



ÅPEN OG SAMORDNET TILGANG TIL KULTURARVEN

**ANBEFALINGER FOR EN VELLYKKET TILSTEDEVÆRELSE
I DEN DIGITALE KULTURELLE VERDEN**

INNHOOLD

1 MANDAT OG AVGRENSINGER.....	8
2 ET PARADIGMESKIFTE.....	10

DEL I – FREMTIDENS KULTURARV, KULTURARVENS FREMTID

3 FREMTIDEN, VISJONER OG TRENDER.....	12
3.1 TRENDER OG VISJONER FOR DIGITAL OG DISTRIBUERT KUNNSKAPSREPRESENTASJON	12
3.2 SOSIALE TRENDER, MOBILITET OG AMATØRGENERERT INNHOOLD	16
3.3 TEKNOLOGISKE TRENDER – SEMANTISK TEKNOLOGI BLIR VIKTIG.....	19
3.4 HVA SKJER RUNDT OSS (PÅ EUROPEISK NIVÅ)?...28	
3.5 HVA SKJER I NORGE?.....	33
3.6 UTFORDRINGER VED DAGENS SATSING PÅ DIGITALISERING AV KULTURARV	39
4 DIGITALISERT KULTURARV OG FORMIDLING: HVORDAN TILRETTELEGGE ICT ARKITEKTUR FOR FREMTIDEN?.....	41
5 VEIEN VIDERE: HVORDAN KAN VI LØFTE ABM- SEKTOREN OG HJELPE TIL MED Å TILRETTELEGGE FOR FREMTIDEN?.....	48
5.1 DEFINERTE KRAV FOR DIGITALISERINGSPROSJEKTER.....	49
5.2 VEILEDNING OG KURS.....	50
5.3 FUNDING / INSENTIVER	50
5.4 DELTAGELSE OG DISKUSJONSFORA: AKTIV OUTREACH!.....	50

DEL II – FORSLAG TIL MINIMALE KRAV FOR DIGITALISERT KULTURARV

1 KRAV OM LOKALISERBARHET	52
2 KRAV OM TILVIRKING OG LAGING AV DIGITALE OBJEKTER	54
3 KRAV OM BESKRIVELSE.....	56
4 KRAV OM DIGITAL VARIGHET.....	59
5 KRAV OM PRESENTASJON	61
6 KONKLUSJON.....	62
7 ORDFORKLARINGER.....	66
8 REFERANSELISTE	68

FORORD

Kulturarven skal på Internett!

Frå 2008 til 2010 kom det fleire stortingsmeldingar som handla om korleis kulturarven skal gjerast meir tilgjengeleg.

- St.meld. nr. 15, (2007-2008), Tingenes tale. Universitetsmuseene
- St.meld. nr. 23 (2008-2009), Bibliotek. Kunnskapsallmenning, møtestad og kulturarena i ei digital tid
- St.meld. nr. 24 (2008-2009) - Nasjonal strategi for digital bevaring og formidling av kulturarv
- St.meld. nr. 49, (2008–2009), Framtidas museum. Forvaltning, forskning, formidling, fornying

Hovudbodskapen i desse meldingane var at institusjonane må opna magasina sine. Brukarane må få innsyn. Dette føreset auka digitalisering, samlingane på Internett og meir aktiv og målretta digital formidling av samlingar og fagkunnskap. Den digitale kunnskapsallmenningen er det kulturpolitiske målet. Det vil krevja stor og samordna innsats over mange år.

Nye tenester på nettet

I åra 2008-2010 har ABM-utvikling prioritert arbeidet med å etablera nasjonale fellesløysingar for arkiv og museum. Digitalt museum ([www.digitaltmuseum](http://www.digitaltmuseum.no)) vart opna i 2009. Her skal brukarane få innsyn i musea sine rike samlingar av foto, gjenstandar og kunst. Digitalt fortalt (www.digitaltfortalt.no) vart opna i 2009 og inneheld digitale forteljingar skapt av institusjonar og privatpersonar. Arkivportalen (www.arkivportalen.no) kom tidleg i 2010. Tenesta skal gje brukarane samla oversyn over arkivmateriale i arkiv, museum og andre institusjonar. Arkivportalen inneheld og ein del digitale kopiar av arkivmateriale. Det tilbodet vil auka mykje i åra som kjem.

Fleire andre viktige tenester har kome til i same tidsrommet. Riksantikvaren har utvikla Kulturminnesøk (www.kulturminnesok.no) som er ein publikumsversjon av fagbasen Askeladden. Universitetsmusea har svært omfattande samlingar av mellom anna foto og arkeologisk materiale. Deler av fotomaterialet kan brukarane no finna på universitetsmusea sin fotoportalt (www.unimus.no/foto/). Deler av dei arkeologiske samlingane er tilgjengeleg på www.unimus.no/arkeologi/.

Nasjonalbiblioteket har opna bokhylla.no med digitale kopiar av litteratur frå dei siste 10-åra i dei tre siste hundreåra. Statens kartverk har og utvikla fleire viktige tenester som er relevante i denne samanhengen. Institusjonen har store samlingar av eldre foto og historiske kart.

I 2009 kom første versjon av Europeana (www.europeana.eu). Ny versjon kjem hausten 2010. Europeana inneheld mykje data levert av norske institusjonar, men og innhald om eller frå Norge i europeiske institusjonar.

Norsk Lokalhistorisk Institutt har utvikla tenesta www.lokalhistoriewiki.no. Pr. september 2010 inneheld tenesta 8300 artiklar skrivne av både faghistorikarar og amatørar. Digitalt fortalt og lokalhistoriewiki er opne for bidrag frå både fagmiljø og amatørar. Kulturminnesøk og Digitalt Museum har lagt til rette for at brukarane kan senda inn sine kommentarar til samlingane. Dette er noko nytt, og markerer eit svært viktig vegskille. Brukargenerert innhald vil verta ein stadig viktigare og meir omfattande del av kulturarvstilbodet på nettet. Me ser og mange døme på at både institusjonar og privatpersonar legg materiale ut på sosiale nettstader som YouTube og Flickr.

Wikipedia innheld og mykje informasjon om historie og kulturarv. Store Norske Leksikon er kome som ei gratisteneste på nettet. Norsk Kunstnarleksikon kjem som netteneste om ikkje så lenge.

Eit sprikande univers

Stortingsmeldingane, som me innleia med, strekar under at denne utviklinga skal halda fram med auka styrke. Fleire og fleire institusjonar publiserer no digitale kopiar eller representasjonar av samlingane sine på Internett gjennom nasjonale fellestener. Det omfattar blant anna foto, gjenstandar, kunstverk,

arkivmateriale, arkeologisk materiale, informasjon om kulturminne, digitale forteljingar med meir.

Sjølv om det finst døme på tenester som samlar data frå fleire område og sektorar, t.d. Europeana, så er dei nye nasjonale tenestene ikkje kopla saman. Data er i praksis lagra i eigne siloar, og det meste av dette vert ikkje indeksert av dei store søkemotorane. Det er data i lukka databasar, ikkje tekstdokument ute på det opne Internettet. Data i desse basane må opnast opp for søkemotorar. Det vert og arbeidd med å opna basane slik at andre kan laga eigne søketenster direkte mot utval av data. Då er dei ikkje lenger avhengige av søkefunksjonane i Arkivportalen eller Digitalt Museum, men kan laga løysingar ut frå eigne behov.

Data frå dei omfattande nasjonale tenestene som veks fram må opnast for dei store søkemotorane.

Data må synast og vera der brukaren er, t.d. Google. Men det er ikkje sikkert at løysinga berre er å dumpa mange millionar datapostar ut på nettet slik at dei kan indekserast av t.d. Google. Denne rapporten skisserer ein mogleg veg som me trur vert viktig framover. Opne lenka data eller Linked Open Data er ein del av det som vert kalla den semantiske veven. Data skal kunne knytast saman basert på slektskap i innhald. Slik vil ein og ta vare på eksisterande kontekst og kunna knyta data saman med nytt innhald som kjem til eller som finst andre stader på nettet.

Kvar, kva, kva tid og kven.

På engelsk talar ein gjerne om where, what, when and who. Eit foto, ein gjenstand, ei forteljing, eit arkivdokument, eit kunstverk eller ein artikkel er på ulikt vis knytt til ein stad, handlar gjerne om ein person, eit tema og utspelar seg i eit avgrensa tidsrom. Alle dei tenestene me har lista opp ovanfor kan og vil ofte ha data om same hendinga, same tidsrommet, same

garden, same gata, same personen, same området eller same kommunen. Desse samanhengane er ikkje synlege i dag. Dei er usynlege for brukarane og dei er usynlege for fagfolk som skal nytta materialet til dokumentasjon, forskning og formidling. Dette er eit tankeskors og ei stor utfordring. Det er i dette biletet denne rapporten kjem inn.

Korleis kan data løftast ut av siloane, koplast saman med kvarandre slik at dei samla kan kasta lys over stader, historiske hendingar og saker. Korleis kan data og innhald frå abm-institusjonane koplast saman med data i wikipedia og liknande ressursar? Kva tenester kan etter kvart byggjast og utviklast ved hjelp av semantisk teknologi? Kva må abm-institusjonane gjera for å leggja tilhøva tilrette for ei slik utvikling?

Dette er hovudspørsmåla ABM-utvikling ville ha faglege innspel på frå ESIS og Robert Engels.

Rapporten skulle og formidla status og trendar på dette feltet på ein lettfatteleg måte.

Målgruppene for rapporten er tilsette i institusjonane og beslutningstakarar på ulike nivå. ABM-utvikling vil nytta rapporten som grunnlag for gode prosessar og drøftingar fremover. Me trur og vonar den vil bli eit verdfullt tilskot i ein debatt som abm-sektoren må ta på alvor i tida som kjem.

Opne og koplata data

Dei siste åra har det skjedd ei haldningsendring både hjå institusjonar og privatpersonar. Mine data vert til våre data. Lukka data vert til opne data. Det har vakse fram ein delingskultur. Offentlege data skal opnast og nyttast fritt. Kulturarven er vår, ikkje min eller din. Dette er eit godt utgangspunkt. Dette fører oss inn i den sosiale veven. ABM-utvikling engasjerte i vår Even Westvang til å sjå på korleis bruk av sosiale media kan koplast mot data i nasjonale fellesløysingar som t.d.

Digitalt Museum og Arkivportalen. Hans rapport finn du på www.slideshare.com. Den semantiske veven heng nøye saman med den sosiale og opne veven der brukarane spelar ein hovudrolle.

Oslo

Stein Slyngstad

direktør

Gunnar Urtegaard

avdelingsdirektør



Futhark var den første runerekken og inneholdt 24 tegn. Info og foto: wikipedia

OM RAPPORTEN

For å få oversikt og innsikt i spørsmål knyttet til den (delvis teknologiske) fremtiden på et internasjonalt nivå, er det foretatt en rekke intervjuer og møter med sentrale personer innenfor feltet «cultural heritage» i Norge og resten av Europa. Etter en telefonisk runde og undersøkelser av eksisterende initiativer viste det seg at viktige internasjonale trender kommer sammen i Den Haag, NL. Her samles det mange gode miljøer i spennende (Europeiske) initiativer. På akslingen Den Haag, Amsterdam og Luxembourg finner man mye av «state of the art» innenfor abm-sektoren. I tillegg til utlandet har vi også brukt tid på å få oversikt over de mange spennende prosjekter og initiativer i Norge, og har tatt en sammenligning med det som etter alt å dømme kommer til å bli «beste praksis» for infrastruktur og informasjonsarkitektur for formidling av digital kulturarv.

Jeg vil derfor gjene takke Gunnar Urtegaard, Lars Wenaas (ABM utvikling), Yvo Volman (EU commission / Directorate for Information Society and Media, Deputy Head of Unit «Information Access»), Luxembourg), Harry Verwayen, Borys Omelayenko og Anette Friberg (Europeana), Jon Peter van den Oever, Judit Moortgat (Beelden van de Toekomst og Nationaal Archief, NL), Paul Doorenbos, Huibert Crijns (Het Geheugen van Nederland og Koninklijke Bibliotheek NL) samt mange

nåværende og tidligere samarbeidspartnere i forskjellige digitaliseringsprosjekter (NRK, Nasjonalbiblioteket Mo i Rana, Popsenteret Oslo, Rockheim, etc.) for viktig informasjon, innsikt og diskusjoner som har gjort det mulig å skrive denne rapporten.

Robert H.P Engels
ESIS Norge AS



Foto: istockphoto.com

1. MANDAT OG AVGRENSINGER



Hvilken trender finnes og hva er forventningene som ABM-sektoren skal forholde seg til?

Hvordan kan ABM-sektoren best tilrettelegge for fremtiden?

Illustrasjon 1: Spørsmål undersøkt i rapporten

Hva er mandatet, hva skal behandles og hva skal ikke behandles i denne rapporten

Den foreliggende rapporten vil formidle en visjon om veier til en digitalisert kulturarv i Norge. Visjonen bygger på samtaler og diskusjoner med personer som har bakgrunn fra ABM-sektoren, men også med forskere og deltagere i nasjonal og internasjonal politikk og samarbeidsprosjekter, samt med representanter for opplevelsessentre.

Mandatet for rapporten er å komme med en visjon for fremtiden, sett med «frie» øyne og uten å være bundet av den historiske utviklingen. Det ligger også et potensial i det faktum at fremtidens bruk av den digital kulturarven stort sett er ukjent, og derfor skal rapporten ikke komme med konkrete anbefalinger for hvordan denne kulturarven skal formidles til fremtidens publikum. Dette skal være så «åpent» som mulig. Derimot skal det utredes hvordan ABM-sektoren kan ruste seg for å møte fremtidige krav på en så «åpen» måte som mulig som tillater og åpner opp for mange forskjellige anvendelser.

Hva kan ABM-sektoren gjøre for å dekke kravet om standardisering, åpenhet og samhandling innenfor det offentlige, som omtalt i diverse stortingsmeldinger?

Hvilken vei tar internasjonal forskning mot digital informasjonsforvaltning og spesielt digital representasjon av kulturarv for tiden?

Hva gjør de store organisasjonene og aktørene som f.eks. Rijksmuseum i Nederland, Louvre i Paris, the British National Library, the Library of Congress og BBC?

Hva er deres visjon for fremtiden, og hvordan kan man støtte mindre regionale museer, biblioteker og arkiver slik at de har tilgang til nødvendig kunnskap og kapital for å «være med» og dele og dra nytte av

andres informasjon?

En undersøkelse av politisk historikk og tidligere vedtak er ikke noen del av denne rapporten. Derimot tar den for seg en overordnet gjennomgang av

noen trender på nasjonalt og internasjonalt nivå, refererer til en del initiativer og portaler som finnes per dags dato og bruker konkrete eksempler fra diverse anvendelser våren 2010 som illustrasjon.



Egyptiske hieroglyfer. Foto: Wikipedia Commons

2. ET PARADIGMESKIFTE

Kulturarvens bidrag til forståelsen av et samfunns identitet og fortid kan knapt overvurderes. Som samfunn har vi derfor satt som krav at gjenstander og informasjon om fortid og nåtid blir bevart for fremtidige generasjoner.

Siden begynnelsen av nittitallet er det lagt betraktelig større press på ABM-sektoren for at den skal digitalisere sine samlinger. Å lage digitale kopier av åndsverk, kunstobjekter og historisk materiale har vist seg å være en effektiv måte å lage «sikkerhetskopier» på. Når mange gjenstander tidligere ble oppbevart i godt kontrollerte omgivelser og bare en brøkdel av materialet var tilgjengelig for publikum, ser vi nå at mulighetene for å lage digitale representasjoner av kulturarven fører til uante muligheter.

Nå er det ikke bare viktig at det lages «digitale kopier» av kulturarven, men at det også følger med digitale metadata, som er enkelt å kopiere, dele og distribuere. Dermed blir det enklere å slå sammen informasjon som stammer fra forskjellige steder. For kulturarven gjelder at summen av slik sammenslått informasjon er vesentlig større enn delene. Relevansen av integrert informasjon er høy, både for eksperter og samfunnet generelt.

I tillegg til en betraktelig innsats for å digitalisere det meste og beste av verdens kulturarv, foreligger en god del av nye kulturelle uttrykk allerede i digital form («born-digital»). Økt tilgang til data/digital fotografi mm. og økt kompetanse i bruk av hard-/software har ført til at mye mer materiale enn før blir oppbevart. Der hvor man tidligere tok hensyn til kostnadene for å dokumentere en hendelse (på åtti-tallet kostet et polaroid-

bilde f.eks. ca. NOK 3,- og tre minutters Super 8 film kostet gjerne NOK 80,-), koster det p.d.d. praktisk talt like mye å ta ett bilde som det koster å ta tusenvis. Siden hardware har blitt så mye billigere i anskaffelse og bruk, samt at tilgjengeligheten av lagringsplass er nesten ubegrenset, øker den tilgjengelige informasjonsmengden eksponentielt. Til sammenligning er kostnadene knyttet til digital publisering/distribusjon gjennom et datanettverk meget lave.

Det er klart at dette fører til et paradigmeskifte. Publikum er blitt vant til mobilitet og deltagelse. Det har kommet fram nye former for informasjons-spredning. Informasjon (og kunnskap?) blir gjort tilgjengelig gjennom portaler som lever av dugnadsånden (wikipedia, youtube, flickr). Kvaliteten av slik informasjon kan i mange tilfeller være høy, men er det ikke alltid. Det er blitt vanskelig å skille kvalifisert informasjon fra ukvalifisert informasjon. Bærbare musikk- og videoavspillere, smartphones og lignende produkter benytter seg av mange distribuerte informasjonskilder og bruker nye teknikker som f.eks. utvidet virkelighet (augmented reality) og 3D fremvisning som igjen øker populariteten og tilgjengeligheten.

Kort sagt er det både mengden av og tilgangen til informasjon i samfunnet som har økt dramatisk de siste 15 år, og det gjenspeiler seg i økende forventninger hos de kommende generasjoner om hva som skal være tilgjengelig i hverdagen (helst alt) og når (helst 24/7) og hvordan man skal kunne bruke det.

Men hvordan har man egentlig tenkt å imøtekomme disse forventningene?

¹ *Rapporten skal ikke gå dypt inn i spørsmålet om man «skal» imøtekomme alle forventninger. Lovverket, datasikkerhet osv. er gjenstand for en samfunns- og politisk debatt som bare såvidt har kommet i gang.*



DELL 1

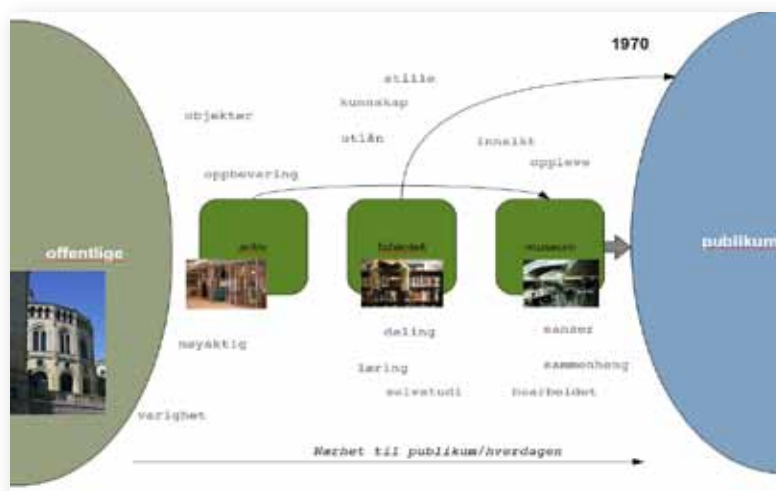
3. FREMTIDEN, VISJONER OG TRENDER

3.1 Trender og visjoner for digital og distribuert kunnskapsrepresentasjon

På globalt nivå ser man en trend som kan oppfattes som en direkte følge av en høy grad av tilknytning («degree of connectiveness») på internett, god tilgjengelighet av verktøy for publisering samt oppblomstringen av det sosiale nettet. Trenden er at man i stor grad forventer en døgnåpen mulighet for kommunikasjon med andre, døgnåpen tilgang til alle tjenester og at alt mulig materiale skapt i alle verdens tidssoner er direkte tilgjengelig.

Endrede forventninger til Arkiv, Bibliotek og Museum

Arkiver har også i mye større grad blitt gjort tilgjengelig for publikum. Mye av materialet er digitalisert eller skal digitaliseres i de nærmeste årene som et tiltak for å sikre kulturarven og gjøre den lettere tilgjengelig. Der hvor man tidligere foretok et fysisk besøk til et offentlig arkiv for å aktivt lete etter det man var ute etter, kan man i dag få døgnåpen tilgang til komplette samlinger på internett og undersøke skattene som ligger der. Publikum har høye forventninger når det gjelder brukbarhet og presentasjonen av innholdet. Dette er ikke noen naturlig rolle for et arkiv, som historisk sett hadde andre institusjoner som «front-end» mot publikum.



Illustrasjon 2: Enkelt, tradisjonelt perspektiv på ABM-sektoren

Arkiver representerer steder hvor kulturarven blir profesjonelt lagret og forvaltet. Arkivets hovedoppgave er å sikre at våre etterkommere har tilgang til samme kulturarv-materiale som vi selv har. Det er logisk at en slik funksjon ligger tett innpå «staten» og lenger bort fra publikum. Biblioteker har også samlinger, men disse er mye mer tilgjengelige og kan ofte lånes ut. Man kan gå til et bibliotek og få tilgang på alt mulig av informasjon. Viktig informasjon er ofte lagret på en redundant måte, slik at biblioteket kan låne ut flere eksemplarer av et medium samtidig. En viktig rolle for bibliotekene er å tilby allmenn tilgang til informasjon og kunnskap, men brukerne må selv gjøre en aktiv innsats for å erverve den. Museene står enda nærmere sluttbrukerne og tilbyr sine samlinger på en helt annen måte enn arkiver og biblioteker gjør. I museene kan man ikke låne ut objektene, men får derimot oppleve en historie som er kvalitativt høyverdig og laget av eksperter som videreformidler et bestemt syn på objektene. I den forstand tilbyr et museum «beriket» informasjon som allerede er sammensatt og interpretert for brukeren, i motsetning til et arkiv som ikke gjør noen interpretasjons-jobb og et bibliotek som heller ikke interpreterer, men som organiserer objektene etter tema/språk/land osv.

Demarginalisering mindre samlinger

Digitaliseringsbølgen har ført til at en del veldig spesialiserte samlinger har blitt tilgjengelig for et tynt spredd publikum som tidligere var avhengig av post, aviser, telefon og kollegaer/venner for at de i det hele tatt kunne innsamles og bygges opp. Prosjekter som Norges Pattedyratlas bruker internett aktivt for å hente inn metadata om forekomster, arter osv. på en direkte, døgnåpen og de-sentralisert måte². Det forval-

tes mye interessant materiale rundt omkring i landet som kanskje er blitt betraktet som «lokal-kunnskap» eller av lokal interesse, men som kan bli populært og veldig mye brukt når det åpnes opp for digital tilgang. Et godt eksempel er de mange emigrantene fra Norge som gjennom tilgang til slike baser kommer i kontakt med sin fortid, utfører slektsforskning eller rett og slett lærer mer om norsk kultur og det norske samfunnet.

Lokal tilstedeværelse vs opplevelser i det offentlige rommet

For å kunne tilby (kunnskapsrike) opplevelser til brukerne er ABM-sektoren avhengig av å kunne definere den konteksten en bruker befinner seg i. Konteksten blir ofte definert gjennom en lokal setting (utstillingslokaler, bibliotekrom, temaer) og brukeren må som regel være fysisk tilstede.

Yngre generasjoner (og i økende grad eldre generasjoner) forventer derimot gjerne en form for mobil/stedsuavhengig tilgang til informasjon. En slik tilgang er ofte basert på bruk av bærbare data (med verktøy som telefon, GPS, sub-notebooks osv.), hvilket innebærer relativt begrensede muligheter for kommunikasjon (små skjermer, liten plass). For å imøtekomme slike motstridende krav (verktøy med liten størrelse, men brukere med store forventninger til innhold) har industrien vært opptatt av å bruke flere og flere sensorer for å unngå begrensningene små tastaturer innebærer. De nyeste mobile apparatene bruker GPS-systemer, bevegelsessensorer, akselerasjonsmålere, trykkskjermer, kameraer og mikrofoner for å kommunisere med og til sine brukere. Det ligger et enormt potensial for ABM-sektoren i å utnytte slik informasjon om «bruker-kontekst» for å tilby skreddersydde og overraskende opplevelser.

2 Artsbanken. <http://artsobservasjoner.no>

Dermed er det satt i gang en trend hvor utlån av fysiske gjenstander blir mindre viktig enn før (ref. videoteker, biblioteker osv). Med oppkomsten av 3D-teknologi er det ikke usannsynlig at også skulpturer osv. kan oppleves digitalt i fremtiden. Dette skal ikke fjerne behovet for å stille ut originaler, men det setter en trend hvor «utlånsinstitusjoner» blir mindre viktige og opplevelser står mer sentralt (uavhengig av tid og sted).

Deltagelse isf mottagelse

Økt deltagelse av publikum i web 2.0 skaper også forventninger ovenfor ABM-institusjoner. Man ønsker rett og slett å kunne dele sine erfaringer med organisasjoner på en helt annen måte enn før. Enten det gjelder opplevelser som museumbesøk, bibliotek eller utstillinger, så ønsker man å dele dem med venner på facebook, linkedIn, Twitter og mange andre steder.

Gode prosesser, sikkert oppbevaring, garantert lesbarhet

Digitalisering er dyrt. Alt arbeid som digitaliseringen innebærer utgjør en viktig bremse for utviklingen, og feil i prosessen kan føre til store tap. Det er ikke gjort noen undersøkelser vi er kjent med som beregner nøyaktig hvor mye av digitaliseringen som er blitt gjort på nittitallet som må gjøres om igjen i dag pga. dårlig kvalitet, men det antas å være en del. I flere land har det blitt vanlig å sette nasjonale minimumskrav til selve digitaliseringsprosessen og registreringen av metadata. Dette løfter resultatene av digitaliseringsinnsatsen opp til minimumsnivået som trengs for å levere gjenbrukbare resultater. En viktig del av slike krav er sikker oppbevaring og fremtidig tilgjengelighet gjennom bruk av standardformater med en publisert spesifisering som garanterer lesbarhet. Mye materiale fra åttitallet er vanskelig å lese i dag, enten

fordi mediumet som er brukt ikke kan avleses lenger, eller fordi det er brukt spesielle og sjeldent brukte formater. Her finnes det mye erfaring å høste fra internasjonale satsinger, og det er altså viktig at man ikke bare ser på den beste tekniske løsningen på et bestemt tidspunkt, for så å nedprioritere åpenheten og den fremtidige tilgjengeligheten av datalagringsformatet som blir brukt.

Utvalg av materiale

Mange institusjoner har blitt mer kritiske til utvalget av materiale som skal digitaliseres og materiale som ikke skal det. Det er nødvendig med en god koordineringsprosess som registrerer hva som er digitalisert og hva som skal digitaliseres på tvers av sektoren. Mye materiale finnes i flere kopier på flere steder, og det er ikke alltid hensiktsmessig å bruke tid og penger på duplisering av allerede eksisterende digitale objekter. Nye teknologier som Semantisk Web bidrar til at slik duplisering kan unngås gjennom en bedre måte å publisere digitale objekter på, slik at alle interessenter kan enkelt sammenligne eksisterende ressurser med sine egne oversikter.

I Nederland ble det rapportert at 32% av materiale eid av ABM-institusjoner ikke trenger digitalisering (De Feiten, 2009).

(A)BM blir opplevelsessenter?

Det er en interessant observasjon at flere og flere ABM-organisasjoner tenker «opplevelser». Det er ikke lenger nok å være museum, bibliotek eller arkiv (ref Illustrasjon 3, neste side). Publikum er blitt mer involvert og deltar aktivt.

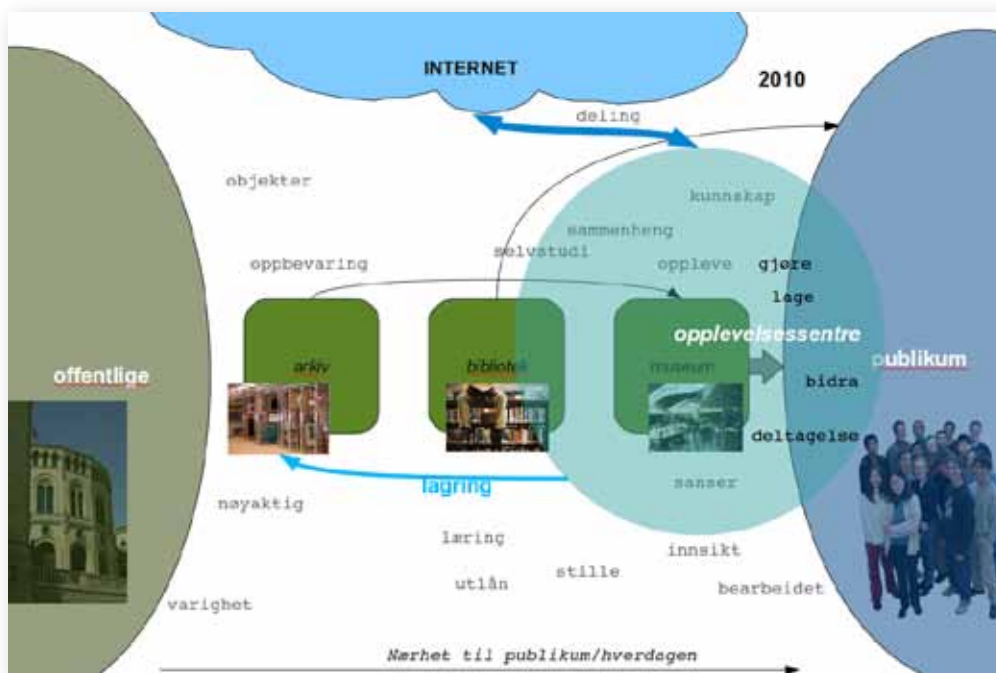
Det er også et større fokus på å «gjøre» ting, å være aktiv. Spillelementet har kommet for alvor. Mange museer ser også behov for å berike sine «opplevelser»

(man kan vel neppe snakke om utstillinger mer?) med informasjon de ikke har selv, f.eks. gjennom internett. Det er også en trend å ha en god representasjon med rike opplevelser på nett, som nettopp gir brukeren lyst å dra til et opplevelsessenter og erfare mer om temaet.

Denne trenden krever at man i fremtiden klarer å dele sine arkiver/samlinger med flere aktører, og at man kan ta imot informasjon (automatisert) fra andre. Siden det er sannsynlig at bransje-interne ABM-aktører i fremtiden blir nødt til å samarbeide med aktører fra andre bransjer (turisme, forlag, mm.), er det viktig at det brukes lagringsformater og publiseringsmetoder som holder mål på (inter-)nasjonalt nivå!

Fri tilgang for alle

Nettopp fordi så mye informasjon (aviser, leksika osv.) og tjenester/produkter (software, lagringsplass og prosessortid) som det tidligere måtte betales for er blitt relativt billig og enkelt tilgjengelig, har det oppstått en trend i retning av at alle data skal være åpne (uten passord) og fritt tilgjengelig (uten kostnader). Film- og musikkindustrien sliter mye med å tilpasse sine forretningsmodeller til denne utviklingen. Noen aviser har også gått fra betaling til gratis innhold og tilbake igjen, som Wall Street Journal og Financial Times varslet våren 2010. Rapporten «Digitale Feiten» rapporterer at 68% av all digital kulturarv i Nederland er gratis tilgjengelig uten registrering, og ytterligere 7% er gratis tilgjengelig etter at man har registrert seg.



Illustrasjon 3: Opplevelsessentere tar museenes plass?

Diskusjonen om tilgjengelighet og eventuelle kostnader krever en dyp og grundig samfunnsdebatt vedrørende kilder som er betalt med offentlige midler. Det finnes en klar global trend i retning av å åpne opp offentlige datakilder med statistiske data, demografiske data mm³. Når man skal åpne opp datakilder, skal man egentlig gjøre det ut ifra tanken om at summen er ofte mer enn delene. En slik tanke innebærer at man ikke skal være redd for at informasjon blir koplet sammen med andre kilder (ref. yr.no, hvor vær, kart og geodata blir kombinert). Det er selvsagt viktig at man ikke publiserer informasjon som er vernet av lovverket (personvern, datasikkerhet), og at ABM-institusjoner har klare retningslinjer for å bestemme om en digital samling kan gjøres åpent tilgjengelig før den blir lagt ut. Hvis den blir lagt ut, skal man helst ikke legge begrensninger på den videre bruken av informasjonen (kanskje unntatt en form for Creative Commons lisens som garanti for at åpne data forblir åpne).

Hvis man drar den linjen som man nå ser i flere land videre, så er det rimelig sannsynlig at digitaliseringsbølgen kommer til å fortsette, med muligens veldig like konsekvenser for biblioteker og bokhandlere som det som ble realiteten i musikkbransjen: færre fysiske enheter som blir formidlet, flere digitale objekter som «lånes» ut eller som man kan abonnere på. Dermed blir et bibliotek i stor grad en digital opplevelse på internett. Mange bokhandlere har begynt å tilby opplevelser som sofa-hjørner, kafearer osv, slik at de kan by på noe mer enn bare salg av bøker. Dette har også noen biblioteker forstått, og fremtidige biblioteker kommer sikkert til å oppfylle en annen og mer integrert rolle i samfunnet, og kombinere kunnskap og opplevelser.

3 <http://data.norge.no> (Norge), <http://data.gov.uk> (UK) og <http://www.whitehouse.gov/open/> (US).

Arkiver kommer også til å få en dobbeltrolle. På den ene siden vil de ta vare på eksisterende fysiske objekter, på den andre siden forvalte digitale kopier av alle disse fysiske objektene. På toppen av dette kommer det et helt nytt fagfelt som vil forvalte «born-digital» objekter, dvs. objekter som er laget digitalt fra starten av (f.eks. musikk, film, tekst, epost). Det er forløpig ikke tydelig hva dette virkelig innebærer og hvordan man skal samle inn den delen av kulturarven som er «born-digital». (Har man f.eks. lagret alle eposter fra Erlend Loe et sted? Hvem samler dem inn og hvor skal de forvaltes?) Det som er sikkert, er at det må stilles mye strengere krav til innsamling og oppbevaring av slike objekter enn man stiller til vanlig digital kulturarv som kan reproduseres mot en kostnad, noe som man nettopp ikke kan gjøre med born-digital objekter.

3.2 Sosiale trender, mobilitet og amatør-generert innhold

Ad-hoc informasjon og troverdighet

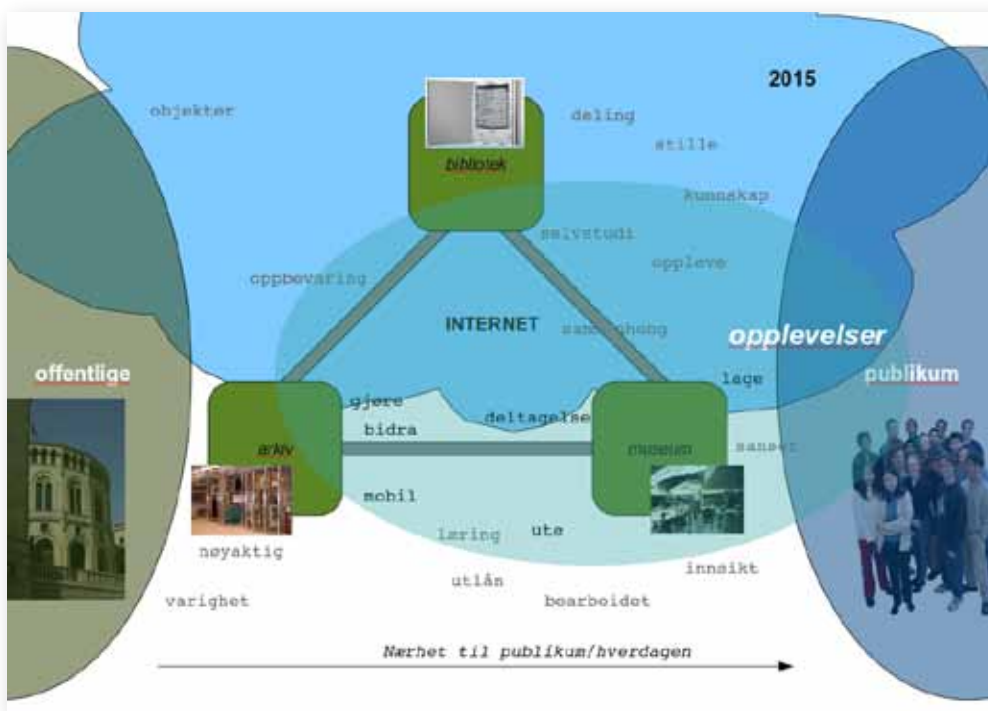
Informasjon blir delt av brukere, og det er en trend at subjektive omtaler av andre brukere nesten blir tatt mer seriøst enn kvalifisert informasjon. Sammenslåing av lokal og kvalifisert informasjon («når var siste besøk av matilsynet til denne restauranten») blir gjerne kombinert med informasjon fra andre brukere («hva synes de om restauranten og menyen?»). ABM-sektoren kan bidra med mye interessant materiale om et sted gjennom by-arkivet eller riksarkivet som f.eks. har bilder fra dette stedet anno 1895, informasjon fra Popsenteret eller Rockheim om artister som har tilknytning til dette stedet (og evt. hva slags tilknytning), Nasjonalbiblioteket har kanskje en videosnutt med et intervju om stedet osv. Det finnes allerede i dag en del

applikasjoner som bruker «Augmented Reality» (utvidet virkelighet) teknologi for å kople slik informasjon sammen i en live-video stream via kamera og skjerm på telefonen.

Slike scenarier og verktøy burde også ABM-sektoren se på som reelle alternativer i mye større grad. En viktig teknisk forutsetning er at denne typen informasjon blir knyttet mot de riktige objektene gjennom et globalt identifikasjonssystem som kan brukes for å knytte metadata til de riktige objektene (nøkkelssystem). Dagens internett byr på denne muligheten gjennom standarden Unified Resource Identifier (URI).

Web 2.0, mashup, deltagels

Tidligere tilbragte brukerne mye tid med å finne informasjon fra forskjellige sider, for så å sette sammen informasjonen selv. Nå forventer man derimot å kunne bruke automatiserte tjenester som integrerer informasjon fra flere steder (booking av hotel, minSide, Gotogate.no, metacrawler.com, google news, etc). Man ønsker ikke bare å få tak i informasjonen på en rask og enkel måte, men man ønsker også å kunne delta og bidra med informasjon selv. Denne utviklingen blir gjerne kalt web 2.0, og automatisert kombinasjon av informasjon fra forskjellige kilder i små applikasjoner online blir ofte referert til som mash-up.



Illustrasjon 4: Fremtiden?

Arkivet lagrer og leverer tjenester til (internett-basert?) bibliotek, opplevelsesentre og museum

Det finnes imidlertid en del utfordringer som har gjort at det neste steget allerede er blitt tatt, dvs. i retning av web 3.0. Typiske web 2.0 applikasjoner er for eksempel dårlig tilgjengelig for søkemotorer og har gjerne API (programmeringsgrensesnitt) som ikke er generiske, og derfor må man tilrettelegge sine applikasjoner for hver kilde man ønsker å ta i bruk. Valget av kildene som brukes av en typisk mash-up applikasjon er ofte forhåndsbestemt, og det er ikke så lett å legge til eller fjerne kilder. Web 2.0 applikasjoner slår sammen informasjon og presenterer det til brukerne, men tilbyr sjelden automatisert tilgang til dataene som blir brukt i applikasjonen.

Web 2.0 er en fellesbetegnelse for den «andre» generasjonen internett. Der hvor Web 1.0 fokuserte på deling og distribusjon av dokumenter (et «read-only web»), så blir 2.0 også kalt «the read-write web». Web 2.0 tillater aktiv deltagelse av brukere for å lage objekter. Denne generasjonen støtter utvikling av interaktive web-applikasjoner som tillater aktiv deltagelse av brukere i skapingsprosessen, samt grunnleggende muligheter for å gjenbruke og slå sammen eksterne informasjonskilder.

Tagging av omverden

Brukere er opptatt av å kunne dele informasjon ad-hoc og til alle døgnets tider. Mens man reiser, lager innlegg i en blogg, tar bilder eller bare har sensorer på mobilen, legger man igjen et digitalt spor. For hvert punkt på jorden hvor mennesker lager slikt innhold eller rett og slett kommer forbi, blir dette punktet markert («tagged») med informasjon som lagres i den virtuelle verden. Dette kan være sensor-informasjon fra bilen (trykk, temperatur, fuktighet), men også bilder (Eiffeltorn, Sinnataggen) eller innlegg («maten her er dårlig»), gjerne i sammenheng med geo-location (fra GPS) og dato+tid.

Denne trenden har et interessant element i seg som burde kunne utnyttes av ABM-sektoren. Med mange brukere der ute som sier mye om mange ting er det godt mulig å fange opp en del uttalelser og informasjon om objekter som tilhører en samling. Det kan være lokalt (i et museum) gjennom f.eks. twitter feeds, men det kan også være på reise (bilder, twitter, blogg om f.eks. Sinnataggen i Frognerparken).



Illustrasjon 5: Wikitude: mobiltelefon med "Augmented Reality" (siddy.wordpress.com)

Ikke kast noe – datalagring ubegrenset tilgjengelig?

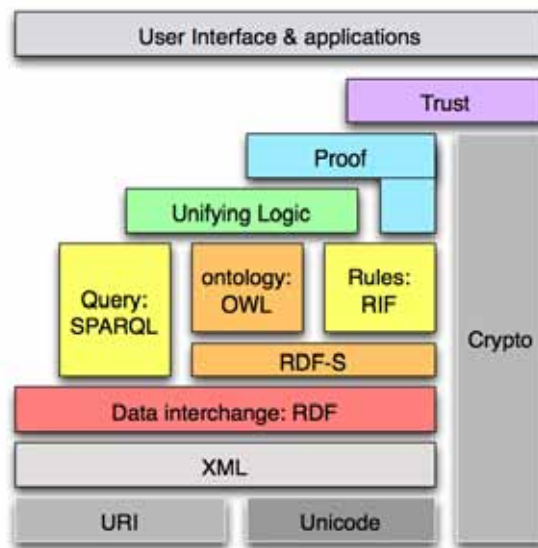
Mens man tidligere valgte materiale til arkivering ut ifra tilgjengeligheten av fysisk lagringsplass, er dette blitt et mindre problem i den digitale verden. Man kan lagre enorme mengder av informasjon per volum som ikke koster veldig mye i anskaffelse og drift. Denne trenden ser man også på internett, det er ikke vanskelig å få 1+ Tb lagringsplass på nett gratis, og organisasjoner som The Internet Archive tilbyr lagringsplass «ubegrenset i størrelse og tid». Brukere blir mer og mer ukritiske, og lager og bevarer mye materiale av lav kvalitet. Slike baser kan få stor betydning om noen år for samfunnsvitenskapelig forskning, men det er kanskje ikke alltid umiddelbart interessant. Hvordan kan ABM-sektoren finne, bruke og kanskje bidra til at de faktiske gullkornene lages og gjenfinnes? Samspillet mellom ABM-innhold og «internett-basert» innhold blir mer og mer viktig i årene som kommer.

3.3 Teknologiske trender – semantisk teknologi blir viktig

De siste årene har det blitt en større fokus på mulighetene for å komme bort fra publikasjon av web-sider som ofte omtaler mange forskjellige temaer eller elementer i ett og samme dokument. Målet er å få til en forvaltning av selve informasjonselementet på det laveste nivået og gjøre det maskininterpreterbart. I 1999 ble tanken om den semantiske veven presentert av Tim Berners Lee. Disse ideene har ført til en stor satsing fra Europa, USA og Asia (Japan, Sør-Korea) for å definere og utvikle et felles rammeverk under paraplyen W3C. Dette har ført til spesifikasjonen av Resource Description Framework (RDF), RDF Schema og Web Ontology Language (OWL).

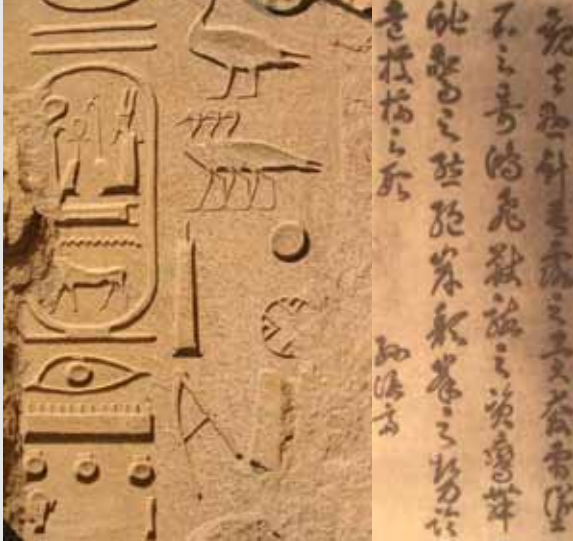
Mange av ulempene med mer tradisjonelle metoder for informasjonslagring lar seg løse med bruk av semantisk teknologi, nettopp fordi semantisk teknologi tilbyr bedre tilgang til kunnskap gjennom bruk av maskin-prosesserbar metadata. I dette kapitlet blir konseptene rundt Semantisk Web Teknologi slik det er foreslått av W3C forklart i litt mer (teknisk) detalj. Det forutsettes her en viss kjennskap til informasjonsteknologi generelt og distribuerte datasystemer spesielt.

Semantisk web teknologi illustreres ofte med bildet fra Illustrasjon 6 som viser en såkalt “stack” (eller stige) av teknologier som bygger på hverandre og til sammen utgjør en viss funksjonalitet. Bildet viser de laveste nivåene, HTTP/internett kommunikasjonsprotokollen (URI+Unicode), XML og RDFS på toppen av denne, samt OWL og noen sikkerhetsmekanismer og



Illustrasjon 6: Semantisk Web “stack” med tekniske nivåer. Bildet er fra W3C.org.

«Executive Summary» av Semantisk Web Teknologi



Suksessen til internett bygger på “konnektivitet” og allmenhetens muligheter til å bidra med informasjon. Alle kan publisere informasjon om hva de er opptatt av. Dessverre er mye av informasjonen på web generert av og for menneskelig interpretasjon, noe som gjør automatisert prosessering av all informasjon

i nåværende generasjon av internett vanskelig.

Semantisk Web Teknologi gjør det derimot mulig å publisere informasjon (og kunnskap) på internett i en form som ikke bare mennesker kan lese og interpretere. Det er forventet at internett, ved siden av å være dagens “bibliotek” og “oppslagstavle”, på denne måten også blir en global “kunnskapsbase” gjennom W3Cs standarder for kunnskapsmodellering på internett.

Et sentral begrep her er ordet “ONTOLOGI”, noe som det finnes mange definisjoner på (fra Sokrates og Aristoteles til internett i dag). Relatert til Semantisk Web defineres Ontologi gjerne som “en formell spesifikasjon av en konseptualisering (gruber, 93)”. I praksis betyr dette at man i fellesskap prøver å lage et begrepssystem som kan beskrive den verden man kjenner. Aktuelle forekomster av objekter (du og meg, bilen min, osv.) blir altså beskrevet

i og lenket til dette systemet, men man kan like godt beskrive konsepter, ideer, meninger osv.

Teknologien som trengs for å implementere et slikt scenario bygger på eksisterende internett og intranett infrastrukturer, og krever derfor ikke noen hardware investeringer for å komme i gang. Ser man på “stacken” i Illustrasjon 6, så ser man at de nederste nivåer er teknologi som allerede finnes og er implementert (URI/Unicode bygger på standard internett og intranett teknologi som HTTP sockets mm.).

Ikke alle nivåer er nødvendig for for-slaget i denne rapporten, og av nivåene over XML er det primært RDF, OWL og SPARQL. Til disse standardene finnes det meget robust Open Source programvare fra ulike miljøer (sourceforge.net osv.) samt kommersielle leverandører som IBM og Oracle. Flere andre globale aktører har signalisert at de vil følge W3Cs anbefalinger.

«trust» mekanismer på toppen.

Det er viktig å huske at de nederste nivåene (URI/Unicode og XML) allerede eksisterer i dagens internett, og grunnlaget til denne “stacken” er dermed allerede kjent og i bruk.

I denne diskusjonen er det først og fremst URI, RDF/RDFS, OWL og SPARQL som er relevant. Resten av kapitlet vil gå litt dypere inn på hva disse termene betyr, samt gi noen eksempler.

3.3.1 Unified Resource Identifier (URI)

Konseptet URI spiller en meget sentral rolle i digitalisering av informasjonskilder. Kort sagt er en URI en måte å identifisere en digital, internett-basert ressurs på en global, unik måte. En URI kan ses på som en «nøkkel» for en digital, virtuell representasjon av et objekt og representerer dermed selve objektet på internett. Det er denne «nøkkelen» som man bruker for å uttrykke noe om dette objektet. Slike uttrykk kan komme fra mange forskjellige steder og applikasjoner, men så lenge de refererer til samme URI burde det være en garanti for at informasjonen gjelder ett og samme objekt og derfor kan slås sammen.

3.3.2 eXtended Markup Language (XML)

Nær bunnen av «stacken» ligger det et nivå som heter eXtensible Mark-up Language (XML). XML skiller selve designet i websider fra deres data eller innhold, som blir representert som en “nøstet” eller rekursiv tre-struktur. Med hjelp av “tags” annoterer man informasjonen. En person med navn, kontor og telefonnummer kan representeres i XML som i Figur 1:

```
<?xml version="1.0"?>
<person>
  <navn>Ola Normann</navn>
  <kontor>1.03</kontor>
  <telefon>23167546</telefon>
</person>
</xml>
```

En slik struktur gir muligheter for strukturert dokumentanalyse, så fremt strukturen er kjent. Dette betyr i praksis ofte at en gruppe aktører må bli enige om den. Det er ingen selvfølge at andre enn de som laget den bestemte XML strukturen forstår innholdet og relasjonene mellom de ulike “tags” som er brukt. Hvis man sender denne strukturen med innhold til en kinesisk samarbeidspartner, så er det sannsynligvis ikke så mye hun kan forstå.

I mange tilfeller er en slik konsensus urealistisk, og man har derfor begynt å tenke på representasjon av struktur på et mer abstrakt nivå, noe som førte til utviklingen av RDF og RDFS.

3.3.3 Resource Description Framework (RDF) og RDF-Schema (RDFS)

For å kunne utveksle data mellom systemer uten at man trenger å bli enige om innhold i datamodellen på forhånd, har W3C kommet med en standard som kalles Resource Description Framework (RDF). Standarden blir “serialisert” eller “skrevet” i XML, noe som

betyr at man bruker XML for å representere data i RDF. Strengt tatt er ikke bruk av XML en forutsetning for RDF. Andre representasjonsmekanismer er også tillatt. Men i praksis har det vist seg at XML dekker de fleste behov, samt at det er et format som er støttet av mange aktører, store som små.

RDF er ment til å beskrive data og ressurser på web, og gjør det gjennom en såkalt Object-Attribute-Value modell. I en slik modell blir en ressurs beskrevet ved hjelp av OAV-»triples»⁴:

- Object: som representerer selve ressursen, f.eks. en webside
- Attribute: som beskriver objektet, f.eks. at websiden har en forfatter

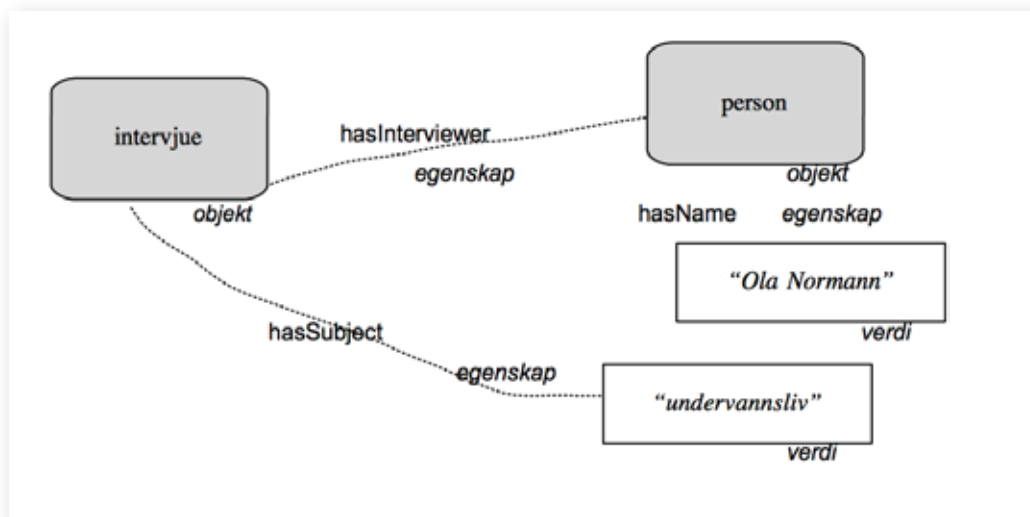
- Value: kan være et annet objekt (URI som beskriver en person) eller en "literal" (tekststreng)

Siden RDF er en formell mekanisme for kunnskapsrepresentasjon, er det mulig å resonnerer og trekke konklusjoner fra modeller i RDF.

En beskrivelse i RDF følger alltid OAV skjemaet:

Object (Objekt) – Attribute (Egenskap) – Value (Verdi)

Hver "verdi" kan enten være enkle verdier (som f.eks. et tall "3.2") eller kan representere et annet objekt. På denne måten er det mulig å bygge opp hele nettverk av relasjoner mellom objekter og verdier (se Figur 2).



Figur 2: Nettverk av AOV-»triples» (RDF)

⁴ Det mangler en god oversettelse for det engelske begreppet «triple» i sammenheng med IKT og logikk. Derfor har vi valgt å bruke det engelske ordet.

En utvidelse av RDF utgjøres av RDFS, hvor “S” står for Skjema. RDFS er et mekanisme som tillater interpretasjon av RDF-uttrykk, og inneholder et ontologisk vokabular med Objekter, Egenskaper og Relasjoner for å kunne beskrive ontologier. Sammen tilbyr RDF og RDFS et system, ofte referert til som RDF(S), som leverer en syntaktisk modell, samt den semantiske strukturen for å definere ontologier som kan bearbeides av maskiner. På denne måten går annotasjonene i RDF(S) lengre enn hva XML kan tilby (David 2007):

- RDF(S) er deskriptivt (beskrivende), ikke preskriptivt. Dette innebærer at der hvor XML bestemmer formateringen til enkelte dokumenter, så tillater RDF(S) oss å beskrive ethvert innhold, og de resulterende RDF(S) annotasjonene må ikke nødvendigvis være noen del av selve dokumentet. De kan like godt forvaltes separat fra det innholdet som beskrives.
- Flere RDF(S) ontologier kan beskrive samme innhold. Dette tillater forskjellige “meninger” eller “synspunkter”, men det kan også brukes til å reflektere forskjellig bruk.
- RDF(S) har en klar og vel-definert semantikk, som tillater resonnering og sammenslåing av uttrykk fra forskjellige systemer, basert på semantikk.
- RDF(S) tillater oss også å definere relasjoner mellom ressurser.

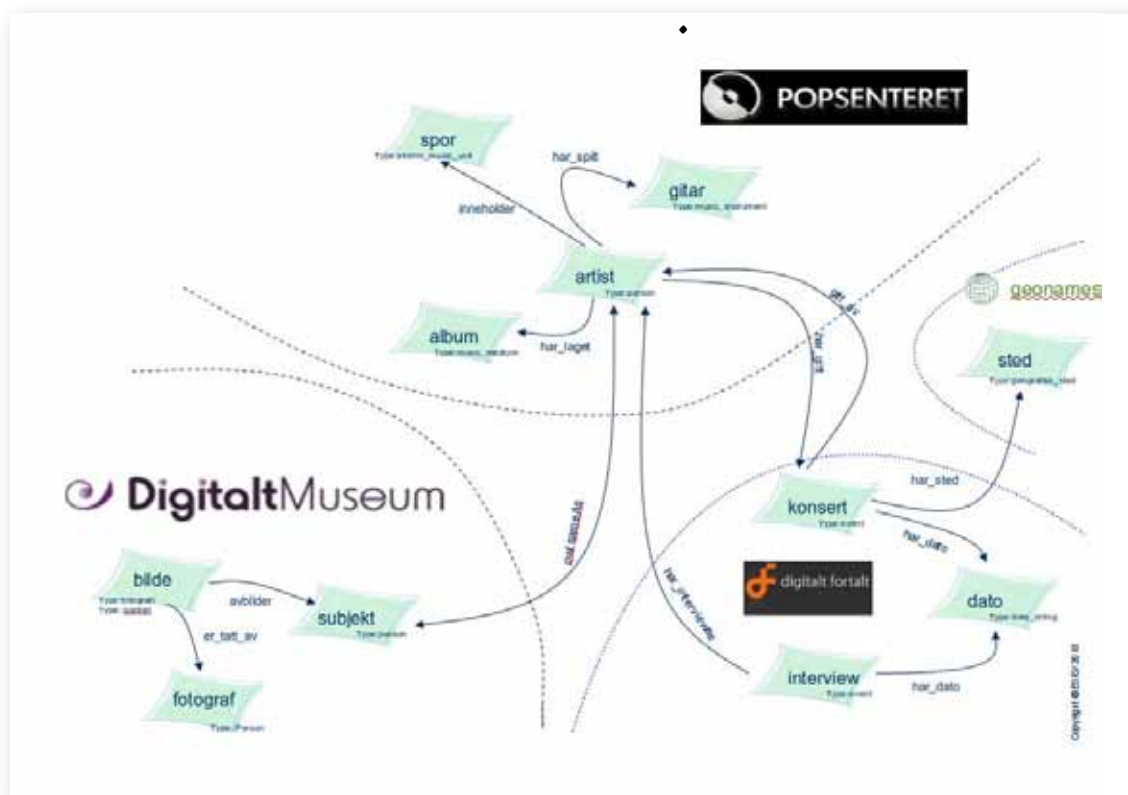
Selv om RDFS er et kraftig verktøy i seg selv, dekker det ikke alle behovene til et system som skal være distribuert og baseres på samhandlingsprinsippet. I RDFS mangler⁵ vi muligheter for å:

- uttrykke relasjoner mellom klasser definert i forskjellige dokumenter
- lage aksiomer for å kunne danne nye klasser fra eksisterende klasser ved å ta i bruk gjennomsnitt (INTERSECT), sammenslåing (UNION) og eventuelt komplementære klasser
- legge restriksjoner på antall og type av egenskapene til klasser
- bestemme hvor mange i en klasse som skal ha en bestemt egenskap; alle, noen eller ingen.

3.3.4 Web Ontology Language (OWL)

Web Ontology Language (OWL) er et relativt nytt forslag fra standardiseringsorganisasjonen W3C. Det har som mål å tilby ekstra funksjonalitet som hverken XML, RDF eller RDFS har. Formålet er selvfølgelig å kunne beskrive kunnskap enda bedre. Der hvor XML leverer et språk for utveksling av innhold og RDF/RDFS gir oss de byggeklossene man trenger for å representere enkle ontologier og instanser, er det OWL som gir oss muligheten for å beskrive verden enda mer nøyaktig ved hjelp av logiske operatører.

5 (Yu, 2007)



Illustrasjon 7: Eksempelvis samhandling mellom forskjellige internett baserte datamodeller.

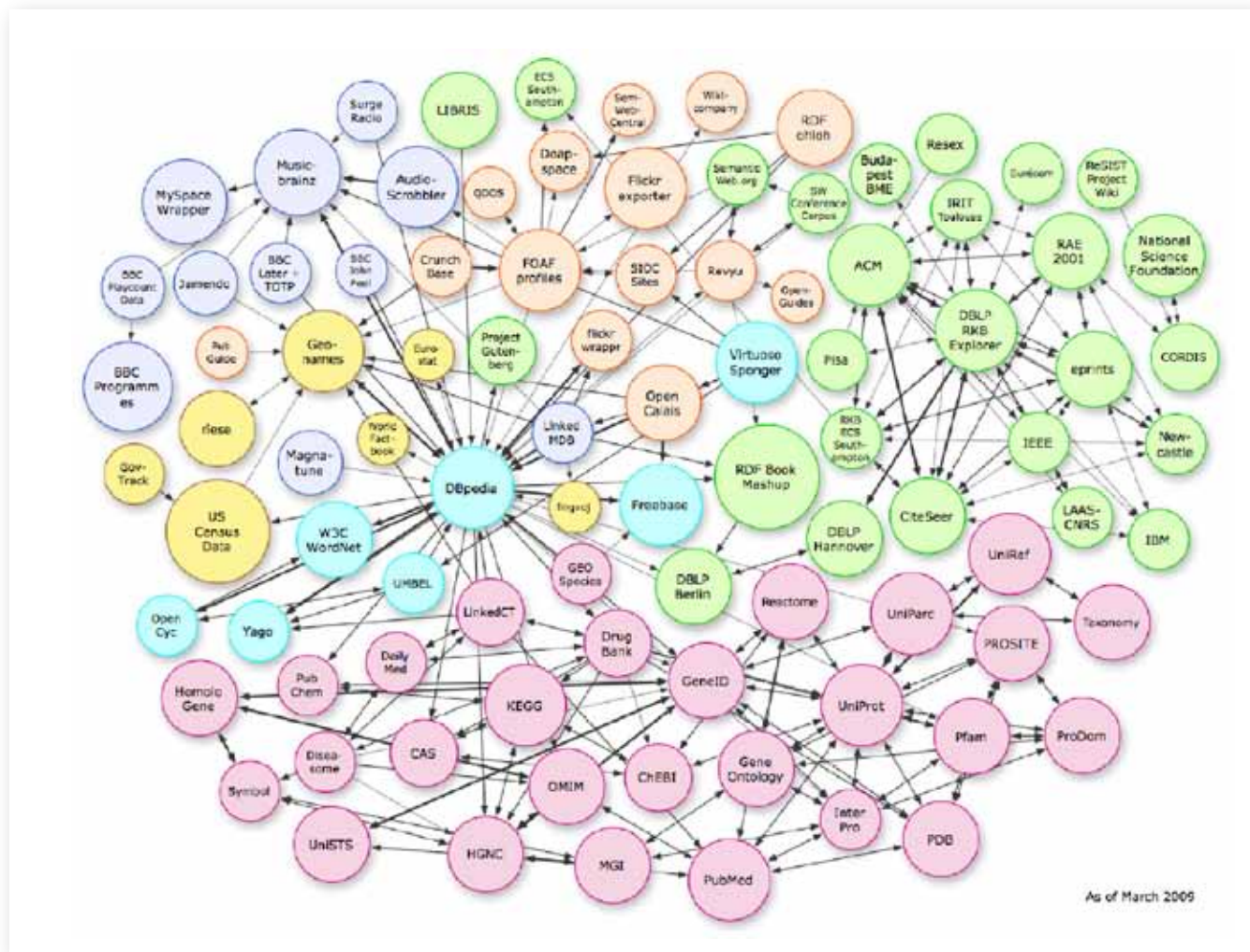
Mens RDFS gir muligheten for å beskrive instanser, klasser, subclasser osv. på en enkel måte, tilbyr OWL avanserte beskrivelser.

Typisk tilleggsfunksjonalitet i RDFS er:

- “range” og “domain” restriksjoner. Man kan f.eks. si at en popgruppe må ha minst én person som har rollen “artist”.
- “kardinalitet” restriksjoner. OWL tilbyr minCardinality, maxCardinality og Cardinality. Med disse aksiomene kan man definere at et symfonisk orkester må ha minst 40 personer med rollen “artist”, og maksimalt én person med rollen “dirigent”.
- “disjointWith” restriksjonen. I OWL kan man si at en planteeter er det motsatte av en kjøtteter.
- definere om en klasse kan relateres til en annen klasse eller en verdi.
- definere om en relasjon kan ha én eller flere verdier på “hver” ende. For eksempel kan hvert musikkalbum bare har én European Article Number (EAN).
- definere at en relasjon er symmetrisk: ethvert EAN kan bare knyttes til ett musikkalbum.
- diverse set-teoretiske aksiomer: unionOf, complementOf, intersectionOf etc.

Disse (og noen flere) gir OWL en kompleksitet som kan brukes for å beskrive de fleste domener. Hvis uttrykkskraften til RDFS/OWL blir kombinert med

styrken i XML til å kunne overføre informasjon i et generisk format over internett, så ser man hvorfor den Semantiske Web er så viktig.



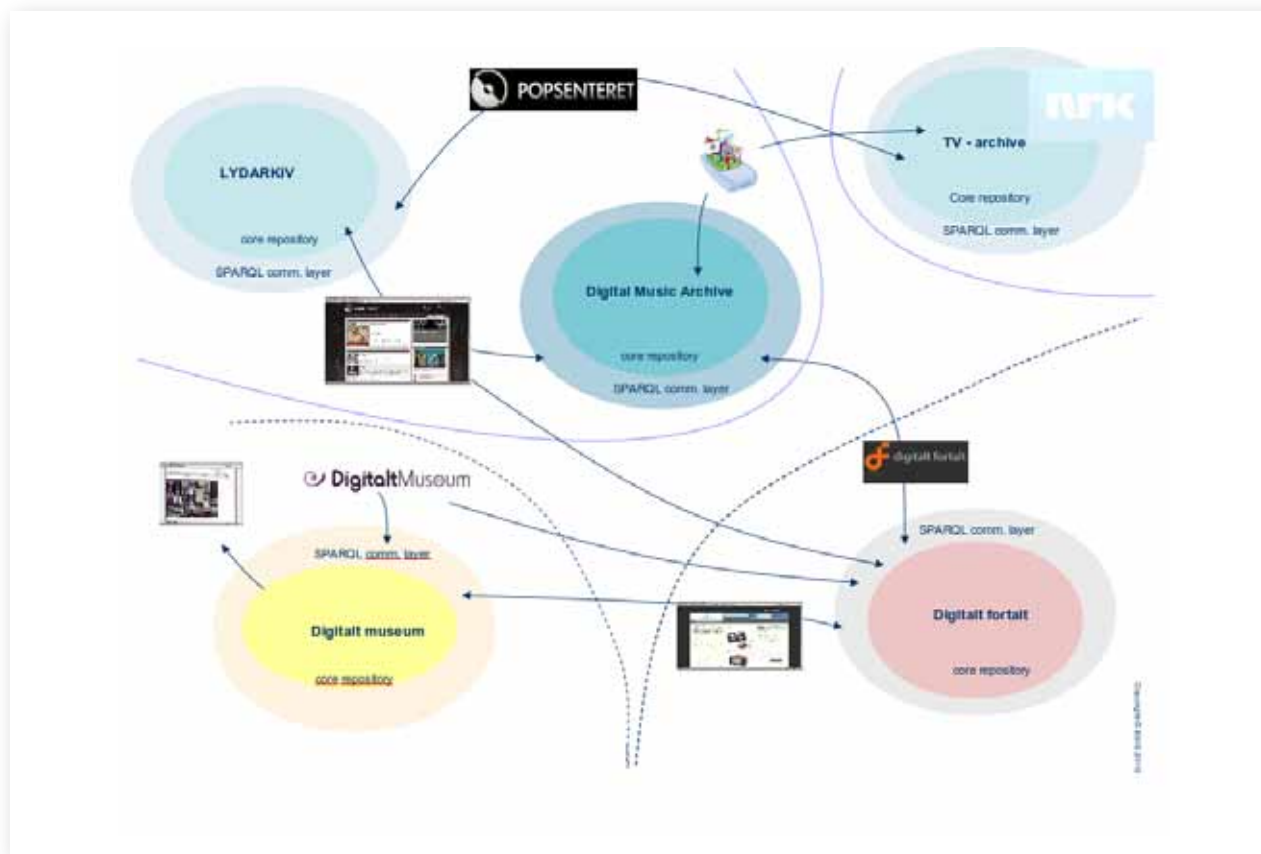
Illustrasjon 8: Linked Open Data Cloud – status 2009

3.3.5 Samkjøring av forskjellige kilder: Web Services og SPARQL

Internett slik vi kjenner det i dag består av mange datamaskiner som er direkte og indirekte koblet sammen med hverandre.

Datamaskiner kan dele informasjon seg imellom, og det finnes flere protokoller for en slik utveksling av data.

Semantic Web Services (SWS eller SWWS) representerer en type teknologi som forbinder to viktige utviklinger i evolusjonen av Semantisk Web.



Illustrasjon 9: SPARQL-enderpunkter som kommuniserer i et nettverk

For det første har Web Service - teknologien utviklet seg slik at det har blitt mulig å distribuere arbeidsoppgaver over flere maskiner, for så å legge sammen resultatene etter at oppgavene er gjort. Dette innebærer også at en applikasjon kan hente informasjon fra flere kilder som hver for seg er "ekspert" på et bestemt område. Hittil har utviklingen av Web Services ment å etablere et leverandøruavhengig rammeverk, dvs. uavhengig av platform, organisasjon eller nettverk.

Den andre trenden i evolusjonen av WWW er den Semantiske Web, som fokuserer på deling av metadata i et delt kunnskapsnett. Den Semantiske Web har kunnskapsrepresentasjon som hovedmål (og dermed representasjon av innhold), noe som kan føre til "intelligent" software og agenter som på uavhengig basis kan samle inn informasjon fra flere kilder.

For å kunne formidle kunnskap til andre datamaskiner trengs det også et "kommunikasjonssystem" som tillater at datamaskiner "snakker" sammen. Vanlige Web Services kan brukes til samhandling datamaskiner imellom, men desverre har ikke Web Services noen deklarativ eller forklarende måte å beskrive sine tjenester på. Det er med andre ord ikke mulig å finne ut hva en Web Service kan tilby. Dermed må man på forhånd kjenne til funksjonen til den Web Service man har tenkt å bruke. Gjennom bruk av SWS beskrives derimot funksjonene deklarativt og i større detalj, noe som tillater bedre verktøy for oppdagelse, integrasjon og samkjøring av Web Services.

SWS bruker et spørrespråk ved navnet "The SPARQL Protocol and RDF Query Language" (SPARQL) for å kunne spørre andre datamaskiner om informasjon. Språket er basert på databasespråket og industristandarden SQL, og tilbyr en enkel måte å lage spørringer på. Datamaskiner som jobber med å løse et problem eller prøver å finne informasjon

til et bestemt formål kan ved hjelp av SPARQL henvende seg til andre datamaskiner som har en publisert Web Service og spørre dem om informasjon (f.eks: "hvem spilte Terje Rypdal med på Molde Jazz 2007 og hva er hans siste plate?"). SPARQL er relativt enkelt å forstå og bruke, og danner grunnlaget for distribuert lagring og forvaltning av informasjon på internett. Linked Open Data initiativet bygger på denne standarden.

3.3.6 Linked Open Data (LOD)

Trenden er at flere og flere organisasjoner beskriver sine datakilder ved hjelp av RDF/RDFS/OWL og tilgjengeliggjør dem for automatisk spørring. Gjennom kryssreferanser mellom slike åpne, standardiserte informasjonskilder kobles innhold sammen til et globalt tilgjengelig kunnskapsnettverk. Linked Open Data er initiativet for å få så mange som mulig forskjellige datakilder koplet sammen. Illustrasjon 8 viser status på LOD skyen i 2009.

Ser man litt nærmere på den, finner man relativt mange velkjente informasjonskilder. Det er ingen som har lagt føringer på bruken av denne skyen. Størrelsen på kildene angir hvor mye data som ligger i dem, tykkelse og retning på lenkene angir hvem som referer til informasjon i hvilke andre baser, og hvor mange relasjoner som har blitt definert.

Sentralt i figuren finner man den semantiske versjonen av Wikipedia, kalt DBPedia. Her ligger alt av informasjon som Wikipedia også publiserer, men tilgjengeliggjort via et semantisk format gjennom et standardisert webgrensesnitt. Ellers så finner man en del kilder til musikk og film (bla. BBC data: lilla farge), offentlige kilder (US census data mm: gullfarge), bio-medical data (i rosa), og mange biblioteks- og forlags-

kilder (i grønt).

Noe å merke seg er at store og viktige initiativer som Europeana ikke har kommet så langt at de publiseres på samme måte (se diskusjon om Europeana i seksjon 3.5).

3.4 Hva skjer rundt oss (på europeisk nivå)?

Europe is proud of its cultural diversity. Language, literature, theatre, visual arts, architecture, crafts, the cinema and broadcasting may belong to a specific country or region, but they represent part of Europe's common cultural heritage. The European Union aims to preserve and support this diversity and to help make it accessible to others.

-- Maastricht Treaty, 1992

Europeana-prosjektet sikter mot en samling av den komplette europeiske digitaliserte kulturarven. I tillegg ble nylig <http://data.gov.uk> lansert som et samlested for offentlig data i Storbritannia, uten bruksbegrensinger. Data.gov.uk bruker semantisk web - teknologi for å tilgjengeliggjøre mange forskjellige datakilder gjennom en felles webtjeneste med standard representasjonsformat. Det samme har Library of Congress i USA gjort for en stund tilbake, og flere land står for tur. Felles for mange av disse er at de forholder seg til anbefalingene fra W3C om standardisering, og RDF/RDFS velges ofte som informasjonsutvekslingsformat. Dette formatet lå f.eks. til grunn i de første RSS implementeringene (RDF Site Summary) som nå er mye brukt i web 2.0 sammenheng, og som er en av de viktigste eksisterende formatene for å publisere oppdateringer på avissider, hjemmesider og ulike informasjon som vær- og trafikkmeldinger.

På europeisk nivå er det blitt satset tungt på initiativer som Framework 6 og 7 med deres fokus på «cul-

tural heritage». I perioden 2007 – 2013 alene er det blitt satt av 400 M€ (NOK 3,2 milliarder) for styrking av en felles europeisk satsing på bevaring og digitalisering av kulturarv.

Vi har gjort en del intervjuer om tilstanden hos forskjellige institusjoner i Europa. Vår undersøkelse viste at mange europeiske tråder innenfor feltet for digital kulturarv kommer sammen i «Directorate General for Information Society and Media» i Luxembourg. Relevante og interessante initiativer fant vi også i Frankrike (Louvre, mm.), UK (the British National Library, data.gov.uk, BBC open data) og Tyskland (hvor det finnes flere profesjonelle, statlige aktører og prosjekter innen ABM markedet). Det eksisterer noen store internasjonale initiativer som «the European Archive» og Europeana som har sine røtter i Amsterdam og Den Haag, hvor også en del andre svært omfattende digitaliseringssatsinger blir koordinert av den Koninklijke Bibliotheek, Nationaal Archief og Digitaal Erfgoed Nederland (DEN Foundation). Av denne grunn har vi valgt å fokusere på diskusjoner med sentrale personer i EU systemet (Luxembourg), Den Haag og Amsterdam.

Selv om nesten alle europeiske land er med i Europeana-prosjektet og som et ledd i dette rapporterer inn resultater innen digitalisering hvert år, er det vanskelig å sammenligne tallene og komme fram til et entydig helhetsbilde⁶. Digital kulturarv er fortsatt et område preget av mange mindre initiativer, noen store initiativer og lite faktisk koordinering på tvers av organisasjoner, regioner og nasjoner. Noen land kan henviser til en større grad av koordinasjon, og Nederland er et eksempel på dette.

⁶ http://ec.europa.eu/information_society/activities/digital_libraries/experts/mseg/reports2010/index_en.htm

Digitalisering av kulturarv i Nederland

I Nederland finnes det mer enn 250 kulturarvinstitusjoner. Digitaal Erfgoed Nederland (DEN Foundation) er et institutt under det nederlandske kulturdepartementet som utelukkende jobber med oppbygging, opplæring og koordinering av digitaliseringsprosjekter i kulturarvsektoren. Gjennom DEN er Nederland blant de få land i Europa som har en relativt detaljert oversikt over digitaliseringsaktiviteter på dette feltet («De Digitale Feiten» 2010). Ifølge innrapporteringen fra Norge mangler en slik oversikt her til lands per i dag⁷.

Rapporten til DEN inneholder mange interessante observasjoner og forslag og er tilgjengelig online, men dessverre bare på nederlandsk. Derfor sammenfatter vi hovedpunktene her:

- Som bakgrunn for rapporten er ca. 245 kulturarvinstitusjoner blitt spurt om tilbakemelding, og av disse har 119 institusjoner (som blir omtalt som en representativ gruppe) gitt tilbakemelding om tilstanden i digitaliseringen.
- Gjennomsnittlig investering i digitalisering ligger på ca. €100.000 (ca. NOK 800.000) per år og institusjon.
- Ca. 4% av totalbudsjettet innenfor disse institusjonene går til digitalisering.
- Ca. 26% av institusjonenes samlinger er blitt digitalisert. Her finnes det store forskjeller: Samlinger med bildemateriale er stort sett digitalisert (65%), mens audiosamlinger (6%) og gamle trykksaker av ulike slag (0.3%) er nesten ikke digitalisert i det hele tatt.
- Arkiver og biblioteker har gjort store deler av sine

digitale samlinger tilgjengelig gjennom internett. Museer gjør det derimot i liten grad, og tilbyr ofte ikke mer enn tilgang til ulike beskrivelser gjennom lokale PCer i sine utstillinger.

- Arkivene melder at store deler av deres samlinger ikke trengs å digitaliseres, mens museene angir at de gjerne vil gjøre størsteparten av sine samlinger tilgjengelige.
- 75% prosent av alle digitaliserte samlinger er tilgjengelig gratis og online(!), 6% av disse etter at man har registrert seg som en bruker. Bare 5% av samlingene er tilgjengelige mot betaling.
- Nesten ingen av institusjonene i Nederland bruker strukturerte metoder for registrering av kostnadene til digitalisering.

Europeana

<http://europeana.eu>



Det EU-finansierte tematiske nettverket Europeana har som mål å bygge opp ekspertise i digitalisering av

⁷ http://ec.europa.eu/information_society/activities/digital_libraries/doc/recommendation/reports_2010/norway.pdf

kulturarv og bygger på EDL Net og erfaring fra The European Library. Som en del av prosjektet har man utviklet en prototype som aggregerer og publiserer mange digitale kulturarv-objekter fra institusjoner fra hele Europa. Fra Norge deltar ABM utvikling og Nasjonalbiblioteket, og en del av NB's samlinger er også tilgjengelig gjennom Europeana. Europeana inneholder mer enn 6 millioner objekter fra hele Europa, med et mål for 2010 på 10 millioner digitale objekter.

Europeanas prototype har til nå vært basert på relativt standard materiale, og mange av samlingene som har blitt tilgjengeliggjort er konvertert manuelt. I mange tilfeller har dataene blitt levert på disk eller via FTP, og deretter konvertert til formatet som Europeana bruker.

Et problem med den måten Europeana er bygd opp på er at de forskjellige samlingene ikke har noen felles referansepunkter. Åndsverk blir beskrevet med tekststrenger i XML-felter som ikke er entydig brukt av forskjellige institusjoner. Felten «creator» er i mange tilfeller blitt brukt for å angi navnet på artisten, men f.eks. Scran (UK) har i sitt bidrag til Europeana brukt «creator» feltet slik:

creator: «David Steel subject David Newell Smith photographer; Hulton Getty»

Her refereres det ikke bare til en artist, men også til subjektet som er avbildet på bildet. En slik fremgangsmåte har sin årsak i at de personer/institusjoner som jobber med slike vedtatte metadata-skjemaer ofte ikke føler at de kan legge inn nok informasjon, alternativt at de ikke har samme tolkning av feltene slik de var definert da skjemaet ble laget. Det kan også oppstå behov for utvidelse av en modell på et senere tidspunkt.

Dette innebærer at Europeana har problemer med å

fortelle deg hvilke ressurser som er laget av en bestemt artist. Et søk på «creator:David Smith» gir mange resultater, men det er basert på søk i creator-feltet i basen. Siden det finnes mange Smiths og Davids får man en blanding av alt som på noen måte har med navnene David Smith å gjøre. Med andre ord: Europeana sliter med samme problemet som de fleste arkiver som skal sammenslås: Et stort fokus på beskrivelsen av selve åndsverkene gjennom veldefinerte felter (Dublin Core, Asta, XMP, etc.), men ofte blir ikke personene, stedene og hendelsene laget som egne objekter med unike referanser. En slik oppbygning av metadata innebærer at man ikke kan filtrere ut alle objekter relatert til en navngitt person, men at man istedenfor må bruke søketeknologi for å finne alle forekomster av en tekststreng i ett eller flere felt i basen. Dermed er de relasjonene som

eksisterer mellom åndsverk ofte basert på sammenligning av tekststrenger, som igjen medfører mange unøyaktigheter og feil. For å unngå de mest åpenbare problemene med stavemåter ved navn osv., bruker mange aktører autoritetslister. En slik fremgangsmåte er dessverre ingen god løsning på problemet, tenk bare på de store internasjonale forskjellene på hvordan man f.eks. staver «Peter Tchaikovski». Europeana klarer ikke å entydiggjøre søkene etter slike navn på lik linje med portaler som bruker semantisk teknologi (jmf. <http://www.evri.com>). Det samme problemet dukker opp i de fleste arkiver når det gjelder stedsnavn, hendelser mm.

Europeana bruker ikke URIs, men søkestrenger. Dette er en direkte følge av problemene med metadata som stammer fra forskjellige steder og som ikke er «løftet» opp på et nivå med tydelig semantikk og klare referanser (URIs) for hvert objekt.

Het Geheugen van Nederland

<http://www.geheugenvannederland.nl>



Illustrasjon 10: Het geheugen van Nederland

Het geheugen van Nederland («Nederlands hukommelse») har som visjon å bli en Europeana på nasjonalt nivå, dvs. en aggregator for alle digitale kulturarvsamlinger i Nederland.

Metadata i «het Geheugen» er basert på en relasjonell database modell, men det brukes ingen URler som definert av W3C for å kunne identifisere objekter i basen på et globalt nivå. I denne portalen kan man navigere fra åndsverk til personer og samlinger på samme måte som i Europeana (dvs. ved hjelp av en søkemotor).

Portalen er meget rik, og har mange forskjellige samlinger allerede integrert. Et meget bredt spektrum av kunstobjekter, dagligdagse objekter, tekst, film og bilder har blitt samlet i portalen. Det gjenstår også i denne portalen å publisere hvert objekt gjennom en URI som gjør en unik identifikasjon mulig.

MultiMedian

<http://www.multimedien.nl>



Illustrasjon 11: Multimedians semantic search - capabilities

Et prosjekt som har tatt de ovenstående utfordringene på alvor er utformet i regi av stiftelsen Multimedien, og demonstrerer hvordan man kan lage metadata ved hjelp av semantiske annotasjoner istedenfor tekstbaserte metadata.

Gjennom å gjøre en grundig og ryddig jobb fra starten av har prosjektet klart å lage metadatabeskrivelser fra bla. samlingen til Rijksmuseum Amsterdam som både kan publiseres og slås sammen med andre baser som bruker samme referanser (f.eks. personreferanser).

Sammenligner man Multimedien med Europeana kommer forskjellene tydelig fram. Brukeropplevelsen i Multimedien er mye bedre, det kan filtreres og søkes på relativt kompliserte spørsmål som «gi meg alle åndsverk av personer som har samme stil som Pablo Picasso». Det som gjør Multimedien spesielt kraftig er relasjonene mellom alle objekttyper og bruk av

unike, globale identifikatorer for objekter av alle typer (åndsverk, personer, stedsnavn osv). En portal med en slik oppbygning har også et enormt potensiale for digital publisering på nettet, f.eks. som en del av Linked Open Data.

Både programvare og samlingene som ble brukt i Multimedien har nå blitt brukt i Europeanas «Thought Lab» demonstrator og prototype – som viser veien videre.

Beelden voor de Toekomst



Det nederlandske fondet FES (Fonds Economische Structuurversterking) har satt av et budsjett på 154 millioner euro (ca. 1,2 milliarder NOK) for å digitalisere det som kalles «den audiovisuelle hukommelsen i Nederland». Pengene er øremerket forebygging av tap av opptak som ligger på gamle og slitte video- og audiobånd. Totalt sett skal det digitaliseres 137.200 timer video, 22.510 timer film, 123.900 timer med audio og 2,9 millioner bilder i perioden 2007-2014. Store deler av dette materialet skal også tilgjengelig-

gjøres online.

Digitaliserte objekter får metadata annotasjoner, men disse er ikke egne objekter. Relasjonene mellom objekter i basen linkes gjennom søkestrenger og fører til vanlige resultatlister som igjen kan brukes til videre navigering.

Et interessant del-prosjekt innenfor Beelden voor de Toekomst er et åpent annotasjonsprosjekt hvor flere hundre bilder ble lagt ut gjennom samarbeidet *Flickr The Commons* og ga publikum muligheten til å legge til annotasjoner i bildefilene. Prosjektet ble en stor suksess og førte til flere tusen annotasjoner av bildene, hvor samtidig mange personer på bildene ble navngitt.

3.5 Hva skjer i Norge?

Det finnes mange forskjellige digitaliseringsinitiativer i Norge. De fleste digitaliserte samlinger er blitt publisert gjennom brukerrettede websider. Det finnes også noen samleportaler som arkivportalen, digitalt museum osv. som er listet opp nedenfor⁸:

Digitalt Museum

<http://www.digitaltmuseum.no>

Denne portalen gir en oversikt over samlinger i norske museer. Samlingene som er tilgjengelige viser gjenstander, bilder og kunst (fra september 2010). Digitalt museum skal ha innhold fra 25-30 institusjoner og omfatter mer enn 1 million digitale objekter.

Teknisk: utviklet av KulturIT, og data i portalen hentes fra Primus. Ingen automatiserte tilgangsmetoder implementert.



Arkivportalen

<http://128.39.57.20/arkivportalen-web/side/forside>

Portalen gir oversikt over materiale fra statlige, fylkeskommunale og kommunale arkivinstitusjoner. Portalen inneholder ca. 2,5 millioner dataposter og digitale kopier av ca. 50.000 protokoller og annet sentralt materiale. Portalen ble ferdigstilt i 2009.

Teknisk: utviklet av Mesan AS. Ukjent teknologi.



Digitalt fortalt

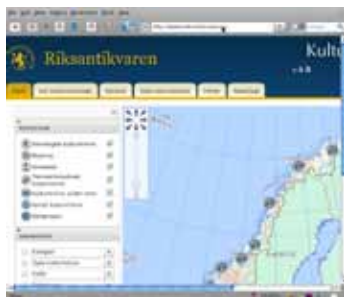
<http://www.digitaltfortalt.no>

Digitale fortellinger fra ABM-institusjoner og privatpersoner. I 2009 ble det lagret 1200 fortellinger, og målet for 2010 er 2000 nye fortellinger. Fortellingene skal bli knyttet opp mot innhold i bla. Digitalt Museum og Arkivportalen.

Teknisk: utviklet av ABM utvikling. Ukjent teknologi, men bruker databasesøk på ID i URLer for å vise objekter og under-objekter.



8 Original liste ble laget av Gunnar Urtegaard og Lars Wenaas – ABM utvikling – Digital ABM.



Kulturminnesøk

<http://www.kulturminnesok.no>

Omfatter ca. 115.000 poster med fredede bygninger, fredede kulturmiljø, verdens kulturarvsteder, kirker, arkeologiske kulturminner og kulturminner under vann. Åpnet høsten 2009.

Teknisk: informasjonen i portalen stammer fra Askeladden.



Universitetsmusea sin fotoportal

<http://www.unimus.no/foto>

Samlingene omfatter 1,6 millioner bilder. Flere hundre tusen er søkbare gjennom fotoportalen.

Teknisk: Ingen direkte tilgang til objektene, veldig tung bruk av Flash, innhold ikke tilgjengelig for søkemotorer eller andre former for automatisk prosessering. Google klarer ikke å indeksere unimus sine sider. Gjenbruk er utelukket i dagens implementasjon.



Arkeologiske samlinger ved Universiteta

<http://www.unimus.no/arkeologi/>

480.000 arkeologiske gjenstander er søkbare. Flotte brukeropplevelser gjennom animert innholdsformidling.

Teknisk: Ingen direkte tilgang til objektene, veldig tung bruk av Flash, innhold ikke tilgjengelig for søkemotorer eller andre former for automatisk prosessering. Google klarer ikke å indeksere unimus sine sider. Gjenbruk er utelukket i dagens implementasjon.



Universitetssamlinger mynter og medaljer

<http://www.unimus.no/numismatikk>

Inneholder samlinger med mynter og medaljer.

Teknisk: Ingen direkte tilgang til objektene, veldig tung bruk av Flash, innhold ikke tilgjengelig for søkemotorer eller andre former for automatisk prosessering. Google klarer ikke å indeksere unimus sine sider. Gjenbruk er utelukket i dagens implementasjon.

Europeana

<http://www.europeana.eu>

Europeana er et europeisk initiativ for å slå sammen flere digitale samlinger. Basen har en størrelse på 10 millioner objekter i 2010, hvorav ca. 500.000-1.000.000 stammer fra Nasjonalbiblioteket samt lokale og regionale ABM-institusjoner i Norge. Siden Norge er en vesentlig bidragsyter i prosjektet er EUROPEANA portalen også oppført under norske innholdsportaler i denne rapporten.

Teknisk: Dagens portal bruker stort sett vanlig teknologi (RDBMS-basert) hvor data fra flere bidragsytere er importert i en felles database som driftes av EUROPEANA sentralt. Prototypen for neste generasjon («Europeana Thought Lab») bygges imidlertid på teknologi fra Multimedias prosjektet og er et vesentlig skritt videre i retning interoperabilitet og samhandling gjennom bruk av URIs, semantikk og referanser mellom objekter.



EuropeanaLocal

<http://158.36.77.203:8080/portal/>

Målet med EuropeanaLocal er å oppnå sammenslåing av nasjonal kulturarv i mindre, lokale Europeana- lignende portaler som igjen samles i en stor europeisk fellesportal. EuropeanaLocal er et nettverksinitiativ hvor de forskjellige land har sine egne aggregasjons-prosjekter basert på samme tekniske plattformen som blir utviklet til Europeana. Tanken er at man slår sammen samlinger på nasjonal nivå i en Europeana Local portal, som i seg selv leverer resultatene videre til Europeana's fellesportal.

Teknisk: Europeana Local bruker den samme tekniske løsning som Europeana



Bokhylla.no og Nasjonalbiblioteket

<http://www.bokhylla.no>

<http://www.nb.no>

Nasjonal tjeneste i regi av Nasjonalbiblioteket. Flere arkiv og museum har registrert sine boksamlinger i nasjonale bibliotekssystemer og åpnet samlingene opp for søk på internett.

Teknisk: Løsningen til bokhylla er identisk med løsningen til Nasjonalbiblioteket. Det brukes både unik identifikasjon for objekter (URN:NBN) og en URI som fører direkte til objektet. Metadata lenkes ikke mot objekter, men er basert på tekststrenger.





Kulturnett.no

<http://kulturnett.no>

Informasjon om kulturinstitusjoner, kunstnere, kunstverk, nettressurser, artikler mm.

Teknisk: Løsningen integrerer «egen data» med tilgang til noen andre portaler. Kulturnett bruker i stor grad identifiserbare objekter og relasjoner mellom disse. Metadataene viser videre til unike objekter gjennom en ID (og ikke vanlig tekstsøk som f.eks. hos «Digitalt Fortalt»). Dette gir en fin navigasjonsopplevelse for brukere som utforsker på egen hånd, med entydige resultater (i motsetning til bruk av tekst-søk, som kan være upresise). Kulturnett bruker spørringer for å publisere informasjonssider, og disse er ikke egnet som URI.

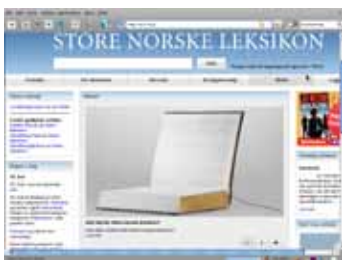


lokalhistoriewiki.no

<http://lokalhistoriewiki.no/>

7400 artikler med hovedsaklig lokalhistorisk innhold. Portalen er bygd opp som en wiki-løsning. Dette er en fin måte å organisere og krysslinke informasjon på som vi kjenner fra Wikipedia.

Teknisk: Løsningen publiserer artiklene gjennom en wiki. Artikkelsidene har en egen unik URI, men i denne implementasjonen inneholder den fortsatt ikke relevant teknisk informasjon (som f.eks. «index.php»).



Store Norske Leksikon

<http://www.snl.no>

Store Norske Leksikon er en portal som sikter mot en fortsettelse av tradisjonen med kvalitetssikret informasjon og kombinerer dette med publikums bidrag. Det er ikke rapportert hva omfanget til SNL er per mai 2010.

Teknisk: En egenutviklet portal basert på åpen kildekode. Systemet fungerer veldig likt vanlige wikier, men har implementert et skille mellom kvalifisert og ikke-kvalifisert data og forskjellige tilgangsnivåer. Det bruker URIs og metadata til å referere til andre identifiserbare objekter, akkurat som i en standard wiki.



Leksikonet wikipedia

<http://no.wikipedia.org>

Wikipedia bruker et standard wiki-opplegg som inkluderer URIer, lenking av metadata til andre objekter og mulighet for å søke i innholdet (tekst).

Teknisk: Wikiløsninger tradisjonelt bruker tanken med unik identifikasjon og lenking av metadata gjennom URI. Ved bruk av wiki-løsninger er det mulig å bruke åpen programvare for å lage semantiske abstraksjoner av wikien som kan publiseres på nett som en del av semantisk web (ref. Dbpedia.org som er en ekstrakt av wikipedia.org i et semantisk format.



Popsenteret

<http://www.popsenteret.no>

Popsenteret publiserer foreløpig ikke informasjonen sin i noen portal. Allikevel er de tatt med i denne listen med portaler, fordi

Popsenteret har valgt å bruke semantisk teknologi for sin driftsløsning (som inkluderer sluttbrukerstasjoner). Arkivsystemet er laget i lisensfri programvare basert på åpen kildekode, og forvalter metadata som RDF i en spesiell database. Basens innhold kan lett utveksles og deles med andre semantiske baser. I tillegg er import fra kilder som musicbrainz og dbpedia.org (som bruker samme teknologi) en enkel sak. Systemet hos Popsenteret er satt opp etter prinsippene som nevnt i denne rapporten, med et åpent backoffice system (arkivet) og grensesnitt tilpasset ulike brukeropplevelser. Disse grensesnittene kan være alt fra websider til interaktive multi-media stasjoner og «digitale signage» systemer.

Teknisk: bruk av Linux, Protégé og MySQL. Webservice med SPARQL for å få tilgang til informasjonen.

Portal	Publisert unik identifikator for objektene?	URI?	Metadata refererer til andre unike objekter?	Automatisert tilgang mulig?
Digitalt museum	Ja	Nei	Nei	ukjent
Arkivportalen	Ja	Nei	Ja	ukjent
Digitalt fortalt	Ja	Nei	Nei (tvetydig streng søk)	RSS/søk
Kulturminnesøk	Ja (location)	Nei	Nei	ukjent
Fotoportal til Universitetsmusea	Sannsynlighvis (class_id)	Nei	Nei	ukjent
Arkeologiske samlinger Universiteter	Nei	Nei	Nei	ukjent
Europeana Local	Ikke tilgjengelig			
Bokhylla.no / NB	Ja	Ja	Nei	ukjent
Kulturnett.no	Ja	Nei	Ja (ID søk)	ukjent
Lokalhistoriewiki.no	Ja	Ja (men trenger opprydding)	Ja	Ukjent
Store norske leksikon	Ja	Ja	Ja	Ukjent
Leksikon Wikipedia (NO)	Ja	Ja	Ja	Ukjent
Popsenteret	Ja	Ja	Ja	Ja
Europeana	Ja	Ja	Nei (Prototype Ja)	Ja

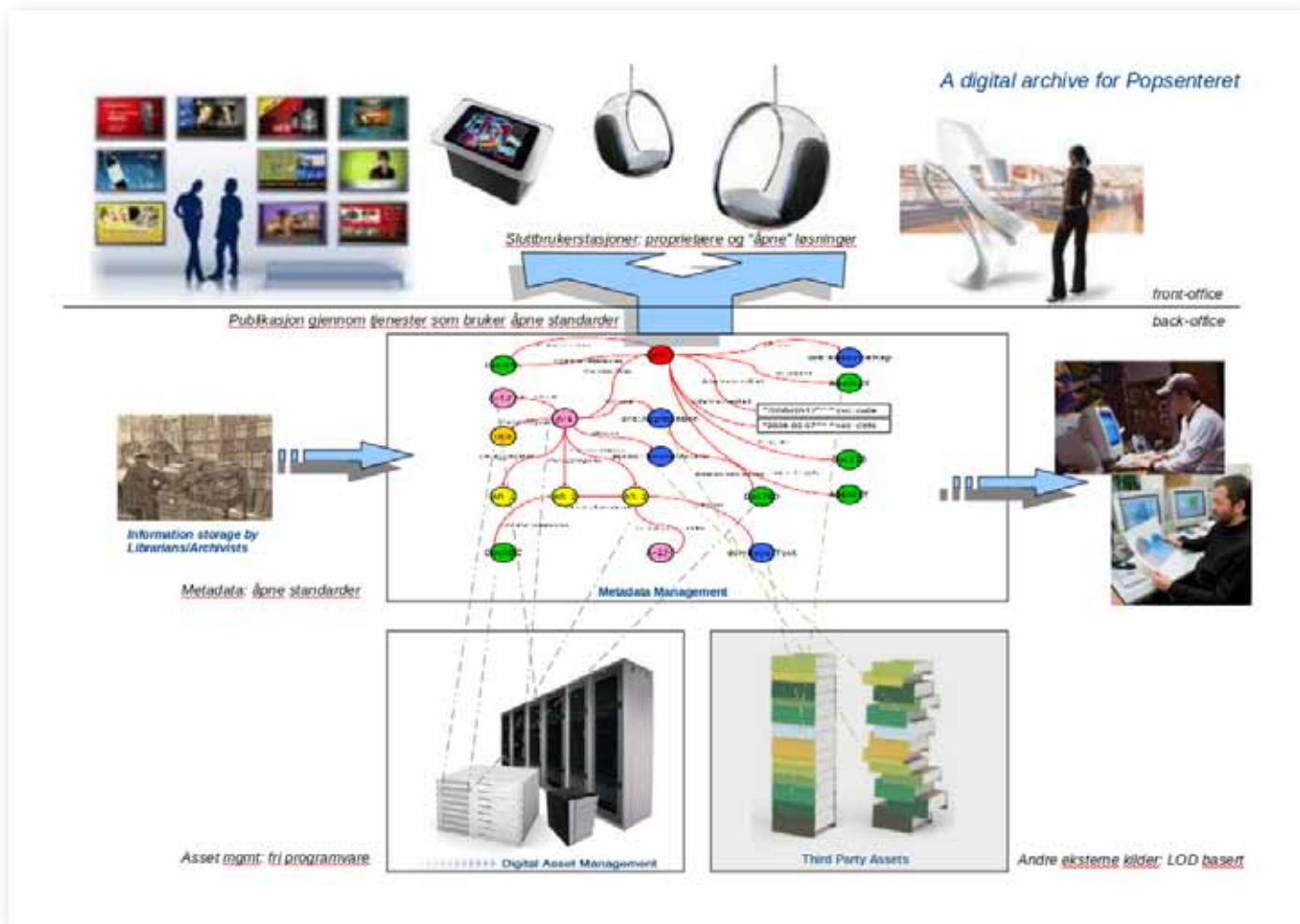
Tabell 1: Sammenligning av forskjellige portaler i forhold til fire spørsmål (publiseres det nøkler til objekter, URler, inneholder metadataene referanser til andre objekter og mulighetene for automatisert tilgang til dataene gjennom web services o.l.).

3.6 Utfordringer ved dagens satsing på digitalisering av kulturarv

- Man baserer seg i for stor grad på en typisk arkiv/bibliotek-tanke, som gjør at fokuset gjerne ligger på åndsverk, og man tar ikke i samme grad hensyn til hendelser, aktører og steder. I mange arkivportaler er for eksempel en person som er forfatter fortsatt bare representert av en tekststreng med et navn, og ikke et refererbart objekt. Det samme gjelder steder og hendelser/datoer. Det er viktig å ta nettverkstanken innover seg når det gjelder å utforske metadata, slik at man kan begynne i en hvilken som helst ende og komme fram til interessante steder. Dagens tanke bak mange portalløsninger er at gjenfinning er viktigere enn utforskning. Det er ofte viktig at man vet hvem (aktøren) eller hva (objektet, tidsperioden etc.) man er interessert i, ellers får man lite ut av en portal. En slik fremgangsmåte er ikke særlig hjelpelig og tillater bare i liten grad å finne relasjoner og forbindelser som ikke er åpenbare (f.eks. hvilke andre malere som laget kunst i samme by som Picasso da han malte «Guernica» i 1937).
- URler begynner å bli brukt som pekere til åndsverk og objekter (ref. Koninklijke Bibliotheek, Nasjonalbibliotek, etc). For å bli brukt i stor skala burde slike portaler helst ha en visjon om varighet (at man f.eks. garanterer at et bestemt objekt blir tilgjengelig/referert til under denne URlen i de neste 100 årene).
- Open Access er en stor utfordring. Mange begrensninger på tilgang og bruk gjør en relativt god start vanskelig å videreføre.

- Metadata i nesten alle løsningene (inkl. Europeana) refererer ikke fra objekt til objekt, men beskriver et objekt og lar søketeknologi lage (dynamiske og ikke-kvalifiserte) lenker. Det er viktig at utgangspunktet for internett (lenker mellom dokumenter) som har ført til dagens visjon om «the internet of things» følges opp. Dette innebærer at man lager digitale representasjoner av aktører, hendelser, objekter, steder osv. som det aktivt refereres til når en (digital) uttalelse framføres. (Mao. skal metadata referere til unike og globalt identifiserbare objekter.)
- Standardiserte «ontologier» for metadata i ABM-sektoren blir publisert kun i liten grad. Den mest kjente er Dublin Core (DC). DC er relativt generell, og passer på langt fra alle objekttyper (steder, aktører, hendelser, film, bilder, lydfragmenter osv).
- Alt for mange prosjekter setter av alt for lave budsjetter til å lage og kvalitetssikre metadata.
- Så lenge det ikke finnes veiledning og klare krav til ABM-organisasjoner som ønsker å starte opp med digitalisering, blir det vanskelig å delta i den globale «sammensmeltningen» av digital kulturarv.

Spesielt dette siste punktet gjør at vi foreslår at man lager en veiledning og et rammeverk for digitaliseringsprosjekter som er løse nok til å reflektere forskjellige behov innenfor ABM-sektoren, men som også er strenge nok til å kunne lede til noen minimumsgarantier.



Illustrasjon 12: Driftsløsning hos Popsenteret basert på prinsipper som omtalt i ABM rapporten

4. DIGITALISERT KULTUR- ARV OG FORMIDLING: HVORDAN TILRETTE- LEGGE ICT ARKITEKTUR FOR FREMTIDEN?¹

Begrepene **back-office** og **front-office systemer** refererer til et mye brukt perspektiv på software-utvikling hvor systemene er bygd opp lagvis. En tjener (server) har en sentral rolle i forvaltning av og tilgang til informasjon. Servere er som regel sikret og står på bakrommet («back-office»), mens flere mindre (eksterne?) maskiner kommuniserer med den samtidig. «Front-office» maskiner står ute på arbeidsgulvet, hos publikum, er mobile eller bruker vanlige web-browsere for å få tilgang til systemene på «back-office». Design, det visuelle inntrykket og interaksjonsmulighetene (knapper, sensorer) er alle en del av front-office.

«Det som ikke ses, finnes ikke»

Georg Berkeley

Allerede på syttitallet ble de første digitaliserte systemene utviklet i ABM-sektoren. Disse systemene førte til en ny type tjenester mot publikum, hvor det ble mye enklere og raskere å foreta bestemte søk innenfor en samling eller et arkiv og få opp digitaliserte kartotek-kort. Noen av systemene hadde mulighet for søk i eksterne baser på et lukket nettverk av datasystemer.

Da det ble mer vanlig med kontor-automatisering på begynnelsen av 80-tallet, ble terminalene gradvis erstattet med «personal computers», eller arbeids-

¹ Deler av dette kapitlet er overtatt fra og inspirert av «Business Model Innovation – Cultural Heritage» som er utgitt under en Creative Commons lisens, samt at vi har fått godkjenning av forfatterne.

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize.

Sir Tim Berners-Lee, 1999

Tim Berners-Lee står bak oppfinnelsen av internett slik vi kjenner det i dag. Han leder standardiseringsorganisasjonen World Wide Web Consortium (W3C) og står bak mange av tankene bak sidene som publiserer «open data» som data.gov.uk.



Illustrasjon 13: Papirbasert kartotek. Foto: Wikipedia



Illustrasjon 14: MAP2 - Samtidsmuseet i Roskilde 2009 - søk i arkivet

stasjoner med mer kapasitet. Dermed ble det også mulig for mindre ABM-organisasjoner å åpne opp sine samlinger digitalt. Mange ABM-organisasjoner designet og utviklet egne databaser, ofte basert på standard og/eller spesiallaget software. Slik utvikling baserte seg ofte på egne erfaringer og eksisterende (kartotek-) systemer. Etterhvert, når også «flat-bed scanners» og digitale kameraer ble tilgjengelige, ble det vanligere å digitalisere bilder selv og knytte dem opp mot beskrivelser av objekter i databasen. Ved siden av slike databaser av metadata om samlingene, ble det mer og mer vanlig å lage kunnskapsbaserte innholdstjenester som referanselister, indekser, bilder, tekstbaser og utstillingskataloger.

På nittitallet kom internettet, og dermed eksploderte mulighetene for informasjonsutveksling. ABM-institusjoner som var knyttet til webben ble plutselig del av et globalt kunnskapsnett. «World Wide Web» tilbød standardisert, enkelt tilgjengelig teknologi som et uniformt og integrert grensesnitt mot heterogene informasjonssystemer.

Kulturarv og teknologi – dagens tilstand

Som et resultat av den raske teknologiske utviklingen siden tidlig på åttitallet finnes det mange veldig forskjellige IKT-løsninger i ABM-sektoren. På syttitallet ble det utviklet databaser som skulle «overta» for kartoteker og gjøre dem tilgjengelige i digitalt format. Løsninger som SIFT («Søk I FriTekst» - et mye brukt system i Norge utviklet av Norsk Data) finnes fortsatt i en del institusjoner, selv om innholdet er rent filbasert og ikke-relasjonelt, og bruken er begrenset til søk i kartotekene. Slike systemer er ikke lett å bruke sammen med eller integrere i andre systemer. Spesielt SIFT-lignende, tekst- og filbaserte databaser er det vanskelig og dyrt å modernisere, siden grunnleggende relasjoner mellom informasjonselementene ikke finnes og må legges til (semi-) manuelt. Slike baser finnes fortsatt i mange organisasjoner.

Siden Relational Database Management Systems (RDBMS) ble kommersialisert tidlig på åttitallet, har man sett en stadig

utvikling av informasjonsbaser som dekker et spesifisert behov innenfor de institusjonene som de ble laget for. Denne utviklingen er preget av spesialløsninger og tilpasninger. Felles for alle disse løsningene er at de er fra «pre-internett» tiden (dvs. før 1992-93) og ble utviklet på helt andre premisser og under helt andre forutsetninger enn det som skulle bli virkeligheten bare noen få år senere.

Og fremtiden kom raskt. Det oppsto et stadig større behov for kopling av disse egenutviklede løsningene til web-grensesnitt. Det viste seg i mange tilfeller å være vanskelig, dyrt eller umulig. Integrasjon med andre systemer som var eller ble utviklet var ofte heller ikke mulig, ble dyrt eller var vanskelig. Dermed ble det for brukerne ofte en hverdag med mange forskjellige systemer, ofte gjennomført som prosjektbasert utvikling med et eget mål som dekket et spesifikt behov. Til slutt endte man opp med en jungel av systemer og et forvaltningproblem.

En direkte følge av utviklingen av slike hybride og sameksisterende IKT-systemer er ofte en lav kvalitet ved de ulike digitale løsningene i en organisasjon. Det kan være vanskelig å hente ut informasjon fra forskjellige systemer, noe man prøver å løse med felles spørre-standarder som SRU.

Som følge av disse tekniske utfordringene finnes samlinger ofte ikke online i det hele tatt (eller bare delvis), selv om de faktisk har blitt digitalisert. Og selv om de er tilgjengelige online, så er det ikke dermed sagt at man kan integrere dem i Web 2.0 eller Web 3.0 (semantisk web) løsninger.

Den videre utviklingen av IKT-arkitekturer må derfor ta hensyn til at det kreves en mye større fleksibilitet og (semantisk) interoperabilitet i dagens og fremtidens løsninger. Denne tanken stemmer godt overens med tanken som formidles i *St. meld. nr. 24*

(2008-2009) *Nasjonal strategi for digital bevaring og formidling av kulturarv*. I denne meldingen står det at:

"Visjonen for regjeringens IKT-politikk på kulturfeltet er å gjøre mest mulig av samlingene i våre [arkiv, bibliotek og] museer tilgjengelige for flest mulig ved fremtidsrettet bruk av IK-teknologiske løsninger. Samlingene skal være søkbare og tilgjengelige på tvers av hele abm-feltet, og innholdet skal formidles på en brukerorientert måte."

Vi forslår at denne tanken løftes fra å ha nasjonale ambisjoner til å være rettet mot den internasjonale utviklingen.

Drivkraften bak den raske utviklingen innen IKT-sektoren er internettet. ABM-institusjoner står nå ved et veikryss. Med dagens hybride IKT-miljøer er det både tungvint og dyrt å bruke, vedlikeholde og videreutvikle alle de forskjellige systemene som ikke er godt integrert.

Standardisering, bruk av generisk infrastruktur og samarbeid mellom institusjoner gir muligheter for bedre balanserte kostnader, større uavhengighet og en bedre utnyttelse av IKT-infrastruktur.

Fornuftige løsninger

Det krever kontinuerlige investeringer for å få til en verdiskaping ved hjelp av IKT-infrastruktur. Det finnes flere faktorer som spiller en rolle, men de viktigste dimensjonene innen digital kulturarv - sektoren er sannsynligvis kostnadsbesparelse og verdiskaping.

Gjennom utvikling av en god IKT-arkitektur kan en organisasjon redusere sine kostnader gjennom økt effektivitet. Samtidig må det investeres for å modernisere og opprettholde standarden i systemet, slik at verdiskapingen kan forbli stor. Det er ofte modernisering og oppgradering av eksisterende løsninger som

utsettes og utsettes. En formodentlig årsak til dette kan være at man har bestilt eller laget spesielle «one-off» løsninger, som ikke har tatt nok hensyn til standardisering.

Bruk av standarder og generisk teknologi

Kulturarv-sektoren sitter på mye unikt og ofte sjeldent materiale. Det er ofte disse unike objektene som gir en institusjon eller organisasjon sin identitet eller eksistensberettigelse. Det er også nettopp disse unike objektene som utgjør den tilleggsverdien som organisasjonen kan tilby i et digitalt informasjons-samfunn. Hvis en organisasjon velger å ikke gjøre sine kulturarv-objekter tilgjengelige for andre, så er det få andre måter man kan bli kjent med objektene på.

Samtidig vil kulturorganisasjoner gjerne beholde sin egen identitet og skille seg fra andre. Utviklingen av digitale tjenester er en spesielt god måte å få oppmerksomhet på. Det er en interessant observasjon at de organisasjonene som blir sett på som mest innovative, er nettopp de som har innsett at man må kunne tilby mer enn brukerne forventer. Samtidig er det nødvendig å vise fram hva man har til et så bredt publikum som mulig. Jo flere som ser på og erfarer om (deler av) samlingene, jo mer relevante blir disse samlingene. Dette innebærer publikasjon av informasjon om samlingens objekter utenfor organisasjonens vegger. Verdien av en samling øker betraktelig når den blir tilgjengeliggjort for mangesidig bruk av mange.

Kanskje det viktigste elementet for å få til en spredning av informasjon om ulike samlinger er at teknologien tilrettelegger for det. Dette innebærer at man må tenke standardisering av IKT ifølge åpne og helst globale standarder. For å selv beholde et unikt konsept og ansikt utad er det fullt mulig med differensiering gjennom egenutviklede og spesiallagde «front-office»

systemer. Selve «back-end» (eller «back-office») skal/burde derimot standardiseres så mye som mulig, slik at man har mulighet for fleksibel (gjen-)bruk av informasjon innenfor og utenfor organisasjonen. Brukerne krever tilgang til mer informasjon enhver organisasjon har tilgjengelig, uansett hvor mye det er. Brukerne krever tilgang til flere baser og informasjonskilder samtidig, og det er ofte uvesentlig for henne hvor informasjonen kommer fra (web, lokalt osv). Søk av informasjon på tvers av systemer er viktig og kan bare implementeres ved hjelp av standarder. I tillegg er det viktig at resultatene av søket kan slås sammen etterpå. Dette innebærer at resultatene av et søk må inneholde gode annotasjoner som setter objektet i riktig kontekst og muliggjør unik identifikasjon.

Popsenteret i Oslo har valgt å bruke en slik oppdeling for sitt arkiv og sin driftsløsning (ref. Illustrasjon 12) som baserer seg på nettopp bruk av åpne (W3C) standarder og proprietære front-office systemer. Erfaringene har vært positive så langt, og mye data om norsk populærmusikks historie er blitt lagt inn og er søkbar, navigerbar og publiserbar ved hjelp av unike identifikatorer og relasjoner mellom objekter i tråd med anbefalingene i denne rapporten.

Dessverre mangler det ofte en slik standardisering hos mange kulturarv-organisasjoner, og det samme gjelder kommunikasjonsprotokoller mellom systemer samt datamodellering. Det har selvsagt mye å gjøre med dagens IKT-standard innenfor sektoren, og fører til at man taper mye, spesielt når det gjelder tilgjengelighet av informasjon. Det er f.eks. fortsatt rimelig vanskelig å finne digital kulturarv gjennom de største søkemotorene som Google og Yahoo. Dette er klart en godt dokumentert effekt av mangel på standardisering (van Kersen, 2004).

Det er vanskelig å tilpasse digitale tjenester til

endrede behov hos sluttbrukerne, hvis tilpassinger i IKT-systemene må skje på en ad-hoc basis. Man kommer til å trenge dyr spesialkunnskap for å videreutvikle systemene hvis man ikke klarer å linke informasjon på tvers av informasjonssystemer ved hjelp av (åpne) standarder.

Et viktig mål for fremtiden må være å **gjøre metadata og datamodeller som gjelder kulturarv tilgjengelig på en standardisert og åpen måte**, slik at gjenbruk blir stimulert. I den ideelle situasjonen forvalter kulturarv-organisasjoner sine data selv, slik at kvaliteten blir garantert og brukerne kan henvende seg til kilder man kan stole på for å få den kvalitet man ønsker.

Siden brukere og utviklere på internett har kastet seg over integrerte applikasjoner, mye som følge av Web 2.0 utviklingen, er ikke spørsmålet *om* informasjonen på internett gjenbrukes av andre, men spørsmålet er *hvem er med* å levere data til publikum. Her ligger det klart en rolle for kilder av høy kvalitet som er betalt av og bygd opp ved hjelp av offentlige midler.

For å **tilrettelegge innhold for deling og interoperabilitet mens hver institusjon kan ha sin egen identitet** anbefales det at det kreves at løsninger for ABM-sektoren har et klart skille mellom front-end og back-end systemer som kommuniserer ved hjelp av protokoller definert som (åpne) standarder.

Et slikt skille gjør det mulig å gjenbruke innhold i systemer gjennom standardiserte protokoller, på lik linje med HTML og HTTP som tillater weblesere å gjengi websider på en standard måte. I tillegg kan bruk av åpne standarder gjøre IKT billigere ved å utvikle funksjoner i fellesskap som ikke trengs å utvikles separat. Enda viktigere er det at man, gjennom en slik standardisering av tilgang, kan åpne opp mer og større deler av de digitaliserte samlingene.

En vel så viktig men uavklart utvikling for kulturarv-sektoren innenfor IKT er den mye omtalte satsing på «cloud computing». «Cloud computing» tilbyr storskala, skalerbare og online tjenester, hvor selve forvaltningen av disse tjenestene er sentralisert hos noen store bedrifter som Google eller Amazon. For mange tjenester innenfor ABM-sektoren er det ikke aktuelt å legge ut sin data i «skyen», mest fordi det er uavklart hva dette innebærer i forhold til rettigheter, kontroll og problemer med å ha viktige nasjonale ressurser utenfor landets grenser.

For andre tjenester er det ikke så problematisk, og man har allerede sett at kulturarv-institusjoner har tatt i bruk cloud baserte tjenester for å a) heve kvaliteten på sine samlinger som i «Beelden voor de Toekomst» som bruker Flickr for å få metadata lagt til på sine bildesamlinger (Moortgat, 2009) og b) nå ut til publikum med tjenester som er like de opplevelsene som brukerne allerede er vant med fra andre hold (iPhone, internett, etc).

Det som er interessant er at IKT er blitt så modent og generelt tilgjengelig at man kan lage gode storskala digitale tjenester uten dyp kunnskap om IKT. Suksessen er for en stor del blitt til pga. åpne standarder og generisk teknologi.

Samhandling og identitet

Standardisering og generisk teknologi bidrar til verdiskaping gjennom fleksibel gjenbruk av digital kulturarv og gjennom kostnads-reduksjon for tjenester. Samarbeid har blitt et kjernekonsept, også når det gjelder IKT-infrastruktur. Både sosiale og teknologiske utviklinger har gjort det vanskelig for kulturarv-institusjoner å jobbe uavhengig. Man har blitt del av et stort digitalt økosystem i en region, sektor, det offentlige eller internett, det globale kunnskapsnettverket.

En observasjon er at det er ett område hvor institusjoner stoler mye på hverandre, og det er når man skal sette bort arbeidet med langtids-oppbevaring av digital data. I flere europeiske land ser man at institusjoner går sammen for å danne felles fasiliteter for dette formålet.



Illustrasjon 15: Søk på "edvard grieg" i Europeana.eu

I alle våre samtaler og diskusjoner er det én sak som ble gjentatt flere ganger. Det som er sett på som meget sentralt og viktig i den videre utviklingen av digital kulturarv - sektoren og samarbeid mellom partnere er noe som heter «persistent identifikasjon» av digitale kulturarvobjekter. Dagens praksis hos mange organisasjoner hvor det f.eks. brukes tekst-baserte «autoritetslister» for personnavn og stedsnavn er ikke tilstrekkelig. Det er viktig at alle ting på en slik liste refereres med en unik identifikator.

Det er i alles interesse å sørge for at alle objekter har **unike og permanente navn/identifikatorer** som publiseres og brukes av portaler/tjenester. Dette er den beste og enkleste måten å tilrettelegge på for

fremtiden.

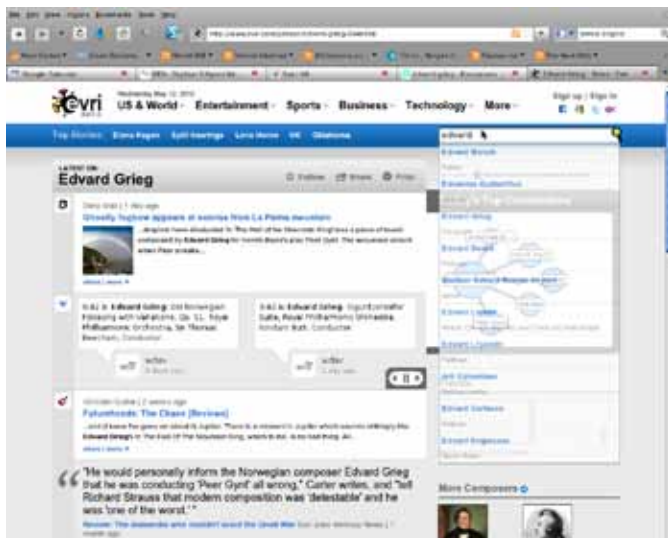
Får man til en konsensus om et slikt rammeverk, blir det mulig for alle å hente inn, slå sammen og publisere data, samt at alle har muligheten til å bidra med ny kunnskap eller nye uttalelser om unikt identifiserbare objekter. En slik permanent identifikasjon tillater «long term traceability», men viser også opprinnelsen til objektet i tilfeller hvor det finnes flere like digitale objekter. En forutsetning er at alle «persistente identifikatorer» er unike, og det finnes flere toneangivende forskningsinstitusjoner som har kommet med forslag om hvordan dette kan garanteres.

Eksempel: La oss si at man vil vite mer om Edvard Grieg og hans verker. Går man til ABM-sektoren sine verktøy og søker på «Edvard Grieg» på Europeana.eu, så får man som første resultat «Edvard Grieg +» av en ukjent forfatter og levert av «Narodna in univerzitetna knjižnica/Slovenia». Man får ingen beskrivelse av selve mannen, og intet bilde som viser ham. Man får heller ikke noen god oversikt over hvilke ressurser om ham som er tilgjengelig (se Illustrasjon 15).



Illustrasjon 16: Søk på "edvard grieg" på sidene til Nasjonalbiblioteket

Foretar man samme søk på nb.no, blir resultatet likt. Denne gangen får man heller ikke tilbakemelding om at han faktisk er en person med en historie. Det første resultatet er en bok «Edvard Grieg, 1843-1907: et idehefte for grunnskolen». Selve metadataene i begge portalene (som på mange måter representerer dagens «state of the art») baserer seg på tekststrenger og tillater ikke at man identifiserer personer på en unik måte (se Illustrasjon 16).



Foretar man derimot samme søk på Evri.com, vil systemet identifisere søkeordene som en mulig komponist (se Illustrasjon 17). Portalen (som henter sin informasjon bla. indirekte fra wikipedia) vet at finnes en person som heter «edvard grieg» og at denne personen har som rolle «komponist». Dermed kan brukeren enkelt skille flere personer fra hverandre. Har brukeren selektert en person, kan systemet automatisk filtrere bort all informasjon som har syntaktiske likheter, men som ikke er relatert til den personen brukeren er interessert i. Interessant er også den grafiske gjengivelsen av nettverket av metadata rundt denne personen. Edvard Grieg har her forbindelser til Peer Gynt, Ole Bull, Henrik Ibsen, Norway og «in the Hall of the Mountain King», alle unikt identifiserbare objekter i seg selv med metadata og igjen andre objekter knyttet til seg.

Sammenfattet

Forskjellige tilnærminger medfører behov for endringer som er nødvendig å gjennomføre for å bli «fremtidsrettet». Det er hovedsaklig to faktorer som står sentralt i denne diskusjonen:

- 1) hvorvidt ABM-institusjonene har et ønske å forvalte sin IKT-teknologi selv,
- 2) relasjonene mellom intern IKT-infrastruktur og delte tjenester.

Hver institusjon kan lage seg en distinkt profil gjennom å bruke eksklusive og spesialutviklede løsninger. Dette kan gi mye oppmerksomhet, men også store forvaltningsproblemer. Faren er at muligheten for gjenbruk i disse tilfellene ikke kan garanteres når teknologien blir utdatert eller leverandøren forsvinner fra markedet.

Bruken av åpne standarder og unike identifikatorer (URI) på alle metadata objekter i «back-office» systemer og bruk av generisk teknologi som langtidstrategi virker til å være det fornuftigste valget når man skal få uttelling for sine digitale investeringer.

5. VEIEN VIDERE: HVORDAN KAN VI LØFTE ABM-SEKTOREN OG HJELPE TIL MED Å TILRETTELEGGE FOR FREMTIDEN?

Det har vært en utvikling de siste årene hvor brukerne har blitt involvert i oppbygging av informasjonskilder og portaler, og digitaliseringsbølgen har sørget for at mange mindre men interessante samlinger har blitt gjort tilgjengelig for publikum. I en slik verden har ABM-sektoren behov for å fornye seg og finne nye veier videre. For mange institusjoner som har mindre samlinger og begrensede budsjetter, blir det ikke mulig å lage selvstendige, gode løsninger som skal nå et stort publikum. Derimot kan de veldig godt slå seg sammen med andre, komplementære samlinger og lage spennende tema/fagfeltbaserte opplevelser som formidler kunnskap. Dette betyr at man må kunne publisere, dele og integrere metadata om samlingenes objekter på en enkel og standardisert måte.

Viktig er det at man i denne fornyelsesprosessen ikke mister helhetstenkingen. Teknologier videreutvikles og endres, men det er blitt utviklet noen basisprinsipper for fremtidens informasjonsutveksling som man må ta hensyn til så tidlig som mulig i sine digitaliseringsprosesser.

Det er disse prinsippene om (automatisert) tilgang til og deling av metadata om objekter som står sentralt i fremtidens kunnskapsformidling.

Forvalt dine egne samlinger, men tilrettelegg for deling

Det er ikke hensiktsmessig å prøve å lage «sentrale datavarehus». Dagens teknologi tillater at informasjonen ligger på de stedene hvor kompetansen som vedlikeholder dem også ligger. Data trengs ikke flyttes hvis man ikke ser behov for det. Det viktigste er at man legger tilrette for deling av innhold på en måte som tillater at man klarer å identifisere objekter og relasjoner mellom objektene og andre baser.

Sørg for at basisprinsippene er på plass for kunnskapsdeling

Unik identifikasjon og automatiserte internett-tjenester som tillater at man «plukker» fra hverandres hyller.

Billigere å dele

Der hvor formidling av innholdet i databaser før var en jobb som krevde både tid og (mye) penger, har digitaliseringsbølgen blant annet innenfor kulturarv-feltet ført til en helt annen situasjon. Tidligere var det å tilgjengeliggjøre og distribuere en gjenstand (bok, bilde) fra et arkiv, bibliotek eller museum en ressurs- og kostnadsintensiv oppgave, men det blir mye enklere og billigere når ressurser blir digitalisert. Uten store kostnader til distribusjon og publisering blir det også enklere å nå et større publikum og samarbeide med andre organisasjoner/institusjoner.

Behovet for kvalifisert informasjon

Det er et stort behov for spesialisert og kvalifisert informasjon. Det blir viktig å kunne skille mellom informasjon man kan stole på og informasjon som sannsynligvis er subjektiv. Her har ABM-sektoren en rolle som kommer til å bli viktig i årene fremover.

Gi dine brukere tilgang til informasjon fra hele ABM-sektoren

Skal man bruke tid og krefter på noe, bør man helst bruke det på egen kompetanse for å bli enda bedre på sitt eget fagområde.

Kompletter opplevelsene du skal lage til dine brukere med kvalifisert informasjon fra andre organisasjoner. Ingen er best på alt, men hver organisasjon har sin egen spesialkompetanse.

Det finnes mange situasjoner hvor f.eks. Wikipedia ikke dekker behovet for kvalifisert informasjon, og publikum er klar over dette. Derfor er det viktig at ABM-sektoren presenterer kvalifisert informasjon på en lett tilgjengelig måte med så mange «innganger» som mulig. Det er bare noen få steder som klarer å bli

«the place to go» når man ønsker informasjon om et bestemt tema, og derfor er det meget viktig at man i fellesskap formidler denne tilgangen.

Eksempel: En person som leter etter informasjon om Rembrandt van Rijn ser sannsynligvis først på Wikipedia eller kanskje på sidene til Rijksmuseum. Det er få som har hørt om Rembrandtshuis, men det er sannsynligvis der man lærer mest om livet til Rembrandt. Hvordan kan vi sørge for at brukeren finner denne informasjonen?

Mest sannsynlig får man slik informasjon bare gjennom videreformidling fra andre portaler, og det er derfor riktig og viktig at man tilrettelegger ABMs kunnskapsformidlingsportaler slik at de kan ha en slik «veiviserfunksjon».

ABM-institusjoner trenger støtte til å definere og gjennomføre en digitaliseringsstrategi som henger sammen

Det finnes en del ABM-organisasjoner som skal foreta eller nettopp har begynt på en digitalisering av sine objekter/ressurser. Det viser seg ofte at man sitter med de samme spørsmålene angående «den riktige veien fram». Hva skal vi ta hensyn til? Hva må på plass? Hva skal vi prioritere? Må vi tilrettelegge/passe på noe mht. IKT-arkitekturen/satsingen vår før vi setter i gang med selve digitaliseringen? Det er ofte spesielt de mindre aktørene som ikke nødvendigvis sitter på all kunnskap rundt digitalisering av kulturarv. Maskinvare som trenges til digitalisering er dyrt, og derfor kan det være fornuftig å finne leverandører som overtar en del av jobben. Hvilke krav burde man stille til leverandørene?

5.1 Definerede krav for digitaliseringsprosjekter

For å få til en felles retning på digitaliseringen i f.eks.

kulturarv-sektoren og for å gi institusjoner og organisasjoner litt støtte i ryggen når man diskuterer løsninger, foreslår vi å bygge opp et sett med anbefalinger og minimale krav for digitaliseringsprosessen. Dette må oppfølges med veiledning og kurs, og burde helst koples til funding/insentiver for å bli tatt raskt i bruk. På europeisk nivå er det flere land som har begynt å tilby slik støtte i samarbeid med ABM-sektoren i hvert land. Erfaringene er så langt veldig positive, og de første “fruktene” kan plukkes (se f.eks. Europeana-prosjektet som aggregerer digital kulturarv fra mange forskjellige steder).

Forskjellige kravpakker for digitalisering av kulturarv

For å være forståelige og brukbare burde slike minimumskrav helst deles opp i forskjellige “pakker”. Det burde også være mulig å få ytterligere informasjon og hjelp gjennom veiledning og kurs. Basert på erfaringer rapportert av DEN⁹ foreslås det å starte med fem forskjellige pakker som inneholder følgende:

- krav om lokalisbarhet
- krav om tilvirking, lagring av digitale objekter
- krav om beskrivelse / metadata
- krav om digital varighet
- krav om presentasjon

5.2 Veiledning og kurs

Slike krav skal følges opp med veiledning og kurs. Siden bruk av åpne standarder anbefales, er en god del kvalifisert informasjon allerede tilgjengelig på internett, inklusive kurs og forelesninger. Denne informasjonen kan utvides med lokale kurs og hjelp, f.eks. gjennom ABM-utvikling.

5.3 Funding/insentiver

En mye diskutert sak er hvordan man klarer å koor-

dinere slike større, fremtidsrettede initiativer. Vi har diskutert dette punktet spesielt med alle institusjoner vi har snakket med og har fått mange forskjellige tilbakemeldinger. Kunnskapsdepartementet i Nederland har valgt å kople finansieringsmuligheter for digitaliseringsprosjekter opp mot noen minimumskrav (se Del II). Denne strategien har hatt som følge at institusjonene fikk tydelig tilbakemelding om hva som i alle fall måtte være på plass, og fikk hjelp til å evaluere og igangsette prosjektene. Denne strategien virker vellykket på mange fronter, med de første museene som har begynt å formidle samlingene på en felles måte.

På europeisk nivå er det valgt en strategi rundt Europeana. Europeana er en stor fellesportal som aggregerer nasjonalt innhold fra de deltakende land. Europeana fungerer som ett stort nettverk og insentivene for deltakere består av økonomisk kompensasjon samt en bedre innsikt i hvordan egne samlinger henger sammen med andre europeiske samlinger.

5.4 Deltagelse og diskusjonsfora: aktiv outreach!

For at slike «kravpakker» skal kunne fungere, er det viktig at man blir enige om kravene og ikke setter ambisjonene alt for høyt i begynnelsen. De må være mulige å gjennomføre selv for mindre institusjoner. En god strategi virker til å være at man inviterer til deltagelse i forskjellige diskusjonsfora, og at man oppretter noen ekspertgrupper på hvert område som skal komme med forslag til diskusjon. Disse forslagene burde så presenteres/diskuteres på nasjonale konferanser eller arbeidsseminarer på en aktiv måte.

Etterpå burde et organ som ABM-utvikling, sammen med Kulturdepartementet, formalisere og publisere kravene og informasjon om hvordan man kan bidra til diskusjon – og evt. hjelp til å komme i gang.

DEL 2

FORSLAG TIL MINIMALE KRAV FOR DIGITALISERT KULTURARV

I denne seksjonen blir det skissert noen minimumkrav som er foreslått som «rettningslinje» for bruk i ABM sektoren. Kravene i denne seksjonen er basert på anbefalinger og erfaringer fra Digitaal Erfgoed Nederland. Nederlandske myndigheter krever at et prosjekt eller en institusjon forholder seg minst til disse 5 krav for å få statlig støtte til digitaliseringsprosjekter. Flere

krav på andre områder blir diskutert. Det er viktig å huske at disse minimumkrav er laget av ABM-sektoren i samarbeid med departementet og at alle diskusjoner er åpent tilgjengelig for bidrag gjennom en wikiløsning⁷.

1. KRAV OM LOKALISERBARHET

Lokaliserbarhet betyr at man gjør objekter «digitalt tilgjengelige» gjennom standard teknologi, slik at objektene kan gjenbrukes av mennesker og maskiner uten forhåndsinformasjon.

1. **HTTP** er en standard protokoll som gjør informasjon tilgjengelig på web. Websider blir vanligvis publisert gjennom HTTP (HyperText Transfer Protocol).
2. **URI** gjør ethvert objekt i en samling identifiserbar på en unik måte (Unified Resource Identifier).
3. **XML** strukturerer metadata på en måte som er raskt og enkelt å interpretere av datamaskiner (eXtensible Markup Language).
4. **UTF8** er et internasjonalt kodesystem for karakterer (tegn). Ved bruk av UTF8 kan man gjengi internasjonale tegnsett.
5. **Dublin Core** er en metadata beskrivelsesstandard som blir brukt på globalt nivå for å beskrive samlinger av objekter og objektene i dem.
6. **OAI-PMH (eller W3C standarder?)** OAI standarden er laget for å gjøre overføringer mellom databasesystemer enklere. (Open Archive Initia-

tive Protocol for Metadata Harvesting). Digital Erfgoed i NL definerer denne standarden som et minimumskrav.

7. **SRU** SRU (Search and Retrieval via URL) protokollen gjør at man kan søke i forskjellige baser gjennom et standard grensesnitt basert på HTTP protokollen og spørresyntaksen CQL (Common Query Language).

Anbefalinger: Det anbefales å gjenbruke 6 av kravene (1-5, 7).

Siden krav 6 om OAI-PMH er tilrettelagt for bruk med HTTP og XML og det finnes nyere standarder på dette området, anbefales det å utrede nytteverdien av å bruke de nyere W3C internettstandardene som definerer formater for informasjonsutveksling (RDF/RDFS, evt. OWL).

Hvis man gjør databaser tilgjengelige på internett i formater som RDF/RDFS, blir det naturlig å bruke en web-service med spørrespråket SPARQL som generisk grensesnitt. SPARQL er et spørrespråk som er spesielt tilrettelagt for RDF/RDFS/OWL-basert

metadata. Gjennom disse to W3C standardene blir alle informasjonskildene som bruker dem «transparent» (gjennomsiktige) og kan adresseres og gjenbrukes direkte og automatisk fra hvor som helst på internettet. Pga. dette utgjør tilgjengeliggjøring av data gjennom en SPARQL-basert nett-tjeneste et interessant forslag om en utvidelse av ovennevnte krav. Det finnes muligheter for automatisk oversettelse av mange CQL spørringer til RDF/RDFS, slik at man kan forvente at en del av dagens systemer kan bygges om til bruk av SPARQL. Fordelen med bruk av SPARQL er at resultatene på spørringer fra forskjellige kilder lett kan slås sammen, siden representasjonsformatet er standardisert på tvers av domener og applikasjoner. Dette i motsetning til XML-baserte resultatsett som f.eks. CQL- implementasjoner leverer tilbake.

2. KRAV OM TILVIRKING OG LAGING AV DIGITALE OBJEKTER

Kravet om tilvirking definerer minimumskrav for digital reproduksjon av originalt analogt kulturarv-materiale. Dette garanterer at sluttresultatet av digitaliseringen har en optimal kvalitet og en informasjons-verdi lik originalens (innhold, struktur, formgivning) samt en fremtidig tilgjengelighet mot en akseptabel kostnad (dette gjelder både for produksjon og konservering). Selve prosessen med å komme frem til en digital reproduksjon burde også være en del av dette kravet.

I ABM-sektoren jobbes det med mange forskjellige objekter og typer. Det er også et stort sprik i kunnskap om teknikker til digitalisering og digitale formater. Gjennom et slikt «krav om tilvirking og laging av digitale objekter» kan sektoren stille bedre og mer begrunnede og rettede krav internt, men også mot leverandører. På denne måten blir det enklere å forsikre seg om at jobben blir utført riktig og at resultatet blir gjenbrukbart på tvers av sektorer og tid. Utgangspunktene burde være:

- **Tilvirk én gang:** sluttresultatet av digitaliseringsprosessen burde være en digital master. Det koster mest å ta fram originalen og digitalisere den, derfor er det viktig at den digitale masteren har en maksimum kvalitet med minimalt tap av informasjon. Denne masteren kan brukes til å skape alle mulige varianter til fremstillingen, men selve masteren røres ikke.
- **Konservering og bruk:** også en digital master må konserveres. Derfor er det viktig at man setter som minimumskrav at det skal brukes åpne standarder til dette formålet. Slike standarder er dokumentert, kjent og tilgjengelig for inspeksjon, noe som gjør at de mest sannsynlig også kan leses i fremtiden, evt. med formål om å konvertere til bedre alternativer hvis de skulle dukke opp. På denne måten sikrer man seg også at de fleste typer av senere bruk blir ivaretatt.
- **Erstatning:** en viktig diskusjon er om «kravet om tilvirking» skal kunne brukes i situasjoner

hvor man ønsker å erstatte analoge objekter med digitale, eller for å sikre seg mot tap av originalen gjennom ulike uhell. Det foreslås å la denne diskusjonen ligge, og heller konsentrere seg om kravene som gjelder kopiering med gjenbruk som formål.

- **Metadata:** Det finnes flere typer av metadata. I «kravet om lokalisbarhet» blir metadata definert som tilrettelegger for gjenfinning og sammenslåing av informasjon om objekter. Denne typen metadata kalles også «sosiale metadata». Metadata i «kravet om tilvirking» er metadata som beskriver de tekniske detaljer om objektet, som filnavn, kvalitet, oppløsning, format osv, og refereres derfor gjerne som «teknisk metadata».
- **unik, global identifikator (også som filnavn):** I praksis skjer det ofte at filer får samme navn (f.eks. DSC000478.JPG, eller innhold.txt). En slik praksis fører ofte til problemer senere i prosessen og det skal derfor være et minimumskrav at filnavnet er unikt.
- **sammensatte objekter:** Ofte er det behov for digitalisering av sammensatte objekter. Det kan være tekster med bilder, filmer av historier, men også malerier som finnes i en historisk verdifull liste eller annet objekt. Kompleksiteten i dette gjør at det foreslås å først og fremst komme til enighet om de andre kravene ovenfor før man kan jobbe med dette kravet.

Det er hensiktsmessig å dele opp kravet etter typen av materiale som skal digitaliseres. For hver av objekttypene foreslås det at en spesialistgruppe som består av personer fra det offentlige (ABM), men også noen (uavhengige) eksperter fra bransjen lager og vedlike-

holder kravene.

Det er viktig at man kommer med klare minimumskrav til objekter av type *Bilder, Tekst, Gjenstander (skulpturer, fysiske objekter i flere dimensjoner osv), Audiovisuelt materiale men også Geo-informasjon*.

Geo-informasjon er ofte viktig i kulturarv-sektoren og har tre praktiske tilpasninger: å kople romslig informasjon (spatial informasjon) til fysiske objekter (metadata), digitalisering av analogt kartmateriale og tilvirking av digitale kart med informasjonselementer plottet inn.

3. KRAV OM BESKRIVELSE

Tilgjengeligheten av gode beskrivelser av kulturarv er essensielt i et godt tjenestetilbud rundt formidling. Dessverre er mye av teknologien til audio- og bildesøk bare tilgjengelig som forskningsprototyper og kan ikke brukes i stor skala. Derfor kan det forventes at det meste av (digitalisert) kulturarvmateriale foreløpig beskrives ved hjelp av symboler og tekstlige uttrykk.

På lik linje med behovet for standardformater for digitalisert kulturarvmateriale, er det også viktig å tilby standardmetoder for metadata-annotasjon, tilgjengeliggjøring av innhold og unik identifikasjon av og i digitale kulturarvsamlinger. Bare gjennom en slik gjennomtenkt prosess kan man sikre seg at kulturarven kan formidles til mange.

3.1 Metadata

Metadata er definert som informasjon om informasjon og foreligger ofte som strukturert informasjon som fasiliterer arkivering, kategorisering, gjenfinning og gjenbruk. Det finnes hovedsaklig fire viktige former for metadata:

- **beskrivende metadata:** disse tillater identifikasjon og gjenfinning av et objekt, f.eks. på disk. I praksis kan dette være et filnavn på disk eller en URI på internett.

- **strukturell (sosial) metadata:** beskriver objektene og deres karakteristikk og sammenhenger / relasjoner.
- **teknisk metadata:** gir informasjon om tilvirking, format osv.
- **administrativ metadata:** er rettet mot forvaltning og lagring av objekter. Det kan også inneholde juridisk metadata som rettigheter, eierskap osv.

Det finnes forskjellige modeller for metadataforvaltning. Det er vanlig at metadata lagres uavhengig av objektene de beskriver, og vedlikeholder bare en referanse til originalen eller den digitale master-filen. For å sikre at metadataene følger med et digitalt objekt i alle tilfeller (også ved distribusjon og kopiering) er det i enkelte tilfeller ønskelig å lagre metadata i selve filen. For eksempel er det slik at EXIF data som beskriver metadata om tilstanden til kameraet når et bilde ble tatt ofte lagres i JPG filen som bildet består av. Nyere og mer avanserte teknologier er METS og MPEG21-DIDL.

3.2 Tematisk indeksering

Ofte legger man til kategorier, temaer, stikkord osv. når et (digitalt) objekt skal beskrives. Mange institu-

sjoner har klare tanker om dette og har ofte også laget skjemaer eller autoritetslister med temaer og deres koplinger. Gjenbruk er selvsagt ønskelig, og det finnes de facto standarder på en del områder. Viktige momenter man burde ta med når man velger tematisk indeksering inkluderer:

- hvis man skal gjenbruke: har rammeverket for tematisk indeksering rettigheter knyttet til seg?
- Hvem forvalter rammeverket?
- Blir det vedlikeholdt? Kan man selv gjennomføre eller foreslå endringer?
- Er det mulig å samkjøre rammeverket og dets innhold med andre kilder?
- Hva er målgruppen til rammeverket?
- Er det flerspråklig?
- Hvilket domene blir behandlet?

Det er for tiden stor interesse for gjenbruk av slike rammeverk, og et av de viktigste momentene i så måte er muligheten for å slå sammen terminologi og uttrykk. Et teknisk rammeverk basert på semantisk web teknologi som kan være interessant for utveksling av slik informasjon er SKOS rammeverket til W3C (Simple Knowledge Organisation System).

3.3 Identifikasjon

På to nivåer er det viktig at man bruker metoder og teknikker som muliggjør unik identifikasjon. På det første nivået er det viktig at alle «digitale masters» bruker en unik identifikator som filnavn. Hvis de blir publisert på (intra)nett trenger de også en URI.

Det samme gjelder også for metadata om objektet. Ofte beskrives kulturarv gjennom å beskrive relasjonen til andre (virtuelle) objekter. Disse må i slike tilfeller

være identifiserbare gjennom en indikator og ikke en tekststreng. Et velkjent problem er at personer (artister, forfattere, rettighetshavere) blir beskrevet med fornavn og etternavn i en vanlig tekststreng.

En slik tilnærming tillater ikke at man på en unik måte kan identifisere hvilken person det dreier seg om (det kan tross alt være flere som heter Ola Nordmann). En slik bruk av metadata tillater heller ikke at man sier noe om (digitale representasjoner av) personer, og dermed heller ikke at man kan søke og gjenfinne personer som et unikt objekt. Mange av dagens løsninger på web tillater ikke en slik unik identifikasjon og gjenfinning av opphavspersoner, organisasjoner, steder osv. som unike objekter med egne metadata. De fleste kulturarv-portaler er veldig sentrert rundt kulturarv-objektene og har bare unike identifikatorer til disse.

3.4 Anvendelse

I kravet om «lokalisierbarhet» ble det blant annet foreslått krav om bruk av Dublin Core, XML og URI. Et minimumskrav for «beskrivelse» av kulturarv burde være bruken av XML/Dublin Core for å beskrive samlinger, objekter, personer osv. Viktigste i dette kravet er det å identifisere standarder og teknologier som kan heve kvaliteten til Dublin Core/XML. For at metadata skal være fleksible nok til å kunne bli gjenbrukt er det viktig at selve informasjonselementene er brukt på riktig måte, at all viktig informasjon er laget og at det minimalt kan svares på spørsmålene: *hvem, hva, hvor, når og hvordan*. De fleste institusjoner har sannsynligvis mer informasjon enn dette, og derfor er dette minimumskrav som skal oppfylles for å kunne garantere samhandling og interoperabilitet mellom samlinger.

DEN foreslår en «mapping» fra disse krav mot Dublin Core som ser slik ut:

hvem	dc.creator , dc.contributor , dc.publisher
hva	dc.title , dc.description , dc.subject
hvor	dc.coverage
når	dc.date
hvordan	dc.type , dc.format

Som minimumskrav for identifikasjon ble URIen foreslått (HTTP + varig identifikator). Den foreslås publisert som dc.identifiser.

Som nevnt tidligere er helhetstenkningen viktig her, og selv om kompleksiteten i de forskjellige samlingene kan være høyere enn det minimumskravene tilsier, er det viktig at alle samlinger er på minst dette minimumsnivået. På denne måten blir det tilrettelagt for en felles formidling av norsk digital kulturarv på nasjonalt nivå.

3.5 Semantisk Web teknologi

Vanligvis er det mennesker som lager og publiserer informasjon om objekter og deres relasjoner. Mengden med tilgjengelig informasjon øker betraktelig for hvert år som går, og dermed blir det også vanskeligere og vanskeligere å gjøre et slikt arbeid manuelt.

Internettet tillater at man publiserer data og at man bidrar med informasjon og kunnskap gjennom deltagelse (blog, chat, innlegg). Denne informasjonen er søkbar gjennom datamaskiner, men informasjonen er vanligvis ikke eller dårlig tilgjengelig for automatisk prosessering. En slik prosessering kan bidra til bedre forståelse av store datamengder og kan hjelpe

til med å finne mer relevant informasjon om et objekt enn man ellers kunne ha fått til. Også formidlingsfunksjoner får et løft når mer og bedre informasjon blir tilgjengelig for applikasjoner og dermed brukere.

Semantisk web er fortsatt under utvikling, men ved å holde seg til de foreslåtte minimumskravene og lage og publisere metadata for alle digitale objekter, legger man et viktig fundament for fremtidens semantiske web.

4. KRAV OM DIGITAL VARIGHET

Til tross for omfattende investeringer finnes det ingen vannrette løsninger for langtidslagring og varig tilgang til informasjon. Å garantere varig tilgang til digital informasjon er derfor en av de største teknologiske utfordringene på dette feltet. Det er viktig at man bruker tid og krefter på digital varighet, slik at nyere og bedre løsninger blir utviklet raskt. Dette er også viktig for å kunne oppnå fremtidig avkastning på digitale investeringer som blir gjort i dag. Det ligger i kortene at digitale masters kommer til å ta rollen som digitale erstatninger til skjorbare originaler.

Beslutninger som tas nå har implikasjoner inn i fremtiden. Institusjoner som tok i bruk håndscannere på nittitallet for å scanne bilder med lav oppløsning ser nå at det kreves mer og bedre kvalitet. Den gangen var ikke det mulig, og beslutningene tatt da var fullstendig realistiske og reflekterte tidens kunnskaps- og teknologiske nivå. Valg of digitale formater er like viktig, men man vet aldri hva som blir utviklet i fremtiden. En kommersiell film på DVD med sikkerhetsalgoritmer er mest sannsynlig ulesbar om 20 år (selv om selve platen fortsatt er lesbar, er ikke formatet det lenger). Derfor er det viktig å lagre filer på formater

som er så «åpent» dokumentert som mulig.

Strategi

Det anbefales å lage og publisere en strategi for digital varighet i det norske samfunnet. En slik «varighetsstrategi» burde for hver institusjon som skal digitalisere minst inneholde fire elementer:

- **Dokumentasjon:** om strategien og planene. Hva skal gjøres og hva skal resultatene av dette være? Er strategien forankret i organisasjonen og hvis ja, på hvilken måte?
- **Organisasjon:** fordel ansvar og roller. Det er god praksis å sørge for at minst én person fra ledelsen er direkte involvert i hvert digitale prosjekt. Dette hever både troverdigheten til prosjektet og motivasjonen. Samtidig får organisasjonsledelsen direkte eierskap og resultatene blir mer synlig for alle som er involvert.
- **Budsjett:** for å kunne garantere digital varighet, må det også avsettes et (varig?) budsjett, mao. Må det ikke kuttes underveis når et budsjett er god-tatt. Slike endringer betyr ofte at man strammer

inn på hardware, software eller timebruk, noe som kan bety at man må gjøre hele jobben om igjen.

- **Tekniske løsninger:** en god teknisk løsning må på plass for å kunne garantere digital varighet. Det burde være opp til institusjonene å definere hvilke deler av løsningen som skal driftes selv og hvilke som skal legges andre steder. Det er også ofte et spørsmål om man skal ha et komplett digitalt arkiv, eller om man henter data fra andre kilder ved behov. Semantisk teknologi gjør det enklere å slå sammen flere baser til en virtuell base, og dermed blir det mer interessant for hver partner å konsentrere seg på «det man er best på». Annen informasjon kan i så fall hentes inn fra andre partnere som har forskjellig spesialkompetanse. På denne måten blir både vedlikehold og synkronisering av veldig spesialisert kunnskap som skal brukes flere steder mye enklere.

Planlegging

Det anbefales at institusjoner lager enkle og tydelige planer for digitaliseringen av sine samlinger og gjør eksplisitte valg som dokumenterer følgende:

- **hva som skal bevares?** Det viser seg at det finnes en del samlinger som faktisk ikke krever fullstendig digitalisering. I den nederlandske undersøkelsen «de Digitale Feiten» ble det f.eks. rapportert at 32% av samlingene ikke trenger eller krever å bli digitalisert.
- **hvilke formater som kan brukes/ikke brukes.** Definer kriteriene som er viktige og bestem hvilke formater som kan brukes eller ikke skal brukes.
- **Hvordan langtidsoppbevaring og tilgjengelighet er ivarettatt**
- **kunnskap om eksterne krav.** Alle kulturarv-

institusjoner må sette seg inn i kravene som er viktig for jobben som skal gjøres. Dette kan være nasjonale tjenester og retningslinjer som regulerer langtidsoppbevaring, varighet og tilgjengelighet.

Gjennomføring

Det foreslås at det kreves at alle institusjoner som skal digitalisere kulturarv må ha en *kontaktperson* som kan formidle erfaringer og kunnskap til andre institusjoner, og at prosjektet skal forholde seg til en *referansemodell* (f.eks. OAIS).

Evaluering av kvalitet

Det foreslås at det kreves at alle institusjoner som lager egne løsninger evaluerer resultatene av prosjektet mot et kvalitetsanalyse verktøy som TRAC (Trustworthy Repositories Audit & Certification).

5. KRAV OM PRESENTASJON

Å formidle informasjon og kunnskap gjennom websider eller informasjonstjenester er både en kunst og en kreativ prosess. Det er vanskelig å komme med klare retningslinjer som er generelt gyldige, siden det finnes så mange variabler i denne prosessen.

Noen minimumskrav foreslås det riktignok å definere, spesielt det om et klart skille mellom innhold og presentasjon. Det er fort gjort at websider blir laget spesifikt for et bestemt formål og kanskje et bestemt objekt. Men for å kunne tilby informasjon gjennom flere kanaler (tenk universell utforming), er det viktig at innholdet er tilgjengelig uten at man har tilrettelagt det for en bestemt type bruk.

Den 1. januar 2009 trådte diskriminerings- og tilgjengelighetsloven i Norge i kraft. Loven gir et rettslig vern mot diskriminering på grunn av nedsatt funksjonsevne, og dette gjelder på alle samfunnsområder. All ny informasjons- og kommunikasjonsteknologi (IKT) rettet mot allmennheten skal fra 1. januar 2011 være universelt utformet. For eksisterende IKT gjelder kravet for første versjon etter denne fristen, senest fra 1. januar 2012¹¹.

For utforming av websider foreslår Direktoratet for forvaltning og IKT i denne forbindelse å bruke retningslinjer for websider som er utformet av W3C i «Web Content Accessibility Guidelines (WCAG)»¹².

Her i landet er det organisasjonen Standard Norge som jobber med standardisering. En god oversikt over standardiseringsrammeverk for universell utforming nasjonalt og internasjonalt finnes på standard.no sine hjemmesider:

<http://www.standard.no/no/fagomrader/Universell--utforming/Fagartikler/Standarder-og-universell-utforming-IKT-og-hjelpemidler/>

11 sitat Senter for IKT i utdanningen. <http://www.uninettabc.no>

12 <http://www.w3.org/TR/WCAG20/>

6. KONKLUSJON

Visjonen om langtidslagring og online tilgjengelighet av digital kulturarv og vitenskaplig informasjon ble nedfelt av EU kommisjonen i dokumentet *i2010 - A European Information Society for growth and employment* allerede i 2005. Det har blitt satt av mer en NOK 3.2 milliard (400 M€) i perioden 2007-2013 til prosjekter innenfor digitalisering og forbedring/tilretteleggelse av teknologien for det.

Det er mange fordeler ved å ha en felles og samkjørt satsing for digitalisering av kulturarv og vitenskaplig kunnskap. Ikke bare er det viktig å tilrettelegge for enklere kunnskapsoppbygging i befolkningen slikt at vi får økt konkurransedyktighet og stimulerer kunnskapsintensiv produksjon og tjenester. Det er gjennom en bedre forståelse av vår historie, kultur, identitet og sammenheng av den med den globale utviklingen de siste århundre at vi forhåpentligvis klarer å kreere et samfunn hvor folk har kulturell og historisk forståelse av seg selv og hverandre.

Gjennom å tilpasse ABM sektoren til den nye hverdagen i denne digitale verden blir det enklere for individer å få tak i kunnskap og informasjon, samt at man kan enklere bidra med sitt kunnskap om digital kulturarv. Det har oppstått en helt ny kultur hvor mennesker gir fra seg mye informasjon og erfaring gjennom (delvis anonymiserte) portaler. Klarer man å fange opp slik erfaring og knytte dette til relevant

kildemateriale av høy kvalitet (som ABM sektor kan levere), har man skapt grunnlag for verdiskaping og øking av konkurranseevnen. Digitalisert kulturarv må derfor gjøres tilgjengelig på en slik måte at hver organisasjon kan jobbe på sine fagområder, og at kunnskap og kildemateriale som ikke tilhører dette området kan (automatisk) hentes inn fra de aktører på markedet som jobber med det. Dette innebærer sterke krav for samhandlingsformer mellom løsninger innenfor ABM sektoren.

Men hvordan kommer man ditt? Det er tydelig at en perspektiv på langtidslagring og tilgjengelighet av digital kulturarv krever standarder, åpenhet og teknikk som støtter opp. Samtidig er det viktig at man ikke «låser» seg opp i alt for spesielle løsninger. Digital kulturarv må bli gjort tilgjengelig på riktig måte, en måte som tar hensyn til at verden ikke forandres på en dag, men som samtidig forebygger at den innsatsen man gjør nå holder mål om 10-20 år også.

På internasjonal nivå er det fremmet metodikk og retningslinjer for å gi institutter nettopp den støtte og trygghet man trenger når det skal settes i gang større digitaliseringsprosjekter. Det er viktig at en slik metodikk blir laget i samarbeid mellom institutter og politikken, slikt at det bygges opp en felles forståelse, samtidig at man kan knytte økonomisk støtte til bruk av metodikken og kravene den inneholder. Ikke i det minste gir en slik fremgangsmåte offentlige aktører

en profesjonell ledetråd i forhandlinger med kommersielle aktører. Poengen er at man, gjennom å bruke metodikken, kan være sikre på at den jobben som blir utført er av en vis kvalitet og tar hensikt til noen viktige krav i fremtiden.

De viktigste krav man må bli enige om gjelder standarder og teknologien for å formidle kunnskap gjennom distribuerte systemer. En viktig konsept som må på plass er at alle digitale objekter skal ha en eller flere unike referanse i den digitale verden som også er kalt Unified Resource Identifier (URI). Hvis den er på plass, blir det mulig å uttale seg om slike objekter og legge til informasjon om den som er tilgjengelig for alle. Uttalelsen man gjør må være interpreterbar for maskiner, og derfor trenges det enighet om eller standarder for språk til kunnskapsrepresentasjon. Mye av det som trenges av infrastruktur er allerede tilgjengelig i dag i form av OSI og TCP/IP standarden som implementerer internettet. W3C har de siste 10 år kommet med flere anbefalinger «recommendations» for standarder til kunnskapsrepresentasjon. Disse standarder (bla RDF/RDFS og SPARQL) tillater at man både representerer kunnskap digitalt og distribuert, samt at man kan spørre kilder for informasjon man ønsker. En viktig utgangspunkt i standardene er at det må være mulig å slå sammen kilder og informasjon etter behov og uten tunge konverteringsproses-

ser.

For å kunne slå sammen uttalelser og kunnskap fra forskjellige kilder er det viktige at man referer til hverandres informasjon, bare på denne måte kan visjonen om «Linked Open Data (LOD)» bli virkelighet. LOD tanken har oppstått for noen år siden, og noen av de viktigste og mest brukte internett ressurser er del av den. Det nevnes bla. omfangsrike kilder som wikipedia, Flickr, Musicbrainz, men også DBLP (bibliotek), Wordnet, BBC, bio-medical sources og Geonames.

I Norge finnes det en god del portaler hvor man får tilgang til digital kulturarv og samlinger. Rapporten behandler en del av disse og viser for eksempel at det ikke finnes portalløsninger som implementerer kravet om unik identifikasjon kombinert med bruk av åpne W3C standarder. Hadde det vært gjort, kunne man for eksempel koble Digitalt Museum mot Digitalt Fortalt. Dette tillater at en besøker av Maihaugen som vet noen historier om en av de artefakter i museet kan faktisk lage en liten film/lydspor om den og publisere den gjennom Digitalt fortalt. Hvis Digitalt Museum bruker URIer for sine objekter, så kan innlegget på Digitalt Fortalt knyttes direkte til den, selv om den digitale artefakten og fortellingen ligger i to forskjellige baser. For å følge tankeeksperimentet videre: si at EUROPEANA begynner å referere til objekter i Digitalt Museum, så hadde de objektene i Digitalt For-

talt automatisk blitt gjort tilgjengelig til et mye større publikk samtidig.

På denne måten støtter man kravet om samordning, utveksling, innhenting og felles formidling av kulturarv.

For å komme ditt fra dagens løsning må portalene som publiserer digital kulturarv være konsekvent med å publisere data som objekter med en unik referanse på nett. I tillegg må disse knyttes til publiserte beskrivelser (metadata) som bruker aksepterte og felles rettnings-linjer (som f.eks W3C tilbyr). Fordelen med å gjennomføre slike endringer er at man enkelt kan hente inn informasjon fra andre steder om sine egne ressurser, at man kan dele sine datamodeller med andre (semi-automatisk) uten høge kostnader og at en hvis grad for automatisk resonering blir mulig.

Legger man til kravet om åpen tilgang, så blir det mulig for alle å innhente, analysere, slå sammen og gjenbruke informasjon i nye, felles tjenester som er til nytte for alle og som har potentiale å inkludere flere i den spennende og viktige historien som kulturarv representerer.

7. ORDFORKLARINGER

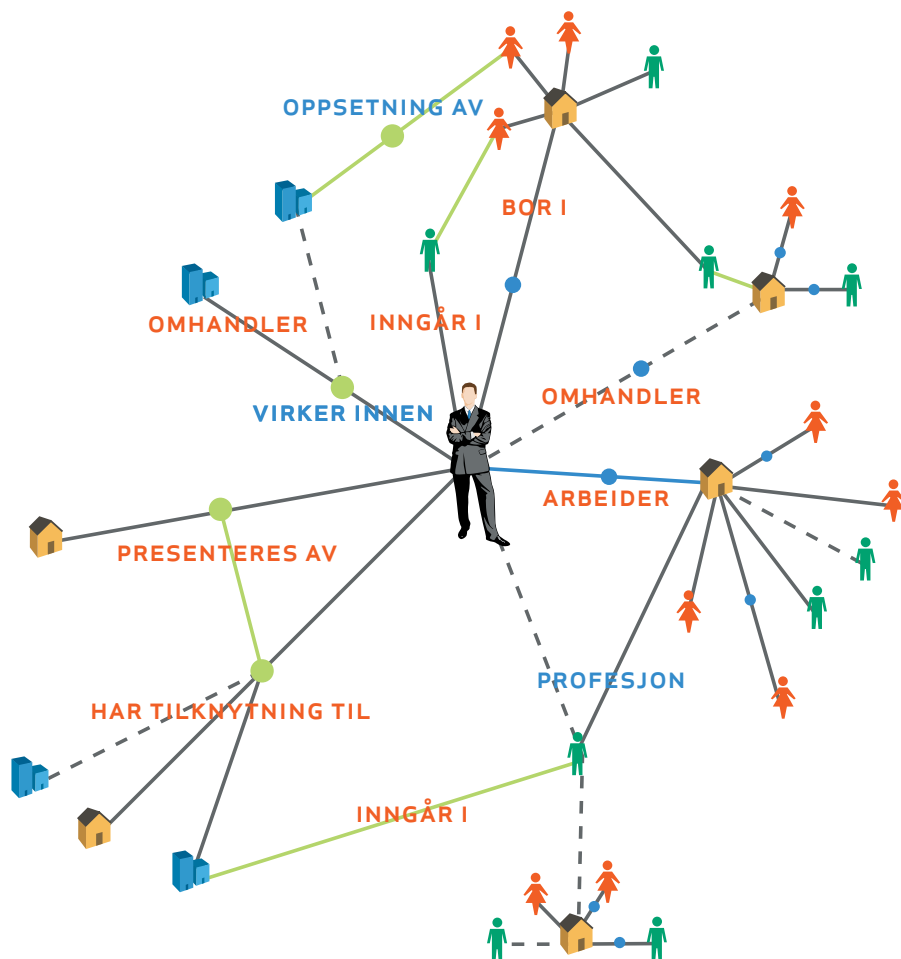
Denne listen med ordforklaringer er lagt ved for enkelhets skyld. Mange av definisjonene er tatt fra no.wikipedia.org og lignende portaler. Noen er tilpasset og forkortet.

Back-end system	Det blir ofte referert til en server som en «back-end» eller «back-office» system.
Back-office	Se back-end system.
Cloud Computing	Norsk: «nettskyen». Dette er en betegnelse for dataprosessering og datalagring på servere som står i eksterne serverparker tilknyttet internett. Oppgaver blir automatisk fordelt over mange maskiner for raskt behandling.
Front-end system	Applikasjoner og systemer som står hos brukeren og som er avhengig av en server i «back-enden» for å kunne fungere.
Ontologi	Ontologi betyr noe slikt som «slik ting faktisk er». I diskusjonen om semantisk teknologi blir det definert som “en formell spesifisering av en konseptualisering” (Gruber ‘93), og er en definisjon av et begrepssystem formulert i noen form for logikk som kan representeres i en datamaskin.
OWL	Web Ontology Language. Dette er en familie med språk for kunnskapsrepresentasjon for å forfatte ontologier. OWL er foreslått av World Wide Web konsortiet W3C.
RDBMS	En relasjonsdatabase er en database som bygger på den såkalte relasjonsmodellen. Relasjonsdatabaser består av tabeller som er forbundet med henvisninger, eller nøkler mellom seg, som utgjør relasjoner.
RDF / RDFS	Resource Description Framework. En W3C standard som er opprinnelig ment som en metadata datamodell. Den har blitt en standard metode for konseptuell beskrivelse av web ressurser, hvor det brukes forskjellige syntakser.
Semantisk Web	Semantisk web er en fremskreden utvikling av World Wide Web, hvor betydningen av informasjonen er definert. Dette gjør det enklere for datamaskiner å prosessere informasjonen.
SPARQL	SPARQL Protocol and RDF Query Language er et spørrespråk i RDF.
SQL	Structured Query Language. Standardisert spørrespråk brukt for å hente ut informasjon fra Relational Database Management Systems (RDBMS).

SRU
URI

Search/Retrieval via URL (Standards, Library of Congress)

En URI (Uniform Resource Identifier) er en kompakt streng av tegn som identifiserer en abstrakt eller fysisk ressurs. En URI kan videre klassifiseres som en lokator eller et navn, eller begge deler.



8. REFERANSELISTE

I tillegg til informasjon fra diskusjoner med fagpersoner fra hele Europa på konferanser og studieturer (se forord) er det hentet informasjon fra følgende publikasjoner:

Allemang, D., Hendler, T. *Semantic Web for the Working Ontologist: effective modeling in RDFS and OWL*. Morgan Kaufman 2007.

Berners-Lee, Tim; Fischetti, Mark. *Weaving the Web*. HarperSanFrancisco. chapter 12. ISBN 9780062515872. 1999.

Codds, E.F. *A Relational Model of Data for Large Shared Data Banks*. 1970.

Davies, J., Studer, R. and Warren, P. (eds). *Semantic Web Technologies; trends and research in ontology-based systems*. Wiley, 2007.

DEN Foundation. *Business Model Innovation – Cultural Heritage*. Ministry of Education, Culture and Science (NL). 2010.

EU commission. *i2010 - A European Information Society for growth and employment*. 2005.

i2010 Digital Libraries Initiative. *Digital Libraries: recommendations and challenges for the future*. Desember 2009.

Kersen, van. 2004. *Aaibaar Erfgoed*. <http://www.den.nl/docs/20050825135226>

Lowndes, M. *An Introduction to the Semantic Web for Museums*. At: *Museums and the Web, the int. Conference for culture and heritage online*. Albuquerque, New Mexico. 2006

NDU-Utvalget. *Nasjonalt digitalt universitetsmuseum (NDU)*. Rapport 2009

Norway. *Implementation of the commission recommendation on digitisation and online accessibility of cultural material and digital preservation*. Norges rapport til EU 2010.

Kruk, S.R., McDaniel, B. (eds). *Semantic Digital Libraries*. Springer 2009.

Stichting DEN. De BASIS. Sidene besøkt: 15.05.2010. <http://den.nl/debasis>.

Stichting DEN. *De Digitale Feiten, Onderzoek naar de omvang en kosten van gedigitaliseerd cultureel erfgoed*. Den Haag, 2009.

St.meld. nr. 15 (2007-2008). *Tingenes Tale*.

St.meld. nr. 23 (2008-2009). *Bibliotek. Kunnskapsallmenning, møtestad og kulturarena i ei digital tid*.

St.meld. nr. 24 (2008-2009). *Nasjonal strategi for digital bevaring og formidling av kulturarv*.

Warberg Løssing, A. S. *Digital Museumsformidling – i brukerperspektiv*. Kulturarvstyrelsen (DK), 2009.

Web Content Accessibility Guidelines (WCAG). <http://www.w3.org/TR/WCAG20/>

Yu, L. *Semantic Web and Semantic Web Web Services*. 2007.

