



# ÅPEN OG SAMORDNET TILGANG TIL KULTURARVEN- 2. UTGAVE

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



**ABM-UTVIKLING**

POSTBOKS 8145 DEP  
N-0033 OSLO

TELEFON: 23 11 75 00  
TELEFAKS: 23 11 75 01

POST@ABM-UTVIKLING.NO  
WWW.ABM-UTVIKLING.NO

TRYKK: 07 GRUPPEN

FORSIDEFOTO:  
ISTOCKPHOTO.COM

2. UTGAVE  
ISSN 1503-5972 (TRYKT UTG.)  
ISSN 1504-9167 (ONLINE)  
ISBN 978-82-8105-089-1

ABM-UTVIKLING 2010



# INNHold

|                              |    |
|------------------------------|----|
| MANDAT OG AVGRENSINGER ..... | 8  |
| ET PARADIGMESKIFTE.....      | 10 |

## **DEL I – FREMTIDENS KULTURARV, KULTURARVENS FREMTID**

|  |    |
|--|----|
| 1 VISJONER, TRENDER OG TEKNOLOGIUTVIKLING..  | 12 |
| 1.1 DIGITAL PRODUKSJON OG DISTRIBUSJON AV KUNNSKAP OG OPPLEVELSER.....             | 12 |
| 1.2 SOSIALE MEDIER, MOBILITET OG BRUKERSKAPT INNHold.....                          | 16 |
| 1.3 SEMANTISK WEB FRA VISJON TIL REALITET.....                                     | 19 |
| 1.4 DIGITALISERING AV KULTURARVEN I EUROPA....                                     | 28 |
| 1.5 DIGITALISERING AV KULTURARVEN I NORGE.....                                     | 33 |
| 1.6 UTFORDRINGER VED DAGENS SATSING PÅ DIGITALISERING AV KULTURARV .....           | 39 |
| 2 IKT-STRATEGI FOR FREMTIDSRETTET DIGITALISERING OG FORMIDLING AV KULTURARVEN..... | 41 |
| 3 VEIEN VIDERE: HVORDAN KAN VI LØFTE ABM-SEKTOREN?.....                            | 48 |
| 3.1 DEFINERING AV KRAV TIL DIGITALISERINGSPROSJEKTER.....                          | 49 |
| 3.2 VEILEDNING OG KURS.....  | 49 |
| 3.3 FINANSIERING OG INSENTIVER.....  | 50 |
| 3.4 DELTAGELSE OG DISKUSJONSFORA.....  | 50 |

## **DEL II – MINIMUMSKRAV TIL DIGITALISERING AV KULTURARV**

|   |    |
|---|----|
| 1 KRAV OM LOKALISERBARHET .....                 | 52 |
| 2 KRAV TIL PRODUKSJON AV DIGITALE OBJEKTER .... | 54 |
| 3 KRAV TIL BESKRIVELSE.....                     | 56 |
| 4 KRAV TIL VARIGHET .....                       | 59 |
| 5 KRAV TIL 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 magasinane 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](http://www.digitaltfortalt.no)) vart opna i 2009 og inneheld digitale forteljingar skapt av institusjonar og privatpersonar. Arkivportalen ([www.arkivportalen.no](http://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](http://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/](http://www.unimus.no/foto/)). Deler av dei arkeologiske samlingane er tilgjengeleg på [www.unimus.no/arkeologi/](http://www.unimus.no/arkeologi/).

Nasjonalbiblioteket har opna [bokhylla.no](http://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](http://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](http://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](http://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 fagfeltet «cultural heritage» i Norge og resten av Europa. Innledende undersøkelser av eksisterende initiativer viste at Haag i Nederland er et knutepunkt for viktige internasjonale trender. Her er mange gode miljøer og spennende initiativer samlet. På aksene Haag-Amsterdam-Luxembourg finner man mye av «state of the art» innenfor abm-sektoren i dag. I tillegg til internasjonale miljøer har vi også brukt tid på å få oversikt over de mange spennende prosjekter og initiativer i Norge, og sammenlignet dem med det som etter alt å dømme kommer til å bli «beste praksis» for infrastruktur og informasjonsarkitektur for formidling av digital kulturarv.

Jeg vil gjerne takke Gunnar Urtegaard, Lars Wenaas (ABM-utvikling), Yvo Volman (EU-kommisjonens generaldirektorat for informasjonssamfunn og medier, 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, Nederland), Paul Doorenbos, Huibert Crijns (Het Geheugen van Nederland og Koninklijke Bibliot-

heek) 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

# MANDAT OG AVGRENSINGER



Hvilke 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

Denne 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 aktører i nasjonale og internasjonale institusjoner 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 digitale kulturarven stort sett er ukjent, og derfor skal rapporten ikke komme med konkrete anbefalinger for hvordan kulturarven skal formidles til fremtidens publikum. Dette skal være så «åpent» som mulig. Derimot skal det utredes hvordan abm-sektoren best kan ruste seg for å møte fremtidige krav. Målet må være at valgene som tas i dag tillater og åpner opp for et størst mulig mangfold av anvendelser i fremtiden.

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

Hvilken retning tar internasjonal forskning på digital informasjonsforvaltning for tiden, spesielt innen digital representasjon av kulturarv?

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 sikres tilgang til nødvendig kunnskap og kapital til å «være med», 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

# 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 1990-tallet 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 form for «sikkerhetskopiering». Mens 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 skaper 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 enkle å kopiere, dele og distribuere. Dermed blir det lettere å knytte sammen informasjon fra forskjellige kilder. For kulturarven gjelder at summen av slik sammenkoplet informasjon er vesentlig større enn delene. Relevansen av integrert informasjon er høy, både for eksperter og samfunnet generelt.

En betydelig innsats gjøres for å digitalisere det meste og beste av verdens kulturarv. I tillegg produseres mange nye kulturelle uttrykk digitalt i utgangspunktet («born-digital»). Økt tilgang til datamaskiner, digital fotografi mm., og økt kompetanse i bruk av maskin og programvare har ført til at mye mer materiale enn før blir oppbevart. Der hvor man tidligere måtte ta hensyn til kostnadene for å dokumentere en hendelse (på 1980-tallet kostet et polaroidbilde f.eks. ca. 3 kroner og tre minutter Super 8-film kostet gjer-

ne 80 kroner), er prisen i dag praktisk talt den samme om man tar ett bilde som hvis man tar tusenvis. Siden maskinvare har blitt så mye billigere i anskaffelse og bruk, samt at tilgjengeligheten av lagringsplass er nesten ubegrenset, øker den tilgjengelige informasjonsmengden eksponensielt. Sammenlignet med den analoge verden er kostnadene knyttet til digital publisering og distribusjon gjennom et datanettverk meget lave.

Det er klart at dette fører til et paradigmeskifte. Publikum er blitt vant til mobilitet og deltagelse. Nye former for informasjonsspredning vokser frem. Informasjon (og kunnskap?) blir gjort tilgjengelig gjennom portaler som lever av dugnadsånden (Wikipedia, YouTube, Flickr). Kvaliteten på slik informasjon kan i mange tilfeller være høy, men er det ikke alltid. Det er blitt vanskeligere å skille kvalifisert informasjon fra ukvalifisert informasjon. Bærbare musikk- og videospillere, smarttelefoner 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 har både mengden av og tilgangen til informasjon i samfunnet 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 abm-sektoren egentlig tenkt å imøtekomme disse forventningene?

<sup>1</sup> Rapporten går ikke 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.

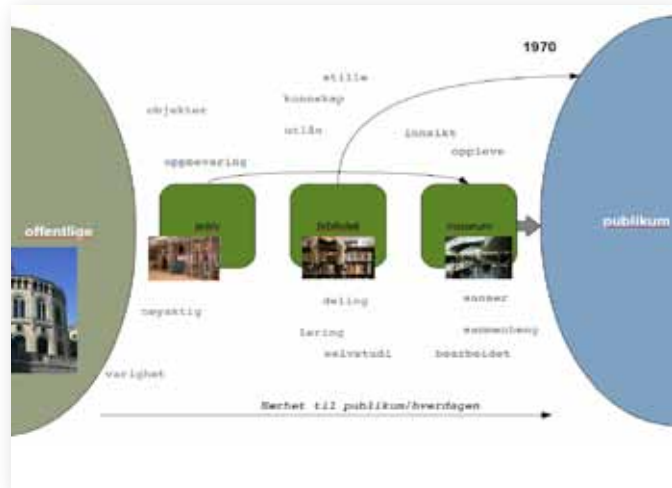


DELL 1

# 1. VISJONER, TRENDER OG TEKNOLOGI-UTVIKLING

## 1.1 Digital produksjon og distribusjon av kunnskap og opplevelser

I økende grad forventer folk døgnåpen mulighet for kommunikasjon med andre, døgnåpen tilgang til all slags tjenester og at innhold og data skapt i alle verdens tidssoner er umiddelbart tilgjengelig. Denne globale trenden kan oppfattes som en direkte følge av at en stadig større andel av befolkningen har nett-tilgang, tar i bruk lett tilgjengelige verktøy for publisering og deler informasjon og innhold med hverandre.



Illustrasjon 2: Enkelt, tradisjonelt perspektiv på abm-sektoren

Arkiver er i mye større grad enn før 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. Mens man tidligere oppsøkte et offentlig arkiv for å lete etter det man var ute etter i det fysiske arkivet, 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 brukervennligheten og presentasjonen av innholdet. Dette er ikke noen naturlig rolle for et arkiv, som historisk sett har hatt andre institusjoner som «front-end» mot publikum.

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 et vidt spekter 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

Endrede forventninger til Arkiv, Bibliotek og Museum

og kunnskap, men brukerne må selv gjøre en aktiv innsats for å erverve seg 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 objektene, men får derimot oppleve en historie som er av høy kvalitet og laget av eksperter som videreformidler et bestemt syn på objektene. I den forstand tilbyr et museum «beriket» informasjon som allerede er redigert og tolket for brukeren, i motsetning til et arkiv som ikke gjør noen redigeringsjobb og et bibliotek som heller ikke tolker og redigerer, men som organiserer objektene etter tema/språk/land osv.

Demarginalisering av mindre samlinger

Digitaliseringsbølgen har ført til at en del veldig spesialiserte samlinger har blitt tilgjengelig for et tynt spredd publikum. Teknologiutviklingen har gjort oppbyggingen av slike samlinger atskillig enklere. Prosjekter som Norges Pattedyratlas bruker internett aktivt for å hente inn metadata om forekomster, arter osv. på en direkte, døgnåpen og desentralisert måte. Det forvaltes mye interessant materiale rundt omkring i landet som kanskje er blitt betraktet som «lokalkunnskap» eller av kun lokal interesse, men som kan bli populært og veldig mye brukt når det formidles digitalt. 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.

For å kunne formidle kunnskap og opplevelser til brukerne er abm-sektoren avhengig av å kjenne konteksten en bruker befinner seg i. Konteksten blir ofte definert gjennom en lokal setting (utstillingslokaler, bibliotekrom, temaer) og brukeren må som regel

være fysisk til stede.

Lokal tilstedeværelse vs opplevelser i det offentlige rommet

Yngre generasjoner (og i økende grad også eldre) forventer derimot gjerne mobil og stedsuavhengig tilgang til informasjon. En slik tilgang er ofte basert på bruk av bærbare og håndholdte data-duppeditter - som avanserte mobiltelefoner, GPS, sub-notebooks osv.) - som har 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), tar IKT-bransjen i bruk flere og flere sensorer for å unngå begrensningene ved små tastaturer. De nyeste mobile produktene bruker GPS-systemer, bevegelsessensorer, akselerasjonsmålere, trykkskjermer, kameraer og mikrofoner for å kommunisere med sine brukere. Det ligger et enormt potensial for abm-sektoren i å utnytte slik informasjon om bruker-kontekst til å tilby skreddersydde og overraskende opplevelser.

Dermed er det satt i gang en trend hvor utlån av fysiske gjenstander blir mindre viktig enn før. Med fremveksten av 3D-teknologi er det ikke usannsynlig at også skulpturer og andre tredimensjonale objekter kan oppleves digitalt i fremtiden. Dette vil ikke fjerne behovet for å stille ut originaler, men det skaper en trend hvor «utlånsinstitusjoner» vil bli mindre viktige og steds- og tidsuavhengige opplevelser vil stå mer sentralt.

Deltagelse i stedenfor mottagelse

Økt deltagelse av publikum i ulike brukerstyrte nett-tjenester (web 2.0) skaper også forventninger til abm-institusjoner. Man ønsker rett og slett å kunne dele sine erfaringer med og om organisasjoner på en helt annen måte enn før. Enten det gjelder opplevelser som museums- og biblioteksbesøk eller utstillinger, så ønsker man å dele erfaringene med

venner via Facebook, LinkedIn, Twitter og mange andre nett-tjenester. Dette må abm-sektoren legge til rette for.

*Gode prosesser, sikker oppbevaring, garantert lesbarhet*

Digitalisering av objekter er dyrt. Institusjonene må sette av ressurser til å utføre arbeidet digitaliseringen innebærer, og feil i prosessen kan føre til betydelige økonomiske tap. Det er ikke gjort noen undersøkelser vi er kjent med som beregner nøyaktig hvor mye av digitaliseringen som ble gjort på 1990-tallet som må gjøres om igjen i dag på grunn av 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 om objektene. 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 spesifikasjon som garanterer teknisk lesbarhet. Mye materiale fra 1980-tallet er vanskelig å få teknisk tilgang til i dag, enten fordi mediet som er brukt ikke kan avleses lenger, eller fordi det er brukt spesielle og sjeldne formater. Her finnes det mye erfaring å høste fra internasjonale satsinger. 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.

Mange institusjoner har blitt mer kritiske i vurderingen av hvilket materiale som skal digitaliseres. 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, som står sentralt i denne rapporten, bidrar til at slik duplisering kan unngås gjennom en bedre måte å publisere digitale objekter på. Slik kan alle interessenter enkelt sammenligne eksisterende ressurser med sine egne oversikter. I Nederland ble det rapportert at 32 prosent av materialet eid av abm-institusjoner ikke trenger å digitaliseres (Stichting DEN, 2009).

*(A)BM blir opplevelsessenter?*

Det er en interessant observasjon at flere og flere abm-institusjoner tenker opplevelser. Det er ikke lenger nok å være museum, bibliotek eller arkiv (se illustrasjon 3). Publikum er blitt mer involvert.

Institusjonene vektlegger sterkere at folk vil «gjøre» ting, være aktive. Spillelementet har kommet for alvor. Mange museer ser også behov for å berike opplevelsene de tilbyr (man kan vel neppe snakke om utstillinger mer?) med informasjon de ikke har selv, for eksempel gjennom internett. Det er også en trend å være til stede 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 og samlinger med flere aktører, og at man kan ta imot informasjon (automatisert) fra andre. Siden det er sannsynlig at bransjeinterne abm-aktører i fremtiden blir nødt til å samarbeide med aktører fra andre bransjer (som turisme, forlag), er det viktig at det brukes lagringsformater og publiseringsmetoder som holder mål på internasjonalt nivå.

Nettopp fordi så mye informasjon (som aviser, leksika) og tjenester og produkter (programvare,

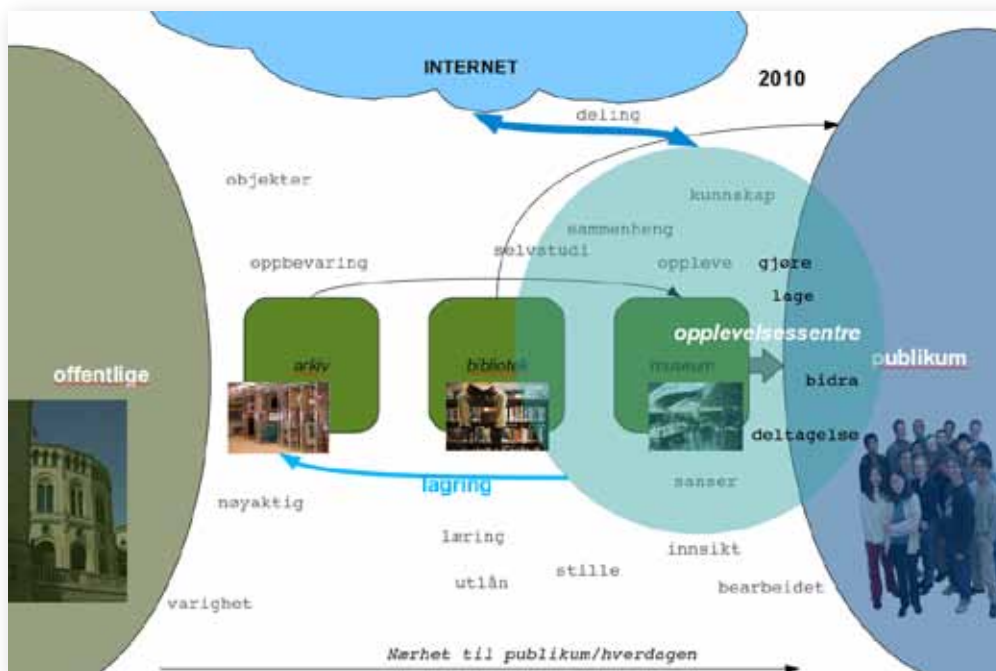


lagringsplass, prosessortid) som det tidligere måtte betales for er blitt relativt billig og enkelt tilgjengelig, har det oppstått en trend i retning av at alt innhold og alle data skal være åpne (uten passord) og fritt tilgjengelig (uten kostnader). I rapporten «Digitale Feiten» går det frem at 68 prosent av all digital kulturarv i Nederland er gratis tilgjengelig uten registrering, og ytterligere 7 prosent er gratis tilgjengelig etter at man har registrert seg.

Diskusjonen om tilgjengelighet og eventuelle kostnader krever en dyp og grundig samfunnsdebatt om informasjons- og kunnskapskilder som er betalt med offentlige midler. Internasjonalt er det en klar trend å åpne opp offentlig sektors datakilder for viderebruk. Eksempler er statistiske data, demografiske data, geodata. Også Norge har sluttet seg til utviklingen<sup>2</sup>. Når

<sup>2</sup> <http://data.norge.no> (Norge), <http://data.gov.uk> (UK) og <http://www.whitehouse.gov/open/> (US).

man skal åpne opp datakilder, skal man egentlig gjøre det ut fra tanken om at summen ofte er større enn delene. En slik tanke innebærer at man ikke skal være redd for at informasjon blir koplet sammen med andre kilder. Et kjent eksempel er 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 legge så få begrensninger som mulig på den videre bruken av informasjonen (dette bør klargjøres i lisensbetingelser, der et alternativ er å ta i bruk en form for Creative Commons-lisens som garanti for at åpne data forblir åpne).



Illustrasjon 3: Opplvessentere tar museenes plass?

Hvis man trekker den linjen som man nå ser i flere land videre, så er det sannsynlig at digitaliseringsbølgen kommer til å fortsette. Konsekvensene for biblioteker og bokhandlere kan komme til å ligne mye på 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 utvidet med sofahjørner og kaffebarer, slik at de kan by på noe mer enn bare salg av bøker. Dette har også enkelte biblioteker forstått. Biblioteker kommer i fremtiden sikkert til å oppfylle en annen rolle i samfunnet, der kombinasjon av kunnskap og opplevelser står sentralt.

Arkiver kommer også til å få flere roller. 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 objekter som er «born-digital» (f.eks. musikk, film, tekst, epost). Det er foreløpig ikke tydelig hva dette virkelig innebærer og hvordan man skal samle inn den delen av kulturarven som er «born-digital». (Har man for eksempel 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 digitalisert kulturarv. Disse objektene kan reproduseres mot en kostnad, noe man nettopp ikke kan gjøre med born-digital-objekter.

## 1.2 Sosiale medier, mobilitet og brukerskapt innhold

Ad-hoc informasjon og troverdighet  
Nettbrukere produserer og deler informasjon og kunnskap. Dette har nådd et slikt omfang at subjektive innlegg skrevet av amatører til tider blir tatt mer seriøst enn informasjon produsert av fagfolk og -institusjoner. Sammenkobling av lokal og offisiell informasjon («når var siste besøk av mattilsynet til denne restauranten?») blir gjerne kombinert med informasjon produsert av brukere («hva synes de om restauranten og menyen?»). Abm-sektoren kan bidra med mye interessant stedsbasert materiale. Byarkivet eller Riksarkivet kan ha bilder av stedet anno 1895, Popsenteret eller Rockheim kan ha informasjon om artister som har tilknytning til dette stedet (og eventuelt hva slags tilknytning), Nasjonalbiblioteket har kanskje en videosnutt med et intervju om stedet - og slik kan det fortsette.

Dette er ikke fremtidsmusikk. Det finnes allerede i dag en del applikasjoner som bruker augmented reality-teknologi for å kople stedsbasert informasjon sammen i en live-videostream via kamera og skjerm på mobiltelefonen (se illustrasjon 5).

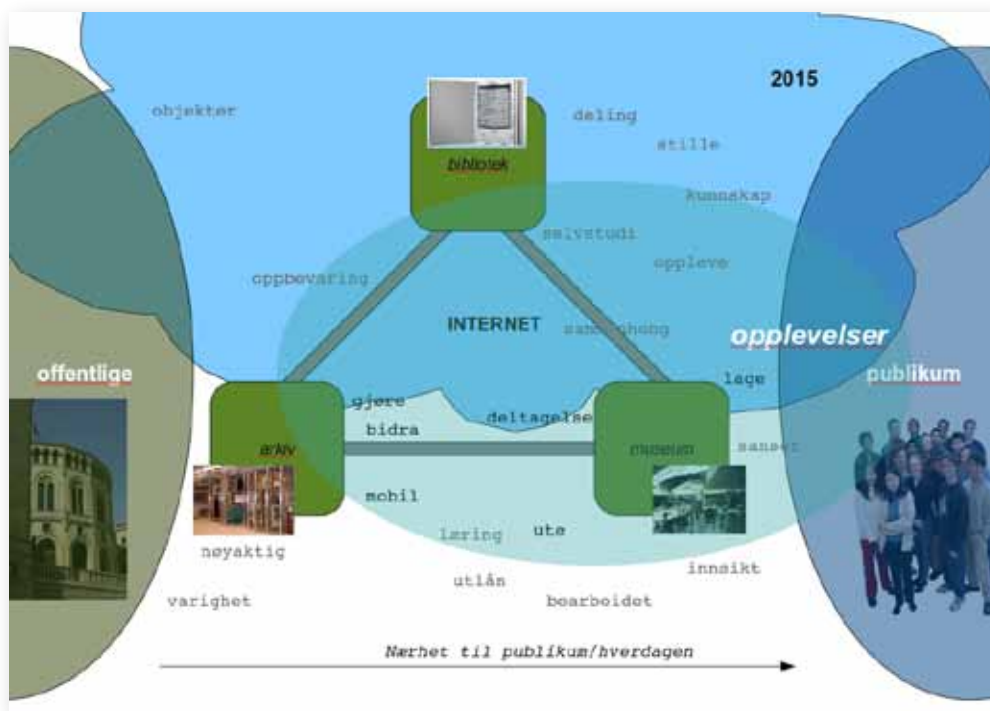
Slike scenarier og verktøy burde også abm-sektoren vurdere som reelle alternativer. En viktig teknisk forutsetning er at informasjon blir knyttet til de riktige objektene gjennom et globalt identifikasjonssystem som kopler metadata til objektene (nøkkel-system). Dagens internett byr på denne muligheten gjennom standarden Unified Resource Identifier (URI).

Tidligere brukte nettbrukerne mye tid på 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 kilder (booking av hotell, minSide, Gotogate.no, metacrawler.com, Google News, etc). Man ønsker ikke bare å få tak i informasjonen på en rask og enkel måte, 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 nettapplikasjoner blir ofte referert til som mash-up.

Web 2.0-løsninger har imidlertid en del utfordrin-

ger. Typiske web 2.0-applikasjoner er for eksempel dårlig tilgjengelig for søkemotorer og har gjerne API (programmeringsgrensesnitt) som ikke er generiske. Det skaper problemer for dem som lager applikasjoner, fordi de må tilrettelegge applikasjonene for hver ny kilde de 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 for brukerne å legge til eller fjerne kilder. Web 2.0-applikasjoner kople sammen informasjon og data og presenterer det for brukerne, men tilbyr sjelden automatisert tilgang til dataene som blir brukt i applikasjonen.

Det arbeides aktivt for å gjøre noe med disse be-



Illustrasjon 4: Fremtiden? Arkivet lagrer og leverer tjenester til (internett-basert?) bibliotek, opplevelsessenter og museum

grensningene og ta neste steg. Resultatene betegnes gjerne som web 3.0.

**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 innhold. Web 2.0 legger til rette for utvikling av interaktive web applikasjoner som tillater aktiv deltagelse av brukere i produksjonsprosessen, samt grunnleggende muligheter for å gjenbruke og slå sammen eksterne informasjonskilder.



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

Tagging av omverdenen  
Brukere er opptatt av å kunne dele informasjon ad hoc og til alle døgnets tider. Mens man reiser, lager innlegg til en blogg, tar bilder eller bare har sensorer på mobilen, legger man igjen et digitalt spor. Hver gang mennesker lager innhold som kan knyttes til et geografisk punkt (eller bare passerer forbi), blir dette punktet markert («tagget») med informasjon som lagres i den virtuelle verden. Dette kan være sensorinformasjon fra bilen (trykk, temperatur, fuktighet), men også bilder (Eiffeltårnet, Sinnataggen) eller innlegg («maten her er dårlig»), gjerne i sammenheng med geo-lokasjonsdata (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 mangt, er det godt mulig å fange opp en del ytringer og informasjon om objekter som tilhører en samling. Dette kan gjøres lokalt (i et museum) gjennom for eksempel feeds fra en tjeneste som Twitter. Men institusjonene kan også samle slike ytringer fra folk som er på reise (bilder, twittermeldinger, blogginnlegg om f.eks. Sinnataggen i Frognerparken).

Ikke kostnøe – data lagring ubegrenset tilgjengelig?  
Mens man tidligere valgte ut materiale til arkivering ut fra fysisk lagringskapasitet, 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 over 1 terabyte 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 for samfunnsvitenskapelig

forskning om noen år, men innholdet er kanskje ikke alltid umiddelbart interessant. Hvordan kan abm-sektoren finne og bruke de virkelige gullkornene, og kanskje bidra til at de lagres og gjenfinnes? Samspillet mellom abm-innhold og brukerskapt, nettbasert innhold vil bli stadig viktigere i årene som kommer.

### 1.3 Semantisk web fra visjon til realitet

Tim Berners-Lee, oppfinneren av world wide web, formulerte visjonen om en semantisk web i 1999. Grunttanken er at datamaskiner skal kunne analysere og tolke dataene som finnes på allverdens web-sider. Målet er å få til en forvaltning av selve informasjonselementet på det laveste nivået og gjøre det maskinlesbart. Disse ideene har inspirert til videre arbeid fra forskere og utviklere i Europa, USA og Asia med å definere og utvikle et felles rammeverk under paraplyorganisasjonen World Wide Web Consortium (W3C). Blant resultatene er spesifikasjonen av Resource Description Framework (RDF), RDF Schema og Web Ontology Language (OWL).

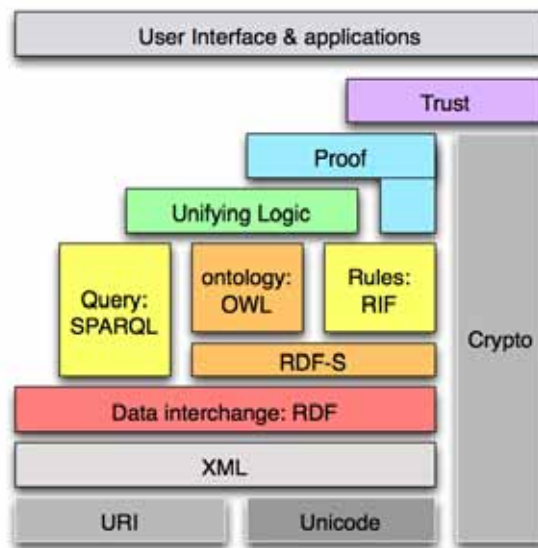
Mange av ulempene med mer tradisjonelle metoder for informasjonslagring lar seg løse med bruk av semantisk teknologi, nettopp fordi denne teknologien tilbyr bedre tilgang til kunnskap gjennom bruk av maskin-prosesserbare metadata. I dette kapitlet blir konseptene rundt semantisk webteknologi slik de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

Semantisk webteknologi illustreres ofte med bildet gjengitt i illustrasjon 6, som viser en såkalt "stack" (eller stabel) av teknologier som bygger på hverandre og til sammen utgjør en viss funksjonalitet. Bildet viser de laveste nivåene, internett-kommunikasjons-

protokollen HTTP (URI+Unicode), XML og RDFS på toppen av denne, samt OWL og noen sikkerhetsmekanismer og «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.



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

## Kort innføring i semantisk webteknologi



Internett legger til rette for kommunikasjon mellom mennesker og gir alle muligheten til å bidra med informasjon. Alle kan publisere informasjon om det de er opptatt av. Nettets suksess har vært enorm, men dessverre er mye av informasjonen på web generert av og for menneskelig fortolkning. Det gjør

automatisert prosessering av all informasjon i nåværende generasjon av internett vanskelig.

Semantisk webteknologi gjør det derimot mulig å publisere informasjon (og kunnskap) på internett i en form som ikke bare mennesker, men også datamaskiner kan lese og tolke. 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.

Et sentralt begrep her er "ontologi". Relatert til semantisk web defineres ontologi gjerne som "en formell spesifisering av en konseptualisering". 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

konsepser, ideer, meninger osv.

Teknologien som trengs for å implementere et slikt scenario bygger på den eksisterende internett-infrastrukturen, og krever derfor ikke investeringer i maskinvare for å komme i gang. Illustrasjon 6 viser at de grunnleggende elementene i semantisk web allerede finnes og er implementert.

Ikke alle nivåer er nødvendig for å gjennomføre forslagene i denne rapporten. Av nivåene over XML er det primært RDF, OWL og SPARQL som trengs. Til disse standardene finnes det meget robust open source programvare fra ulike miljøer samt kommersielle leverandører som IBM og Oracle. Flere andre globale aktører har signalisert at de vil følge W3Cs anbefalinger.

### 1.3.1 Unified Resource Identifier (URI)

Konseptet URI spiller en meget sentral rolle i digitalisering av informasjonskilder. Kort sagt har en URI som funksjon å identifisere en digital, internettbasert ressurs på en global, unik måte. En URI kan ses på som en nøkkel for en digital, virtuell representasjon av et objekt. URI-en 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.

### 1.3.2 eXtended Markup Language (XML)

Nær bunnen av «stacken» ligger det et nivå som heter eXtensible Markup 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>
```

XML gir muligheter for strukturert dokumentanalyse, såfremt 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 begynte derfor å tenke på representasjon av struktur på et mer abstrakt nivå, noe som førte til utviklingen av RDF og RDFS.

### 1.3.3 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 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»<sup>3</sup>.

<sup>3</sup> *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.*

- 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)

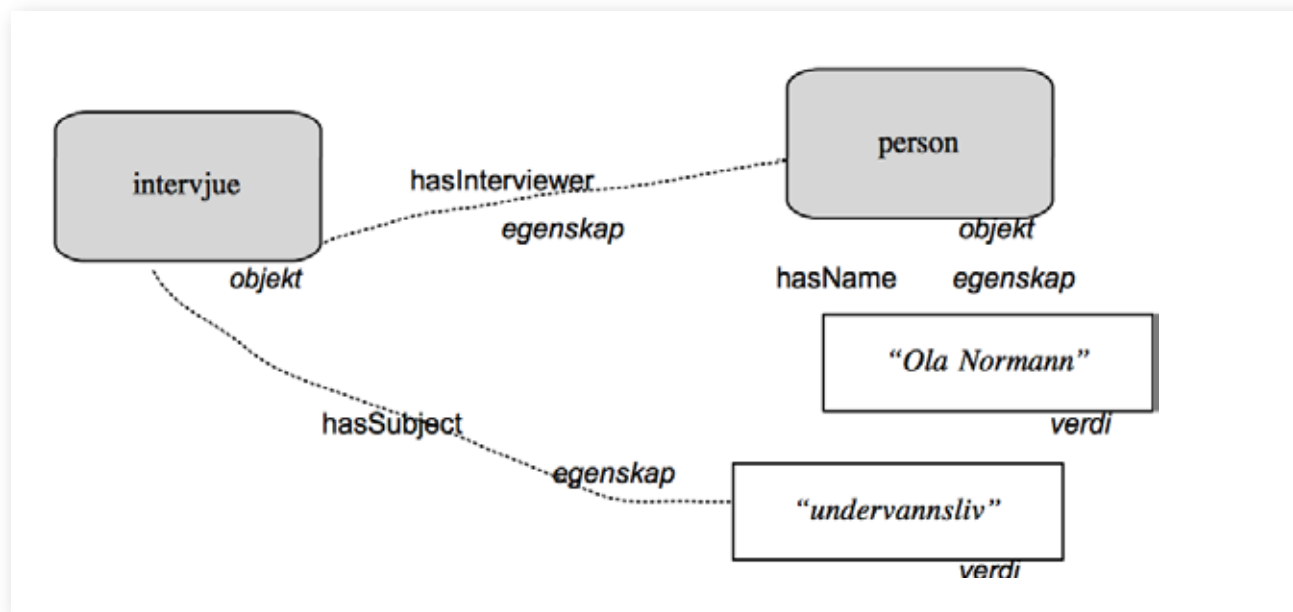
Siden RDF er en formell mekanisme for kunnskapsrepresentasjon, er det mulig å resonnerer og trekke konklusjoner fra modeller i RDF.

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 illustrasjon 7).

RDFS er en utvidelse av RDF, der "S" står for Skjema. RDFS er en mekanisme som tillater tolkning 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 bearbejdes av maskiner. På denne måten går annotasjonene i RDF(S) lengre enn hva XML kan tilby (Davies 2007):



Illustrasjon 7: Nettverk av OAV-»triples« (RDF)



- 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 veldefinert semantikk som tillater resonnering og sammenslåing av uttrykk fra forskjellige semantisk baserte systemer.
- 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 mangler 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

### 1.3.4 Web Ontology Language (OWL)

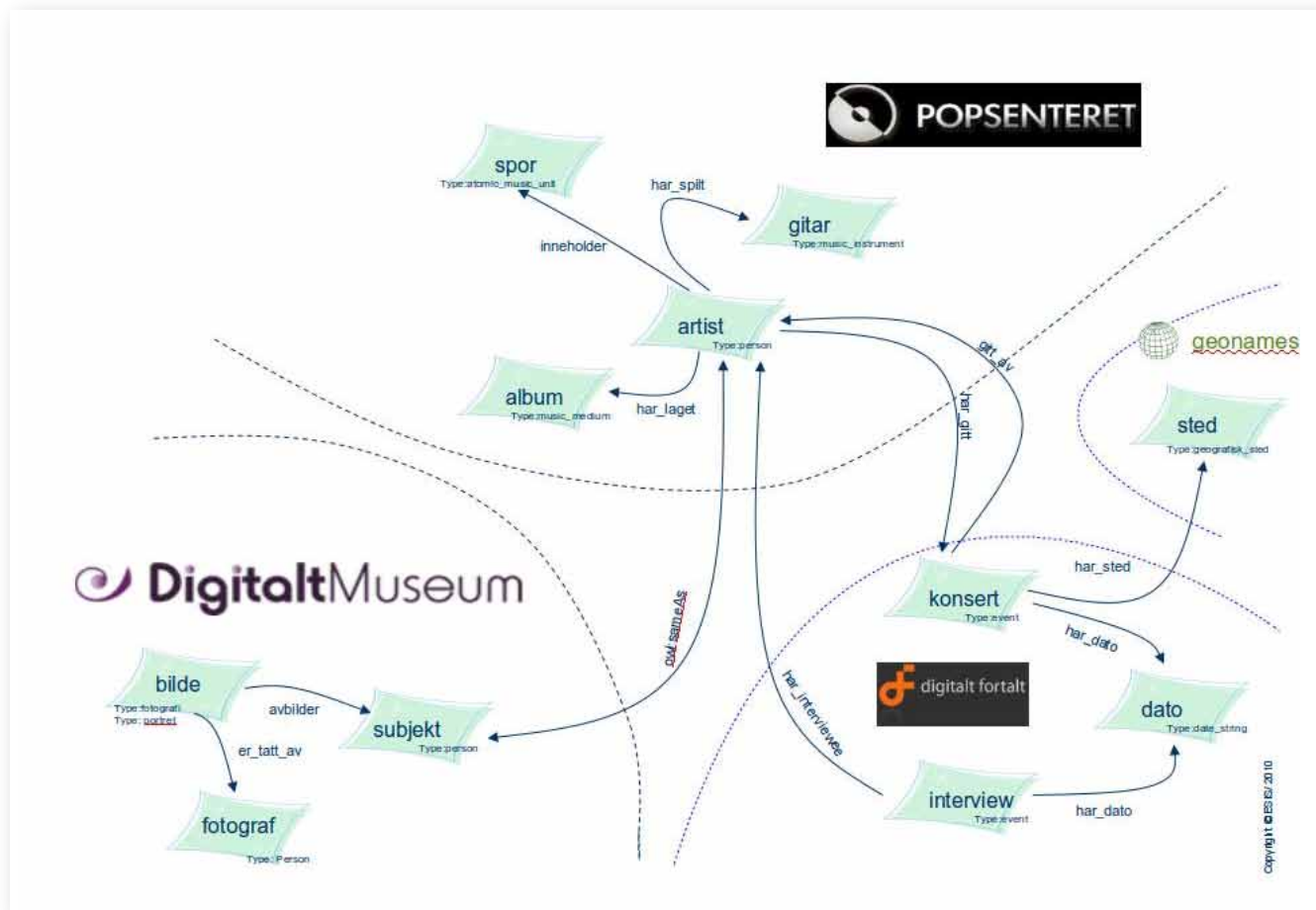
Web Ontology Language (OWL) er et relativt nytt forslag fra W3C. Det har som mål å tilby ekstra funksjonalitet som hverken XML, RDF eller RDFS har. Formålet er selvfølgelig å kunne beskrive kunnskap enda bedre. Der XML leverer et språk for utveksling av innhold og RDF/RDFS gir oss de byggeklossene vi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

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

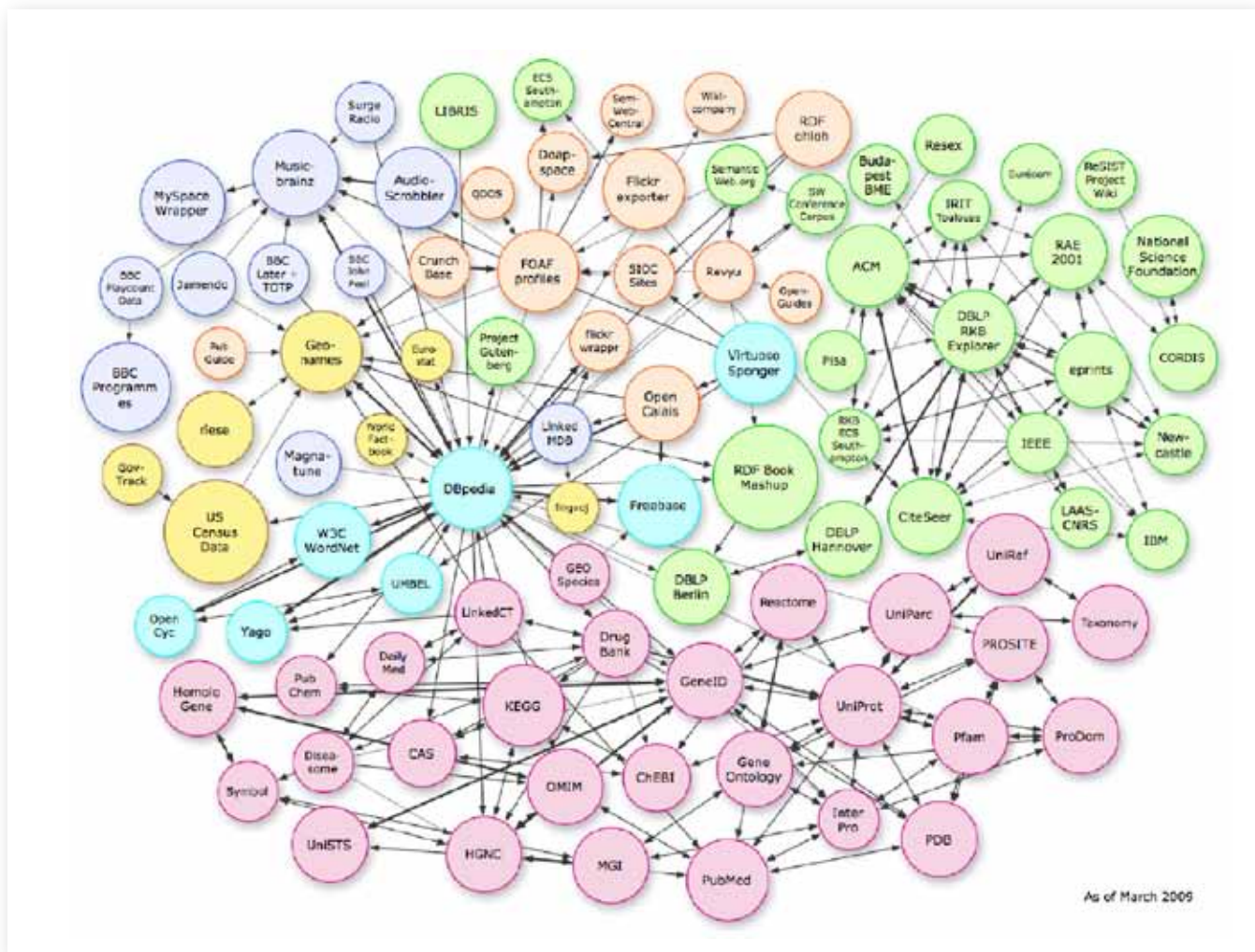
- “range”- og “domain”-restriksjoner. Man kan f.eks. si at en popgruppe må ha minst én person som har rollen “artist”
- “kardinalitet”-restriksjoner. OWL tilbyr minCardinality, maxCardinality og Cardinality. Med disse aksiomene kan man definere at et symfonisk orkester må ha minst 40 personer med rollen “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 ha 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

Dette gir OWL en kompleksitet som kan brukes for å beskrive de fleste domener. Når 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: Eksempel på samhandling mellom forskjellige internettbaserte datamodeller



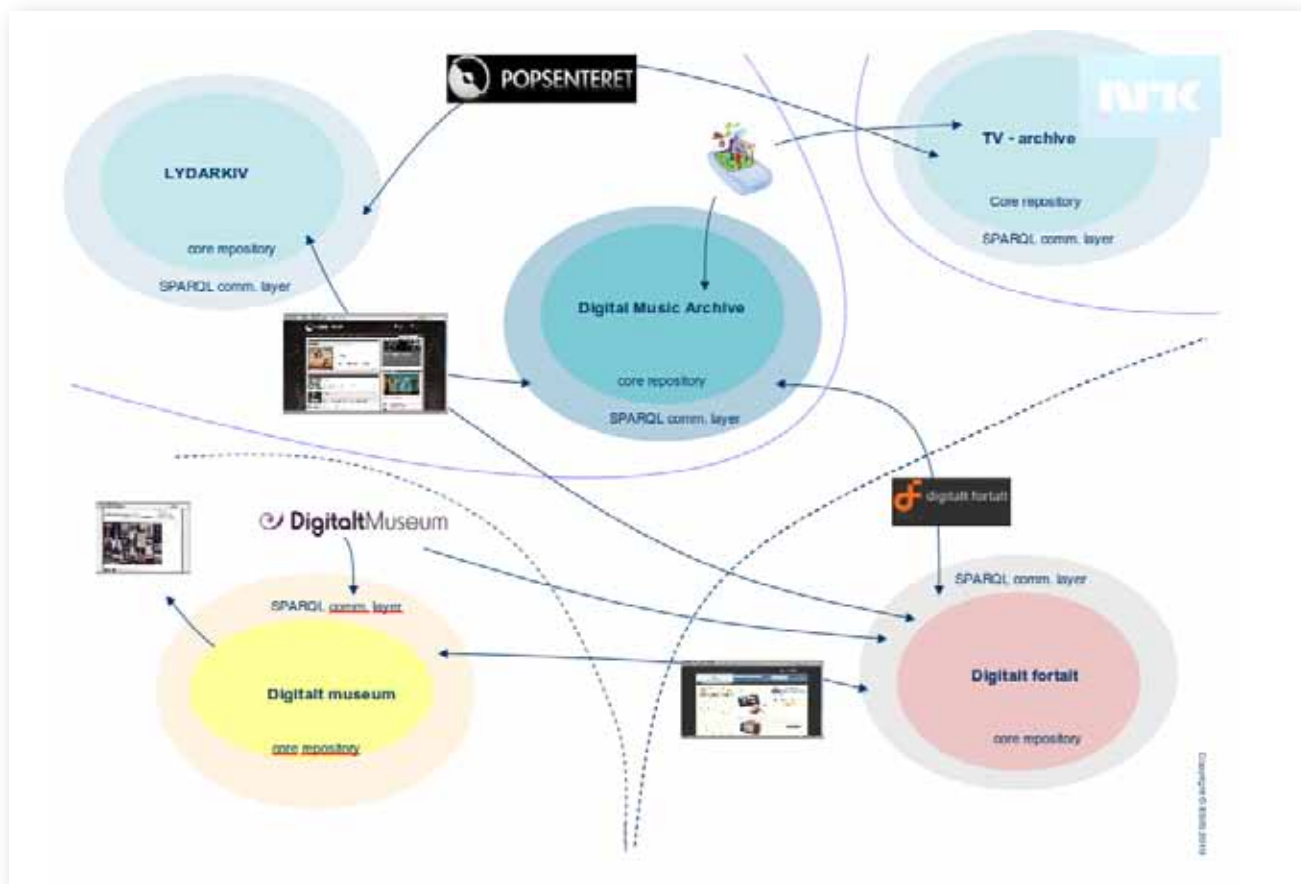
Illustrasjon 9: Linked Open Data- "skysten" - status 2009

### 1.3.5 Samkjøring av forskjellige kilder: Web Services og SPARQL

Internett slik vi kjenner det i dag består av mange datamaskiner som er direkte og indirekte koplet sammen med hverandre. Datamaskiner kan dele in-

formasjon seg imellom, og det finnes flere protokoller for 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 10: SPARQL-endepunkter som kommuniserer i et nettverk

For det første har den etablerte web service-teknologien utviklet seg slik at det har blitt mulig å fordele (distribuere) arbeidsoppgaver på 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 hatt til hensikt å etablere et leverandøruavhengig rammeverk, det vil si uavhengig av plattform, organisasjon eller nettverk.

Den andre trenden i evolusjonen av weben er den semantiske web, som fokuserer på deling av metadata. Den semantiske web har kunnskapsrepresentasjon som hovedmål (og dermed representasjon av innhold). Det åpner for utviklingen av “intelligent” programvare og agenter som 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 mellom datamaskiner, men dessverre 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 spørrespråket “The SPARQL Protocol and RDF Query Language” (SPARQL) for å kunne spørre andre datamaskiner om informasjon. Språket er basert på databasespørrespråket 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 Moldejazz 2007 og hva heter 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.

### 1.3.6 Linked Open Data (LOD)

Flere og flere organisasjoner beskriver sine datakilder ved hjelp av RDF/RDFS/OWL og tilgjengeliggjør dem slik for automatisk spørring. Gjennom kryssreferanser mellom slike åpne, standardiserte informasjonskilder koples innhold sammen til et globalt tilgjengelig kunnskapsnettverk. Linked Open Data er initiativet for å få koplet sammen så mange og forskjellige datakilder som mulig (se illustrasjon 10). )

Ser man litt nærmere på denne ”skyen”, finner man relativt mange velkjente informasjonskilder. Det er ingen som har lagt føringer på bruken av dem. Størrelsen på kildene angir hvor mye data som ligger i dem, tykkelse og retning på lenkene angir hvem som refererer 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 finner man en del kilder til musikk og film (bl.a. data fra BBC: lilla farge), offentlige kilder (bl.a. US census data: gullfarge), biomedisinske data (i rosa), og mange biblioteks- og forlagskilder (i grønt).

Det er verdt å merke seg at store og viktige ini-

tiativer som Europeana ikke har kommet så langt at de publiseres på samme måte (se egen diskusjon om Europeana under punkt 1.5).

## 1.4 Digitalisering av kulturarven i Europa

*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-traktaten, 1992

Europeana-prosjektet sikter mot en samling av den komplette europeiske digitaliserte kulturarven. På europeisk nivå er det de siste årene blitt satset tungt på initiativer som EUs 6. og 7. rammeprogram for forskning og utvikling, der kulturarv er et av satsingsområdene. I perioden 2007–2013 er det blitt satt av 400 millioner euro (3,2 milliarder kroner) til styrking av en felles europeisk satsing på bevaring og digitalisering av kulturarv.

Som ledd i arbeidet med rapporten har vi utført intervjuer ved forskjellige europeiske institusjoner. Vår undersøkelse viste at mange europeiske tråder innenfor feltet digital kulturarv samles i EU-kommisjonens generaldirektorat for informasjonssamfunn og medier i Luxembourg. Relevante og interessante initiativer fant vi også i Frankrike (Louvre, m.fl.), Storbritannia (British National Library, data.gov.uk, BBC open data) og Tyskland (hvor det finnes flere statlige aktører og prosjekter innen abm-sektoren).

Noen store internasjonale initiativer som The European Archive og Europeana har sine røtter i Amsterdam og Haag, hvor også en del andre svært

omfattende digitaliseringssatsinger blir koordinert av 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), 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 frem til et entydig helhetsbilde<sup>4</sup>. Digital kulturarv er fortsatt et område preget av mange mindre initiativer, noen store satsinger og lite faktisk koordinering på tvers av organisasjoner, regioner og nasjoner. Noen land kan vise til en større grad av koordinering, og Nederland er et eksempel på dette.

### 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 (Stichting DEN, 2009). Ifølge innrapporteringen fra Norge mangler en slik oversikt her til lands per i dag<sup>5</sup>.

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

4 [http://ec.europa.eu/information\\_society/activities/digital\\_libraries/experts/mseg/reports2010/index\\_en.htm](http://ec.europa.eu/information_society/activities/digital_libraries/experts/mseg/reports2010/index_en.htm)

5 [http://ec.europa.eu/information\\_society/activities/digital\\_libraries/doc/recommendation/reports\\_2010/norway.pdf](http://ec.europa.eu/information_society/activities/digital_libraries/doc/recommendation/reports_2010/norway.pdf)

- Som bakgrunn for rapporten er en spørreundersøkelse utført blant ca. 245 kulturarv-institusjoner. Av disse har 119 institusjoner (som blir omtalt som en representativ gruppe) gitt tilbakemelding om digitaliseringsaktiviteter.
- Gjennomsnittlig investering i digitalisering ligger på ca. 100.000 euro (ca. 800.000 kroner) per år og institusjon.
- Ca. 4 prosent av totalbudsjettet innenfor disse institusjonene går til digitalisering.
- Ca. 26 prosent av institusjonenes samlinger er blitt digitalisert. Her finnes det store forskjeller: En stor andel av samlinger med bildemateriale er digitalisert (65 prosent), mens audiosamlinger (6 prosent) og gamle trykksaker av ulike slag (0,3 prosent) nesten ikke er digitalisert i det hele tatt.
- Arkiver og biblioteker har gjort store deler av sine digitale samlinger tilgjengelig via internett. Museer gjør det derimot i liten grad, og tilbyr ofte ikke mer enn tilgang til ulike beskrivelser gjennom lokale PC-er 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 tilgjengelig.
- 75 prosent av alle digitaliserte samlinger er tilgjengelig gratis og på nett, 6 prosent av disse etter at man har registrert seg som bruker. Bare 5 prosent av samlingene er tilgjengelige mot betaling.
- Nesten ingen av institusjonene i Nederland bruker strukturerte metoder for registrering av kostnadene ved digitalisering.

## *Europeana*

<http://europeana.eu>



Det EU-finansierte tematiske nettverket Europeana har som mål å bygge opp ekspertise innen digitalisering av kulturarv. Det bygger på EDLnet (European Digital Library Network) og erfaringer fra prosjektet The European Library. Som en del av Europeana-prosjektet har man utviklet et nettsted (foreløpig som prototype) som aggregere og publiserer digitale kulturarv-objekter fra institusjoner fra hele Europa. Fra Norge deltar ABM-utvikling og Nasjonalbiblioteket, og en del av NBS 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 filoverføring (ftp), og deretter konvertert til formatet som Europeana bruker.

Et problem med 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 (Storbritannia) 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 og 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 fører til 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 Smith-er og David-er får man en blanding av alt som på en eller annen måte har med navnene David Smith å gjøre. Med andre ord sliter Europeana med det samme problemet som de fleste arkiver som skal knyttes sammen: Man er svært opptatt av 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. Istedenfor må man 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 av 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 staver et navn som «Peter Tchaikovski». Europeana klarer ikke å entydiggjøre søkene etter slike navn like godt som portaler som bruker semantisk teknologi (jfr. [www.evri.com](http://www.evri.com)). Det samme problemet dukker opp i de fleste arkiver når det gjelder stedsnavn, hendelser m.m.

Europeana bruker ikke URI-er, 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 (URI-er) for hvert objekt.



## *Het Geheugen van Nederland*

<http://www.geheugenvannederland.nl>



Het geheugen van Nederland

Het Geheugen van Nederland («Nederlands hukommelse») har som visjon å bli en Europeana på nasjonalt nivå, altså en aggregator for alle digitale kulturarv-samlinger i Nederland.

Metadata i «het Geheugen» er basert på en relasjonell databasemodell, men det brukes ingen URI-er 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 (ved hjelp av en søkemotor).

Portalen er meget innholdsrik, og mange forskjellige samlinger er allerede integrert. Et 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>



MultimediaN's semantic search capabilities

Et prosjekt som har tatt utfordringene skissert ovenfor på alvor er utformet av stiftelsen MultimediaN. Det viser 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 metadata-beskrivelser fra bl.a. samlingen til Rijksmuseum Amsterdam som både kan publiseres og kombineres med andre baser som bruker samme referanser (f.eks. personreferanser).

Sammenligner man MultimediaN med Europeana, kommer forskjellene tydelig frem. Brukeropplevelsen i MultimediaN 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 MultimediaN 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 potensial for republisering på nettet, f.eks. som en del av Linked Open Data.

Både programvaren og samlingene fra MultimediaN er tatt i bruk i Europeanas «Thought Lab». Her er det laget en testversjon av en semantisk søkemotor for Europeana - et prosjekt 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 kroner) for å digitalisere det som kalles «Nederlands audiovisuelle hukkommelse». Pengene er øremerket til å redde opptak som ligger på gamle og slitte video- og audiobånd. Totalt skal det digitaliseres 137.200 timer video, 22.510 timer film, 123.900 timer audio og 2,9 millioner bilder i perioden 2007-2014. Store deler av dette materialet skal også tilgjengeliggjøres på nett.

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

Et interessant delprosjekt innenfor Beelden voor de Toekomst er et åpent annotasjonsprosjekt hvor flere hundre bilder ble lagt ut på fotodelingstjenesten Flickr gjennom samarbeidet Flickr Commons. Her ble publikum invitert til å legge til annotasjoner i bildefilene - det vil si skrive inn sine egne tags (stikkord) og kommentarer til bildene. Prosjektet ble en stor suksess og førte til flere tusen annotasjoner av bildene, hvor samtidig mange tidligere ukjente personer på bildene ble identifisert.

## 1.5 Digitalisering av kulturarven 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 og Digitalt Museum, som er beskrevet nedenfor<sup>6</sup>:

### ***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, et system for forvaltning av museumssamlinger. 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. 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. Det går ikke frem av publiseringssystemet hvilken teknologi som er brukt.*



### ***Digitalt Fortalt***

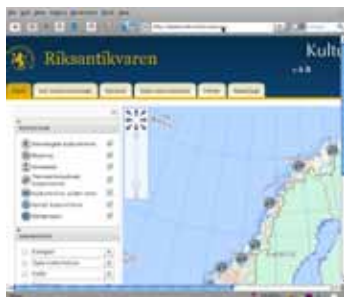
<http://www.digitaltfortalt.no>

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 bl.a. Digitalt Museum og Arkivportalen.

*Teknisk: utviklet av ABM-utvikling. Det går ikke frem av publiseringssystemet hvilken teknologi som er brukt. Bruker databasesøk på ID i URL-er for å vise objekter og underobjekter.*



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 kulturminnedatabasen Askeladden.*



### **Universitetsmuseenes 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-sidene. Gjenbruk er utelukket i dagens løsning.*



### **Universitetsmuseenes arkeologisamlinger**

<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-sidene. Gjenbruk er utelukket i dagens løsning.*



### **Universitetsmuseenes mynt- og medaljesamlinger**

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

*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-sidene. Gjenbruk er utelukket i dagens løsning.*

## **Europeana**

<http://www.europeana.eu>

Et europeisk initiativ for å presentere flere digitale samlinger på ett sted. 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») bygger imidlertid på teknologi fra det nederlandske MultimediaN-prosjektet og representerer et stort skritt videre i retning interoperabilitet og samhandling gjennom bruk av URI-er, semantikk og referanser mellom objekter.*



## **EuropeanaLocal**

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

Målet med EuropeanaLocal er å presentere 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 aggregasjonsprosjekter basert på samme tekniske plattform som blir utviklet til Europeana. Tanken er at man knytter sammen samlinger på nasjonalt nivå i en Europeana Local-portal, som i seg selv leverer resultatene videre til Europeanas 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 Nasjonalbibliotekets. 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 «egne 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 tekstsøk, som kan være upresise). Kulturnett bruker spørringer for å publisere informasjonssider, og disse er ikke egnet som URI.*



### ***lokalhistoriewiki.no***

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

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



### ***Store Norske Leksikon***

<http://www.snl.no>

Med snl.no vil Store Norske Leksikon videreføre den klassiske leksikontradisjonen der fagredaktører skriver artikler som kvalitetssikres av en redaksjon. Dette kombineres med bidrag fra publikum. Det er ikke rapportert hva omfanget av 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 "kvalifiserte" og "ikke-kvalifiserte" data og forskjellige tilgangsnivåer. Det bruker URI-er og metadata til å referere til andre identifiserbare objekter, akkurat som i en standard wiki.*



## **Wikipedia**

<http://no.wikipedia.org>

<http://no.wikipedia.org>

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

*Teknisk: Wikiløsninger bruker tradisjonelt tanken med unik identifikasjon og lenking av metadata gjennom URI-er. Ved bruk av wikiløsninger 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. Se DBpedia.org, som er et "ekstrakt" av Wikipedia.org i et semantisk format.*



## **Popsenteret**

<http://www.popsenteret.no>

Popsenteret, et prosjekt finansiert av Oslo kommune, har bygd opp en innholdsbase som foreløpig ikke er tilgjengelig i sin helhet for eksterne brukere (kun noe av senterets informasjon er publisert i en offentlig portal). Senteret er likevel tatt med i denne listen med portaler, fordi det har valgt å bruke semantisk teknologi i sin driftsløsning (som inkluderer sluttbrukerstasjoner). Arkivsystemet er laget med 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 multimediasasjoner og digital skilting.

*Teknisk: bruk av Linux, Protégé og MySQL. Web Service 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  | RSS/søk                     |
| Kulturminnesøk                                 | Ja (location)                               | Nei                         | Nei  | Ukjent                      |
| Universitetsmuseenes fotoportal                | Sannsynligvis (class_id)                    | Nei                         | Nei  | Ukjent                      |
| Universitetsmuseenes arkeologisamlinger        | Nei   | Nei                         | Nei  | Ukjent                      |
| Universitetsmuseenes mynt- og medaljesamlinger | Sannsynligvis (class_id)                    | Nei                         | Nei  | Ukjent                      |
| Europeana Local                                | Ja  | Ja                          | Nei  | Ukjent                      |
| 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                      |
| Wikipedia (no)                                 | Ja  | Ja                          | Ja   | Ukjent                      |
| Popsenteret                                    | Ja  | Ja                          | Ja   | Ja                          |
| Europeana                                      | Ja  | Ja                          | Nei (Prototype: Ja)                          | Ja                          |

Illustrasjon 11: Sammenligning av forskjellige portaler i forhold til fire spørsmål: publiseres det nøkler til objekter, URI-er, inneholder meta-dataene referanser til andre objekter og mulighetene for automatisert tilgang til dataene gjennom web services o.l.

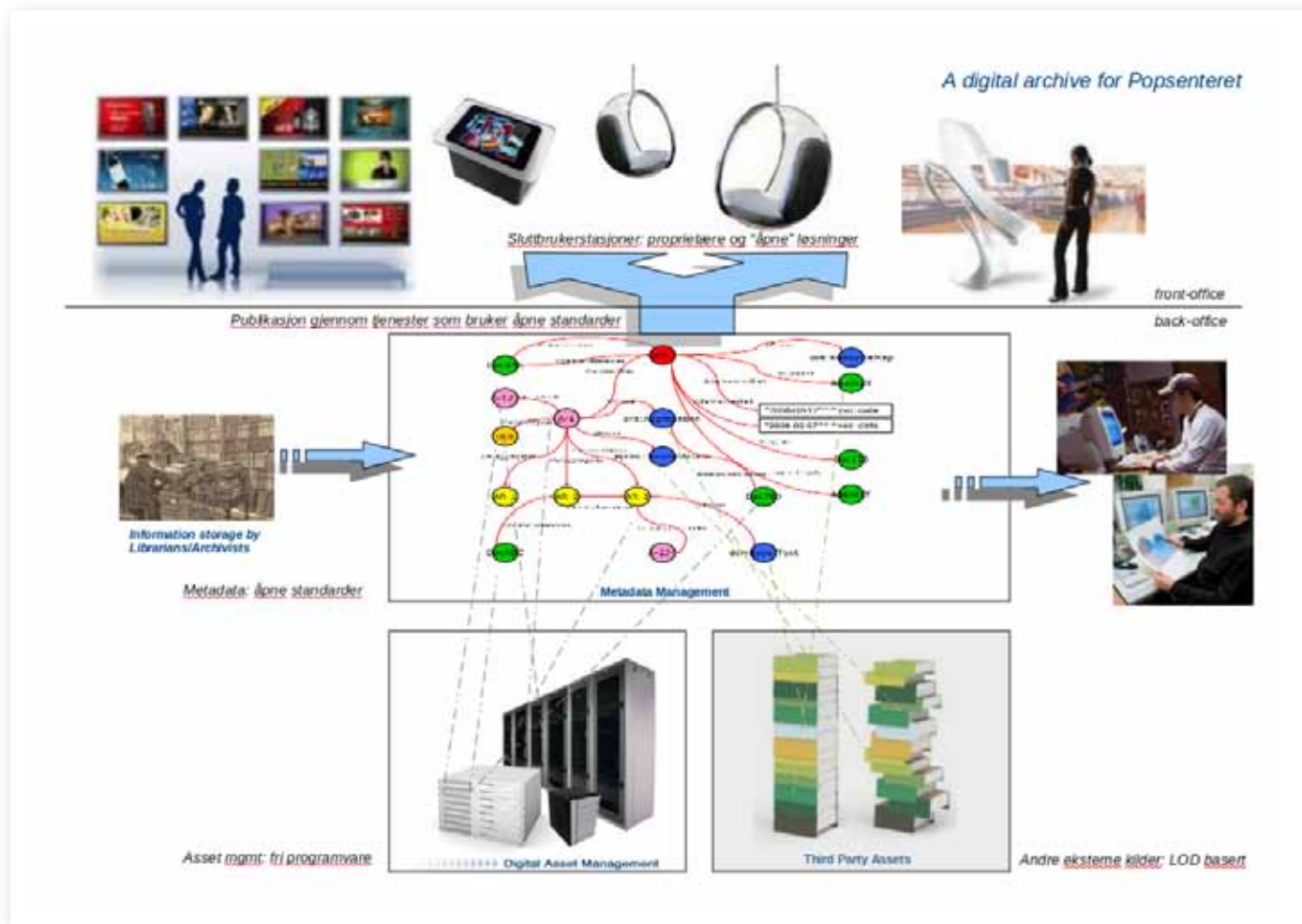


## 1.6 Utfordringer ved dagens satsing på digitalisering av kulturarv

Når gjennomgangen av europeiske og norske initiativer ses i sammenheng med den teknologiske utviklingen i retning av økt bruk av semantisk webteknologi, avdekkes noen utfordringer ved måten digitalisering av kulturarven gjøres på i dag.

- Man baserer seg i for stor grad på en typisk arkiv/bibliotek-tankegang, der fokuset gjerne ligger på åndsverk. Man er ikke i samme grad opptatt av 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. Dagens tanke bak mange portalløsninger er at gjenfinning er viktigere for brukerne enn utforskning. Det er ofte nødvendig for brukeren å vite på forhånd hvem (aktøren) eller hva (objektet, tidsperioden etc.) hun er interessert i, ellers får hun lite ut av en slik portal. En slik fremgangsmåte tillater bare i liten grad brukeren å finne relasjoner og forbindelser som ikke er åpenbare (f.eks. hvilke andre malere som var aktive i samme by som Picasso da han malte «Guernica» i 1937).
  - URI-er begynner å bli brukt som pekere til åndsverk og objekter (se f.eks. Koninklijke Bibliotheek, Nasjonalbiblioteket). For å bli brukt i stor skala burde slike portaler helst ha en visjon om varighet. Man kan f.eks. garantere at et bestemt objekt vil være tilgjengelig/bli referert til under denne URI-en i de neste 100 årene.
  - Åpen tilgang er en stor utfordring. Brukerne skal sikres tilgang til tjenesten uavhengig av hvilken teknologi/leverandør man måtte velge. Mange begrensninger på tilgang og bruk gjør en relativt god start vanskelig å videreføre.
  - Metadata i nesten alle løsningene (inkludert 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 visjonen om den semantiske web (eller «the internet of things»), følges opp. Dette innebærer at man lager digitale representasjoner av aktører, hendelser, objekter, steder osv. som det aktivt refereres til i digitale uttrykk. Med andre ord skal metadata referere til unike og globalt identifiserbare objekter.
  - Standardiserte «ontologier» for metadata i abm-sektoren blir kun i liten grad publisert. 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).
  - Altfor mange prosjekter setter av altfor 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 «sammensmeltingen» 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 for-

skjellige behov innenfor abm-sektoren, men også strenge nok til å kunne lede til noen minimumsgarantier.



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

## 2. IKT-STRATEGI FOR FREMTIDSRETTET DIGITALISERING OG FORMIDLING AV KULTURARVEN<sup>1</sup>?

Begrepene **back-office** og **front-office-systemer** refererer til et mye brukt perspektiv på programvare-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 brukes vanlige nettlesere 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

Deler av dette kapitlet er oversatt fra og inspirert av rapporten *Business Model Innovation - Cultural Heritage*, utgitt av DEN Foundation (2010) under Creative Commons-lisensen "Navngivelse-Del på samme vilkår" (3.0/nl). I tillegg har vi fått godkjenning av forfatterne.

Allerede på 1970-tallet ble de første digitaliserte systemene utviklet i abm-sektoren. Disse systemene førte til en ny type publikumsrettede tjenester, hvor det ble mye enklere og raskere å foreta bestemte søk

<sup>1</sup> 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 world wide web slik vi kjenner det i dag. Han leder standardiseringsorganisasjonen World Wide Web Consortium (W3C) og har engasjert seg sterkt i arbeidet for frigivelse av offentlig sektors data.



Illustrasjon 13: Papirbasert kartotek. Foto: Wikipedia



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

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 1980-tallet, ble terminalene gradvis erstattet med PC-er eller arbeidsstasjoner med mer kapasitet. Dermed ble det også mulig for mindre abm-institusjoner å åpne opp sine samlinger digitalt. Mange institusjoner designet og utviklet egne databaser, basert på standard eller spesiallaget programvare. Slike prosjekter var ofte basert på egne erfaringer og eksisterende (kartotek-) systemer. Etter hvert, når også planskannere og digitale kameraer ble tilgjengelig, ble det vanligere å digitalisere bilder selv og knytte dem til beskrivelser av objekter i databasen. Ved siden av slike databaser med metadata om samlingene, ble det mer og mer vanlig å lage kunnskapsbaserte innholdstjenester som referanselister, indekser, bilder, tekstbaser og utstillingskataloger.

På 1990-tallet kom world wide web, og dermed eksploderte mulighetene for informasjonsutveksling. Abm-institusjoner ble plutselig del av et globalt kunnskapsnett. ”Verdensveven” tilbød standardisert, enkelt tilgjengelig teknologi som et enhetlig og integrert grensesnitt mot mange ulike typer informasjonssystemer.

## Kulturarv og teknologi – dagens tilstand

Som et resultat av den raske teknologiske utviklingen siden tidlig på 1980-tallet finnes det mange veldig forskjellige IKT-løsninger i abm-sektoren. På 1970-tallet ble det utviklet databaser som skulle «overta» for kartoteker og gjøre dem tilgjengelig i digitalt format. Løsninger som SIFT («Søking I Fri Tekst» - 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 grunn-

leggende relasjoner mellom informasjonselementene ikke finnes og må legges til (semi-) manuelt.

Siden relasjonsdatabaser (Relational Database Management Systems, RDBMS) ble kommersialisert tidlig på 1980-tallet, har man sett en kontinuerlig utvikling av informasjonsbaser som dekker et spesifisert behov innenfor de institusjonene de ble laget for. Denne utviklingen er preget av spesielløsninger og tilpasninger. Felles for alle disse løsningene er at de er fra «pre-www»-tiden (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. Resultatet ble ofte en hverdag der brukerne måtte forholde seg til mange forskjellige systemer, som igjen gjerne var utformet gjennom prosjektbasert utvikling med et eget mål som dekket et spesifikt behov. Til slutt endte man opp med en jungel av systemer og et forvaltningsproblem.

En direkte følge av utviklingen av slike hybride og sameksisterende IKT-systemer er ofte at de ulike digitale løsningene i en organisasjon holder lav kvalitet. 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 på nett i det hele tatt (eller bare delvis), selv om de faktisk har blitt digitalisert. Og selv om de er tilgjengelig online, så er det ikke dermed sagt at man kan integrere dem i web 2.0- eller semantisk web- løsninger (web 3.0).

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 konklusjonen stemmer godt overens med tanken som formidles i *St. meld. nr. 24 (2008-2009) Nasjonal strategi for digital bevaring og formidling av kulturarv*:

*"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 foreslår at perspektivet løftes fra det nasjonale til det internasjonale nivået.

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-infrastrukturen.

#### Fornuftige løsninger

Det krever kontinuerlige investeringer for å få til verdiskaping ved hjelp av IKT-infrastruktur. Det finnes flere faktorer som spiller en rolle, men mest relevant i diskusjonen om digital kulturarv 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. Modernisering og oppgradering av eksisterende løsninger blir ofte utsatt. En sannsynlig årsak til dette kan være at man har bestilt eller laget spesielle engangsløsninger, som ikke har tatt nok hensyn til standardisering.

## Bruk av standarder og generisk teknologi

Abm-sektoren sitter på mye unikt og ofte sjeldent materiale. Det er ofte disse unike objektene som gir en institusjon eller organisasjon dens identitet eller eksistensberettigelse. De unike objektene utgjør tilleggsverdien som institusjonen kan tilby i et digitalt informasjonssamfunn. Hvis en institusjon 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 frem hva man har til et så bredt publikum som mulig. Jo flere som blir kjent med samlingene, jo mer relevante blir de. Dette innebærer at informasjon om samlingens objekter publiseres også utenfor institusjonens vegger, og ikke bare av institusjonen selv. Verdien av en samling øker betraktelig når den blir tilgjengeliggjort for mangesidig bruk av mange.

Den kanskje viktigste faktoren for å få til spredning av informasjon om ulike samlinger er at teknologien legger til rette for det. Dette innebærer at man må

tenke standardisering av IKT i tråd med åpne og helst globale standarder. For selv å 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») bør 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 enn en enkelt organisasjon forvalter, uansett hvor mye det måtte være. Brukerne forventer tilgang til flere baser og informasjonskilder samtidig, og det er ofte uvesentlig for dem hvor informasjonen kommer fra (web, institusjonen selv osv). Søk i 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 kombineres etterpå. Dette innebærer at resultatene av et søk må inneholde gode beskrivelser som plasserer objektet i riktig kontekst og muliggjør unik identifikasjon.

Popsenteret i Oslo har valgt å bruke en slik struktur for sitt arkiv og sin driftsløsning som baserer seg på bruk av åpne (W3C-) standarder og proprietære "front-office" systemer. Erfaringene har vært positive så langt, og mye data om norsk populærmusikkshistorie 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 abm-institusjoner, 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 går glipp av mye, spesielt når det gjelder tilgjengelighet av informasjon. Det er for eksempel fortsatt ganske vanskelig å finne digitaliserte kulturarvobjek-

ter gjennom de største søkemotorene som Google og Yahoo. Dette er en godt dokumentert effekt av mangel på standardisering (van Kersen, 2004).

Det er vanskelig å tilpasse digitale tjenester til endrede behov hos sluttbrukerne hvis tilpasninger i IKT-systemene må skje på ad hoc-basis. Man vil trenge dyr spesialkunnskap for å videreutvikle systemene hvis man ikke klarer å kople 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. Det finnes eksempler på at digital kulturarv ligger i baser som driftes og forvaltes av kommersielle aktører. Ideelt sett bør kulturarv-organisasjoner forvalte 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.

Mye som følge av web 2.0-utviklingen har brukere og utviklere på internett kastet seg over integrerte applikasjoner, der informasjon hentes fra mange ulike kilder. Derfor er ikke det interessante spørsmålet om informasjonen på internett vil bli gjenbrukt av andre - det kan vi ta for gitt. Spørsmålet er snarere: hvem er med på å levere innhold og data til publikum? Her bør kilder av høy kvalitet som er betalt av og bygd opp ved hjelp av offentlige midler ha en viktig rolle.

For å **tilrettelegge innhold for deling og interoperabilitet samtidig som hver institusjon beholder 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 nettlesere å gjengi websider på en standard måte. I tillegg kan bruk av åpne standarder gjøre IKT billigere ved at funksjoner som ikke trengs å utvikles separat kan utvikles i fellesskap. Enda viktigere er det at man, gjennom en slik standardisering av tilgang, kan åpne opp mer av de digitaliserte samlingene.

En annen viktig utvikling innenfor IKT er den mye omtalte satsingen på «cloud computing» eller nettskyen. Med «cloud computing» tilbys storskala, skalerbare og nettbaserte tjenester, der forvaltningen er sentralisert hos noen store bedrifter som Google eller Amazon. For mange aktører innenfor abm-sektoren vil det ikke være aktuelt å legge ut sine data i nettskyen, mest fordi det er uavklart hva dette innebærer i forhold til rettigheter, kontroll og problemer med å plassere viktige nasjonale ressurser utenfor landets grenser.

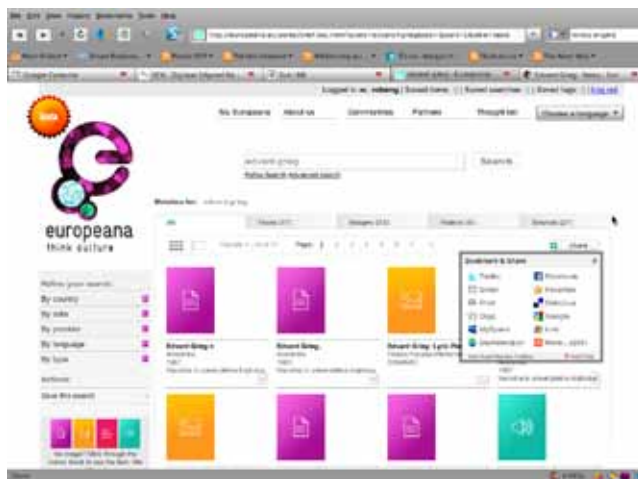
Visse typer bruk av nettskyen er ikke like problematisk. Kulturarv-institusjoner har allerede tatt i bruk nettskybaserte tjenester for å heve kvaliteten på samlingene. «Beelden voor de Toekomst» bruker Flickr for å få hjelp av brukerne til å legge metadata til sine bildesamlinger, en mulighet Bergen Offentlige Bibliotek og Fylkesarkivet i Sogn og Fjordane også har benyttet seg av (Moortgat, 2009). En annen motivasjon for å bruke nettskybaserte tjenester er å nå ut til publikum med tilbud som er like de opplevelsene som brukerne allerede er vant med fra andre sammenhenger (iPhone er ett aktuelt eksempel).

IKT er blitt så modent og generelt tilgjengelig at man i dag kan lage gode storskala digitale tjenester uten dyp kunnskap om IKT. Suksessen er for en stor del bygget på åpne standarder og generisk teknologi.

## Samhandling og identitet

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

Vi har merket oss ett område hvor institusjoner har stor tillit til hverandre, og det er når man skal sette bort arbeidet med langtidssoppbevaring av digitale data. I flere europeiske land går institusjoner sammen for å danne felles fasiliteter for dette formålet.



Søk på "edvard grieg" i Europeana.eu

Ett tema gikk igjen og ble fremhevet som sentralt i alle våre samtaler og diskusjoner om den videre utviklingen av digital kulturarv-sektoren. Det var behovet for «persistent identifikasjon» av digitale kulturarv-objekter. Dagens praksis, der mange organisasjoner for eksempel bruker tekstbaserte «autoritetslister»

for person- og stedsnavn, er ikke tilstrekkelig. Det er viktig at alle ting på en slik liste refereres til 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 og tjenester. Dette er den beste og enkleste måten å legge til rette for fremtidige bruksmåter.

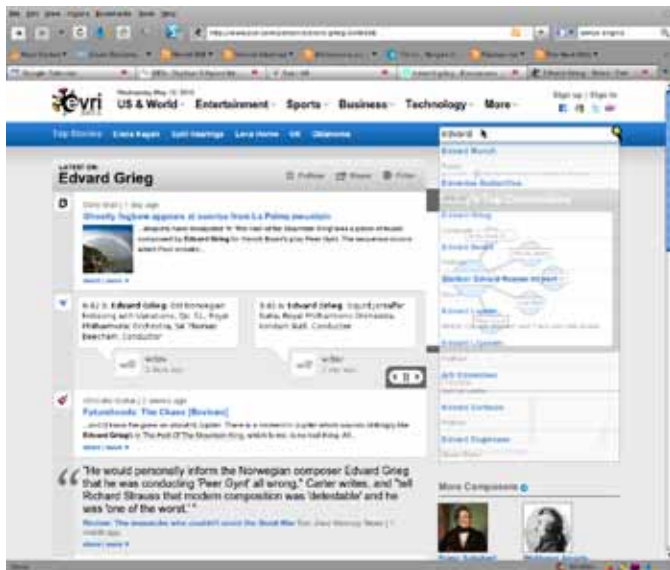
Får man til konsensus om et slikt rammeverk, blir det mulig for alle å hente inn, kombinere og publisere data. Samtidig gis alle muligheten til å bidra med ny kunnskap eller nye ytringer om unikt identifiserbare objekter. En slik permanent identifikasjon sørger for at objektene og relasjonene mellom dem er sporbare over tid («long term traceability»), og dokumenterer objektets opprinnelse i tilfeller hvor det finnes flere like digitale objekter. En forutsetning er at alle «persistente identifikatorer» er unike, og det finnes flere toneangivende forskningsmiljøer som har kommet med forslag om hvordan dette kan garanteres.

Eksempel: La oss si at man ønsker å lære mer om Edvard Grieg og hans verker. Et søk på «Edvard Grieg» på Europeana.eu ga 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 tilgjengelige ressurser om komponisten.





Samme søk på nb.no ga et tilsvarende resultat. Heller ikke denne gangen blir man fortalt 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.



Foretar man derimot samme søk på den kommersielle portalen Evri.com, vil systemet identifisere søkeordene som en mulig komponist (se illustrasjon 17). Portalen (som henter sin informasjon bl.a. indirekte fra Wikipedia) «vet» at det 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 valgt 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 andre objekter knyttet til seg.

## Oppsummering: IKT-veivalg i abm-sektoren

Det er hovedsakelig to faktorer som står sentralt i diskusjonen om IKT-strategi i abm-sektoren:

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 skape 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.

Bruk av åpne standarder og unike identifikatorer (URI) på alle metadataobjekter i «back-office»-systemer, og bruk av generisk teknologi som langtidsstrategi, fremstår som alternativet som gir størst uttelling for digitale investeringer.

# 3. VEIEN VIDERE: HVORDAN KAN VI LØFTE ABM-SEKTOREN?

De siste årene har brukere i økende grad blitt involvert i oppbygging av informasjonskilder og portaler. Digitaliseringsbølgen har sørget for at mange mindre, men interessante samlinger har blitt gjort tilgjengelig for publikum. Endringene gjør det nødvendig for abm-sektoren å fornye seg, og skaper også nye muligheter. For mange institusjoner som har mindre samlinger og begrensede budsjetter, er det ikke aktuelt å lage selvstendige, gode løsninger som skal nå et stort publikum. Derimot kan de veldig godt samarbeide med andre, komplementære samlinger om å lage spennende tema- og fagfeltbaserte opplevelser som formidler kunnskap. Dette forutsetter at man må kunne publisere, dele og integrere metadata om samlingsenes objekter på en enkel og standardisert måte.

I denne fornyelsesprosessen er det viktig å tenke helhetlig. 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 vil stå 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 den også befinner seg. Data behøver ikke flyttes hvis man ikke ser spesielle behov for det. Det viktigste er at man legger til rette 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 for kunnskapsdeling er på plass**

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

## **Gi dine brukere tilgang til informasjon fra hele abm-sektoren**

En institusjon bør konsentrere ressursene om å bli enda bedre på eget fagområde. Kompletter opplevelsene du skal lage til dine brukere med informasjon på høyt kvalitativt nivå fra andre organisasjoner. Ingen er

best på alt, men hver organisasjon har sin egen spesialkompetanse.

Det finnes mange situasjoner hvor de mest brukte ressursene som f.eks. Wikipedia ikke dekker behovet for god og pålitelig informasjon, og publikum er klar over dette. Derfor er det viktig at abm-sektoren presenterer kvalitetsinformasjon på en lett tilgjengelig måte med så mange innganger som mulig. Det er bare noen få kilder som klarer å bli «the place to go» når man ønsker informasjon om et bestemt tema, og derfor er det meget viktig at institusjoner bidrar til å gjøre disse kildene kjent.

*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 Rembrandts liv. Hvordan kan vi sørge for at brukeren finner denne informasjonen?*

Mest sannsynlig får man slik informasjon bare gjennom viderefremming fra andre portaler, og det er derfor riktig og viktig at man tilrettelegger abm-sektorens kunnskapsformidlingsportaler slik at de kan ha en slik veiviserfunksjon.

### Gi abm-institusjoner støtte til å definere og gjennomføre en god digitaliseringsstrategi

Abm-institusjoner som planlegger eller nettopp har begynt å digitalisere sine objekter og ressurser, må ofte finne svar på de samme spørsmålene. Hva skal vi ta hensyn til? Hva må på plass? Hva skal vi prioritere? Må vi tilrettelegge eller passe på noe ved IKT-arkitekturen eller -satsingen vår før vi setter i gang med selve digitaliseringen? Spesielt de mindre aktørene har ikke nødvendigvis all nødvendig kompetanse på digitalisering av kulturarv i egen organisasjon. Maskinvare som

trengs 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

## 3.1 Definerer av krav til digitaliseringsprosjekter

For å få til en felles retning på digitaliseringen i 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 minimumskrav til digitaliseringsprosessen. Disse må følges opp med veiledning og kurs, og burde helst koples til finansiering og insentiver for å sikre at de blir tatt raskt i bruk. Flere europeiske land har begynt å tilby slik støtte i samarbeid med abm-sektoren. Erfaringene er så langt veldig positive, og de første fruktene kan plukkes (se f.eks. Europeana-prosjektet)

### Forskjellige kravpakker for digitalisering av kulturarv

For å være forståelige og anvendelige burde slike minimumskrav helst deles opp i forskjellige ”pakker”. Basert på erfaringer rapportert av nederlandske DEN foreslås det å starte med fem forskjellige pakker som inneholder følgende:

- krav til lokalisbarhet
- krav til produksjon av digitale objekter
- krav til beskrivelse/metadata
- krav til varighet
- krav til presentasjon

## 3.2 Veiledning og kurs

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

### 3.3 Finansiering og incentiver

Hvordan koordinerer man slike større, fremtidsrettede initiativer? Vi har diskutert dette punktet spesielt med alle institusjoner vi har snakket med, og fått mange forskjellige tilbakemeldinger. Kunnskapsdepartementet i Nederland har valgt å kople finansieringsmuligheter for digitaliseringsprosjekter til 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. Strategien virker vellykket på mange fronter. De første museene har begynt å formidle samlingene på en felles måte.

På europeisk nivå står Europeana sentralt i strategien. Europeana fungerer som et stort nettverk og incentivene for deltakerne består av økonomisk kompensasjon samt bedre innsikt i hvordan egne samlinger henger sammen med andre europeiske samlinger.

### 3.4 Deltagelse og diskusjonsfora

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

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

# DEL 2

## MINIMUMSKRAV TIL DIGITALISE- RING AV KULTURARV

I denne seksjonen blir det skissert noen minimumskrav som foreslås brukt som retningslinjer for abm-sektoren. Kravene er basert på anbefalinger og erfaringer fra Digitaal Erfgoed Nederland. Nederlandske myndigheter krever at et prosjekt eller en institusjon som et minimum forholder seg til disse fem kravene før statlig støtte til digitaliseringsprosjekter kan gis. Flere krav på andre områder blir diskutert.

Det er viktig å huske at disse minimumskravene er formulert av den nederlandske abm-sektoren selv i samarbeid med departementet. Alle diskusjoner om krav og grunnlag føres i en åpen wikiløsning<sup>7</sup>.

---

<sup>7</sup> Wikiløsning for å diskutere og bidra til minimumskrav for digitaliseringsprosjekter: [http://wiki.dan.nl/Main\\_Page](http://wiki.dan.nl/Main_Page)

# 1. KRAV TIL LOKALISERBARHET

Lokaliserbarhet betyr at man gjør objekter digitalt tilgjengelig 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 tolke for datamaskiner (eXtensible Markup Language).
4. **UTF8** er et internasjonalt codesystem for karakterer (tegn). Ved bruk av UTF8 kan man gjengi internasjonale tegnsett.
5. **Dublin Core** er en beskrivelsesstandard for metadata som blir brukt på verdensbasis for å beskrive samlinger av objekter.
6. **OAI-PMH, alternativt W3C-standarder.** OAI-standarden er laget for å gjøre overføringer mellom databasesystemer enklere. (Open Archive Initia-

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

7. **SRU** SRU-protokollen (Search/Retrieval via URL) 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 seks 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 først å utrede nytteverdien av å bruke de nyere internetstandardene fra W3C som definerer formater for informasjonsutveksling (RDF/RDFS, evt. OWL).

Hvis man gjør databaser tilgjengelig 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-baserte meta-

data. Gjennom disse to W3C-standardene blir alle informasjonskildene som bruker dem «transparente» (gjennomsiktige) og kan adresseres og gjenbrukes direkte og automatisk fra hvor som helst på internettet. På grunn av 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-spøringer 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 av spøringer 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 TIL PRODUKSJON AV DIGITALE OBJEKTER

Kravet til produksjon definerer minimumskrav for digital reproduksjon av originalt analogt kulturarv-materiale. Dette skal garantere at sluttresultatet av digitaliseringen har optimal kvalitet og en informasjonsverdi lik originalens (innhold, struktur, formgivelse) samt sikre fremtidig tilgjengelighet. Kostnaden skal være akseptabel (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 typer objekter. Kunnskapen om digitaliseringsteknikker og digitale formater er ujevnt fordelt. Gjennom et slikt «krav til produksjon av digitale objekter» kan sektoren stille bedre og mer begrunnede og rettede krav internt, men også til 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 over tid. Utgangspunktene burde være:

- **Produsér én gang:** Sluttresultatet av digitaliseringsprosessen burde være en digital master. Det koster mest å ta frem originalen og digitalisere

den, derfor er det viktig at den digitale masteren har maksimal 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, eventuelt med formål om å konvertere til bedre alternativer hvis slike skulle dukke opp. På denne måten sikrer man seg også at de fleste typer av senere bruk blir ivarettatt.
- **Erstatning:** En viktig diskusjon er om «kravet til produksjon» 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 lokaliserbarhet» blir metadata definert som tilrettelegger for gjenfinning og sammenkobling av informasjon om objekter. Denne typen metadata kalles også «sosiale metadata». I «kravet om produksjon» er metadata forstått som data som beskriver de tekniske detaljer om objektet, som filnavn, kvalitet, oppløsning, format osv, og refereres derfor gjerne til som «tekniske metadata».

- **Unik, global identifikator (også som filnavn):** I praksis skjer det ofte at filer får samme navn (f.eks. DSC000478.JPG eller innhold.txt). Dette fører gjerne til problemer senere i prosessen. 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 tar for seg 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-sektoren), men også noen (uavhengige) eksperter fra IKT-bransjen, lager og vedlikeholder kravene.

Det er viktig at man stiller klare minimumskrav til objekter av typene *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 anvendelser: å kople romlig (spatial) informasjon til fysiske objekter (metadata), digitalisering av analogt kartmateriale og produksjon av digitale kart med informasjonselementer plottet inn.

# 3. KRAV TIL BESKRIVELSE

Gode beskrivelser av kulturarven er essensielt for å få til gode formidlingstilbud. 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 at kulturarven kan formidles til mange.

## 3.1 Metadata

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

- **Beskrivende metadata:** tillater identifikasjon og gjenfinning av et objekt. I praksis kan dette være et filnavn på lokal harddisk eller en URI på internet.

- **Strukturelle (sosiale) metadata:** beskriver objektene og deres karakteristikk og sammenhenger/relasjoner.
- **Tekniske metadata:** gir informasjon om produksjon, format osv.
- **Administrative metadata:** er rettet mot forvaltning og lagring av objekter. Kan også inneholde juridiske metadata som rettigheter, eierskap osv.

Det finnes forskjellige modeller for metadataforvaltning. Det er vanlig at metadata lagres uavhengig av objektene de beskriver, og bare vedlikeholder 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 kameraets innstillinger, dato og tid for når et bilde ble tatt med mer, ofte lagres i 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 institusjoner arbeider godt med 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 vurdere 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 webteknologi som kan være interessant for utveksling av slik informasjon er W3Cs rammeverk SKOS (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 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 legger ikke til rette for 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

Under kravet om «lokaliserbarhet» ble det blant annet foreslått at Dublin Core, XML og URI må brukes. Et minimumskrav for «beskrivelse» av kulturarv burde være bruken av XML/Dublin Core for å beskrive samlinger, objekter, personer osv. Viktigst i dette kravet er det å identifisere standarder og teknologier som kan heve kvaliteten på 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 som et minimum 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.

Nederlandske DEN foreslår en tilpasning av disse kravene til Dublin Core som ser slik ut:

|         |  |
|---------|--|
| hvem    | <a href="#">dc.creator</a> , <a href="#">dc.contributor</a> , <a href="#">dc.publisher</a> |
| hva     | <a href="#">dc.title</a> , <a href="#">dc.description</a> , <a href="#">dc.subject</a>     |
| hvor    | <a href="#">dc.coverage</a>  |
| når     | <a href="#">dc.date</a>  |
| hvordan | <a href="#">dc.type</a> , <a href="#">dc.format</a>  |

Som minimumskrav for identifikasjon blir URI-en 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 webteknologi

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

Internettet gjør det mulig å publisere data og at man bidrar med informasjon og kunnskap i offentligheten gjennom deltagelse i ulike sjangre som blogg, chat og debattfora. Denne informasjonen er søkbar via datamaskiner, men informasjonen er vanligvis dårlig tilgjengelig (eller utilgjengelig) for automatisk prosessering. En slik prosessering kan bidra til bedre forståelse av store datamengder og hjelpe oss å finne mer relevant informasjon om et objekt enn vi ellers ville klart. Også formidlingsfunksjoner får et løft når mer

og bedre informasjon blir tilgjengelig for applikasjoner, og dermed brukere.

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 TIL VARIGHET

Til tross for omfattende investeringer finnes det ingen vanntette 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 erstatninger av skjøre originaler.

Beslutninger som tas nå legger føringer inn i fremtiden. Institusjoner som tok i bruk håndskannere på 1990-tallet for å skanne bilder i lav oppløsning, ser nå at det kreves bedre kvalitet. Den gangen var ikke det mulig, og beslutningene tatt da var fullstendig realistiske og reflekterte tidens kunnskaps- og teknologiske nivå. Valg av 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 abm-sektoren. En slik «varighetsstrategi» burde for hver institusjon som skal digitalisere minst inneholde fire elementer:

- **Dokumentasjon:** av strategien og planene. Hva skal gjøres og hva skal resultatene av arbeidet være? Er strategien forankret i organisasjonen? 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 budsjett. Med andre ord må det ikke kuttes underveis når et budsjett er godtatt. Slike endringer betyr ofte at man strammer inn på maskinvare, programvare 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 de skal drifte 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 kombinere flere baser til en virtuell base, og dermed blir det mer interessant for hver partner å konsentrere seg om «det man er best på». Informasjon fra andre fagområder kan i så fall hentes inn fra partnere som har annen spesialkompetanse.

## 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 prosent 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 brukes.
- **Hvordan langtidsoppbevaring og tilgjengelighet er ivarettatt.**
- **Kunnskap om eksterne krav.** Alle kulturarvinstitusjoner 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 med et kvalitetsanalyseverktøy som TRAC (Trustworthy Repositories Audit & Certification).

# 5. KRAV TIL PRESENTASJON

Å formidle informasjon og kunnskap gjennom websider eller informasjonstjenester er både et fagfelt 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 viktig er krav til 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.

1. januar 2009 trådte diskriminerings- og tilgjengelighetsloven i Norge i kraft. Loven gir et rettslig vern mot diskriminering på grunn av nedsatt funksjons- evne, 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 2012<sup>8</sup>.

---

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

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)»<sup>9</sup>.

Her i landet er det Standard Norge som jobber med standardisering. En god oversikt over standardiseringsrammeverk for universell utforming nasjonalt og internasjonalt finnes på organisasjonens nettsted [standard.no](http://standard.no).

Universell utforming og tilgjengelighet er også utførlig behandlet i ABM-skrift # 50 "På nett?" og netthåndboken "*Universell utforming innen abm-sektoren*", også gitt ut av ABM-utvikling.

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

---

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

# 6. KONKLUSJON

Som beskrevet i denne rapporten, har europeiske land alene og gjennom EU-prosjekter satset på langtidslagring og nettbasert tilgjengelighet av digital kulturarv og vitenskapelig informasjon. For perioden 2007-2013 er 400 millioner euro (3,2 milliarder kroner) bevilget til felleseuropeiske prosjekter innenfor digitalisering. Forbedring og tilrettelegging av teknologi står sentralt.

En felles og samkjørt satsing på digitalisering av kulturarv og vitenskapelig kunnskap har mange fordeler. Ikke bare er det viktig å tilrettelegge for enklere kunnskapsoppbygging i befolkningen for å forbedre konkurransedyktigheten og stimulere kunnskapsintensiv produksjon og tjenester. Gjennom en bedre forståelse av vår historie, kultur, identitet og hvordan den henger sammen med siste århundres globale utvikling, kan vi forhåpentligvis klare å skape et samfunn hvor folk har en god kulturell og historisk forståelse av seg selv og hverandre.

Dersom abm-sektoren blir tilpasset til den nye digitale hverdagen, blir det enklere for enkeltmennesket å få tilgang til kunnskap og informasjon. Samtidig kan den enkelte lettere bidra med sin egen kunnskap om kulturarven. I den nye nettkulturen deler mennesker mye informasjon og erfaring med hverandre gjennom (delvis anonymiserte) portaler. Klarer man å fange opp disse erfaringene og knytte dem til relevant kildemate-

riale av høy kvalitet (som abm-sektoren forvalter), har man lagt et godt grunnlag for verdiskaping og øking av konkurranseevnen. Digitalisert kulturarv må derfor gjøres tilgjengelig på en slik måte at hver organisasjon kan konsentrere seg om sine fagområder, og at kunnskap og kildemateriale som ikke tilhører dette området (automatisk) kan hentes inn fra de aktører som jobber med det. Dette innebærer at det må stilles tydelige krav til samhandling mellom løsninger innenfor abm-sektoren.

Men hvordan kommer man dit? Det er tydelig at et slikt perspektiv på langtidslagring og tilgjengelighet av digital kulturarv må støttes av standarder, åpenhet og teknikk. Samtidig er det viktig at man ikke «låser» seg til altfor 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 sikrer at den innsatsen man gjør nå holder mål om 10-20 år også.

Internasjonalt er det utviklet metodikk og retningslinjer for å gi institusjoner nettopp den støtte og trygghet de trenger når det skal settes i gang større digitaliseringsprosjekter. Det er viktig at en slik metodikk blir utviklet i samarbeid mellom institusjoner og politiske myndigheter, slik at det bygges opp en felles forståelse. Samtidig kan man knytte økonomisk støtte til bruk av metodikken og kravene den innehol-



der. Ikke minst gir en slik fremgangsmåte offentlige virksomheter et profesjonelt grunnlag for forhandlinger med kommersielle aktører. Poenget er at man, gjennom å bruke metodikken, kan være sikre på at den jobben som blir utført er av en viss kvalitet og tar høyde for noen viktige utfordringer som vil dukke opp i fremtiden.

De viktigste kravene man må bli enige om gjelder standarder og teknologi for å formidle kunnskap gjennom distribuerte systemer. Et viktig konsept som må på plass er at alle digitale objekter utstyres med en eller flere unike referanser i den digitale verden (Unified Resource Identifier, URI). Hvis dette er på plass, blir det mulig å ytre seg om slike objekter og legge til informasjon som er tilgjengelig for alle. Disse uttrykkene må være maskinlesbare, og derfor trengs det enighet om eller standarder for kunnskapsrepresentasjon.

Mye av infrastrukturen som behøves er allerede tilgjengelig i dag i form av standardene som ligger til grunn for internettet. Standardiseringsorganisasjonen W3C har de siste ti år kommet med flere anbefalinger til standarder på feltet kunnskapsrepresentasjon. Disse standardene (bl.a. RDF/RDFS og SPARQL) tillater at man representerer kunnskap digitalt og distribuert, samt at man kan gjøre spørringer mot andre kilder om informasjon man ønsker. En viktig utgangspunkt

for standardene er at det må være mulig å kombinere kilder og informasjon etter behov og uten tunge konverteringsprosesser.

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 ingen av dagens portalløsninger har implementert kravet om unik identifikasjon kombinert med bruk av åpne W3C-standarder. Hadde det vært gjort, kunne man for eksempel kople tjenestene Digitalt Museum og Digitalt Fortalt. Dette ville gjøre det mulig for en person som har spesiell kunnskap om et objekt i et museum som Maihaugen faktisk kan lage en liten film eller lydspor om det og publisere den gjennom Digitalt Fortalt. Hvis Digitalt Museum bruker URI-er for sine objekter, så kan innlegget på Digitalt Fortalt knyttes direkte til den, selv om det digitale objektet og fortellingen ligger i to forskjellige baser. For å følge tankeeksperimentet videre: si at Europeana begynner å referere til objekter i Digitalt Museum. Da ville disse objektene i Digitalt Fortalt automatisk blitt gjort tilgjengelig for et mye større publikum.

På denne måten møter man kravet om samordning, utveksling, innhenting og felles formidling av kulturarven.

For å komme dit fra dagens løsninger, må portalene som formidler den digitale kulturarven konsekvent

publisere data som objekter med en unik referanse på nett. I tillegg må disse knyttes til publiserte beskrivelser (metadata) som bruker aksepterte og felles retningslinjer (som f.eks W3C tilbyr). Fordelene med å gjennomføre slike endringer er at man enkelt kan hente inn informasjon fra andre kilder om sine egne ressurser og at man kan dele sine datamodeller med andre (semi-automatisk) uten høye kostnader. Legger man til kravet om åpen tilgang, så blir det mulig for alle å innhente, analysere, knytte sammen og gjenbruke informasjon i nye, felles tjenester som er til nytte for alle og som har potensial til å inkludere flere i den spennende og viktige historien som kulturarv representerer.

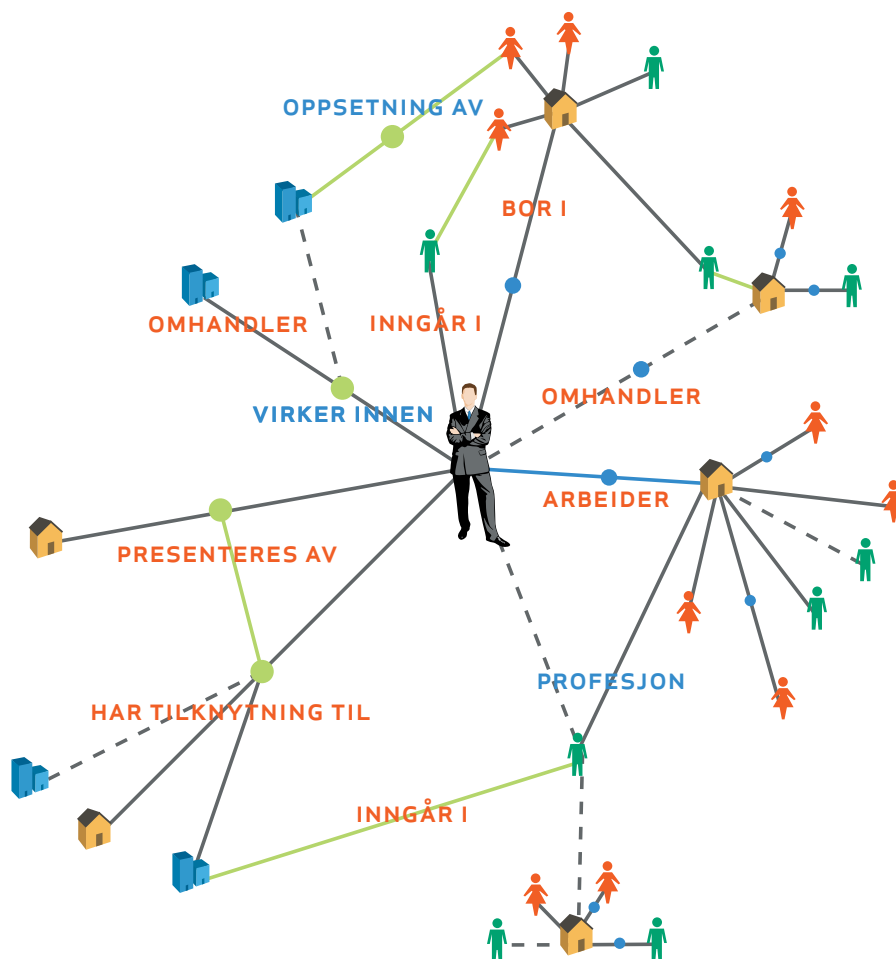


# 7. ORDFORKLARINGER

Mange av definisjonene her er tatt fra no.wikipedia.org og andre portaler. Noen er tilpasset og forkortet. Begrepene er også forklart i teksten.

|                  |  |
|------------------|--|
| Back-end system  | En server blir ofte referert til som et «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 rask 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 spesifikasjon av en konseptualisering” (Gruber '93), og er en definisjon av et begrepssystem formulert i en 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 standardiseringsorganisasjonen W3C.  |
| RDBMS            | Relational Database Management Systems, på norsk relasjonsdatabase. En database som bygger på den såkalte relasjonsmodellen. Relasjonsdatabaser består av tabeller som er forbundet med henvisninger, eller nøkler, som utgjør relasjoner.   |
| RDF / RDFS       | Resource Description Framework. En W3C-standard som er opprinnelig ment som en datamodell for metadata. Den har blitt en standard metode for konseptuell beskrivelse av webressurser hvor det brukes forskjellige syntakser.   |
| Semantisk Web    | Grunntanken bak ideen om den semantiske web er at datamaskiner skal kunne analysere og tolke dataene som finnes på allverdens websider. Dette innebærer en videreutvikling av den opprinnelige world wide web. Semantisk webteknologi gjør det mulig å definere selve betydningen av informasjonsenhetene på en webside. Dette gjør det enklere for datamaskiner å prosessere informasjonen - den blir maskinlesbar. |
| 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 | Search/Retrieval via URL. Protokoll som gjør at man kan søke i forskjellige baser gjennom et standard grensesnitt basert på HTTP-protokollen og spørresyntaksen CQL (Common Query Language).   |
| URI | En URI (Uniform Resource Identifier) er en kompakt streng av tegn som identifiserer en abstrakt eller fysisk ressurs. En URI har som funksjon å identifisere en digital, internettbasert ressurs på en global, unik måte. En URI kan ses på som en nøkkel for en digital, virtuell representasjon av et objekt. URI-en representerer dermed selve objektet på internett. |



# 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.

Gruber, T.R. "A translation approach to portable ontologies". *Knowledge Acquisition*, 5(2):199-220, 1993.

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

Moortgat, J. "Taking pictures to the public. Evaluatieverslag Nationaal Archief & Spaarnestad Photo op Flickr The Commons". 2009. [http://www.nationaalarchief.nl/images/3\\_16370.pdf](http://www.nationaalarchief.nl/images/3_16370.pdf)

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.









ISBN 978-82-8105-089-1