Liste over opprinnelige Galleri Nor elementer som er tilgjengelig for den allmenne bruker	64
Tabell 6-2. Bildekategorier	71
Tabell 6-3. Navnetypeskjema	76
Tabell 6-4. Forslag til samlingsmetadata for Galleri Nor basert på Dublin Core.	83
Tabell 6-5. Forslag til metadata på ressursnivå for Galleri Nor basert på Dublin Core. 83	
Tabell 7-1. Eksempel fra tabell bilde_sted i databasen maro_gn_metadata.....	101
Tabell 7-2. Eksempel fra tabellen steder i databasen maro_GAZETTEER.....	102
Tabell 7-3. Eksempel fra tabellen navne_varianter fra databasen maro_GAZETTEER.	102
Tabell 7-4. Eksempel fra tabellen typer i databasen maro_GAZETTEER.	102
Tabell 7-5. Eksempel fra tabellen relasjoner i databasen maro_GAZETTEER.....	103
Tabell 7-6. Eksempel fra tabellen relasjonstyper i databasen maro_GAZETTEER. ...	103
Tabell 7-7. Eksempel fra tabellen Overordna_typer i databasen maro_GAZETTEER.	103
Tabell 8-1. Søketermer for testscenarier.....	106
Tabell 8-2. Viser antall treff ved søk etter navnetyper / søkeord i eksisterende Galleri Nor.....	107
Tabell 8-3. Viser antall treff ved søk etter stedsnavn.	107
Tabell 8-4. Viser antall treff ved søk etter stedsnavn og navnetype / søkeord.....	108
Tabell 8-5. Testresultat scenario 1.....	110
Tabell 8-6. Testresultat scenario 2.....	110
Tabell 8-7. Testresultat scenario 3.....	111
Tabell 8-8. Testresultat scenario 4.....	111
Tabell 8-9. Testresultat scenario 5.....	112
Tabell 8-10. Testresultat scenario 6.....	112
Tabell 8-11. Testresultat scenario 7.....	113
Tabell 8-12. Testresultat scenario 8.....	113

Figuroversikt

Figur 4-1. Bilde med MARC-post fra bildebasen ved Trondheim folkebibliotek	35
Figur 4-2 Skjerm bilde fra ADL Gazetteer som selvstendig web-tjeneste.....	50
Figur 4-3 Skjerm bilde fra ADL med integrert ADL Gazetteer	51
Figur 4-4 Skjerm bilde fra Norgesglasset.....	52
Figur 5-1 Bildebeskrivelse på to nivåer.....	54
Figur 5-2 Bilde av nybyggerfamilie. Region 1 er en sirkel rundt kvinnens hode til høyre i bildet.....	55
Figur 5-3 ADL Arkitektur	56
Figur 6-1 Skjerm bilde fra Galleri Nors Søkeside.....	63
Figur 6-2 Hovedpunkter i ADL Gazetteer Content Standard (elementer med uthevet skrift er obligatoriske, (R) betyr at de er repeterbare)	67
Figur 6-3. Yttergrenser for avgrensede boks som rammer inn en del av Nord-Norge. Kartutsnitt fra ADL Gazetteer.	75
Figur 7-1. Overordna systemmodell.....	89
Figur 7-2. Aktivitetsdiagram som viser mulige veier gjennom systemet.....	90
Figur 7-3. Bruker kommuniserer med systemet via et web-grensesnitt ved å sende en forespørsel og får presentert søkeresultat.	91
Figur 7-4. Brukergrensesnittets søkeform.	92
Figur 7-5. Hierarkisk indeks over navnetyper i søkebildet.	93
Figur 7-6. Del av navnetypesaurus som er tilgjengelig via en lenke fra søkesiden.....	94
Figur 7-7. Presentasjon av søkeresultat.	95
Figur 7-8. Brukergrensesnittet kommuniserer med klassene Sporing og Databasesoker.	96
Figur 7-9. Klassen Databasesoker kommuniserer med klassene Gazetteersoler og Bildesoker.....	97
Figur 7-10. Databaseskjema for Gazetteeren.	98
Figur 7-11. Databaseskjema for bildemetadata.	99
Figur 7-12. Synlige metadata for bilde_o3 slik de framstilles for bruker i eksisterende Galleri Nor.....	100
Figur 8-1. Søkeresultat fra Galleri Nor ved søk via termen <i>byer</i>	108

1. Introduksjon til oppgaven

Hovedtema for denne avhandlingen er gazetteers og tesauri og hvordan strukturer som disse kan integreres og brukes for å støtte gjenfinningsprosessen i digitale bibliotek. Innholdet i avhandlingen er organisert med tanke på å gi leseren en grunnleggende forståelse for aktuelle problemområder som er sentrale i avhandlingen og hvilke egenskaper gazetteers og tesauri har som gjør at de muligens kan utnyttes i håndteringen av disse. En prototyp er utviklet og implementert med tilstrekkelig funksjonalitet til at viktige prinsipper er testet og evaluert.

1.1. Oppgaveformulering

Kan vi utnytte egenskaper ved strukturer som gazetteers og tesauri til å støtte gjenfinning av dokumenter i samlinger?

I denne avhandlingen skal jeg med utgangspunkt i en norsk fotosamling og identifiserte problemområder, undersøke om vi kan bedre gjenfinningsmuligheten ved å integrere samlingen i et system med en gazetteer. Jeg skal foreslå en norsk gazetteer til fotosamlingen og et typeskjema for en navnetypesaurus for gazetteeren. Videre skal jeg undersøke om et system som gjør automatisk oppslag i gazetteeren, kan utnytte geografisk informasjon til å delvis kompensere for inkonsistens i eksisterende metadatabeskrivelser.

For å støtte interoperabilitet mellom samlinger og systemer skal jeg foreslå et egnet metadatasjema for aktuelle fotosamling.

1.2. Problembeskrivelse

Å benytte geografiske navn som søketermer ved søk etter informasjonsobjekter er ikke uvanlig, men gjenfinningsmuligheter begrenses ofte av at en er avhengig av eksakt match mellom søketermer og registrerte søkbare metadata i databaser.

Konseptbaserte motivbeskrivelser av fotografier er komplisert og arbeidskrevende. Mye av arbeidet gjøres ofte manuelt og innebærer en fare for at indekseringen påvirkes av indekserers subjektive oppfatning av motivet som skal indekseres. Manuell indeksering er tidkrevende og flere personer kan være involvert i indeksering for en og samme samling. Detaljeringsgraden i indeksering kan variere og i tillegg kan mangel på vokabularkontrollerende virkemidler medføre inkonsistens i indekseringen. Å søke etter og finne relevante fotografier kan, for mange samlinger, være svært vanskelig dersom bruker ikke vet noe om bakgrunnen for indekseringen og hvilke begreper indekserer forholder seg til. En kan ha optimale databaseløsninger og søkesystemer, men oppleve å allikevel komme til kort i søkeprosessen som følge av inkonsistens i indeksering. Med den enorme mengden av informasjon som finnes samt variasjon i indekseringspraksis for ulike systemer og samlinger står en overfor store utfordringer på dette området.

1.3. Mål for oppgaven

Et overordna mål for denne oppgaven er å vise hvordan integrasjon og bruk av gazetteers og tesauri i digitale bibliotek kan påvirke mulighetene for gjenfinning av dokumenter generelt og bilder spesielt.

1.4. Delmål

Jeg skal gjennom arbeidet med denne avhandlingen:

- Få en grunnleggende forståelse for hva en digital gazetteer er og hvilke egenskaper som kan utnyttes i digitale bibliotek
- Bruke Galleri Nor som eksempel for å konkretisere problemer og mulige løsninger
- Gi en grunnleggende beskrivelse av problemområder i forbindelse med søk og gjenfinning i Galleri Nor
- Foreslå en gazetteer og en navnatypetesaurus for Galleri Nor
- Foreslå et gjenfinningssystem for Galleri Nor som integrerer en gazetteer
- Implementere en prototyp for å teste noen prinsipper og for å vise noen muligheter et slikt integrert system kan gi
- Benytte fotografier fra Galleri Nor for testing og evaluering av prototypen.
- Vurdere Dublin Core kan brukes som utvekslingsformat for å støtte interoperabilitet mellom Galleri Nor og andre systemer og samlinger.

1.5. Avgrensing av oppgaven

Denne oppgaven har fokus på hvordan vi kan ta i bruk strukturer som gazetteers og tesauri for å påvirke muligheter for gjenfinning av dokumenter generelt og bilder spesielt i digitale bibliotek. Bildegjenfinning er i denne sammenhengen basert på konseptbaserte beskrivelser.

Bilder fra fotosamlinga Galleri Nor benyttes som eksempel for å konkretisere problemer og mulige løsninger. Beskrivende informasjon som benyttes vil være begrenset til geografisk relatert informasjon som allerede finnes i eksisterende beskrivelser av bildene i Galleri Nor.

Prototyping vil ha fokus på å vise muligheter framfor tekniske løsninger. Prototype vil være et forenklet system, ikke et operasjonelt system

Geografisk referert informasjon og geografisk relaterte søketermer er tema i denne avhandlingen, men den omfatter ikke tradisjonelle geografiske informasjonssystemer (GIS).

2. Digitale Bibliotek

I vår del av verden er bibliotek noe de fleste av oss har et forhold til. I Norge har vi et lovbestemt offentlig bibliotektilbud som fordeler ansvar for å forvalte informasjon og kunnskap på ulike forvaltningsnivåer og som i prinsippet skal sikre tilgjengelighet til informasjon og kunnskap for alle.

I lovens *kapittel 1. § 1* heter det [83] :

”Folkebibliotekene skal ha til oppgave å fremme opplysning, utdanning og annen kulturell virksomhet gjennom informasjonsformidling og ved å stille bøker og annet egnet materiale gratis til disposisjon for alle som bor i landet.”

Nå er det slik at vi lever i et vidstrakt land med mange avkroker, det bor folk på fjellet og i daler, på øyer og i områder der det kan være stor fysisk avstand mellom bosettinger og nærmeste offentlige bibliotek. Bokbusser og bokbåter kan til dels styrke tilbudet i utkantene, men det er ingen tvil om at tilgangen til offentlige bibliotek er større i sentrale områder.

Et viktig prinsipp som ligger til grunn for den offentlige bibliotekvirksomheten er at tilbudet skal være gratis. I andre deler av verden er tilgang til informasjon og kunnskap forbeholdt de som i utgangspunktet er ressurssterke og kan betale for seg og derigjennom bygges det opp under sosiale forskjeller i befolkningen. I Norge kan alle i prinsippet ha mulighet til å benytte seg av de offentlige bibliotektilbudene. Retten til undervisning og opplæring i skolen er lovfestet og mange av oss blir kjent med bibliotek gjennom skolebibliotek om ikke før.

For å finne informasjonsressurser i et tradisjonelt bibliotek kan vi søke manuelt, bøker er gjerne ordnet i et system, enten alfabetisk eller tematisk. Vi kan gå langs hyllene og la blikket vandre opp og ned, hit og dit, kanskje finner vi noe spennende eller interessant. Dersom vi vet akkurat hva vi vil ha kan vi gå direkte til det stedet vi forventer å finne aktuelle ressurs, eller vi kan kontakte bibliotekaren og be om hjelp til å finne den. Bibliotekaren kan gi oss anbefalinger og forslag på bakgrunn av tidligere lån eller fordi vi forteller hva vi er interessert i.

Utvikling av digitale teknologier, ikke minst nettverksteknologier har i stadig større tempo påvirket og fortsetter å påvirke ulike strukturer i samfunnet. Organisering av næringsvirksomhet, utforming av varer og tjenester, muligheter for menneskelig kommunikasjon og samhandling er områder som påvirkes. I takt med denne utviklingen har stadig lavere priser på da datamaskiner og programvare gitt flere av oss muligheten til å ta disse nye teknologiene i bruk. Det er en form for gjensidighet i denne utviklingen. Vi ønsker å utforske de muligheter som data og informasjonsteknologien gir, vi utvikler nye systemer og finner stadig nye bruksområder. Samtidig gjør vi oss avhengige av teknologien for å håndtere de endringer som utviklingen fører med seg. Rask og rimelig tilgang til internett åpner dører ut i verden og gir oss muligheter til å få tak i informasjon fra mange kilder når som helst. Det er jo en lovende tanke, men snur vi den på hodet kan vi si at hvem som helst kan dytte sitt materiale etter oss i håp om at vi vil stoppe opp og ta det med oss. Alle har i prinsippet mulighet til å produsere og publisere materiale over www og det sier seg selv at kvaliteten på den informasjon som er tilgjengelig er av svært varierende karakter.

Vår kunnskap og erfaring påvirker vår mulighet til å orientere oss i informasjonsvrimmelen, men vi er avhengige av at noen gjør informasjonen tilgjengelig på en måte som gjør det mulig for oss å finne det vi er ute etter.

Oppbygging av samlinger, bevaring og tilgjengeliggjøring av informasjon har vært og er fortsatt sentrale virksomhetsområder for alle typer bibliotek, virksomhetsområder som skal ivaretas på en slik måte at de bidrar til forskning, undervisning og formidling. At enorme mengder informasjon nå er tilgjengelig over et verdensomspennende nettverk gjør er med på å tydeliggjøre at behovet for samlingsbygging, bevaring og tilgjengeliggjøring er høyst presserende for å kunne ha mulighet til å håndtere informasjonen. Vi har behov for digitale bibliotek.

Begrepet *digitale bibliotek* favner stort og det er ikke enkelt å gi en entydig definisjon av begrepet. Som fagfelt berører digitale bibliotek mange store og komplekse emner innenfor både datateknikk, kommunikasjons- og informasjonsvitenskap og begrepet er da også brukt i mange forskjellige miljøer og har mange forskjellige betydninger alt etter hvilket miljø man beveger seg i. Etter en grundig gjennomgang av begrepet definert ut fra forskjellige perspektiv oppsummerer Borgman [79] at ulike definisjoner er formulert for å tjene ulike formål. Hun hevder at mange av disse definisjonene berører to hovedtema, forskning og bibliotekpraksis. Fra et forskningsperspektiv ser man på *digitale bibliotek* som innhold som er samlet inn og organisert på vegne av brukergrupper og definisjonene skal tjene til å identifisere, fokusere oppmerksomhet på og spre interesse for forskningsaktuelle problemområder. De som har sitt virke i tradisjonelle bibliotek har en mer praktisk orientert tilnærming til begrepet. De ser *digitale bibliotek* som organisasjoner eller institusjoner som tilbyr informasjonstjenester i digital form. De er i følge Borgman, mest opptatt av de praktiske utfordringene som medfører overgangen fra manuelle til digitale tjenester.

Hva er så digitale bibliotek? Digitale bibliotek har mange sammenfallende interesser med tradisjonelle bibliotek. Noen er ansvarlige for å skape eller samle inn informasjon, organisere denne og gjøre den tilgjengelig for oss som brukere. Digitale bibliotek inneholder digitale dokumenter og beskrivelser av dokumenter. *Dokument* kan defineres i bred forstand og kan sees på som en enhet som kan dokumenteres.

Informasjonsressurser og *informasjonsobjekter* er begreper som brukes om de enheter som forvaltes i digitale bibliotek. For at en enhet skal være et informasjonsobjekt må den være informasjonsbærende og den må kunne identifiseres og organiseres i digitale bibliotek. Dokumentbegrepet inkluderer begrepet *dokumentlignende objekt*, et interessant begrep som fikk sin opprinnelse i forbindelse med det første møtet for Dublin Core Metadata Initiative (DCMI) i 1995. Dette initiativet førte til utvikninga av Dublin Core Metadata Element Set (DC) for ressursbeskrivelser på internett. Begrepet *dokumentlignende objekt* var opprinnelig definert som [80]:

”En entitet som kan sammenlignes med et dokument med det utgangspunkt at det i hovedsak er tekstbasert og deler andre egenskaper som et dokument har, for eksempel elektronisk mail eller regneark.”

Under en workshop for bildebeskrivelser på internett året etter, ble denne definisjonen utvidet til å omfatte alle typer objekter som er relativt stabile. Med stabilt menes at et objekt har samme innhold for alle, uavhengig hvor i verden man befinner seg. Eksempler på dokumentlignende objekter kan være tekst, bilde eller film.

Digitale bibliotek har ikke fysisk form som et tradisjonelt bibliotek, det er ingen bygning eller en adresse vi kan gå til for å hente informasjon. Begrensninger i tid og rom eksisterer ikke på samme måte som for vanlige bibliotek. Digitale bibliotek kan være tilgjengelig når som helst og fra hvor som helst og vi kan videreutvikle metoder for å gi enhetlig tilgang til flere systemer og samlinger.

Selv om digitale bibliotek inneholder digitale dokumenter vil mye av den informasjon eller de informasjonsobjekter vi kan få tilgang til være dokumenter som representerer originale informasjonskilder. For bilder kan dette være metadata om bildet, men det kan også være digitale kopier av opprinnelige fotografier beregnet på papirkopiering. Muligheten for å gjøre dokumentсурогater tilgjengelige kan ha flere fordeler. Lesk [1] påpeker tre åpenbare fordeler

- En kan preservere sjeldne og verdifulle objekt uten å nekte adgang for dem som er interessert i å studere disse. Som eksempel nevner Lesk et gammelt manuskript i British Museums varetakt som ikke var tilgjengelig for andre enn noen få inntil det ble scannet og lagt ut på Internet til almen beskuelse. I tillegg kunne en, ved å benytte forskjellige lyskilder ved scanningen, få fram detaljer som normalt ikke er synlig for det blotte øye.
- Dokumenter kan gjøres tilgjengelig for bruker i det øyeblikk de er digitalisert. Flere kan ha tilgang til samme dokument på samme tid og bruker kan få tilgang til dokumenter som de ikke ville ha mulighet til å oppsøke fysisk.
- Lagringskostnadene som vil være ”*av en annen verden*” for digitale dokumenter enn for fysiske eksemplarer av for eksempel bøker.

Samlinger i tradisjonelle bibliotek består stort sett av fysiske objekter, men bibliotekene har tradisjon for å bruke automatiserte og etter hvert digitale systemer for å organisere samlingene. Dersom beskrivelsene som representerer fysiske objekter kan sies å være digitale dokumenter har man et hybrid system, altså en kombinasjon av digitale og ikke digitale bibliotek.

Vi kan se digitale bibliotek bibliotek i en større sammenheng, som et informasjonsrom som er sammensatt av mange systemer og komponenter. Disse fungerer sammen uten synlige grenser for å gjøre samlinger og tjenester tilgjengelige for å dekke vårt informasjonsbehov..

Uavhengig av hvordan vi skal bygge framtidens digitale bibliotek er det aspekter fra mange ulike fagområder som skal ivaretas.

Samlingsbygging vil fortsette å være sentral i utviklinga av digitale bibliotek. Vi må definere mål og målgrupper for samlinger. Vi trenger metoder og strukturer for å organisere innholdet i samlinger slik at vi på best mulig måte kan gjøre det tilgjengelig for brukeren. Metadata og vokabularkontroll er nøkkelford, vi må katalogisere, indeksere og beskrive for ressursoppdaging og vi må ta vare på informasjonsobjektet gjennom hele dets levetid.

Informasjonsobjektene må gjøres tilgjengelige via systemer og tjenester for gjenfinning og bruk.

Vi må arbeide for å styrke mulighetene for å dele data både mellom ulike systemer og mellom komponenter internt i systemer. Utvikling av standarder og metoder for å kombinere formater vil være viktige oppgaver.

Digitale bibliotek kan ha flere funksjoner og tjenester. Digitale teknologier gjør det mulig å tilby andre typer funksjonalitet i tillegg til funksjonalitet som forbindes med samlingsorienterte bibliotek. Vi kan for eksempel tilby tjenester for å bearbeide informasjon, for å gi beslutningsstøtte i arbeidsprosesser og for nyttiggjøre oss informasjon slik at den omdannes til kunnskap. Pasientjournaler og læringssystemer kan være eksempler på systemer som kan benytte slike tjenester.

Nettverksbaserte gazetteers og tesauri er eksempler på systemer for kunnskapsorganisering som er tilgjengelige for bruk som selvstendige enheter, men vi ser nå flere eksempler på at slike systemer integreres i digitale bibliotek [71][72][73]. Slike systemer kan integreres som generelle komponenter med definerte tjenesteprotokoller for å dekke forskjellige behov eller som mer spesialiserte komponenter for å dekke spesielle behov.

Antall brukere og bruksområder for digitale bibliotek vil øke i framtida. Vi trenger brukerundersøkelser og metoder for å ivareta brukbarhet for grupper med forskjellige behov. Gode brukergrensesnitt er en forutsetning for at systemer skal bli brukt. Vi kan jo tenke oss en konkurranse på dette markedet også og da er det særdeles viktig å ta hensyn til brukere og brukbarhet.

Vi har folkebibliotek, fag og forskningsbibliotek, bedriftsinterne bibliotek eller såkalte corporate memory og private samlinger som alle vil kunne tjene forskjellige brukergrupper med forskjellige behov.

I tradisjonelle bibliotek kan vi gå til bibliotekaren og få anbefalinger og hjelp til å finne informasjon. Behov for skreddersydde løsninger vil være tilstede også for digitale bibliotek. Personalisering er et begrep som brukes i sammenheng med personlige tilpasninger av systemer og tjenester.

I ethvert bibliotek er det viktig å ha mekanismer for å ivareta de rettigheter som reguleres av åndsverkloven. Vi trenger mekanismer for rettighetshåndtering av digitalt materiale og vi trenger digital rettighetshåndtering av materiale, men det er viktig at håndtering av rettigheter baseres på forretningsmodeller som er forenlig med bibliotekenes formål om å være kilder til informasjon og kunnskap som skal være fritt tilgjengelig for alle samtidig som åndsverkhaverne skal beskyttes. Personvern må ivaretas på brukerens vegne.

For å sikre at digitale bibliotek utleverer god og riktig informasjon til rett person til rett tid er det nødvendig å ha gode metoder for å kvalitetssikre både informasjon og tjenester. Dette er et stort problem med www i dag. Vi har ingen garanti for at den informasjonen vi finner er god eller brukbar. Dette er et komplisert tema, det er vanskelig å vurdere kvalitet på informasjon. Derfor er det viktig å utvikle metoder og verktøy som kan støtte evalueringsprosesser. Men det er ikke bare

informasjonsobjektene som bør kvalitetssikres. Tjenester og brukbarhet er eksempler på andre aspekter der kvalitet spiller en avgjørende rolle.

Med tanken på digitale bibliotek som et globalt informasjonsrom er det nærliggende å tenke på www. Vi kan også se på www som et globalt informasjonsrom. Mange systemer og komponenter fungerer sammen, tilsynelatende uten altfor mange synlige grenser. Vi får tilgang til informasjon og tjenester.

Men, er informasjon på www svaret på visjonen om global tilnærming til digitale bibliotek? NEI!

Hvem som helst kan produsere og publisere materiale ukontrollert på www, det er fritt fram for alle. Mange av aspektene ved digitale bibliotek som nettopp er omtalt blir ikke ivarettatt for www. Resultatet er at altfor mange useriøse aktører er representert og store mengder ustrukturert informasjon gjør det vanskelig for oss som brukere å skille ut den informasjonen vi egentlig har behov for. Mange digitale bibliotek-prosjekter er tilgjengelige på www så vi kan kanskje si at www er et globalt informasjonsrom som rommer mange informasjonsrom der mange av disse kan være digitale bibliotek.

Med digitale bibliotek som med forskning generelt er det slik at en stor del av veien blir til mens vi går. Og mens vi går møter vi stadig vekk problemer og nye muligheter. Det er fortsatt et stykke igjen på denne veien så det kan bli spennende å se hva som skjer framover.

Det er mange utfordringer i vente og mange spørsmål som vi kanskje får svar på om noen år.

Et spørsmål er om digitale bibliotek vil erstatte dagens tradisjonelle bibliotek i framtida. Det er ikke godt å si, men jeg vil tro at spørsmålet er mest aktuelt for fag og forskningslitteratur der brukeren gjerne trenger informasjon med en gang. Vi kan ikke se bort fra at en utvikling kan gå i retning av at digitale bibliotek i hvert fall delvis vil erstatte slike typer bibliotek. En forutsetning for en slik utvikling er at vi kan laste ned, skrive ut, kopiere og bruke uten altfor mange restriksjoner og tekniske sperrer. Jeg kan vanskelig se for meg at en slik utvikling kan være aktuell for tradisjonelle folkebibliotek. Jeg tror vi har et forhold til bøker som er vanskelig å endre og jeg kan vanskelig forestille meg at et digitalt dokument skal kunne erstatte den følelsen det er å sitte med et eksemplar av en god bok i fanget. Da har jeg vel mer tro på hybride systemer som kan kombinere det beste fra to områder.

Geografisk referert informasjon i digitale bibliotek

Geografisk referert eller georeferert informasjon er ikke nødvendigvis synonymt med geografisk informasjon, men kan være informasjon som på en eller annen måte refererer til ett eller flere geografiske områder på jordas overflate.

I rapporten *geoXwalk - A Gazetteer Server and Service for UK Academia* [2] sier James Reid at Britiske "National Geographic Data Framework" (NGDF) estimerer at omtrent 80% av all informasjon som er registrert i UK er georeferert på en eller annen måte. Jeg har ikke klart å finne originalkilden for denne påstanden og vil på ingen måte ta stilling til om dette medfører riktighet eller ikke, men uansett hvor stor andel av informasjonsressursene som er georefererte vil det være knyttet spesielle egenskaper til disse ressursene. Disse egenskapene bærer med seg potensielle muligheter og det vil da

være et spørsmål om hvordan en kan benytte disse mulighetene til å utnytte georeferering i digitale bibliotek.

Vi kan skille mellom *direkte* og *indirekte* geografisk referert informasjon [61]. Over hele verden samles økende mengder informasjonsressurser som satellittfotografier, flyfotografier, fotografier fra ulike overvåkingsenheter rundt om på og utenfor jorda, digitale kartografiske produkter osv. Disse kan være geospasiale i den forstand at de refererer *direkte* til geografiske koordinater, lengdegrad og breddegrad, som lokaliserer deres avtrykk på jorda.

Indirekte georeferert er alle de informasjonsressurser som på en eller annen måte refererer til en lokasjon via geografiske navn og disse er det mange av rundt om i forskjellige samlinger. For eksempel kan et litterært verk handle om et eller flere steder i tillegg til at det kan være skrevet et eller flere steder, utgitt ett eller flere steder. Verket har minst en forfatter som er fra ett sted og bor et sted osv.

Når det gjelder fotografier generelt og de som tilhører Galleri Nor spesielt vil det være slik at motivet på en eller annen måte alltid vil kunne knyttes til ett sted. Eksisterende metadata for en stor del av fotografiene i Galleri Nor inkluderer da også lokasjonsinformasjon i form av geografiske navn.

Geografisk lokasjon for en forekomst kan uttrykkes *direkte* i form av koordinater for breddegrad og lengdegrad eller *indirekte* i form av relasjoner til andre forekomster.

Et geografisk avtrykk representerer en geografisk lokasjon. Et geografisk avtrykk representert ved koordinater kan være:

- et punkt – et enkelt sett koordinater
- en avgrenset boks – et dobbelt sett koordinater som representerer minimum og maksimum lengde og breddegrad
- en linje – et sett av punkter som ikke innrammer et rom
- et polygon – et sett av punkter som innrammer et rom

En geografisk forekomst kan knyttes til flere representasjoner av geografiske avtrykk. Disse kan være:

- av forskjellig type
- fra forskjellige kilder
- for forskjellige perioder – kan vise for eksempel utvidelse av bygrenser, sammenslåing av kommuner
- passende til forskjellige løsninger – kan for eksempel være svært detaljert eller mer generalisert

Geografisk referert informasjon er et aktuelt tema for flere prosjekter i digitale bibliotek, også for denne avhandlingen.

3. Søk og gjenfinning

3.1. Informasjonssøking

Marchionini definerer informasjonssøking som en ”*prosess som drives av livet selv*”. Mer spesifikt sier han at ”*informasjonssøking er en prosess som drives av menneskers behov for informasjon slik at vi kan samhandle med omgivelsene*” og at denne prosessen ”*er en del av menneskers kontinuerlig pågående innsats for å forstå og handle i verden*”. [18]

Han ser på informasjonssøking som en aktivitet for problemløsning der denne baseres på aktivitet for kommunikasjon. Marchionini hevder at dette perspektivet på informasjonssøking deles av flere modeller for søkeoppførsel .

Søkesystemer representerer kunnskap og tilbyr verktøy og regler for å gi oss tilgang og mulighet til å bruke kunnskapen [18].

Søkesystemet støtter informasjonssøking ved å strukturere kunnskap og begrense tilgang. Måten kunnskapen er organisert og gjort tilgjengelig påvirker i hvilken grad informasjonssøkeren er i stand til å skaffe seg tilgang til kunnskapen og derigjennom vedkommendes utførelse av informasjonssøk..

Som informasjonssøkere er vi opptatt av innholdet i de databasene vi søker i og i hovedsak det domene som databasen representerer. Marchionini kaller dette for *aboutness*. Vi er ikke bare interessert i innholdet, men i det som innholdet handler om. Dersom innholdet ikke er primært, men i form av dokumentsurrogater, må en i mange tilfeller til andre kilder for å finne det primære innholdet som en er på jakt etter.

3.2. Søkeprosessen

Marchioninis rammeverk for informasjonssøking er brukersentrert og handlingsorientert. Søk og gjenfinning beskrives som en iterativ prosess. Han deler søkeprosessen inn i flere delfaser som krever at brukeren utfører ulike former for handlinger. Det begynner med en anerkjennelses og aksepteringsfase og fortsetter gjennom ulike subprosesser helt til problemet er løst. Informasjonssøk-prosessen vil gjerne bestå av flere iterasjoner gjennom hendelsesforløpet før søket avsluttes.

Fase 1. Anerkjenne/ akseptere et informasjonsproblem.

Problemet kan oppstå på flere ulike måter, først og fremst internt eller eksternt motivert i forhold til informasjonssøkeren. Personen må altså bli klar over at det eksisterer et problem. Han kan da velge å fortrenge eller overse det, om *kostnadene* ved et eventuelt søk anses å bli for store. Dette gjelder både fysisk og mentalt.

Blir problemet akseptert, er det ulike faktorer som har spilt inn, som for eksempel kunnskap om søkesystemer, kunnskap rundt problemets tema, settingen og søkerens tiltro til hans informasjonsinfrastruktur.

Fase 2. Definere og forstå problemet.

Dette er en meget viktig fase under søkeprosessen. Fase 2 kan man si forblir aktiv, så lenge søket gjør fremgang.

Forståelsen er avhengig av kunnskap om oppgavens område, og kan også bli påvirket av settingen.

Den kognitive prosessen som identifiserer nøkkelkonsepter og forbindelser fører til en definisjon av problemet. For å forstå og definere problemet, må det være klart avgrenset og ha en slags form eller ramme for hva utfallet bør bli.

Det å ha en oppfatning av hva svaret bør bli vil påvirke vårt valg av søkesystem, som er den neste subprosessen i denne modellen.

Fase 3. Valg av søkesystem.

Denne fasen avhenger av søkerens tidligere erfaringer rundt temaet, rekkevidden av hans egen informasjonsinfrastruktur, og forventninger rundt hva svaret bør bli, som muligens ble formet under defineringen av problemet og oppgaven.

Kunnskapen rundt temaet vil være med å påvirke valg av system, da det sannsynligvis betyr at man har erfaring også fra ulike søkekilder rundt nettopp det feltet.

Ut fra domenekunnskap, generelle kognitive betingelser og tidligere søkeerfaring prøver søkerne å velge et passende system. I praksis vil søkeren konsultere flere ulike søkesystem samtidig som de nærmer seg løsningen på problemet.

Fase 4. Formulering av spørring

Spørreformuleringer krever nøyaktig forståelse av det valgte systemet, da det er store variasjoner innenfor den mengden tilbud som finnes. Det er vanlig å starte med et noe generelt søk, og deretter raffinere og reformulere spørringen.

I statiske søkesystemer som f.eks bøker har søkeren full kontroll. Her prøver vi selv å finne matcher, mens i dynamiske systemer som for eksempel mennesker vil selve intelligensen i systemet spille inn.

I elektroniske søkesystemer er spørringen delvis dynamisk. Et slikt system kan ha et strengt kontrollert vokabular, eller et fulltekst vokabular og man kan ha mulighet til å regulere kombinasjoner av søketermer ved å bruke logiske operatører.

Marchionini mente at elektroniske systemer har gitt verdifulle bidrag nettopp til denne fasen i søkeprosessen.

Det er vanlig at man går fra denne fasen og tilbake til definisjonsfasen og repeterer det hele, og forhåpentligvis oppnår et mer tilfredstillende resultat.

Fase 5. Utfør søk.

Utførelsen er basert på det som ble utviklet i fase 4, formuleringen av spørringen.

Det kan blant annet være å slå opp i et verk eller stille et spørsmål verbalt, trykke på en knapp, finne frem et leksikalsk bind osv.

Kommunikasjons- og datateknologien har hatt stor påvirkning på hvordan søk blir utført. Telefon, telefaks og elektronisk mail er kilder som kan benyttes i dynamiske søkesystemer. Statiske systemer har også blitt lettere tilgjengelige med utviklingen av nettverksteknologier.

Under denne subprosessen er det ikke uvanlig at man går ett steg tilbake og reformulerer spørringen.

Fase 6. Eksaminere resultatet.

Tilbakemeldingene etter utført spørring må vurderes av informasjonssøkeren før man kan si seg ferdig med søket. Har brukeren fått det svaret og den informasjonen vedkommende var ute etter?

Dette bør vurderes ut fra kvantitet, type, og formatet av det som ble gitt på tilbakemeldingen av søket. Et høyt antall treff er ikke det samme som at man har funnet mye relevans, men heller at man kanskje må gå tilbake og formulere en ny spørring.

Eller, faktisk gå helt tilbake og definere problemet og informasjonsbehovet på nytt igjen.

Fase 7. Trekke ut informasjonen.

Informasjon blir sett på som relevant, når den gir løsningen på det problemet man i søkeprosessen har prøvd å løse. Vurderinger i forhold til relevans avgjør om informasjon ignorert eller om den blir sett på som nyttig.

Denne fasen vil være knyttet til foregående fase der informasjonen ble eksaminert, og man vil ofte gå tilbake til fase 6.

Når informasjonssøkeren har valgt ut informasjon, blir den manipulert og integrert inn i søkerens kunnskap rundt temaet, og benyttet til å løse problemet som førte til denne søkeprosessen.

Det er ikke uvanlig at denne fasen itererer tilbake til å definere problemstillingen på nytt, eller til å formulere en ny spørring som er bedre egnet.

Fase 8. Refleksjoner. Iterasjoner. Avslutte.

Et søk blir sjelden fullført og avsluttet ved kun en enkel spørring. Utfallet brukes kanskje like ofte som tilbakemeldinger til videre søk.

Når og hvor man i søkeprosessen skal iterere krever en vurdering av selve søkeprosessen, hvordan det kan relateres til aksepteringen av hva som er problemet, den forventede kostnaden av å utføre søket og hvor godt den utvalgte informasjonen passer til problemet som man prøver å løse.

Man kan altså begynne med å definere problemet på nytt igjen, formulere en ny spørring, eller gå tilbake og eksaminere resultatet på nytt igjen.

Det kan være interessant å se på Marchioninis modell av søkeprosessen i forhold til søk og gjenfinning av bilder. To faser er spesielt interessante her. Formulering av query og eksaminasjon av søkeresultat.

Fasen der en formulerer query er en kritisk fase uansett hvilke typer medier det her er snakk om. For at en skal ha mulighet til å finne en relevant ressurs er en avhengig av å formulere query som på en eller annen måte treffer de data som er registrert om ressursen. Kontroll av vokabular mellom bruker og system er særdeles viktig i denne fasen. Når ressursene først er funnet må man velge strategi for å eksaminere resultatet. Målet med eksaminasjon er å velge ut eventuelle relevante dokumenter. Bilder har egenskaper som gjør at de egner seg for browsing på en annen måte enn tekstlige ressurser. Disse egenskapene kan gi grunn for å anta at et menneske vil være i stand til å eksaminere et søkeresultat bestående av et antall bilder ved browsing på atskillig kortere tid enn de kan gjøre med samme antall tekstlige ressurser. Dette innebærer det samme som at en på en gitt tid kan eksaminere et resultatsett av bilder som er betydelig større enn et resultatsett av tekstdokumenter. Dersom dette stemmer kan vi anta at vi, i sammenheng med bildegjenfinning kan godta søkeresultater med høyere recall og lavere presisjon enn en kan ved gjenfinning av tekstlige ressurser. Det er dessuten vanskeligere å rangere bilder i forhold til presisjon i et søkesystem. Rangeringen gjøres best av den som søker etter bilder og det er derfor vanskelig å måle presisjon.

3.3. Gjenfinning

Et overordna mål med søk etter og gjenfinning av informasjon er å skaffe riktig informasjon til rett person til rett tid. *Relevance feedback* er et begrep som sier noe om hva vi sitter igjen med som resultat etter et søk og hvor relevant dokumentene i resultatlista er for oss. *Recall* og *presisjon* er begreper som kan brukes for å si noe om graden av gjenfinning. *Relevans* sier mer om hva som er relevant for den enkelte informasjonssøker [19]. Vanligvis ønsker vi å oppnå lav recall og høy presisjon, men mulighetene for å rangere og måle resultater er forskjellig for ulike typer medier. Det er mulig at vi kan, ikke bare godta, men også ønske høyere recall for bilder enn for tekstdokumenter.

3.3.1. Recall

Recall er et uttrykk for gjenfinningseffektivitet som viser den prosentvise andel av de relevante dokumenter som er funnet via et søk i forhold til alle relevante dokumenter som finnes i databasen. Total recall vil altså lokalisere alle dokumenter som samsvarer med søkekriteriet for dokumentbasen. Det kan være vanskelig å beregne store databaser med mange dokumenter, men kan la seg beregne for testsamlinger der innholdet er kjent og oversiktlig.

3.3.2. Presisjon

Presisjon uttrykker den prosentvise andel dokumenter fra en søking som brukeren er interessert i. Begrepet fokuserer på de mest relevante, mest brukbare dokumenter i samlingen. Det brukes forskjellige teknikker for å rangere søkeresultater med det formål å gi høyest rang til de mest relevante dokumentene.

3.3.3. Relevans

Relevans uttrykker hvilke dokumenter som virkelig er relevante for bruker. Dette er vanskelig å måle og det er i grunnen bruker selv som må vurdere hvorvidt et dokument virkelig er relevant eller ikke. For en slik vurdering kan en ta i bruk ulike strategier ut fra hvilke type dokumenter man har og hvor store mengder man har.

For tekstdokumenter kan en ha mulighet til å lese abstracts eller skimme gjennom dokumentet. Dette er en prosess som kan kreve ganske mye tid. Ved veldig høy recall kan en vurdere å raffinere søket for å minske mengden av resultatdokumenter.

Når det gjelder visuelle medier som bilder vil det være mulig å browse resultatmengden for å evaluere søkeresultatet. En kan selvfølgelig ha kriterier for utvelgelse som er av en slik art at en kan gjøre utvelgelse på bakgrunn av metadata om bildene, men en kan ikke være helt sikker på at en har funnet det en ønsker før en kan inspisere objektet ved å faktisk se på motivet.

3.4. Brukerundersøkelser

For å kunne tilby gode forhold for informasjonsgjenfinning er det viktig å vite noe om hvordan mennesker søker etter informasjon og det er gjort mange studier som belyser dette temaet. Det har imidlertid vært vanskelig å finne dokumentasjon for studier som er

gjort for å undersøke hvordan mennesker søker etter bilder. Det hevdes av flere at dette er et område som er lite belyst og der det er behov for ytterligere forskning [55][56].

3.5. Søkesystemer

Søkesystemer er forskjellige og dette kan skape problemer for den som søker etter informasjon. Ekspertene på området er gjerne kjent med mange ulike typer systemer og er i stand til å utføre søk i de ulike systemene i tillegg til at de gjerne er eksperter på domene som det søkes i forhold til. De vet gjerne hvor de skal søke og hvilken terminologi som egner seg best for formålet, mens brukere som er mindre vant til slike søkesystemer vil være nødt til å bruke tid og energi på å finne fram til og forstå de ulike systemer. Ikke alle systemer er like intuitive.

Grensesnittet mellom bruker og system er avgjørende for hvordan systemet vil bli brukt og er dermed avgjørende for hvordan mulighetene er for bruker til å finne fram til den informasjon som er ønsket.

Ulike måter å kommunisere med grensesnittet på for ulike typer brukere:

Kommandobaserte grensesnitt for eksperter som kjenner kommandospråket kan være effektivt og raskt og kan gi muligheter for spesialiserte oppgaver.

Fill in former er mye brukt, også i web baserte systemer. Selv om formene er definerte i forhold til hva som kan fylles inn står bruker fritt til å skrive inn hva som helst.

Menystyrte grensesnitt lar gjerne bruker velge blant alternative handlinger og legger mer grenset for de handlinger bruker kan gjøre.

Grensesnittet inkluderer også hvordan resultatet av informasjonssøk presenteres for søkeren og elektroniske systemer har potensiale til å tilby flere ulike eller alternative representasjoner i forhold til brukers behov og ønsker.

Domene vil i dette tilfelle være geografisk referert eller relatert informasjon. For samlinga Galleri Nor vil det være mulig å kategorisere bilder. Rent intuitivt er det kategorier som peker seg ut slik som landskap/sted, portrett, aktivitet, hendelser/begivenheter. Dette er de som umiddelbart åpenbarer seg, men en kan også vurdere egne kategorier for transportmidler, arbeidsliv osv.

Setting er en faktor som påvirker søkeprosessen. Både søkerens mentale tilstand, arbeidsminne og fysiske omgivelser spiller en rolle her

Utkomme er både produkt og prosess. Som produkt er resultatet av å bruke et informasjonsgjenfinningssystem, feedback fra systemet. Resultatet tjener også som objekt for evaluering av søket eller systemet. Utkomme kan også sees på som mentale refleksjoner omkring episoder i forbindelse med informasjonssøkingen.

3.6. Bildegjenfinning

Forskning omkring emnet bildegjenfinning dreier seg i hovedsak om to retninger, automatisk innholdsbasert gjenfinning og konseptbasert eller semantisk innholdsbasert gjenfinning.

3.6.1. Automatisk innholdsbasert gjenfinning (Content-Based Image Retrieval, CBIR)

CBIR kan beskrives som gjenfinning av bilder basert på egenskaper som kan ekstraheres automatisk fra selve bildet så som spesielle fargenyanser, overflateegenskaper, form, mønstre osv. Dette er et spennende forskningsområde der ulike teknikker er tatt i bruk. Et eksempel kan være "search by image" som er en teknikk der bruker sender et bilde inn i systemet og kan få lignende bilder tilbake. Et eksempel på bruk av "search by image" kan være for identifikasjon av ansikter. En kan da fotografere et ansikt for så å søke gjennom et antall ansiktsbilder som er lagret for formålet for å kontrollere om et bilde av vedkommende finnes i databasen. I andre tilfeller der bilder ikke er lagret for spesielle formål kan resultatet være bilder som har samme fargenyanser eller lignende former, men som ikke har noen likheter i semantisk innhold. Dette er et mye omtalt problem som gjerne omtales som "det semantiske gap" og indikerer at det er et gap mellom det slike system kan tilby og det bruker faktisk er interessert i som er gjenfinning basert på semantiske egenskaper. Hybride systemer som kombinerer CBIR med tekstbasert semantisk gjenfinning kan få mer oppmerksomhet i framtida[32].

3.6.2. Konseptbasert gjenfinning

Dette er den tradisjonelle måten å gjenfinne bilder på og kan beskrives som gjenfinning basert på tekstlige beskrivelser av bildet i forhold til et konsept. Denne formen for bildegjenfinning omtales også som semantisk innholdsbasert bildegjenfinning [33] eller kontekstbasert bildegjenfinning [34].

Det er denne formen for bildegjenfinning de fleste av oss er kjent med og som er mest utbredt.

3.6.3. Motivindeksering

Motivindeksering er en nødvendighet for at en skal kunne benytte gjenfinningssystemer som baseres på tekstlige motivbeskrivelser. Ut fra interesser, kunnskaper, erfaringer og kulturell bakgrunn kan vi ha ulike oppfatninger om hva som er et bildes motiv. En svensk rapport om bildedatabaser og digitalisering *Plattform for ABM-samvirke* gir følgende definisjon av begrepet motiv [35]:

"motiv det som en bild ger en synlig framställning av. Anmärkning: Motiv är det som en bildframställning åskådliggör – genom att avbilda, representera, illustrera eller symbolisera – och som med hjälp av synsinnet kombinerat med kunskap om omvärlden och tidigare erfarenhet av bildframställningar kan uppfattas som "en bild av" något."

Dette er en omfattende definisjon som åpner for ulike roller et motiv kan ha og som også sier noe om at kunnskap om verden og erfaringer påvirker menneskets oppfatning av noe. Selv om definisjonen er omfattende synes jeg den er god fordi den viser kompleksiteten som begrepet omfatter og indikerer derigjennom at motivindeksering er en komplisert prosess.

I tillegg til at vi kan oppfatte en bildeframstilling som et bilde av noe skal den som indekserer omsette denne oppfatningen til beskrivende termer som skal være utgangspunkt for at bildet skal kunne gjenfinnes av andre.

Rapporten sier videre at arbeidsgruppens enhetlige syn er at målet med motivindeksering er å kunne søke etter bilder som inneholder en visuell framstilling av det begrep som indekseringstermen refererer til. For at søkingen skal bli så effektiv som mulig bør man ved indekseringen ta utgangspunkt i et kontrollert vokabular, en tesaurus, et klassifikasjonssystem eller en emneordliste. Systemet som brukes bør være lett å tilpasse og kunne forstås både av den som søker og den som indekserer. Det er viktig at den som søker får et forventet resultat, det vil si at et søk bør gi kun relevante treff. Derfor bør et system for motivindeksering også inneholde en metode for hvordan indekseringen skal gjøres samtidig som det er utformet slik at den enkelte institusjon kan velge hvor detaljert man vil at indekseringen skal være [35].

Hvilken informasjon som er nødvendig å registrere for å ivareta tekniske og bevaringsmessige hensyn til informasjonsobjekter må vurderes av de som er eksperter på området. Det som er interessant i sammenheng med denne oppgaven, er beskrivelse av motiv. Uansett hvilken standard eller format en velger å bruke så vil det framdeles være en utfordring å beskrive fotografiers motiv. De fleste modeller for beskrivelser er nokså detaljerte på området tekniske data, men når det kommer til motiv så er det stort sett nokså åpne elementer.

Fotografiers motiv kan vise både tilsiktede og utilsiktede elementer som er med på å formidle kulturhistorisk informasjon. Registreringsopplysninger om bildets opphavssituasjon så som fotograf, tid og sted for eksponeringer stort sett enkel og grei informasjon dersom den er kjent. Motiv derimot krever en viss grad av subjektiv menneskelig tolkning for å kunne oversettes til søkbare termer.

4. Viktige strukturer

For å beskrive informasjonsressurser generelt og bilder spesielt er metadata et nøkkelord. For at brukere skal ha muligheter til å finne igjen bilder som er beskrevet av andre kan en si at beskrivende metadata er kritisk, men beskrivende metadata er ikke nok. Noen viktige forutsetninger må være tilstede for at søkesystemer basert på beskrivende metadata skal kunne gi gode gjenfinningsmuligheter. Vi har behov for å ha en viss kontroll med de metadata som brukes slik at det kan være mulig for bruker å finne fram til de samme begreper som brukes ved indeksering av ressursene. Videre har vi behov for konsistens i beskrivelsene av ressursene. Forskjellige termer som beskriver samme begrep kan være en kilde til misforståelser og vanskeliggjøre gjenfinning i et system. Ulik detaljeringsgrad i beskrivelser kan også være problematisk. Dette er eksempler på problemområder som kan være gjeldende både internt i systemer og for samvirke mellom ulike systemer.

Stilt overfor disse problemene kan en velge en av flere strategier. Den enkleste er å overse problemet og late som det ikke eksisterer. Man må i så fall lære seg å leve med problemene og takle konsekvensene på en eller annen måte. En annen strategi er å forsøke å oppnå enighet mellom utviklere og brukere om detaljer så vel som hovedprinsipper for utvikling og bruk av systemer. En slik strategi vil sannsynligvis skape flere problemer enn den kan løse.

En mulighet er å ta i bruk prinsipper og mekanismer for å kontrollere vokabular og til en viss grad kompensere for inkonsistens i beskrivelser. Strukturer for å organisere og definere terminologi, som vi bruker for å representere konsepter og konkrete ressurser, kan integreres i digitale bibliotek. Gazetteers og tesauri er eksempler på slike strukturer som kan integreres i systemer med samlinger og utnyttes i den hensikt å påvirke gjenfinningsmuligheter. Disse kan benyttes til automatisk queryutvidelse som kan kompensere noe for manglende konsistens i beskrivelser. Tesauros er et verktøy for å kontrollere vokabular og kan lede både bruker og indekserer til brukbare termer. Hierarkisk indeks kan fungere som et virkemiddel for å kontrollere vokabular mellom bruker og system.

4.1. Metadata

I sin enkleste form kan en definere metadata som data om data. Metadata beskriver data og brukes for å sette data i en sammenheng eller i et system som gjør data mer tilgjengelig for brukere.

Metadata benyttes for å [9] :

- Forbedre tilgang til informasjonsobjekter
- Dokumentere og vedlikeholde sammenhenger og relasjoner mellom ulike informasjonsobjekter, mellom objekter og mennesker, steder, hendelser og bevegelser
- Muliggjøre forskjellige bruksområder for samme informasjonsobjekt
- Lenke sammen og samtidig skille mellom ulike versjoner av et informasjonsobjekt
- Sikre at et informasjonsobjekt kan eksistere uavhengig av det systemet som opprinnelig brukes for lagring og gjenfinning av objektet.

- Medvirke til evaluering og forbedring av informasjonssystemer

Vi omgir oss med og bruker metadata i mange sammenhenger i vårt daglige liv uten at vi nødvendigvis kjenner og bruker begrepet *metadata*. Uten metadata kan vi få problemer med å forstå verden omkring oss.

Metadata kan brukes for å beskrive informasjonsobjekter av ulike karakterer. Vi bruker metadata for å katalogisere og klassifisere informasjonsobjekter i den hensikt at man, på enklest mulig måte, kan finne, identifisere, velge ut og få tak i den informasjon/ de informasjonsobjekter man er interessert i. En kan si det slik at jo bedre man er i stand til å beskrive informasjonsobjekter ved hjelp av metadata, jo større mulighet har man for å finne tilbake til aktuelle lagrede informasjonsobjekter og samtidig slippe å gå igjennom mengder av uønsket informasjon.

Termen *metadata* er en tvetydig term som tillegges forskjellig betydning avhengig av hvem som bruker termen og i hvilken sammenheng den brukes. Definisjoner av begrepet *metadata* spenner fra spesifikke til svært generelle. Enkelte definerer metadatabegrepet kun i forbindelse med elektroniske ressurser [5], mens andre definisjoner er mer generelle og kan omfatte alle typer ressurser [6]

Tradisjonelt har termen *metadata* vært brukt for å referere til katalogiserings- eller indekseringsinformasjon som skapes i den hensikt å organisere, beskrive og forbedre tilgang til informasjonsobjekter. Dette har tradisjonelt vært forbundet med møysommelig og tidkrevende manuelt arbeid.

Fra 1960-årene har bibliotekarer begynt å dele beskrivende metadata ved å bruke automatiserte systemer som offentlig tilgjengelige kataloger online. Dette har vært gjort med støtte fra internasjonalt kjente katalogiseringsregler og struktur og innholdsstandarder som MARC [7] og spesialiserte emneoverskrifter som LCSH [8].

Metadata kan brukes i henhold til definerte standarder eller formater som finnes i mange ulike varianter alt fra veldig enkle formater som kan brukes av hvem som helst, til kompliserte systemer som er utviklet av eksperter for eksperter.

Behov for beskrivende metadata varierer for ulike grupper alt etter hva slags informasjonsobjekter som skal beskrives og hva som er intensjonen med dette. Samtidig vil en del behov være de samme for alle typer informasjonsobjekter. Som følge av dette har det vokst fram et økende antall nye formater som forsøker å ta vare på det som er felles for ulike behov samtidig som en har mulighet til å få fram forskjeller i måten å beskrive informasjon på.

Et informasjonsobjekt tilordnes metadata på forskjellige vis og av forskjellige involverte ressurser. Gjennom ulike faser i et objekts levetid kan det tilordnes metadata som kan assosieres med objektet på forskjellige måter.

- Metadata kan knyttes direkte til informasjonsobjektet og lagres som en del av dette.
- Metadata kan knyttes til informasjonsobjektet via pekere eller hyperlenker.
- Et metadataregister kan dokumentere sammenhenger mellom metadata og informasjonsobjekter og mellom ulike aspekter ved metadata.

4.1.1. Tilnærminger til metadata

Ulike behov for metadata

Digitale informasjonssystemer inkluderer gjerne digitale versjoner av informasjonsobjekter i tillegg til beskrivelse av innholdet. Digitale informasjonsobjekter i et informasjonssystem krever andre typer vedlikehold enn fysiske objekter i for eksempel et museum og vil dermed også ha behov for andre typer metadata.

Ulike typer metadata

Metadata brukes for å identifisere og beskrive informasjonsobjekter, men kan brukes i videre forstand og kan si noe om hvordan et informasjonsobjekt oppfører seg, dets funksjon og bruk, hvordan det skal håndteres og ikke minst hvilke relasjoner et informasjonsobjekt har til andre informasjonsobjekter.

For å forstå metadata som konsept i bred forstand kan det være nyttig å skille mellom ulike typer metadata [9]:

Beskrivende – metadata som brukes for å beskrive eller identifisere informasjonsressurser. Her kan en også skille mellom metadata som beskriver indre og ytre egenskaper ved en ressurs. Indre egenskaper er egenskaper ved selve ressursen som for eksempel fotograf eller motiv for et bilde. Ytre egenskaper er egenskaper som ikke kan trekkes ut av selve ressursen så som for eksempel rettigheter.

Administrative – metadata som brukes for å forvalte og administrere informasjonsressurser

Tekniske – metadata som relateres til hvordan et system fungerer eller hvordan metadata oppfører seg

Preservering – metadata som relateres til preservering av informasjonsressurser

Bruk av samlinger – metadata som relateres til nivå og type bruk av informasjonsressurser

Vi snakker her om ulike typer metadata, men metadata uttrykker også forskjellige karakteristikk. Vi kan knytte ulike attributter til metadata for å uttrykke ulike karakteristikk.

Struktur for metadata

Metadata kan tilordnes manuelt eller automatisk. For å kunne automatisere tilordning av metadata er det nødvendig at involverte ressurser utvikler en felles forståelse for metadata og at en benytter felles vokabular samt entydige regler for tilordning av metadata. Utvikling av internasjonale standarder for metadataformater er nyttig i denne sammenheng. Mange ulike formater eksisterer. Noen er generelle og kan benyttes til forskjellige typer informasjonsobjekter mens andre er tilpasset spesielle typer informasjonsobjekter eller tiltenkt spesielle bruksområder.

En kan skille mellom metadataformater ved å dele dem i grupper av enkle, strukturerte og rike formater [6].

Enkle formater har ingen spesiell struktur. Et enkelt format kan for eksempel være en automatisk generert fulltekstindeks som består av en samling termer fra dokumenter. Termene kan knyttes til dokumenter via lenker eller identifikatorer. Ved søk i slike

ustrukturerte metadata er alle termene likeverdige slik at en ikke har mulighet for å avgrense delmengder av termer det skal søkes i.

Strukturerte formater er formater som har en struktur selv om denne kan være forholdsvis enkel. Disse er gjerne basert på enkle attributt-verdi-par og gir muligheter for å legge kriterier til grunn for et søk. Man kan avgrense delmengder av søketermer på grunnlag av semantisk innhold og kan derved påvirke søking og gjenfinningsprosessen. Rike formater er formater der en tar i bruk delfelt for å beskrive detaljer ved ressursen. Slike formater gir mulighet for å spesifisere søk enda mer enn strukturerte formater. MARC er et eksempel på et rikt og strukturert format med mulighet for å ivareta mange forskjellige opplysninger om en ressurs. Hvor Dublin Core [10] kan plasseres i denne sammenhengen kan diskuteres. DC har attributt-verdi-par som gjør at en kan plassere formatet i gruppa for strukturerte formater. Muligheter for gjentakelse av elementer og utvidelser ved bruk av kvalifikatorer gjør DC til et fleksibelt format som det kan være vanskelig å plassere i en ren gruppe.

Det å tilordne metadata kan være en møysommelig og kompleks prosess. I noen tilfeller kan denne prosessen automatiseres, men ikke alltid. Prosessen kan være overkommelig også for ikke-eksperter i og med at det finnes formater med ulik grad av kompleksitet fra svært enkle formater til kompliserte formater som krever ekspertise for å kunne brukes.

4.1.2. Interoperabilitet mellom metadata

Med framvekst og bruk av nettverksteknologi har det oppstått et økende behov for å fremme interoperabilitet eller deling av data mellom applikasjoner.

I medisinske miljø har en i lange tider hatt tradisjon for å benytte latin som en verdensomspennende fellesnevner i terminologien. Dette gjør at mennesker i slike miljø kan utveksle informasjon seg imellom og til en viss grad forstå hverandre på tvers av landegrenser og egne språk, men det er ikke nødvendigvis nok å bruke samme språk for å oppnå felles forståelse av begrepene.

Miller peker på nettopp dette problemet i artikkelen ”*Jeg sier hva jeg mener, men mener jeg hva jeg sier?*” [67]

”Interoperabilitet er lett – digitaliser noe innhold og legg det ut på en webside, for at noen skal finne det bruk denne coole tinge metadata, en beskrivelse av innholdet ditt i noe som kalles metatagger.”

Miller hevder at altfor mange har dette synet på interoperabilitet, også i miljøer der en skulle vente bedre! Han hevder at ”*slaget om metadata*” er nesten vunnet, men at dette ikke er nok til å oppnå interoperabilitet. Man kan ha muligheter til samvirke mellom ulike systemer, men manglende bruk av mekanismer for kontroll av vokabular kan redusere brukbarheten.

Registreringsformater for metadata kan variere sterkt fra system til system. Institusjoner og bedrifter bruker gjerne egne proprietære formater for å beskrive sine informasjonsressurser. Formelle standarder og formater kan ligge til grunn for tilpassinger til den enkelte institusjons registreringsbehov. Registreringsformatet som benyttes for Galleri Nor er et eksempel på dette. Det er en variant av NKKM’s Feltkatalogen. Med stort mangfold i ulike registreringsformater oppstår også større problemer med å ivareta interoperabilitet mellom disse. Integreerte systemer kan ha

behov for å samle informasjon ved å søke i ulike heterogene kilder eller de kan ha behov for å dele data ved å utveksle disse seg imellom.

Semantikk, struktur og syntaks

Disse tre begrepene er særdeles viktige for å muliggjøre interoperabilitet mellom metadata.

Semantikk sier noe om meningen med elementene. Det er viktig at applikasjoner som skal dele data benytter en klart definert semantikk for elementene som brukes

Metadata må organiseres i en struktur som er lesbar og forståelig for maskiner og tolkbar for mennesker

Syntaks handler om grammatikker for å håndtere semantikk og struktur. Vi trenger regler for å håndtere semantikk og struktur på en entydig måte

Metadatakonflikter

Konflikter kan oppstå mellom metadata. Dette kan være ulike typer konflikter som for eksempel navnekonflikter i form av synonymer eller homonymer, granularitetskonflikter som skyldes ulik detaljeringsgrad av metadata, syntakskonflikter i forhold til datarepresentasjon eller strukturelle konflikter. Konflikter kan for eksempel oppstå mellom ulike formater i et system eller mellom ulike systemer, men kan også forekomme innad i ett og samme system. Jeg skal ikke gå nærmere inn på ulike konflikter her, men det er viktig å være klar over at konflikter mellom metadata kan medføre interoperabilitetsproblemer og krever strategier for håndtering på en eller annen måte.

Kostnader ved metadata

Prosesen med å tilordne metadata til informasjonsobjekter er både tidkrevende og kostbart enten det nå foregår manuelt eller automatisk. Det kan være vanskelig å definere grensene for om dette er lønnsomt. En kan spesifisere problemstillingen ved å stille følgende spørsmål:

Hvor går grensen mellom for lite og nok metadata og hvor går grensen mellom nok og for mye metadata? Hva kan en oppnå ved å bruke gode metadata og hva vil konsekvensene være ved legge mindre vekt på metadata?

Problemstillingen må sees i sammenheng med hva sags informasjonsobjekter det er snakk om og hva som er intensjonen med dem – hva de skal brukes til. Det finnes ikke noen universell løsning på denne problemstillingen. En må finne sin løsning basert på hvilke behov en har og hvilke ressurser en har til rådighet.

4.1.3. Metadata for bilder

Standarder og formater som er i bruk er i stor grad generelle og kan brukes som rammeverk for egne tolkinger og utvikling av egne modeller. Enkelte formater er utviklet spesielt med tanke på visuelle informasjonsressurser [15] mens andre er mer generelle i den forstand at de er utviklet for å kunne beskrive ulike typer medier. Denne delen omtaler noen metadataformater som er i bruk i dag og en kort vurdering i forhold til om disse kan brukes for å beskrive bilder, da spesielt med bildene i Galleri Nor i tankene.

MARC

MARC er akronym for Machine Readable Catalogue eller Cataloguing og er internasjonalt anerkjent for utveksling av bibliografiske data mellom ulike systemer. Det er gjort en rekke oversettelser av formatet, også til norsk. NORMARC er et format for utveksling av bibliografiske data i maskinleselig form for utveksling mellom datasystemer i Norge.

Disse oversettelsene er ikke nødvendigvis direkte delmengder av MARC slik at det vil kreve konverteringer i forbindelse med utveksling av data mellom ulike land. Etterhvert finnes ca. 20 forskjellige oversettelser av formatet og UNIMARC er utviklet for å fungere som en slags bro mellom disse. Det kan da fungere slik at alle kan konvertere til og fra UNIMARC i stedet for at en oversettelse skal konverteres til alle oversettelsene for de land en skal utveksle data med.

Den Norske Katalogkomiteen, som står for den norske oversettelsen understreker viktigheten av at NORMARC [11] er et utvekslingsformat og ikke et registreringsformat.

En biblioteksystembruger må ikke nødvendigvis kunne noe om MARC /NORMARC og det kan være for komplisert å bruke som registreringsformat.

Systemer kan konstrueres slik at de kan importere fra og eksportere til MARC format tilpasset lokale behov. Lokale registreringsformat må da mappes direkte til/fra utvekslingsformat.

Den Norske Katalogkomiteen skisserer noen av fordelene ved å utveksle data slik:

”Ved import av data til eget lokalt system kan det rasjonalisere eget katalogiseringsarbeid å utnytte andres data.

Det er til fellesskapets beste at data kan leveres til samkataloger og andre felleskataloger, evt. til andre bibliotekers kataloger.

Ved skifte fra ett EDB-system til et annet kan NORMARC benyttes til å hente over de registrerte dataene.”

Slik har en da muligheter for å konvertere data mellom flere formater, men en risikerer samtidig å miste data som er særegne for ulike lokale formater.

MARC formatet er omfattende og svært detaljert og det brukes mest i tradisjonell biblioteksammenheng. MARC-formatet har høy status, er mye brukt og er et format det ofte refereres til i ulike metadataprojekter.



```

*000 c
*0010000052
*019 $bff
*100 $aWickstrøm, Jørgen$yfotograf$310000800
*245 $aJernbanestasjonen$hfortografi
*260 $c1893
*300 $a1 fotogr. : positiv$bs./h.$310000500
*690 $aTrondheim Sentralstasjon$cTrondheim$310015600
*715 $aTrondheim kommune Byarkivet$nnr. 76$310035400
*856 $u/bilder/lokal/vanlig_200302071448.jpg
^

```

Figur 4-1. Bilde med MARC-post fra bildebasen ved Trondheim folkebibliotek

MARC for bilder

Det er fullt mulig å benytte MARC-formatet for å beskrive bilder, men for å bruke dette som registreringsformat vil det kreve at indekserer har detaljert kunnskap om formatet, noe som i mange tilfeller kan virke uhensiktsmessig. Som utvekslingsformat derimot kan MARC-formatet være godt egnet for å utveksle data om stillbilder likeså vel som tekst, lyd og levende bilder. Figur 4-1 viser eksempel på et bilde med tilhørende MARC-post fra bildebasen ved Trondheim Folkebibliotek [75]. Dette er en relativt liten og enkel MARC-post, men den er komplisert nok til at en bør ha detaljkunnskap om formatet for å kunne forstå betydningen av de forskjellige feltene. Felt 245 viser tittelinformasjon der subfelt \$a viser objektets tittel mens \$h viser hvilket medium som benyttes. Felt 856 brukes når en ressurs eller en del av en ressurs er tilgjengelig elektronisk og sier noe om lokasjons og tilgangsinformasjon. Subfelt \$u angir en URI (Uniform Resource Identifier) mens \$q er feltet for å angi elektronisk format. For å angi geografisk navn kan en bruke felt 651 mens felt 752 kan brukes for hierarkiske beskrivelser av stedsnavn.

Dublin Core (DC)

Dublin Core Metadata Initiative (DCMI) startet som en workshop i Dublin, Ohio i 1995 og hadde som mål å forbedre standarder for oppdaging av dokumenter og dokumentlignende objekter på Internet. Initiativet samlet interessenter med ulik bakgrunn fra bibliotek, forskning på digitale bibliotek, innholdstilbydere og eksperter på tekst-markup.

Det opprinnelige DC element sett bestod av 13 kjerneelementer og initiativet fikk stor oppmerksomhet fra mange hold.

En ny workshop i 1996 med spesiell fokus på DC og bildebeskrivelser førte til en interessant diskusjon om hvor bilder kan plasseres i diskusjonen om dokumentlignende objekter og hvorvidt bilder er så forskjellige fra tekstdokumenter at de krever et eget sett elementer for beskrivelser. Konklusjonen på denne diskusjonen var at selv om tekst og bilder er svært forskjellige typer ressurser som trenger spesielle eksperter for å bli studert og tolket så er hovedkategoriene som brukes for klassifikasjon og søking omtrent de samme. I stedet for å utvikle et eget sett elementer for bildebeskrivelser førte denne diskusjonen til noen endringer i det opprinnelige DC element sett [13].

Enkelte elementnavn er modifisert for å gjøre dem mindre tekstsentrert og to elementer er lagt til de opprinnelige 13 elementer.

Elementene *Subject* og *Description* er separert, først og fremst fordi disse er forskjellige konsepter for bilder, men en atskillelse av disse to kan også være nyttig for andre typer media som for eksempel lyd.

Et enkelt felt for å håndtere rettigheter ble ansett for å være nødvendig i en kjernebeskrivelse. Elementet *Rights* ble innført samtidig som det ble foreslått hvordan feltet kan brukes, med verdier og definisjoner av disse for å utvide anvendelsesmuligheter. DC element sett, slik det tilbys nå, består av 15 kjerneelementer.

DC element sett har siden september 2001 hatt status som amerikansk nasjonal standard (NISO Standard Z39.85-2001) og har nå status som internasjonal standard (ISO 15836 - 2003) fra februar 2003. Standardiseringen omfatter kun de 15 kjerneelementene som utgjør DC element sett.

I tillegg tilbyr DCMI en spesifisering for alle metadatatermer som holdes vedlike av DCMI. DCMI Metadata Terms inkluderer fire forskjellige typer termer: elementer, element-raffinement, kodeskjema og vokabulartermer fra DCMI Type Vokabular [12]. En oversatt oversikt over alle tilgjengelige DCMI termer (19.11.2003) finnes i vedlegg A.ii.

DC Termer av type element

Et element er en egenskap ved ressursen, et attributt som ressursen kan ha så som tittel eller forfatter osv.

Elementer inkluderer, i tillegg til Dublin Core Metadata element sett 1.1 med 15 elementer som anbefales av DCMI, også andre elementer som anbefales av DCMI.

Elementer kan benyttes som de er eller de kan kombineres med kvalifikatorer som kan være av typen element-raffinement eller av typen kodeskjema.

DC Termer av type element – raffinement

Element -raffinement tilbys for å gi mulighet til å spesifisere betydningen av et DC element.

Et raffinement bør være av en slik art at det kun spesifiserer og ikke utvider det semantiske innholdet i elementet. Dette innebærer at en skal ha mulighet til å fjerne eventuelle raffinement, bruke det opprinnelige kjerneelement som om det ikke var

raffinert og fremdeles opprettholde en mening mellom element og verdi som er korrekt for elementet [76]. Et eksempel kan være elementet *coverage* som skal si noe om utstrekning eller omfang til ressursen. For å spesifisere om en snakker om omfang i tid eller rom kan en benytte de kvalifiserte elementene *coverage.spatial* og *coverage.temporal*.

I tillegg til raffinement av elementer som anbefales av DCMI inkluderer spesifikasjonen også termer for raffinement som er av en slik art at en implementasjonsenhet har demonstrert behov for termen uten at den nødvendigvis oppfyller kriteriet om at den må være nyttig på tvers av domener eller nyttig for oppdaging av ressurser. Disse termene har status som *conforming* (<http://dublincore.org/usage/documents/process/#conforming>).

DC Termer av type kodeskjema

Kodeskjema innføres som vokabularkontrollerende virkemiddel eller for å indikere hvordan verdien i et element skal tolkes. Et eksempel på kontrollert vokabular kan være å si at dette geografiske navnet er valgt fra en registrert tesaurus over geografiske navn. Et annet eksempel kan være datoformat, som kan uttrykkes på mange forskjellige måter. Ved å benytte et standardisert format som ISO 8601 (YYYY-MM-DD) fjerner en muligheten for misforståelser. Skjema kan benyttes også for å tolke elementverdien i kvalifiserte elementer som for eksempel *coverage.spatial schema=DCMI point*, som sier at dette elementet dekker omfang i rom angitt ved et punkt. Kodeskjema har status som *registrert*, det vil si at DCMI tilbyr informasjon om disse, men ikke nødvendigvis spesifikke anbefalinger. Kodeskjema kan brukes for å kvalifisere elementer som er raffinert så vel som de som ikke er det.

DC Termer av type vokabular – term

DCMI Type Vocabulary er en generell og domeneuavhengig liste over anbefalte termer som kan brukes for å velge verdi til DC elementet Type. Termene brukes for å identifisere ressursens natur eller sjanger.

DC elementene er repeterbare og valgfrie og muligheter for bruk av kvalifikatorer gjør dette til et fleksibelt format samtidig som det er enkelt og det legges vekt på at en ikke skal behøve å være ekspert for å ta formatet i bruk.

DC for bilder

DC er utviklet for å beskrive dokumentlignende informasjonsressurser. Dette innebærer også ikke tekstlige ressurser som bilder, lyd og multimedia. En grundig diskusjon av elementene i DC element sett for Galleri Nor finnes i vedlegg A.i mens en oppsummering av denne diskusjonen finnes i kapittel 6.3.9.

MPEG-7

MPEG-7 er en ISO/IEC standard utviklet av Moving Picture Experts Group (MPEG) [14]. Standarden kalles formelt *Multimedia Content Description Interface* og tilbyr et omfattende sett verktøy for å beskrive multimedia innhold. Et mål er å gjøre *audiovisuelt materiale like søkbart som tekst*.

MPEG-7 søker å tilby et fleksibelt og utvidbart rammeverk for beskrivelser av audiovisuelle data for å kunne rettes mot mange ulike applikasjoner i mange forskjellige sammenhenger og er ment å være så generelt som mulig. Dette innebærer at brukere av standarden kan definere sine egne sett beskrivende elementer innenfor rammen av basis strukturer og kategorier som standarden definerer.

Standarden er svært omfattende og generell, designet for beskrivelse av multimediaobjekter.

Visual Resources Association Data Standards Committee (VRA)

VRA er en amerikansk korporasjon som har eksistert formelt siden 1982. Før dette har organisasjoner som er medlemmer av korporasjonen hatt uformelle samarbeidsgrupper på tvers av organisasjoner i USA og Europa. Et VRA-prosjekt har utviklet et sett metadataelementer for å beskrive visuelle ressurser.

VRA Core Categories v. 3.0 [15] tilbyr et sett bestående av 17 repeterbare valgfrie elementer som kan brukes for å beskrive *verk* fra visuell kultur og bilder som dokumenterer disse.

VRA definerer verk slik:

"A Work is an object, i.e., any physical entity that exists or has existed at some time in the past or an ephemeral event (happening, performance art) that may be captured in physical form as a visual document of the original work. Examples include works of art, architecture, and artifacts or structures from material, popular, and folk culture. The Work Description Categories are intended for description of the original object, work, or event that is depicted in the visual document (e.g., photograph, slide, digital image) of that object, work, or event. The term "work" is used throughout to refer to the object or event described."

VRA tilbyr en oversikt over elementene der disse samtidig mappes til DC [10], til Categories for the Descriptions of Work of Art (CDWA) [16] samt til en tidligere versjon av VRA Core. Elementsettet inneholder kun enkle forskjeller fra DC for å ivareta både originale verk og surrogater av disse. VRA Core 3.0 er utviklet for å beskrive visuelle medier, og skiller seg lite fra DC når det gjelder beskrivende metadata. For å dekke behovet for beskrivende metadata for bildene i Galleri Nor kan en komme like godt ut ved å anvende et noe utvidet DC.

Outline klassifikasjonssystem

Dette er opprinnelig et amerikansk klassifikasjonssystem, Outline of Cultural Materials som har fått en viss utbredelse i Norden fra 1970-årene. Da det opprinnelige systemet ble tatt i bruk viste det seg at enkelte kategorier ble mye brukt slik at disse etter hvert ble store og det oppstod behov for finelig av kategorier. I Norge har det vært jobbet med dette i flere år og i 2001 kom en revidert og oppdatert versjon av Outline på norsk [17]. Outline er et delvis hierarkisk system som presenteres i tematiske grupper med tosifrede hovedkategorier og tresifrede emne-kategorier

Det gjøres rede for en teori der det påstås at en kan betrakte alle kulturelementer ut fra syv ulike aspekter og hvert av disse kan være med å danne hovedgrunnlaget for å definere kategoriene i Outline. De syv aspektene er:

Alle kulturelementer inngår i et *aktivitetsmønster* og kan for eksempel danne grunnlag for klassifisering for gruppa 484 *Reise*.

Visse aktiviteter forekommer bare av og til, under visse *omstendigheter*, for eksempel 527 *Fridager og helligdager*.

Enkelte aktiviteter utføres av en spesiell *gruppe personer*, det kan være personer som har status eller tilhører en sosial gruppe, for eksempel 793 *Presteskap*.

Noen aktiviteter er rettet mot et *objekt*, dette kan være en ting, et dyr eller et menneske, for eksempel 235 *Høsehold*.

Mange aktiviteter forbindes med bruk av spesielle *hjelpemidler*, for eksempel 411 *Våpen*.

Aktiviteter har som regel en *hensikt* eller et mål og det er som regel hensikten som bestemmer kategori som for eksempel 861 *Oppdragelsesmetoder*.

En aktivitet har vanligvis et konkret *resultat* som ofte er i overensstemmelse med hensikten, for eksempel 396 *Skipsbygging* og 681 *Straff*.

Outline for bilder

Med denne versjonen av Outline klassifisering for museum og arkiv følger veiledning i klassifisering av fotografier. Med utgangspunkt i de syv aspektene ovenfor velger forfatterne å betrakte fotografier ut fra følgende aspekter:

Fotografens konkrete *motiv*, for eksempel gruppe, portrett, landskap bygninger, redskaper osv.

Hva personer på bildet *gjør*, for eksempel arbeider, leker, spiser osv.

Under hvilke *omstendigheter* aktivitetene skjer, for eksempel fritid, selskap, skolegang, krigstid, 17. mai osv.

Type gruppe som utfører aktiviteter, for eksempel kvinner, lærere, prester osv.

Det gis et eksempel på en beskrivelse av et bilde av tre personer som er på sykkel / bærtur med ryggsekker og kasser, der syklene er framtrekkende.

Registreringsopplysninger i tillegg til person- og stedsnavn er at personene er på bærtur. I dokumentet om Outline er selve bildet ikke presentert, men denne følgeteksten:

På bærtur. 1941. Fotograf ukjent. Akershus fylkesmuseums billedsamling.

Beskrivelsen som presenteres ser slik ut:

Motiv: personer, sykler, kasser ryggsekker, landskap

Aktivitet: Bærtur

Omstendighet: Fritid/rekreasjon

Type gruppe: ingen spesiell

Ut fra disse stikkordene velges Outline kategoriene *Personer 159*, *Bærtur 222* og *Sykkel 493,8*

Forfatterne har her i utgangspunktet et ønske om å beskrive en fritidsaktivitet og tenker at bildet kan være interessant for en som er interessert i sykler.

De fire aspektene som er valgt som grunnlag for å betrakte fotografiene er interessante. Slik jeg ser det vil *Motiv* være et overordna aspekt De tre andre aspektene vil være nyttige i den grad motivet utgjøres av personer som utfører handlinger.

Feltkatalog for NKKM's EDB prosjekter

I regi av Norges Kunst og Kulturhistoriske Museer (NKKM) ble det i 1992 utviklet en feltkatalog for fotografier, kulturhistorisk gjenstandsmateriale, billedkunst, kunstindustri og museumsbibliotek. [54]. Arbeidet ble utført som en del av NKKMs edb-prosjekt etter oppdrag fra Humanistisk datasenter, Universitetet i Bergen.

Feltkatalogen skal være retningsgivende som standard for registrering. Den er generell og omfattende og har som mål å definere felt for de behov for registreringsdata som er kjent. Det er opp til de enkelte institusjonene å velge ut nødvendige felt og tilpasse formatet til sitt behov. I tillegg til å definere felt og subfelt omfatter feltkatalogen kodelister for sted og personroller og for typebestemmelse av fotomateriale. Feltkatalogen ligger til grunn for det registreringsformatet som brukes for Nasjonalbibliotekets fotodatabase Galleri Nor.

FRBR-modellen, en konseptuell modell for bibliografiske beskrivelser

FRBR-modellen definerer ikke et metadataformat, men er en modell som kan fungere som et rammeverk for bibliografiske beskrivelser. Modellen er omfattende og vil bare bli kort beskrevet her, men den har elementer som er interessante i en diskusjon om bildebeskrivelser og motivindeksering [59].

I følge denne modellen er målet med bibliografiske beskrivelser:

- Å finne alle manifestasjoner som inneholder
 - De verk som en gitt person eller korporasjon er ansvarlig for
 - De ulike uttrykksformer som et verk har
 - Verk om ett gitt emne
 - Verk i en gitt rekke
 - En bestemt manifestasjon når tittelen er kjent
 - En bestemt manifestasjon når identifikatoren er kjent
- Å identifisere
 - Et verk
 - Et verks uttrykk
- Å velge
 - Et verk
 - Et uttrykk
 - En manifestasjon
- Å få tak i
 - En manifestasjon

Modellen er bygget opp av tre hovedelementer. Det første er entiteter som, i denne modellen, består av tre komponenter. Disse er produktentiteter (selve produktene), ansvarsentiteter (de som har ansvar for produktene) og emneentiteter (emne for produktene). Det andre elementet er entitetenes egenskaper så som et verks tittel og et verks skaper og det tredje elementet er relasjoner mellom entiteter så som at en forfatter har skapt et verk eller at en person eier et eksemplar av et verk.

Produktentitetene er firedelt og omfatter verk, uttrykk, manifestasjon og eksemplar. Ansvarsentitetene er todelt og omfatter person og korporasjon. For emneentitetene skiller modellen mellom fire entiteter: begrep, objekt, hendelse og sted.

En entitets egenskaper er i denne modellen ment å fange de generelle og spesifikke egenskaper som kan knyttes direkte til produktet så som et verks tittel, skaper, form osv. Elementer som hører til disse egenskaper er elementer som vil være viktig å ta i betraktning i en metadatabeskrivelse av ethvert informasjonsobjekt.

Vel så interessant i denne sammenhengen er komponenten som består av emneentitetene. Dette omfatter her de elementer som brukes for å beskrive emne for det intellektuelle eller det kunstneriske produkt.

Modellen skiller disse i fire entiteter, begrep, objekt, hendelse og sted.

Begrep – er definert som og kan fungere som entitet i den grad det omfatter en abstraksjon som kan fungere som emne for et verk.

Objekt – er definert som en fysisk gjenstand som kan være emne for et verk.

Hendelse – er definert som historiske hendelser, epoker og tidsperioder som kan være emne for et verk.

Sted – er definer som stedsnavn, på og utenfor jorda, nåværende eller historiske, geografiske posisjoner eller geopolitiske jurisdiksjoner som kan være emne for et verk.

Ved å trekke noen paralleller fra denne modellens emneentiteter til de fire aspektene som ligger til grunn for Outline-beskrivelsene kan man si at disse ikke er så langt fra hverandre, men om den ene modellen skal kunne dekke alle aspektene ved den andre vil jeg anta at FRBR- modellen er noe mer generell og kan dekke flest behov.

FRBR-modellen åpner for at emneentitetene kan omfatte alle entiteter som inngår i modellen slik at også et verk, en manifestasjon, en person eller en korporasjon kan være emne for et verk.

Med denne modelleringen av entiteter følger også omfattende muligheter for å angi relasjoner mellom ulike komponenter i modellen uten at jeg vil gå nærmere inn på dette her.

Ad Hoc formater

Formelle formater og standarder er gjerne omfattende og kan være krevende å ta i bruk. Mange institusjoner og bedrifter har utviklet sine egne proprietære løsninger som dekker akkurat deres behov. Slike løsninger kan godt fungere utmerket internt i institusjonen, men kan skape problemer dersom man har behov for å bruke sine data i sammenheng med andre systemer eller aktører.

En spørreundersøkelse fra 1998/99 blant medlemsinstitusjoner i the Art Libraries Society (UK) – ARLIS har fokus på praksis for beskrivelser og indeksering av bilder i biblioteksamlinger [62]. Undersøkelsen omfatter 181 ulike organisasjoner og inkluderer akademiske, offentlige og private institusjoner. Når det gjelder bruk av katalogiseringsverktøy og standarder for registrering av fotografier viser undersøkelsen at mer enn 70 % av de forespurte benyttet sine egne *in-house* regler for dette arbeidet.

4.2. Tesaurus

En tesaurus er et verktøy for vokabularkontroll og kan brukes til å lede både indekserere og informasjonsøkere til brukbare termer for å beskrive konsepter og ressurser.

Kompleksiteten i ulike tesauri kan variere og det er ikke nødvendigvis enkelt å gi en fullstendig definisjon av hva en tesaurus er. To definisjoner som utfyller hverandre er framsatt av henholdsvis Guinchat / Menou og Rowley [21].

Guinchat / Menou [22]:

”Tesauri er verktøy som består av et kontrollert termsett lenket sammen med hierarkiske eller assosiative relasjoner og som markerer nødvendige ekvivalente relasjoner med termer fra naturlig språk og konsentreres om et spesifikt kunnskapsområde.”

Rowley [23]:

“En thesaurus er en kompilering av ord og fraser som viser synonymer og hierarkiske og andre relasjoner og avhengigheter. Funksjonen er å tilby standard vokabular for lagrings og gjenfinningsystemer. Hovedmål for thesaurus er å ha kontroll over terminologien i indeksering og til å gi støtte ved søking ved å gjøre søkeren oppmerksom på valgte indekseringstermer.”

Den første internasjonale standard for konstruksjon av monolingual thesaurus kom i 1974 [69] og ifølge Aitchison [68] var det i sammenheng med dette at man først begynte å betrakte thesaurus som et viktig verktøy i indekseringssammenheng.

Nå finnes flere standarder og retningslinjer for utvikling av både ett- og fler-språklig thesaurus.

Britisk standard, BS 5723 [57], definerer thesaurus i forhold til formål og struktur og sier at en thesaurus tjener fire hovedformål. En thesaurus skal:

- Kontrollere termer for indeksering, tilby en måte å oversette naturlig språk til et mer begrensa språk for indeksering og søk/gjenfinning.
- Sikre konsistens i praksis mellom ulike indekserere i samme bedrift eller i ulike bedrifter som samarbeider.
- Begrense antall termer som må tilordnes et dokument – term som tilordnes bør representere konseptet slik det er beskrevet av forfatter så spesifikt som mulig.
- Tjene som søkehjelp for gjenfinning, inkludert gjenfinning fra fritekstsystemer.

Gjennom sin struktur viser en thesaurus synonymer, hierarkiske og andre relasjoner mellom termene som til sammen utgjør et indekseringsspråk.

En thesaurus har tre hovedegenskaper [21]. Disse er:

- **Vokabularkontroll**
- **Tesaural sammenheng (relasjoner)**
- **Thesaurus display**

4.2.1. Vokabularkontroll gjennom thesaurus

I tesauri representeres konsepter av indekseringstermer som er enten foretrukne eller ikke foretrukne termer. Foretrukne termer brukes ved indeksering og søking mens ikke foretrukne termer skal lede til foretrukne termer.

I følge BS 5723 [57] er det to hovedmåter å oppnå vokabularkontroll i thesaurus:

- Termer er begrenset i omfang for utvalgt betydning. Hver term i thesaurus er begrenset til den betydningen som best tjener indekseringsbehovet. Strukturen på thesaurus, muligheten for å vise hierarkiske relasjoner, indikerer ofte betydningen av en term. Dersom dette ikke er nok kan en legge til en definisjon

eller en notis (Scope Note (SN)). Disse kan også angi betydninger av termer som en finner i naturlig språk, men som er utelukket for indekseringsformål.

- For konsepter som kan representeres av to eller flere synonymer, velges normalt en av disse, som brukes konsistent som foretrukket term i indekseringen. Synonymer som kan fungere som brukers aksesspunkt skulle ha referanser til foretrukket term.

4.2.2. Relasjoner mellom termer i tesaurus

Tre generelle klasser

- Ekvivalensrelasjon
- Hierarkisk (hel-part) relasjon
- Assosiativ relasjon

Ekvivalensrelasjoner

Ekvivalensrelasjoner dekker to typer termer, synonymer og kvasisynonymer. Synonymer er termer der meningen kan betraktes som den samme i ulike kontekster mens kvasisynonymer er termer der meningen generelt betraktes som forskjellig i vanlig bruk, men som behandles som synonymer for indekseringsformål, for eksempel termene hardhet og mykhet.

Ekvivalensrelasjoner angir relasjoner mellom foretrukne og ikke foretrukne termer i indekseringsspråk. Relasjonene angis ved å bruke USE og UF.

Foretrukket term USE FOR (UF) ikke foretrukket term.

Ikke foretrukket term USE (USE) foretrukket term.

Hierarkiske relasjoner

En hierarkisk relasjon er en basisrelasjon som skiller en systematisk tesaurus fra andre organiserte lister av termer, for eksempel Subject Heading lister

Termpar representeres i sin superordinale eller subordinale status

Superordinat representerer en helhet, subordinat representerer et medlem eller en del

Superordinat representeres av Broader Term (BT), subordinat representeres av Narrower Term (NT)

BT og NT representeres gjensidig, for eksempel:

Hovedsteder **BT** Byer

Byer **NT** Hovedsteder

BS 5723 [57] identifiserer tre relaterte situasjoner som representerer hierarkiske relasjoner: generisk, hierarkisk og polyhierarkisk relasjon. *Generisk relasjon* er en relasjon som identifiserer relasjoner mellom en klasse / kategori og deres medlemmer. Generiske relasjoner har hierarkisk styrke, det vil si at alt som gjelder for en gitt klasse gjelder også for alle subklasser under denne. *Hierarkisk (hel part) relasjon* er relasjoner som dekker begrensede antall klasser av termer der navnet til delen impliserer navnet på det hele uavhengig av kontekst, slik at termer kan ordnes som logisk hierarki.

Det er identifisert kun fire situasjoner der hel-part relasjoner kan betraktes som hierarkiske ellers er de assosiative [21]. De fire situasjonene er:

- Systemer og organer i kroppen
- Geografiske lokasjoner
- Disipliner eller områder for discourse
- Hierarkiske sosiale strukturer

En *Polyhierarkisk relasjon* oppstår når et konsept hører til mer enn en kategori

Assosiative relasjoner

Angir relasjoner som ikke er hierarkiske eller ekvivalente, termene er mentalt assosiert i en slik grad at lenkene mellom dem gjøres eksplisitt i thesaurus og vil synliggjøre alternative termer som kan brukes til indeksering og gjenfinning. Disse relasjonene representeres ved Related Term (RT).

Assosiative relasjoner er de som er vanskeligst å definere og derfor også å bestemme mellom et termpar.

BS 5723 [57] tilbyr en generell retningslinje som slår fast at en av termene alltid bør impliseres i henhold til de referanserammer som er felles for brukere av en indeks såfremt den andre brukes som indekseringsterm. Standarden identifiserer to brede termkategorier som bindes av assosiative relasjoner. Den ene er termer som hører til samme kategori og som ofte er termer med overlappende mening som båt og skip. De kan defineres presist, men er ofte brukt løselig og kan utveksles. Den andre er termer som tilhører forskjellig kategori. Dette kan for eksempel være termpar som er lenket sammen av årsaksavhengighet som for eksempel lidelse og sykdom. Andre eksempler på termpar som tilhører forskjellig kategori kan være konsepter relatert med deres egenskaper, konsept relatert til opprinnelse, en handling og produktet som er resultat av denne.

4.2.3. Display av termer i tesaurus

Det er tre basismetoder for display av termer i tesaurus:

- Alfabetisk visning med scope notes og relasjoner for hver term
- Systematisk visning med alfabetisk indeks
- Grafisk visning med alfabetisk indeks

Alfabetisk visning

Alle indekstermer, foretrukne og ikke foretrukne, er organisert i en enkelt alfabetisk sekvens. Underordnet informasjon, Scope Notes (SN) og relasjoner listes under foretrukne termer på følgende måte [21]:

- 1 SN eller definisjon
- 2 UF refererer til ekvivalente ikke foretrukne termer
- 3 TT refererer til topptermer
- 4 BT refererer til bredere termer
- 5 NT refererer til smalere termer
- 6 RT refererer til relaterte termer

Ikke foretrukne termer følges kun av referanser (USE) til foretrukket term
multinivå hierarkisk display er mulig

Den alfabetiske formen er den enkleste å utvikle, men den kan synes å være mangelfull fra brukers syn da det vil være vanskelig å ha oversikt over alle bredere og smalere termer fra en enkelt posisjon.

Ekstra relevant informasjon kan legges til alfabetisk display eks. Toppterm i hierarki som et spesifikt konsept hører til

Likewise kan en vise nivå for super / subordinasjon ved å bruke BT1, BT2, NT1, NT2 osv.

Systematisk visning

Systematisk organisert thesaurus bør ha to deler

- Kategorier eller hierarki av termer arrangert i forhold til betydning og logiske sammenhenger
- Alfabetisk indeks som leder bruker til riktig del av systematisk seksjon

Lenking mellom disse vedlikeholdes gjennom notasjon.

Systematisk display er bra både for indekserer og bruker, den gir et fugleperspektiv over tema og setter det i kontekst med hele emnefeltet.

Grafisk visning

Viser indekstermer og relasjoner i form av todimensjonale figurer, supplert av alfabetiske seksjoner. Trestrukturer og pilgrafer er de mest brukte strukturer for å framstille tesauri grafisk. En slik grafisk framstilling har den fordel at den viser en overordnet oversikt over konseptets omgivelser men ulempen er at den ikke viser ekvivalente termer eller SN, den skiller ikke mellom hierarkiske og assosiative relasjoner.

4.2.4. Prinsipper for valg av foretrukne termer

Retningslinjer for valg av foretrukne termer er hentet fra Tim Carvens introduksjonstutorial til tesauruskonstruksjon.[70]

Retningslinjer	Eksempler
Bruk	Cooking UF cookery (cooking er mer brukt)
Bredde	Plastics UF polyethylene (plastic mener all plastikk hvorav polyethylene er en variant)
Entydighet	American library association UF ALA (ALA kan bety noe annet)
Colokasjon	Railway stations UF train stations (I en alfabetisk sekvens vil railway station forekomme nært railways og andre termer relatert til railways)
Konsisitet	Muckrakers UF muckraking movement (ett ord i stedet for to)
Flertallsform for tellbare objekt	Geese UF goose (gås er tellbare)
Intern konsistens	Hvis du har bestemt å bruke latinske plantenavn, gjør det konsistent
Ekstern konsistens	Du kan foretrekke enkelte termer fordi det samsvarer med bruk i andre thesaurus

Tabell 4-1 Retningslinjer for valg av foretrukne termer (uf = use for)

4.3. Gazetteers

En gazetteer kan være en liste over geografiske navn sammen med informasjon om geografisk lokasjon og eventuelt annen beskrivende informasjon. Et geografisk navn er et korrekt navn for et sted og en navnetype. Navnetype sier noe om stedets natur. Geografiske steder og forekomster inkluderer politiske og administrative områder, naturfenomener, menneskebygde forekomster og upresise områder. I tillegg til geospasiale relasjoner mellom forekomster kan relasjoner uttrykkes eksplisitt som for eksempel *erDelAv*, *harDeler* eller andre typer relasjoner.

En gazetteer kan brukes som en frittstående referanse-tjeneste og som en innebygget søkehjelp i informasjonstjenester (*Indirect Spatial Referencing*). Ved å bruke gazetteer ved indirekte spatial referering kan en, ved å starte med et stedsnavn, finne informasjon som er beskrevet enten ved hjelp av geografiske navn eller geospasiale koordinater. [3] I tillegg til informasjon om geografiske navn, navnetyper og lokasjoner kan en knytte annen informasjon til forekomster i gazetteeren. Hva slags type informasjon en gazetteertjeneste skal tilby kan avhenge av hva man ønsker å bruke gazetteeren til. Informasjon kan inkluderes som attributter til den geografiske forekomsten, men en kan også tenke seg muligheten for å integrere gazeetters med andre lignende systemer slik at de kan utfylle hverandre.

Lister med geografiske navn kan være tilgjengelige som hierarkiske termsett / tesauri som for eksempel *the Getty Thesaurus of geographic names* [66]. Relasjoner som kan uttrykkes eksplisitt mellom geografiske forekomster er da innebygget i den hierarkiske strukturen. Ofte er dette administrativt oppdelte hierarki. Relasjoner mellom forekomster kan også modelleres og implementeres som eksplisitte attributter til selvstendige poster for forekomstene i en gazetteer [65]. Hvilken modell en skal velge for en gazetteer kan være avhengig av hvilke rolle gazetteeren skal ha. En tesaurusmodell kan fungere utmerket for å vise hierarkiske relasjoner slik som *the Getty Thesaurus of geographic names* [66], men kan være mindre hensiktsmessig for anvendelsesområder der en legger vekt på deling av informasjon. Tesauri er som regel domenespesifikke og det kan, på grunn av dette samt den innebygde strukturen, være vanskelig å kombinere data fra flere lignende kilder. En metadatalignende modell som knytter eksplisitte relasjoner til selvstendige poster kan være mer fleksibel i så måte. Relasjoner som uttrykkes eksplisitt i en slik modell kan utnyttes til å generere hierarkiske strukturer og kan dermed tilby de samme egenskaper som en geografisk tesaurus.

Generelt for gazetteers og lignende systemer er at det finnes ulikheter i tilnærminger til beskrivelse av geografiske steder. Bortsett fra alternative kilder for geografiske navn i seg selv har det ikke foregått noe standardisering på dette området. Dette ble avdekket som et problemområde under utvikling av ADL Gazetteer (kapittel 6.2) som inneholder data fra flere ulike amerikanske kilder. Dette førte i sin tur til utvikling av ADL Gazetteer Content Standard (ADL GCS) (kapittel 6.2.1) som kan være veiledende for utvikling av nye gazetteers.

Alle navngitte geografiske steder kan representeres i en gazetteer. Detaljeringsgraden på innholdet kan variere fra gazetteer til gazetteer og dette kan også sees i sammenheng med hva man skal bruke gazetteeren til. For eksempel må man ta stilling til om man skal definere navngitte bygninger som geografiske forekomster i en gazetteer. Noen av eksemplene under overskriften *Hva en gazetteer kan gjøre for oss* i dette kapitlet, viser at dette kan være nyttig.

I Norge spiller Statens kartverk en viktig rolle i forhold til forvaltning av stedsnavn og navnetyper. Norsk *Lov om stadnamn* fastsetter at det skal opprettes et sentralt stedsnavnregister. Dette er lagt til Statens kartverk (SK) og kalles Sentralt stedsnavnregister (SSR).

SSR bygger opprinnelig på navnedatabasen over alle navn som forekom på hovedkartserien Norge 1:50 000 i 1989, men er senere oppdatert med navn også fra andre kartserier (Økonomisk kartverk, sjøkart etc). Registeret oppdateres kontinuerlig av SK med alle navnevedtak som gjøres i henhold til lov om stadnamn - av f.eks vegvesenet, kommunene, SK og andre vedtaksinstanser.

Lov om stadnavn trådte i kraft 01.07.1991 og Sentralt stedsnavnregister inneholder i underkant av 400 000 navneforekomster.

SSR skal inneholde alle stedsnavn som er innmeldt til registeret, hvor skrivemåten er vedtatt etter *Lov om stadnamn* (se loven §10 og forskriftene til loven kap VII [77]).

I tillegg kan registeret inneholde andre navn som var i offentlig bruk før *Lov om stadnamn* trådte i kraft 01.07.91.

I tillegg til opplysninger om skrivemåten av navn, inneholder registeret blant annet informasjon om språk og språkform som norsk, nordsamisk, lulesamisk, sørsamisk og finsk, administrativ tilhørighet som fylke, kommune og sokn, koordinater, objekttype (hva navnet er navn på), status for skrivemåten, vedtaksinstans, vedtaksdato, samt forekomst på kart, skilt osv. [31]

Hva en gazetteer kan gjøre for oss

En gazetteer kan støtte flere funksjoner. Noen eksempler på hva en gazetteer kan gjøre:

- svare på spørsmål som *Hvor er ...?*
- typebestemmelse av sted
- oversette mellom geografiske navn og lokasjoner slik at en kan finne samlingsobjekter ved å matche geografisk avtrykk for et geografisk navn med avtrykk for samlingsobjekter
- finne avstanden mellom geografiske forekomster ved å regne ut avstand mellom geografiske koordinater
- en toveis oversettelse kan finne sted, et geografisk navn kan oversettes til et geografisk avtrykk og et slikt avtrykk kan oversettes til et sett av geografiske navn i området
- når geografiske navn også er kategorisert til typer kan også spesifiserte typer av steder identifiseres innenfor et område.

I tillegg til et minimum sett informasjon for hvert sted som geografisk navn, navnetype og geografisk avtrykk kan en gazetteer:

- tilby flere variable navn for samme lokasjon
- inneholde forskjellige spatiale representasjoner for samme sted så som punkt, avgrensende boks, detaljerte grenser eller spatiale avtrykk fra forskjellige kilder og målbare detaljer
- peke til andre typer informasjon
- peke til relaterte gazetteer-innganger med lenker som *er del av* og *er hovedstad i*
- kombinere informasjon fra varierte kilder med henvisning til kilden for hver bidragsyter

Trykte gazetteers

Tradisjonelt kjenner de fleste av oss til gazetteers slik de gjerne forekommer i trykte atlas. Trykte atlas har ofte en egen gazetteer-seksjon der en kan slå opp geografiske navn og finne lokasjonsinformasjon. I tillegg til informasjon om lokasjon kan slike seksjoner i ulike atlas tilby forskjellig annen informasjon om steder og deres natur eller egenskaper. For eksempel kan dette være informasjon om historie, populasjonsdata, fysiske data, uttale av navn.

Digitale gazetteers

En gazetteer i kontekst av en online informasjonstjeneste kan sees på som en kunnskapsbase som minimum definerer geografiske navn ved spatial representasjon. Kjerneelementene i digitale gazetteers er selve stedsnavnet, hvilken type sted det er og et geografisk avtrykk som representerer stedets lokasjon og mulig utvidelse.[24]

En digital gazetteer kan defineres som en geospatial dictionary av geografiske navn med kjernelementer for en forekomst består av:

- et geografisk navn for et sted – et sted kan ha flere navn
- en lokasjon – i form av koordinater som representerer et punkt, en linje eller et areal, eller i form av en relasjon til en annen forekomst
- en navnetype – valgt fra et type-skjema over kategorier for steder

Digitale gazetteers kan brukes på samme måte som trykte gazetteers. En kan slå opp på eller søke på geografiske navn og finne informasjon om lokasjon og eventuell annen informasjon om steder.

Digitale gazetteers får nye roller å spille i informasjonsarkitektur for geo-lokasjons aksess til informasjon og data der vanlig brukte stedsnavn, som vi bruker for å identifisere en lokasjon, kan brukes til å få direksjoner fra et datasystem for å finne fram eller få informasjon om et sted eller et område. Strukturer for interoperabilitet gir løfter om nettverkstilgang til forskjellige sett av gazetteer-data for å øke tilgjengelighet til global, lokal og spesialisert beskrivelse av steder og deres lokasjon [24].

Det finnes flere veletablerte gazetteer webtjenester samt et økende antall nye. Det finnes også flere stedsnavntjenester som ikke kaller seg gazetteers, men som tilbyr lignende tjenester, for eksempel gule sider for kommersiell lokasjon, digitale atlas, stedsnavnaksess til online kart og ruting service. Websteder tilbyr pekere til online gazetteers.

Karakteristisk for gazetteers og lignende tjenester er at der er ingen standardisering av noen aspekter vedrørende recordformat, innhold og tjenestegrensesnitt.

Det er vanskelig å dele data mellom disse ressursene. Standardisering for offisielt arbeid med gazetteers har vært fokusert på å etablere autoritative navneformer for ikke-kommersielle steder. Offisielle gazetteerbyråer har ikke vært involvert i å skape autoritative grenserepresentasjoner for yttergrenser for steder. Deres formål har vært tilstrekkelig tjent med punktlokasjoner som skiller et sted fra et annet.

Hver gazetteer og lignende tjenester har skapt sitt eget sett av kategorier for stedsnavn, deres eget recordformat og måter å beskrive detaljer [24].

Digitale gazetteertjenester

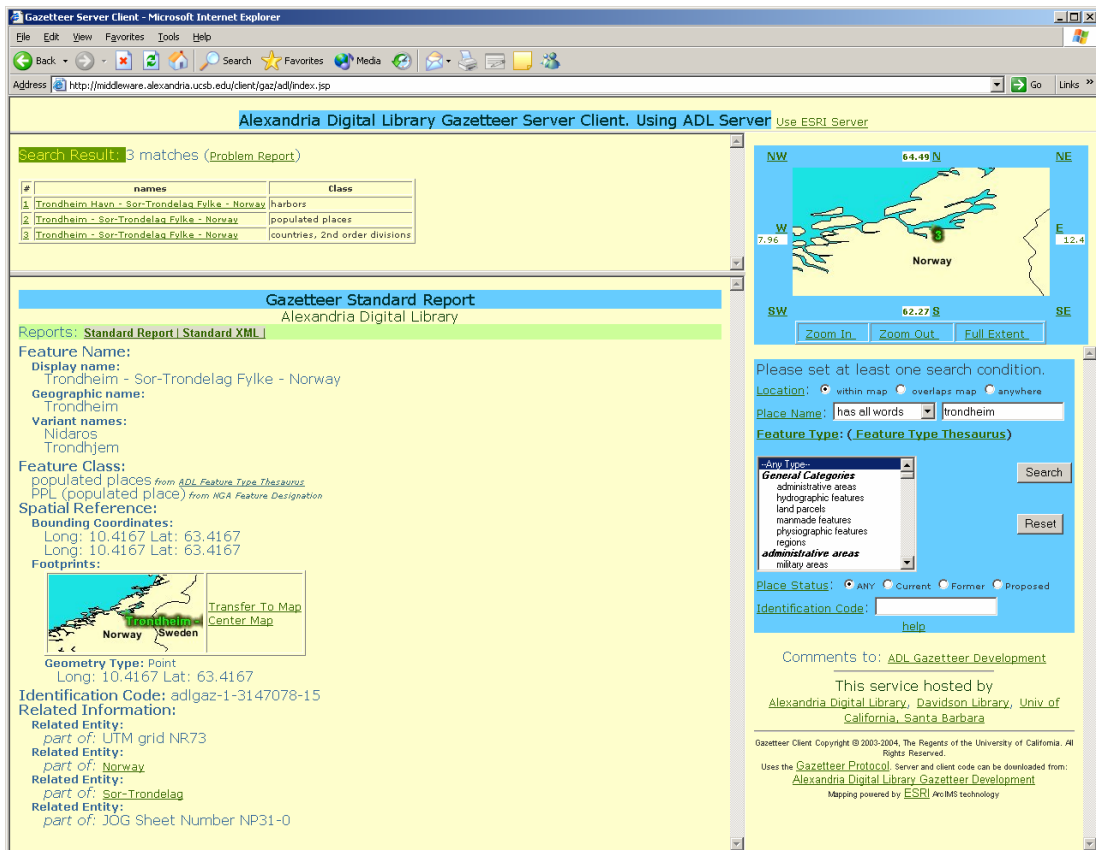
Det er utviklet en rekke digitale gazetteertjenester og mange av disse er nå fritt tilgjengelige på www. [25][28]. Felles for disse er at en kan søke etter geografisk navn og finne lokasjonsinformasjon om aktuelle navn. Lokasjonsinformasjon er som regel i form av geografiske koordinater, grafisk framstilling av merke på kart eller begge deler. I tillegg tilbyr de ulike tjenestene forskjellig annen informasjon om aktuelle sted.

ADL Gazetteer

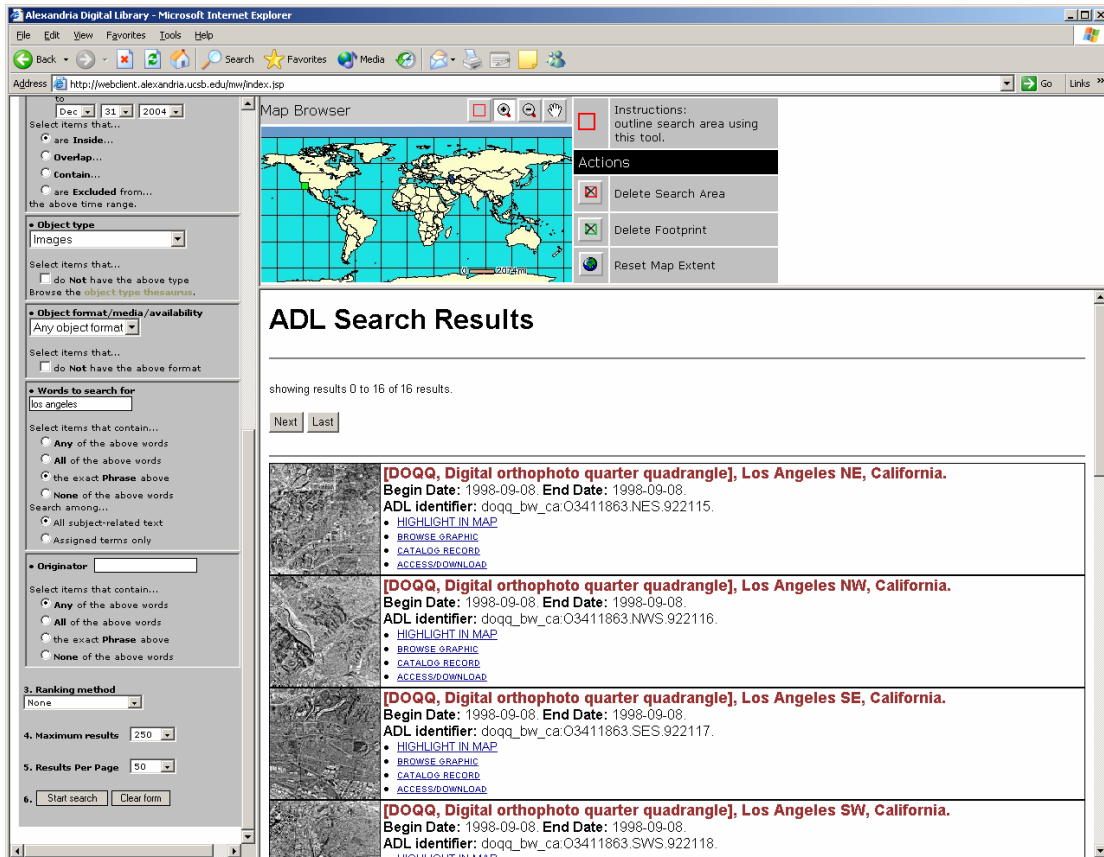
Alexandria Digital Library Project (ADL) [71] er et prosjekt ved University of California, Santa Barbara. Mål for prosjektet er å utvikle et globalt distribuert georeferert digitalt bibliotek. Ett av prosjektets hovedfokus er på formalisering av strukturer og tjenester for systemer for kunnskapsorganisering (KOS) og integrering av KOS i digitale bibliotek. Oppgaver knyttet til utvikling av digitale gazetteers har vært og er fortsatt en viktig del av prosjektet. Arbeidet inkluderer utvikling av en innholdsstandard og en navnetypesaurus for digitale gazetteers. ADL Gazetteer er tilgjengelig både som selvstendig webservice og som en integrert komponent av ADL [27].

ADL Gazetteer tilbyr brukeren flere alternativer for å søke etter stedsnavn og / eller navnetype i forhold til et kartområde. Kartområdet kan defineres av bruker. Navnetype må velges fra en liste i brukergrensesnittet og en navnetypesaurus er tilgjengelig for oppslag via en lenke.

Som integrert komponent i ADL kan du søke etter stedsnavn og få aktuelle forekomster markert på kart. I tillegg kan bruker her velge å søke etter ressurser i en eller flere samlinger. Bruker kan regulere flere elementer som vil påvirke søkeresultatet som for eksempel geografisk omfang, tidsperiode, type ressurs, format osv. ADL Gazetteer er nærmere omtalt i kapittel 6.2.



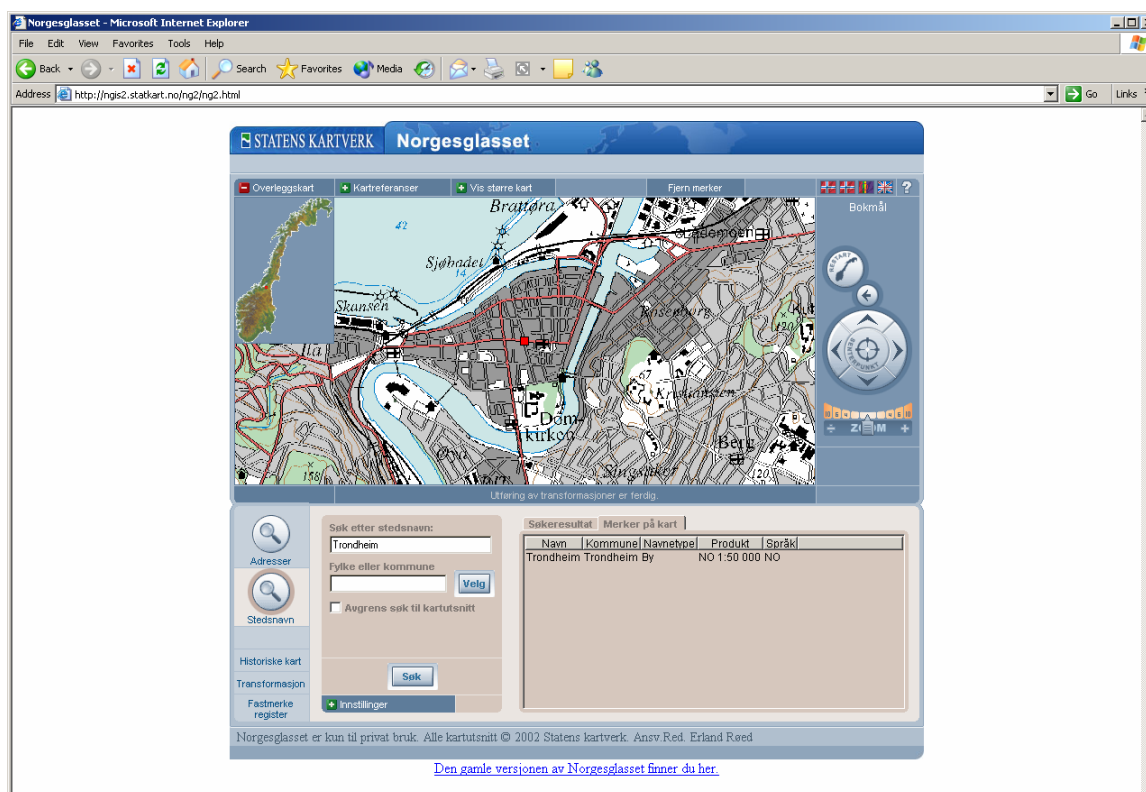
Figur 4-2 Skjerm bilde fra ADL Gazetteer som selvstendig web-tjeneste



Figur 4-3 Skjermbilde fra ADL med integrert ADL Gazetteer

Norgesglasset

Statens Kartverk [29] tilbyr tjenesten Norgesglasset der en kan søke etter norske stedsnavn og finne disse samt informasjon om kommune, navnetype, språk og kartprodukt [30]. I tillegg til å søke etter stedsnavn kan en gjøre detaljerte søk etter gateadresser. En kan velge å vise søkeresultater på kart som kan justeres i forhold til detaljeringsgrad.



Figur 4-4 Skjerm bilde fra Norgesglasset

Andre digitale gazetteertjenester

The National Imagery and Mapping Agency (NIMA) Geonames Server [28] er en av hovedkildene for informasjon i ADL gazetteer. NIMA tilbyr en tjeneste der en kan søke etter stedsnavn, navnetype og/eller lokasjon i form av en avgrensende boks som bestemmes ved å angi geografisk koordinater. Et søkeresultat gir informasjon om stedsnavn, navnetype, navnevarianter og lokasjonsinformasjon i form av geografiske koordinater. Informasjon om relasjoner mellom forekomster er begrenset til en angitt regiontilhørighet for forekomsten. Regionene er omfattende, For eksempel region 1, som Trondheim tilhører, omfatter Amerika og Vest-Europa

The Getty Thesaurus of Geographic Names [66] tilbyr muligheter for søk etter geografiske navn og navnetyper. Et søkeresultat viser en grafisk framstilling av et hierarki der aktuelle forekomst har sin plass og der en kan navigere mellom lenker til informasjon om ulike forekomster i hierarkiet.

(USGS) Geographic Names Information System (GNIS) [25] tilbyr en rekke tjenester i tilknytning til å søke på geografiske navn og navnetyper.

5. Systemer og samlinger

Denne delen gir en oversikt over noen prosjekter som er relevante for oppgavens problemstilling. *From Digital Library to multimedia application* og *Schoolnet* er prosjekter som, ut fra forskjellige behov, utforsker bruken av Dublin Core som metadataformat for bilder.

Alexandria Digital Library (ADL), *Adept*, *GeoXwalk* og *Textual – Geospatial Integration Services for the National SMETE Digital Library* er prosjekter som, på forskjellig vis er opptatt av hvordan en kan utnytte geografisk referering i forvaltning av informasjonsressurser.

Enkelte av nasjonalbibliotekets samlinger er omtalt til slutt i dette kapittel.

ADL og Galleri Nor er spesielt viktige for denne oppgaven. Disse er nærmere omtalt i kapittel 6.1 og 6.2 og er derfor kort omtalt i dette kapittel.

5.1. From Digital Library to multimedia application

Vercoustre og Paradis (V&P) har i sin artikkel *Metadata for Photographs: From Digital Libraries to Multimedia Applications* [42] redegjort for hvordan de har brukt Dublin Core som et rammeverk for å beskrive fotografier til et undervisningsprosjekt i form av en CD-ROM om restaurering av gamle bygninger. Behovet for metadata var tilstede for å gjøre det mulig å finne fotografier som elementer til automatisk genererte websider. Som et argument for å benytte Dublin Core peker V&P på nytten av å anvende standardiserte metadatabeskrivelser for å fremme interoperabilitet og muligheter for gjenbruk av fotografiene for andre prosjekter. De har tatt i bruk kvalifikatorer og skjema for å oppnå et utvidbart metadataskjema og for å tilpasse formatet til sitt behov. De konkluderer med at Dublin Core, med visse tilpasninger, er brukbart for å beskrive fotografier.

5.2. SchoolNet

Metadata Searching in a Multimedia Database Environment er tittel på et kanadisk prosjekt for utvikling av metadataskjema for beskrivelser av multimediaobjekter for Internet som er tenkt brukt for undervisning i det kanadiske SchoolNet [43].

Prosjektgruppa har designet et konseptkart som et rammeverk for databasesøking og har brukt dette for å søke etter relevant informasjon i et utvalg forskjellige databaser.

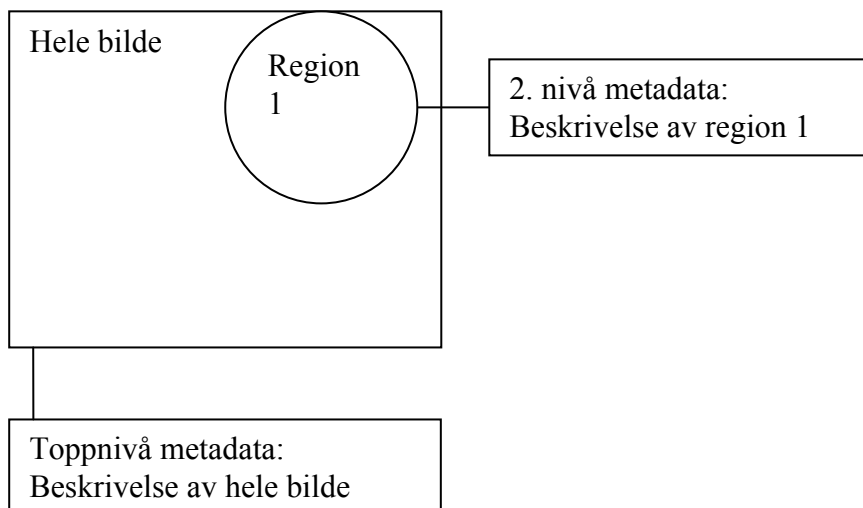
En oppsummering av søkeresultatene slår fast at DC er en klar leder for ressursbeskrivelser for bruk i nettverk, mye på grunn av utstrakt aksept og på grunn av dets muligheter for utvidelser. Jane Hunter Fra Distributed Systems Technology Centre (DSTC) Pty Ltd i Australia er en anerkjent spesialist på indeksering, querying og browsing av multimedia databaser og teknikker for ressursoppdaging for multimedia. Hunter har bidatt til dette prosjektet med en rapport der hun definerer metadataskjema basert på DC for beskrivelser av multimedia objekter [43]. Hunter har tidligere gjort mye arbeid knytta til video-indeksering, noe som, ifølge prosjektgruppa, taler sterkt for å benytte et utvida DC for å beskrive multimedia objekter. Videre antar de, på bakgrunn

av forskning på området, at DC vil fortsette å definere standarder for denne type objekter.

Metadataskjema for bildebeskrivelser er på to nivå med en toppnivå beskrivelse knyttet til bildet som et hele og valgfrie 2. nivå beskrivelser som kan assosieres til avgrensede regioner av bildet som en kan ønske å knytte informasjon til.

2. nivå beskrivelser består av 5 de DC elementene title, identifiser, description, relation som er av typen IsPartOf til hele bildet og elementet coverage som her anbefales brukt på en fiffig måte slik at en kan bruke metadata også til å konstruere bildekart eller ”klikkbare” bilder.

Eksempel på en bildebeskrivelse av et bilde fra the William Boag Photograph Collection i the John Oxley library in Brisbane, Qld. Det er et digitalt surrogat for et fotografi av en nybyggerfamilie [43].



Figur 5-1 Bildebeskrivelse på to nivåer



Figur 5-2 Bilde av nybyggerfamilie. Region 1 er en sirkel rundt kvinnens hode til høyre i bildet.

Toppnivå Metadata for bildet som helhet

Toppnivå skjema består av de 15 DC elementer med nødvendige kvalifikatorer i tillegg til domenespesifikke elementer knyttet til at bildene skal brukes i undervisning.

dc>Title: A selector and his family, probably in the Beenleigh district, 1872

dc:Creator: William Boag

dc:Subject: Photograph collection - Queensland

dc:Description: The difficulties faced by a family in the Queensland bush included poor roads, an unreliable mail service and dense, vine-matted scrub.

For many years, a selector's staple diet was salted meat (salt horse) and pumpkins. For several months, a woman and her children might be alone in their stringy-bark hut while her husband went off to split shingles or to earn extra money on a cattle property.

dc>Date.created: 1872

dc>Date.recordCreated: 1996 *dc>Date.placedOnline:* 1997

dc:Publisher: State Library of Queensland

dc>Type: image.photograph

dc:Format: image/jpeg

dc:Format.fileSize: 50.6Kb

dc:Format.dimensions: 672 x 512

dc:Format.colorpalette: grayscale

dc:Identifier: <http://archive.dstc.edu.au/RDU/SLQ/boag/20248.jpg>

dc:Source: BOAG negative no. 906

dc:Language: en

dc:Relation.isPartOf: <http://www.slq.qld.gov.au/jol/boag.htm>

dc:Relation.hasParts: region1

dc:Relation.hasFormat: <http://archive.dstc.edu.au/RDU/SLQ/boag/20248.gif>

dc:Coverage: Beenleigh region, Queensland, 1872

dc:Rights: <http://www.slq.qld.gov.au/cright.htm>

etc:Audience: Secondary school teachers, students

etc:Grade: grades 7-13 (approx age 12-17), post-secondary (approx age 18+)

etc:Discipline: History, Australia

2. nivå Metadata for Region 1

dc>Title: The mother of the family

dc:Identifier: <http://archive.dstc.edu.au/RDU/SLQ/boag/20248.jpg#region1>

dc:Description: Although only in her mid-30's the mother of this family appears much older. The difficulties of bringing up 6 children in the bush, the frequent loneliness and the constant need to guard the crop against marauding birds, bandicoots and scrub pademelons (small forest wallabies), have clearly taken their toll.

dc:Coverage.circle: 496, 256, 24

dc:Relation.isPartOf: <http://archive.dstc.edu.au/RDU/SLQ/boag/>

5.3. ADL og ADEPT

Alexandria Digital Library Project er et prosjekt ved Universitetet i California, Santa Barbara. Prosjektet startet i 1994 med utvikling av Alexandria Digital Library (ADL) [4], et distribuert, digitalt bibliotek med samlinger av georeferert materiale og tjenester for å få tilgang til samlingene. Biblioteket er distribuert fordi komponentene kan være spredt på Internett likeså som de kan være samlet på et skrivebord. Samlingsmaterialet er georeferert i den forstand at objektene i samlingene kan assosieres med et område på jordas overflate. En kan si at ADL er et geobibliotek som inneholder organiserte samlinger der et primært attributt ved samlingsobjekter er deres lokasjon på jorda, representert ved geografiske avtrykk.

Utvikling av ADL inkluderer utviklingen av ADL Gazetteer. Denne er tilgjengelig som komponent av ADL, men også som en frittstående webtjeneste. Som komponent av ADL kan gazetteertjenesten benyttes for å finne fram i samlinger med georeferert materiale. Samlingene i ADL inkluderer kart, flyfotografier og satelittfotografier av jordas overflate.

Alexandria Digital Earth Prototype (ADEPT) er et paraplybegrep på en videreføring av ADL Project som samtidig utvider fokus fra et geospatialt bibliotek til et integrert miljø for forvaltning og bruk av geospatial informasjon og da særlig for læringsapplikasjoner [63].

Aktuelt forskningsfokus som del av ADEPT er å bygge en konsept-rom-modell for å representere konsepter i naturfagfeltet og deres relasjoner til bruk i undervisning. Konsept-rom-modellen utvider tesaurusmodellen ved å karakterisere sett av konsepter til å representere kunnskapsdomener mer komplett. En slik modell anbefaler/ oppmuntrer til representasjon av relasjoner som separate sett av komponenter – et system for kunnskapsorganisering i seg selv [65].

5.3.1. ADL arkitektur

ADL har en trelags klient-server arkitektur som består av klient, mellomvare og databaseservere. [3]

Klient	Støtter interaktive queries og evaluering av søkeresultat, inkl. interaktivt kart. Ekstra søkeparameter kan settes ved å benytte et sett av høy-nivå søkebuckets. Disse er designet for å støtte enkle søk via multiple metadatasett av varierende format.
Middelvare	Utfører kompleks kommunikasjon og oversetter mellom mange klienter og mange databaseservere. Oppretter og avslutter sessions.
Databaseservere	

Figur 5-3 ADL Arkitektur

Eksisterende ADL samlinger inkluderer blant annet en ADL Gazetteer og som del av ADL kan denne brukes som et verktøy i et georeferert digitalt bibliotek. ADL Gazetteer inkluderer en Feature Type Thesaurus (FTT), en Gazetteer Content Standard (GCS) og en operasjonell interaktiv gazetteer service. [4]

5.4. GeoXwalk

GeoXwalk er et felles prosjekt mellom et skotsk nasjonalt datasenter EDINA og History Data Services. Prosjektet er støttet av The Joint Information Systems Committee (JISC). Ref [2] [72]

Hovedmål for prosjektet er å utvikle en felles tjeneste innenfor JISC Information Environment som kan understøtte geografisk søking. Prosjektet går over flere faser der mål for fase 1 var å utrede muligheter for å utvikle og tilby en Britisk og Irsk Gazetteertjeneste som vil være viktig for å støtte geografisk søking i Distributed National Electronic Resource (DNER). Den skulle være online, Z39.50 compliant, rask og utvidbar. Prosjektet definerer geospasiale data som data som har en form for spatial eller geografisk referanse som gjør det mulig å lokalisere disse i et to eller tredimensjonalt rom. En geoparser er programvare som kan identifisere forekomster av geografiske navn i tekstdokumenter slik at en kan få tilgang til lokasjonsinformasjon om disse via en gazetteer. Prosjektets definisjon av en gazetteer:

”en gazetteer som en liste over geografiske navn sammen med deres assosierte lokasjon. En geografisk lokasjon kan uttrykkes på flere måter for eksempel som koordinater for breddegrad og lengdegrad eller rett og slett i form av navn på et annet sted som aktuelle sted er en del av”.

Mål for fase 2 var å utvikle en gazetteertjeneste med muligheter for å utvide tjenestetilbudet, vurdere hvordan gazetteerdata kunne gjøres tilgjengelig som felles tjeneste innenfor JISC Information Environment og å promotere mulighetene for en fullt fungerende tjeneste og gå foran som et bevis på konseptet. Fase 3 bygger videre på fase 2 og utforsker ulike aspekter som er identifisert i fase 2.

Prosjektet vurderte muligheten for å bruke gazetteertjenester til å støtte andre tjenester i DNER og fant at dette kunne gjøres på to måter:

- Støtte geografiske søk
- Støtte geografisk indeksering av informasjonsressurser

Prosjektet har identifisert behov for en slik tjeneste for å virke som en generell referansekilde for steder og forekomster i UK og Irland.

Noe av informasjonen om dette prosjektet er hentet fra prosjektets nettsider og manglende dateringer på disse gjør det vanskelig å si hvor langt prosjektet er kommet. Jeg har ikke vært i stand til å finne informasjon om erfaringer eller evaluering av prosjektet. Jeg synes allikevel at dette er interessant fordi det handler om å gjøre informasjonsressurser, som i utgangspunktet ikke er definert som geografisk informasjon, søkbare via geografiske referanser. Prosjektet har utforsket nytten og brukbarheten av ADL Gazetteer Content Standard inkludert ADL Feature Type Thesaurus og har basert sin modell på denne. (Se kapittel 6.2)

5.5. Textual – Geospatial Integration Services for the National SMETE Digital Library

Dette er et amerikansk prosjekt ved National Science, Mathematics, Engineering and Technology Education Digital Library (SMETE Digital Library) [73]. Målet med dette prosjektet er å gi publikum enhetlig tilgang til relatert informasjon fra ressurser som kart, tekst og bilder, spesielt informasjon som er semantisk relatert i forhold til geospasiale forekomster og regioner. Prosjektet er støttet av Universitetet i California, Santa Barbara og integrerer geospasiale søketjenester fra ADL med kraftige informasjonsgjenfinningstjenester som er utviklet ved Illinois Institute of Technology's Information Retrieval Laboratory [74]. Nøkkelkomponenten er ADL Gazetteeren med tjenester. Prosjektets definisjon av en gazetteer:

”En gazetteer tilbyr mapping mellom tekstlig representasjon av steder og forekomster og koordinatbasert representasjon av ders geografiske avtrykk og tillater videre en semantisk mapping mellom innhold i tekst og kart/bilder, i begge retninger.”

Avanserte tjenester skal støtte brukere til bedre å kunne evaluere konsistens i informasjon fra ulike kilder mens tjenestene samlet skal gjøre det mulig å indeksere store mengder tekst slik at denne kan aksesseres geospasialt selv om det ikke finnes geospasiale referanser i opprinnelig metadatapost.

Konkret arbeides det med å integrere en gazetteer inn i en IR-motor, integrere IR-motoren inn i et geospasialt digitalt bibliotek og utvikle tjenester som sjekker konsistensen ved alternative geospasiale referanser. Tjenestene skal tillate geospasiale spørringer i distribuerte heterogene digitale bibliotek etter både tekstlige og ikke tekstlige ressurser og vil gjøre det mulig å georeferere tekst som ellers ikke vil være katalogisert med tanke på geografisk informasjon. Dette er, når dette skrives, et pågående prosjekt som etter planen skal avsluttes i september 2004.

5.6. Samlinger I Nasjonalbiblioteket

Blant de av nasjonalbibliotekets samlinger som er tilgjengelige på www finnes flere bildearkiver. Et betydelig arbeid er nedlagt for å gjøre disse tilgjengelige for internettbrukere. Felles for disse er at de er søkbare via web-grensesnitt der brukeren har mulighet for å gjøre avanserte søk og får tilgang til digitale dokumentsurrogater. Hver samling har sitt eget grensesnitt mot brukeren og søkemulighetene er litt forskjellige. Jeg har ikke studert disse samlingene og systemene grundig, men ved å gjøre noen enkle søk ser man at problemområder som er nevnt tidligere også gjør seg gjeldende her. For eksempel gir søk etter henholdsvis *person* og *personer*, i feltet *Motiv* forskjellig resultat i samlingene *Roald Amundsen bildearkiv* og *Fridtjof Nansen bildearkiv*. Det gjør det for så vidt ved søk i feltet *Tekst/Motiv* i samlingen *Nazisme og motstandskamp 1933-45 i Norsk krigstrykksamling* også, men i denne samlingen omfatter feltet *Tekst/Motiv* den teksten som utgjør innholdet i det originale trykket. Søkegrensesnittet for sistnevnte samling kontrollerer søketermene for to av feltene, *Materialtype* og *Sted* ved å tilby indekser for disse. For å søke i disse feltene er brukeren nødt til å velge søketermer fra en liste. Informasjon om disse samlingene og systemene finnes via Nasjonalbiblioteket sine websider [83].

Nasjonalbiblioteket avdeling Rana (NBR) ble opprettet i 1989 og har bygd opp en egen avdeling for å sikre film og fotografi [46]. *Galleri Nor – møteplassen for fotografi i Norge* er resultat av en videreutvikling av prototypen BAUTA som ble utviklet gjennom et samarbeidsprosjekt mellom NBR og flere andre aktører i tida 1993-1995. Galleri Nor med Norsk Folkemuseums Wilse-samling som den første større samling på nettet, ble satt i drift fra 1996. Galleri Nor ble samtidig lansert som et tilbud til andre institusjoner om tilgjengeliggjøring av norske fotosamlinger.

Fotosamlingen av Anders Beer Wilse utgjør en av de mest betydningsfulle delene av vår nasjonale fotoarv og teller 110.000 negativer og 30.000 originalkopier, tatt i tidsrommet 1900 - 1945. Samlingen inneholder også omlag 3.500 negativer av Aksel Lindahl, tatt i 1880-årene. Wilse reiste mye og var blant annet en av pionerene i Norsk Turistforening. Sannsynligvis har han besøkt hver eneste kommune i landet, og han brukte fotoapparatet flittig der han for. I samlingen finnes landskapsbilder og prospekter fra by og land, dokumentasjonsbilder fra arbeidsliv og friluftsliv, samt en mengde portretter. Det er ingen underdrivelse å si at fotosamlingen er en visuell framstilling av norgeshistorien i denne perioden. Galleri Nor er en viktig komponent i denne oppgaven og blir derfor nærmere beskrevet i kapittel 6.1.

6. Mitt system

Mitt forslag går ut på å integrere en gazetteer og tesaurus i et system med en bildesamling. Målet med dette er å utnytte egenskaper ved slike strukturer for å skape gode gjenfinningsmuligheter. Bilder fra fotodatabasen Galleri Nor brukes som eksempel for å konkretisere problemområder og mulige løsninger. Konkretisering av forslaget begrenses til å omfatte en gazetteer for Galleri Nor og en navnetypesaurus som er en del av denne. Gazetteer og navnetypesaurus kan brukes for å utnytte geografisk relatert informasjon som allerede eksisterer i opprinnelige beskrivelser av informasjonsressursene.

Galleri Nor sammen med ADL Gazetteer Content Standard og ADL Feature Type Thesaurus er viktige komponenter som har betydning for innholdet i forslaget. Det er derfor nødvendig å gå noe nærmere inn på Galleri Nor og ADL Gazetteer.

- Forslaget har fokus på mulighetene for gjenfinning ved å integrere en gazetteer og en navnetypesaurus i et system med samlinger og her spesielt med samlingen Galleri Nor.
- Forslaget inkluderer beskrivelse av en gazetteer, beskrivelse av en navnetypesaurus og hvordan et slikt system kan fungere.
- Forslaget vil også inkludere en beskrivelse av et noe forenklet system som er implementert og testet på en subsamling av Galleri Nor og en beskrivelse av hvordan dette fungerer og eventuelt påvirker søkeresultatet.
- Som et ledd i tilnærmingen til Galleri Nor og med tanke på muligheten for interoperabilitet med andre systemer eller samlinger har jeg vurdert muligheten for å bruke Dublin Core som utvekslingsformat for de metadata som er interessante for brukere av Galleri Nor. Jeg legger fram forslag til mapping av deler av Galleri Nors metadata til Dublin Core med en gjennomgang og diskusjon av aktuelle elementer.

6.1. Galleri Nor i dag

Galleri Nor er brukt som eksempel på en bildesamling og skal være utgangspunkt for konkretisering av problemer og mulige løsninger. Det er derfor nødvendig å si noe om Galleri Nor i dag, hva Galleri Nor er, hva som tilbys, hvordan søkesystemet fungerer og noe om hva som eventuelt er bra og mindre bra med systemet slik det er i dag¹.

Generelt om samlingen Galleri Nor

En av Nasjonalbibliotekets fotodatabaser, Galleri Nor, er en digital katalog over viktige fotografi i Norge. Etter en oppdatering 03.01.03 inneholdt basen 68.053 digitale versjoner av fotografier med beskrivelser. Mer enn 65.000 av disse kan føres tilbake til fotografen Anders Beer Wilse og er produsert i tida 1880 til 1950. Wilse var etablert fotograf i Oslo, den gang Christiania, men han brukte også mye av sin tid på å reise rundt i Norge for å fotografere. Han var først og fremst kjent for sine landskapsbilder og

¹ Galleri Nor har i 2004 lansert en ny versjon av sitt system. Dette har skjedd etter at det meste av dette er skrevet. Se *Nye Galleri Nor* sist i kapittel 6.1. Henvisning med url viser til nye galleri Nor.

portretter. Basen er tilgjengelig for publikum via Internet og det tilbys et system for gjenfinning og visning av digitale versjoner av fotografier. [53]

Søkesystem og grensesnitt

Informasjon om søkesystem er hentet fra Galleri Nor's brukerinformasjon som er tilgjengelig via lenken *søketips* på søkesiden.

Søk bygges opp ved å fylle ut ett eller flere søkefelt i brukergrensesnittet. Det skilles ikke mellom liten og stor forbokstav slik at søk etter *bergen* vil automatisk også søke etter *Bergen*.

Logiske operatører OG (&), ELLER (|) og IKKE (^) kan benyttes i alle søkefelt utenom dateringsfeltene. Automatisk høyre- og venstretrunkering benyttes i systemet, men kan utelukkes ved at en bruker skilletegn (space) foran eller bak søkeordet for å angi om man søker etter starten eller slutten av et ord.

Søkemuligheter

Fire typer kriterier kan legges til grunn for søk etter fotografier. Disse er *motiv*, *dato*, *sted* og *egennavn* og har egne felt i søkebildet for å legge inn søkekriterier. Figur 6-1 viser søkebildet fra Galleri Nor.

Søkeord for motiv

Her kan en angi frie emneord for søk etter motiv og det søkes etter metadataelementet motiv i databasen. Stikkord er registrert i form av verbalsubstantiv og substantiv i entall og flertall. Emneord av overordna abstrakt karakter er ikke registrert slik at en ved å søke etter *idrett* ikke vil få noen treff, mens søkeord som *fotball* og *skiløping* vil kunne gi treff.

Dato

Dateringer er gjort i form av årstall, ikke måned og dag. Her kan en søke på årstall for eksponering av bildet. En kan søke på et eksakt årstall eller angi et tidsrom i form av to årstall. Ved å angi tidsrom vil en finne alle bilder som har et av årstallene innenfor intervallet.

Sted

Søkefeltet for søk etter geografisk sted er todelt. Ett felt er reservert for å angi land, fylke eller kommune, mens det i det andre feltet kan søkes etter andre typer stedsnavn. En kan søke etter navn på områder som dalfører og fjorder eller angi mer spesifikke navn som bydeler eller gatenavn. En fylkesvis liste over kommuner er tilgjengelig for oppslag. Kommunelista er fra 1960-tallet.

Egennavn

Her kan søkes på personnavn, navn på byggverk, institusjoner, egennavn på dyr. For bilder av privatpersoner vil en av og til oppleve at bildet er skjult, men at en kan få fram noe tekstinformasjon. Personvernlovgivningen setter begrensninger for visning av bilder av privatpersoner. Det tilbys to like felt for søk etter egennavn der en samtidig kan spesifisere hvilket forhold navnet skal ha til bildet. Dette gjør en ved å velge fra en liste med fire valg: fotograf, eier, avbildet eller annen relasjon. Uspesifisert navnesøk fører til søk i alle fire kategoriene.

Presentasjon av søkeresultat

Brukeren kan til en viss grad bestemme hvordan søkeresultatet skal presenteres. En kan velge mellom bilder i frimerkeformat med noe metadata, bilder i kvartformat med de samme metadata eller frimerkebilder uten metadata. En kan også velge antall bilder som skal vises og hvilket nummer en skal starte med.

Fra resultatlista kan en velge enkeltbilder som da vises i fullt format med all metadata som er tilgjengelig for bruker.

GALLERI NOR - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address http://www.nb.no/gallerinor/

Galleri nor - Søkeside -

Bygg opp søket ved å fylle ut et eller flere av søkefeltene under.
Rett under søkefeltene kan du velge hvordan du ønsker å få søkeresultatene dine presentert ("Vis treff som").
Når du trykker på ledeteksten til et felt, får du nederst på siden hint om hvordan feltet kan fylles ut.
For mer informasjon om hvordan søket kan utføres, se på [søketipsene](#).

Søkeord til motivet **Datering** -

Land, Fylke eller Kommune **Stedsprecisering**

Egennavn til bildet **Egennavn til bildet**

Uspesifisert **Uspesifisert**

Fotograf Eier Fotograf Eier

Vis treff som: Frimerkebilder Kvartformat Frimerkebilder w/tekst

Frimerkebilder pr. side:
Kvartbilder pr. side:
Start med bilde nr.:

Trykk på ledeteksten over feltet for en nærmere forklaring på hvordan du skal bruke det.
Boolske operatører for å bygge opp et søk er: "&" = OG, "|" = ELLER og "^" = IKKE.
Se søketips for utfyllende beskrivelse av hvordan du skal søke.

Figur 6-1 Skjermbilde fra Galleri Nors Søkeside.

Datamodell og ulike typer metadata

Galleri Nor har utviklet sin datamodell med basis i NKKM's registreringsformat. Galleri Nors databasemodell finnes i vedlegg C. Modellen ivaretar ulike typer metadata for ulike formål. En stor del av registrerte metadata vil ikke være synlige for andre enn de som administrerer samlingen og derigjennom har full tilgang til alle metadata. De metadata som er tilgjengelige for den allmenne bruker, kan kalles informative metadata og det er disse, eller deler av disse, som har betydning for gjenfinning i den forstand at

de er søkbare for bruker. En liste over de metadata som er tilgjengelig for den allmenne bruker og hvilke av disse som er søkbare er vist i Tabell 6-1.

Fotograf, eier og navn er søkbare via feltene *Egennavn til bildet*, frie søkeord via feltet *søkeord til motivet*, datering via feltene *datering* og sted via feltene *Land, Fylke eller Kommune* og *Stedspresisering*. Stedsnavn kan i enkelte tilfeller også være registrert som søkeord og dermed være søkbare i feltet *søkeord til motivet*. Internnummer og tilvekstnummer er synlige, men ikke søkbare og elementene tittel og samling er ikke i bruk.

Galleri Nor element * kan gjentas		Søkbart element
Tittel	[ikke i bruk]	Nei
Fotograf		Ja
Eier		Ja
Søkeord *		Ja
Utfyllende opplysninger		Nei
Datering fra - til		Ja
Sted *		Ja
Navn *		Ja
Samling	[ikke i bruk]	Nei
Internnr.		Nei
Tilvekstnr.		Nei

Tabell 6-1 Liste over opprinnelige Galleri Nor elementer som er tilgjengelig for den allmenne bruker

Stedsnavn i Galleri Nor

På forespørsel om oppbygging av stedsnavnregister sier en medarbeider ved Galleri Nor:

”Stedsregistreret bygger vi selv opp etter hvert som nye steder/adresser blir lagt til. Vi benytter oss av ulike kart i arbeidet med å identifisere bildene, men prøver etter beste evne å unngå at et enkelt sted blir skrevet på forskjellige måter (Navn fra Statens Kartverk har førsteprioritet). Vi har også jevnlig runder med korrekturlesning av stedsregisteret. Når vi registrerer i stedsfeltene har vi hele tiden mulighet til å søke opp tidligere brukte stedsangivelser.”

Emneord i Galleri Nor

På forespørsel om oppbygging av emneordregister sier en medarbeider ved Galleri Nor:

”Også her er det vi selv som bygger opp et stikkordsregister etter hvert som nye ord blir lagt til. I det daglige arbeid har vi ingen hjelp av dette registeret fordi det ikke vises under registreringen, men ved korrekturlesning av de registrerte

postene kjører vi ut lister over begrep som er benyttet. Det hadde vært svært ønskelig med ei liste som foreslår stavemåter etter hvert som man skriver inn ordene. Emneordene som benyttes er forøvrig substantiv (ubest. form entall og flertall) og verb i -ing form. Egennavn skal ikke forekomme i dette feltet.”

Styrke og svakheter i eksisterende system

Samlingen består av historisk billedmateriale og er omfattende både når det gjelder størrelse og motiv. Dette kan gi oss muligheter for å dokumentere vår historie på flere områder som for eksempel transportmidler, klesdrakter, bebyggelse osv.

- En har mulighet for, som tidligere nevnt, å søke etter motiv i form av frie søkeord, egennavn og forhold dette skal ha til bildet, sted i form av land, fylke kommune og spesifisert sted og datering i form av ett eller to årstall.
- Muligheter for å benytte logiske operasjoner og mekanismer for å begrense automatisk høyre- og venstretrunkering av søketermer er positive egenskaper ved systemet for de som vet å utnytte disse.

Det er ingen tvil om at det her er gjort et stort stykke arbeid for å gjøre fotosamlinga tilgjengelig for allmennheten. Mulighetene for å finne aktuelle fotografier er i utgangspunktet mange, men det er enkelte forhold som ikke styrker systemets brukbarhet.

- Brukergrensesnittet er lite intuitivt og dette i kombinasjon med systemet bak krever at bruker må være interessert i å sette seg inn i hvordan systemet fungerer for å kunne utnytte eksisterende muligheter.
- Motivene er ikke kategorisert, men søkeordene er tilordnet i et visst system fra generelle til spesielle. Svært generelle søketermer vil kunne gi et urimelig stort antall treff. Termen *landskap* kan være et eksempel på dette.
- Systemet gjør bruk av både entall og flertall for tellbare objekter avhengig av antall objekter i motivet. Bruker må søke på begge formene for å være sikker på å finne aktuelle fotografier.

Registrering av stedsnavn og typiske navnetypetermer som søkeord er spesielt interessant for denne oppgaven. Denne type informasjon er registrert for en stor del av fotografiene med tanke på at geografiske navn og begreper skal være søkbare.

- Oppdelingen av søkefeltet for sted kan være med på å skape usikkerhet på det vis at en må vite om et stedsnavn er navn på en kommune, et fylke eller et spesifisert sted. Mange bilder er landskapsbilder der dette kan være problematisk.
- Detaljeringsgraden i registrering av stedsnavn varierer, spesielt i feltet *stedspresisering*. Eksempelvis kan ett fotografi fra Trondheim være registrert med *Trondheim* som *stedspresisering* mens et annet kan være registrert med *Ila* og et tredje med *Bispegata* i samme felt.
- Det gjøres ikke systematisk bruk av navnetyper noe som medfører at det ikke nødvendigvis er sammenheng mellom typiske navnetypetermer i søkeord og registrerte stedsnavn. Et eksempel på dette er bilde_41 i testsamlinga. Her er *Trondheim* registrert som eneste stedsnavn i feltet *stedspresisering* mens vi blant søkeordene finner *bydel* som er en typisk navnetypeterm.

- Det er ikke nødvendigvis noen sammenheng mellom relaterte termer som for eksempel bygninger og kirker. Dette bekreftes ved at et søk via termen kirke gir høyere treff enn et søk via termen bygning.

Nevnte problemområder indikerer inkonsistens i registreringen av stedsnavn så vel som andre metadata. Inkonsistens i registrering av metadata vil skape konflikter i systemet som i stor grad kan påvirke gjenfinningsmulighetene.

Nye Galleri Nor

Etter at dette arbeidet kom i gang og mye av grunnlaget for hele oppgaven var lagt, lanserte Nasjonalbiblioteket Rana, vinteren 2004, en ny versjon av Galleri Nor.

Ny versjon har fått nytt brukergrensesnitt og det er nå mulighet for å søke i flere samlinger. Det nye brukergrensesnittet er tiltalende og virker ryddig.

Angående søkefeltene er det ingen andre endringer enn at det er lagt til et nytt felt *nummer*. I følge søketipsene som er tilgjengelige via en lenke fra søkesiden kan de som kjenner ressursens tilvekstnummer søke etter dette i feltet *nummer*. Tidligere nevnte kommuneliste fra 1960-tallet er ikke tilgjengelig i det nye søkebildet.

Ellers er det endringer i bruk av symboler for boolske operatører samt noen få endringer som angår presentasjon av søkeresultat.

Nye Galleri Nor ser ikke ut til å medføre endringer som berører identifiserte problemområder i denne oppgaven og vil dermed heller ikke ha noen betydning for resultatene. Henvisning til Galleri Nor med url viser til nye Galleri Nor.

6.2. ADL Gazetteer

Gazetteers har lenge vært sentarale i ADL. En tid var Gazetteeren sett på som en samling og behandlet som andre samlinger av flyfotografier, satelittbilder, kart og andre georefererte dokumenter. Men Gazetteerens funksjon i ADL arkitekturen er spesiell. Den svarer på spørsmål som *hvor er Bakersfield?* Den tilbyr en oversetterfunksjon i prosesseringen av et query som *hvilke satelittbilder har biblioteket som dekker Bakersfield?* der stedsnavnet *Bakersfield* blir oversatt til koordinater for lengdegrad og breddegrad for å finne de bildene som dekker dette området. Gazetteerdata kan være veldig viktig for geospasiale bilder og kart for å identifisere og merke forekomster og for å tilby kontekst som er nødvendig for å evaluere geospasiale data. Gazetteerer støtter geoparsing av tekstdokumenter, hvor igjennom geografiske områder som omhandles i dokumentet kan representeres av koordinater og slik gjøre dem til objekter som er passelig for et geospasialt digitalt bibliotek [65].

ADL Gazetteer er utviklet ved å kombinere data fra to store amerikanske databaser, Geographic Names Information System (GNIS) og Geographic Names Processing System (GNPS). Disse tilbyr god dekning av geografiske stedsnavn. Begge databasene kan søkes i og returnerer rapporter – noe er likt og noe er forskjellig i rapportene som returneres.

Begge opererer under the Board of Geographic Names (BGN), en amerikansk navneautoritet som håndterer navnebruk for både amerikanske og ikke-amerikanske navn, undersjøiske og antarktiske steder. BGN håndterer problemer, spørsmål og forslag om navn både for offentlige organer og for generelle publikum [26].

GNIS og GNPS har sine egne karakteristikk og under utvikling av ADL Gazetteer har staben ved ADL lagt vekt på å ivareta strukturer og detaljer fra begge. Kombinert datasett holder nærmere 6 millioner innganger.

ADL Gazetteer data er beskrevet i henhold til ADL Gazetteer Content Standard (GCS) og ADL Feature Type Thesaurus (FTT). Disse blir beskrevet nærmere i kapittel 6.2.1 og 6.2.2.

ADL Gazetteer er tilgjengelig både som webservice og som en komponent av ADL[27]. I tillegg til å lokalisere stedsnavn på kart viser et søkeresultat fra ADL gazetteer annen informasjon som for eksempel navnetype og relasjoner til andre geografiske navn som fylke og land og annen relatert informasjon.

Kjerneelementer i ADL Gazetteer

- Geografiske navn
- Lokasjonsinformasjon
- Navnetyper

Kjerneelementer for digitale gazetteers er definert under utvikling av ADL Gazetteer. Disse er tidligere omtalt under avsnittet *Digitale gazetteers* i kapittel 4.3.

6.2.1. ADL Gazetteer Content Standard (ADL GCS)

Prosjektene ADL og ADEPT har bygget sin gazetteer basert på geografiske data fra flere kilder og peker på dette som en utfordring. På bakgrunn av dette har de utviklet en innholdsstandard, Gazetteer Content Standard (GCS) [60] som de bygger sin gazetteer på og som kan benyttes for utvikling av nye gazetteers. ADL GCS har 14 hovedelementer i tillegg til kildeinformasjon (Figur 6-2). Av disse er det 5 elementer i tillegg til kildeinformasjon som er obligatoriske. Hovedelementene har null eller flere sub-elementer. Enkelte av disse er også obligatoriske. En oversikt over elementene i ADL GCS finnes i vedlegg B.

<ol style="list-style-type: none">1. Geographic Feature ID2. Geographic Name3. Variant Geographic Name (R)4. Type of Geographic Feature (R)5. Other Classification Terms (R)6. Geographic Feature Code (R)7. Spatial Location (R)8. Street Address9. Related Feature (R)10. Description11. Geographic Feature Data (R)12. Link to Related Source of Information (R)13. Supplemental Note14. Metadata Information <p style="text-align: center;">Source Information</p>

Figur 6-2 Hovedpunkter i ADL Gazetteer Content Standard (elementer med uthevet skrift er obligatoriske, (R) betyr at de er repeterbare)

6.2.2. ADL Feature Type Tesauros (ADL FTT)

ADL FTT inneholder termer i et hierarkisk skjema. Termene skal brukes for å typebestemme navngitte geografiske steder slik at de sier noe om stedets natur. ADL FTT foreligger i to versjoner. Felles for disse er at antall topptermer er 6 og antall foretrukne termer er 210. Antall ikke foretrukne termer er rundt 1000, men er noe høyere for den seneste versjonen.

6.2.3. Identifiserte kriterier for å integrere gazetteers i digitale bibliotek

Under arbeidet med utviklingen av ADL Gazetteer ble det identifisert en del kriterier for å integrere gazetteers i digitale bibliotek [26].

Innholdsstandard (Content Standard)

De mange kildene for spatial geografisk refererte navn er for det meste beregnet brukt for spesielle formål og er ikke designet for interoperabilitet og deling. Dette avdekket et behov for et standard konseptuelt skjema for gazetteer informasjon for at slik informasjon bedre kan skapes og deles. Etablering av en slik standard for deling av data vil støtte høsting av data fra flere kilder.

Behovet for en innholdsstandard førte til utviklingen av ADL Gazetteer Content Standard (ADL GCS) som kan være veiledende for utvikling av metadata for gazetteers. Denne er ganske omfattende, men man kan velge å bruke deler av den. Forslag til metadata for en gazetteer i denne oppgaven vil ta utgangspunkt i ADL GCS.

Navnetyper (Feature types)

Behov for et typeskjema for å kategorisere individuelle forekomster for delte gazetteers Skjema må være:

- hierarkisk
- rikt i termvarianter
- utvidbart for å kunne gi større dybde i terminologi hvis nødvendig

Skjema må inkludere termvarianter fra etablerte forekomst-type-skjema slik at det kan tilby mappinger mellom de forskjellige skjema.

ADL Feature Type Tesauros (ADL FTT) er utviklet for å dekke dette behovet for ADL Gazetteer. ADL FTT inneholder termer for å typebestemme geografiske steder. ADL FTT brukes for å begrense antallet og kontrollere hvilke termer som kan brukes. ADL FTT er tilgjengelig for oppslag ved søk i ADL Gazetteer for å veilede søkeren til å finne de søkbare termene. En norsk gazetteer til Galleri Nor vil ha behov for et lignende typeskjema som egner seg for norske forhold. Navnetyper fra SOSI standarden, som brukes av Statens Kartverk og Sentralt Stedsnavnregister, kan være et godt utgangspunkt for en norsk navnetypesaurus.

Tidsaspekter (Temporal aspects)

Gazetteer må ha mekanismer som håndterer endringer i gazetteer-data over tid. En må kunne knytte tidsaspektet til gazetteer-data.

Endringer i geografisk informasjon skjer hele tiden. Grenser flyttes, nye forekomster oppstår mens andre forsvinner. Hvorvidt, og i så fall hvordan, en gazetteer skal håndtere dette aspektet må vurderes ut fra hva gazetteeren skal brukes til. ADL GCS tilbyr mekanismer for å registrere dateringer i forhold til gyldighet for stedsnavn. Tidsaspektet kan være vesentlig også i forhold til gjenfinning av informasjonsressurser, ikke minst dersom ressursene er knyttet til geografiske steder og strekker seg over et visst tidsrom.

Uklare avtrykk (Fuzzy footprints)

Geografiske avtrykk er ofte unøyaktig definerte. Det vil være behov for regler og metoder for å framstille slike uklare grenser og lokasjoner som igjen framstilles for bruker.

Hvor nøyaktig en skal angi lokasjonsinformasjon vil være avhengig av hvilken sammenheng denne informasjonen skal brukes i. For å vise en korrekt plassering på et kart vil en ha behov for nøyaktig informasjon, mens en i andre tilfeller kan tenke seg at en skal forklare noen hvor et sted er. Vedkommendes lokalkunnskap vil da være avgjørende for hvor detaljert en slik beskrivelse kan være. Jo mindre lokalkjennskap vedkommende har jo mer uklare må grensene være for å ha noen nytte. ADL GCS tilbyr elementet "related feature" som en måte å angi slike uklare avtrykk. I denne oppgaven vil en slik mulighet for å angi uklare avtrykk danne grunnlaget for en viktig del av gazetteerens funksjonalitet.

Kvalitetsaspekter (Quality aspects)

Flere kvalitetsaspekter

- hvordan indikere nøyaktighet for lengde og breddegrad
- sikre at rapporterte koordinater stemmer med andre elementer i beskrivelsen
- generelt bør kvalitetskontroll forekomme for alle dataelementer der det er mulig

Dette er aspekter som vil ha mer eller mindre betydning alt etter hva en gazetteer skal brukes til. For forslaget i denne oppgaven er disse aspektene av mindre betydning.

Spatiale aspekter (Spatial aspects)

- punktlokasjoner representerer ikke yttergrenser for geografiske lokasjoner
- avgrensede bokser inkluderer for store områder

Punktlokasjoner og avgrensede bokser er de enkleste formene for geografiske avtrykk. Hvilken type avtrykk som skal brukes må vurderes i forhold til hva man skal bruke denne type informasjon til. For en gazetteer som skal brukes for å vise omtrentlig lokasjon på et kart kan punktlokasjoner være greit nok, men som oppslagstjeneste i et gjenfinningssystem vil en ikke ha særlig nytte av denne typen avtrykk. Avgrensede bokser vil kunne gjøre nytte, men kan i mange tilfeller inkludere for store områder.

Internasjonale og flerspråklige gazetteers

- bruk av non ascii-characterset
- oversette fra et tegnsatt til et annet
- flerspråklige kategorisett

Bruk av koordinater skjer på tvers av grenser og språk

Dette er aspekter som er veldig interessante og som har blitt mer og mer aktuelle med økende bruk av nettverksteknologi og de muligheter dette innebærer for deling av data og informasjon på tvers av landegrenser og språk. Spesielt for norsk er tegnene æ, ø og å. Vi ser ofte at disse erstattes av ae, o eller oe og aa, men dette skjer gjerne uten spesielle regler og ikke helt uten problemer.

Navnetypesauri kan oversettes mellom flere språk slik at en kan knytte relasjoner mellom termer fra ulike språk. Navnetyper fra Sentralt Stedsnavnregister kan for eksempel mappes til ADL feature types.

6.3. Systemforslag

Forslaget går ut på å integrere en gazetteer og tesauri i et system med en bildesamling. Målet med dette er å utnytte egenskaper ved slike strukturer for å skape gode gjenfinningsmuligheter. Konkretisering av forslaget begrenses til å omfatte en gazetteer for Galleri Nor og en navnetypesaurus som er en del av denne. Gazetteer og navnetypesaurus kan brukes for å utnytte geografisk relatert informasjon som allerede eksisterer i opprinnelige beskrivelser av informasjonsressursene.

6.3.1. Brukere – hvem er de og hvilke ønsker/behov har de

Brukere vil i denne sammenheng være den allmenne bruker som kan tenkes å være interessert i bilder fra aktuelle kategorier og tidsrom. Beskrivelsene tar i hovedsak for seg egenskaper knytta til fotografiernes motiv. Metadata for tekniske og administrative egenskaper ved bildene så som type utstyr for produksjon, digitalisering, preservering og så videre er ikke tatt hensyn til. Fotografiene er, for det meste, produsert i tida 1880 til 1950 og for enkelte fotografier som har motiv fra todimensjonale objekter som for eksempel tegninger eller malerier kan motivet dateres enda lengre tilbake i tid.

Hvordan mennesker oppfører seg når vi søker etter informasjon på Internet er forskjellig ut fra en rekke forutsetninger. Denne oppgaven søker å vise hvilke muligheter bruker har for å finne fram til aktuelle bilder i Galleri Nor via stedsnavn og navnetyper slik systemet er nå og slik det vil kunne fungere dersom man integrerer en gazetteer og en navnetypesaurus i systemet. Det er et viktig poeng her at det å ta i bruk gazetteers og tesauri i et system med samlinger ikke er ment å erstatte det opprinnelige søkesystemet der en kan søke i tradisjonelle felter som emne, dato, person osv., men at dette er teknologier som kan tas i bruk for å styrke eksisterende systemer. En bør fortsatt ha muligheten til å søke på andre felter i tillegg til stedsnavn og navnetyper.

6.3.2. Bildekategorier

Fotografiernes motiv kan intuitivt deles i fire kategorier eller konsepter: *portrett*, *landskap/geografisk sted*, *aktivitet* og *hendelse/begivenhet* (Tabell 6-2). Med en vid tolkning av begrepene ser vi at disse godt kan dekkes av emneentitetene i FRBR-modellen. Bygninger er motiv for en del av fotografiene i samlinga og en kunne ha valgt å se på disse som en egen kategori, men jeg har valgt å inkludere disse i kategorien *Sted*. Når det gjelder institusjoner ser jeg det slik at institusjonens navn må knyttes til et konkret objekt for å kunne utgjøre et motiv. I denne sammenhengen mener jeg det er mest naturlig også disse kommer inn under samme kategori som bygninger. Det er stort mangfold i motiver og hvilke metadataelementer som er nødvendige for beskrivelser av

de ulike kategorier vil kunne variere noe. Felles for en stor del av fotografiene er at de kan knyttes til minst ett geografisk navn som igjen kan knyttes til en navnetype. Det som kan være interessant videre i denne oppgaven er å se om denne informasjonen kan utnyttes til å påvirke søkeresultatet i et foreslått søkesystem.

Bildekategori:	Beskrivelse:
Portrett	En eller flere personer
Sted	Geografisk sted med stedsnavn og navnetype, sted inkluderer også navngitte bygninger / institusjoner
Hendelse / Begivenhet	For eksempel 17.maifeiring
Aktivitet	For eksempel skihopping, innhøsting

Tabell 6-2. Bildekategorier

6.3.3. Hva skal vi kunne søke etter

- **Geografisk sted via stedsnavn**
Herunder kommer også navngitte bygninger
- **Et steds natur via navnetype – feature type**
En navnetype sier noe om et steds natur. Vanligvis er typiske navnetypetermer registrert som emneord.
- **Andre søkekriterier**
 - **Person via egennavn for fotograf, avbildet/portrettert**
Eventuelt også for kunstner dersom avbildet objekt er todimensjonalt)
 - **Institusjon/korporasjon via stedsnavn**
Eventuelt via egennavn
 - **Navngitt bygning via stedsnavn**
Navngitte bygninger er registrert som stedsnavn i Statens Stedsnavnregister, men som egennavn i Galleri Nor
 - **Dato i form av ett eller to årstall**
For at en skal kunne finne alle bilder fra ett spesielt år eller innenfor et tidsrom
 - **Egennavn**
 - **Emneord**

Gazetteertjenesten skal integreres slik at den konsulteres automatisk ved søk etter informasjon som kan knyttes til sted/navnetype.

6.3.4. En norsk gazetteer for Galleri Nor

Fram til utviklingen av ADL Gazetteer har mangel på standardisering vært karakteristisk for gazetteers og lignende tjenester. Dette gjelder generelle aspekter vedrørende utvikling av slike tjenester som for eksempel registreringsformat, innhold, tjenestegrensesnitt.

Det er vanskelig å dele data mellom disse ressursene. Standardisering for offisiell gazetteer-arbeid har vært fokusert på å etablere autoritative navneformer for ikke-kommersielle steder. Offisielle gazetteerbyråer har ikke vært involvert i å skape

autoritative grenserepresentasjoner for yttergrenser for steder. Deres formål har vært tilstrekkelig tjent med punktlokasjoner som skiller et sted fra et annet.

Hver gazetteer og lignende tjenester har skapt sitt eget sett av kategorier for stedsnavn, deres eget registreringsformat og måter å beskrive detaljer. I lys av dette kan det standardiseringsarbeidet som gjøres i regi av ADL-prosjektet få stor betydning for utvikling av nye gazetteers. Hvorvidt ADL GCS tilfredsstillende de krav man måtte ha til en gazetteer må diskuteres i forhold til brukere og bruksområde, men så lenge det ikke finnes et bedre forslag til standard så er det etter mitt syn bedre å forholde seg til dette ene forslaget enn ingen. Man kan velge å gjøre egne tilpasninger for å dekke lokale behov.

Kjerneelementer for gazetteer

Kjerneelementer i foreslåtte gazetteer er geografiske navn, navnetyper og lokasjonsinformasjon. Kjerneelementene beskriver en forekomst av et fysisk objekt.

- **Geografiske navn**

En forekomst skal ha minst ett geografisk navn, men en forekomst kan også ha flere varianter av navn. Dette forslaget inkluderer navnevarianter slik at det skal være mulig at gazetteeren, ved søk etter stedsnavn, skal finne en forekomst selv om søketermen er en variant av forekomstens gyldige navn.

- **Navnetyper**

En forekomst skal ha tilordnet minst en navnetype. En navnetype skal si noe om forekomstens natur eller egenskaper. En navnetype kan ha en overordna navnetype som er en generalisering av aktuelle navnetyper. Forekomster i foreslåtte gazetteer tilordnes navnetyper som velges fra et foreslått navnetypeskjema.

- **Lokasjonsinformasjon**

Lokasjonsinformasjon ivaretas av avgrensede bokser samt uklare avtrykk i form av relasjoner mellom forekomster.

Innhold i gazetteer

En gazetteer kan inneholde mye forskjellig informasjon og kan tjene mange ulike bruksområder. Et minstekrav til en gazetteer er geografiske navn og lokasjonsinformasjon om disse. Som et skjult hjelpemiddel i et gjenfinningssystem kan det være interessant å se om en gazetteer kan ha nytteverdi selv med et minimum av innhold. Som for forskjellige typer metadata kan det også for denne typen informasjon være nyttig å basere et forslag på et etablert format eller en standard som kan gi mulighet for kommunikasjon og utveksling av data med andre gazetteers. Innholdsstandarden som er utviklet som en del av ADL Gazetteer, ADL GCS [60] kan være et godt utgangspunkt for å planlegge en gazetteer. ADL GCS definerer 14 hovedelementer der 5 av disse er obligatoriske. Kildeinformasjon kommer i tillegg og er også definert som obligatorisk. En gazetteer som skal fungere som en selvstendig samling vil nødvendigvis ha noe større krav til innhold enn en som primært ska gi støtte for et gjenfinningssystem. For et digitalt bibliotek kan det være ideelt å bygge en gazetteer som kan fungere både som selvstendig samling og integrert i systemer med andre samlinger.

En gazetteer for Galleri Nor er i denne sammenhengen tenkt å brukes som et verktøy for gjenfinning av bilder via geografiske navn og navnetyper. Den skal konsulteres automatisk ved behov i en søkeprosess og den som søker etter informasjon trenger ikke å vite noe om dens eksistens.

En implementasjon av en gazetteer for Galleri Nor vil i denne oppgaven ha et minimum av nødvendig informasjon for at den skal kunne fungere som oppslagstjeneste for et gjenfinningssystem. Datamodellen tar allikevel utgangspunkt i de obligatoriske hovedelementene som er en del av ADL GCS[60]. Enkelte andre elementer er også tatt med.

6.3.5. Elementer fra ADL GCS som inngår i forslaget

Systemet som er implementert i denne oppgaven bruker et minimum av elementer som er nødvendig for å vise noe funksjonalitet. Jeg har valgt å ta med noen flere elementer i dette forslaget da jeg mener disse bør være med for et operasjonelt system. Kildeinformasjon er ikke tatt med. Fullstendig oversikt over ADL GCS finnes i vedlegg B.

- Forekomst_id
- Geografiske navn
 - Stedsnavn
 - Gyldig navn(boolean)
 - Dato_fra
 - Dato_til
- Forekomststype
 - Navnetypeskjema
 - Navnetype
 - Gyldig
 - Dato_fra
 - Dato_til
- Spatial lokasjon
 - Spatial representasjon
 - Avgrensende boks
 - Vest
 - Øst
 - Sør
 - Nord
 - Gyldig(boolean)
- Metadatainformasjon
 - Innført dato
 - Endret dato

Ikke obligatoriske elementer som inngår i forslaget:

- Navnevariant

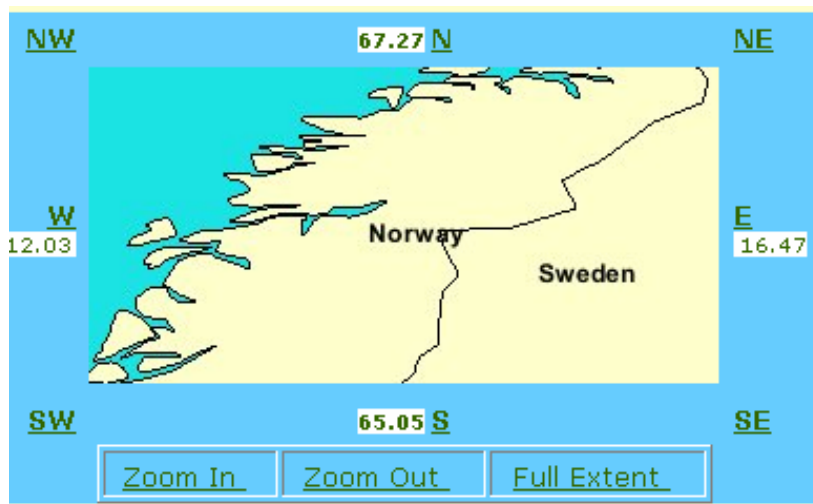
- Stedsnavnvariant
- Gyldig
- Dato_fra
- Dato_til
- Relatert forekomst
 - Type relasjon (er_del_av eller har_deler)
 - Stedsnavn på relatert forekomst
 - Forekomst_id på relatert forekomst
 - Gyldig
 - Dato_fra
 - Dato_til

For dette forslaget er det også tatt med noen av de ikke obligatoriske elementene. Det kan være en fordel å ha muligheten for å registrere navnevarianter. Fotografiene som utgjør samlinga i Galleri Nor er tatt over mange år og for en tid tilbake slik at det kan ha skjedd navneendringer på den tida. I et system som legger ekstra vekt på geografiske navn kan det uansett være en fordel å kunne finne fram også via varierte navn. Elementet *navnevariant* er derfor tatt med. Videre er elementet *relatert forekomst* tatt med. Dette er her begrenset til relasjoner av typen *ErDelAv* og kan gjøre nytte særlig i situasjoner der samme stedsnavn forekommer på flere ulike lokasjoner. Relasjoner av denne typen kan også spille en rolle for selve gjenfinningen da en kan ha mulighet til å finne fram til steder via relaterte forekomster.

Lokasjonsinformasjon

Lokasjonsinformasjon kan være geografiske avtrykk i form av geografiske koordinater og/eller relasjoner mellom geografiske forekomster. Et operasjonelt system som skal tilby tjenester fra en gazetteer bør inkludere både koordinater og relasjoner. Dette er også i tråd med ADL Gazetteer Content Standard.

Maksimum og minimum lengde og breddegrad kan definere avgrensede bokser som kan benyttes for å avgrense det geografiske omfanget av et søk. For et system som skal implementeres i denne oppgaven kan en slik løsning ha uheldige effekter på den måten at det geografiske omfanget blir for stort. Søker man etter kirker i Nordland fylke risikerer man å få med like mange fra Sverige (Figur 6-3). Det er mulig å basere søk på andre typer koordinatbaserte avtrykk, for eksempel polygoner, som kan gi mer nøyaktige grenser for en lokasjon, men dette krever kompliserte databasestrukturer og søkealgoritmer og kan være vanskelig å implementere. Prototypen i denne oppgaven baserer lokasjonsinformasjon på relasjoner mellom forekomster i gazetteeren. Det er hovedsakelig to grunner til det. Først, som nevnt over, ved å bruke avgrensede bokser kan det geografiske omfanget for søket bli for stort, noe som kan påvirke søkepresisjonen. Den andre grunnen, som kanskje er vel så viktig, er at jeg ønsker å undersøke om det er mulig å påvirke søkeresultatet ved å ta i bruk forholdsvis enkle strukturer med et minimum av innhold.



Figur 6-3. Yttergrenser for avgrensede bokser som rammer inn en del av Nord-Norge. Kartutsnitt fra ADL Gazetteer.

6.3.6. Navnetypesaurus for Galleri Nor

Navnetyper

ADL Gazetteer inkluderer ADL FTT. Under utvikling av denne ble det utviklet et navnetypeskjema over foretrukne termer [26]. ADL FTT har 6 topptermer som angir hovedkategorier for navnetyper. Disse er:

- Administrative areas
- Hydrographic features
- Land parcels
- Manmade features
- Physiographic features
- Regions

I Norge er Statens Kartverk ansvarlig for Samordnet Opplegg for Stedfestet Informasjon (SOSI). SOSI - standarden [82], en omfattende standard som omhandler teknikk for datadefinisjoner av geografisk informasjon. I SOSI del 2 finnes blant annet informasjon om stedsnavn [31]. Disse registreres i Sentralt Stedsnavnregister (SSR) som ligger under Statens Kartverk. SSR skal benyttes i navneforvaltning i Statens Kartverk og i offentlig virksomhet som er pålagt å bruke vedtatte skrivemåter. SOSI – standarden definerer navnetyper for å typebestemme geografiske forekomster i 9 hovedkategorier.

Disse er:

- Terrengformer
- Vannkonturer
- Markslag
- Kystdata
- Bebyggelse
- Samferdsel
- Eiendommer
- Administrative områder
- Andre

Navnetypeskjema

Jeg har foreslått et navnetypeskjema for gazetteeren etter modell fra ADL FTT [26]. Dette skal brukes for testing og er tilpasset et subsett av Galleri Nor. Det er ikke gjort forsøk på å utvikle et komplett navnetypeskjema.

Navnetyper fra SOSI-standard benyttes som navnetyper for topptermene i forslaget. SOSI-standard brukes av Statens Kartverk. De øvrige navnetypene er delvis fra SOSI-standard og delvis oversatte navnetypetermer fra ADL typeskjema [31] [26]. Valgte termer for typeskjema er samsvar med de prinsipper for valg av termer som er skissert i kapittel 4.2.4. For eksempel er *religiøse fasiliteter* valgt som foretrukket term som kan være dekkende for mange ulike typer bygninger som kan knyttes til religion, alle foreslåtte navnetyper som representerer tellbare objekter er representert i flertallsform. Foreslåtte typeskjema har 4 topptermer og 26 foretrukne termer i tillegg til topptermene. Oversikt over navnetypehierarki, foretrukne og ikke foretrukne termer finnes i vedlegg E.i, E.ii og E.iii.

ADMINISTRATIVE OMRÅDER
Fylker
Kommuner
Land
BEBYGGELSE
Bebygde områder
Byer
Hovedsteder
Bygninger
Bibliotekbygninger
Institusjonelle steder
Religiøse fasiliteter
Utdanningsfasiliteter
Kommersielle områder
Industriområder
Jordbruksområder
Samferdsel
Brygger
Gater
Jernbanefasiliteter
Havneområder
TERRENGFORMER
Dalfører
Fjell
VANNKONTURER
Elver
Innsjøer
Kystdata
Fjorder
Øyer

Tabell 6-3. Navnetypeskjema

6.3.7. Interoperabilitet med andre samlinger

Registreringsformater for ulike samlinger kan variere sterkt fra system til system, noe som kan gjøre det vanskelig å anvende data på tvers av systemer og samlinger. For å kunne ha muligheten til kommunikasjon mellom ulike systemer og samlinger kan det være behov for å integrere metadata fra ulike kilder. Allmenne brukere som søker etter informasjon for å dekke et spesielt behov er generelt ikke interessert i tekniske eller administrative metadata. De er interessert i metadata i den grad de kan være søkbare og gjøre bruker i stand til å finne fram til den informasjon han eller hun har behov for. De som forvalter samlinger og har ansvar for administrasjon og vedlikehold har gjerne større behov for detaljert informasjon om de samme informasjonsobjektene. For gjenfinningsformål kan det være tilstrekkelig å sørge for interoperabilitet for et subsett av lokale metadataskjema.

En måte å gjøre dette på kan være å benytte et formelt format som en fellesnevner for ulike systemer. En slik fellesnevner kan virke som et globalt format for søk eller for utveksling av data mellom ulike formater.

NKKM's feltkatalogen, som Galleri Nor sitt metadataskjema har som basis, er forholdsvis detaljert og har felter for å angi både beskrivende og administrative metadata. Et subsett av metadata fra Galleri Nors metadataskjema kan stille brukers behov for gjenfinning, og gi nødvendig informasjon om de informasjonsressursene som er funnet til at bruker kan vurdere om ressursene er relevante for vedkommende. Ved å mappe lokale metadataelementer til et formelt format kan en ha mulighet for kommunikasjon med andre systemer og samtidig beholde sine egne lokale data og metadata uforandret.

Kan virke som et globalt format eller for utveksling til andre formater

6.3.8. Dublin Core for Galleri Nor

Galleri Nor sin bruk av stedsspesifisering er et eksempel på at inkonsistens i registreringspraksis fører til navnekonflikter i form av homonymer. Spesifisering av sted for en by kan variere fra stedsnavn for by til bydel eller gatenavn. Dette medfører konflikter i metadata og kan føre til redusert gjenfinningsmulighet.

Grunner for å velge DC er nettopp den brede enigheten om hvordan elementene skal brukes samtidig som DC er fleksibelt i den forstand at en både kan gjenta elementer samt raffinere for å spesifisere disse. En raffinering av elementene kan gjøres slik at en ikke påvirker semantikken i forhold til kjerneelementene. Det at DC er et forholdsvis enkelt format er ikke ensbetydende med at det er mindre brukbart enn andre formater. Det motsatte kan noen ganger vise seg å være tilfelle, både fordi terskelen for å registrere metadata kan senkes til et ikke-ekspert-nivå og fordi mulighetene for elementraffinement gjør at en kan tilpasse løsningen til sitt behov. Jeg vil gå gjennom elementene for å se om en kan dekke behovet for å beskrive de metadata for Galleri Nor som kan være av interesse for brukeren ved å bruke DC.

6.3.9. Diskusjon og forslag til DC metadata for Galleri Nor

Er det mulig å bruke DC for beskrivende metadata for fotografiene i Galleri Nor? DC er opprinnelig utviklet med tanke på oppdaging av dokumenter og dokumentlignende objekter på internett og det er bred enighet blant spesialister på området om at bilder kan defineres som dokumentlignende objekter. Spørsmålet er om DC kan dekke Galleri Nors behov for beskrivende metadata.

Vercoustre og Paradis (V&P) har i sin artikkel Metadata for Photographs: From Digital Libraries to Multimedia Applications [42], redegjort for hvordan de har benyttet DC til å beskrive objekter i en samling fotografier av bygninger. Formålet deres med å beskrive fotografier ved hjelp av DC metadataelementer var å bruke disse for å finne bilder som elementer til automatisk genererte websider til et prosjekt om restaurering av gamle bygninger.

Bildene i Galleri Nor har et større mangfold når det gjelder motiv enn samlinga som V&P refererer til. Wilse-samlinga slik den er i skrivende stund inkluderer 68053 bilder. Mer enn 65000 av disse er produsert av fotografen Anders Beer Wilse (1865-1949) som i sin tid var kjent som Norges fremste fotograf. Han var berømt i inn- og utland, særlig for sine landskapsbilder, stemningsbilder og portretter.

Jeg har, på samme måte som V & P, gått gjennom elementene i DC for å se hvordan disse kan anvendes med tanke på fotografiene i Galleri Nor og spesielt med tanke på å ivareta geografisk referert informasjon. Denne diskusjonen tar ingen hensyn til hvordan dette kan implementeres.

Mange aktuelle problemstillinger dukket opp underveis i denne diskusjonen og den ble noe lengre enn jeg hadde forutsett. Diskusjonen er interessant og jeg vil ikke skjære ned på innholdet, derfor er de viktigste hovedpunktene og beslutningene gjengitt under her mens hele diskusjonen er lagt til vedlegg A.i.

Gjennomgang av elementene i DC v. 1.1 med tanke på fotografiene i Galleri Nor.

Element: Tittel (*Title*)

Definisjon: et navn som ressursen er formelt kjent som

Forslag til bruk:

Ressursnivå:

Samlingsnivå: tittel

Kommentar:

Et bilde / fotografi har ikke nødvendigvis en tittel. Galleri Nor bruker ikke feltet *tittel* i sine beskrivelser.

En tilordning av tittel på ressursnivå må vurderes ut fra hensikt og nytteverdi. På samlingsnivå kan det være greit å kunne angi tittel på samlinga.

Element: Skaper (*Creator*)

Definisjon: en entitet som er hovedansvarlig for innholdet i ressursen

Forslag til bruk:

Ressursnivå: skaper.fotograf

Samlingsnivå:

Kommentar:

Dette elementet brukes for å angi hvem som er fotografen bak bildet. I tilfeller der fotografens mål har vært å avbilde motivet i todimensjonale objekter som tegninger eller maleri kan man diskutere om man yter noen urett ved å si at fotografen er ansvarlig for innholdet i ressursen. Elementet raffineres for å unngå tvilstilfeller som dette.

Element: Emne (*Subject*)

Definisjon: emne for innhold av ressursen

Forslag til bruk:

Ressursnivå: emneord

Samlingsnivå:

Kommentar:

Det er vanligvis ikke vanskelig å foreslå emneord for et bildemotiv. Problemet er å sette kriterier for hva som skal være emne for motivet. Bruker gruppe og bruksområde bør tas i betraktning når man velger emneord.

Element: Beskrivelse (*Description*)

Definisjon: en oppsummering av innholdet i ressursen

Forslag til bruk:

Ressursnivå: beskrivelse

Samlingsnivå: beskrivelse

Kommentar:

Å oppsummere innholdet i et bilde krever andre metoder enn å oppsummere innholdet i en tekst. En fritekst med utgangspunkt i informasjon som allerede finnes i Galleri Nor elementet utfyllende opplysninger kan være et godt utgangspunkt. På samlingsnivå kan en gi en kort beskrivelse av samlinga.

Element: Utgiver (*Publisher*)

Definisjon: en entitet som er ansvarlig for å gjøre innholdet i ressursen tilgjengelig

Forslag til bruk:

Ressursnivå:

Samlingsnivå: utgiver

Kommentar:

Wilsesamlinga administreres og gjøres tilgjengelig av Nasjonalbiblioteket avdeling Rana. Dette vil være felles for alle bildene i samlinga og det vil derfor være tilstrekkelig å registrere dette elementet på samlingsnivå.

Element: Bidragsyter (*Contributor*)

Definisjon: bidragsyter til innholdet i ressursen

Forslag til bruk:

Ressursnivå: bidragsyter.navn
bidragsyter.rolle

Samlingsnivå:

Kommentar:

Dette elementet skal brukes med kvalifikatoren navn for å angi personer som har bidratt til innholdet. Kvalifikatoren rolle brukes for å si noe nettopp om hvilken rolle bidragsyter har for dette bildet.

Element: Dato (*Date*)

Definisjon: en dato som assosieres med en hendelse i ressursens livssyklus

Forslag til bruk:

Ressursnivå:

Samlingsnivå:

Kommentar:

Dateringsinformasjon for bildene i Galleri Nor dekkes av elementet *coverage.temporal*.

Element: Type (*Type*)

Definisjon: innholdet i ressursens natur eller genre

Forslag til bruk:

Ressursnivå:

Samlingsnivå: type

Kommentar:

Element: Type – innholdet i ressursens natur eller genre

Ifølge DCMI Type Vocabulary vil dette, for alle ressursene, være et *image*, altså et bilde.

For Galleri Nor vil det være tilstrekkelig å registrere dette på samlingsnivå.

Element: Format (*Format*)

Definisjon: den fysiske eller digitale manifestasjon av ressursen

Forslag til bruk:

Ressursnivå:

Samlingsnivå: format.medium skjema = IMT

Kommentar:

DC tilbyr her kvalifikatorene extent og medium.

Medium vil være bildeformatet som angis i henhold til skjema Internet Media Type (IMT). Dette vil være JPG for alle bildene i Galleri Nor. Extent sier noe om størrelse eller varighet for ressursen.

Element: Identifikator (*Identifier*)

Definisjon: en entydig referanse til ressursen i en gitt kontekst

Forslag til bruk:

Ressursnivå: identifikator.tilvekstnummer

Samlingsnivå: identifikator.url

Kommentar:

DCMI anbefaler å identifisere en ressurs i henhold til et formelt identifikasjonssystem. Eksempler på slike er URI, URL, DOI og ISBN.

Bildene i Galleri Nor er tilgjengelige via et søkegrensesnitt på www og skal fortsette å være det slik at en kan bruke aktuelle URL, men denne vil i så fall være gyldig på samlingsnivå. Et eget element for å angi samlingstilhørighet kan være nyttig, men dette kan også ivaretas via elementet *relasjon.erDelAv*.

På ressursnivå kan tilvekstnummer gjøre nytten som intern identifikator.

Element: Kilde (*Source*)

Definisjon: en referanse til en ressurs som ligger til grunn for aktuelle ressurs

Forslag til bruk:

Ressursnivå:

Samlingsnivå: kilde

Kommentar:

Ifølge V & P vil kilde være en referanse til det originale foto eller dets negativ. De foreslår å spesifisere et skjema for å holde orden på om originalen arkiveres i en personlig samling eller i en arkiveringsinstitusjon. I denne sammenhengen må man vurdere om man kan registrere dette elementet på samlingsnivå eller om man må registrere det på ressursnivå. I skrivende stud kommer alle bildene fra samme kilde og det vil være tilstrekkelig å registrere elementet på samlingsnivå.

Element: Språk (*Language*)

Definisjon: språk for det intellektuelle innholdet i ressursen

Forslag til bruk:

Ressursnivå:

Samlingsnivå: språk.metadataspråk skjema=ISO 639-2

Kommentar:

Dette elementet er ikke aktuelt å bruke for bilder og deres innhold, men det kan være aktuelt å angi språk for metadata. DCMI tilbyr kvalifikatorer til dette elementet kun i form av kodeskjema for å representere språk, ikke for å raffinere elementet. Det vil være nødvendig å legge til et raffinement her for å spesifisere at det er snakk om metadataspråk. Dette kan registreres på samlingsnivå.

Element: Relasjon (*Relation*)

Definisjon: referanse til en relatert ressurs

Forslag til bruk:

Ressursnivå: relasjon.erDelAv

Samlingsnivå: relasjon.erDelAv

Kommentar:

Aktuelle relasjoner til samlingen som helhet kan her være: Is Part Of og / eller Is Referenced By. Begge relasjonselementene vil ha samme verdi som vil være URL til Galleri Nor. Jeg finner det mest hensiktsmessig å benytte Is Part Of siden alle bildene er del av Galleri Nor.

Element: Dekning (*Coverage*)

Definisjon: omfang / dekningsgrad av innholdet i ressursen

Forslag til bruk:

Ressursnivå: dekning.spatial.stedsnavn

dekning.temporal.fra skjema=W3C-DTF

dekning.temporal.til skjema=W3C-DTF

Samlingsnivå: dekning.temporal.fra skjema=W3C-DTF

dekning.temporal.til skjema=W3C-DTF

Kommentar:

DCMI tilbyr kvalifikatorene Spatial og Temporal for dette elementet og anbefaler, så langt det er mulig, å benytte stedsnavn og navngitte tidsperioder framfor numeriske identifikatorer som geografiske koordinater og datoer.

Begge kvalifikatorene vil sannsynligvis være viktige for alle typer samlinger, men spesielt viktig for en samling som skal være en del av et integrert system med en tilgjengelig gazetteertjeneste. Når det gjelder spatial dekning foreslår jeg å bruke elementet med anbefalt kvalifikator spatial.stedsnavn. Elementet må gjentas tilstrekkelig antall ganger til at en får registrert alle nødvendige stedsnavn for en ressurs.

Når det gjelder Temporal Coverage vil det, for denne bildesamlinga, være vanskelig å følge DCMI's anbefaling om å bruke navngitte tidsperioder. Jeg foreslår å bruke ett element for periodens start og ett element for periodens slutt. Periodene er ikke navngitt, men kun tidfestet med årstall. For bilder som er knyttet til kun ett årstall brukes ikke elementet for periodens slutt. W3C-DTF brukes her.

Element: Rettigheter (*Rights*)

Definisjon: angir hvem som har rettigheter i forhold til ressursen

Forslag til bruk:

Ressursnivå:

Samlingsnivå: rettigheter

Kommentar:

Dersom de samme rettighetene gjelder for alle bildene i samlingen kan dette elementet brukes ukvalifisert på samlingsnivå for å angi rettighetshaver.

6.3.10. Oversikt over DC metadataforslag til Galleri Nor

Metadataelementer på samlingsnivå

Galleri Nor er Nasjonalbibliotekets fotodatabase og tar sikte på å presentere materiale fra flere sentrale aktører i norsk fotohistorie. Da mesteparten av dette ble skrevet var kun en samling tilgjengelig for søk i Galleri Nor. Etter relansering av i 2004 inneholder Galleri Nor fotografier fra flere forskjellige fotografer og forskjellige eiere. En kan søke etter fotografiene etter hver fotograf som egen samling, etter hver eier som egen samling eller på hele Galleri Nor som en samling.

Fotografiene er tilgjengelige fra fotodatabasen via galleri Nor sin søkeside. For å kunne gi gode forslag til metadataskjema er det viktig å definere samlingsbegrepet. Hva er en samling i denne sammenhengen? Er Galleri Nor en samling eller er det en samling for hver fotograf eller en for hver eier? Et forslag til samlingsmetadata i denne sammenhengen tar ikke stilling til dette problemet. Det er mulig at en her skulle hatt et skjema for et overordna samlingsnivå for hele Galleri Nor på toppen og et for hver samling.

Forslag til metadataskjema på samlingsnivå

Dc Element	Verdi
Tittel	
Beskrivelse	
Utgiver	Nasjonalbiblioteket Rana
Type	fotogarafo
Format.medium skjema=IMT	jpg
Kilde	
Identifikator.url	
Språk.Metatataspråk skjema=ISO 639-2	Norsk
Relasjon.erDelAv	
Dekning.Temporal.fra skjema=W3C-DTF	
Dekning.Temporal.til skjema=W3C-DTF	
Rettigheter	

Tabell 6-4. Forslag til samlingsmetadata for Galleri Nor basert på Dublin Core.

Metadataelementer på ressurnivå

Forslag til metadata på ressurnivå for Galleri Nor. Forslaget er begrenset til de metadata som er søkbare og /eller synlige for bruker ved visning av en valgt ressurs.

Forslag til metadataskjema på ressurnivå

Dc Element	Verdi
Skaper.Fotograf	
Emne.Navnetype	
Emne.Emneord	
Beskrivelse	
Bidragster.Navn	
Bidragster.Rolle	
Identifikator.tilvekstnummer	
Relasjon.erDelAv	
Dekning.Spatial.stedsnavn	
Dekning.Temporal.fra skjema=W3C-DTF	
Dekning.temporal.til skjema=W3C-DTF	

Tabell 6-5. Forslag til metadata på ressurnivå for Galleri Nor basert på Dublin Core

Foreslåtte metadataskjema er ikke tatt direkte i bruk i mitt system, men de kan legges til grunn for beskrivelser av informasjonsobjekter ved en eventuell utvidelse av prototypen.

7. Prototyping

Prototypen integrerer en gazetteer i et system med en bildesamling. Gazetteeren inkluderer en navnetypesaurus. Gazetteeren er integrert i systemet slik at bruker ikke behøver å vite noe om dens eksistens og den konsulteres automatisk dersom en forespørsel fra bruker inkluderer spørring etter stedsnavn eller navnetype. Forespurt stedsnavn alene eller i kombinasjon med navnetype påvirker det geografiske omfanget for et søk. Konsultasjon av gazetteer genererer automatisk utvidelse av spørring. Det er implementert funksjonalitet for å kunne søke etter bilder via geografiske navn og/eller navnetyper. For å kunne søke via navnetyper må bruker velge søke-term fra en hierarkisk indeks i brukergrensesnittet.

Søk som resulterer i funn av ett eller flere bilder, returnerer aktuelle bilder til bruker med metadata som angår geografiske steder og navnetyper i tillegg til en identifikator for hvert bilde. Søk som ikke resulterer i funn av bilder, returnerer en melding om dette til bruker sammen med en lenke tilbake til søkesiden.

Brukergrensesnitt for forespørsler og resultatservice er implementert med tanke på å vise muligheter og har begrenset funksjonalitet.

JSP og Java er valgt som programmeringsspråk for prototypen. Java tilbyr gode muligheter for gjenbruk av kode og ferdige klasser og metoder er tilgjengelige for bruk som de er eller med egne modifiseringer. JSP utgjør grensesnittet mot bruker mens det meste av kontroll og logikk håndteres av et bakenforliggende system som består av Java-filer. Her foregår også kommunikasjon med databaser.

Testdokumenter i prototypen er 38 bilder fra Galleri Nor som alle tilhører Wilse-samlinga. Bildene i testsamlinga er valgt ut spesielt med tanke på at de skal brukes i et system med en tilgjengelig gazetteer og alle bildene kan knyttes til steder i Norge. Et bilde er tilordnet en *sted_id* for hver av de geografiske forekomster som, representert av et stedsnavn, er registrert i feltet *Sted* eller i feltet *Egennavn* i opprinnelige metadata. Stedsnavn som er registrert i feltet *Egennavn* i Galleri Nor representerer stort sett navngitte bygninger. Ingen nye metadata er direkte tilordnet bildene. Mer informasjon om bildene i testsamlinga og om metadata om disse finnes i kapittel 7.8.5 og kapittel 7.8.6. Samtlige bilder i testsamlinga og opprinnelige metadata fra Galleri Nor finnes i vedlegg D.

Opprinnelige metadatabeskrivelser for bildene i Galleri Nor, som angår geografiske steder og navnetyper, danner grunnlaget for 122 geografiske forekomster i gazetteeren. Galleri Nor benytter ikke navnetyper systematisk, men i en del av beskrivelsene kan en finne søkeord som kan knyttes til navngitt sted og som sier noe om stedets natur. I denne oppgaven er det lagt vekt på at ethvert navngitt sted skal tilordnes en navnetype som registreres i gazetteeren. Vedlegg F inneholder resultat fra en SQL-spørring til gazetteer-databasen som viser alle forekomster med stedsnavn og navnetype.

Navnetyper er tilordnet og vurdert i forhold til tidligere foreslåtte navnetypeskjema (se kapittel 6.3.6). Navnetypesaurus har 31 foretrukne termer og 39 ikke foretrukne termer. Ikke foretrukne termer er begrenset til termer som finnes som søkeord i opprinnelige metadatabeskrivelser for bildene og som kan brukes for å typebestemme et

navngitt sted. Mer detaljert informasjon om stedsnavn og navnetyper i gazetteeren finnes i kapittel 7.8.7 og kapittel 7.8.8.

Lokasjonsinformasjon i form av koordinater for geografiske avtrykk er ikke implementert i prototypen. Systemet benytter relasjoner mellom geografiske forekomster for å avgrense det geografiske omfanget av et søk.

7.1. Implementasjon i forhold til identifiserte kriterier

Identifiserte kriterier for å integrere gazetteers og tesauri i digitale bibliotek (se kapittel 6.2.3) er vurdert og tatt hensyn til på følgende måte:

- Forslag til gazetteer metadata tar utgangspunkt i ADL GCS
- Navnetyper tar utgangspunkt i navnetyper fra SOSI standard som brukes i Sentralt Stedsnavnregister.
- Tidsaspekt kan ivaretas ved dateringer for gyldighet, dette gjelder flere forhold som navn, navnetype, relasjon osv. Dette er også tatt med i forslaget i kapittel 6.3.5, men er ikke implementert da det ikke vil ha noen betydning for prototypens funksjonalitet.
- Uklare avtrykk (fuzzy footprints, se kapittel 6.2.3) blir ivaretatt ved å opprette relasjoner mellom forekomster. Relasjoner mellom forekomster er av typen *er_del_av* og gir de geografiske forekomstene en hierarkisk struktur som kan utnyttes i søkesystemet.
- Kvalitetsaspekter som går på nøyaktighet for lokasjonsinformasjon har mindre betydning i denne oppgaven og er ikke tatt hensyn til.
- Spatiale aspekter kan ivaretas ved å benytte koordinater for geografiske avtrykk i form av avgrensede bokser. Lokasjonsinformasjon i form av geografiske koordinater er ikke implementert i prototypen, men er tatt med i forslaget i kapittel 6.3.5.
- Flerspråklighet kan muliggjøres ved for eksempel å oversette mellom SSR navnetyper og adl feature types. Det er ikke gjort noe forsøk på å oppnå flerspråklighet i prototypen.

7.1.1. Andre forhold

Det kan forekomme unøyaktigheter og mangler i forbindelse med tilordning av navnetyper til geografiske navn, men alle geografisk navn som er registrert har fått tilordnet en navnetype slik at det skal være registrert tilstrekkelige data til å kunne teste systemet.

Tilordning av navnetyper er gjort først og fremst for å ha mulighet til å teste systemet og omfatter kun deler av et navnetypeskjema som et operasjonelt system vil ha behov for. Dersom det skulle være aktuelt å utvikle et operasjonelt system basert på et prosjekt som dette anbefales det å diskutere og eventuelt revurdere deler av de beslutningene som er tatt for denne prototypen. For eksempel er *Nidaros* tatt med som navnevariant til *Trondheim*, mens det ikke er registrert som gyldig navn med navnetype som indikerer at det er et bispedømme. Videre er *Nidarosdomen* registrert med relasjon til *Bispegata* mens den kanskje heller skulle vært registrert med relasjon til *Nidaros bispedømme*, eventuelt begge deler. Denne type feil og mangler kan det være flere av i prototypen, men dette er forhold som ikke behøver å ha noen innvirkning på det som er målet med denne oppgaven.

Prototypen er en forenkling av et system slik det kan framstå som et operasjonelt system. Implementert funksjonalitet er begrenset til et minimum som er nødvendig for å vise hvordan en kan påvirke gjenfinningsmulighetene ved å integrere en gazetteer i et system med samlinger.

7.2. Funksjonelle krav til prototypen

Krav til funksjonalitet er begrenset til den funksjonalitet som er nødvendig for å vise noen av mulighetene et system som dette kan gi. Der er ikke definert egne krav til gazetteer og tesaurus, noe som vil være nødvendig å gjøre for et operasjonelt system. Prototypen viser kun muligheter som kan knyttes til integrasjon av gazetteer og navnetypesaurus. Det er ikke implementert funksjonalitet for å søke via andre søkeelementer som *datering*, *egennavn*, *emneord* osv.

7.2.1. Søk via stedsnavn

En skal kunne søke via stedsnavn og skal da ha mulighet til å finne alle bilder som er registrert med steder som representeres av aktuelle stedsnavn. Dette innebærer at en også skal finne bilder som er registrert med steder som kan sies å tilhøre eller være del av sted som stedsnavnet representerer. For eksempel ved søk via *Oslo* skal en ha mulighet for å finne alle bilder med motiv fra Oslo uavhengig om det er Oslo fylke, Oslo kommune eller Oslo by. Bilder som ikke er registrert med sted som har stedsnavn *Oslo*, men for eksempel med sted som har stedsnavn som *Nordstrand*, *Stortorvet* eller *Nationaltheatret* skal også finnes ved å søke via stedsnavn *Oslo*.

7.2.2. Søk via navnetype

En skal kunne søke via navnetype og skal da ha mulighet til å finne alle bilder som er registrert med sted som er av aktuelle navnetype eller smalere type. For eksempel ved å søke via navnetype *bygninger* skal en finne alle bilder som er registrert med stedsnavn som er av type *bygninger* samt de som er av type *bibliotekbygninger*, *religiøse fasiliteter*, *utdanningsfasiliteter* og andre spesialiserte typer bygninger som er registrert. Ved for eksempel å søke via *by* skal en finne alle bilder fra alle byer, også bilder som er registrert med stedsnavn som representerer bydeler, gateadresser eller navngitte bygninger som er del av en by.

7.2.3. Søk via stedsnavn og navnetype

En skal kunne søke via stedsnavn i kombinasjon med navnetype og skal da ha mulighet til å finne alle bilder som er registrert med sted som har aktuelle stedsnavn og er av aktuelle navnetype eller smalere type samt alle bilder som er registrert med sted som kan sies å være del av sted som representeres av det stedsnavn det søkes via og som er av aktuelle type eller smalere type. For eksempel ved søk etter *Oslo og byer* skal en ha mulighet til å finne alle bilder med motiv fra Oslo by, også de som er registrert med stedsnavn som *Nordstrand* eller *Stortorvet*, har navnetype *bydeler* og som kan sies å tilhøre eller være del av Oslo by.

7.2.4. Navnevarianter

Det skal være mulig å finne bilder også ved søk via navnevarianter. For eksempel ved søk via *Trondhjem* eller *Nidaros* skal en finne de samme bildene som om en søker via *Trondheim*. Et søk via stedsnavn skal automatisk inkludere oppslag i navnevarianter da et gyldig navn for en forekomst kan være navnevariant for andre forekomster.

7.3. Ikke funksjonelle krav

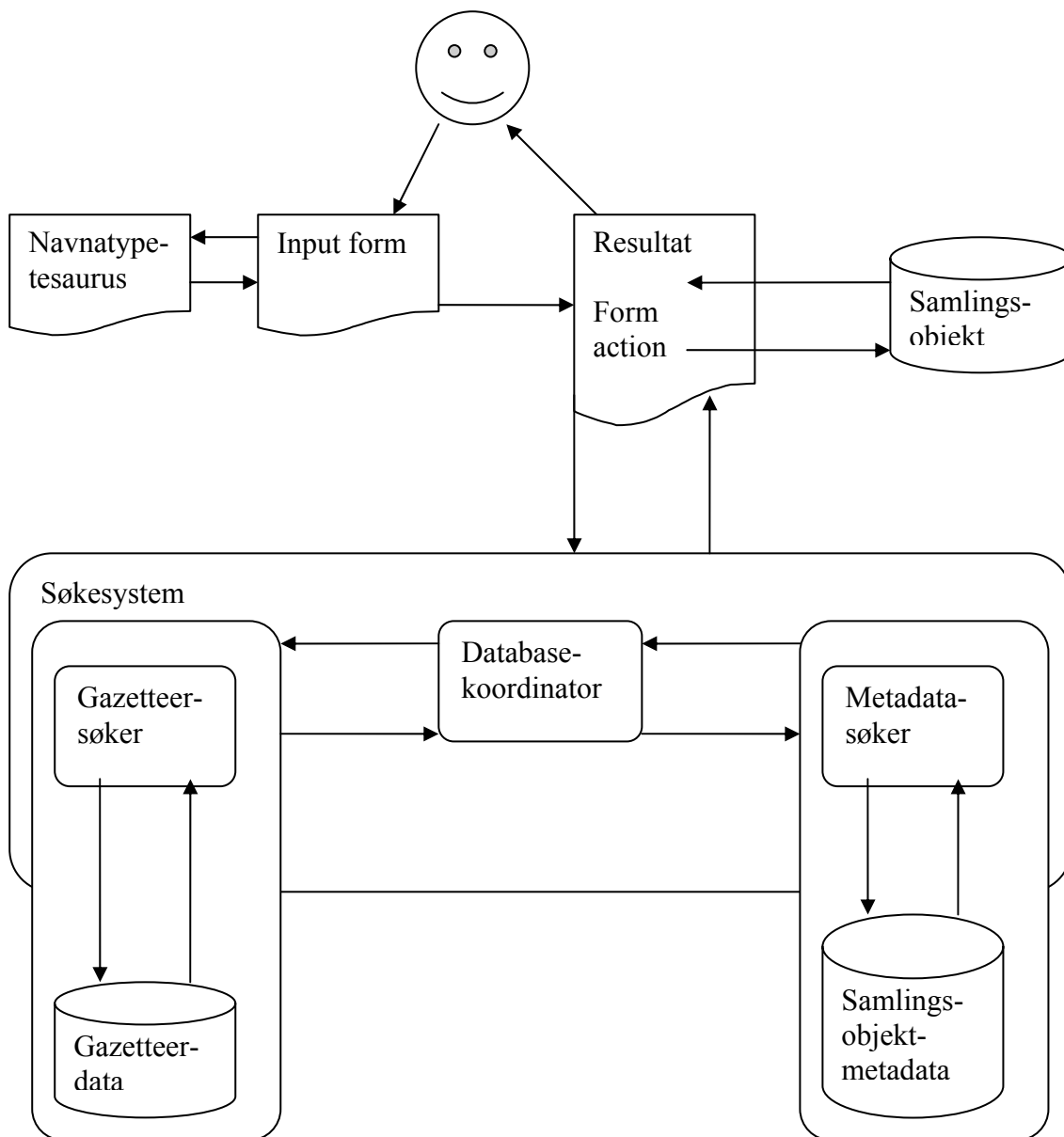
Prototypen er implementert for å kunne teste noen av mulighetene et slikt system kan gi. Det er ikke satt spesielle krav til brukergrensesnitt utover de som er nevnt over. Det er heller ikke spesielle krav til responstid og ytelse for systemet. Det er heller ikke spesielle krav til systemarkitektur, men noen prinsipper ligger til grunn for at systemet er blitt som det er.

7.4. Teknologi for implementasjon

Java er valgt som programmeringsspråk for implementasjon av prototypen. Objektorientert tilnærming og en god del tilgjengelig åpen kildekode for Java-pakker motiverer til gjenbruk av kode samtidig som den objektorienterte tilnærmingen gir gode forutsetninger for modularisering av koden. Java kan gjøres tilgjengelig for bruk via Internet ved å bruke servlets eller Java Server Pages (JSP). JSP er valgt som grensesnitt mot bruker. I JSP-filer kan html-kode kombineres med Java-kode som igjen kan kommunisere med Java-klasser i et bakenforliggende system. Relasjonsdatabaser og MySQL er valgt som databaseløsning. All kommunikasjon med databaser foregår fra klasser i det bakenforliggende system.

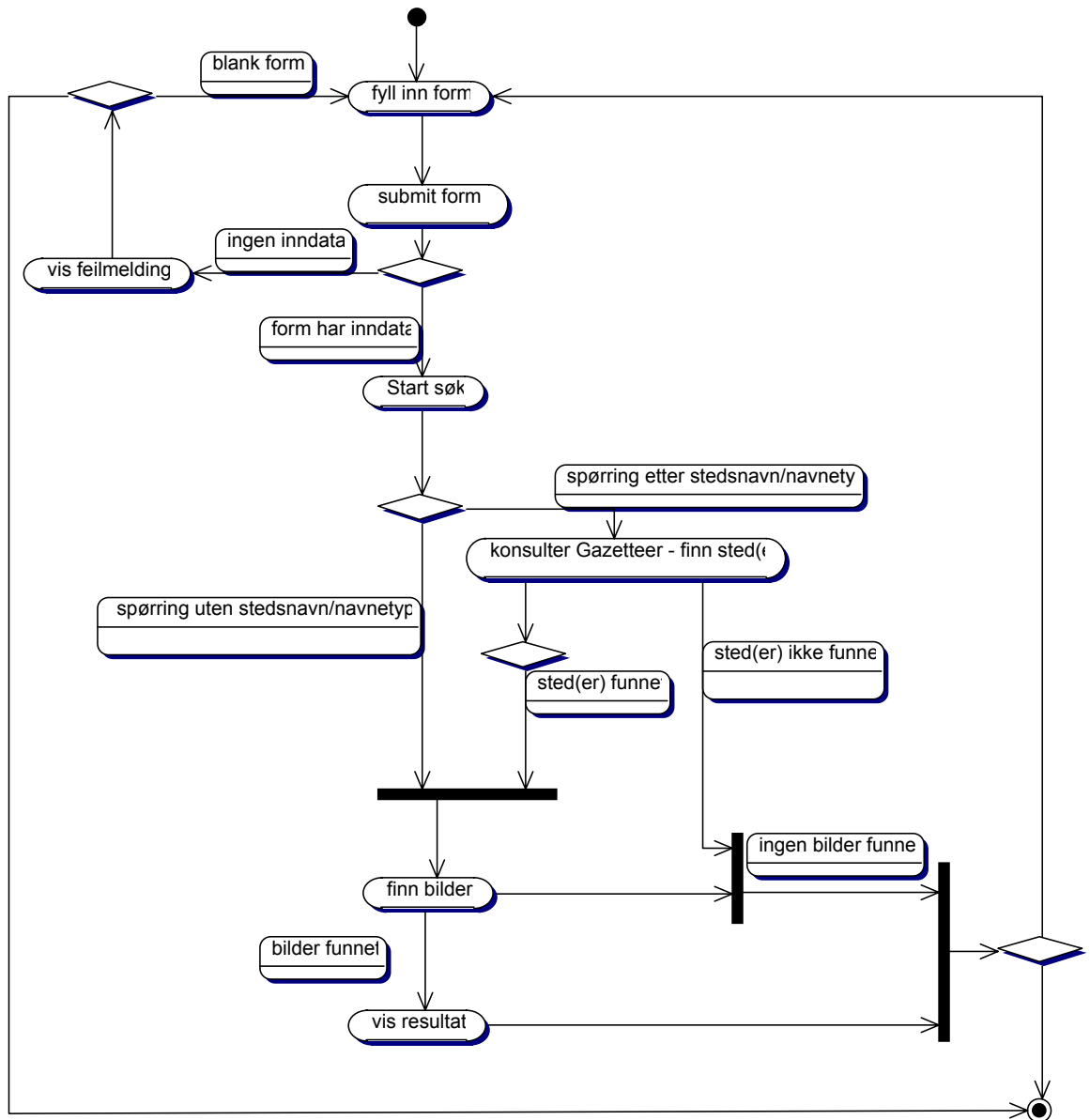
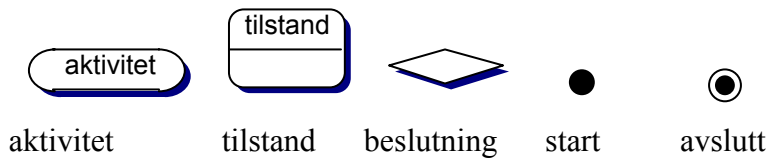
7.5. Arkitektur for prototypen

Jeg har lagt noe vekt på å gjøre systemet modulbasert slik at eventuelle endringer kan gjøres uten at en må gjøre om på hele systemet. Jeg er klar over at dette kan gjøres bedre, men det fungerer for mitt formål. Ved å gjøre noen endringer og legge til noen metoder i et par av klassene skal det være mulig å utvide søkefunksjonaliteten til systemet uten altfor mye arbeid. Meningen er at det skal være mulig å utvide gazetteeren og bruke denne uavhengig av resten av systemet. Det bør også være mulig å legge til flere samlinger. Dette vil nok kreve noe mer omfattende redesign av systemet. Figur 7-1 viser en overordna systemmodell. Brukergrensesnittet kommuniserer med et underliggende system. I dette systemet fungerer en klasse som en koordinator for videre søk. All kommunikasjon mellom brukergrensesnitt og søkesystem går via koordinatoren. Kommunikasjon mellom gazetteeren og resten av søkesystemet går også via denne koordinatoren. En egen klasse fungerer som søkeenhet for gazetteeren, mens en klasse fungerer som søkeenhet for bildemetadata. bakgrunn av de metadata som eventuelt returneres fra søkesystemet genereres en url som peker til aktuelle samlingsobjekt



Figur 7-1. Overordna systemmodell

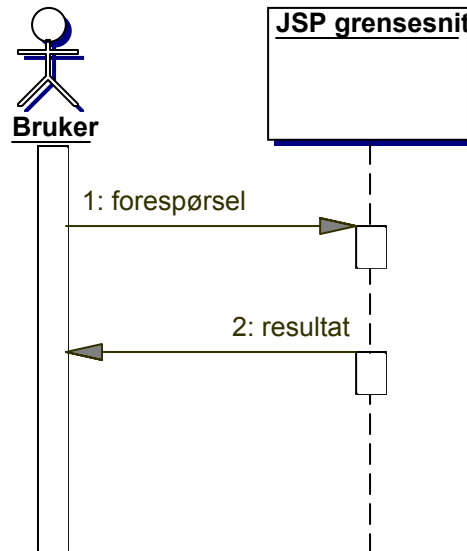
De neste delkapitlene forklarer de forskjellige lagene noe nærmere. Et aktivitetsdiagram (se Figur 7-2) viser logiske veier gjennom systemet slik det fungerer nå. Symbolene som brukes i diagrammet er som følger:



Figur 7-2. Aktivitetsdiagram som viser mulige veier gjennom systemet

7.6. Grensesnitt mot bruker

To JSP-filer, `form.jsp` og `resultat.jsp`, utgjør grensesnittet mot bruker. JSP-filene ligger i en katalog på `stud.ntnu.no/~maro/public_html/gn_system` og kommuniserer med Java-klasser i et bakenforliggende system.



Figur 7-3. Bruker kommuniserer med systemet via et web-grensesnitt ved å sende en forespørsel og får presentert søkeresultat.

7.6.1. Søkeform for forespørsel fra bruker

JSP-filen `form.jsp` inneholder blanding av JSP-kode og HTML-kode og utgjør grensesnittet for formulering av forespørsel fra bruker. Fullstendig skript for brukergrensesnittet finnes i vedlegg H.iii.

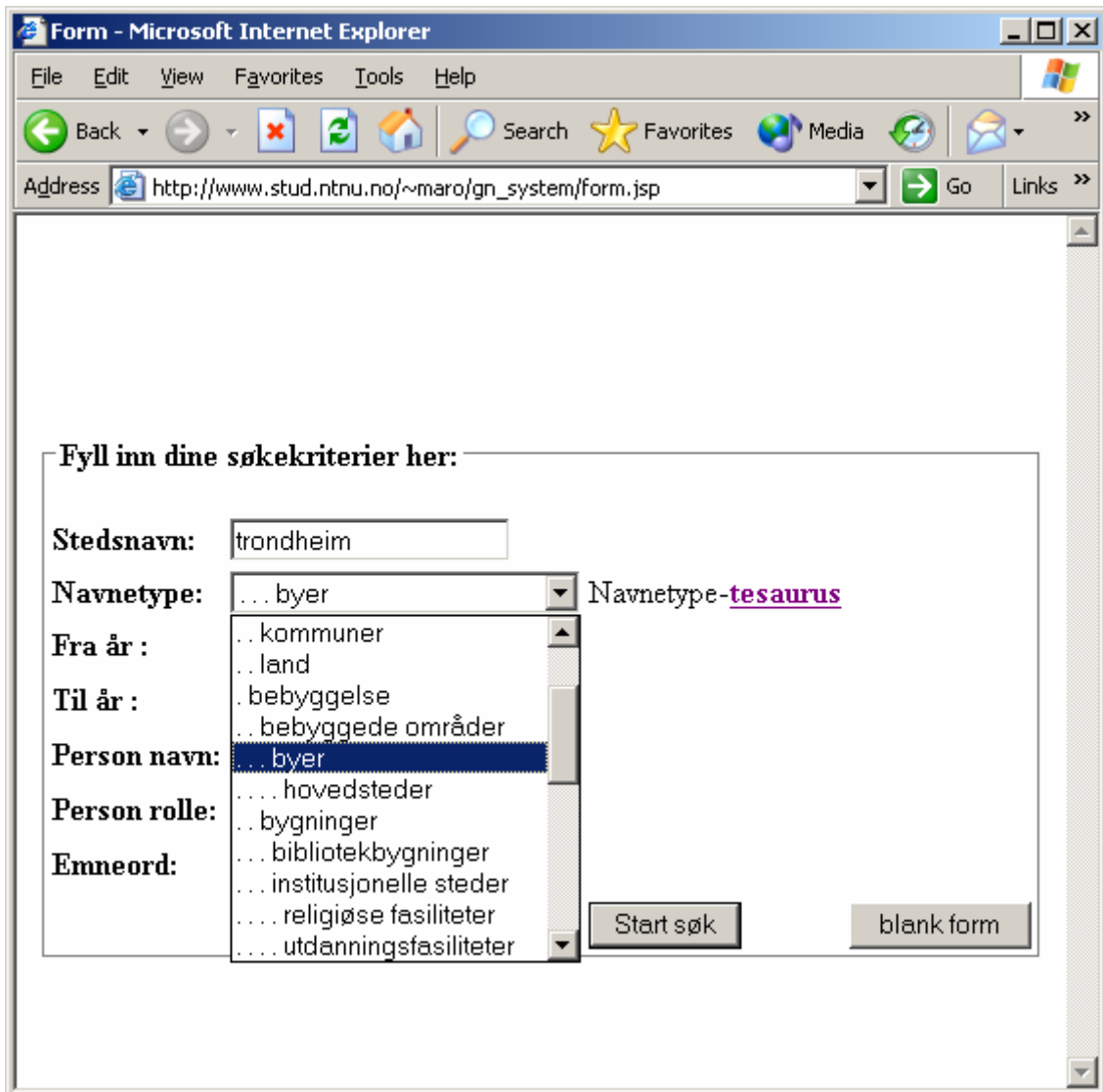
En html-form har felt der bruker kan fylle inn søkekriterier. Formen har to knapper, en for å sette i gang søk og en for å tømme feltene i formen.

The screenshot shows a Microsoft Internet Explorer window titled "Form - Microsoft Internet Explorer". The address bar contains the URL "http://www.stud.ntnu.no/~maro/gn_system/form.jsp". The main content area displays a search form with the following elements:

- A heading: "Fyll inn dine søkekriterier her:"
- A text input field for "Stedsnavn:".
- A dropdown menu for "Navnetype:" with the current selection "velg type" and a link "Navnetype-tesaurus".
- Text input fields for "Fra år :", "Til år :", and "Person navn:".
- A dropdown menu for "Person rolle:" with the current selection "alle roller".
- A dropdown menu for "Emneord:" with the current selection "alle roller" and other options: "fotograf", "eier", and "avbildet".
- Two buttons at the bottom right: "Start søk" and "blank form".

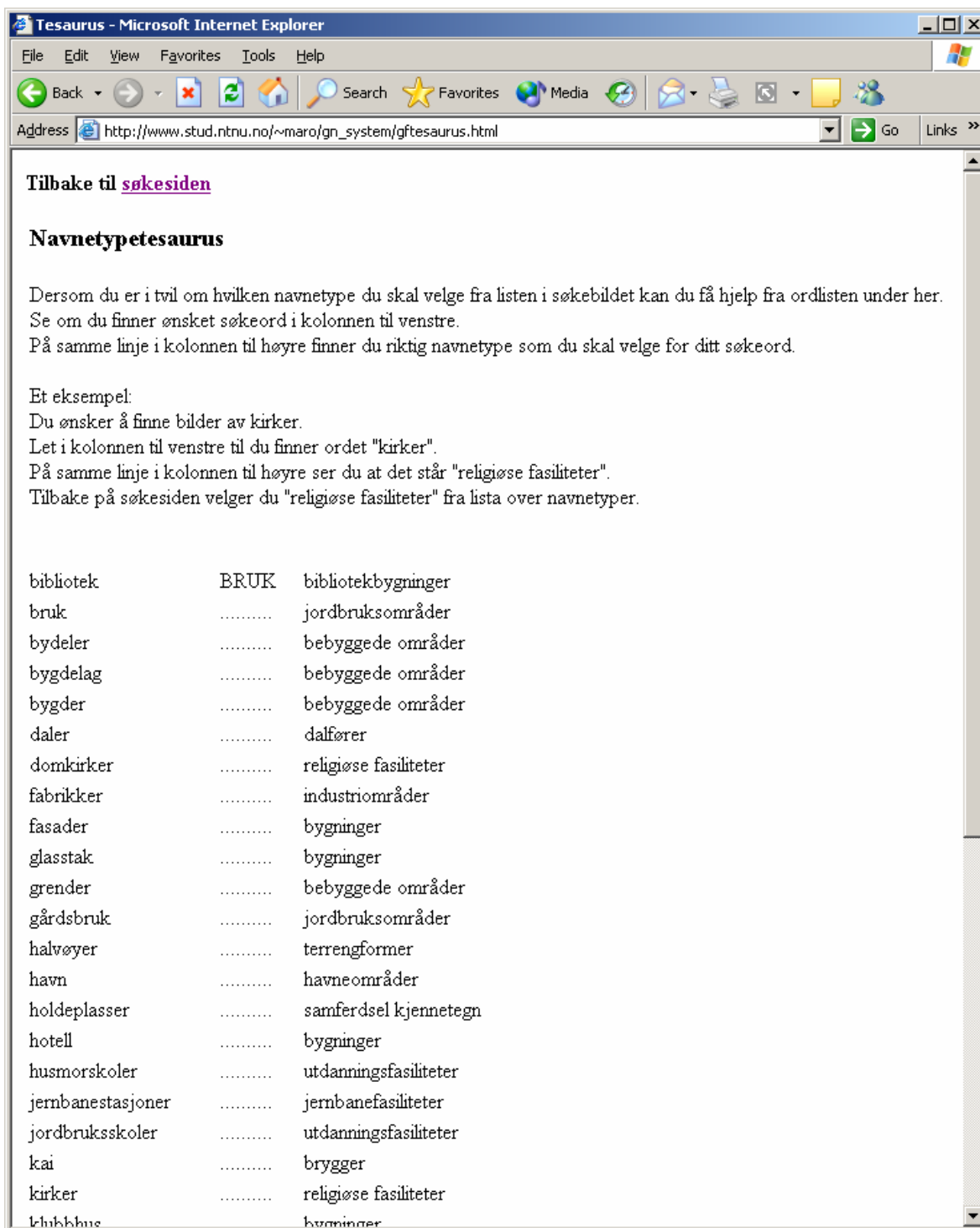
Figur 7-4. Brukergrensesnittets søkeform.

Feltet *Person rolle* kan være ubegripelig for bruker dersom vedkommende ikke vet noe om hva slags roller det her er snakk om, men det kan allikevel være nyttig å ha muligheten til å knytte roller til personer. En rolle i denne sammenhengen kan for eksempel være fotograf, avbildet eller eier. Det kan derfor være nyttig å la bruker velge fra en liste her. Dette er også mulig i eksisterende system for Galleri Nor. Foretrukne navnetyper er listet i en hierarkisk indeks som bruker må velge fra dersom det skal søkes via navnetyper.



Figur 7-5. Hierarkisk indeks over navnetyper i søkebildet.

Søkesiden tilbyr også en lenke til en navnetypesaurus. Navnetypesaurus er her begrenset til en html-fil som viser en liste over ikke foretrukne termer med tilsvarende foretrukne termer.



Figur 7-6. Del av navnetypesaurus som er tilgjengelig via en lenke fra søkesiden.

7.6.2. Presentasjon av søkeresultat


Ved å aktivere knappen for å starte søk aktiveres filen `resultat.jsp`. Denne filen inkluderer filen `incl1.jsp` via kommandoen `include`. Dette gjøres for å sende med passord som er nødvendig for å opprette kontakt med databasene. Filen `resultat.jsp` kommuniserer med de bakenforliggende Javaklasser, behandler søkeresultat og presenterer dette for bruker. Presentasjon for bruker forberedes i denne filen. Ut fra søkeresultatet presenteres tilgjengelige metadata for bruker. Metadata er her begrenset til informasjon om identifikatorer samt registrerte stedsnavn og navnetyper for de bilder som er funnet. Ut fra identifikatorer til de bilder som er funnet genereres en lenke til lokasjon for aktuelle bilde slik at dette kan presenteres sammen med tilgjengelig metadata.

Søkeresultat:

Antall bilder funnet : 4 [Tilbake til søkesiden](#)


Bilde 1 av 4

DC.Identifikator.Bilde Id: 3
DC.Dekning.Spatial.Stedsnavn : Trondheim Jernbanestasjon
DC.Emne.Navnetype : jernbanefasiliteter
DC.Dekning.Spatial.Stedsnavn : Trondheim Jernbanestasjon
DC.Emne.Navnetype : bygninger
DC.Dekning.Spatial.Stedsnavn : Trondheim
DC.Emne.Navnetype : kommuner
DC.Dekning.Spatial.Stedsnavn : Trondheim
DC.Emne.Navnetype : byer
DC.Dekning.Spatial.Stedsnavn : Fosenskaia
DC.Emne.Navnetype : havneområder
DC.Dekning.Spatial.Stedsnavn : Sentrum
DC.Emne.Navnetype : bebyggede områder



Bilde 2 av 4

DC.Identifikator.Bilde Id: 41
DC.Dekning.Spatial.Stedsnavn : Bispegata
DC.Emne.Navnetype : gater
DC.Dekning.Spatial.Stedsnavn : Nidarosdomen
DC.Emne.Navnetype : religiøse fasiliteter
DC.Dekning.Spatial.Stedsnavn : Trondheim
DC.Emne.Navnetype : kommuner
DC.Dekning.Spatial.Stedsnavn : Sør-Trøndelag
DC.Emne.Navnetype : fylker
DC.Dekning.Spatial.Stedsnavn : Trondheim
DC.Emne.Navnetype : byer
DC.Dekning.Spatial.Stedsnavn : Sentrum
DC.Emne.Navnetype : bebyggede områder



Figur 7-7. Presentasjon av søkeresultat.

7.7. Java-klasser

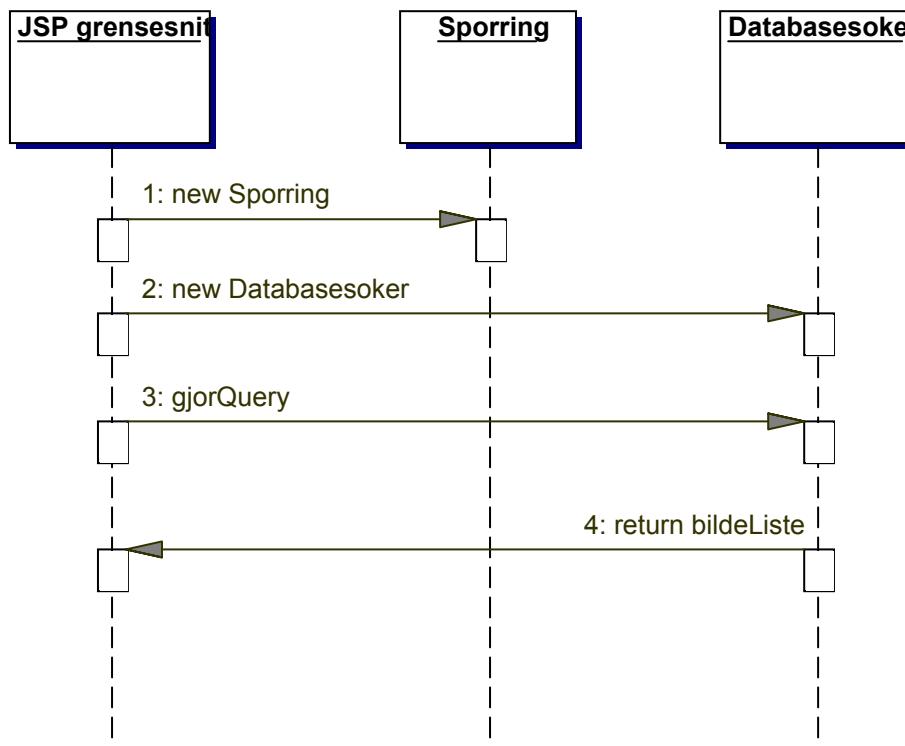
Java-klasser tar seg av mesteparten av kontroll og logikk og håndterer all kommunikasjon med databasene i systemet. For å opprette kontakt mellom JSP-koden og Java-klassene er disse lastet opp til en *web-inf* katalog på *Resin JSP-server* som kjører på *jsp-stud.idi.ntnu.no:8080*.

Java-koden er organisert i syv klasser der noen av disse kun har til oppgave å bære og gi tilgang til informasjon.

Alle klassene ligger i samme pakke, men ikke alle klassene kommuniserer med hverandre. Funksjonaliteten er begrenset til kun å ta imot forespørsler om stedsnavn og navnetyper, men systemet er konstruert slik at det ikke skal kreve mye arbeid å utvide funksjonaliteten. Klassediagram og dokumentasjon av Java-kode finnes i vedlegg G og fullstendig kildekode finnes i vedlegg H.iv.

7.7.1. Kommunikasjon mellom JSP og Java-klasser

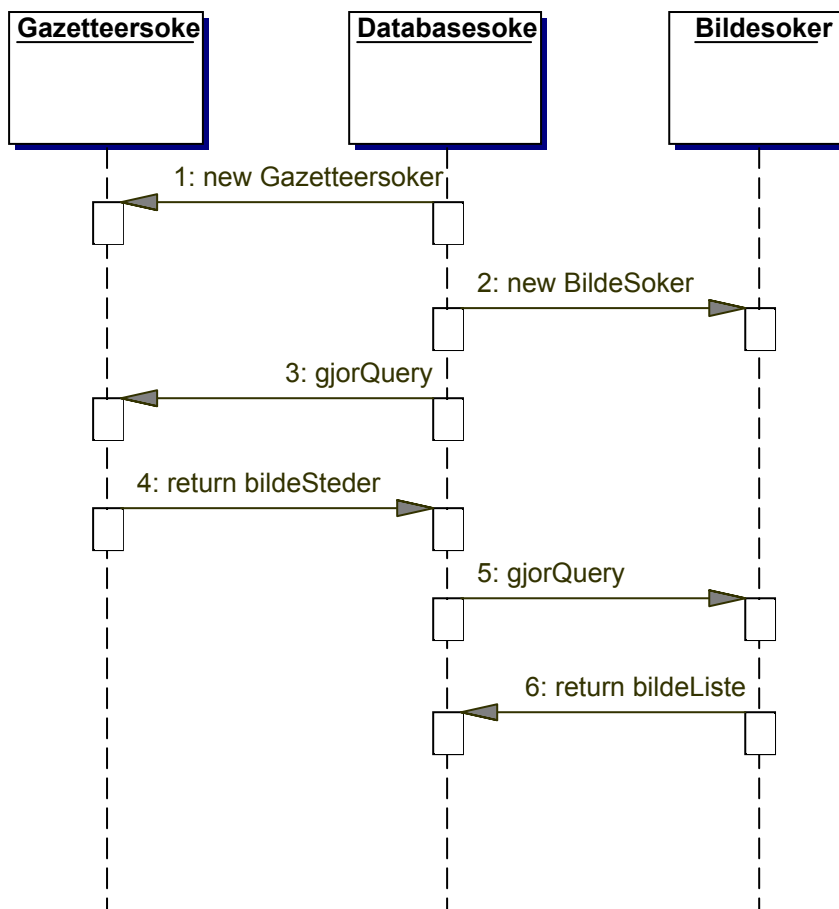
Ved forespørsel fra bruker kalles konstruktoren i klassen *Sporring* i det bakenforliggende system. Et objekt som holder informasjon fra brukers forespørsel returneres til JSP-koden før det sendes som argument til metoden *gjorQuery* i klassen *Databasesoker* for å sette i gang en søkesevens. Dette kunne muligens vært gjort på en bedre måte, men det fungerer greit nok i denne sammenhengen. Bortsett fra instantiering av objektet av typen *Sporring* kommuniserer JSP-koden kun med klassen *Databasesoker*. Klassen *Databasesoker* returnerer en *ArrayList* som holder objekter av typen *Bilde* tilbake til JSP-koden i fila *resultat.jsp*.



Figur 7-8. Brukergrensesnittet kommuniserer med klassene Sporning og Databasesoker.

7.7.2. Kommunikasjon mellom Java-klassene

Klassen `Databasesoker` fungerer som et bindeledd mellom gazetteerdelen og den delen som søker etter ressursene som bruker etterspør. Ved spørring etter stedsnavn og /eller navnetype kalles metoden `gjorQuery` i klassen `Gazetteersoker` først. Denne returnerer en `ArrayList` som inneholder null, ett eller flere objekt av typen `Sted`. Dersom listen er tom, det vil si at det ikke finnes bilder som er registrert med steder eller typer som er etterspurt, returneres en tom `ArrayList` tilbake til resultat.jsp. Bruker får da melding om at ingen bilder er funnet og kan eventuelt endre sine søkekriterier. Dersom listen ikke er tom sendes den som argument til metoden `gjorQuery` i klassen `Bildesoker` som da kan hente ut informasjon om steder som legges til kriteriene for søk etter metadata for bilder.



Figur 7-9. Klassen `Databasesoker` kommuniserer med klassene `Gazetteersoker` og `Bildesoker`.

7.8. Databaser

MySQL er valgt som databaseløsning. Det er opprettet to relasjonsdatabaser, en for gazetteeren og en for bildesamlinga. Databasene ligger på serveren *mysql.stud.ntnu.no* og aksesseres fra Java-klassene *Gazetteersoker.java* og *Bildesoker.java* via *mysql:jdbc* som er en Java databaseforbindelse.

Databasene er noe forenklet i forhold til forslaget i kapittel 6.3.5. Det er kun opprettet tabeller som er nødvendige for å ivareta de funksjonelle kravene til systemet slik at det kan testes. Databaseskript finnes i vedlegg H.i og H.ii.

7.8.1. Gazetteer

Databasen *maro_GAZETTEER* har følgende seks tabeller:

Steder (*sted_id*, *stedsnavn*, *navnetype*)

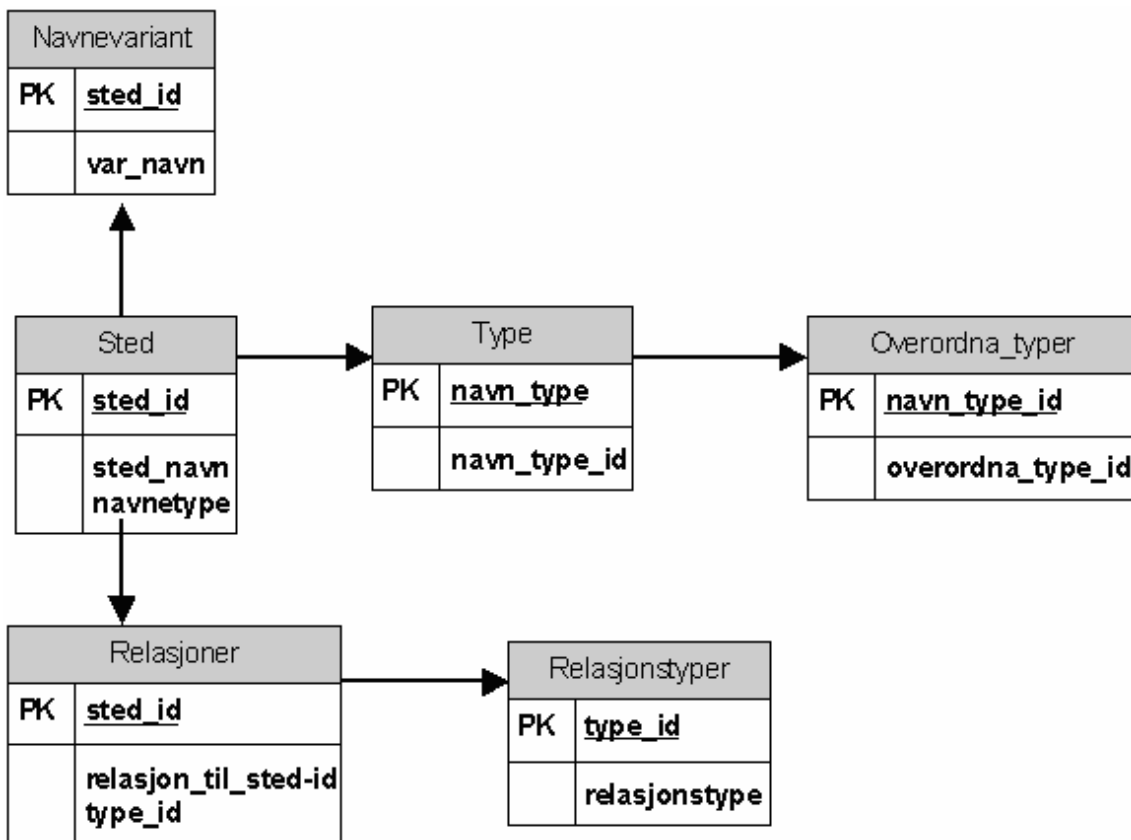
Navne_varianter (*navn*, *sted_id*)

Typer(*type_id*, *type_navn*)

Overordna_typer(*type_id*, *overordna_type_id*)

Relasjoner(*sted_id*, *relasjon_til_sted_id*, *relasjonstype_id*)

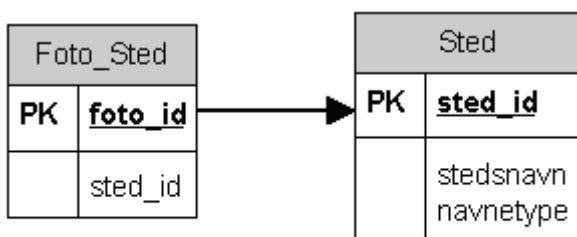
Relasjonstyper(*relasjonstype_id*, *relasjonstype*)



Figur 7-10. Databaseskjema for Gazetteeren.

7.8.2. Samling

Databasen `maro_gn_metadata` har kun to tabeller som benyttes i prototypen:
`Bilde_sted(bilde_id, sted_id)`
`Sted(sted_id, stedsnavn, navnetype)`
Tabellen `Sted` er identisk med tabellen `Steder` i gazetteeren og kan gjøres overflødig med litt omskriving av kode i to av Javaklassene. Den har ingen betydning for gjenfinning av bildene, kun for å hente ut informasjon om steder.



Figur 7-11. Databaseskjema for bildemetadata.

7.8.3. Data i dtabasene

Det er lagt inn noe data i databasene for å ha mulighet til å teste systemet. Data er lagt inn manuelt, dette har vært mulig da testsamlingen er forholdsvis liten. Dette har foregått på to måter. Noe data er lagt inn direkte gjennom et phpMyAdmin webgrensesnitt mens en del data først ble listet i Excel regneark, konvertert til kommaseparerte filer og deretter importert til databasetabeller. Data som omtales videre i dette kapitlet er data som er begrenset til eksisterende metadata for bildene i testsamlinga.

7.8.4. Navnetypesaurus

ADL Feature Type Tesaurus og SOSI navnetyper fra Sentralt Stedsnavnregister danner grunnlaget for et forslag til en navnetypesaurus for dette systemet. Typiske navnetypetermer i eksisterende metadata for Galleri Nor og navnetyper som er tilordnet etter en subjektiv vurdering av motiv er registrert som foretrukne eller ikke foretrukne termer. Foretrukne termer er organisert i et navnetypehierarki. Disse er tilgjengelig for bruker i form av en hierarkisk indeks i brukergrensesnittets søkeside. En alfabetisk visning av ikke foretrukne termer med tilsvarende foretrukne termer er gjort tilgjengelig via en lenke fra søkesiden slik at det skal være mulig for bruker å finne fram til foretrukne termer. Alle navnetyper i gazetteeren skal være foretrukne termer.

7.8.5. Bildene i testsamlinga

Det er 38 bilder i testsamlinga. Antallet er ikke veldig viktig, men det må være stort nok til å kunne teste funksjonalitet og effekt av implementerte strukturer slik at en kan evaluere systemet i forhold til målet for oppgaven. Samtidig må testsamlinga være liten nok til at en faktisk kan måle testresultater.

Utvelgelsen av bilder ble gjort under en forholdsvis tidlig utforsking av eksisterende system for Galleri Nor. En forutsetning for å velge et bilde var at det fantes noe

informasjon om sted i de opprinnelige synlige metadata for dette bilde. Bak utvelgelsen ligger også en tanke om at eventuell inkonsistens i beskrivelsene for bildene i eksisterende Galleri Nor skal gjenspeiles i testsamlinga. Utover dette ligger ikke spesielle kriterier til grunn for det utvalget som er gjort. Bildene i testsamlinga ligger i katalogen *public_html* på *stud.ntnu.no/~maro/gn_system/gn_bilder*. I prototypen har bildene fått tilordnet en *bilde_id* som kan representeres av en intern id som for eksempel tilvekstnummer i et operasjonelt system.

7.8.6. Metadata om bildene

Metadata om bildene er lagret i databasen *maro_gn_metadata* og er begrenset til metadata som angår sted og navnetyper. I tabellen *bilde_sted* knyttes forbindelsen mellom bildene i samlinga og steder i gazetteeren. En *id* som representerer et bilde er i denne tabellen knyttet til *id* som representerer de stedsnavn som, ved søk i Galleri Nor, er synlige for bruker i feltet *sted* samt de steder som er representert i andre felt og som er registrert i gazetteeren. Figur 7-12 viser et eksempel på synlige metadata for ett bilde i eksisterende Galleri Nor og Tabell 7-1 viser hvilke metadata som er lagret for dette bilde i tabellen *bilde_sted*. I prototypen har dette bildet fått *id* 3. Metadata for dette bildet danner grunnlaget for utarbeidelsen av eksemplene under, som i Tabell 7-2 til Tabell 7-7 viser eksempler fra databasetabellene. I databasen *maro_gn_metadata* er det opprettet en tabell *sted* som er identisk med tabellen *steder* i gazetteeren (Tabell 7-2). Denne kan gjøres overflødig ved å legge til noen metoder i to av klassene. Den har ingen påvirkning på selve gjenfinningsresultatet, men brukes for å hente ut metadata om stedsnavn og navnetyper for bildene etter at systemet har identifisert hvilke bilder som skal returneres til bruker. Det er ikke opprettet noen tabell som holder lokasjonsinformasjon om selve bildene. Funksjonalitet for å lokalisere bildene er lagt til JSP-skriptet som genererer lenker til bildene før søkeresultatet presenteres for bruker. Denne løsningen er sannsynligvis ikke den beste så dette må vurderes ved eventuell utvikling av et operasjonelt system. Ingen nye metadata er direkte tilordnet bildene. Databasemodellen som brukes for Galleri Nor ved Nasjonalbiblioteket Rana har også en kobling mellom bilde og *sted_id*. Forskjellen ligger i at en *sted_id*, i prototypen er i overensstemmelse med *sted_id* for en forekomst i gazetteeren. Dette medfører at det, ved oppslag i gazetteeren, indirekte også tilordnes navnetyper. Disse vil være synlig for bruker som del av metadata for bildene ved presentasjon av søkeresultat.

Bilde_03

Eier: Norsk Folkemuseum

Datering: 1880 - 1890

Fotograf: Lindahl, Axel

Tittel:

Søkeord: jernbanestasjon, stasjonsbygning, båter, havn

Utfyllende opplysninger: Prot: Trondheim Jernbanestationen

Navn:

Sted:

Trondheim jernbanestasjon - (avbildet)

71 Trondheim 71 Sentrum Fosenkaia

Samling:

Tilvekstnr: NF.WL 01847

Internnr: NBR9312:02330

Figur 7-12. Synlige metadata for bilde_03 slik de framstilles for bruker i eksisterende Galleri Nor.

Bilde id	Sted id
3	14
3	86
3	83
3	111

Tabell 7-1. Eksempel fra tabell bilde_sted i databasen maro_gn_metadata

7.8.7. Stedsnavn i Gazetteeren

Et stedsnavn er et geografisk navn som representerer en geografisk forekomst. Alle geografiske navn som er registrert i feltet *sted* i opprinnelige beskrivelser for bildene i testsamlinga er tatt med unntatt de kommunenavn som ikke lenger er gyldige.

Geografisk navn som forekommer i andre felt i galleri Nor er tatt med der det tydelig går fram at dette er knyttet direkte til motivet og navnet ikke er tatt med i feltet *sted*. For eksempel bilde_03, der *Trondheim Jernbanestasjon* forekommer i feltet *Navn*, men ikke i feltet *sted*. Det er gjort oppslag i *Norgesglasset* for å finne gyldig kommunetilhørighet slik at alle kommunenavn som er representert i gazetteeren skal være gyldige. Navn på fylker og land er ikke synlige metadata i Galleri Nor, men *Norge* er tatt med som stedsnavn i gazetteeren og for fylker der disse er representert via andre stedsnavn i testsamlinga er navn på fylker også tatt med i gazetteeren.

Geografisk navn som forekommer i andre felt enn feltet *sted* er registrert i gazetteeren som alternativt navn der det tydelig går fram at dette er en variant av et navn som er registrert i feltet *sted*. For å kunne teste dette er enkelte navnevarianter lagt til i gazetteeren uten at de er registrert i eksisterende metadata for bildene. For eksempel er også *Nidaros* lagt til som navnevariant for *Trondheim by*.

Geografisk navn som er indirekte knyttet til motiv, for eksempel fødested til en som er portrettert, og som ikke er registrert i feltet *sted* er ikke tatt med.

Tabellene Tabell 7-2 og Tabell 7-3 viser sammenhengen mellom registrerte metadata for bilde_03 og data registrert i tabellene *steder* og *navne_varianter* i gazetteeren. Vi ser her at forekomstene med id 14, 86, 83 og 111 er forekomster i gazetteeren som representerer steder med aktuelle stedsnavn som er registrert i opprinnelige metadata for bilde_03. Disse forekomstene kan ha direkte eller indirekte relasjoner til forekomstene med id 75, 2 og 1 i gazetteeren. Relasjoner mellom geografiske forekomster omtales nærmere i kapittel 7.8.9.

Sted_id	Stedsnavn	Navnetype
14	Trondheim	Kommuner
86	Sentrum	Bebyggede områder
83	Fosenskaia	Havneområder
111	Trondheim Jernbanestasjon	Jernbanefasiliteter
75	Trondheim	Byer
2	Sør-Trøndelag	Fylker
1	Norge	Land

Tabell 7-2. Eksempel fra tabellen steder i databasen maro_GAZETTEER.

Tilsvarende eksempel fra tabellen sted i databasen maro_gn_metadata er identisk med eksempel i Tabell 7-2.

Navn	Sted id
Trondhjem	14
Trondhjem	75
Nidaros	75

Tabell 7-3. Eksempel fra tabellen navne_varianter fra databasen maro_GAZETTEER.

7.8.8. Navnetyper i Gazetteeren

I tilfeller der navnetypertermer er benyttet som søkeord og det kommer tydelig fram at disse har tilknytning til registrerte stedsnavn i Galleri Nor er disse tilordnet som navnetyper til aktuelle steder i gazetteeren. For stedsnavn der det ikke er registrert typiske navnetypertermer i eksisterende metadata for Galleri Nor er det gjort en subjektiv vurdering ut fra motivet på bildet for å tilordne navnetyper til steder. Tilordnede navnetyper er så vurdert i forhold til tidligere foreslåtte navnetypeskjema og har fått status som foretrukket eller ikke foretrukket term. For eksempel er *Trondheim* registrert i gazetteeren som stedsnavn for en forekomst med navnetype *kommuner* og en forekomst med navnetype *byer*. Tabell 7-4 viser et eksempel fra tabellen *typer*. Relasjoner mellom navnetyper omtales nærmere i kapittel 7.8.10.

Type id	Type navn
2	Alle typer
3	Administrative områder
4	Fylker
5	Kommuner
6	Land
7	Bebyggelse
8	Bebyggede områder
9	Byer
22	Jernbanefasiliteter
23	Havneområder
19	Samferdsel

Tabell 7-4. Eksempel fra tabellen typer i databasen maro_GAZETTEER.

7.8.9. Relasjoner mellom forekomster

En geografisk forekomst i gazetteeren har en relasjon til en annen forekomst. Relasjonene er av typen *er del av* og danner et hierarki av forekomstene. For eksempel er det slik at forekomsten *Trondheim Jernbanestasjon* er del av *Fosenkaia*, *Fosenkaia* er del av *Sentrum*, *Sentrum* er del av *Trondheim (by)*, *Trondheim (by)* er del av *Trondheim (kommune)*, *Trondheim (kommune)* er del av *Sør-Trøndelag* og *Sør-Trøndelag* er del av *Norge*. Tabellene Tabell 7-5 og Tabell 7-6 viser eksempler fra tabellene *relasjoner* og *relasjonstyper*. Tabellen *relasjonstyper* har kun ett innslag, men er opprettet slik for å vise at det skal være mulig å legge til andre typer relasjoner.

Sted_id	Relasjon_til_sted_id	Type_id
1	0	1
2	1	1
14	2	1
75	14	1
86	75	1
83	86	1
111	83	1

Tabell 7-5. Eksempel fra tabellen relasjoner i databasen maro_GAZETTEER.

Type_id	Type
1	Er del av

Tabell 7-6. Eksempel fra tabellen relasjonstyper i databasen maro_GAZETTEER.

7.8.10. Relasjoner mellom typer

På samme måte som relasjoner mellom geografiske forekomster skaper et hierarki skaper relasjoner mellom navnetyper et navnetypehierarki. En navnetype har en overordna navnetype som igjen kan ha en overordna navnetype. *Alle typer* (Type_id="2") er øverst i hierarkiet og er for eksempel overordna type til *Administrative områder* (Type_id="3") og til *Bebyggelse* (Type_id="7"). *Administrative områder* (Type_id="3") er overordna type til *Land* (Type_id="6"), *Fylker* (Type_id="4") og *Kommuner* (Type_id="5"). *Bebyggelse* (Type_id="7") er overordna type til *Bebyggede områder* (Type_id="8") som igjen er overordna type til *Byer* (Type_id="9"). *Bebyggelse* (Type_id="7") er også overordnet type til *Samferdsel* (type_id="19") som igjen er overordna type til *Jernbanefasiliteter* (Type_id="22") og *Havneområder* (Type_id="23").

Type_id	Overordna_type_id
2	0
3	2
4	3
5	3
6	3
7	2
8	7
9	8
22	19
23	19
19	7

Tabell 7-7. Eksempel fra tabellen Overordna_typer i databasen maro_GAZETTEER.

8. Testing og evaluering

8.1. Testing

Hva skal testes og evalueres

Testing av foreslått system er todelt. Prototypen testes og evalueres for å sikre at den fungerer og at den gjør det den er ment å gjøre i kapittel 8.1.1. Prototypen testes versus Galleri Nor i kapittel 8.1.2. Dette gjøres ved å utføre søk og sammenligne søkeresultater. Fokus er å undersøke om søkeresultatene påvirkes ved å integrere en gazetteer og en navnetypesaurus i et gjenfinningssystem og hva en eventuelt kan oppnå med tanke på recall og presisjon. Recall og presisjon vurderes ut fra tilgjengelige metadata om bildene i testsamlinga.

Hva skal ikke testes

Systemets brukervennlighet og ytelsesevne er aspekter som vil være viktige å teste for et operasjonelt system, men dette er aspekter som ikke er viktige for denne prototypen og som ikke er testet. Kvalitetsaspekter er heller ikke testet.

8.1.1. Testing av prototypen

Testing av kode

Systemet er modulbasert og metodene i Java-klassene er testet fortløpende underveis i arbeidet med prototypen, men det er ikke skrevet testlogg for testing av koden. Det ble skrevet en ekstra klasse med en mainmetode for å kunne kjøre programmet fra et kommandovindu underveis. Syntaksfeil som er oppdaget ved kompilering og kjøretidsfeil, er rettet opp fortløpende. For å sjekke at metodene gjør det de er ment å gjøre, er det lagt inn utskriftskommandoer fortløpende i koden. Disse er ikke fjernet fra kildekoden, men står som kommentarer. JSP- kode og HTML-kode er også testet fortløpende. All kode fungerer og det er skrevet Java-dokumentasjon til all Java-kode.

Systemtest og evaluering

Javaklassene ble lastet opp på JSP-serveren som nevnt i kapittel 7.7 for å kunne teste at systemet som helhet fungerer. Funksjonalitet for det ferdige systemet er testet som følger:

- 1) **Start søk uten data i inputform. Systemet returnerer en feilmelding som leder bruker tilbake til søkeform.**
- 2) **Start søk med spørring etter stedsnavn uten navnetype.**
 - a. Spørring etter *Trondheim* returnerer 5 bilder
 - b. Spørring etter *Oslo* returnerer 9 bilder
 - c. Spørring etter *Harstad* returnerer 2 bilder
 - d. Spørring etter *Troms* returnerer 4 bilder

Systemet returnerer alle aktuelle bilder fra forespurt sted eller sted som kan sies å være del av eller tilhøre forespurt sted uavhengig av navnetype.

3) Start søk med spørring etter navnevariant uten navnetype.

- Spørring etter *Trondhjem* returnerer de samme 5 bilder som spørring etter *Trondheim* i pkt. 2
- Spørring etter henholdsvis *Kristiania* og *Christiania* returnerer de samme 9 bilder som spørring etter *Oslo* i pkt 2

Systemet returnerer de samme bilder som spørring etter gyldig navn.

4) Start søk med spørring etter navnetype uten stedsnavn.

- Spørring etter *alle typer* returnerer alle 38 bilder i testsamlinga
- Spørring etter *kystdata* returnerer 8 bilder
- Spørring etter *fjorder* returnerer 6 bilder
- Spørring etter *øyer* returnerer 2 bilder
- Spørring etter *byer* returnerer 20 bilder
- Spørring etter *hovedsteder* returnerer 9 bilder
- Spørring etter *bygninger* returnerer 24 bilder

Systemet returnerer de samme bilder som spørring etter gyldig navn.

5) Start søk med spørring etter stedsnavn og navnetype.

- Spørring etter *Trondheim* og *byer* returnerer 4 bilder
- Spørring etter *Trondheim* og *bygninger* returnerer 4 bilder
- Spørring etter *Trondheim* og *religiøse fasiliteter* returnerer 3 bilder
- Spørring etter *Harstad* og *byer* returnerer 1 bilde
- Spørring etter *Troms* og *kystdata* returnerer 3 bilder
- Spørring etter *Trondheim* og *hovedsteder* returnerer ingen bilder

Systemet returnerer alle bilder som er registrert med sted som har aktuelle stedsnavn eller som kan sies å være del av sted som representeres av det stedsnavn det søkes via, og som er av aktuelle navnetype eller smalere type.

6) Start søk med stedsnavn som ikke finnes i gazetteeren.

- Spørring etter *Tromsø* med navnetype *byer*, deretter uten navnetype gir ingen resultat.

Systemet returnerer en melding til bruker og leder bruker tilbake til søkesiden.

Testing av systemet som helhet viser at prototypen har implementert ønsket funksjonalitet og at systemet oppfører seg som forventet.

8.1.2. Prototypen vs. Galleri Nor

Prototypen har implementert nok funksjonalitet til at en kan gjøre enkle søk for å sammenligne med Galleri Nor. Søkescenarier er formulert med tanke på å dekke implementert funksjonalitet i prototypen og er basert på å søke etter stedsnavn og

navnetyper/søkeord hver for seg og i kombinasjon. Søketermer som er benyttet er listet i Tabell 8-1.

For å klargjøre noen av problemområdene og gi et bilde av hvordan Galleri Nor fungerer i dag, har jeg først gjennomført søkescenariene i Galleri Nor sitt system som er tilgjengelig på www. Analyse av antall treff for de ulike søketermene kan være med på å konkretisere problemområdene. Deretter er scenariene gjennomført med testsamlinga som grunnlag for sammenligning mellom prototypen og Galleri Nor. Disse søkeresultatene er kommentert fortløpende.

Følgende søkescenarier er utført:

1. Jeg vil se alle bilder fra alle byer
2. Jeg vil se alle bilder av alle kirker
3. Jeg vil se alle bilder av alle bygninger
4. Jeg vil se alle bilder fra Trondheim
5. Jeg vil se alle bilder fra Trondhjem
6. Jeg vil se alle bilder fra byen Trondheim
7. Jeg vil se alle bilder av alle kirker i Trondheim
8. Jeg vil se alle bilder av alle bygninger i Trondheim
- 9.

Scenario:	Stedsnavn	Søkeord Galleri Nor	Navnetyper i prototypen
1		Byer / By	Byer
2		Kirker / Kirke	Kirker
3		Bygninger / Bygning	Bygninger
4	Trondheim		
5	Trondhjem		
6	Trondheim	Byer / By	Byer
7	Trondheim	Kirker / Kirke /	Religiøse fasiliteter
8	Trondheim	Bygninger / Bygning	Bygninger / Bygning

Tabell 8-1. Søketermer for testscenarier.

Søk i eksisterende Galleri Nor

Antall bilder i Galleri Nor er så stort at en vanskelig kan måle søkeresultater i forhold til recall og presisjon, men antall treff for de ulike søketermene kan allikevel si oss noe. Tabell 8-2 viser at entall og flertall brukes om hverandre, men at entall er hyppigere representert for disse søkeordene. Ved nærmere undersøkelse av de 11 ressursene som ble funnet ved søk etter *byer* viser det seg at dette slett ikke er bilder med motiv fra byer, men bilder der motivet er babyer (se Figur 8-1). Videre undersøkelser viser at disse ressursene også er en del av resultatlista ved søk etter *by* der også ressurser som er beskrevet med søkeord som *bygninger* og *bydeler* er representert. Faktorer som dette er med på å redusere presisjon ved søk i samlinga.

Videre viser tallene i denne tabellen at det ikke nødvendigvis er noen sammenheng mellom termene *kirke* og *bygning* eller *kirker* og *bygninger*. Det at antall treff er høyere ved søk etter *kirke* enn ved søk etter *bygning* tyder på at kirker ikke nødvendigvis er definert som bygninger. Faktorer som dette reduserer recall for søk i samlinga.

Navnetype / Søkeord	Antall treff
By	7120
Byer	11
Kirke	2577
Kirker	27
Bygning	1545
Bygninger	346

Tabell 8-2. Viser antall treff ved søk etter navnetyper / søkeord i eksisterende Galleri Nor

Tabell 8-3 viser at en ikke vil finne bilder fra Trondheim ved å søke etter stedsnavn som *Trondhjem* eller *Nidaros*. Dette er ikke overraskende da disse navnevariantene ikke er offisielle gyldige navn. Det er ikke uvanlig at en geografisk forekomst kan knyttes til flere forskjellige geografiske navn, det være seg både tidligere navn og lokale varianter av samme navn. *Trondhjem* er et navn som har lang tradisjon og som fortsatt er i bruk blant mange ”Trondhjemmere”, mens *Nidaros* er et navn som er sterkt knytta til Trondheims historie og som kan tenkes brukt i flere sammenhenger.

Det at et søk etter *Trondheim* gir vidt forskjellig resultat alt etter om en søker i feltet for *område* eller som *stedspresisering* sier egentlig ikke så mye uten å se konkret på de ressursene som finnes og de metadata som er tilgjengelige for disse. Dersom en søker etter *Trondheim* i kombinasjon med *by* vil en derimot få resultater som kan være noe vanskelige å forklare (Tabell 8-4). Ved å plassere stedsnavnet i feltet *område* får en 245 treff. Det er greit og kan oppfattes som at 245 bilder er fra Trondheim kommune og har motiv fra Trondheim by. Jeg vet at Trondheim kommune har kun en by og det vil være logisk for meg å tenke at plassering av stedsnavnet i feltet *stedspresisering* vil gi samme resultat som i forrige forsøk. Det gjør det ikke, søkeresultatet viser 73 treff. Dersom jeg ikke visste at Trondheim er en kommune som omfatter kun en by ville min logiske tanke være at her er det 73 bilder fra Trondheim by og de andre 172 bildene er fra andre byer som ligger innenfor et område som heter Trondheim.

Stedsnavn	Område	Stedspresisering
Trondheim	787	96
Trondhjem	0	0
Nidaros	0	0

Tabell 8-3. Viser antall treff ved søk etter stedsnavn.

Også ved å kombinere søk etter *Trondheim* med søkeordene *kirke* eller *bygning* vil valg av felt for plassering av stedsnavn som søketerm være avgjørende for hvilke bilder man finner. Ved å undersøke noen av bildene fra søkeresultatene nærmere viser det seg at det også her er inkonsistens i forhold til beskrivelser av bildene. Noen av bildene som har motiv fra Trondheim by er beskrevet med *Trondheim* både som *område* og som *stedspresisering*, men mange er beskrevet med *Trondheim* som *område* og mer detaljerte stedsnavn for *stedspresisering*. Dette kan være navn på bydel, gatenavn osv. Disse bildene finner jeg ikke dersom jeg søker etter *Trondheim* i feltet *stedspresisering*.

Stedsnavn og navnetyper / Søkeord	Område		Stedspresisering	
	Navnetype entall	Navnetype flertall	Navnetype entall	Navnetype flertall
Trondheim og by / byer	245	0	73	0
Trondheim og kirke / kirker	125	0	5	0
Trondheim og bygning / bygninger	51	0	8	1

Tabell 8-4. Viser antall treff ved søk etter stedsnavn og navnetype / søkeord.

The screenshot shows the Galleri Nor website interface. At the top, there is a navigation menu with links for 'forsiden', 'info', 'søk i arkivet', 'fotografer', and 'ei'. The main content area displays a grid of search results for the term 'byer'. Each result consists of a small thumbnail image and a text block containing the following information:

- Fotograf:** Wilse, Anders Beer (for most results) or Kvantoland, Lyder (for the last result).
- Sted:** Sørfold, Røsvik.
- Datering:** 1927 - 1928, 1944, or 1958.
- Eier:** Norsk Folkemuseum or Sørfold Lokalhistorielag.
- Tilvekstnr.:** NF.WB 21638* B, NF.WB 21639, NF.WB 21640, NF.WB 47752* A, NF.WB 47753* A, NF.WB 47754* A, NF.WB 47755* A, NF.WB 47756* A, or SLHL 1958/017.

The page also includes a search bar at the top, a 'Format' dropdown menu set to 'Frimerkebilder', and a 'Viser bilde 1 - 11 av 11' indicator.

Figur 8-1. Søkeresultat fra Galleri Nor ved søk via termen *byer*.

Søkeresultater ved søk i prototypen vs. Galleri Nor

Søkescenariene er her utført med fotografiene i testsamlinga som registrerte søkbare ressurser. Søk i Galleri Nor innebærer her at det er gjort manuelle søk i en tekstfil med alle synlige metadata for fotografiene i testsamlinga. Grunnlag for sammenligning vil være funn av stedsnavnene *Trondheim* eller *Trondhjem* i feltene for *sted* samt funn av *termene by, byer, kirke, kirker, bygning* eller *bygninger* i feltet *søkeord*. Her vil et registrert søkeord som *bygning* også gi treff når det søkes etter *by*. Bilder og metadata finnes som vedlegg D.

For hvert scenarie har jeg valgt å utføre manuelt søk i Galleri Nor først for deretter å utføre søk ved hjelp av prototypen. Oversikt over testsøk og resultater er gjengitt i Tabell 8-5 – Tabell 8-12.

Scenario 1

Søk etter alle bilder fra alle byer	
Søk i Galleri Nor	
Søkeord: <i>Byer</i>	Antall funnet: 0
Søkeord: <i>By</i>	Antall funnet: 17
Funnet bilde_id: 2, 3, 4, 5, 6, 7, 8, 9, 14, 15, 19, 22, 26, 27, 28, 41, 42	
Søk i prototyp	
Navnetype: <i>Byer</i>	Antall funnet: 20
Funnet bilde_id: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 16, 22, 24, 30, 36, 38, 41, 42, 44	

Tabell 8-5. Testresultat scenario 1.

Kommentar til scenario 1:

Her ser vi at søkeordet *by* gir 17 treff i Galleri Nor, mens navnetype *byer* gir 20 treff i prototypen. Den største forskjellen her ligger i hvilke bilder som finnes og ikke finnes. 6 av bildene som er funnet i Galleri Nor har ikke motiv fra byer, men er beskrevet med søkeord som for eksempel *bygning*. 10 av bildene i testsamlinga som har motiv fra byer er ikke funnet i Galleri Nor. Ved hjelp av prototypen finnes alle, og kun de, bilder som har motiv fra byer.

Scenario 2

Søk etter alle bilder fra alle kirker	
Søk i Galleri Nor	
Søkeord: <i>kirker</i>	Antall funnet: 0
Søkeord: <i>kirke</i>	Antall funnet: 13
Funnet bilde_id: 11, 12, 13, 14, 15, 16, 17, 18, 19, 23, 41, 42, 44	
Søk i prototyp	
Navnetype: <i>religiøse fasiliteter</i>	Antall funnet: 12
Funnet bilde_id: 11, 12, 13, 14, 15, 16, 17, 18, 19, 41, 42, 44	

Tabell 8-6. Testresultat scenario 2.

Kommentar til scenario 2:

Her ser vi at søk i Galleri Nor og prototypen gir tilnærmet samme resultat. Ett bilde finnes i Galleri Nor som ikke finnes ved hjelp av prototypen, bilde_id 23. Dette har motiv fra flere bygninger der man også kan se en kirke. Denne er ikke registrert i gazetteeren da det er en ikke navngitt kirke i Galleri Nor.

Scenario 3

Søk etter alle bilder fra alle bygninger	
Søk i Galleri Nor	
Søkeord: <i>bygninger</i>	Antall funnet: 0
Søkeord: <i>bygning</i>	Antall funnet: 7
Funnet bilde_id: 3, 4, 5, 6, 7, 8, 9	
Søk i prototyp	
Navnetype: <i>bygninger</i>	Antall funnet: 24
Funnet bilde_id: 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 24, 41, 42, 43, 44	

Tabell 8-7. Testresultat scenario 3.

Kommentar til scenario 3:

Her er det stor forskjell på trefflista fra Galleri Nor og ved hjelp av prototypen. Av de bildene som finnes i Galleri Nor er det ett som ikke finnes ved hjelp av prototypen, bilde_id 3. Dette er registrert med Trondheim Jernbanestasjon og viser absolutt en flott bygning. Her er det tilordning av navnetype og overordna navnetype i prototypen som gjør at dette bildet faller utenfor. Til gjengjeld finner en ved hjelp av prototypen 18 bilder som en ikke finner ved søk i Galleri Nor.

Scenario 4

Søk etter alle bilder fra Trondheim	
Søk i Galleri Nor	
Område: <i>Trondheim</i>	Antall funnet: 5
Funnet bilde_id: 3, 41, 42, 43, 44	
Stedspresisering: <i>Trondheim</i>	Antall funnet: 1
Funnet bilde_id: 41	
Søk i prototyp	
Stedsnavn: <i>Trondheim</i>	Antall funnet: 5
Funnet bilde_id: 3, 41, 42, 43, 44	

Tabell 8-8. Testresultat scenario 4.

Kommentar til scenario 4:

Her ser vi at vi finner alle bildene fra Trondheim i begge systemene. Forskjellen er at bruker i Galleri Nor må vite hvilket felt søketermen skal plasseres i, eller en må prøve seg fram med flere forespørsler til systemet, mens det i prototypen holder med en forespørsel for å finne alle bildene.

Scenario 5

Søk etter se alle bilder fra Trondhjem	
Søk i Galleri Nor	
Område: Trondhjem	Antall funnet: 0
Stedspresisering: Trondhjem	Antall funnet: 0
Søk i prototyp	
Stedsnavn: Trondhjem	Antall funnet: 5
Funnet bilde id: 3, 41, 42, 43, 44	

Tabell 8-9. Testresultat scenario 5.

Kommentar til scenario 5:

Ved hjelp av prototypen finnes de samme bildene som i scenario 5. Søk i Galleri Nor gir ingen treff.

Scenario 6

Søk etter alle bilder fra byen Trondheim	
Søk i Galleri Nor	
Område: <i>Trondheim</i> Søkeord: <i>by</i>	Antall funnet: 3
Funnet bilde id: 3, 41, 42	
Stedspresisering: <i>Trondheim</i> Søkeord: <i>by</i>	Antall funnet: 1
Funnet bilde id: 41	
Område: <i>Trondheim</i> Søkeord: <i>byer</i>	Antall funnet: 0
Stedspresisering: <i>Trondheim</i> Søkeord: <i>byer</i>	Antall funnet: 0
Søk i prototyp	
Stedsnavn: <i>Trondheim</i> Navnetype: <i>byer</i>	Antall funnet: 4
Funnet bilde id: 3, 41, 42, 44	

Tabell 8-10. Testresultat scenario 6.

Kommentar til scenario 6:

Søk i Galleri Nor finner 3 av 4 aktuelle bilder mens søk ved hjelp av prototypen finner de 4 aktuelle bildene.

Scenario 7

Søk etter alle bilder fra alle kirker i Trondheim	
Søk i Galleri Nor	
Område: <i>Trondheim</i> Søkeord: <i>kirke</i>	Antall funnet: 3
Funnet bilde id: 41, 42, 44	
Område: <i>Trondheim</i> Søkeord: <i>kirker</i>	Antall funnet: 0
Stedspresisering: <i>Trondheim</i> Søkeord: <i>kirke</i>	Antall funnet: 1
Funnet bilde id: 41	
Stedspresisering: <i>Trondheim</i> Søkeord: <i>kirker</i>	Antall funnet: 0
Søk i prototyp	
Stedsnavn <i>Trondheim</i> Navnetype: <i>religiøse fasiliteter</i>	Antall funnet: 3
Funnet bilde id: 41, 42, 44	

Tabell 8-11. Testresultat scenario 7.

Kommentar til scenario 7:

Resultatet er det samme for de to systemene. Ved søk i Galleri Nor må en vite hvor søketermene skal plasseres eller så må man gjøre flere forespørsler. Ved hjelp av prototypen finnes de aktuelle bilder ved første forespørsel.

Scenario 8

Søk etter alle bilder fra alle bygninger i Trondheim	
Søk i Galleri Nor	
Område: <i>Trondheim</i> Søkeord: <i>bygning</i>	Antall funnet: 1
Funnet bilde id: 3	
Område: <i>Trondheim</i> Søkeord: <i>bygninger</i>	Antall funnet: 0
Stedspresisering: <i>Trondheim</i> Søkeord: <i>bygning</i>	Antall funnet: 0
Stedspresisering: <i>Trondheim</i> Søkeord: <i>bygninger</i>	Antall funnet: 0
Søk i prototyp	
Stedsnavn: <i>Trondheim</i> Navnetype: <i>bygninger</i>	Antall funnet: 4
Funnet bilde id: 41, 42, 43, 44	

Tabell 8-12. Testresultat scenario 8

Kommentar til scenario 8:

Her ser vi at vi ved søk i Galleri Nor finner ett bilde, bilde_id 3. Søk ved hjelp av prototypen finner 4 bilder, men ikke bilde_id 3. Årsaken til det er den samme som i scenario 3.

Oppsummering av testresultater for prototypen versus Galleri Nor

- I de tilfeller det er noen forskjell på søkeresultatet for de to systemene er det to tilfeller der et relevant bilde finnes i Galleri Nor som ikke finnes ved hjelp av prototypen. Bildet er det samme i begge tilfellene og årsaken har sammenheng med tilordning av navnetyper og overordna typer. Bortsett fra dette viser testresultatene at recall er høyere for prototypen i fire av scenariene, mens ett av disse (scenario 1) også viste adskillig høere presisjon.
- Ved søk via termene *by / byer* viser resultatet fra Galleri Nor at systemet finner fram til 50% av ressursene med motiv fra byer mens 35% av ressursene som er funnet har ikke motiv fra byer, men er beskrevet med termer som for eksempel *bygninger, bydel*
- Ved hjelp av prototypen finnes alle og kun de ressurser med motiv fra byer.
- Ved søk via termene *kirke / kirker / religiøse fasiliteter* viser resultatet at navnetypetermer brukes som søkeord i Galleri Nor uten at de nødvendigvis er satt i noen sammenheng med registrert stedsnavn.
- Ved søk via termene *bygning / bygninger* ser en klare forskjeller på de to systemene. For Galleri Nor sin del ser vi at manglende relasjoner mellom termer som for eksempel *kirke* og *bygning* gir lavere recall.
- Resultatet fra søk ved hjelp av prototypen viser at tilordning av navnetyper til steder og tilordning av overordna typer i navnetypehierarki er vesentlig for resultatet og vil gi lavere recall dersom dette ikke er tilstrekkelig gjennomtenkt.
- Ved søk via stedsnavn må bruker forholde seg til to separate søkefelt i Galleri Nor og må da enten vite mye om hvordan ressursene er registrert eller gjøre søk i begge feltene etter tur for å finne alle ressursene
- Ved søk via navnevarianter ser vi at dette ikke gir treff på ressurser i Galleri Nor mens en ved hjelp av prototypen får treff på de samme ressursene som ved søk via gyldige stedsnavn.
- Søk via *stedsnavn Trondheim* og *navnetype by / byer* viser at en ved søk i Galleri Nor ikke har mulighet for å finne alle ressursene med motiv fra Trondheim by selv om en benytter begge feltene for søking via stedsnavn. Mitt system finner 3 ressurser. Disse er alle registrert med *Trondheim* som *område*, mens en er registrert med *Trondheim* også som *stedspresisering*. Ingen av ressursene er registrert med *by* eller *byer* som søkeord, men gjenfinnes her som et resultat av tilfeldighet da de er beskrevet med søkeord som *bydel* eller *bygning*. Ved hjelp av prototypen gjenfinnes alle bildene fra Trondheim by
- Ved søk via *kirke, kirker / religiøse fasiliteter* i *Trondheim* er resultatet det samme for de to systemene, men ved søk i Galleri Nor må en igjen

enten vite hvilket felt en skal bruke for å søke via stedsnavn eller så må man søke i begge feltene etter tur

- Ved søk via termene *Trondheim* og *bygninger* er resultatene sammenfallende med resultatene ved søk via *navnetype bygning*. Ressurser som er registrert med søkeord som er spesifiserte varianter av bygninger forblir uoppdaget ved søk i Galleri Nor. Ved hjelp av prototypen gjenfinnes de ressurser som er registrert med id for et sted som er av type *bygning* eller som er av en type som er definert som en smalere type av type *bygning*

8.2. Diskusjon og evaluering

8.2.1. Testresultater

Selv om testsamlingen er liten og prototypen forholdsvis enkel viser testresultatene at det på enkelte områder er klare forskjeller mellom mitt system og Galleri Nor. Dette er ikke veldig overraskende. Det er heller ingen grunn til å anta at resultatene fra Galleri Nor er enestående da manglende vokabularkontroll og inkonsistens i indeksering er identifiserte problemområder for mange typer systemer. Sammenligningen mellom de to systemene viser et reelt forbedringspotensiale for et system som Galleri Nor. Integrert i systemer kan strukturer som gazetteers og tesauri støtte gjenfinningen på en måte som øker både recall og presisjon. Automatisk oppslag i gazetteer og tesaurus genererer utvidelse av query, og kan til en viss grad kompensere for inkonsistens i indeksering av metadata .

Bruk av hierarkisk indeks i systemets søkebilde i kombinasjon med en tilgjengelig navnetypesaurus, kontrollerer at brukers søketermer er i samsvar med gazetteerens navnetypeskjema.

Forskjeller i sammenligningsgrunnlaget kan diskuteres da Galleri Nor ikke gjør systematisk bruk av navnetyper slik at det her er termer som er registrert som frie søkeord som danner grunnlaget for sammenligning med prototypens navnetyper. Prototypen har ikke implementert funksjonalitet for å søke via frie emneord og en kan ikke se bort fra at noen av søkeresultatene kunne ha vært annerledes dersom dette hadde vært tilfelle. Det er allikevel ikke grunn til å anta at dette ville ha påvirket de funn som er gjort i forhold til å identifisere problemområder i eksisterende Galleri Nor.

8.2.2. Prototypen som helhet

Som helhetlig system kan en si at prototypen fungerer for sitt formål, men man lærer så lenge man holder det gående og jeg ser at det er ting som kunne vært gjort annerledes. Testsamlinga fungerer for sitt formål, men med en så liten testsamling kunne en muligens ha tydeliggjort resultatene bedre dersom utvelgelsen av fotografier hadde vært konsentrert om et noe mindre geografisk område. Når et gjelder søkefunksjonalitet og navnetyper må en vurdere hvilke typer som er hensiktsmessig å gjøre tilgjengelige for søk. Nytteverdien av å kunne søke på navnetyper som kommuner eller fylker kan diskuteres. Dette er detaljer som jeg registrerer at man må ta stilling til for et operasjonelt system, men som jeg ikke har vært opptatt av for prototypen.

For denne prototypen er tjenestene fra gazetteeren implementert i en klasse i samme pakke som resten av systemet. For et operasjonelt system vil dette neppe være noen god løsning. En bedre løsning ville være å implementere gazetteeren som en selvstendig enhet med et tydelig grensesnitt mot andre komponenter. Aktuelle bruksområder for gazetteeren vil være av gjørende for hvilke tjenester den skal tilby slik at en tydelig definisjon av tjenestetilbud vil være en fordel.

Valgt teknologi for implementasjon har fungert godt. Java som programmeringsspråk med objektorientert tilnærming har gjort det mulig å dele opp koden i små moduler. Ved gjenbruk av en del åpen kildekode har jeg spart noe arbeid. JSP fungerer greit som grensesnitt mot bruker. Jeg vurderte muligheten for å bruke servlets, men fant ut at det sannsynligvis ikke ville gjøre noen annen forskjell enn at det ville kreve atskillig mer arbeid.

Databaseløsningen har også fungert godt for sitt formål. Her er det greit å være klar over at implementerte databaser har kun de tabeller som er tilstrekkelig for å kunne teste begrenset søkefunksjonalitet for systemet i denne oppgaven. Utvidet søkefunksjonalitet vil kreve at en også utvider databasene. Løsningen med å ha identiske tabeller med data om steder i de to databasene er ikke den beste, men det fungerer. For et operasjonelt system vil jeg foreslå å løse dette på en annen måte.

8.2.3. Gazetteeren

Gazetteeren som er implementert i prototypen er noe forenklet i forhold til den som er foreslått i kapittel 6.3.5. Blant annet er lokasjonsdata i form av geografiske avtrykk ikke implementert i prototypen. Gazetteeren er basert på en metadatalignende modell med eksplisitt uttrykte relasjoner mellom forekomster. Relasjoner mellom geografiske forekomster er utnyttet for å skape hierarkiske strukturer som danner grunnlag for automatisk queryutvidelse. Dette er en tilfredsstillende løsning for å vise muligheter i denne oppgaven, men om dette kan være en god løsning for et operasjonelt system må vurderes. Hvorvidt en tilsvarende løsning basert på geografiske avtrykk vil være et bedre alternativ vil avhenge av flere faktorer. Viktig her vil være hva slags type avtrykk som brukes. Velger en å bruke avgrensede bokser kan en anta at yttergrensene for en boks, og dermed det geografiske omfanget for et søk, vil gå utover den forespurte geografiske forekomst. Dette vil i sin tur kunne føre til svakere presisjon. Det er ikke grunn til å anta at det vil være særlige forskjeller når det gjelder recall.

En databasetabell som tabellen *relasjonstype* med bare ett innslag kan virke litt lite smart, men den er tatt med for å vise muligheten for å legge til flere relasjonstyper. Dette bør løses på en annen måte foret operasjonelt system dersom en skal ha med kun en type relasjoner.

Registrerte stedsnavn i opprinnelige metadata for bildene i testsamlinga danner grunnlaget for forekomstene i gazetteeren. Forekomster som representerer kommuner og fylker er lagt til etter behov. Alle forekomster i gazetteeren er tilordnet navnetyper i henhold til foreslåtte navnetypeskjema. Kjerneelementer for en forekomst i gazetteeren er

- Geografisk navn
- Navnetype i henhold til foreslåtte navnetypeskjema
- Lokasjonsinformasjon i form av en relasjon av typen *ErDelAv* til en annen forekomst.

8.2.4. Navnetypesaurus

Forslaget til navnetypeskjema og navnetypesaurus er forenklet i denne oppgaven og omfatter kun de navnetyper som er representert i testsamlinga. Dette er allikevel tilstrekkelig til å teste prinsippene i denne oppgaven. Et komplett navnetypeskjema for et operasjonelt system vil kreve en revurdering av de beslutninger som er tatt i denne oppgaven. En komplett tesaurus for navnetyper vil også kreve atskillig mer arbeid enn den oppgaven gir rom for.

- Valgte termer til navnetypeskjema er i tråd med retningslinjer for valg av tesaurustermer
- Navnetyper er valgt i henhold til SOSI navnetyper og ADL Feature Type Tesaurus
- Navnetypesaurus har to viktige roller i denne oppgaven
 - Lede brukeren til å finne en gyldig søketerm samtidig som
 - Danne en hierarkisk struktur som er grunnlag for queryutvidelse.

8.2.5. Hierarkisk indeks

Ved søk via navnetyper må bruker velge navnetypeterm fra en hierarkisk indeks i brukergrensesnittet. Dette kontrollerer vokabularet mellom bruker og system og styrker brukers reelle mulighet til å finne aktuelle ressurser. I prototypens grensesnitt er alle navnetyper tilgjengelige, men nytten av å kunne søke på enkelte av disse i et system som dette kan diskuteres. Hvorvidt en slik løsning kan være hensiktsmessig for andre systemer må vurderes ut hvilket materiale man skal søke i og hvilken informasjon man ønsker å presentere for bruker.

8.2.6. Automatisk queryutvidelse

Relasjoner mellom forekomster i gazetteeren samt relasjoner mellom navnetyper genererer hierarkiske strukturer som utnyttes av systemet for å automatisk utvide spørringer. En spørring fra bruker etter et stedsnavn vil, når minst en forekomst er funnet, automatisk generere spørringer etter forekomster som kan sies å tilhøre eller være del av aktuelle forekomst. Testresultatene viser at en slik løsning kan veie opp for noe av den inkonsistens som forekommer i registrering av metadata for ressursene. En bruker som ønsker å se bilder fra Trondheim by kan bruke *Trondheim* og *by* som søketermer og få presentert bilder som ikke registrert med verken *Trondheim* som *stedspresisering* eller *by* som *søkeord* i opprinnelige metadata fra Galleri Nor. Registrerte metadata kan f. eks. være *Ila* og *bydel* eller *Nidarosdomen* og *domkirke*.

8.2.7. Dublin Core for bildebeskrivelser

Dublin Core er foreslått som metadataformat for å muliggjøre interoperabilitet med andre systemer og samlinger og det er foreslått en mapping av Galleri Nor elementer til Dublin Core. Forslaget er begrenset til å omfatte informative metadata med hovedvekt på de metadata som er synlige og søkbare for bruker. Bakgrunnen for å velge Dublin Core inkluderer tanken om at det ikke er vesentlige forskjeller på metadata-beskrivelser for fotografier og tekstdokumenter for denne type informative metadata. Av diskusjonen i kapittel 6.3.9 (se også vedlegg A.i) kan en konkludere med at Dublin Core kan brukes til et slikt formål, men ikke aldeles uproblematisk. Et eksempel på problemer man må ta stilling til kan være hvordan man skal bruke elementet *skaper* (*creator*).

Blant fotografiene i Galleri Nor forekommer eksempler på at fotografiets motiv tydelig søker å gjenspeile motivet i et todimensjonalt objekt i form av et maleri eller en tegning, og der metadata som datering og stedsbestemmelse er tilordnet på bakgrunn av det todimensjonale objektet. Hvorvidt skaperen da er fotografen eller den som malte eller tegnet motivet er forhold man må ta stilling til.

Anvendelse av DCMI's anbefalte termer for raffinering av kjerneelementer kan i mange tilfeller være en god måte å spesifisere det semantiske innholdet i elementet på, men eksemplet over viser at disse ikke alltid er tilfredsstillende. DCMI anbefaler ingen raffinering av elementene *skaper (creator)* eller *bidragsyter (contributor)*. NKKM's feltkatalogen tilbyr muligheter for å knytte roller blant annet til personer. I denne oppgaven antydes muligheten for å benytte rolle som raffinering til elementet *bidragsyter (creator)*.

Det er viktig å poengtere at dette forslaget kun omfatter informative metadata. Et system som Galleri Nor vil også ha behov for andre typer metadata for å ivareta informasjon om et objekt gjennom hele dets levetid, metadata som Dublin Core vanskelig kan dekke.

- Dublin Core kan brukes som utvekslingsformat for metadata fra Galleri Nor med vekt på de metadata som er synlige og søkbare for bruker
- Enkelte elementer kan i visse tilfeller kreve nærmere spesifisering av semantisk innhold.
- Dublin Core har liten støtte for å knytte attributter til personer i elementene *creator* og *contributor*
- Et system som Galleri Nor vil ha behov for andre typer metadata som Dublin Core ikke kan dekke.

8.2.8. Oppgaven som helhet

- Beskrivende metadata er kritisk for å skape gode forhold for konseptbasert bildegjenfinning, men beskrivende metadata er ikke nok. Vokabularkontroll er et nøkkelord for å muliggjøre interoperabilitet både internt i systemer og mellom ulike systemer.
- Galleri Nor er brukt som eksempel for å konkretisere problemområder og mulige løsninger. Det er gjort et betydelig arbeid fra Nasjonalbiblioteket Rana for å gjøre en stor fotosamling tilgjengelig for allmennheten, men til tross for at visse prinsipper ligger til grunn for tilordning av indekstermer viser det seg å være inkonsistens i disse.
- Opprinnelige beskrivelser i Galleri Nor er gjort med tanke på at geografiske begreper skal være søkbare, men flere forhold vedrørende disse skaper problemer som påvirker gjenfinningen:
 - varierende detaljeringsgrad i beskrivelsene
 - manglende samsvar mellom stedsnavn og søkeord
 - manglende relasjoner mellom relaterte termer
 - bruk av både entall og flertalls form av tellbare objekter
 - oppdeling av søkefelt for sted i brukergrensesnittet
- Dublin Core er vurdert som metadataformat for beskrivende metadata for bilder og det er lagt fram forslag til ressursmetadata og samlingsmetadata for Galleri Nor basert på Dublin Core.

- Jeg har foreslått en gazetteer til Galleri Nor samt et søkesystem som integrerer og bruker denne. Forslaget til gazetteer inkluderer et forslag til en navnetypesaurus.
- For at foreslått gazetteer skal kunne integreres i et system med Galleri Nor kreves det at en sted_id som tilordnes en ressurs samsvarer med sted_id for en forekomst i gazetteeren. Dette er det eneste metadatakravet i forhold til ressursene i Galleri Nor.
- En prototyp er implementert for å vise noen av mulighetene et integrert system kan gi. Prototypen er en forenkling av foreslått system og den har begrenset søkefunksjonalitet, men den fungerer tilfredsstillende for å teste prinsippene som ligger til grunn for mitt forslag.
- Testresultater for prototypen viser at strukturer som gazetteers og tesauri har egenskaper som kan utnyttes til å påvirke søkeresultater dersom de integreres og brukes i et gjenfinningssystem.
- Strukturer som gazetteers og tesauri kan tilby tjenester til et gjenfinningssystem slik at dette kan produsere resultater som tilsynelatende bærer preg av at ressursene er beskrevet enhetlig og strukturert selv om dette ikke er tilfelle.
- For prototypen i denne oppgaven har det vært nyttig å registrere bygninger som sted i gazetteeren.
- Valgte teknologier og tekniske løsninger for prototypen har fungert tilfredsstillende. Det har ikke oppstått nevneverdige tekniske problemer underveis.

Ved å se tilbake til introduksjonsdelen kan en oppsummere at oppgavens mål og delmål er oppfylt og det er oppnådd et resultat som kan knyttes til problemstillingen. Jeg har gjort det som jeg var bestemt på å gjøre, men jeg ser at jeg sannsynligvis ville ha valgt en del andre løsninger dersom jeg skulle begynt på et slikt arbeid nå. Noe av dette kommer fram under forslag til videre arbeid.

8.2.9. Veien videre

Det at oppgaven berører store og komplekse tema innenfor ulike retninger av fagfeltet har til tider gjort det vanskelig å sette grenser. Det har heller ikke alltid vært enkelt å velge ståsted og stake ut kursen. Nye problemstillinger og muligheter har fristet underveis. Noen av dem danner grunnlag for forslag til videre arbeid.

Det kan være interessant å videreutvikle en prototyp basert på de samme prinsippene som kan brukes på større og eventuelt flere samlinger, også flere typer informasjonsobjekter. Det vil ikke nødvendigvis kreve mye arbeid å utvide funksjonalitet til å kunne søke på andre elementer i tillegg til de som er relatert til geografisk informasjon, men jeg vil allikevel anbefale å redesigne hele systemet. Systemet i sin helhet kan med fordel deles opp i flere moduler, både større og mindre sådanne.

Gazetteer og tesaurus bør få mer oppmerksomhet på et mer overordna nivå. Vi bør definere klare krav til disse komponentene i forhold til hvilken rolle de skal ha i samspill med hverandre og med andre komponenter. Grensesnittene mellom de ulike komponentene må gjøres tydeligere. Det vil være nødvendig å utvikle et mer fullstendig navnetypeskjema og navnetypesaurus må få et større ordforråd når det gjelder termer som ikke er foretrukne. De beslutninger som er tatt for prototypen må revurderes. Blant annet kan en vurdere å gi en navnetype muligheten til å ha flere overordna typer.

Bruk av navnetypesaurus viser seg å være nyttig i prototypen. Det kan være aktuelt å ta i bruk tesauri også innenfor andre områder for bruk i tilsvarende systemer. For eksempel kan en ta utgangspunkt i de foreslåtte kategorier for bildemotiv og ta i bruk tesauri for noen av de ulike kategoriene som for eksempel aktiviteter, hendelser/begivenheter.

Ved å oppsøke ulike nettbaserte systemer og samlinger via www kan en få inntrykk av at geografisk relatert informasjon ofte er representert på en eller annen måte. Det kunne være interessant å vite noe mer om hvordan denne type informasjon er representert i ulike systemer og samlinger, både hvordan informasjonen er organisert internt i systemer og hvordan den presenteres for brukere. I den grad geografisk relatert informasjon er søkbar i form av metadata kan det også være interessant å vite noe om hvor stor del av de søkbare metadata denne type informasjon utgjør.

I kapittel 3.4 kommer det fram at menneskers søkeadferd er lite dokumentert når det gjelder å søke etter bilder spesielt. Dette er et stort område med mange mulige problemstillinger, men jeg kunne godt tenke meg å vite noe om hvilken rolle geografisk relatert informasjon spiller i denne sammenhengen. Kjerneelementer for en gazetteer er geografiske navn, lokasjoner og navnetyper. Dette er elementer som det er grunn til å anta at de fleste mennesker, om ikke alle, har et mer eller mindre bevisst forhold til og dermed en forutsetning for å kunne bruke i søkesammenheng. En adferdsstudie med dette i tankene kunne være interessant.

Det kan være interessant å sammenligne resultater i system basert på avgrensede bokser i forhold til å utnytte relasjoner til å skape hierarkiske strukturer. Det vil være mulig å utvide gazetteeren som er implementert i prototypen til å inkludere lokasjonsinformasjon i form av koordinater. Dette vil ikke kreve mye arbeid med databasen, men vil kreve at en skriver om eller legger til noen metoder i noen av klassene.

9. Referanseliste

1. Michael Lesk. (1997) *Going Digital*. Scientific American, March 1997. ISSN 0036-8733
2. James Reid. *geoXwalk - A Gazetteer Server and Service for UK Academia*. Lecture Notes in Computer Science Vol.2769, Springer Verlag, 2003. ISBN 3-540-40726-X
3. Hill, L. & Zheng, Q. (1999) *Indirect geospatial referencing through place names in the digital library: Alexandria Digital Library experience with developing and implementing gazetteers*. Proceedings of the American Society for Information Science Annual Meeting, Washington, D.C., Oct. 31 – Nov. 4, 1999, pp. 57 – 69. www.alexandria.ucsb.edu/~lhill/paper_drafts/ASIS99_confpaper2_final.pdf (sist sett 24062004)
4. www.alexandria.ucsb.edu (sist sett 24062004)
5. International Federation of Library Associations and Institutions <http://www.ifla.org/II/metadata.htm#general-indices> (sist sett 24062004)
6. Loran Dempsey, Rachel Heery. *DESIRE - RE1004. A review of metadata: A survey of current resource description formats*. http://www.ukoln.ac.uk/metadata/desire/overview/rev_toc.htm (sist sett 24062004)
7. Library of Congress, MARC Standards <http://www.lcweb.loc.gov/marc/> (sist sett 24062004)
8. Library of Congress, Cataloging Distribution Service <http://www.lcweb.loc.gov/cds/lcsh.html> (sist sett 24062004)
9. Anne J. Gilliland-Swetland. *Defining Metadata*. <http://www.getty.edu/research/institute/standards/intrometadata/pdf/swetland.pdf> (sist sett 24062004)
10. Dublin Core Metadata Initiative (DCMI) www.dublincore.org (sist sett 24062004)
11. NORMARC <http://www.nb.no/katkom/NORMARC/normarc.html> (sist sett 24062004)

12. Dublin Core Metadata Initiative (DCMI), DCMI Metadata Terms
<http://dublincore.org/documents/dcmi-terms/>
(sist sett 24062004)
13. Stuart Weibel, Erik Miller. (1997) *Image description on the Internet*. A summary of the 1996 CNI/OCLC Image Metadata Workshop. D-lib Magazine, January 1997. ISSN 1082-9873
<http://www.dlib.org/dlib/january97/oclc/01weibel.html>
(sist sett 24062004)
14. Metadata Resources, MPEG-7 Multimedia Content Description Interface
<http://www.ukoln.ac.uk/metadata/resources/mpeg-7/>
(sist sett 24062004)
15. Visual Resources Association Data Standards Committee, VRA Core Categories, Version 3.0
<http://www.vraweb.org/vracore3.htm>
(sist sett 24062004)
16. Categories for the Description of Works of arts
http://www.getty.edu/research/institute/standards/cdwa/2_overview/index.html
(sist sett 24062004)
17. Outline klassifikasjonssystem for museum og arkiv, norsk oversettelse med utarbeidet findeling, Norsk Museumsutvikling 4:2001, ISBN 82- 90935-94-3, ISSN 1501-0309
<http://www.abm-utvikling.no/publisert/fulltekst/nmu4-2001/nmu4-2001.pdf>
(sist sett 24062004)
18. Gary Marchionini. (1995) *Information Seeking in Electronic Environments*. Cambridge. Unic.press. 1995
http://www.ils.unc.edu/~march/isee_book/web_page.html
(sist sett 24062004)
19. R.Storleer. Søkeprosedyre: Presisjon og søkestrategi. Mai. 1998.
<http://www.ub.ntnu.no/ansatte/roars/it282/penslitt/sokepres1.html>
(sist sett 24062004)
20. Lancaster, F.W. (1972) *Vocabulary control for information retrieval*, 2nd edn, Greenwood 1979. ISBN 0-87815-053-6
21. C. G. Chowdhury. *Introduction to Modern Information Retrieval*. Library Assoc. Publ., London, 1999. ISBN 1856043185
22. Guinchat. C. and Menou, M., *General introduction to the techniques and information and documentation work*, Paris, Unesco, 1983.
23. Rowley J. E., *The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research*, Journal of information science, 20 (2), 1994, 108-19.

24. Hill, L. L. (2000). *Core elements of digital gazetteers: placenames, categories, and footprints*. In J. Borbinha & T. Baker (Eds.), *Research and Advanced Technology for Digital Libraries : Proceedings of the 4th European Conference, ECDL 2000* Lisbon, Portugal, September 18-20, 2000 (pp. 280-290). Berlin: Springer
25. Geographic Names Information System (GNIS)
<http://geonames.usgs.gov/>
(sist sett 24062004)
26. Hill, L. L., Frew, J., & Zheng, Q. (1999). *Geographic names: The implementation of a gazetteer in a georeferenced digital library*. D-Lib, January 1999
<http://www.dlib.org/dlib/january99/hill/01hill.html>
(sist sett 24062004)
27. Tilgjengelige webklienter fra Alexandria Digital Library, University of California, Santa Barbara
<http://webclient.alexandria.ucsb.edu/>
(sist sett 24062004)
28. NGA, GNS, digital gazetteer
<http://gnpswww.nima.mil/geonames/GNS/index.jsp>
(sist sett 24062004)
29. Statens Kartverk sitt websted
<http://www.statkart.no>
(sist sett 24062004)
30. Norgesglasset, Statens Kartverks digitale søketjeneste
<http://ngis2.statkart.no/ng2/ng2.html>
(sist sett 24062004)
31. Statens Kartverk, SOSI standard, Navn
http://www.statkart.no/standard/sosi/html_34/navn/navn.htm
(sist sett 24062004)
32. David J. Harper. (1999) Report on CIR99, the 2nd UK Conference on “*The Challenge of Image Retrieval*”
Newcastle upon Tyne, 25-26 February 1999
http://www.acm.org/sigir/forum/F99/CIR_99.pdf
(sist sett 24062004)
33. Carlo Meghini, Pasquale Savino. Fourth DELOS Workshop on Image Retrieval
http://www.ercim.org/publication/Ercim_News/enw31/meghini.html
(sist sett 24062004)
34. Thijs Westerveld. (2000) Image Retrieval: Content versus Context,
<http://wwwhome.cs.utwente.nl/~westerve/riao.pdf>
(sist sett 24062004)

35. Bilddatabaser og digitalisering – plattform for ABM-samverkan. Ett samverkamsprosjekt mellom Kungl. Biblioteket, Nationalmuseum, Riksantikvarieämbetet og Riksarkivet 2001 – 2002, ISBN 91-631-3387-3 (trykket rapport), ISBN 91-631-3388-1 (CD-ROM) Informasjon om prosjektet finnes på url: <http://abm.kb.se/akt4cd/slutrapporten/webbrapp/motiv.pdf>
(sist sett 24062004)
36. (ref. uaktuell)
37. (ref. uaktuell)
38. (ref. uaktuell)
39. (ref. uaktuell)
40. (ref. uaktuell)
41. (ref. uaktuell)
42. Vercoistre, A.M. og Paradis, *Metadata for Photographs: From Digital Libraries to Multimedia Applications* Lecture Notes in Computer Science, Research and Advanced Technology for Digital Libraries : Proceedings of the third European Conference, ECDL 1999 Paris, France, September 1999. Berlin: Springer (<http://link.springer.de/link/service/series/0558/bibs/1696/16960039.htm>),
(sist sett 24062004)
43. *Metadata Searching in a Multimedia Database Environment*, prosjektinformasjon: http://dopey.hil.unb.ca/Imaging_docs/IC/index.html
(sist sett 24062004)
Multimedia metadata skjema:
http://dopey.hil.unb.ca/Imaging_docs/IC/schemas.html
(sist sett 24062004)
44. (ref. uaktuell)
45. (ref. uaktuell)
46. Norsk Museumsutvikling:
<http://www.abm-utvikling.no/publisert/fulltekst/nmu3-1999/kapittel2.html>
(sist sett 24062004)
47. (ref. uaktuell)
48. (ref. uaktuell)
49. (ref. uaktuell)
50. (ref. uaktuell)
51. (ref. uaktuell)
52. (ref. uaktuell)
53. Galleri Nor, Nasjonalbibliotekets fotoarkiv:
www.nb.no/gallerinor
(sist sett 24062004)
54. Østby, Jon Birger. Feltkatalog for NKKM's EDB-prosjekter
<http://helmer.aksis.uib.no/regimus/feltkode.html>
(sist sett 24062004)

55. Eakins, J.P. and Graham M. E. (1999) *Content-based Image Retrieval*. A report to the JISC Technology Applications Programme. Januar 1999
<http://www.unn.ac.uk/iidr/research/cbir/report.html>
 (sist sett 24062004)
56. Evans, Jane., Shabajee, Paul. *Preliminary Results from the FILTER Image Categorisation and Description Exercise*. Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002: 53-60 © Firenze University Press
57. British Standards Institution. Guidelines for the establishment and development of monolingual thesauri. BS 5723. London, BSI, 1979
58. Fidel, R. (1984) *Online Searching Styles: A Case-Study-Based Model of Searching Behavior*. Journal of the American Society for Information Science, 35 (1984)
<http://www.ischool.washington.edu/fidelr/RayaPubs/OnlineSearchingStyles.pdf>
 (sist sett 24062004)
59. IFLA study group on the functional requirements for bibliographic records. *Functional requirements for bibliographic records : Final report*. Rapport, IFLA, 1998.
<http://www.ifla.org/VII/s13/frbr/frbr.pdf>
 (sist sett 24062004)
60. ADL Gazetteer Content Standard, Versjon fra 06/12/00
http://www.alexandria.ucsb.edu/gazetteer/gaz_content_standard.html
 (sist sett 24062004)
61. Hill, L. L. (2003). *Introduction to Georeferencing in Digital Libraries: Tutorial Document*. Santa Barbara: University of California, Santa Barbara.
 Contact the author at lhill@alexandria.ucsb.edu
62. Margaret E. Graham. (1999) *The description and indexing of images: Report of a survey of ARLIS members, 1998/99*.
<http://www.unn.ac.uk/iidr/ARLIS/>
 (sist sett 24062004)
63. Greg Janee, James Frew. (2002) *The ADEPT Digital Library Architecture*. Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries 2002, Portland, Oregon, USA July 14 - 18, 2002
64. (ref. uaktuell)
65. Hill, L., Buchel, O., Janée,G., Zeng, M. L. *Integration of Knowledge Organization Systems into Digital Library Architectures: Position Paper for 13th ASIS&T SIG/CR Workshop, "Reconceptualizing Classification Research"*
http://www.alexandria.ucsb.edu/paper_drafts/KOSpaper7-2-final.doc
 (sist sett 24062004)

66. Getty Thesaurus of Geographic Names Online
http://www.getty.edu/research/conducting_research/vocabularies/tgn/
(sist sett 24062004)
67. Paul Miller (2000) I say what I mean, but do I mean what I say? Report on outcomes from January's MODELS 11 workshop
<http://www.ariadne.ac.uk/issue23/metadata>
(sist sett 24062004)
68. Aitchison, J.(1996) Indexing languages and indexing. In Dossett, P. (ed.), *Handbook of special librarianship and information work*, 6th edn, London, Library Association Publishing, 1996.
69. National Information Standards Organization, Developing the Next Generation of Standards for Controlled Vocabularies and Thesauri
<http://www.niso.org/committees/MT-info.html>
(sist sett 24062004)
70. Tim Carven. *Thesaurus Construction*. Faculty of Information and Media Studies, University of Western Ontario.
<http://instruct.uwo.ca/gplis/677/thesaur/main00.htm>
(sist sett 24062004)
71. Alexandria Digital Library, University of California, Santa Barbara, Informasjon om forskningsområder
<http://www.alexandria.ucsb.edu/research/index.htm>
(sist sett 24062004)
72. GeoXwalk Gazetteer Project, prosjektets webside:
<http://hds.essex.ac.uk/geo-X-walk/>
(sist sett 24062004)
73. Textual – Geospatial Integration Services for the national SMETE Digital Library
http://www.ehr.nsf.gov/gpra/gpra_award_info.asp?AWD_ID=0121578
(sist sett 24062004)
74. Illinois Institute of Technology's Information Retrieval Laboratory
www.ir.iit.edu
(sist sett 24062004)
75. Trondheim Folkebiblioteks Bildebase, søkeside:
<http://www.trondheim.folkebibl.no/cgi-bin/websok-bilde>
(sist sett 24062004)
76. Carl Lagoze. Keeping Dublin Core Simple, Cross-Domain Discovery or Resource Description? D-Lib Magazine January 2001, Volume 7 Number 1, ISSN 1082-9873
<http://www.dlib.org/dlib/january01/lagoze/01lagoze.html>
(sist sett 24062004)

77. Lov 18. Mai 1990 Om stadnamn
<http://www.sprakrad.no/stadnamn.htm>
(sist sett 24062004)
78. (ref. uaktuell)
79. Christine Borgman. From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World. MIT Press, 2000. ISBN 026202473X.
80. Dublin Core Metadata Glossary
<http://library.csun.edu/mwoodley/dublincoreglossary.html>
(sist sett 24062004)
81. Bok og bibliotek 01-2003
<http://www.abm-utvikling.no/publisert/bob/03-1/s.31-55.pdf>
(sist sett 24062004)
82. Statens Kartverk, SOSI-standarden
<http://www.statkart.no/standard/sosi/html/sosi.htm>
(sist sett 24062004)
83. Nasjonalbiblioteket sitt websted, oversikt over bildebaser
<http://www.nb.no/baser/>
(sist sett 24062004)

Vedlegg

A. Dublin Core

i. Dublin Core for Galleri Nor - Diskusjon

Element: Title – et navn som ressursen er formelt kjent som

DC er opprinnelig utviklet for å beskrive dokumenter og dokumentlignende objekter. Tekstdokumenter har som regel en tittel eller en overskrift som er knyttet til innholdet i selve dokumentet. Et bilde / fotografi har ikke nødvendigvis en tittel. Galleri Nor sine beskrivelser av fotografiene har et metadataelement som kalles *Tittel*, men dette har ikke vært i bruk til nå. Beskrivelsene har også et element som kalles *Navn*. Dette er i bruk for en del av bildene, men ikke for alle. Der feltet *Navn* er i bruk er verdien ofte knyttet til navn på avbildet objekt. Det kan være navnet på en skole, navnet på en kirke, navnet til en person osv.

Dersom et bilde skal tilordnes en tittel vil det være naturlig at denne, på en eller annen måte, kyttes til bildets motiv. En vil også kunne ta i betraktning hvilken sammenheng bildene skal brukes i. V & P velger å gi bildene en tittel i form av en generell beskrivelse som kan brukes i mange ulike sammenhenger.

Dersom en skal tilordne tittel til bildene i denne sammenheng vil det være naturlig å velge tittel ut fra stedsangivelse eller feature type. Stedsnavn dekkes av elementet *coverage.spatial* mens feature type dekkes av elementet *subject* og til en viss grad av elementet *description*.

En kan også bruke informasjon fra Galleri Nor-feltet *Navn* for å angi tittel. Man kan bruke hele eller deler av denne informasjon.

Enkelte todimensjonale objekt, for eksempel malerier og grafikk, kan ha tilordnet tittel til originalverket, men det vil være vanskelig å finne informasjon om dette i eksisterende beskrivelser i Galleri Nor.

En kan stille spørsmål om det i det hele tatt har noen hensikt å tilordne tittel til bilder i en slik samling. Er det trolig at en kan oppnå noen gevinst i forbindelse med søk og gjenfinning ved å tilordne tittel til bildene? Jeg kan ikke se hvilken type informasjon en tittel skal kunne gi som ikke dekkes av noen av de andre metadataelementene.

Element: Creator – en entitet som er hovedansvarlig for innholdet i ressursen

Dette elementet kan brukes for å angi hvem som er fotografen bak bildet. En kan raffinere elementet for å være mer nøyaktig.

Ved å studere et og annet bilde utover de 50 først utvalgte dukker det opp andre problemer knytta til både dette og andre elementer. Det viser seg at en del av fotografiene har tegninger eller malerier som motiv der opphavsmannen til originalen ikke er fotografen. Her kan det da være problematisk å definere hva som er innhold i ressursen. Er det selve objektet som er motiv for fotografiet eller er det motivet for det avbildede objekt?

Slik disse bildene framstår i Galleri Nor er det nokså tydelig at fotografens mål har vært å avfotografere motivet i det avbildede objekt og en kan dermed ikke uten videre si at det er fotografen som er ansvarlig for innholdet i ressursen.

Ved å benytte raffinement som tydeliggjør at *creator* her er fotografen, kan en holde fast ved denne bruken uten å yte urett til tegneren eller maleren. En kan så finne en metode for å ivareta tilgjengelig informasjon om tegneren / maleren og eventuelt annen

informasjon om originalbildet. Dette kan gjøres ved å repetere dette elementet eller ved å benytte andre elementer som for eksempel elementet *Contributor*. I eksisterende beskrivelser i Galleri Nor kan det være vanskelig å identifisere opphavet til originalverket.

Element: Subject – emne for innholdet av ressursen

Her må en også ta i betraktning hvilken sammenheng ressursen skal brukes i og det vil være bildemotivet som ligger til grunn for tilordning av emne. I følge V & P bør en helst, for deres samling, benytte nøkkelord som er hentet fra en standardisert terminologi for bygninger og arkitektur. Selv benytter de vanlige termer, som vindu, tak, fasade, i mangel av et klassifikasjonsskjema som er spesielt relevant for denne type bygninger.

Bildesamlinga i Galleri Nor har et stort mangfold av motiver og hvert motiv kan knyttes til mange emner. Beskrivelsene i Galleri Nor har et element som kalles *Søkeord*. Her angis søketermer som er hentet fra et internt opparbeidet autoritetsregister over anvendte emneord.

Dersom en velger å knytte emne til feature type kan det være komplisert å definere emne for mange av bildene. Enkelte bilder er oversiktsbilder mens andre har ett eller flere tydelige objekt som er framtreddende som motiv for bildet. For bilder som har et framtreddende objekt som motiv kan en bruke objektets feature type som emne mens en, for andre bilder, kan finne en mer generell definisjon av emne, for eksempel en overordna feature type som dekker et større emneområde, eller bruke flere ulike feature types. Det at et oversiktsbilde kan dekke flere ulike feature types behøver ikke å være problematisk i så måte. Jeg vil foretrekke å benytte flere mer presise termer framfor mer generelle termer som skal dekke et større emneområde.

Det anbefales [10] å velge en verdi fra et kontrollert vokabular eller et formelt klassifikasjonsskjema. Et klassifikasjonsskjema kunne her være en feature type thesaurus, men det er et spørsmål om den ville være tilstrekkelig til å dekke det mangfold av motiver som finnes i bildesamlinga. Dette ville kanskje være tilfelle dersom bildesamlinga kun skulle brukes i sammenheng med geografisk referert informasjon, men det kan medføre å utelukke informasjon som kan være nyttig i andre sammenhenger. Det bør ikke være noe i veien for å å benytte et kontrollert vokabular for de termer som beskriver feature types, i kombinasjon med andre typer termer for å beskrive motivet dersom det er behov for dette. Jeg vil ta utgangspunkt i søkeord som allerede benyttes i Galleri Nor, vurdere disse i forhold til nytteverdi, vurdere å fjerne og /eller tilføye termer og så angi emne i form av tilstrekkelig antall nøkkelord. For termer som angir feature types vil jeg så langt som mulig bruke navneobjekttyper i henhold til SOSI standard som brukes av Statens kartverk [31].

Fotografen bak de fleste av bildene var kjent som landskapsfotograf og de fleste bildene er tatt i løpet av hans mange turer rundt omkring i Norge. Det er derfor ikke problematisk å knytte geografisk relaterte søketermer til de fleste av bildene i samlinga. Dette er da også gjort i mange av de opprinnelige beskrivelsene og mange av disse er klare og konsise og kan fungere utmerket som søketermer. Enkelte termer derimot er mer vage og upresise og må vurderes om de fortsatt skal være søketermer eller om en rett og slett skal fjerne disse fra et eventuelt emneordregister. Ett eksempel på en slik

term kan være *landskap*. *Landskap* er etter mitt syn en vag og upresis term, men samtidig er det ikke vanskelig å forstå at den blir brukt som søketerm i en samling som i stor grad består av landskapsbilder. Noe av det som jeg oppfatter som problematisk med å benytte slike vage og upresise termer er å sette grenser for hva som er hva. Hva dekkes for eksempel av termen *landskap*? I en samling med landskapsbilder vil det meste kunne dekkes av denne termen og det ville i så fall være nokså meningsløst å benytte den som søketerm.

Et spørsmål er om vi skal inkludere geografiske navn som søketermer under søk på emne. Hva kan en oppnå med det og hvordan kan en i så fall gjøre det på en god måte? Ved å benytte DC elementer for beskrivelser vil geografiske navn dekkes av elementet *spatial coverage* og en må da vurdere om en skal lagre samme informasjon flere ganger eller om en, på en eller annen måte, kan referere til verdiene i elementet *spatial coverage* fra elementet *subject*. Uansett hva en får ut av dette bringer diskusjonen, etter min mening, klarhet i at det må være en klar sammenheng mellom disse to elementene, *subject* og *spatial coverage*. Denne må være slik at der en entydig feature type er angitt som søketerm i elementet *subject*, må det angis et geografisk navn i elementet *spatial coverage*. Dette kan en selvfølgelig også snu på hodet og si at ethvert geografisk navn som er angitt i elementet *spatial coverage* må tilordnes en feature type som angis som søketerm i elementet *subject*. I denne oppgaven kan det hande at det er tilstrekkelig at gazetteeren holder orden på feature types.

Det er ikke noe i veien for å gjenta elementer i en beskrivelse der man bruker Dublin Core. En mulighet kan være å ha et 2. element *subject* som kan vise til elementet *spatial coverage*.

Søkeord kan eventuelt være knyttet til aktivitet som utføres av objekt i motivet, for eksempel fritidsaktiviteter som slalåm og fiske eller aktiviteter som såing og høsting. Disse vil ikke dekkes av en tesaurus som ADL FTT. Det vil være behov for å bruke en annen form for emneregister i tillegg til / i stedet for en feature type tesaurus. Kan en utvide begrepet informasjonsobjekt til også å omfatte mennesker for så å spesifisere roller som disse kan ha i tilknytning til aktiviteter (slalåmkjørere, fiskere, jordbrukere)? Eller spesifisere feature types som slalåmbakke, fiskeplass, åker for å dekke slike motiv? Hva hvis motivobjektet i bildet er en person, navngitt eller ikke navngitt?

Element: Description – en oppsummering av innholdet i ressursen

For tekstlige ressurser kan en benytte kvalifikatorene abstracts eller innholdsfortegnelse som beskrivelser. For bilder må man finne andre måter å beskrive innholdet på. V & P har i sin artikkel / samlingsbeskrivelse utvidet elementet med et skjema som gir en presisering av stedsangivelse med utfyllende adresse. Dette vil være mulig for noen av bildene i Galleri Nor, men på langt nær for alle. Det ville være bedre å gi en fritekst beskrivelse av motivet, med utgangspunkt i det som finnes i nåværende element *Utfyllende opplysninger* for bildene.

Element: Publisher – en entitet som er ansvarlig for å gjøre innholdet i ressursen tilgjengelig

Samlingen som her skal beskrives administreres og gjøres tilgjengelig av Nasjonalbiblioteket avdeling Rana. Dette vil være felles for alle bildene.

Element: Contributor – bidragsyter til innholdet i ressursen

Mange personer med ulike roller kan være bidragsytere til et fotografi. det kan for eksempel være den som har stått for digitaliseringen, den som er avbildet, den som har malt originalmotivet, tidligere eier osv. Av denne type informasjon behøver ikke alt å være interessant for bruker, men noe kan absolutt være det. Det kan av og til være behov for å angi personers navn og å knytte annen informasjon til dette. DCMI har ingen forslag til kvalifikatorer for dette elementet og om løsningen med kvalifikatorer er den beste skal jeg ikke påstå, men jeg kan vanskelig se noen bedre måte å løse dette på her og nå.

Her kan en regne den / de som har gjort arbeidet med å digitalisere fotografiene og eventuelt den / de som har stått for framkalling av bildene som bidragsytere til innholdet. DC tilbyr ingen kvalifikatorer, men V&P foreslår i sin artikkel at en kan spesifisere roller som knyttes til bidragsytere. I den grad man har flere bidragsytere med forskjellige roller så kan det være et godt forslag, men man må da definere ulike roller tydelig. Når det gjelder bildene i Galleri Nor så vil jeg tro at det vil være mulig å finne ut hvem som har stått for digitaliseringen av materialet, men det kan by på større problemer å finne ut hvem som har framkalt filmene og det er ikke usannsynlig at fotografen gjorde det selv. Slik informasjon vil være mindre interessant i denne sammenhengen.

Også dette elementet kan berøres av tilfeller der det avbildede objekt er en tegning eller et maleri. Dersom en velger å opprettholde fotografen som *creator* kan en her ha muligheten til å angi informasjon om tegneren eller maleren. I tilfeller der bildet er et portrett kan den som er portrettert angis som bidragsyter. Kanskje det ikke er så dumt å spesifisere roller slik som V&P foreslår, men også her må en vurdere i hvilken grad det er behov for å angi personers navn.

Element: Date – en dato som assosieres med en hendelse i ressursens livssyklus V & P anser det opprinnelige fotografiet som ressursen som beskrives og digitale versjoner som surrogater for dette. Dato er derfor knyttet til når bildet er tatt, ikke når det er digitalisert. Dette vil være greit i denne sammenhengen også fordi det vil være bildemotivet knyttet til stedsnavn og feature types som er interessant. Dato for fotografering kan derfor være et viktig element her mens dato for digitalisering er mindre interessant i denne sammenhengen.

Dato for bildene i Galleri Nor er angitt med årstall, ikke måned og dag. Anbefalt datoformat W3C-DTF kan godt brukes for å angi årstall. Dette skaper dermed ikke noe annet problem enn at datering blir upresis innenfor angitt tid. Enkelte bilder er knyttet til en mer presis dato i feltet *Utfyllende opplysninger* og / eller i feltet *Navn*. Det er ikke alltid denne samsvarer med årstallene i datofeltet så dette krever noe nærmere undersøkelser for å finne ut hva som er hva her.

Som svar på direkte forespørsel til Galleri Nor sies det at datering i feltet *Utfyllende opplysninger* angir mer nøyaktig dato for fotografering og at dersom det finnes datering i dette feltet så skal samme datering finnes i feltet *Datering*. Dette er en grei forklaring og det stemmer stort sett for de bildene som jeg har undersøkt. Videre sies det at det ikke skal forekomme datering i feltet *Navn*, noe som ikke stemmer så godt. Av et utvalg på 50 bilder forekommer det en eller annen form for datering i feltet *Navn* i 30 av de opprinnelige bildebeskrivelsene, mens jeg fant kun et tilfelle der datering i feltet *Utfyllende opplysninger* ikke samsvarer med datering i feltet *datering*. Foreløpig er det vanskelig å vite hva dateringene i feltet *Navn* representerer, men det er lite trolig at de har noe med selve bildet eller tidspunkt for fotografering å gjøre.

For de bildene som har en datering i feltet *Utfyllende opplysninger* og der denne samsvarer med datering i feltet *Datering* kan en angi en mer presis datering for når bildet er tatt. For resten av bildene i samlinga vil det være spørsmål om hvor presis man ønsker at informasjonen skal være, eller kanskje det er bedre å spørre om hvor presis må den være for å gjøre noen nytte. Når det gjelder gamle bilder så antar jeg at, i enkelte tilfeller, vil datering i form av årstall være bedre enn ingen datering selv om informasjonen er upresis.

Også for dette elementet dukker det opp problemer i forbindelse med at originalmotiver kan være tegnet eller malt for så å være fotografert. Det er bekreftet fra kilde i NB Rana at dateringer i Galleri Nor sitt felt *datering* for disse bildene er knyttet til den tiden bildet ble tegnet eller malt og ikke når det ble fotografert. Produksjonsdato for fotografiet er ikke tilgjengelig via Galleri Nor sine søkesider.

Med tanke på at disse bildene skal kunne brukes i sammenheng med geografisk referert materiale så er dette egentlig positivt. Datering vil i så fall være knyttet til motivet som bildet viser framfor å være knyttet til framstilling av selve fotografiet. Her kan det være litt vanskelig å følge V&P sin tolkning av dette elementet da de er enige om at fotografiet er originalressursen som beskrives mens digitaliserte versjoner er surrogater. I slike tilfeller her kan det like godt være fotografiet som er surrogatet og tegningen / maleriet som er originalressursen. I og med at dateringene, som ellers angis for fotografering, her angis for når bildet ble malt eller tegnet kan det være vrient å få til en beskrivelse av dette med DC. En mulig løsning kan være å knytte dato sterkere til bildemotivet og ikke skille mellom når det ble fotografert og når det ble tegnet eller malt.

For bildene i Galleri Nor må en vurdere om dateringsinformasjon like godt kan dekkes av elementet *coverage.temporal*.

Element: Type – innholdet i ressursens natur eller genre

Ifølge DCMI Type Vocabulary vil dette, for alle ressursene, være et *image*, altså et bilde.

V&P utvider dette elementet med et 2. typeelement for å angi om det er svart / hvitt eller fargebilde det er snakk om. Jeg kan se at det i enkelte sammenhenger kan være et poeng i at dette kan komme fram, men ikke akkurat i denne sammenheng. Alle fotografiene som her skal beskrives er sort / hvit og det vil ikke være noe stort problem å ta dette med seinere hvis ønskelig, eventuelt på samlingsnivå.

Element: Format – den fysiske eller digitale manifestasjon av ressursen DC tilbyr her kvalifikatorene extent og medium.

Medium vil være bildeformatet som angis i henhold til skjema IMT (Internet Media Type). Dette vil være JPG for alle bildene i Galleri Nor. Extent sier noe om størrelse eller varighet for ressursen. V&P utvider dette elementet med tre subelementer. Orientation forteller om bildet skal sees horisontalt eller vertikalt. Dette har betydning for deres bruk av bildene, for grupperinger og plasseringer som elementer på websider, men har mindre betydning i min sammenheng. Full og small refererer til størrelsen for to forskjellige surrogater for bildene. Disse vil være mulig å benytte også for bildene i Galleri Nor da de kan vises fram i ulike størrelser.

Element: Identifier – en entydig referanse til ressursen i en gitt kontekst DCMI anbefaler å identifisere en ressurs i henhold til et formelt identifikasjonssystem. Eksempler på slike er URI, URL, DOI og ISBN. V&P kan, i sine beskrivelser, ikke bruke URL da deres bilder skal inkluderes i en applikasjon som legges på en CD-ROM med et annet hierarki. De har derfor valgt å bruke en URI som reflekterer den relative organiseringen av applikasjonen.

Bildene i Galleri Nor er tilgjengelige via et søkegrensesnitt på www og skal fortsette å være det slik at en kan bruke aktuelle URL, men denne vil i så fall være gyldig på samlingsnivå. Et eget element for å angi samlingstilhørighet kan være nyttig, men dette kan også ivaretas via elementet *Relasjon.erDelAv*.

Når man beskriver en ressurs for at den skal oppdages på Internet vil en identifikator i form av en URL være en god løsning. For den generelle bruker vil bildene i Galleri Nor være tilgjengelige via et søkegrensesnitt på www. Hva slags identifikator systemet bruker internt i søkesystemet er i utgangspunktet uinteressant for meg som bruker. I Galleri Nor sine eksisterende beskrivelser gjøres både internnummer og tilvekstnummer synlig for bruker, uten at jeg helt kan se poenget med dette.

Dette avsnittet er tilføyd senere: I nyeste versjon (2004) av Galleri Nor sin søkeside er også tilvekstnummer gjort tilgjengelig for søk. Om dette vil ha noen nytteverdi for noen er vanskelig å si sikkert, men slik informasjon er vanligvis av mindre betydning for generelle brukere. Tilvekstnummer kan gjøre nytten her.

Element: Source – en referanse til en ressurs som ligger til grunn for aktuelle ressurs Ifølge V&P vil kilde være en referanse til det originale foto eller dets negativ. De foreslår å spesifisere et skjema for å holde orden på om originalen arkiveres i en personlig samling eller i en arkiveringsinstitusjon. I denne sammenhengen vil dette ikke være nødvendig da kilde for alle bildene vil være negativarkiv i Norsk Folkemuseum. Også her dukker det opp spørsmål i tilfeller der avbildet objekt er en tegning eller et maleri og der motivet for dette er målet for fotografen. Spørsmålet her er om en skal vise til originalen som kilde for ressursen i tillegg til negativarkiv i Norsk Folkemuseum. Denne informasjonen er ikke nødvendigvis tilgjengelig.

Element language - språk for det intellektuelle innholdet i ressursen

Dette elementet er ikke aktuelt å bruke for bilder og deres innhold, men det kan være aktuelt å angi språk for metadata. Dette vil i så fall være et eksempel på at det kan være nødvendig med ulike typer metadata og at det ikke alltid er lett å trekke grenser mellom de ulike typer metadata. DCMI tilbyr kvalifikatorer til dette elementet kun i form av kodeskjema for å representere språk, ikke for å raffinere elementet. Det vil være nødvendig å legge til et raffinement her, men man kan vurdere om dette heller skal være med på samlingsnivå.

Element: Relations – referanse til en relatert ressurs

Ifølge V&P kan det her angis relasjon til enten samlingen som helhet eller til et annet bilde for sammenligning. Elementet er ikke i bruk slik Galleri Nor fungerer nå, men begrepet forekommer av og til som en kommentar til feltet *Navn*.

Aktuelle relasjoner til samlingen som helhet kan her være: *Is Part Of* og / eller *Referenced By*. Begge relasjonselementene vil ha samme verdi som vil være URL til Galleri Nor. Jeg finner det mest hensiktsmessig å benytte *Is Part Of* siden alle bildene er del av Galleri Nor. Å angi relasjoner til andre bilder i eller utenfor samlinga kan være mulig. Det kan være greit dersom man skal organisere bildene for å browse hele eller deler av samlinga.

Kan det være aktuelt å peke til annen tilgjengelig info for eksempel om sted?

Element: Coverage – omfang / dekningsgrad av innholdet i ressursen

DCMI tilbyr kvalifikatorene *Spatial* og *Temporal Coverage*.

Så lenge det er et mål at bildene i Galleri Nor skal kunne inkluderes i en større sammenheng via geografisk referert informasjon vil elementet *coverage*, både *spatial* og *temporal*, være spesielt viktig i denne sammenhengen.

DCMI anbefaler, så langt det er mulig, å benytte stedsnavn og navngitte tidsperioder framfor numeriske identifikatorer som geografiske koordinater og datoer.

Spatial Coverage vil angis i form av stedsnavn og kan i en større sammenheng angis i forma av koordinater for geografisk avtrykk. Utvalgte bilder i Galleri Nor er allerede knyttet til stedsnavn, disse kan brukes i den grad de vil kunne finnes igjen i en liste over geografiske navn. I tillegg er det slik at enkelte av bildene er knyttet til flere stedsnavn fordi motivet inneholder flere ulike feature types. Ett sted kan også ha flere navn eller ha skiftet navn og dermed være knyttet til flere navn. En kan velge ett stedsnavn for hvert bilde. Det vil i så fall være en subjektiv vurdering av hvilket stedsnavn som skal knyttes til bildet. Dette behøver ikke å være noe problem da en kan velge å fortsatt la et bilde være knyttet til flere stedsnavn.

Videre må en ta stilling til hvor detaljert en skal spesifisere den geografisk relaterte informasjonen om bildemotivet. V & P introduserer et adresseskjema for å kunne sortere bilder etter landsby, gate og hus. Det vil være mulig å gjøre noe lignende også for bildene i Galleri Nor, men svært varierende informasjon i de opprinnelige beskrivelsene vil gjøre dette vanskelig for de fleste bildene.

Bildene har angitt område som peker til fylke eller kommune. For noen av bildene angis mer enn en kommune. Noen by-bilder har spesifisert bydel og noen få har gatenavn. Noen bilder har spesifisert gårdsnavn.

For å ivareta mest mulig informasjon om det enkelte bilde må en gjenta dette elementet tilstrekkelig antall ganger slik at en får med alle aktuelle stedsnavn.

Når det gjelder Temporal Coverage vil det, for denne bildesamlinga, være vanskelig å følge DCMI's anbefaling om å bruke navngitte tidsperioder. Datamodellen sier ikke noe om slike perioder og enkelte av bildene er knyttet til kun ett årstall. Det er fortsatt noen uklarheter i forbindelse med dateringer for bildene. Informasjon i dette elementet vil derfor være knyttet til datoangivelsen i elementet *Dato* dersom ikke annen datering er tilgjengelig. Det vil her være mulig å benytte DCMI Period både for bildene som er knytta til ett årstall og for bilder som er knytta til to årstall. Jeg anbefaler ett element for periodens start og ett element for periodens slutt. Periodene er ikke navngitt, men kun tidfestet med årstall. For bilder som er knyttet til kun ett årstall brukes ikke elementet for periodens slutt. W3C-DTF brukes her.

V&P har valgt å ikke bruke dette elementet i sine beskrivelser, men elementet kan være viktig å ha med i beskrivelsen av bildene i Galleri Nor i den sammenhengen de her er tenkt brukt. Dateringer kan gi verdifull informasjon som kan brukes for indirekte referering til andre informasjonsressurser.

Element: Rights – angir hvem som har rettigheter i forhold til ressursen.

V & P nevner at det kan være uklarheter i forhold til rettigheter for originalkilden og for det digitaliserte fotografi, men slår fast at dette elementet dekker både originalkilden og alle mulige digitaliserte versjoner av denne. Denne tolkningen kan greit overføres til Galleri Nor. Alle bildene eies av Norsk Folkemuseum og eieren har alle rettigheter i forhold til bildene.

Ifølge [13] kan det være greit å oppgi en URL til et sted der en kan finne informasjon om rettigheter knyttet til bruk av ressursen.

ii. DCMI Metadata termer pr. 04.03.2003

En oppdatert spesifikasjon for alle metadatermer som holdes vedlike av DCMI er tilgjengelig på [www. fra 04.03.2003](http://www.dublincore.org). DCMI Metadata Terms inkluderer fire forskjellige typer termer: elementer, element-raffinement, kodeskjema og vokabulartermer fra DCMI Type Vokabular [12].

Termer av type Element

Et element er en egenskap ved ressursen, et attributt som ressursen kan ha så som tittel eller forfatter osv.

Elementer inkluderer, i tillegg til Dublin Core Metadata element sett 1.1 med 15 elementer som anbefales av DCMI, også andre elementer som anbefales av DCMI. Elementer kan benyttes som de er eller de kan kombineres med kvalifikatorer som kan være av typen element-raffinement eller av typen kodeskjema.

Termer av type element-raffinement

Element -raffinement tilbys for å gi mulighet til å spesifisere betydningen av et DC element.

Et raffinement bør være av en slik art at det kun spesifiserer og ikke utvider det semantiske innholdet i elementet. Dette innebærer at en skal ha mulighet til å fjerne eventuelle raffinement, bruke det opprinnelige kjerneelement som om det ikke var raffinert og fremdeles opprettholde en mening mellom element og verdi som er korrekt for elementet (keeping DC simple). Et eksempel kan være elementet coverage som skal si noe om utstrekning eller omfang til ressursen. For å spesifisere om en snakker om omfang i tid eller rom kan en benytte de kvalifiserte elementene coverage.spatial og coverage.temporal.

I tillegg til raffinement av elementer som anbefales av DCMI inkluderer spesifikasjonen også termer for raffinement som er av en slik art at en implementasjonsenhet har demonstrert behov for termen uten at den nødvendigvis oppfyller kriteriet om at den må være nyttig på tvers av domener eller nyttig for oppdaging av ressurser. Disse termene har status som "conforming"

(<http://dublincore.org/usage/documents/process/#conforming>).

Termer av type kodeskjema

Kodeskjema innføres som vokabularkontrollerende virkemiddel eller for å indikere hvordan verdien i et element skal tolkes. Et eksempel på kontrollert vokabular kan være å si at dette geografiske navnet er valgt fra en registrert tesaurus over geografiske navn. Et annet eksempel kan være datoformat, som kan uttrykkes på mange forskjellige måter. Ved å benytte et standardisert format som ISO 8601 (YYYY-MM-DD) fjerner en muligheten for misforståelser. Skjema kan benyttes også for å tolke elementverdien i kvalifiserte elementer som for eksempel coverage.spatial schema=DCMI point, som sier at dette elementet dekker omfang i rom angitt ved et punkt.

Kodeskjema har status som "registrert", det vil si at DCMI tilbyr informasjon om disse, men ikke nødvendigvis spesifikke anbefalinger. Kodeskjema kan brukes for å kvalifisere elementer som er raffinert så vel som de som ikke er det.

Termer av type vokabular-term

DCMI Type Vocabulary er en generell og domeneuavhengig liste over anbefalte termer som kan brukes for å velge verdi til DC elementet Type. Termene brukes for å identifisere ressursens natur eller sjanger.

Oversikt over DC Termer

Dc metadata element sett er oversatt til mange språk, deriblant til norsk. Versjon 1.1 er, med tillatelse fra DCMI, oversatt av Frank B. Haugen og Carol van Nuys for den norske katalogkomite datert mars 2002. (<http://www.nb.no/katkom/dublincore.html>)

Elementnavnene i versjon 1.1 har vært stabile over lengre tid, men det forekommer enkelte endringer i forhold til kvalifikatorer og anbefalt bruk av disse. Beskrivelsene under kan derfor avvike noe fra den offisielle norske oversettelsen da de er forsøkt oppdatert i forhold til siste offentliggjorte dokumentasjon fra DCMI fra mars 2003 [12].

DC elementer som inngår i DCMI Metadata Element set v.1.1

DC termnavn	Norsk elementnavn	Elementbeskrivelse
Title	Tittel	Et navn gitt ressursen. Tittelen er vanligvis navnet som ressursen formelt er kjent under.
Creator	Opphavsmann	En enhet som er hovedansvarlig for ressursens intellektuelle innhold. Eksempler på en opphavsmann er en person, en organisasjon eller tjeneste. Vanligvis angir man her opphavsmannens navn.
Subject	Emne- og stikkord	Et uttrykk for ressursens emnemesseige innhold. Vanligvis vil et emne bli uttrykt ved hjelp av et emneord, en frase eller en klassifikasjonskode som beskriver ressursens emne. Det anbefales å velge en verdi fra kontrollerte emneordlister eller fra formelle klassifikasjonsskjema.
Description	Innholdsbeskrivelse	En redegjørelse for ressursens innhold. En beskrivelse av ressursens innhold kan omfatte et sammendrag, en innholdsfortegnelse, en referanse til en grafisk framstilling av innholdet eller en uformell beskrivelse av innholdet, men er ikke avgrenset til dette.
Publisher	Utgiver	En enhet som er ansvarlig for å ha gjort ressursen tilgjengelig. Eksempler på en utgiver er en person, en organisasjon eller en tjeneste. Vanligvis angir man her utgiverens navn.
Contributor	Bidragster	En enhet som har bidratt til ressursens innhold. Eksempler på en bidragster er en person, en organisasjon eller en tjeneste. Vanligvis angir man

Date	Dato	<p>her bidragsyterens navn.</p> <p>En dato knyttet til en hendelse i levetiden til en ressurs.</p> <p>Dato knyttes vanligvis til en ressurs' tilblivelse eller offentliggjørelse. Det anbefales å angi datoverdien slik det er definert av ISO 8601 på formatet ÅÅÅÅ-MM-DD.</p>
Type	Ressurstype	<p>En egenskap ved ressursen eller en sjangerbetegnelse som beskriver ressursens innhold. Type inkluderer termer som beskriver innholdet i ressursen i generelle kategorier, funksjoner, sjangre, eller som angir om det dreier seg om et enkelt stående dokument eller en samling. Det anbefales å velge en verdi fra en kontrollert liste (for eksempel DCMI Type Vocabulary). Bruk elementet FORMAT for å beskrive ressursens fysiske eller digitale form.</p>
Format	Format	<p>En ressurs' fysiske eller digitale manifestasjon. Format inkluderer vanligvis ressursens medietype eller omfang. Format kan brukes til å bestemme hva slags programvare, maskinvare eller annet utstyr som er nødvendig for å vise eller bruke ressursen. Eksempler på omfang er størrelse og varighet. Det anbefales å velge en verdi fra en kontrollert liste (for eksempel Internet Media Types som definerer format for datamedia).</p>
Identifiser	Identifikator	<p>En entydig henvisning til ressursen i en gitt kontekst.</p> <p>Det anbefales å identifisere ressursen ved hjelp av en streng eller et nummer i samsvar med et formelt identifikasjonssystem. Eksempler på formelle identifikasjonssystemer er Uniform Resource Identifier (URI, herunder Uniform Resource Locator (URL)), Digital Object Identifier (DOI) og International Standard Book Number (ISBN).</p>
Source	Kilde	<p>En referanse til en ressurs som foreliggende ressurs er utledet av.</p> <p>Aktuelle ressurs kan stamme helt eller delvis fra "kilderessursen". Det anbefales å identifisere ressursen ved hjelp av en streng eller et nummer i samsvar med et formelt identifikasjonssystem.</p>
Language	Språk	<p>Språk for ressursens intellektuelle innhold. Det anbefales å bruke RFC 3066, som i samsvar med ISO 639, definerer to- og trebokstavers primære språkkoder med valgfrie subkoder. Eksempler kan være "en" eller "eng" for English, "en-GB" for engelsk i Storbritannia, "no" eller "nor" for norsk. Det finnes egne koder for norsk bokmål og nynorsk slik at en har mulighet til å spesifisere dette.</p>

Relation	Relasjon	En henvisning til en relatert ressurs. Det anbefales å henvise til den beslektede ressursen ved hjelp av en streng eller et nummer i samsvar med et formelt identifikasjonssystem.
Coverage	Dekning	Utstrekningen til eller omfanget av ressursens innhold. Dekning inkluderer vanligvis begrep for rom (et geografisk navn eller geografiske koordinater), begrep for tid (navn på en tidsperiode, dato, eller tidsintervall) eller jurisdiksjon (som for eksempel en navngitt administrativ enhet). Det anbefales å velge en term fra en kontrollert liste (for eksempel Thesaurus of Geographic Names) og, der det er mulig, velg navngitte steder eller tidsperioder framfor numeriske identifikatorer som koordinater eller tidsintervall.
Rights	Rettigheter	Informasjon om rettigheter knyttet til ressursen. Elementet inneholder vanligvis en erklæring om hvordan rettigheter knyttet til ressursen skal håndteres, eller en henvisning til en tjeneste som gir slik informasjon. Informasjon om rettigheter omfatter for eksempel Intellectual Property Rights (IPR), Copyright m.m. Dersom elementet ikke er registrert, skal det ikke gjøres noen antakelse om hvilke rettigheter som er knyttet til ressursen.

Andre termer av type element som anbefales av DCMI:

DC	Norsk	Elementbeskrivelse
Elementnavn	Elementnavn	
Audience	Publikum	En klasse av entiteter for hvem ressursen er intendert eller nyttig. En klasse av entiteter kan bestemmes av opphavsmann, utgiver eller en tredje part.

Temer av type element-raffinement som anbefales av DCMI:

DC element	Raffinement	Beskrivelse
termnavn	termnavn	
Title	Alternative	Enhver form av tittelen som kan brukes i stedet for eller som alternativ til ressursens formelle tittel. Dette kan være forkortelser av tittel så vel som oversettelser.
Creator	Ingen raffinement	
Subject Description	tableOfContents Abstract	En liste over subenheter av innholdet i ressursen En oppsummering av ressursens innhold

Publisher		
Contributor		
Date	Created	Dato for skapelsen av ressursen
	Valid	Dato for ressursens gyldighet i tid, angis ofte som et tidsintervall/en periode.
	Available	Dato for når ressursen vil være eller har vært tilgjengelig, angis ofte som et intervall/ en periode.
	Issued	Dato for formell utgivelse (eks. Publisering) av ressursen
	Modified	Dato for når ressursen ble endret
Type		
Format	Extent	Størrelse eller varighet for ressursen
	Medium	Materiale eller fysisk bærer av ressursen
Identifier		
Source		
Language		
Relation	isVersionOf	Beskrevet ressurs er en versjon, utgave eller tilpasning av refererte ressurs
	hasVersion	Beskrevet ressurs har en versjon, utgave eller tilpasning som er den refererte ressurs
	isReplacedBy	Beskrevet ressurs er supplert, forflyttet eller erstattet av refererte ressurs
	Replaces	Beskrevet ressurs supplerer, forflytter eller erstatter refererts ressurs
	isRequiredBy	Beskrevet ressurs kreves av referert ressurs, fysisk eller logisk
	Krever	Beskrevet ressurs krever referert ressurs for å støtte dens funksjon, levering eller sammenheng i innhold
	isPartOf	Beskrevet ressurs er en fysisk eller logisk del av referert ressurs
	hasPart	Beskrevet ressurs inkluderer referert ressurs, fysisk eller logisk
	isReferencedBy	Beskrevet ressurs er referert, plassert eller på annet vis pekes til av referert ressurs
	References	Beskrevet ressurs refererer, plasserer eller, på annet vis, peker til referert ressurs
	isFormatOf	Beskrevet ressurs har samme intellektuelle innhold som referert ressurs, men i et annet format
	hasFormat	Beskrevet ressurs eksisterte før referert ressurs som har samme intellektuelle innhold presentert i et annet format.
	conformsTo	En referanse til en etablert standard som ressursen er konform til
Coverage	Spatial	Spatiale (romlige) karakteristikk av ressursens intellektuelle innhold
	temporal	Temporale (tidsmessige) karakteristikk av ressursens intellektuelle innhold
Rights		
Audience	Mediator	En klasse av entiteter som megler tilgang til ressursen?? og for hvem ressursen er intendert eller

nyttig

Termer av type element-raffinement som har status som konforme:

DC element termnavn som kvalifiseres	Raffinement termnavn	Beskrivelse
Date	dateAccepted	Dato for akseptasjon av ressursen (for eksempel en avhandling av et universitet, en artikkel av et tidsskrift)
	dateCopyrighted	Dato for deklarasjon av rettigheter forbundet med reproduksjon og publisering av ressursen
	dateSubmitted	Dato for innlevering av ressursen
Audience	educationLevel	Generell beskrivelse av kontekst for læring eller øving. Alternativt en mer spesifikk deklarasjon av "hvor" publikum befinner seg i forhold til progresjon gjennom en gitt lærings- eller øvingskontekst.
Rights	accessRights	Informasjon om hvem som har tilgang til ressursen eller en indikasjon på dens sikkerhetsstatus. Tilgangsrettigheter kan inkludere informasjon angående tilgang eller restriksjoner basert på private, sikkerhetsmessige eller andre reguleringer.
Identifiser	bibliographicCitation	En bibliografisk referanse for ressursen. Det anbefales å inkludere tilstrekkelige bibliografiske detaljer for å identifisere ressursen så entydig som mulig, det være seg om "citation" er i en standardiser form eller ikke.

Termer av type kodeskjema:

DC element eller raffinement termnavn	Kodeskjema termnavn	Definisjon
Subject	LCSH	Library of Congress Subject Headings
	MESH	Medical Subject Headings
	DDC	Dewey Decimal Classification
	LCC	Library of Congress Classification
	UDC	Universal Decimal Classification
Type	DCMIType	DCMI TypeVocubular
Format	IMT	The Internet media type of the resource
Language	ISO639-2	ISO 639-2: koder for representasjon av navn på språk.
	RFC1766	Internet RFC 1766 'Tags for the identification of Language' spesifiserer en tobokstavs kode tatt fra ISO 639, fulgt av en tobokstavs landkode tatt fra

		ISO 3166.
	RFC3066	Internet RFC 3066 'Tags for the Identification of Languages' spesifiserer en primær subtag som er en tobokstavs kode tatt fra ISO 639-1 eller en trebokstavs kode fra ISO 639-2, fulgt av en valgfri tobokstavs landkode fra ISO 3166. Når et språk i ISO 639 har både en tobokstavs og en trebokstavs kode skal en bruke tobokstavskoden. Når kun trbokstavskode finnes skal denne brukes. Denne RFC erstatter RFC 1766.
Identifier Source Relation Spatial	URI	Uniform Resource Identifier
	Point	DCMI Point identifiserer et punkt i rom ved hjelp av geografiske koordinater
	ISO3166	ISO 3166 koder for representasjon av navn på land
	Box	DCMI Box identifiserer en region i rom ved hjelp av geografisk grenser
	TGN	The Getty Thesaurus of Geographic Names
Date	Period	Spesifiserer grensene for et tidsintervall
Temporal	W3CDTF	W3C koderegler for dato og tid – en profil basert på ISO 8601

Termer av type vokabular-term

Termnavn	Beskrivelse
Collection	En samling er en aggregering av elementer. Termen ”collection” angir at ressursen er beskrevet samlet som gruppe, dens deler kan være beskrevet og kan navigeres separat.
Dataset	Et datasett er informasjon kodet i en definert struktur (lister, tabeller, databaser), intendert å være nyttig for direkte maskinprosessering.
Event	En event/hendelse er en ikke-varig tidsbasert forekomst. Metadata for en hendelse tilbyr beskrivende informasjon som danner basis for å kunne oppdage informasjon som er relatert til hendelsen. En ressurs av type event kan være av en slik art at den ikke kan gjenfinnes dersom den beskrevne instantiering er utgått eller ikke har forekommet enda. (eksempler: utstilling, direkteoverføring på web, konferanse osv.)
Image	Et bilde er primært en symbolsk visuell representasjon forskjellig fra tekst. For eksempel bilder og fotografier av fysiske objekt, malerier, trykk, tegninger, grafikk, film, kart, noter osv. ”Image” kan inkludere både elektronisk og fysisk rerepresentasjon.
InteractiveResource	En interaktiv ressurs er en ressurs som krever interaksjon med bruker for å bli forstått, utført eller erfart. For eksempel former på websider, applets, multimedia læreobjekter osv.
Service	En tjeneste er et system som tilbyr en eller flere funksjoner som

	er av verdi for brukeren. For eksempel en fotokopieringstjeneste, banktjenester, autentiseringstjenester, en Z39.50 eller en webtjener.
Software	Software er programvare i kilde eller kompilert form som er tilgjengelig for varig installasjon på en annen maskin. For programvare som brukes kun for å skape interaktive omgivelser, bruk "interactive" i stedet.
Sound	En ressurs hvis innhold primært er hørbart. For eksempel musikkfilformat, en lyd-cd, innspilt tale eller annen lyd.
Text	En ressurs hvis innhold primært er lesbare ord. For eksempel bøker, brev, aviser osv. faksimiler og bilder av tekst er av sjangeren tekst.
PhysicalObject	Et ikkelevende tredimensjonalt objekt eller substans. For eksempel en computer, en pyramide, en skulptur. Digitale representasjoner av, eller surrogater for, disse objektene bør bruke "Image", "Text" eller en av de andre typene.

B. ADL GAZETTEER CONTENT STANDARD

Elementer som er markert med en asterisk(*) og uthevet skrift er obligatoriske. Andre elementer er valgfrie. Når et obligatorisk element er et sub-element til et ikke-obligatorisk element, er dette obligatorisk kun såfremt overordna element er i bruk.

Repeterbare elementer etterfølges av (R).

(http://www.alexandria.ucsb.edu/gazetteer/gaz_content_standard.html)

1. **Geographic Feature ID** * < feature_id >
2. **Geographic Name** *
 1. **Name** * < primary_name > [the primary name for feature in a particular gazetteer application]
 2. Name Source < name_source >
 3. Etymology < etymology >
 4. Language < language_used > (default is English)
 5. Pronunciation < pronunciation >
 6. Transliteration Scheme Used < trans_schema >
 7. Character Set (default is ASCII) < character_set >
 8. **Current / Historical Note** * (default is Current)
< is_current > {boolean}
 9. Beginning Date < beginning_date >
 10. Ending Date < ending_date >
 11. Time Period Note < time_period_note >
 12. Source Mnemonic < source_mnemonic >
 13. Entry Date < entry_date >
3. Variant Geographic Name (R)
 1. **Variant Name** * < feature_name >
 2. Name Source < name_source >
 3. Etymology < etymology >
 4. Language < language_used > (default is English)
 5. Pronunciation < pronunciation >
 6. Transliteration Scheme Used < trans_schema >
 7. Character Set (default is ASCII) < character_set >
 8. **Current / Historical Note** * (default is Current) < is_current > {boolean}
 9. Beginning Date < beginning_date >
 10. Ending Date < ending_date >
 11. Time Period Note < time_period_note >
 12. Source Mnemonic < source_mnemonic >
 13. Entry Date < entry_date >
4. **Type of Geographic Feature** *
 1. **Feature Type Schema** * < schema_name > [e.g., ADL Feature Type Thesaurus]
 2. **Feature Type** * (R) < type >
5. Other Classification Terms (R)
 1. Classification Schema < classification_name >
 2. Classification Term (R) < term >
6. Geographic Feature Code (R)
 1. **Feature Code Schema** * < code_schema >
 2. **Feature Code** * < code >

3. Source Mnemonic < source_mnemonic >
4. Entry Date < entry_date >
7. **Spatial Location ***
 1. **Planetary Body *** < planet_body > (default of Earth)
 2. **Spatial Representation * (R)**
 1. **Bounding Box Spatial Geometry Representation ***
 - < primary_bounding_box >
 - 1. **West Bounding Coordinate *** < west_bounding_coor >
 - 2. **East Bounding Coordinate *** < east_bounding_coor >
 - 3. **South Bounding Coordinate *** < south_bounding_coor >
 - 4. **North Bounding Coordinate *** < north_bounding_coor >
 - 5. **Current / Historical Note *** (default is Current)
 - < is_current > {boolean}
 - 6. Beginning Date < beginning_date >
 - 7. Ending Date < ending_date >
 - 8. Time Period Note < time_period_note >
 - 9. Measurement Date, Beginning Date
 - < measurement_begin_date >
 - 10. Measurement Date, Ending Date < measurement_end_date >
 - 11. Method of Measurement < measurement_method >
 - 12. Accuracy of Measurement < measurement_accuracy >
 - 13. Source Mnemonic < source_mnemonic >
 - 14. Entry Date < entry_date >
 2. Detailed Spatial Geometry Representation (R) < g_points > [set of points; dependent on system capabilities and requirements; can represent set of non-contiguous areas]
 1. **Detailed Spatial Geometry Representation ***
 - < geometry_type > {point, bounding box, linear, complex object}
 2. **Number of Points *** < num_points >
 3. **Points Order *** < points_order >
 4. **(Longitude, Latitude) (R) *** < long_lat_pairs >
 5. **Current / Historical Note *** (default is Current)
 - < is_current > {boolean}
 6. Beginning Date < beginning_date >
 7. Ending Date < ending_date >
 8. Time Period Note < time_period_note >
 9. Measurement Date, Beginning Date
 - < measurement_begin_date >
 10. Measurement Date, Ending Date < measurement_end_date >
 11. Method of Measurement < measurement_method >
 12. Accuracy of Measurement < measurement_accuracy >
 13. Source Mnemonic < source_mnemonic >
 14. Entry Date < entry_date >
8. Street Address (Physical Address)
 1. **Address *** < street_address >
 2. **City *** < city >

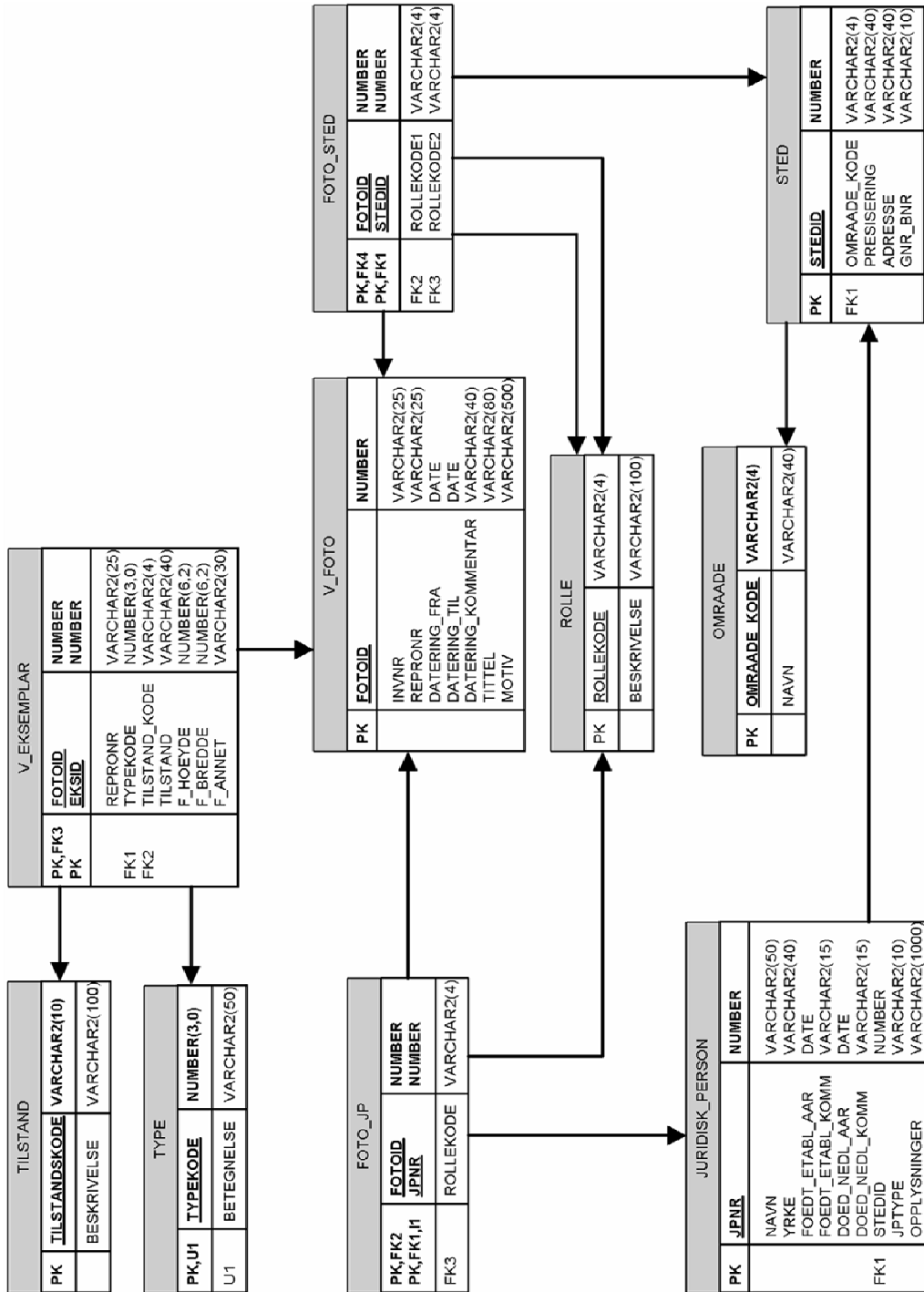
3. **State or Province** * < state_province >
4. Postal Code < postal_code >
5. Country < country >
6. Source Mnemonic < source_mnemonic >
7. Entry Date < entry_date >
9. Related Feature (R)
 1. **Type of Relationship** * < relationship_type > {in-state-of, in-province-of, in-county-of, in-country-of, in-region-of, part-of, formerly-known-as}
 2. **Geographic Name of Related Feature** * < related_name >
 3. Related ADL Feature ID < related_feature_id >
 4. **Current / Historical Note for Relationship** * < is_current >
 5. Beginning Date of Relationship < begin_date_rel >
 6. Ending Date of Relationship < end_date_rel >
 7. Time Period Note < time_period_note >
 8. Source Mnemonic < source_mnemonic >
 9. Entry Date < entry_date >
10. Description
 1. **Short Description** * < short_description >
 2. Source Mnemonic < source_mnemonic >
 3. Entry Date < entry_date >
11. Geographic Feature Data (R)
 1. **Type of Geographic Feature Data** * < label_name >
 2. **Geographic Feature Data Value*** < data_value >
 3. **Geographic Feature Data Value Unit of Measure** * < measurement_unit >
 4. Explanatory Note < explanatory_note >
 5. Beginning Date < beginning_date >
 6. Ending Date < ending_date >
 7. Time Period Note < time_period_note >
 8. Source Mnemonic < source_mnemonic >
 9. Entry Date < entry_date >
12. Link to Related Source of Information (R)
 1. **Description of Linked Item** * < link_description >
 2. **URL** * < link_url >
 3. Source Mnemonic < source_mnemonic >
 4. Entry Date < entry_date >
13. Supplemental Note
 1. **Note** * < note >
 2. Source Mnemonic < source_mnemonic >
 3. Entry Date < entry_date >
14. **Metadata Information** *
 1. Entry Note < entry_note >
 2. **Entry Date** * < entry_date >
 3. **Modification Date** * < mod_date >

Source Information * (R)

1. **Source Mnemonic** * < source_mnemonic >

2. **Contributor Organization** * < organization_name >
3. Contributor Web Site
 1. **Web Site Title** * < web_site_title >
 2. **URL** * < web_site_url >
4. Contact Person < contact_person >
5. Email < email >
6. Telephone Number < telephone >
7. **Contributor Address** *
 1. **Address** * < street_address >
 2. **City** * < city >
 3. **State or Province** * < state_province >
 4. **Postal Code** * < postal_code >
 5. **Country** * < country >
8. **Source Information** * (R)
 1. Author Statement < author >
 2. **Title** * < title >
 3. Edition < edition >
 4. Series Name < series_name >
 5. Series Issue < series_issue >
 6. **Publisher** * < publisher >
 7. **Publication Date** * < publication_date >
 8. Publication Date Note < pub_date_note >
 9. Pages < pages >
 10. Source Identifier < source_identifier > [e.g., ISBN]
 11. Source_URL < source_url >

C. Nasjonalbibliotekets databasemodell for Galleri Nor

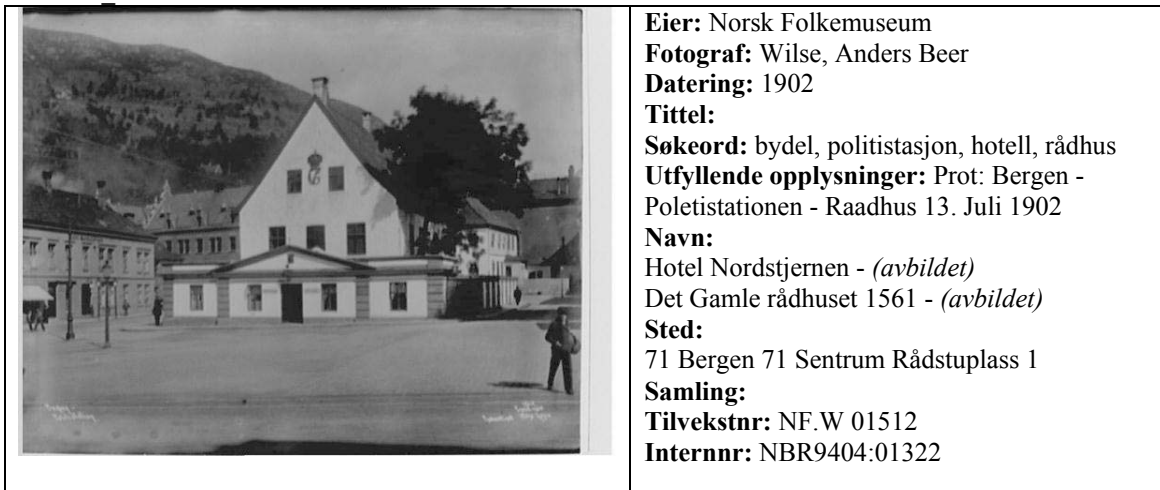


D. Testsamlinga – bilder og metadata²

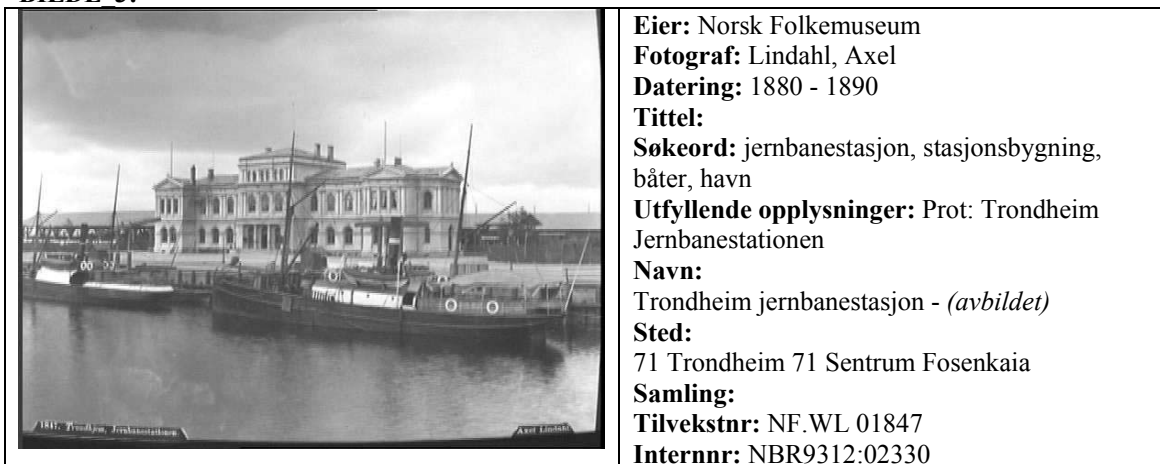
BILDE 1:



BILDE 2:



BILDE 3:



² Etter bilde nummer 30 er det enkelte bilder tatt ut av testsamlingen. Dette er årsak til brudd i nummereringen. Alle bilder som skal være søkbare i prototypen er representert her.

BILDE 4:

Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: bydel, losjebygning, klubbhus
Utfyllende opplysninger: Prot: Bergen Klubbygningen "Den Gode Hensigt" taget fra Nord
Navn:
 Logen 1883 - (*avbildet*)
 Den Gode hensigt (selskapsklubb) 1798 - (*relasjon til bildet*)
Sted:
 71 Bergen 71 Sentrum Ole Bulls plass 14
Samling:
Tilvekstnr: NF.WL 00478
Internnr: NBR9312:00506

BILDE 5:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1902
Tittel:
Søkeord: bygning
Utfyllende opplysninger: Prot: Sarpsborg - Festiviteten og Sandesundveien 11. Juni 1902
Navn:
 Festiviteten - (*avbildet*)
Sted:
 71 Sarpsborg 71 Sarpsborg Sandesundsveien
Samling:
Tilvekstnr: NF.W 01308
Internnr: NBR9404:01121

BILDE 6:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1903
Tittel:
Søkeord: bygning
Utfyllende opplysninger: Prot: Drøbak Biologiske Station Juli 1903
Navn:
 Biologisk stasjon - (*avbildet*)
Sted:
 71 Drøbak 71 Drøbak Biologiveien
Samling:
Tilvekstnr: NF.W 00324
Internnr: NBR9404:00223

BILDE 7:



Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1903
Tittel:
Søkeord: teaterbygning
Utfyllende opplysninger: Prot:
Nationaltheatret bakfra Jan.1903
Navn:
Nationaltheatret 1899 - *(avbildet)*
Sted:
71 Oslo 71 Sentrum Stortingsgata 15
Samling:
Tilvekstnr: NF.W 00943
Internnr: NBR9404:00752

BILDE 8:



Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1910
Tittel:
Søkeord: kontorbygning, fasade
Utfyllende opplysninger: Prot:
Elektrisitetværkets Kontorhygn. Raadhusg
Navn:
Kristiania elektrisitetsverk - *(relasjon til bildet)*
Sted:
71 Oslo Sentrum Rådhusgata
Samling:
Tilvekstnr: NF.WB 02969
Internnr: NBR9102:00983

BILDE 9:



Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1911
Tittel:
Søkeord: papirfabrikk, kontorbygning
Utfyllende opplysninger: Prot: Tinfos
Fabrikker Kontor 18/11 Neg: Tinfos
Papirfabriks Kontor
Navn:
Tinfos papirfabrik 1873 - *(avbildet)*
Sted:
71 Notodden 71 Notodden
Samling:
Tilvekstnr: NF.W 13511
Internnr: NBR9404:09638

BILDE 10:



Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1930
Tittel:
Søkeord: husmorskole, fasade
Utfyllende opplysninger: Prot: Nordstrand husmorskole Extr
Navn:
Nordstrand husmorskole - *(avbildet)*
Sted:
71 Oslo 71 Nordstrand
Samling:
Tilvekstnr: NF.WB 37459* B
Internnr: NBR9205:05194

BILDE 11:



Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: kirke
Utfyllende opplysninger: Prot: Christiania Vor Frelsers Kirke
Navn:
Oslo domkirke 1697 - *(avbildet)*
Sted:
71 Oslo 71 Sentrum Stortorvet 1
Samling:
Tilvekstnr: NF.WL 00065
Internnr: NBR9312:00085

BILDE 12:



Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: kirke
Utfyllende opplysninger: Prot: Christiania Trefoldigheds Kirke
Navn:
Trefoldighetskirken 1858 - *(avbildet)*
Sted:
71 Oslo Sentrum Akersgata
Samling:
Tilvekstnr: NF.WL 00067
Internnr: NBR9312:00088

BILDE 13:**Eier:** Norsk Folkemuseum**Fotograf:** Lindahl, Axel**Datering:** 1880 - 1890**Tittel:****Søkeord:** landskap, fjell, grend, kirke, personer**Utfyllende opplysninger:** Prot: Bakke Kirke ved Nærøfjorden sydfra**Navn:**Bakka kirke 1859 - *(avbildet)***Sted:**

71 Aurland 71 Bakka

71 Aurland 71 Nærøy

Samling:**Tilvekstnr:** NF.WL 00356**Internnr:** NBR9312:00380**BILDE 14:****Eier:** Norsk Folkemuseum**Fotograf:** Lindahl, Axel**Datering:** 1880 - 1890**Tittel:****Søkeord:** landskap, fjord, fjell, bygd, kirke**Utfyllende opplysninger:** Prot: Kvæningstinderne fra Skjervøy**Navn:**Skjervøy kirke 1728 - *(avbildet)***Sted:**

71 Skjervøy 71 Skjervøy

71 Kvæningen 71 Kvæningstindan

Samling:**Tilvekstnr:** NF.WL 01218**Internnr:** NBR9312:01265**BILDE 15:****Eier:** Norsk Folkemuseum**Fotograf:** Lindahl, Axel**Datering:** 1880 - 1890**Tittel:****Søkeord:** landskap, bygd, kirke, prestegård**Utfyllende opplysninger:** Prot: Ringsaker parti ved Kirken #Ekstern kommentar: "Ringsaker prestegård ligger t.v. I forgrunnen er det dampskipsbrygge for Mjøs båter, og Ringsaker handelsforening holdt til husene fra 1870 til ca. 1900."**Navn:**Ringsaker kirke 1240 - *(avbildet)*Ringsaker prestegård - *(avbildet)***Sted:**

71 Ringsaker 71 Ringsaker

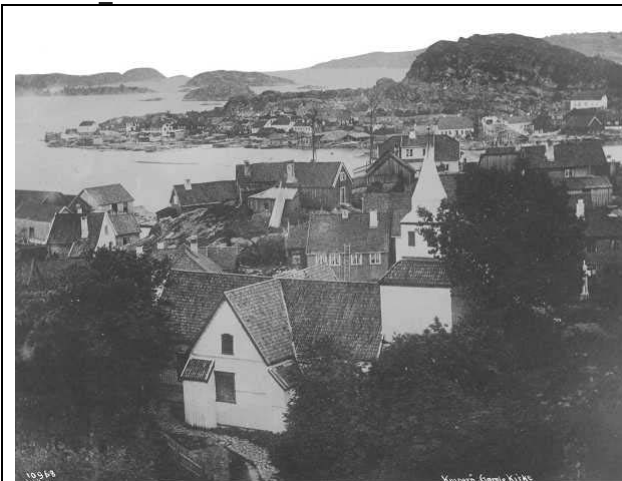
Samling:**Tilvekstnr:** NF.WL 01311**Internnr:** NBR9312:01316

BILDE 16:

Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: tettsted, kirke
Utfyllende opplysninger: Prot: Lillehammer parti af Byen
Navn:
 Hansen, Knut Vigar - (41)
 Lillehammer kirke 1882 - (avbildet)
Sted:
 71 Lillehammer 71 Lillehammer
Samling:
Tilvekstnr: NF.WL 01313
Internnr: NBR9312:01319

BILDE 17:

Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: kirke
Utfyllende opplysninger: Prot: Vaage Kirke
Navn:
 Hansen, Knut Vigar - (41)
 Vågå kirke 1625 - (avbildet)
Sted:
 71 Vågå 71 Vågåmo
Samling:
Tilvekstnr: NF.WL 01356
Internnr: NBR9312:01387

BILDE 18:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1850 - 1869
Søkeord: fotografi: landskap, tettsted, kirke, skjærgård
Utfyllende opplysninger: Prot: Kragerø - Gamle Kirke med Øen 5/12 1909
Navn:
 Kragerø gamle kirke 1652 - 1871 (avbildet)
Sted:
 71 Kragerø 71 Kragerø
 71 Kragerø 71 Kragerøya
Samling:
Tilvekstnr: NF.W 10968
Internnr: NBR9404:07636

BILDE 19:

Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: landskap, fjord, fjell, bygd, kirke
Utfyllende opplysninger: Prot: Lærdalsøren
Navn:
 Hauge kirke 1868 - *(avbildet)*
Sted:
 71 Lærdal 71 Lærdalsøyri
 71 Lærdal 71 Lærdalsfjorden
Samling:
Tilvekstnr: NF.WL 00377* B
Internnr: NBR9312:00407

BILDE 20:

Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: landbruksskole
Utfyllende opplysninger: Prot: Søndfjord
 Mo Landbrugsskole
Navn:
 Hansen, Knut Vigar - *(41)*
 Mo jordbruksskole - *(avbildet)*
Sted:
 71 Førde 71 Mo
Samling:
Tilvekstnr: NF.WL 02514
Internnr: NBR9312:03098

BILDE 21:

Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: landskap, fjord, tettsted
Utfyllende opplysninger: Prot: Drøbak
Navn:
Sted:
 71 Drøbak 71 Drøbak
Samling:
Tilvekstnr: NF.WL 00041
Internnr: NBR9312:00043

BILDE 22:



Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: hovedstad, bydel, landskap, utsikt
Utfyllende opplysninger: Prot: Christiania fra Ekeberg I-IV
Navn:
Sted:
71 Oslo 71 Oslo
71 Oslo 75 Ekeberg
Samling:
Tilvekstnr: NF.WL 00048* E
Internnr: NBR9312:00055

BILDE 23:



Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: landskap, tettsted, vann, kirke
Utfyllende opplysninger: Prot: Voss
Sted:
71 Voss 71 Voss
71 Voss 71 Vangsvatnet
Samling:
Tilvekstnr: NF.WL 00326* A
Internnr: NBR9312:00340

BILDE 24:



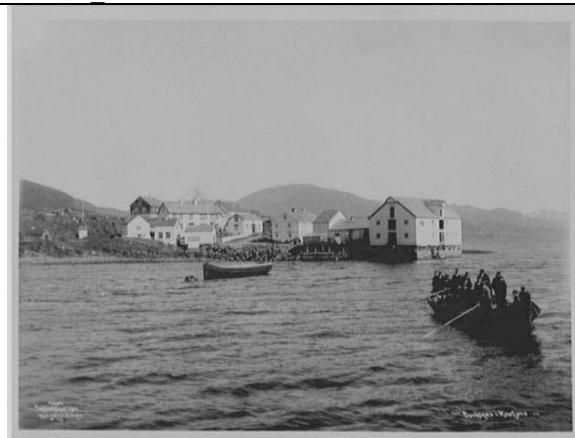
Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: landskap, vann, jenter, restaurant
Utfyllende opplysninger: Prot: Holmenkollen. Peisestuen
Navn:
Peisestuen 1889 - (avbildet)
Sted:
71 Oslo 71 Besserudtjernet
71 Oslo 71 Holmenkollen
Samling:
Tilvekstnr: NF.WL 00093
Internnr: NBR9312:00111

BILDE 25:

Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: fotografi: landskap, dalføre, fjell, elv
Utfyllende opplysninger: Prot: Telemarken
 Gaustadfjeld avspeiling i vandflaten
Navn:
Sted:
 71 Tinn 71 Gaustatoppen
 71 Tinn 71 Vestfjorddalen
 71 Tinn 71 Måna (elv)
Samling:
Tilvekstnr: NF.WL 00171* C
Internnr: NBR9312:00187

BILDE 26:

Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 - 1890
Tittel:
Søkeord: bygd, fjord, landskap
Utfyllende opplysninger: Prot: Parti af Utvik
 med Dampskib
Navn:
Sted:
 71 Stryn 71 Utvik
 71 Stryn 71 Nordfjord
Samling:
Tilvekstnr: NF.WL 00929
Internnr: NBR9312:00960

BILDE 27:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1905 - 1906
Tittel:
Søkeord: landskap, bygd, fjord, robåter, roing,
 personer
Utfyllende opplysninger: Prot: Borkenæs i
 Kvæfjord 1906
Navn:
Sted:
 71 Kvæfjord 71 Borkenes
Samling:
Tilvekstnr: NF.W 04407
Internnr: NBR9404:03613

BILDE 28:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1928
Tittel:
Søkeord: landskap, bygd, fjord
Utfyllende opplysninger: Prot: Bossekop Market - Oversigt Neg: Bossekop. Finmark
Navn:
 Bossekop marked - *(relasjon til bildet)*
Sted:
 71 Alta 71 Bossekop
 71 Alta 71 Altafjorden
Samling:
Tilvekstnr: NF.W 32075
Internnr: NBR9406:11836

BILDE 29:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1930
Tittel:
Søkeord: landskap, fjord, fjell, øy, fiskebåt, hus, robåter
Utfyllende opplysninger: Prot: Kvefjord, Erviken
Navn:
Sted:
 71 Harstad 71 Grytøya
 71 Harstad 71 Ervik
Samling:
Tilvekstnr: NF.W 35841
Internnr: NBR9406:15772

BILDE 30:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1930
Tittel:
Søkeord: landskap, tettsted, båter
Utfyllende opplysninger: Prot: Harstad, panorama
Navn:
Sted:
 71 Harstad 71 Harstad
Samling:
Tilvekstnr: NF.W 35822
Internnr: NBR9406:15751

BILDE 33:



Eier: Norsk Folkemuseum
Fotograf: Lindahl, Axel
Datering: 1880 – 1890
Søkeord: portrett, samer
Utfyllende opplysninger: Prot: Lapper
Navn:
Sted:
72 Finnmark
Samling:
Tilvekstnr: NF.WL 02763* A
Internnr: NBR9312:04012

BILDE 34:



Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1882 - 1883
Tittel:
Søkeord: fotografi: portrett, samer
Utfyllende opplysninger: Prot: Lapper fra Kautokæino
Navn:
Tromholt, Sophus 1851 - 1896, fotograf
(*opphavsperson*)
Gaup, Anna Aslaksdatter - (*avbildet*)
Somby, Anna Johnsdatter - (*avbildet*)
Sted:
71 Kautokeino 71 Kautokeino
Samling:
Tilvekstnr: NF.W 19704
Internnr: NBR9404:14412

BILDE 36:**Eier:** Norsk Folkemuseum**Fotograf:** Wilse, Anders Beer**Datering:** 1902**Tittel:****Søkeord:** portrett, fisker**Utfyllende opplysninger:** Prot: En lun gammel Fisker Bergen 1902 #Ekstern kommentar:

"Mannen på bildet er Elling Hansen (1819-1907). Han var født og bodde på Askøy utenfor Bergen. Bildet skal være tatt på Zakariasbryggen."

Navn:

Hansen, Elling 1819 - 1907 (avbildet)

Sted:

71 Bergen 71 Sentrum Zachariasbryggen

Navn:

Hansen, Elling 1819 - 1907 (avbildet)

Sted:

71 Bergen 71 Sentrum Zachariasbryggen

Samling:**Tilvekstnr:** NF.W 00732**Internnr:** NBR9404:00579**BILDE 38:****Eier:** Norsk Folkemuseum**Fotograf:** Wilse, Anders Beer**Datering:** 1907**Tittel:****Søkeord:** kronprins, portrett, skigåing**Utfyllende opplysninger:** Prot: Kronprinsen paa toppen av bakken 31/1 1907**Navn:**

kronprins Olav 1903 - 1991 (avbildet)

Sted:

72 Oslo 71 Bygdøy

Samling:**Tilvekstnr:** NF.W 06142**Internnr:** NBR9404:04452

BILDE_41:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1926
Tittel:
Søkeord: landskap, bydel, domkirke, oversikt
Utfyllende opplysninger: Prot: Bergensfjords tur – Trondhjem
Navn:
 Bergensfjord (passasjerskip) 1913 - (*relasjon til bildet*)
 Nidarosdomen 1070 - (*avbildet*)
Sted:
 71 Trondheim 71 Trondheim
Samling:
Tilvekstnr: NF.W 28773
Internnr: NBR9406:08476

BILDE_42:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1907
Tittel:
Søkeord: bydel, elv, kirke
Utfyllende opplysninger: Prot: Trondhjem - Elveparti med Ilen Kirke 1907
Navn:
 Ilen kirke 1889 - (*avbildet*)
Sted:
 71 Trondheim 71 Ila Ilevollen
Samling:
Tilvekstnr: NF.W 06315
Internnr: NBR9404:04539

BILDE_43:

Eier: Norsk Folkemuseum
Fotograf: Wilse, Anders Beer
Datering: 1927
Tittel:
Søkeord: fabrikk
Utfyllende opplysninger: Prot: Trøndelagen - Trondhjem, Trolla bruk
Navn:
 Trolla bruk - (*avbildet*)
Sted:
 71 Trondheim 71 Trolla
Samling:
Tilvekstnr: NF.W 31629
Internnr: NBR9406:11492

BILDE_44:



Eier: Norsk Folkemuseum

Fotograf: Lindahl, Axel

Datering: 1880 - 1890

Tittel:

Søkeord: domkirke

Utfyllende opplysninger: Prot: Trondhjem.
Domkirken

Navn:

Hansen, Knut Vigar - (41)

Nidarosdomen 1070 - (avbildet)

Sted:

71 Trondheim 71 Sentrum Bispegata

Samling:

Tilvekstnr: NF.WL 00975* B

Internnr: NBR9312:01013

E. Tabeller for tesaurus

i. Navnetypehierarki

administrative områder

fylker
kommuner
land

bebyggelse

bebyggede områder
byer
hovedsteder
bygninger
bibliotekbygninger
institusjonelle steder
religiøse fasiliteter
utdanningsfailiteter
kommersielle områder
industriområde
jordbruksområder
samferdsel
brygger
gater
jernbanefasiliteter
havneområder

terrengformer

dalfører
fjell

vannkonturer

elver
innsjøer
kystdata

fjorder
øyer

ii. Foretrukne termer USES FOR ikke foretrukne termer

bebyggede områder UF:

bydeler
bygdelag
tettbebyggelser
tettsteder
grender
bygder

bibliotekbygninger UF:

bibliotek
universitetsbibliotek

religiøse fasiliteter UF:

kirker

utdanningsfasiliteter UF:

skoler
husmorskoler
jordbruksskoler
landbruksskoler

industriområde UF:

fabrikker

bygninger UF:

hotell
rådhus
stasjoner
glasstak
fasader
politistasjoner
stasjonsbygninger
losjebygning
klubbhus
teaterbygninger
kontorbygninger
prestegårder
restauranter

BT

NT

jordbruksområder UF:

bruk
gårdsbruk

havneområder UF:

havn

brygger

UF:

kai

gater UF:

samferdsel kjennetegn

UF:	holdeplasser
jernbanefasiliteter UF:	jernbanestasjon
dalfører UF:	daler
terrengformer UF:	halvøyer landskapsområder
innsjøer UF:	vann

iii. Ikke foretrukket term USE foretrukket term

bibliotek	USE	bibliotekbygninger
bruk	USE	jordbruksområder
gårdsbruk	USE	jordbruksområder
bydeler	USE	bebyggede områder
bygdelag	USE	bebyggede områder
bygder	USE	bebyggede områder
daler	USE	dalfører
domkirker	USE	religiøse fasiliteter
fabrikker	USE	industriområder
fasader	USE	bygninger
glasstak	USE	bygninger
grender	USE	bebyggede områder
halvøyer	USE	terrengformer
havn	USE	havneområder samferdsel
holdeplasser	USE	kjennetegn
hotell	USE	bygninger
husmorskoler	USE	utdanningsfasiliteter
jernbanestasjoner	USE	jernbanefasiliteter
jordbruksskoler	USE	utdanningsfasiliteter
kai	USE	brygger
kirker	USE	religiøse fasiliteter
klubbhus	USE	bygninger
kontorbygninger	USE	bygninger
landbruksskoler	USE	utdanningsfasiliteter
landskapsområder	USE	terrengformer
losjebygninger	USE	bygninger
papirfabrikker	USE	industriområder
politistasjoner	USE	bygninger
prestegårder	USE	bygninger
restauranter	USE	bygninger
rådhus	USE	bygninger
skoler	USE	utdanningsfasiliteter
stasjoner	USE	bygninger
stasjonsbygninger	USE	bygninger
teaterbygninger	USE	bygninger
tettbebyggelser	USE	bebyggede områder
tettsteder	USE	bebyggede områder
universitetsbibliotek	USE	bibliotekbygninger
vann	USE	innsjøer

F. SQL resultat Gazetteer

Vert: mysql.stud.ntnu.no

Database : maro_GAZETTEER

Generert den: 22. Apr, 2004 klokka 12:17 PM

Generert av: phpMyAdmin 2.5.3

SQL-spørring: SELECT * FROM `steder` LIMIT 0, 122;

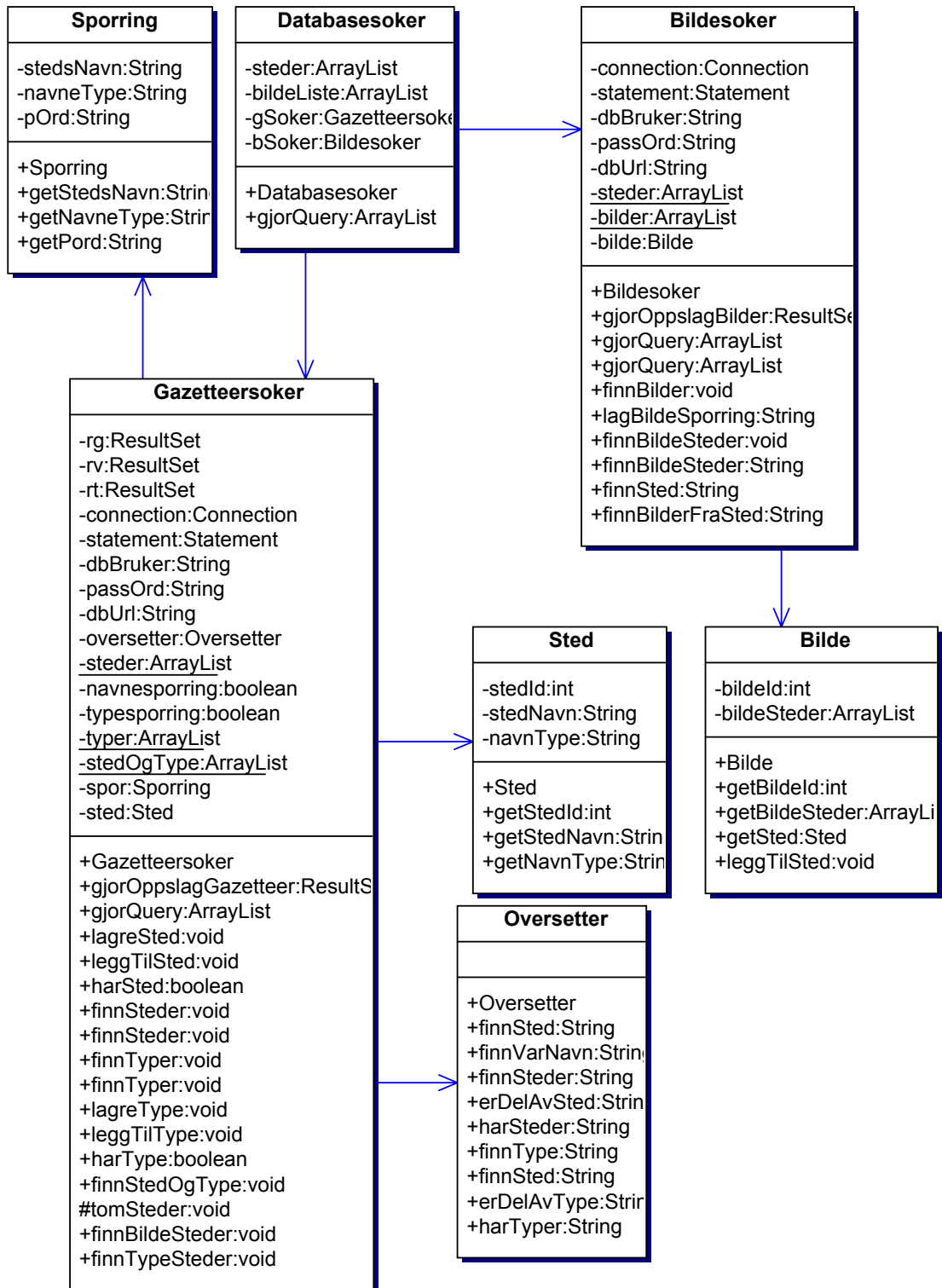
sted_id	stedsnavn	navnetype
1	Norge	land
2	Sør-Trøndelag	fylker
3	Finnmark	fylker
4	Oslo	fylker
5	Sogn og Fjordane	fylker
6	Nordland	fylker
7	Hordaland	fylker
8	Troms	fylker
9	Akershus	fylker
10	Telemark	fylker
11	Oppland	fylker
12	Østfold	fylker
13	Hedmark	fylker
14	Trondheim	kommuner
15	Oslo	kommuner
16	Alta	kommuner
17	Aurland	kommuner
19	Bergen	kommuner
20	Kvæfjord	kommuner
22	Frogn	kommuner
23	Harstad	kommuner
25	Førde	kommuner
26	Tinn	kommuner
28	Lærdal	kommuner
29	Kautokeino	kommuner
30	Kragerø	kommuner
31	Kvænanngen	kommuner
32	Lillehammer	kommuner
34	Stryn	kommuner
35	Notodden	kommuner
36	Ringsaker	kommuner
37	Sarpsborg	kommuner
38	Skjervøy	kommuner
39	Voss	kommuner
40	Vågå	kommuner
41	Bergen	byer
42	Altafjorden	fjorder
43	Bakka	jordbruksområder
45	Borkenes	bebyggede områder
46	Bossekop	bebyggede områder
47	Drøbak	bebyggede områder
48	Ervik	jordbruksområder
50	Gaustatoppen	fjell
51	Grytøya	øyer
53	Harstad	byer

54	Hauge kirke	religiøse fasiliteter
55	Kautokeino	bebyggede områder
56	Kragerø	bebyggede områder
57	Kvæfjorden	fjorder
58	Kvænangen	fjorder
59	Kvængstindan	fjell
60	Lillehammer	byer
61	Lærdalsfjorden	fjorder
62	Lærdalsøyri	bebyggede områder
63	Mo jordbruksskole	utdanningsfasiliteter
64	Måna	elver
65	Nordfjord	fjorder
66	Notodden	byer
68	Nærøyfjorden	fjorder
69	Oslo	hovedsteder
70	Ringsaker kirke	religiøse fasiliteter
71	Ringsaker	bebyggede områder
72	Sarpsborg	byer
73	Skjervøy	bebyggede områder
75	Trondheim	byer
76	Utvik	bebyggede områder
77	Vangsvatnet	innsjøer
78	Vestfjorddalen	dalfører
79	Voss	bebyggede områder
80	Vågå kirke	religiøse fasiliteter
81	Vågåmo	bebyggede områder
82	Øya	øyer
83	Fosenkaia	havneområder
84	Akersgata	gater
85	Sentrum	bebyggede områder
86	Sentrum	bebyggede områder
87	Sentrum	bebyggede områder
88	Rådstuplass	gater
89	Ole Bulls plass	gater
90	Zachariasbryggen	brygger
91	Bygdøy	bebyggede områder
92	Bygdøy	halvøy
93	Drammensveien	gater
94	Biologiveien	gater
96	Ekeberg	bebyggede områder
97	Holmenkollen	bebyggede områder
98	Karl Johans gate	bebyggede områder
99	Kragerø Gamle Kirke	religiøse fasiliteter
100	Lillehammer kirke	religiøse fasiliteter
102	Nationaltheatret	samferdsel
103	Nationaltheatret	bygninger
104	Nordstrand	bebyggede områder
105	Rådhusgata	gater
106	Sandesundsveien	gater
107	Skjervøy kirke	religiøse fasiliteter
108	Stortingsgata	gater
109	Stortorvet	gater
110	Trefoldighetskirken	religiøse fasiliteter
111	Trondheim Jernbanestasjon	jernbanefasiliteter

112	Universitetsbiblioteket	bibliotekbygninger
113	Det Gamle rådhuset	bygninger
114	Hotel Nordstjernen	bygninger
115	Logen	bygninger
116	Nordstrand husmorskole	utdanningsfasiliteter
117	Oslo domkirke	religiøse fasiliteter
118	Festiviteten	bygninger
119	Drøbak Biologiske Stasjon	bygninger
120	Kristiania elektrisitetsverk	bygninger
121	Tinfos papirfabrik	industriområder
122	Peisestuen	bygninger
123	Ila	bebyggede områder
124	Ilevollen	gater
125	Trolla	bebyggede områder
126	Trolla bruk	bygninger
127	Ilen kirke	religiøse fasiliteter
128	Bispegata	gater
129	Nidarosdomen	religiøse fasiliteter
130	Ringsaker prestegård	bygninger
131	Skillebekk	bebyggede områder
133	Besserudtjernet	innsjøer
134	Bakka kirke	religiøse fasiliteter

G. Dokumentasjon av Java-kode

i. Klassediagram



ii. Klasser og metoder

maro.geofinner

Class Sporning

java.lang.Object

|
+--maro.geofinner.Sporring

```
public class Sporring
extends java.lang.Object
```

Class that has methods to receive and access input information from user requests

Constructor Summary

[Sporring](#)(java.lang.String pOrd, java.lang.String stedsNavn, java.lang.String navneType)

Constructor to receive password information and input information from user requests and store this information as an object

Method Summary

java.lang.String	getNavneType () Method to access the future type for the current object
java.lang.String	getPord () Method to access the password
java.lang.String	getStedsNavn () Method to access the place name for the current object

Methods inherited from class java.lang.Object

clone, equals, finalize, getClass, hashCode, notify, notifyAll, toString, wait, wait, wait

Constructor Detail

Sporring

```
public Sporning(java.lang.String pOrd,
                java.lang.String stedsNavn,
                java.lang.String navneType)
```

Constructor to receive password information and input information from user requests and store this information as an object

Parameters:

pOrd - a string representing the password that is required for establishing the database connection

stedsNavn - a string representing the placename requested by the user

navneType - a string representing the feature type requested by the user

Method Detail

getStedsNavn

```
public java.lang.String getStedsNavn()
```

Method to access the place name for the current object

Returns:

a string representing the place name

getNavneType

```
public java.lang.String getNavneType()
```

Method to access the feture type for the current object

Returns:

a string representing the feature type

getPord

```
public java.lang.String getPord()
```

Method to access the password

Returns:

a string representing the password

maro.geofinner

Class Databasesoker

java.lang.Object

|

+-maro.geofinner.Databasesoker

```
public class Databasesoker
extends java.lang.Object
```

Constructor Summary

Databasesoker (java.lang.String s)	
--	--

Method Summary

java.util.ArrayList	gJorQuery (maro.geofinner.Sporring sporrng) Method to control the search.
---------------------	--

Methods inherited from class java.lang.Object

clone, equals, finalize, getClass, hashCode, notify, notifyAll, toString, wait, wait, wait

Constructor Detail

Databasesoker

```
public Databasesoker(java.lang.String s)
```

Method Detail

gJorQuery

```
public java.util.ArrayList gJorQuery(maro.geofinner.Sporring sporrng)
throws java.sql.SQLException
```

Method to control the search. Consulting the gazetteer or not depends on the user request. Invokes methods in the classes Gazetteersoker and Bildesoker.

Parameters:

sporrng - an object of type Sporrng that brings information about the user request together with required password information

Returns:

an ArrayList holding objects of type Bilde. If no objects is retrieved the value of the ArrayList is null.

java.sql.SQLException

maro.geofinner

Class Gazetteersoker

java.lang.Object

|

+-maro.geofinner.Gazetteersoker

public class **Gazetteersoker**

extends java.lang.Object

Constructor Summary

[Gazetteersoker](#)(java.lang.String s)

Constructor to create an object of this class type.

Method Summary

void	finnBildeSteder () Method to retrieve information about all places that may be part of places retrieved after matching placenames and feature types
void	finnSteder (java.sql.ResultSet rg) A recursive method to retrieve data about all places that may be part of one of the places represented in the resultset object returned from the last database query
void	finnSteder (maro.geofinner.Sporring sporing) Method to retrieve places with the requested placename or places for which the requested placename is a variant name
void	finnStedOgType () Method to match placenames to feature types to retrieve the objects of type Sted that is of interest for further retrieval
void	finnTyper (java.sql.ResultSet rt) A recursive method to retrieve all feature types that may be a narrower type of one of the types represented in the resultset object returned from the last database query
void	finnTyper (maro.geofinner.Sporring sporing) Method to retrieve the requested feature type
void	finnTypeSteder () Method to retrieve information about places with certain feature types.
java.sql.ResultSet	gjorOppslagGazetteer (java.lang.String sql) Method to access the gazetteer database.
java.util.ArrayList	gjorQuery (maro.geofinner.Sporring sporing) Method to control the gazetteer search.
boolean	harSted (maro.geofinner.Sted sted)

	Method to check if a list contains a certain object
boolean	harType (java.lang.String typeNavn) Method to check if a list contains a certain string
void	lagreSted (java.sql.ResultSet rs) Method to store information about places in objects of type Sted
void	lagreType (java.sql.ResultSet rt) Method to store information about feature types
void	leggTilSted (maro.geofinner.Sted sted) Method to add an object of type Sted to an ArrayList
void	leggTilType (java.lang.String typeNavn) Method to add a string representing the feature type ArrayList if it's not already present
protected void	tomSteder () Method to remove all objects from the ArrayList steder

Methods inherited from class java.lang.Object

clone, equals, finalize, getClass, hashCode, notify, notifyAll, toString, wait, wait, wait

Constructor Detail

Gazetteersoker

```
public Gazetteersoker(java.lang.String s)
```

Constructor to create an object of this class type.

Parameters:

s - a string representing the required password information for creating the database connection

Method Detail

gjorOppslagGazetteer

```
public java.sql.ResultSet gjorOppslagGazetteer(java.lang.String sql)
```

Method to access the gazetteer database.

Parameters:

sql - a string representing the sql query

Returns:

a ResultSet object with the data retrieved from the database.

gjorQuery

```
public java.util.ArrayList gjorQuery(maro.geofinner.Sporring sporing)
```

Method to control the gazetteer search. Which tables to search depends on user request This method communicates with the class Databasesoker

Parameters:

`sporring` - an object of type `Sporring` that keeps information about the user request

Returns:

an `ArrayList` containing objects of type `Sted` if not null.

lagreSted

```
public void lagreSted(java.sql.ResultSet rs)
    throws java.sql.SQLException
```

Method to store information about places in objects of type `Sted`

Parameters:

`rs` - a `ResultSet` object holding data retrieved from the database
`java.sql.SQLException`

leggTilSted

```
public void leggTilSted(maro.geofinner.Sted sted)
```

Method to add an object of type `Sted` to an `ArrayList`

Parameters:

`sted` - representing the object to be added to the list

harSted

```
public boolean harSted(maro.geofinner.Sted sted)
```

Method to check if a list contains a certain object

Parameters:

`sted` - representing the object to be checked

Returns:

a boolean value, if the list already contains the object the value is true, otherwise it's false

finnSteder

```
public void finnSteder(maro.geofinner.Sporring sporring)
```

Method to retrieve places with the requested placename or places for which the requested placename is a variant name

Parameters:

`sporring` - an object of the type `Sporring` that keeps information about the user request

finnSteder

```
public void finnSteder(java.sql.ResultSet rg)
```

A recursive method to retrieve data about all places that may be part of one of the places represented in the resultset object returned from the last database query

Parameters:

rg - a ResultSet object returned from the last query

finnTyper

```
public void finnTyper(maro.geofinner.Sporring sporrings)
```

Method to retrieve the requested feature type

Parameters:

sporrings - an object of the type Sporrings that keeps information about the user request

finnTyper

```
public void finnTyper(java.sql.ResultSet rt)
```

A recursive method to retrieve all feature types that may be a narrower type of one of the types represented in the resultset object returned from the last database query

Parameters:

rt - a ResultSet object returned from the last query

lagreType

```
public void lagreType(java.sql.ResultSet rt)
    throws java.sql.SQLException
```

Method to store information about feature types

Parameters:

rt - a ResultSet object holding data retrieved from the database
java.sql.SQLException

leggTilType

```
public void leggTilType(java.lang.String typeNavn)
```

Method to add a string representing the feature type ArrayList if it's not already present

Parameters:

typeNavn - representing the string to be added to the list

harType

```
public boolean harType(java.lang.String typeNavn)
```

Method to check if a list contains a certain string

Parameters:

typeNavn - representing the string to be checked

Returns:

a boolean value, if the list already contains the string the value is true, otherwise it's false

finnStedOgType

```
public void finnStedOgType()
```


Method to match placenames to feature types to retrieve the objects of type Sted that is of interest for further retrieval

tomSteder

`protected void tomSteder()`

Method to remove all objects from the ArrayList steder

finnBildeSteder

`public void finnBildeSteder()`

Method to retrieve information about all places that may be part of places retrieved after matching placenames and feature types

finnTypeSteder

`public void finnTypeSteder()`

Method to retrieve information about places with certain feature types. This method invokes only when the user requests a feature type but no placename

maro.geofinner

Class Oversetter

java.lang.Object

|
+--maro.geofinner.Oversetter

```
public class Oversetter  
extends java.lang.Object
```

Class to create sql statements. This is used by the class Gazetteersoker but may be used by the class Bildesoker too The methods takes different parameters but they all returns a string representing a sql statement to be executed

Constructor Summary

Oversetter ()	
-------------------------------	--

Method Summary

java.lang.String	erDelAvSted (int id)
java.lang.String	erDelAvType (int id)
java.lang.String	finnSted (maro.geofinner.Sporring sporing)
java.lang.String	finnSted (java.lang.String type)
java.lang.String	finnSteder (int id)
java.lang.String	finnType (maro.geofinner.Sporring sporing)
java.lang.String	finnVarNavn (maro.geofinner.Sporring sporing)
java.lang.String	harSteder (int id)
java.lang.String	harTyper (int id)

Methods inherited from class java.lang.Object

clone, equals, finalize, getClass, hashCode, notify, notifyAll,

```
toString, wait, wait, wait
```

Constructor Detail

Oversetter

```
public Oversetter()
```

Method Detail

finnSted

```
public java.lang.String finnSted(maro.geofinner.Sporring sporing)
```

finnVarNavn

```
public java.lang.String finnVarNavn(maro.geofinner.Sporring sporing)
```

finnSteder

```
public java.lang.String finnSteder(int id)
```

erDelAvSted

```
public java.lang.String erDelAvSted(int id)
```

harSteder

```
public java.lang.String harSteder(int id)
```

finnType

```
public java.lang.String finnType(maro.geofinner.Sporring sporing)
```

finnSted

```
public java.lang.String finnSted(java.lang.String type)
```

erDelAvType

```
public java.lang.String erDelAvType(int id)
```

harTyper

```
public java.lang.String harTyper(int id)
```

maro.geofinner

Class Sted

java.lang.Object

|
+--maro.geofinner.Sted

```
public class Sted
    extends java.lang.Object
```

Class that instantiates and gives access to information about objects of this class

Constructor Summary

[Sted](#)(int sId, java.lang.String sNavn, java.lang.String nType)

Constructor that instantiates an object of the type of this class The object represents a geographic feature

Method Summary

java.lang.String	getNavnType () Method to give access to the feature type for the current object
int	getStedId () Method to give access to the identifier for the current object
java.lang.String	getStedNavn () Method to give access to the placename for the current object

Methods inherited from class java.lang.Object

clone, equals, finalize, getClass, hashCode, notify, notifyAll, toString, wait, wait, wait

Constructor Detail

Sted

```
public Sted(int sId,
            java.lang.String sNavn,
            java.lang.String nType)
```

Constructor that instantiates an object of the type of this class The object represents a geographic feature

Parameters:

sId - an int representing the identifier for the current feature

sNavn - a string representing the placename for the current feature

nType - a string representing the feature type for the current feature

Method Detail

getStedId

```
public int getStedId()
```

Method to give access to the identifier for the current object

Returns:

an int representing the identifier

getStedNavn

```
public java.lang.String getStedNavn()
```

Method to give access to the placename for the current object

Returns:

a string representing the placeneme

getNavnType

```
public java.lang.String getNavnType()
```

Method to give access to the feature type for the current object

Returns:

a string representing the feature type

maro.geofinner

Class Bildesoker

java.lang.Object

|
+--maro.geofinner.Bildesoker

```
public class Bildesoker
extends java.lang.Object
```

Constructor Summary

Bildesoker (java.lang.String s) Constructor to create an object of this class type.
--

Method Summary

void	finnBilder (java.util.ArrayList steder, maro.geofinner.Sporring sporing) Method to retrieve information about all the collection objects with connection to the places returned from the gazetteer
java.lang.String	finnBilderFraSted (java.util.ArrayList steder) Method to create a sql statement to retrieve collection objects that is connected to places represented by objectys of type Sted in a list
void	finnBildeSteder (maro.geofinner.Bilde bilde) This method is created just to retrieve some metadata for presentation for the user This should be done in another way, may be through communication with the class Databasesoker and further via the class Gazetteersoker
java.lang.String	finnBildeSteder (int id) Method to create an sql statement to retrieve information about places for a spesific collection object
java.lang.String	finnSted (int id) Method to create a sqlstatement to retrieve information about places This method should be unnescassary with another solution for the method public void finnBildeSteder(Bilde bilde){} in this class
java.sql.ResultSet	gJOROppslagBilder (java.lang.String sql) Method to access the collection database.
java.util.ArrayList	gJORQuery (java.util.ArrayList bSteder, maro.geofinner.Sporring sporing) Method to retrieve the collection objects This method is invoked from the class Databasesoker
java.util.ArrayList	gJORQuery (maro.geofinner.Sporring sporing) This method is not fully implemented.

java.lang.String	lagBildeSporring (java.util.ArrayList steder, maro.geofinner.Sporring sporring) Method to create a sql statement for the collection database This should be the method to extend the statement depending on the user request
------------------	---

Methods inherited from class java.lang.Object
clone, equals, finalize, getClass, hashCode, notify, notifyAll, toString, wait, wait, wait

Constructor Detail

Bildesoker

public Bildesoker(java.lang.String s)
 Constructor to create an object of this class type.

Parameters:

s - a string representing the required password information for creating the database connection

Method Detail

gjorOppslagBilder

public java.sql.ResultSet gjorOppslagBilder(java.lang.String sql)
 Method to access the collection database.

Parameters:

sql - a string representing the sql query

Returns:

a ResultSet object with the data retrieved from the database.

gjorQuery

public java.util.ArrayList gjorQuery(java.util.ArrayList bSteder, maro.geofinner.Sporring sporring)

Method to retrieve the collection objects This method is invoked from the class Databasesoker

Parameters:

bSteder - an ArrayList containing objects of type Sted

sporring - an object of type Sporring that keeps information about the user request

Returns:

bilder an ArrayList containing objects of type Bilde

gjomQuery

```
public java.util.ArrayList gjomQuery(maro.geofinner.Sporring sporrng)
```

This method is not fully implemented. This method should be the one to invoke from the class Databasesoker in cases that the user request doesn't require the gazetteer to be consulted

finnBilder

```
public void finnBilder(java.util.ArrayList steder,  
                      maro.geofinner.Sporring sporrng)
```

Method to retrieve information about all the collection objects with connection to the places returned from the gazetteer

Parameters:

- `steder` - a list of objects of type Sted
- `sporrng` - an object of type Sporrng that keeps information about the user request

lagBildeSporring

```
public java.lang.String lagBildeSporring(java.util.ArrayList steder,  
maro.geofinner.Sporring sporrng)
```

Method to create a sql statement for the collection database This should be the method to extend the statement depending on the user request

Parameters:

- `steder` - a list of objects of type Sted
- `sporrng` - an object of type Sporrng that keeps information about the user request

finnBildeSteder

```
public void finnBildeSteder(maro.geofinner.Bilde bilde)
```

This method is created just to retrieve some metadata for presentation for the user This should be done in another way, may be through communication with the class Databasesoker and further via the class Gazetteersoker

Parameters:

- `bilde` - an object representing the collection object to retrieve some information about

finnBildeSteder

```
public java.lang.String finnBildeSteder(int id)
```

Method to create an sql statement to retrieve information about places for a specific collection object

Parameters:

- `id` - an int representing the identifier for the specific object

Returns:

- sql a string representing the sql statement

finnSted

```
public java.lang.String finnSted(int id)
```

Method to create a sqlstatement to retrieve information about places This method should be unnecessary with another solution for the method `public void finnBildeSteder(Bilde bilde){}` in this class

Parameters:

`id` - an int representing an identifier for a place

Returns:

sql a string representing the sql statement

finnBilderFraSted

```
public java.lang.String finnBilderFraSted(java.util.ArrayList steder)
```

Method to create a sql statement to retrieve collection objects that is connected to places represented by objects of type `Sted` in a list

Parameters:

`steder` - an ArrayList containing the objects of type `Sted`

Returns:

sql the string representing the sql statement

maro.geofinner

Class Bilde

java.lang.Object

|
+--maro.geofinner.Bilde

```
public class Bilde  
extends java.lang.Object
```

Class that instantiates and gives access to information about objects of this class

Constructor Summary

[Bilde](#)(int id)

Constructor that instantiates an object of the type of this class

Method Summary

int	getBildeId () Method to access the identifier for the current object
java.util.ArrayList	getBildeSteder () Method to access a list of objects of the type Sted that holds information about places for the current object
maro.geofinner.Sted	getSted (int i) Method to access an object in the list holding objects of type Sted
void	leggTilSted (maro.geofinner.Sted sted) Method that adds an object of type Sted to the list of places for the current object Bilde.

Methods inherited from class java.lang.Object

clone, equals, finalize, getClass, hashCode, notify, notifyAll, toString, wait, wait, wait

Constructor Detail

Bilde

```
public Bilde(int id)
```

Constructor that instantiates an object of the type of this class

Parameters:

`id` - an int that represents the identifier for the current object

Method Detail

getBildeId

```
public int getBildeId()
```

Method to access the identifier for the current object

Returns:

an int that represents the identifier for the current object

getBildeSteder

```
public java.util.ArrayList getBildeSteder()
```

Method to access a list of objects of the type `Sted` that holds information about places for the current object

Returns:

an `ArrayList` that holds objects of type `Sted`

See Also:

`Sted`

getSted

```
public maro.geofinner.Sted getSted(int i)
```

Method to access an object in the list holding objects of type `Sted`

Parameters:

`i` - an int representing a position in the list

Returns:

an object of type `Sted`

leggTilSted

```
public void leggTilSted(maro.geofinner.Sted sted)
```

Method that adds an object of type `Sted` to the list of places for the current object `Bilde`. The object of type `Sted` is added only if it's not already present in the list

Parameters:

`sted` - the object to be added to the list

H. Kildekode

i. Databasescript for gazetteeren

```
CREATE TABLE `steder` (  
  `sted_id` int(11) NOT NULL default '0',  
  `stedsnavn` varchar(255) NOT NULL default "",  
  `navnetyper` varchar(255) NOT NULL default "",  
  PRIMARY KEY (`sted_id`)  
) TYPE=MyISAM COMMENT='navnetyper i hh til tesaurus';
```

```
CREATE TABLE `typer` (  
  `type_id` int(11) NOT NULL default '0',  
  `type_navn` varchar(255) NOT NULL default "",  
  PRIMARY KEY (`type_id`)  
) TYPE=MyISAM COMMENT='navnetyper';
```

```
CREATE TABLE `navne_varianter` (  
  `navn` varchar(100) NOT NULL default "",  
  `sted_id` int(11) NOT NULL default '0'  
) TYPE=MyISAM;
```

```
CREATE TABLE `overordna_typer` (  
  `type_id` int(11) NOT NULL default '0',  
  `overordna_type_id` int(11) NOT NULL default '0',  
  PRIMARY KEY (`type_id`)  
) TYPE=MyISAM;
```

```
CREATE TABLE `relasjoner` (  
  `sted_id` int(11) NOT NULL default '0',  
  `relasjon_til_sted_id` int(11) NOT NULL default '0',  
  `type_id` int(11) NOT NULL default '0',  
  PRIMARY KEY (`sted_id`)  
) TYPE=MyISAM;
```

```
CREATE TABLE `relasjonstyper` (  
  `type_id` int(11) NOT NULL default '0',  
  `relasjonstype` varchar(255) NOT NULL default "",  
  PRIMARY KEY (`type_id`)  
) TYPE=MyISAM;
```

ii. Databasescript for bildemetadata

```
CREATE TABLE `bilde_sted` (  
  `bilde_id` varchar(15) NOT NULL default "",  
  `sted_id` varchar(15) NOT NULL default "",  
  PRIMARY KEY (`bilde_id`,`sted_id`)  
) TYPE=MyISAM;
```

```
CREATE TABLE `sted` (  
  `sted_id` int(15) NOT NULL default '0',  
  `stedsnavn` varchar(50) NOT NULL default "",  
  `navnetype` varchar(50) default NULL,  
  PRIMARY KEY (`sted_id`)  
) TYPE=MyISAM;
```

iii. Script for brukergrensesnitt

form.jsp

```
<!doctype html public "-//W3C//DTD HTML 4.0 //EN">

<% if(request.getParameter("sted")==null&&
request.getParameter("navnetype")==null&&
request.getParameter("fra_aar")==null&&
request.getParameter("til_aar")==null&&
request.getParameter("person")==null
){ %>
<html>
<head>
  <title>Form</title>
</head>
<body>
  <form action="resultat.jsp" method="post" >
  <br><br><br><br><br><table><tr><td>
  <fieldset><legend><b>Fyll inn dine s&oslash;slashkekriterier her: </b></legend><br>

  <table>

  <tr><td><b>Stedsnavn:</b></td><td><input type="text" name="sted"></td></tr>
  <tr><td><b>Navnetype:</b></td><td> <select name = "navnetype" >
  <option value = "null">velg type
  <option value = "alle typer">alle typer
  <option value = "administrative omr&aringder"> . administrative omr&aringder
  <option value = "fylker"> . . fylker
  <option value = "kommuner"> . . kommuner
  <option value = "land"> . . land
  <option value = "bebyggelse"> . bebyggelse
  <option value = "bebyggede omr&aringder"> . . bebyggede omr&aringder
  <option value = "byer"> . . . byer
  <option value = "hovedsteder"> . . . . hovedsteder
  <option value = "bygninger"> . . bygninger
  <option value = "bibliotekbygninger"> . . . bibliotekbygninger
  <option value = "institusjonelle steder"> . . . institusjonelle steder
  <option value = "religi&oslash;se fasiliteter"> . . . . religi&oslash;se fasiliteter
  <option value = "utdanningsfasiliteter"> . . . . utdanningsfasiliteter
  <option value = "kommersielle omr&aringder"> . . . kommersielle omr&aringder
  <option value = "industriomr&aringder"> . . . . industriomr&aringder
  <option value = "jordbruksomr&aringder"> . . jordbruksomr&aringder
  <option value = "samferdsel"> . . samferdsel
  <option value = "brygger"> . . . brygger
  <option value = "gater"> . . . gater
  <option value = "jernbanefasiliteter"> . . . jernbanefasiliteter
  <option value = "terrengformer"> . terrengformer
  <option value = "dalforer"> . . dalf&oslash;shrer
  <option value = "fjell"> . . fjell
```

```

<option value = "vannkonturer"> . vannkonturer
<option value = "elver"> . . elver
<option value = "innsjøer"> . . innsjøer
<option value = "kystdata"> . . kystdata
<option value = "fjorder"> . . . fjorder
<option value = "øyer"> . . . øyer
</select></td>
<td>
Navnetype-<a href="gftesaurus.html"><b>tesaurus</b></a>

</td></tr>
<!--<input type="text" name="navnetype">-->

<tr><td><b>Fra år :</b></td> <td><input type="text" name="fra_aar"></td></tr>

<tr><td><b>Til år :</b> </td> <td><input type="text" name="til_aar"></td></tr>

<tr><td><b>Person navn:</b></td> <td> <input type="text"
name="person"></td></tr>
<tr><td><b>Person rolle:</b> </td> <td> <select name = "rolle" >
<option value = "alle">alle roller
<option value = "fotograf">fotograf
<option value = "eier">eier
<option value = "avbildet">avbildet
</select></td></tr>
<tr><td><b>Emneord:</b> </td> <td><input type="text" name="rolle"></td></tr>

<tr></tr>
<tr><td></td><td></td></tr>
<td><input type="submit" name="startSok" value="Start søk"></td>
<td><input type="reset" value="blank form"></td></tr>
</table>
</fieldset>
</td></tr></table>

</form>
</body>
</html>
<%} %>

<html>
<body>

</body>

```

```

</html>

</body>
</html><!doctype html public "-//W3C//DTD HTML 4.0 //EN">
<html>
<head>
  <title>Title here!</title>
</head>
<body>

</body>
</html>

```

resultat.jsp

```

<!doctype html public "-//W3C//DTD HTML 4.0 //EN">
<%@ page import = "maro.geofinner.*"%>
<%@ page import = "java.lang.*"%>
<%@ page import = "java.util.ArrayList"%>
<%@ page import = "java.io.*"%>

<html>
<head>
  <title>Resultat!</title>
</head>
<body>
</body>
</html>
<% if(request.getParameter("sted")==null&&
request.getParameter("navnetype")==null&&
request.getParameter("fra_aar")==null&&
request.getParameter("til_aar")==null&&
request.getParameter("person")==null
){ %>

<html>
<body>

  <b>Fyll ut s&oslash;keformen på
  <a href = "form.jsp"> s&oslash;kesiden</a></b>
</body>
</html>
<%} else {%>
  <%@ include file = "inc1.jsp"%>

```



```

    <fieldset><legend><b><%= ("Ingen bilder funnet !")%></b></legend><br>
    <br><br>
    <b>Tilbake til
    <a href = "form.jsp"> s&oslash;slashkesiden</a></b><br><br>
    </fieldset>

    <%=}%> <!--else-->

    <%=}%> <!--else-->
</body></html>

```

inc1.jsp

```

<!doctype html public "-//W3C//DTD HTML 4.0 //EN">
<html>
<head>
    <title>incfil1</title>
</head>
<body>
<%=String pso = "passord";%>

</body>
</html>

```

inc2.jsp

```

<!doctype html public "-//W3C//DTD HTML 4.0 //EN">
<html>
<head>
    <title>Inc2</title>
</head>
<body>
<table> <tr><td width="50%">
    <fieldset><legend><b><%= ("S&oslash;slashket kan ikke gjennomf&oslash;slashres
    !")%></b></legend><br>
    <br><br>
    Fyll ut s&oslash;slashkeformen p&aring;ring
    <b>
    <a href = "form.jsp"> s&oslash;slashkesiden</a></b><br><br>
    </fieldset>
</td></tr></table>
</body>
</html>

```

gftesaurus.html

```
<!doctype html public "-//W3C//DTD HTML 4.0 //EN">
<html>
<head>
  <title>Tesaurus</title>
</head>
<body>
```

```
<tr><td><b>Tilbake til
  <a href = "form.jsp"> s&oslash;skesiden</a></b></td></tr><br><br>
```

```
<table>
```

```
<tr><td><h3>Navnetypesaurus</h3></td></tr>
```

```
<tr><td>Dersom du er i tvil om hvilken navnetype du skal velge fra listen i søkebildet
kan du få hjelp fra
```

```
ordlisten under her.<br> Se om du finner ønsket søkeord i kolonnen til venstre.<br> På
samme linje i kolonnen
```

```
til høyre finner du riktig navnetype som du skal velge for ditt søkeord.<br><br>
```

```
Et eksempel:<br>
```

```
Du ønsker å finne bilder av kirker.<br>
```

```
Let i kolonnen til venstre til du finner ordet "kirker".<br>
```

```
På samme linje i kolonnen til høyre ser du at det står "religiøse fasiliteter".<br>
```

```
Tilbake på søkesiden velger du "religiøse fasiliteter" fra lista over navnetyper.
```

```
</td></tr>
```

```
</table><br><br>
```

```
<table>
```

```
<tr><td>bibliotek&nbsp;</td><td>BRUK&nbsp;</td><td>&nbsp;&nbsp;</td><td>bib
liotekbygninger&nbsp;</td></tr>
```

```
<tr><td>bruk&nbsp;</td><td>.....&nbsp;</td><td></td><td>jordbruksområder&nbsp;
sp;</td></tr>
```

```
<tr><td>bydeler &nbsp;</td><td>.....&nbsp;</td><td></td><td>bebyggede
områder&nbsp;</td></tr>
```

```
<tr><td>bygdelag&nbsp;</td><td>.....&nbsp;</td><td></td><td>bebyggede
områder&nbsp;</td></tr>
```

```
<tr><td>bygder&nbsp;</td><td>.....&nbsp;</td><td></td><td>bebyggede
områder&nbsp;</td></tr>
```

```
<tr><td>daler&nbsp;</td><td>.....&nbsp;</td><td></td><td>dalfører&nbsp;</td><
/tr>
```

```
<tr><td>domkirker&nbsp;</td><td>.....&nbsp;</td><td></td><td>religiøse
fasiliteter&nbsp;</td></tr>
```

```
<tr><td>fabrikker&nbsp;</td><td>.....&nbsp;</td><td></td><td>industriområder&
nbsp;</td></tr>
```

```
<tr><td>fasader&nbsp;</td><td>.....&nbsp;</td><td></td><td>bygninger&nbsp;</t
d></tr>
```

<tr><td>glasstak </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>grender </td><td>..... </td><td></td><td>bebyggede områder </td></tr>
 <tr><td>gårdsbruk </td><td>..... </td><td></td><td>jordbruksområder </td></tr>
 <tr><td>halvøyer </td><td>..... </td><td></td><td>terrengformer </td></tr>
 <tr><td>havn </td><td>..... </td><td></td><td>havneområder </td></tr>
 <tr><td>holdeplasser </td><td>..... </td><td></td><td>samferdsel kjennetegn </td></tr>
 <tr><td>hotell </td><td>..... </td><td></td><td>bygninger </td>></tr>
 <tr><td>husmorskoler </td><td>..... </td><td></td><td>utdanningsfasiliteter </td></tr>
 <tr><td>jernbanestasjoner </td><td>..... </td><td></td><td>jernbanefasiliteter </td></tr>
 <tr><td>jordbruksskoler </td><td>..... </td><td></td><td>utdanningsfasiliteter </td></tr>
 <tr><td>kai </td><td>..... </td><td></td><td>brygger </td></tr>
 <tr><td>kirker </td><td>..... </td><td></td><td>religiøse fasiliteter </td></tr>
 <tr><td>klubbhus </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>kontorbygninger </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>landbruksskoler </td><td>..... </td><td></td><td>utdanningsfasiliteter </td></tr>
 <tr><td>landskapsområder </td><td>..... </td><td></td><td>terrengformer </td></tr>
 <tr><td>losjebygninger </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>papirfabrikker </td><td>..... </td><td></td><td>industriområder </td></tr>
 <tr><td>politistasjoner </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>prestegårder </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>restauranter </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>rådhus </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>skoler </td><td>..... </td><td></td><td>utdanningsfasiliteter </td></tr>
 <tr><td>stasjoner </td><td>..... </td><td></td><td>bygninger </td></tr>
 <tr><td>stasjonsbygninger </td><td>..... </td><td></td><td>bygninger </td></tr>

```
<tr><td>teaterbygninger&nbsp;</td><td>.....&nbsp;</td><td></td><td>bygninger&nbsp;</td></tr>
<tr><td>tettbebyggelser&nbsp;</td><td>.....&nbsp;</td><td></td><td>bebyggede
områder&nbsp;</td></tr>
<tr><td>tettsteder&nbsp;</td><td>.....&nbsp;</td><td></td><td>bebyggede
områder&nbsp;</td></tr>
<tr><td>universitetsbibliotek&nbsp;</td><td>.....&nbsp;</td><td></td><td>bibliot
ekbygninger&nbsp;</td></tr>
<tr><td>vann&nbsp;</td><td>.....&nbsp;</td><td></td><td>innsjøer&nbsp;</td><
/tr>
<tr><td></td></tr>
<tr><td></td></tr>
<tr><td><b>Tilbake til
  <a href = "form.jsp"> s&oslash;kesiden</a></b></td></tr>
```

```
</table>
</body>
```

```
</html>
```

iv. Javakode

Klasse Sporning

```
package maro.geofinner;

/**Class that has methods to recieve and access input information from user requests */
public class Sporning {

    private String stedsNavn;
    private String navneType;
    private String pOrd;

    /**
    *Constructor to recieve password information and input information from user requests
    and store this information as an object
    *@param pOrd a string representing the password that is required for establishing the
    database connection
    *@param stedsNavn a string representing the placename requested by the user
    *@param navneType a string representing the feature type requested by the user
    */
    public Sporning(String pOrd, String stedsNavn, String navneType ){
        this.stedsNavn = stedsNavn;
        this.navneType = navneType;
        this.pOrd = pOrd;
    } //Sporring

    /**
    *Method to access the place name for the current object
    *@return a string representing the place name
    */
    public String getStedsNavn(){
        return stedsNavn;
    }

    /**
    *Method to access the feture type for the current object
    *@return a string representing the feature type
    */
    public String getNavneType(){
        return navneType;
    }

    /**
    *Method to access the password
    *@return a string representing the password
    */
    public String getPord(){
        return pOrd;
    }
}
```

```

    }
} //Sporning

```

Klasse Databasesoker

```

package maro.geofinner;

import java.sql.SQLException;
import java.util.ArrayList;
public class Databasesoker {

    /**
    *Class with only one method to connect and invoke methods in the gazetteer part and
    the collection part of this system
    *This class communicates with the file resultat.jsp that is part of the user interface.
    *@Author Marit Olsen maro@idi.ntnu.no
    *@versjon 1.0
    */
    private ArrayList steder;
    private ArrayList bildeListe;
    private Gazetteersoker gSoker;
    private Bildesoker bSoker;

    public Databasesoker(String s){
        this.steder = new ArrayList();
        this.gSoker = new Gazetteersoker(s);
        this.bSoker = new Bildesoker(s);
        this.bildeListe = new ArrayList();
    }

    /**
    *Method to control the search. Consulting the gazetteer or not depends on the user
    request.
    *Invokes methods in the classes Gazetteersoker and Bildesoker.
    *@param sporning an object of type Sporning that brings information about the user
    request together with
    *required password information
    *@return an ArrayList holding objects of type Bilde. If no objects is retrieved the value
    of the ArrayList is null.
    */
    public ArrayList gjorQuery(Sporring sporning)throws SQLException {

        if ((sporning.getStedsNavn() != null) || (sporning.getNavneType() !=
null)){
            steder = gSoker.gjorQuery(sporning);
            System.out.println("DATABASESOKER SIN STEDER HAR
ANTALL "+steder.size()+"");

```

```

        if (steder.isEmpty()){
            return bildeListe;
        }
        else{
            bildeListe = bSoker.gjorQuery(steder, sporing);
            System.out.println("DATABASESOKER SIN BILDELISTE
HAR ANTALL : '"+bildeListe.size()+"");
            //skrivUtBildeId(bildeListe); for testing
            }//else if
        }//if
        else{
            bildeListe = bSoker.gjorQuery(sporing);
            System.out.println("DATABASESOKER SIN BILDELISTE
HAR ANTALL : '"+bildeListe.size()+"");
            //skrivUtBildeId(bildeListe); to write to the command window for
testing the program
        }

        return bildeListe;
    }//gjorQuery

}//Databasesoker

```

Klasse Gazetteersoker

```
package maro.geofinner;
```

```
import java.util.ArrayList;
import java.io.*;
import java.lang.Object;
import java.sql.*;
```

```
/**
```

```
*This class provides methods to create connection to and search the gazetteer database.
```

```
*@Author Marit Olsen maro@idi.ntnu.no
```

```
*@versjon 1.0
```

```
*/
```

```
public class Gazetteersoker{
```

```
    private ResultSet rg = null;
    private ResultSet rv = null;
    private ResultSet rt = null;
```



```

private Connection connection;
private Statement statement;
private String dbBruker = "maro_admin";
private String passOrd = "";
private String dbUrl = "jdbc:mysql://mysql.stud.ntnu.no/maro_GAZETTEER";
private Oversetter oversetter;
//private String sql = "select";
private static ArrayList steder;
private boolean navnesporring = false;
private boolean typesporring = false;
private static ArrayList typer;
private static ArrayList stedOgType;
private Sporrying spor;

/**
 *Constructor to create an object of this class type.
 *@param s a string representing the required password information for creating the
database connection
 */
public Gazetteersoker(String s){
    oversetter = new Oversetter();
    steder = new ArrayList();
    typer = new ArrayList();
    stedOgType = new ArrayList();
    this.passOrd = s;
}

/**
 *Method to access the gazetteer database.
 *@param sql a string representing the sql query
 *@return a ResultSet object with the data retrieved from the database.
 */
public ResultSet gjorOppslagGazetteer(String sql){
    ResultSet rs = null;
    try{
        Class.forName("org.gjt.mm.mysql.Driver").newInstance();
        connection =
DriverManager.getConnection(dbUrl,dbBruker,passOrd);
        statement = connection.createStatement();
        rs = statement.executeQuery(sql);
        //connection.close;
    }//try
    catch (Exception e){
        e.printStackTrace();
    }//catch
    return rs;
}

```

```

/**
 *Method to control the gazetteer search. Which tables to search depends on user request
 *This method communicates with the class Databasesoker
 *@param sporing an object of type Sporing that keeps information about the user
 request
 *@return an ArrayList containing objects of type Sted if not null.
 */
    public ArrayList gjorQuery(Sporring sporing){
        //steder = new ArrayList();
        if (sporing.getStedsNavn() != "null"){
            navnesporring = true;
            System.out.println("navnesporring er sant!");
            finnSteder(sporring);
            System.out.println("ANTALL STEDER ER "+steder.size()+"");

            if (sporing.getNavneType() != "null"){
                typesporring = true;
                System.out.println("typesporring er sant!");
                finnTyper(sporring);
                finnStedOgType();
                tomSteder();
                steder = stedOgType;
                finnBildeSteder();
            }
            else {
                return steder;
            }
        }//if
        else {
            finnTyper(sporring);
            finnTypeSteder();
        }

        System.out.println("ANTALL BILDESTEDER ER
"+steder.size()+"");
        //return steder;
        return steder;
    }//gjorQuery

/**
 *Method to store information about places in objects of type Sted
 *@param rs a ResultSet object holding data retrieved from the database
 */
    public void lagreSted(ResultSet rs)throws SQLException{
        System.out.println("er inne i lagre sted");
        rs.beforeFirst();
        while (rs.next()){
            Sted sted = new Sted(rs.getInt(1), rs.getString(2), rs.getString(3));
            System.out.println("nytt sted er "+(rs.getString(2)));
            leggTilSted(sted);
        }
    }

```

```

        }//while
    }//lagreSted

/**
 *Method to add an object of type Sted to an ArrayList
 *@param sted representing the object to be added to the list
 */
    public void leggTilSted(Sted sted){
        if (!harSted(sted)){
            steder.add(sted);

            System.out.println("Dette er lagt til i stedlista: ");
            System.out.println("stedId: "+(sted.getStedId()));
            System.out.println("stedsnavn: "+(sted.getStedNavn()));
            System.out.println("navnetype: "+(sted.getNavnType()));
            //System.out.println("erDelAvId: "+(sted.getErDelAvId()));
            System.out.println("Steder har antall  " + steder.size());
        }
    }

/**
 *Method to check if a list contains a certain object
 *@param sted representing the object to be checked
 *@return a boolean value, if the list already contains the object the value is true,
 otherwise it's false
 */
    public boolean harSted(Sted sted){
        boolean retur = false;
        for (int i = 0; i < steder.size(); i++){

            if ((sted.getStedId()) == (((Sted)steder.get(i)).getStedId()))
            {
                retur = true;
                System.out.println("steder har sted fra før");
                return retur;
            }
        }
        System.out.println("steder har ikke sted fra før");
        return retur;
    }//harSted

/**
 *Method to retrieve places with the requested placename or places for which the
 requested placename is a variant name
 *@param spurring an object of the type Spurring that keeps information about the user
 request

```

```

*/
    public void finnSteder(Sporring sporing){
        try{
            rg = gjorOppslagGazetteer(oversetter.finnSted(sporring));
            System.out.println("Spørring:
"+oversetter.finnSted(sporring)+"");
            //System.out.println("rg antall rader: "+rg.getFetchSize()+"");
            if (rg.next()){
                System.out.println("rg er forskjellig fra null");
                rg.beforeFirst();
                //lagreSted(rg);
                //System.out.println("går til finnSteder(rg)");
                finnSteder(rg);
            }
        }//try
        catch (Exception e){
            System.out.println("Koblingsfeil!!");
            e.printStackTrace();
        }//catch
        try{
            rv = gjorOppslagGazetteer(oversetter.finnVarNavn(sporring));
            if (rv.next()){
                rv.beforeFirst();
                while (rv.next()){
                    rg =
gjorOppslagGazetteer(oversetter.finnSteder(rv.getInt(2)));
                    lagreSted(rg);
                    finnSteder(rg);
                }//while
            }//if
        }//try

        catch (Exception e){
            System.out.println("Koblingsfeil!!");
            e.printStackTrace();
        }//catch
    }//finnSteder

```

/**

*A recursive method to retrieve data about all places that may be part of one of the places represented in the resultset object

*returned from the last database query

*@param rg a ResultSet object returned from the last query

*/

```

    public void finnSteder(ResultSet rg){
        System.out.println("er inne i finnSteder(rg)");
        ResultSet nyttRg;
        int id;
        try{

```

```

        lagreSted(rg);
        rg.beforeFirst();
        while (rg.next()){
            id = rg.getInt(1);
            System.out.println("id er "+id+"");
            //lagreSted(rg);
            //rg.beforeFirst();
            nyttRg = gjørOppslagGazetteer(oversetter.harSteder(id));
            System.out.println(oversetter.harSteder(id));
            finnSteder(nyttRg);
        }//while
    }//try
    catch (Exception e){
        System.out.println("Koblingsfeil!");
        e.printStackTrace();
    }//catch

} //finnSteder

```

```

/**
 *Method to retrieve the requested feature type
 *@param sporing an object of the type Sporing that keeps information about the user
 request
 */

```

```

public void finnTyper(Sporring sporing){
    System.out.println("er inne i finnTyper(sporring)");
    try{
        rt = gjørOppslagGazetteer(oversetter.finnType(sporring));
        if (rt.next()){
            System.out.println("rt er forskjellig fra null");
            rt.beforeFirst();
            //lagreType(rt);
            System.out.println("går til finnTyper(rg)");
            finnTyper(rt);
        }
    }
    catch (Exception e){
        System.out.println("Koblingsfeil!");
        e.printStackTrace();
    }
}

```

```

/**
 *A recursive method to retrieve all feture types that may be a narrowe type of one of the
 types
 *represented in the resultset object returned from the last database query

```

```

*@param rt a ResultSet object returned from the last query
*/
public void finnTyper(ResultSet rt){
    System.out.println("er inne i finnTyper(rt)");
    ResultSet nyttRt;
    int id;
    try{
        lagreType(rt);
        rt.beforeFirst();

        while (rt.next()){
            id = rt.getInt(1);
            System.out.println("id er '"+id+"'");
            //lagreType(rt);
            nyttRt = gjorOppslagGazetteer(oversetter.harTyper(id));
            System.out.println(oversetter.harTyper(id));
            finnTyper(nyttRt);
        }//while
    }//try
    catch (Exception e){
        System.out.println("Koblingsfeil!");
        e.printStackTrace();
    }//catch
}//finnTyper

```

```

/**
*Method to store information about feature types
*@param rt a ResultSet object holding data retrieved from the database
*/

```

```

public void lagreType(ResultSet rt)throws SQLException{
    System.out.println("er inne i lagreType(rt)");
    String type;
    rt.beforeFirst();
    while (rt.next()){
        type = rt.getString(2);
        leggTilType(type);
    }//while
}//lagreType

```

```

/**
*Method to add a string representing the feature type ArrayList if it's not already
present
*@param typeNavn representing the string to be added to the list
*/

```

```

public void leggTilType(String typeNavn){
    System.out.println("er inne i leggTilType");
    if (!harType(typeNavn)){

```

```

        typer.add(typeNavn);
        System.out.println("Navnetype: "+typeNavn+ " er lagt til
typelista");
    }
    System.out.println("Typer har antall  " + typer.size());
}

} //leggTilType

/**
 *Method to check if a list contains a certain string
 *@param typeNavn representing the string to be checked
 *@return a boolean value, if the list already contains the string the value is true,
 otherwise it's false
 */
public boolean harType(String typeNavn){
    System.out.println("er inne i harType");
    boolean ligner = false;
    for (int i=0; i<typer.size(); i++ ){
        if (typeNavn.equalsIgnoreCase((String)typer.get(i))){
            ligner = true;
            System.out.println("typer har type fra før");
            return ligner;
        } //if
    } //for
    System.out.println("typer har ikke type fra før");
    return ligner;
}

/**
 *Method to match placenames to feature types to retrieve the objects of type Sted
 *that is of interest for further retrieval
 */
public void finnStedOgType(){
    System.out.println("er inne i finnStedOgType()");
    for (int i=0; i<steder.size() ;i++ ){
        Sted sted = (Sted) steder.get(i);
        for (int j=0; j<typer.size() ;j++ ){
            if
(sted.getNavnType().equalsIgnoreCase((String)typer.get(j))){
                stedOgType.add(sted);
                System.out.println("Dette er lagt til i stedOgType:
");
                System.out.println("stedId: "+(sted.getStedId()));
                System.out.println("stedsnavn:
"+(sted.getStedNavn()));
                System.out.println("navnetype:
"+(sted.getNavnType()));
            }
        }
    }
}

```

```

//System.out.println("erDelAvId:
"+(sted.getErDelAvId()));
stedOgType.size());
        }//if
    }
} //finnStedOgType

/**
 *Method to remove all objects from the ArrayList steder
 */
protected void tomSteder(){
    for (int i = 0; i < steder.size();){
        steder.remove(i);
        System.out.println("Steder har antall " + steder.size());
    }//for
} //tomSteder

/**
 *Method to retrieve information about all places that may be part og places retrieved
 after matching
 *placenames and feature types
 */
public void finnBildeSteder(){
    for (int i = 0; i < stedOgType.size(); i++){
        rg = gjorOppslagGazetteer(oversetter.harSteder(((Sted)
stedOgType.get(i)).getStedId()));
        finnSteder(rg);
    }
}

/**
 *Method to retrieve information about places with certain feature types.
 *This method invokes only when te user requests a feature type but no placename
 */
public void finnTypeSteder(){
    for (int i = 0; i < typer.size(); i++){
        rg =
gjorOppslagGazetteer(oversetter.finnSted((String)typer.get(i)));
        finnSteder(rg);
    }
}

} //Gazetteersoker

```


Klasse Oversetter

```
package maro.geofinner;

import java.util.ArrayList;

/**
 *Class to create sql statements. This is used by the class Gazetteersoker but may be
used by the class Bildesoker too
 *The methods takes different parameters but they all returns a string representing a sql
statement to be executed
 *@Author Marit Olsen maro@idi.ntnu.no
 *@versjon 1.0
 */

public class Oversetter {

    public Oversetter() {
    }

    public String finnSted(Sporring sporing) {
        String sql = "select * from steder where stedsnavn =
"+sporing.getStedsNavn()+"";
        return sql;
    }

    public String finnVarNavn(Sporring sporing) {
        String sql = "select * from navne_varianter where navn =
"+sporing.getStedsNavn()+"";
        return sql;
    }

    public String finnSteder(int id) {
        String sql = "select * from steder where sted_id = "+id+"";
        return sql;
    }

    public String erDelAvSted(int id) {
        String sql = "SELECT * FROM steder WHERE sted_id = "+id+"";
        return sql;
    }

    public String harSteder(int id) {
```

```

        String sql = "select steder.* from steder, relasjoner, relasjonstyper where
(steder.sted_id = relasjoner.sted_id) and (relasjon_til_sted_id = '"+id+"')and
(relasjonstype = 'er_del_av' ) and (relasjoner.type_id = relasjonstyper.type_id)";
        return sql;
        //finner underordna steder
    }

    public String finnType(Sporring sporing){
        String sql = "select * from typer where type_navn =
 '"+sporing.getNavneType()+"'";
        return sql;
    }

    public String finnSted(String type){
        String sql = "select * from steder where navnetype = '"+type+"'";
        return sql;
    }

    public String erDelAvType(int id){
        String sql = "SELECT * FROM typer WHERE type_id = '"+id+"'";
        return sql;
        //finner overordna type
    }

    public String harTyper(int id){
        String sql = "select typer.* from typer, overordna_typer where
(overordna_type_id = '"+id+"') and (typer.type_id = overordna_typer.type_id)";
        return sql;
        //finner underordna typer
    }
}
} //Oversetter

```

Klasse Sted

```
package maro.geofinner;
```

```
/**
```

```
*Class that instantiates and gives access to information about objects of this class
```

```
*@author Marit Olsen
```

```
*version 1.0
```

```
*/
```

```
public class Sted{
```

```
    private int stedId;
```

```
    private String stedNavn;
```

```

        private String navnType;

/**
 *Constructor that instantiates an object of the type of this class
 *The object represents a geographic feature
 *@param sId an int representing the identifier for the current feature
 *@param sNavn a string representing the placename for the current feature
 *@param nType a string representing the feature type for the current feature
 */
    public Sted(int sId, String sNavn, String nType){

        this.stedId = sId;
        this.stedNavn = sNavn;
        this.navnType = nType;
    }

/**Method to give access to the identifier for the current object
 @return an int representing the identifier
 */
    public int getStedId(){
        return stedId;
    }

/**Method to give access to the placename for the current object
 @return a string representing the placeneme
 */
    public String getStedNavn(){
        return stedNavn;
    }

/**Method to give access to the feature type for the current object
 @return a string representing the feature type
 */
    public String getNavnType(){
        return navnType;
    }

} //Sted

```

Klasse Bildesoker

```

package maro.geofinner;

import java.sql.*;
import java.util.ArrayList;
import java.lang.Object;

```

```

public class Bildesoker{

/**
 *This class provides methods to create connection to and search the collection database.
 *@Author Marit Olsen maro@idi.ntnu.no
 *@versjon 1.0
 */

    //private ResultSet resultatMeta = null;
    private Connection connection;
    private Statement statement;
    private String dbBruker = "maro_admin";
    private String passOrd = "";
    private String dbUrl = "jdbc:mysql://mysql.stud.ntnu.no/maro_gn_metadata";
    private static ArrayList steder;
    private static ArrayList bilder;
    //private static String sql = "null";
    //private static String kolonner = "null";
    //private static String tabeller = "null";

/**
 *Constructor to create an object of this class type.
 *@param s a string representing the required password information for creating the
 database connection
 */

    public Bildesoker(String s){
        steder = new ArrayList();
        bilder = new ArrayList();
        this.passOrd = s;
    }

/**
 *Method to access the collection database.
 *@param sql a string representing the sql query
 *@return a ResultSet object with the data retrieved from the database.
 */

    public ResultSet gjorOppslagBilder(String sql){
        ResultSet rs = null;
        try{
            Class.forName("org.gjt.mm.mysql.Driver").newInstance();
            connection =
DriverManager.getConnection(dbUrl,dbBruker,passOrd);
            statement = connection.createStatement();
            rs = statement.executeQuery(sql);
            //connection.close;
        }//try

```

```

        catch (Exception e){
            e.printStackTrace();
        }//catch
        return rs;
    }//gjorOppslagBilder

/**
 *Method to retrieve the collection objects
 *This method is invoked from the class Databasesoker
 *@param bSteder an ArrayList containing objects of type Sted
 *@param sporing an object of type Sporing that keeps information about the user
 request
 *@return bilder an ArrayList containing objects of type Bilde
 */
    public ArrayList gjorQuery(ArrayList bSteder, Sporing sporing){
        steder = bSteder;
        System.out.println("BILDESOKER SIN STEDER HAR ANTALL :
""+steder.size()+"");
        finnBilder(steder, sporing);
        return bilder;
    }

/**
 *This method is not fully implemented. This method should be the one to invoke from
 the class Databasesoker
 *in cases that the user request doesn't require the gazetteer to be consulted
 */
    /*
    public ArrayList gjorQuery(Sporing sporing){

        return bilder;
    }
    */

/**
 *Method to retrieve information about all the collection objects with
 *connection to the places returned from the gazetteer
 *@param steder a list of objects of type Sted
 *@param sporing an object of type Sporing that keeps information about the user
 request
 */
    public void finnBilder(ArrayList steder, Sporing sporing){
        Bilde bilde = null;
        try{
            ResultSet rb = gjorOppslagBilder(lagBildeSporing(steder,
sporing));

            //finner alle bilder fra alle steder i lista steder
            if (rb.next()){
                rb.beforeFirst();
            }
        }
    }

```

```

        while (rb.next()){
            //Integer      nyInt = new Integer(rb.getInt(1));
            //bilder.add(nyInt);
            bilde = new Bilde(rb.getInt(1));
            //finnBildeSteder(bilde, rb.getInt(1));
            finnBildeSteder(bilde);
            bilder.add(bilde);
            System.out.println("har lagt til bilde");
        }
    }
    catch (Exception e){
        System.out.println("Koblingsfeil!!");
        e.printStackTrace();
    }
}

/**
 *Method to create a sql statement for the collection database
 *This should be the method to extend the statement depending on the user request
 *@param steder a list of objects of type Sted
 *@param sporing an object of type Sporing that keeps information about the user
 request
 */
    public String lagBildeSporring(ArrayList steder, Sporing sporing){
        String sql = finnBilderFraSted(steder);
        return sql;
    }

/**
 *This method is created just to retrieve some metadata for presentation for the user
 *This should be done in another way, may be through communication with the class
 Databasesoker and further via the
 *class Gazetteersoker
 *@param bilde an object representing the collection object to retrieve some information
 about
 */
    public void finnBildeSteder(Bilde bilde){
        try{
            //ResultSet rb = gjorOppslagBilder(finnBildeSteder(id));
            ResultSet rb =
            gjorOppslagBilder(finnBildeSteder(bilde.getBildeId()));
            //finner alle steder for ett bilde
            if (rb.next()){
                rb.beforeFirst();
                while (rb.next()){
                    ResultSet rs =
                    gjorOppslagBilder(finnSted(rb.getInt(2)));
                    if (rs.next()){

```



```

        for (int i = 0; i < steder.size(); i++){
            if (sql == "null"){
                sql = "select distinct bilde_id from b_sted where (sted_id
= ""+(Sted)steder.get(i).getStedId()+""");
            }
            else {
                sql = sql + "|| (sted_id =
""+(Sted)steder.get(i).getStedId()+""");
            }
        }
        System.out.println(sql);
        return sql;
    }

```

```

/**
 *Method not implemented but this should be the method to create sql statement in case
 the gazetteer is not consulted
 */
/**
public String finnBilder(){}
*/

} //Bildesoker

```

Klasse Bilde

```

package maro.geofinner;

import java.util.ArrayList;

/**
 *Class that instantiates and gives access to information about objects of this class
 *@author Marit Olsen
 *version 1.0
 */
public class Bilde{

    private int bildeId;
    private ArrayList bildeSteder;

    /**
 *Constructor that instantiates an object of the type of this class
 *@param id an int that represents the identifier for the current object
 */
    public Bilde(int id){

```



```

        this.bildeId = id;
        this.bildeSteder = new ArrayList();

    }

    /**
    *Method to access the identifier for the current object
    *@return an int that represents the identifier for the current object
    */
    public int getBildeId(){
        return bildeId;
    }

    /**
    *Method to access a list of objects of the type Sted that holds information about places
    for the current object
    *@return an ArrayList that holds objects of type Sted
    *@see class Sted
    */
    public ArrayList getBildeSteder(){
        return bildeSteder;
    }

    /**
    *Method to access an object in the list holding objects of type Sted
    *@param i an int representing a position in the list
    *@return an object of type Sted
    *@calls the get(int index) method inherited from class ArrayList
    */
    public Sted getSted(int i){
        return (Sted)bildeSteder.get(i);
    }

    /**
    *Method that adds an object of type Sted to the list of places for the current object
    Bilde.
    *The object of type Sted is added only if it's not already present in the list
    @param sted the object to be addad to the list
    */
    public void leggTilSted(Sted sted){
        if (!bildeSteder.contains(sted)){
            bildeSteder.add(sted);
        }
    }

} //Bilde

```

Klasse SokeProgram

```
/**
 *This class was written just for testing during programming session
 */

import java.sql.SQLException;
import java.util.ArrayList;

public class SokeProgram{

private ArrayList subSamling;
public SokeProgram(){
}

public static void main(String[] args){

    try{
        Sporning sporning = new Sporning("null",/*0,0,*/"fjorder");
        Databasesoker dbSoker = new Databasesoker();
        //this.subSamling = new ArrayList();
        /*subSamling =*/ dbSoker.gjorQuery(sporring);

    }
    catch(SQLException e){
        e.printStackTrace();
    }
} //Main

public void utskrift(){

    // lag noen utskriftsmetoder her eller i en annen klasse
}

} //Sokeprogram
```

