Kulturdepartementets strategi for åpne data

Innledning

Kulturdepartementet legger med dette frem en strategi for å gjøre offentlige data tilgjengelig for bruk. Strategien er en oppfølging av Meld. St. 27 (2015-2016) "Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet". I meldingen gjør regjeringen det klart at den ønsker å styrke arbeidet med å gjøre offentlige data tilgjengelig for viderebruk og gjenbruk. Om kultursektoren heter det:

"Data fra denne sektoren er verdifulle for kulturbaserte næringer. Denne typen data vil være lette å formidle og ha bred appell til publikum. Data herfra vil være lett å integrere med andre typer offentlige og private data. Samtidig må man ta hensyn til utfordringer knyttet til tredjeparts opphavsrettigheter. Åpne data om ordninger på kulturområdet vil kunne bidra til en mer transparent kulturforvaltning."

Det vises også til regjeringens digitaliseringsrundskriv med pålegg og anbefalinger om digitalisering av offentlig sektor.[[1]](#footnote-1)

Formålet med offentleglova er "å legge til rette for at offentlig verksemd er open og gjennomsiktig, for slik å styrkje informasjons- og ytringsfridommen, den demokratiske deltakinga, rettstryggleiken for den enkelte, tilliten til det offentlege og kontrollen frå ålmenta. Lova skal òg leggje til rette for vidarebruk av offentleg informasjon."

I tillegg til å bidra til åpenhet og innsyn, vil åpne data også kunne bidra til bedre tjenester, rikere opplevelser og økt effektivisering og verdiskaping. Forutsetningen er at dataene finnes, er lett tilgjengelige både for mennesker og maskiner, og at de enkelt kan viderebrukes og sammenstilles med andre data i nye kontekster. Det viktigste kulturinstitusjonen kan bidra med i dette bildet, er å tilby mer data og bedre data, dvs. mer konsistente og standardiserte data, og å gjøre dataene åpent tilgjengelig.

På kulturfeltet setter opphavsretten grenser for økt tilgjengeliggjøring av data med nyere kulturinnhold, som digitale fotografier, digitale bøker mv. Personvernhensyn setter også grenser for økt tilgjengeliggjøring, og er særlig aktuelt på arkivfeltet. Alle kulturdata som kan viderebrukes bør imidlertid være lisensiert/merket for videre bruk.

Dagens situasjon

Mange kulturinstitusjoner har data som kan tilbys åpent og det er en rekke eksempler på arbeid med å åpne datasett i sektoren allerede.

Store mengder digitaliserte gjenstander, fotografier, bøker og dokumenter tilbys i dag i tjenester som Digitalarkivet, nb.no og DigitaltMuseum. Tjenestene og innholdet tilfredsstiller i begrenset grad definisjonen[[2]](#footnote-2) til åpne data per i dag, men representerer et potensial i form av data, infrastruktur og brukere. Disse dataene kalles i denne sammenheng for samlingsdata. Dataene kultursektoren produserer gjennom forvaltningen av ulike roller, oppgaver og funksjoner, som for eksempel tilskuddsforvaltning, statistikk og rapportering, kalles i denne sammenheng for forvaltningsdata.

I vedlegget til denne strategien beskrives status for både samlings- og forvaltningsdata, status for tilgjengeliggjøring og bruk, utviklingstrekk, planer, eksisterende strategier og utfordringer og muligheter for enkelte utvalgte kulturinstitusjoner. Kort oppsummert viser situasjonen et begrenset, men målrettet arbeid i sektoren, både etats- og sektorvis, og i samarbeid på tvers av strukturelle skiller. Et flertall av etatene har bygget kompetanse og nettverk, og viktig infrastruktur er på plass eller under utvikling. Det er etablert konkrete tjenester, og flere datasett er tilgjengeliggjort som åpne data. Selv om omfanget så langt er moderat, viser aktiviteten at det finnes miljøer og erfaringer det kan bygges videre på.

Det er likevel ingen tvil om at arbeidet er i en tidlig fase. Kunnskapen om åpne data er begrenset, og brorparten av kulturdataene er ikke å betrakte som åpne. Videre er relativt få nye tjenester utviklet av tredjepart på de datasettene som er åpent tilgjengelige.

For å oppnå en effektiv og hensiktsmessig ressursbruk i arbeidet med å gjøre kulturdata mer tilgjengelige, må aktørene koordinere sin innsats på viktige områder, og ta ansvar for fellestjenester og funksjoner utover egen etat eller institusjon.

Det legges opp til følgende visjon og mål:

Visjon: Kultursektoren skal ha en kultur for åpenhet og transparens der data som hovedregel aktivt gjøres åpent tilgjengelig

Mål 1: Prioriterte datasett i sektoren identifiseres, dokumenteres og gjøres tilgjengelige som åpne data

Prioriterte samlings- og forvaltningsdata skal gjøres åpent tilgjengelige og merkes for videre bruk.

Datasett må kartlegges og prioriteres av hver enkelt etat og institusjon. Kartleggings- og prioriteringsprosesser på sektor- og institusjonsnivå bør være samordnet for å forhindre dobbeltarbeid. Sentrale aktører[[3]](#footnote-3) vil få et særskilt ansvar for denne samordningen.

Bruk og etterspørsel, både internt og eksternt, er viktige kriterier for prioritering. Brukspotensial og informasjonsverdi er eksempler på andre kriterier som bør tillegges vekt.

Prioriterte datasett skal være dokumentert og tilgjengeliggjort på data.norge.no og eventuelt andre felles kanaler for tilgjengeliggjøring. Virksomhetene skal gjøre egnet informasjon tilgjengelig i maskinlesbare formater.

Konkrete mål og planer for digitalisering skal formuleres der slike ikke finnes. En styrking av eksisterende produksjonsmiljøer bør også vurderes, slik at de kan tilby effektive og spesialiserte fellestjenester.

Mål 2: Standarder og autoritetsregistre er identifisert og implementert

De sentrale aktørene skal i fellesskap definere hvilke standarder som skal brukes i dialog med Difi, og hvordan disse skal kommunisere med hverandre. Det bør legges til rette for datautveksling på tvers av landegrensene ved å følge internasjonale standarder og være kompatibel med internasjonal praksis.

Identiteter knyttet til personer og steder må være unike og hentes fra felles registre. Andre felles innholdsressurser må også være unike og hentes fra felles kontrollerte ordlister.

Hvilke registre og ordlister sektoren har behov for, og hvilke som skal brukes, må kartlegges, prioriteres og defineres. Ansvaret for å etablere, tilgjengeliggjøre (tjenester/infrastruktur) og vedlikeholde disse ressursene plasseres hos de sentrale aktørene.

Anbefalinger som følger opp tverrfaglige og/eller fagspesifikke standarder og beste praksis bør formuleres i konkrete veiledninger og anbefalinger fra de sentrale aktørene.

Mål 3: Felles ansvars- og arbeidsdeling knyttet til utvikling og drift av felles dataressurser og tjenester er etablert

Det etableres et samarbeidsforum for åpne kulturdata i 2017. Forumet skal definere felles behov og foreslå løsninger. Forumet skal ha som mål å minimere dobbeltarbeid og skal jobbe for å få til en hensiktsmessig arbeidsdeling knyttet til utvikling og drift av felles dataressurser.

Forumets form avklares i dialog med Nasjonalbiblioteket, Arkivverket og Norsk kulturråd.

Kulturdepartementet utarbeider mandat for arbeidet og innarbeider strategien i styringsdialogen med etatene og andre aktuelle virksomheter.

Mål 4: En kultur for åpenhet, deling og samarbeid er etablert

Arbeidet med åpne data skal være basert på et prinsipp om effektiv oppgaveløsning, åpen tilgjengeliggjøring og en tydelig ansvars- og arbeidsdeling.

Behov for indikatorer og statistikk utredes nærmere i dialog mellom sentrale aktører og brukermiljøer.

Felles møteplasser, aktiviteter mv. for å utvikle kompetanse, holdninger og tjenester etableres.

Brukerundersøkelser benyttes for å ha oppdatert kunnskap om brukerne og deres behov og ønsker.

Tiltak som innebærer samhandling med brukerne, der brukerne aktivt oppmuntres til bruk, iverksettes.

Det skal søkes oppnådd gode rammeavtaler (avtalelisenser) med rettighetsforvaltere. Det forutsettes at rettighetsbelagte kulturdata i større grad enn i dag enkelt kan brukes og deles i stort omfang, og om mulig også viderebrukes, mot en effektiv klarerings- og vederlagsordning.

Vedlegg til Kulturdepartementets strategi for åpne data

Avgrensinger

Definisjon

På Norges offisielle nettsted for åpne offentlige data, data.norge.no, defineres åpne data på denne måten: «Åpne data er digitale data som er gjort tilgjengelige med de tekniske og juridiske egenskaper som er nødvendige for at de skal kunne brukes fritt, og bli distribuert av hvem som helst, når som helst og hvor som helst».[[4]](#footnote-4)

Åpne data skal altså kunne brukes fritt av hvem som helst, også til kommersielle formål, uten forespørsel, klarering eller vederlag. Data som bare tilgjengeliggjøres på forespørsel omfattes ikke av definisjonen. Data som gjøres tilgjengelig som digitale tjenester med søkbart og nedlastingsbart innhold er heller ikke uten videre å anse som åpne data, selv ikke når det er lagt til rette for fri viderebruk. Slike data og tjenester omtales likevel også i strategien, fordi de kan ha et betydelig brukspotensial som åpne data. I et slikt videre perspektiv har også konvertering av analoge data og dataenes kvalitet betydning for den fremtidige tilgangen til dataene. Betaling kan ikke kreves for tilgang til og bruk av det som er definert som åpne data, men eventuelt for spesialtilpassede tjenester, som ulike former for brukerstøtte eller sammenstilling av data.

Kulturdata – aktører og institusjoner

I Meld. St. 27 (2015-2016) "Digital agenda for Norge – IKT for en enklere hverdag og økt produktivitet er kultursektorens data beskrevet på denne måten:

«Data fra denne sektoren er verdifulle for kulturbaserte næringer. Denne typen data vil være lette å formidle og ha bred appell til publikum. Data herfra vil være lett å integrere med andre typer offentlige og private data. Samtidig må man ta hensyn til utfordringer knyttet til tredjeparts opphavsrettigheter. Åpne data om ordninger på kulturområdet vil kunne bidra til en mer transparent kulturforvaltning.»

Under Kulturdepartementet er det særlig Nasjonalbiblioteket, Arkivverket og Norsk kulturråd som har store mengder kulturdata. Strategien har derfor lagt hovedvekt på disse tre etatenes virksomhet.

Samtidig er det viktig å påpeke at kulturdata og relaterte data finnes i en rekke andre statlige etater og offentlig finansierte institusjoner, både under Kulturdepartementet og andre departementer og forvaltningsnivåer. Det gjelder for eksempel Norsk lyd- og blindeskriftbibliotek, Den norske kirke, Språkrådet, Norsk lokalhistorisk institutt og NRK, Den norske opera & ballett og andre teatre og konsertarenaer. Som eksempler på kulturdata på andre forvaltningsområder og -nivåer kan nevnes:

|  |  |  |
| --- | --- | --- |
| Type data | Institusjon | Ansvarlig departement |
| Kulturminnedata | Riksantikvaren | Klima- og miljødepartementet/KLD |
| Museumsdata | Universitetsmuseene | Kunnskapsdepartementet/KD |
| Museumsdata | Etatsmuseer (flere) | Samferdselsdepartementet m.fl |
| Arkivkataloger | Fylkesarkivene | Fylkeskommuner |
| Arkivkataloger | Kommunearkiv/byarkiv | Kommuner |
| Språksamlinger | Universitetet i Bergen | Kunnskapsdepartementet/KD |
| Ordbøker | Universitetet i Bergen | Kunnskapsdepartementet/KD |

Andre offentlige virksomheter kan også ha stor betydning for en brukervennlig utnyttelse av kulturdata gjennom sammenstilling av data. Et eksempel er Kartverket med sitt rike tilfang av kart- og eiendomsdata. Men også Folkeregisteret, Brønnøysundregistrene, kjøretøyregisteret, skipsregisteret og andre registerdata vil kunne sammenstilles med kulturdata.[[5]](#footnote-5) Det er derfor av betydning at en strategi for åpne kulturdata ikke opererer langs skillelinjer som forhindrer en formålstjenlig sammenstilling av data på tvers av både fag-, sektor- og departementsskiller. Dette reflekteres også i digitaliseringsrundskrivet (rundskriv H-7/14) som påpeker at virksomheter med grenseoverskridende ansvarsområder bør samarbeide og gi brukerne et hensiktsmessig og helhetlig tilbud, uavhengig av måten staten er organisert på.

Samlingsdata vs. forvaltningsdata

Strategien skiller mellom *samlingsdata* (digitalt tilgjengeliggjort samlingsinnhold), og *forvaltningsdata* (data generert som ledd av etatenes virksomhet). Det er stor forskjell i graden av og forutsetninger for åpenhet på slike data, både grunnet rettigheter, material- og filtyper.

*Samlingsdata* består som regel av et digitalt objekt, gjerne i form av et bilde av en gjenstand, et foto eller et dokument (innholdsdata), samt beskrivende og tekniske metadata.

*Forvaltningsdata* er data som etatene og kulturinstitusjonene produserer som en del av sin virksomhet med forvaltning av tilskuddsordninger, kulturarvsobjekter, statistikk o.a. Slike data kan inneholde mye relevant informasjon om norsk kunst- og kulturliv. Også forvaltningsdata kan bestå av innholdsdata og tilhørende metadata. Forvaltningsdata fra statsforvaltningen blir til samlingsdata når de etter hvert avleveres til Arkivverket. Det er viktig å påpeke at skillet mellom samlings- og forvaltningsdata først og fremst har betydning for konteksten de produseres, og til dels tilgjengeliggjøres i. Som åpne data er skillet av mindre betydning.

Arkivverket, Nasjonalbiblioteket og museene er de store forvalterne av samlingsdata, i tillegg til NRK. Både Nasjonalbiblioteket, Arkivverket, Norsk filminstitutt og Norsk kulturråd har dessuten til dels store mengder forvaltningsdata (bl.a. søknadsforvaltning og statistikk) som kan generere kunnskap om kunst- og kultursektoren og bidra til økt transparens. Alle etatene kan i tillegg ha viktige roller i å legge til rette for tilgjengeliggjøring av åpne data gjennom standardisering og infrastruktur.

Beskrivelse av nåsituasjonen

I denne drøftingen og presentasjonen av tilgjengelige data har vi valgt å inkludere datasett som finnes tilgjengelige via søkbare tjenester. Slike data bidrar til å illustrere potensialet i hva som kan gjøres tilgjengelig som åpne data, og er av den grunn inkludert her.

Innholds-/samlingsdata

ABM-sektoren (arkiv, bibliotek og museum) har arbeidet med digitalisering og tilgjengeliggjøring av innhold fra samlingene siden begynnelsen av 1990-tallet. De siste årene har det vært økende oppmerksomhet i sektoren rundt fri viderebruk av samlingsdata som nedlastbare filer, og etter hvert også som maskinlesbare datasett. Utgangspunktet er derfor at det finnes relativt mye innhold, det finnes infrastruktur, og det finnes et utgangspunkt for et felles ramme-/avtaleverk. Samtidig er det viktig å påpeke at det fremdeles finnes mye innhold som ikke er digitalisert, mye data som ikke er åpne, og mye arbeid som gjenstår for at verktøy og standarder kan brukes mer effektivt på tvers av etats- og departementsskiller til åpen tilgjengeliggjøring av kultursektorens data.

Opphavsrettigheter – metadata og «selve innholdet»

For enkelte materialtyper med stort brukspotensial, bl.a. foto, kan både opphavsrettslig vernetid og rettighetshavere være vanskelig å fastslå. Den lange vernetiden for åndsverk (70 år etter opphavers død) vil dessuten kunne skape et fravær av nyere innhold i åpne tjenester og data.

Klarerings- og vederlagsordninger kan gjøre dette landskapet lettere å manøvrere i, og samtidig også gi inntekter til rettighetshavere til materiale som i dag ellers ikke blir brukt. Avtaler og vederlagsordninger etter modell av Nasjonalbibliotekets «bokhylla-avtale» kan muliggjøre en begrenset tilgang og bruk. Det finnes derimot ikke eksempler på avtaler som muliggjør noen grad av fri gjenbruk.

Løsningen per i dag er å skille innholdsdata (digitale objekter) og metadata. Både Arkivverket, Nasjonalbiblioteket og museene publiserer sine metadata mer eller mindre åpent, mens det legges en lisens på det digitale objektet, f.eks. et fotografi (bildefil), en skannet bokside eller et dokument. Det er derimot ingen tradisjon for å markere metadataene med åpne lisenser. I Norge kan det virke overflødig, men i internasjonale sammenhenger kan det være både fordelaktig og til dels nødvendig.

Personvern

Personvernhensyn kan begrense tilgjengeliggjøringen av kulturdata og adgangen til å publisere personbilder og film for viderebruk. For Arkivverket vil personvernet legge til dels store begrensninger for hvilke datasett som kan publiseres som åpne data.

Status for tilgjengeliggjøring

Nasjonalbiblioteket

Nasjonalbibliotekets nettbibliotek (tidligere NBdigital) er inngangen til Nasjonalbibliotekets digitale samlingsdata. Her finner brukeren visningen av digitale objekter i alle medietyper. De tilhørende metadataene er tilgjengelige via Nasjonalbibliotekets søketjeneste eller søk i Oria[[6]](#footnote-6). Språkbanken har en egen ressurskatalog med åpne datasett der metadata om språkressurser og selve datasettene er tilgjengelige.

Samlingsdata som ikke er opphavsrettslig beskyttet, kan fritt lastes ned av brukerne. Dette gjelder aviser, bøker, tidsskrift, foto, musikkmanuskripter og privatarkivmateriale. Per i dag er omlag 450 000 norske bøker digitalisert. 30 000 av dem er opphavsrettslig falt i det fri og er nedlastbare.

Gjennom en avtalelisens med Kopinor (Bokhylla-avtalen) kan Nasjonalbiblioteket gjøre tilgjengelig  alle bøker utgitt i Norge før år 2001, selv om de er beskyttet av opphavsrett. Ved inngangen til 2017 vil Bokhylla.no inneholde ca. 250 000 rettighetsbelagte norske bøker med indeksert og søkbar tekst. Dataene er imidlertid kun tilgjengelige for brukere med norsk IP-adresse, og kan dermed ikke uten videre klassifiseres som åpne data. Bokhylla-avtalen åpner heller ikke for fri gjenbruk av rettighetsbelagt materiale.

Nasjonalbiblioteket tilbyr API-er eller andre former for tilgjengeliggjøring for alle datasett som har falt i det fri. Dataene er derimot ikke synliggjort på data.norge.no. Det er nødvendig å få på plass lisensiering for å legge bedre til rette for bruk av datasettene.

Arkivverket

Folketellingen for 1910 er åpent tilgjengelig med et API[[7]](#footnote-7) dokumentert på data.norge.no. I tillegg er emigrantprotokollene i Digitalarkivet og kjøretøyregistret Autosys 1980-90 i ferd med å åpnes som «linked open data» (LOD). Dataene og nytt API vil bli dokumentert på data.norge.no. Videre er katalogdata fra Asta (data publisert i Arkivportalen) tilgjengelige via Norvegianas[[8]](#footnote-8) API.

*Digitalarkivet* er Arkivverkets kildenettsted – en tjeneste primært for tradisjonell tilgjengeliggjøring av innhold. De mest populære kildene er kirkebøker, folketellinger og panteregistre. Kildene representeres som transkriberte lister eller skannede sider/dokumenter som bildefiler/pdf-filer, og er med få unntak ikke å anse som åpne data. Det er store mengder data i Digitalarkivet, deriblant over 30 mill. personposter, nærmere 1.3 mill. bosteds- og eiendomsposter og skannede versjoner av ca. 50 000 kilder, til sammen ca. 26 millioner bilder av dokumenter/sider. Dataene er tilgjengelige for publikum gjennom enkle eller avanserte person- eller eiendomssøk, og egne navigerings- og blatjenester for ulike kildekategorier. PDF-filer kan hentes fram og lastes ned fra Digitalarkivets nettsider.

Dette illustrerer noe av potensialet ved åpning av data som allerede er tilgjengeliggjort. Brukerne av Digitalarkivet er i hovedsak privatpersoner som slektsgranskere, lokalhistorikere, grunneiere og det allmenne publikum. Det planlegges å åpne Digitalarkivet for bidrag fra hele arkivsektoren, og det vil da øke både i innhold og bruk.

Kulturrådet/forvaltningsdata

Kulturrådet behandler hvert år ca. 20 000 søknader til over 60 søkbare tilskuddsordninger, og fordeler til sammen ca. 1,2 mrd. kroner til kunst- og kulturfeltet i Norge. Disse søknadsdataene med tilhørende oppfølging og rapportering er ikke tilgengelige p.t., men Kulturrådet har en målsetning om å publisere forvaltningsdata som åpne data innenfor rammen av eksisterende lover og forskrifter innen 2018.

Kulturrådet er i gang med et prosjekt for statistikk og analyse (SANK prosjektet) som har som målsetting, blant annet, å tilgjengeliggjøre forvaltningsdata fra Norsk kulturråd som åpne data (også på datanorge.no). Digitaliseringen av søknads- og saksbehandlingssystemet i Norsk kulturråd representerer et stort potensial for datafangst, strukturering, presentasjon og analyse av forvaltningsdata. Etableringen av eRapport innebærer strukturering av rapporteringsdata fra tilskuddsmottakere. SANK prosjektet har en målsetting om å øke informasjonsutveksling, samhandling og kunnskapsproduksjon, til gode for hele samfunnet. Kulturrådet mottar årlig opp mot 20 000 søknader fra kunst- og kulturvirksomheter, artister og kunstnere. I tillegg til å gjøre forvaltningsdata tilgjengelige, ligger det i dette prosjektet også et potensial for å kombinere data fra Norsk kulturråd med data fra andre kilder (blant annet offentlige datakilder, kulturinstitusjoner og kulturaktører) for videre statistikkutarbeiding og analyser som vil ha verdi for forskning, politikkutforming og verdiskapning. Åpne data er nøkkelordet her.

Fra og med 2017 legges søker- og tildelingslister ut på Kulturrådets nettsider.

Kulturrådet produserer og publiserer statistikk fra museene (og for arkivene fram til 2013/2014). [[9]](#footnote-9)

Kulturrådet/samlingsdata

Kulturrådet forvalter ikke egne samlingsdata, men har siden 1990-tallet (som Norsk museumsutvikling og senere ABM-utvikling) hatt et utviklings- og tilretteleggingsansvar for museene. Dette omfatter også museenes arbeid med digitalisering og publisering, bl.a. gjennom støtte til utviklingen av digitale verktøy og infrastruktur, og gjennom informasjon og holdningsskapende arbeid knyttet til åpenhet og deling. Det er brukt betydelige ressurser i museene til digitalisering og tilgjengeliggjøring de siste årene, og mye innhold er publisert på nett. Likevel gjenstår mye arbeid på veien til representative og åpne museumsdata.

Digitalisert og publisert i museene per materialtype (alle tall i mill. per 1.1.2015)[[10]](#footnote-10):

|  |  |  |  |
| --- | --- | --- | --- |
| Materialtyper | Antall | Digitalisert | Publisert med bilde |
| Kunsthistoriske gjenstander | 0,8 | 0,2 | 0,1 |
| Kulturhistoriske gjenstander | 3,8 | 1,3 | 0,7 |
| Arkeologiske gjenstander | 6,7 | 4,9 | 4,7 |
| Naturhistoriske gjenstander | 10,6 | 2,1 | 2,0 |
| Fotografier | 33,9 | 3,1 | 1,6 |
| I alt | **55,8** | **11,6** | **9,1** |

Samlet sett er ca. 16 % av museenes samlinger tilfredsstillende publisert på nett, jf. kriteriene i museumsstatistikken. Dette innebærer ikke publisering med åpne lisenser.

De viktigste digitale tjenestene som publiserer og tilgjengeliggjør museumsdata er DigitaltMuseum.no og universitetsmuseenes samlingsportaler UNIMUS.no.

DigitaltMuseum er publikums inngang til digitaliserte museumssamlinger fra museene som sorterer under Kulturdepartementet. DigitaltMuseum inneholder både kunst- og kulturhistoriske data. Noen nøkkeltall for DigitaltMuseum per oktober 2016:

* 1,2 mill foto
* 0,6 mill gjenstander
* 35 000 kunstverk
* 174 museer/visningssteder

Kun ca. 2 pst. av materialet på DigitaltMuseum, eller ca. 340 000 poster, er merket med åpne lisenser. Selv om hensynet til opphavsrettigheter - og/eller personvern begrenser tilgangen til deler av museumsinnholdet, vil en langt større andel av museenes samlinger kunne merkes med åpne lisenser. En enkel stikkprøve viser f.eks. at for ca. 2.600 kunstverk som har falt i det fri fra et tilfeldig utvalg kunstnere som døde før 1945, er kun 35 objekter merket med åpne lisenser, dvs. godt under 2 prosent.

Alle metadata i museumskatalogene er fritt tilgjengelig.

UNIMUS[[11]](#footnote-11) omfatter følgende tjenester:

* [Fotoportalen](http://www.unimus.no/foto)
* [Arkeologisøk](http://www.unimus.no/arkeologi/forskning/index.php)
* [Etnografisøk, Kulturhistorisk museum, Universitetet i Oslo](http://www.unimus.no/etnografi/khm/samling)
* [Kultursøk, Tromsø museum](http://www.unimus.no/etnografi/tmu/samling)
* [Etnografi-/kultursøk](http://www.unimus.no/etnografi/forskning)

Noen nøkkeltall for UNIMUS (universitetsmuseene) per oktober 2016:

* 1 mill. foto av arkeologiske gjenstander
* 0,7 mill. fotografier
* 250 000 gjenstander
* 174 museer / visningssteder

Alle objektene i UNIMUS-systemene er publisert med Creative Commons (CC) -lisenser[[12]](#footnote-12). Mange objekter, bl.a. alle arkeologiske gjenstander, er publisert med åpne lisenser som muliggjør viderebruk.

Data fra DigitaltMuseum publiseres via DigitaltMuseums eget API, og via Norvegiana. Data fra UNIMUS er tilgjengelige som nedlastbare filer (hele databasen) og via eget søke-API. I tillegg er hele arkeologisamlingen tilgjengelig via Norvegiana.

Norvegiana

Norvegiana er Kulturrådets verktøy for å gjøre kunst- og kulturdata lettere tilgjengelig som åpne data. Norvegiana er en database med tilhørende tjenester, med kulturdata fra arkiv, museer og andre kultur-institusjoner; fra i alt ca. 300 virksomheter eller avdelinger[[13]](#footnote-13). Norvegiana inneholder 7,4 mill. poster totalt (per august 2016). Av dette er 1,9 mill. bilder, 16 000 lydklipp og 1 400 videoer.

Norvegiana er et strategisk virkemiddel for å oppnå flere mål:

1. Arbeide for mer og bedre åpne data direkte fra eierinstitusjonene
2. Tilby en samlet tilgang til åpne data på ett sted; “one stop shopping”
3. Tilby en forenklet tilgang til mer komplekse originaldata via en felles data- og begrepsmodell, et forenklet datainnhold og et felles API
4. Tilby tjenester om API-basert tilgang for de institusjonene som ikke selv har slike tjenester fra sine systemer
5. Levere norske kunst- og kulturdata til Europeana[[14]](#footnote-14)

Norvegiana drives av Norsk kulturråd i samarbeid med flere nasjonale og regionale kulturinstitusjoner (Arkivverket, Riksantikvaren, Fylkesarkivet i Sogn og Fjordane) og kulturarvssystemer (bl.a. DigitaltMuseum /KulturIT, UNIMUS/MUSIT og Arkivportalen/Stiftelsen Asta).

Data i Norvegiana er også tilgjengelige i Europeana. Datasettene er dokumentert på egen nettside [data.norvegiana.no](https://norvegianablog.wordpress.com/), og på [data.norge.no](http://data.norge.no/).

Medietilsynet

Medietilsynet forvalter ulike støtteordninger og har oversikt over medie- og avishus i Norge, i tillegg til økonomien og eierskapsforholdene i norske medium. Medietilsynet har blant annet et film- og videogramregister som inneholder 195 500 titler. Registeret er ikke tilgjengeliggjort i sin helhet. Medietilsynet har ingen åpne datasett i dag. Så langt er ikke åpne data omtalt i strategi- eller plandokumenter.

Rapportene som publiseres jevnlig, som årsrapport, allmennkringkastingsrapporten, medieøkonomirapporten, barn og medier-undersøkelsen og rapporteringer til EU publiseres digitalt. Rapportene tilpasses for lesing både på internett, nettbrett og mobil.

Tilgjengelige data [[15]](#footnote-15)(ikke åpne):

* informasjon om kinofilm (i hovedsak aldersgrenser)
* tilskuddsordninger (mediestøtte og støtte til lokalkringkasting)
* radio- og tv konsesjoner
* medieeierskap
* aldersgrense på spill (PEGIs spillbase)
* veiledere for barn og voksne om digitale medier
* veiledningsfilmer på youtube

Norsk Filminstitutt

Norsk Filminstitutt har ingen åpne datasett p.t, men det arbeides med et API for Filmdatabasen. Filmdatabasen består av filmer som Norsk Filminstitutt har eller har hatt i sin festivaldistribusjon. Databasen er et oppslagsverk for norske spillefilmer, kortfilmer og dokumentarfilmer som er produsert siden 2002. Den er ikke en komplett oversikt over norsk film, da den først og fremst inneholder filmer som Filminstituttet har jobbet med. Hvor egnet databasen er når det gjelder deling er noe usikkert, da den først og fremst er et arbeidsverktøy for lansering av film i utlandet. Totalt inneholder databasen ca. 1500 filmer, og antallet øker hele tiden. Filminstituttet er i ferd med å oppgradere sine saksbehandlingssystemer, og i den forbindelse er det ønskelig å se på hvordan data kan utveksles mellom ulike systemer. Data det her er snakk om er først og fremst knyttet til forvaltningsoppgaver mot fond – dvs. tilskuddordninger og tildelinger i disse.

Retningslinjer for tilgjengeliggjøring

KMD har utarbeidet retningslinjer[[16]](#footnote-16) for tilgjengeliggjøring av offentlige data. Disse retningslinjene er et godt utgangspunkt for å vurdere tilgjengeligheten til åpne data. Utvalgte datasett innenfor kultursektoren er i tabellen nedenfor vurdert opp mot Difis 15 kriterier. Datasettene har fått en karakter for hvert av kriteriene. Karakterene 0 og 1 er markert med rødt for å synliggjøre nødvendige forbedringer. Vi har benyttet følgende skala i vurderingen: 0: Følger ikke kriteriet, 1: Følger kriteriet i svært liten grad, 2: Følger kriteriet i noen grad, 3: Følger kriteriet i svært stor grad.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Kriterier | N-grammer [[17]](#footnote-17) NbDigital | Nb.no søkeAPI | NB Boklistene | Folketellingen for 1910 (Digitalarkivet), søkeAPI | Katalogdata fra ASTA | Digitalt­Museum | Norvegiana |
| 1a. Åpne standardlisenser – som rammeverk | 3 | 0 | 0 | 3 | 3 | 3 | 3 |
| 1. Åpne standardlisenser – i faktisk bruk | 3 | 0 | 0 | 3 | 3 | 1 | 1 |
| 2. Maskinlesbare formater | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 3. Gratis data | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4. Program­merings­grensesnitt | 1 | 3 | 1 | 3 | 3 | 3 | 3 |
| 5. Komplett nedlasting | 3 | 0 | 1 | 1 | 1 | 3 | 3 |
| 6. Faste adresser og unike ID | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 7. Dokumentasjon | 2 | 1 | 1 | 2 | 1 | 2 | 2 |
| 8. Informasjon | 2 | 0 | 1 | 1 | 1 | 1 | 2 |
| 9. Synlig | 1 | 1 | 1 | 3 | 2 | 1 | 2 |
| 10. Brukerbehov | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 11. Tilbake­meldinger | 1 | 1 | 1 | 2 | 2 | 1 | 1 |
| 12. Uten registrering | 2 | 2 | 2 | 3 | 3 | 1 | 3 |
| 13. Oppdatert | 1 | 1 | 1 | 2 | 1 | 3 | 3 |
| 14. Oppmuntre | 2 | 1 | 1 | 1 | 1 | 0 | 0 |
| 15. Oversikt | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| Totalsum | 33 | 21 | 21 | 34 | 29 | 27 | 31 |

På kriteriene *gratis data* og *maskinlesbare formater* skårer de utvalgte datasettene høyt, mens på kriteriene *åpne standardlisenser- som rammeverk* og *åpne standardlisenser i faktisk bruk* er det mer ujevnt mellom datasettene. For de øvrige kriteriene gis det jevnt over lavere skår. Selv om tabellen viser svakheter og forbedringspotensial, viser samtidig aktiviteten at det finnes erfaringer, kompetanse og infrastruktur det kan bygges videre på.

Brukerstatistikk

Tabellen under viser noen nøkkeltall for 2015 for de mest brukte tjenestene fra etatene (basert på tall fra Google Analytics).

|  |  |  |  |
| --- | --- | --- | --- |
| Tjeneste | Antall besøk (brukersesjoner) | Antall unike brukere | Antall sidevisninger |
| DigitaltMuseum | 1,3 mill. | 0,8 mill.  | 15 mill. |
| UNIMUS (universitetsmuseene) | 0,06 mill. | 0,03 mill. |  0,35 mill.  |
| Digitalarkivet | 4,8 mill. | 0,8 mill.  | 190 mill. |
| NBdigital | 3,5 mill. | 1,6 mill. | 15,8 mill. |

Status for bruk av åpne data

Generelle betraktninger

Vi har lite kunnskap om bruken av åpne kulturdata. Det finnes foreløpig ingen statistikk som viser bruken av APIene, og det er ikke gjort brukerundersøkelser som kan si noe om hvem som bruker hvilke data til hva. Ingen av etatene har implementert løsninger for logging av bruken av de åpne APIene. Basert på den begrensede informasjonen man har er det grunn til å anta at det er relativt lite bruk av åpne kulturdata, spesielt i form av nye tjenester fra tredjepart.

Erfaringen viser også at det er nødvendig med tiltak for å øke kjennskapen til og bruken av åpne kulturdata.

Direkte tilgjengeliggjøring vs. aggregatorer/data.norge.no

Kulturdata er tilgjengelige gjennom API-er, web-tjenester, filer for nedlasting, bruk av åpne formater, etc. Institusjonene har i varierende grad valgt å bruke fellestjenester som data.norge.no og Norvegiana for dokumentasjon og tilgjengeliggjøring. Det er også noe varierende hvordan datasett og tjenester er dokumentert på institusjonenes egne nettsider.

Tabellen under illustrerer dette.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Datasett | Institusjon | Dokumentasjon på institusjonens egen nettside | Dokumentasjon på data.norge.no | Tilgjengelig i Norvegiana |
| Arkivportalen | Stiftelsen ASTA | Nei | Nei | Ja |
| Folketellingen 1910 | Arkivverket | Nei | Ja | Ja |
| NBdigital | Nasjonalbiblioteket | Ja | Nei | Nei |
| Bokhylla | Nasjonalbiblioteket  | Ja | Nei | Nei |
| Digitalt Museum | KulturIT | Nei | Ja | Ja |

Bruk av åpne samlingsdata

Digitalarkivet (Arkivverket) og Bokhylla (Nasjonalbiblioteket) er sektorens mest brukte webtjenester i dag. Populære data i Digitalarkivet er folketellinger, kirkebøker, adressebøker og andre «slekts­granskning­relaterte» data. Av disse er det kun folketellingen fra 1910 som er tilgjengelig som åpne data på data.norge.no. Det finnes ingen statistikk for bruken av denne.

Bokhylla.no vil innen 2017 inneholde omlag 250 000 bøker som er utgitt i Norge til og med år 2000. Teksten er indeksert og søkbar. Dataene er derimot ikke maskinlesbare, og er kun tilgjengelige for norske IP-adresser. Bokhylla er dermed ikke å anse som åpne data. Begrensningene for det rettighetsbeskyttede materialet dikteres av avtalen Nasjonalbiblioteket har gjort med rettighetshaverne.

Bruk av åpne forvaltningsdata

Museumsstatistikken er et eksempel på forvaltningsdata som brukes i økende grad. I tillegg til at statistikken brukes av forvaltningen selv, publiserer Statistisk sentralbyrå museumsstatistikk og bruker tallene videre i sin egen presentasjon av kulturstatistikken. Det private forskningsinstituttet Telemarksforsking er også i gang med å lage en database over museumstatistikken, basert på rådata levert fra Norsk kulturråd. Rådgivningsselskapet Rambøll utarbeider en oversikt over visuell kunst i tall og bruker tall fra museumsstatistikken som en del av bakgrunnsmaterialet.

Nasjonalbiblioteket har ansvar for å samle og presentere den norske bibliotekstatistikken. Her finnes et åpent datasett som viser årlig bestand, tilvekst og bruk av alle fag- og forskningsbibliotek, folke- og fylkesbibliotek i Norge.

Arkivverket, Nasjonalbiblioteket og Norsk kulturråd samler og publiserer statistikkene på henholdsvis arkiv-, bibliotek- og museumsfeltet. De tre statistikkene overlapper til en viss grad på hvilke institusjoner som leverer tall. De har også berøringspunkter der typene materiale i samlingene overlapper. Det er etablert en møteplass for koordinering av spørreskjema, utveksling av erfaringer og diskusjon av arbeidet. Statistiske data fra de tre sektorene bør kunne publiseres som åpne data, og gjerne også med lenking av institusjoner, geografi, material- og samlingstyper.

Hvilke data er etterspurt men ikke tilgjengelige i dag?

Data fra Norsk kulturråds søknadsforvaltning er etterspurt både fra media, forvaltning og forskning (evaluerings- og utredningsprosjekter). Det finnes ingen statistikk over denne etterspørselen, men den oppfattes som relativt stabil de siste årene. Det vanligste er forespørsler om en konkret region eller kategori, og da om hvem som har søkt, hva de har søkt om (både prosjekt og sum) og om de har fått tilskudd eller ikke. Slike data må i dag i stor grad hentes ut og tilgjengeliggjøres manuelt på forespørsel.

Det er generelt stor etterspørsel etter diverse samlingsinnhold i Arkivverket, Nasjonalbiblioteket og museene som ikke er digitalisert eller digitalt tilgjengeliggjort. Eksempler på populære materialtyper er aviser, tidsskrifter, kart, tegninger og foto. Noe av dette materialet, primært aviser og tidsskrifter, er ikke tilgjengeliggjort grunnet komplekse og uavklarte rettighets- og avtaleforhold.

All tekst som digitaliseres i Nasjonalbiblioteket OCR-behandles[[18]](#footnote-18). Denne tekstgjenkjenningen gjør det mulig å søke på ord og fraser i en scannet bok, en avis eller et tidsskrift. OCR-teksten er dessuten etterspurt av brukere med synshemming eller lesevansker, fordi den kan brukes på leselist og til å generere syntetisk tale. Slike løsninger støter imidlertid på en del utfordringer. Kvaliteten på OCR-teksten bestemmer leseopplevelsen, og materiale som er dekket av avtaler med rettighetshavere er ikke uten videre nedlastbart.

Har tilgjengeliggjøringen gitt nye tjenester?

Tilgjengeliggjøringen av åpne kulturdata har så langt gitt flere nye tjenester. Tjenestene er utviklet gjennom intern bruk av APIer (Digitalarkivet på mobil og Historisk befolkningsregister i Arkivverket), gjennom samarbeidsprosjekter som Kultur- og naturreise eller på direkte initiativ fra eksterne interessenter. De viktigste kjente brukerne, dvs. løsningene/tjenestene som bruker åpne kulturdata frermgår av tabellen nedenfor:

|  |  |  |  |
| --- | --- | --- | --- |
| Tjeneste / løsning | Tema | Utvikler | Data som inngår |
| [**Kultursøk**](http://www.kultursok.no/) | Innsyn; søke i flere samlinger | Webløft; Buskerud fylkesbibliotek | Norvegiana, NBdigital, folkebibliotekkataloger m.m. |
| [**Kultursøk**](http://mr.kulturnett.no/delving/search) | Innsyn; søke i flere samlinger | Regionale kulturnett | Norvegiana |
| [**Norgeskart**](https://play.google.com/store/apps/details?id=no.avinet.norgeskart&hl=no) | Mobilapplikasjon; innsyn; søke i flere samlinger | Avinet | Norvegiana, FT 1910 |
| [**En blå tråd**](http://xn--enbltrd-hxac.no/kart/) | Innsyn; søke i flere samlinger | Trondheim kommune, byplankontoret | Norvegiana |
| [**Kartdemonstratorer**](http://knreise.no/demonstratorer/) | Mobil og nett. Kartvisning av innhold fra flere kilder | Kultur- og naturreise-prosjektet | Norvegiana, og flere andre kilder |
| [**KNappen**](http://knreise.org/index.php/mobil-formidling/knappen/) | Mobilapplikasjon | Kultur- og naturreise-prosjektet | Norvegiana, og flere andre kilder |
| [**Rallarvegen**](https://play.google.com/store/apps/details?id=no.kamikazemedia.rallarvegen&hl=no) | Mobilapplikasjon; opplevelser langs Rallervegen | Hordaland fylkeskommune | Norvegiana |
| [**Prøysenvandring**](http://proysenvandring.no/#first-page) | Mobilapplikasjon; opplevelser i “Prøysenland” | Hedmark fylkesbibliotek | Norvegiana |
| [**Ætteforsker**](https://itunes.apple.com/no/app/tteforsker-norway-around-1900/id630809223?mt=8) | Mobilapplikasjon  | Privat utvikler | Folketelling 1910 og Norvegiana |
| [**Stedr**](http://stedr.blogspot.no/) | Mobilapplikasjon  | Prosjektet TAG CLOUD/Sintef | Norvegiana |
| [**Historiske vandreruter**](http://knreise.no/demonstratorer/demonstratorer/flyktningeruta.html) | Nett og mobil | Riksantikvaren | Norvegiana |
| [**Oslo i gamle dager**](https://itunes.apple.com/no/app/oslo-i-gamle-dager/id962249492?mt=8) | Mobilapplikasjon  | Privat utvikler | DigitaltMuseum |
| [**Nasjonalmuseets samlingssøk**](http://samling.nasjonalmuseet.no/no/) | Søk i museets samlinger | Nasjonalmuseet | DigitaltMuseum |
| [**Digitalarkivet på mobil**](http://arkivverket.no/arkivverket/Digitalarkivet/Om-Digitalarkivet/Organisasjon-og-tenester/Tenester/App-for-Digitalarkivet) | Mobilapplikasjon  | Riksarkivet  | Digitalarkivet |
| [**Søkemotor for bibliotek**](http://www.bibliotekarensbestevenn.no/nbsok/) | Lenkegenerator | [Bibliotekarens Beste Venn](http://www.bibvenn.no/) | NBdigital |

Detaljert oversikt og mer informasjon om de enkelte løsningene som bruker data fra Norvegiana er publisert på <https://norvegianablog.wordpress.com/losninger-som-bruker-norvegiana-data/>.

Den generelle erfaringen er at tilgjengeliggjøring og bruk påvirker datakvaliteten positivt gjennom brukerkommentarer på feil og nødvendige forbedringer. Gitt at disse blir fulgt opp vil datakvaliteten bli bedre. Nasjonalbiblioteket har som mål å ta i bruk egne APIer internt, bl.a. for å selv kunne finne og rette feil i egne data og dermed heve datakvaliteten raskere.

Planer for nye tilgjengeliggjørings- og formidlingsløsninger for åpne data

Arkivverket

I 2017 vil Arkivverket utbedre, dokumentere og åpne sitt eksisterende søke-API mot de transkriberte person- og eiendomsdataene (innholdsdata) og kildedataene (metadata) fra alle kildekategoriene i hele Digitalarkivet. Noe senere vil indeksdata og andre metadata til de skannede bildene (arkivdokumentene) i Digitalarkivet få et åpent API. Stiftelsen Asta vil antakelig også åpne sitt søkeAPI mot de landsomfattende arkivkatalogdataene (metadata) i Arkivportalen.

Kulturrådet /museene

Norsk kulturråd vil tilby museumsstatistikken som åpne data via et API. Både intern og ekstern tilgang til dataene skal kanaliseres gjennom API-et og leveres som JSON[[19]](#footnote-19) (alternativ til XML). Videre behandling, f.eks. som lenkede data, er ikke planlagt, men vurderes videre når API-et er på plass. For øvrige forvaltningsdata er det ikke konkrete planer, utover en generell målsetning om å tilby søknadsdata som åpne data i den grad det er mulig innen 2018.

Museenes digitale infrastruktur er i kontinuerlig utvikling. I 2017 er følgende aktiviteter sentrale:

KulturITs økosystem «eKultur» videreutvikles, bl.a. for å forbedre tilgangen til data fra Primus og DigitaltMuseum via felles API-er

MUSIT (universitetsmuseene) kjører et omfattende prosjekt med restrukturering av grunnleggende IT-arkitektur, som bl.a. skal ivareta behovene for effektiv tilgjengeliggjøring av samlingene som åpne data

KulturNav.org, som er KulturIT sitt verktøy for å skape, forvalte og dele felles åpen terminologi og autoritsregistre) skal videreutvikles for å bedre tilgangen til autoritetsregistre fra eksterne brukere og datasett, bl.a. gjennom et API

Nasjonalbiblioteket

Det følger av Nasjonal bibliotekstrategi 2015-2018 at Nasjonalbiblioteket skal sørge for at folkebibliotekene får tilgang til frie metadata fra en sentral kilde. Nasjonalbiblioteket har derfor inngått en avtale med Bokbasen om å produsere metadata for norske utgivelser fortløpende. Disse vil bli gjort åpent tilgjengelig i flere formater.

Nasjonalbiblioteket lanserer i løpet av 2017 nye nettsider for tilgjengeliggjøring av samlingene. Det utvikles også nye API som brukes som utgangspunkt for formidlingstjenesten som skal finnes på nye nb.no.

Nasjonalbiblioteket skal i samarbeid med Biblioteksystemer lage en ny løsning som erstatter dagens Biblioteksøk (<http://www.nb.no/bibsok/start.jsf>). Tjenesten er en nasjonal søketjeneste som viser oversikt over hvilket bibliotek som har tilgang til den aktuelle publikasjonen.

Utviklingstrekk

De siste fem årene har åpne og maskinlesbare data gradvis fått mer oppmerksomhet i kultur- og kanskje spesielt ABM-sektoren. Det er likevel fremdeles slik at begreper som åpenhet, tilgjengeliggjøring og viderebruk i de fleste sammenhenger knyttes til tradisjonell tilgjengeliggjøring av samlingsrelatert innhold i tjenester som NBdigital, Digitalarkivet og DigitaltMuseum. Samtidig har utviklingen i den tradisjonelle tilgjengeliggjøringen representert et viktig grunnlag for tilnærmingen til åpne data. Holdninger til deling, åpenhet og viderebruk er endret gjennom de siste 10–15 årene. Behovet for felles autoritetsregistre og vokabularer har vært kjent lenge, men er ytterligere konkretisert gjennom samarbeid og infrastruktur (bl.a. Kultur- og naturreiser, K-lab og KulturNav). Kunnskap om teknologi og lovverk har også utviklet seg positivt gjennom det mer tradisjonelle digitaliserings- og tilgjengeliggjøringsarbeidet.

Produkter og brukere

Den tydeligste «trenden» de siste fem årene er kanskje nettopp åpne og maskinlesbare data, som har gitt et annet perspektiv både på hvilke kulturdata som er relevante i et brukerperspektiv, hvilke målgrupper vi henvender oss til og hvilke sluttprodukter som skal produseres. Sluttproduktet er ikke lenger «bare» en søkbar representasjon av et objekt eller et dokument med beskrivende metadata, men også en åpen tilgang til våre interne forvaltningsdata *som data* – maskinlesbare, sammenstillbare og gjenbrukbare med andre data i nye kontekster.

Sluttbrukeren kan, som før, være en journalist, forsker eller allment interessert kulturkonsument, men kan også være en kommersiell tjeneste- eller produktutvikler, en kollega i forvaltningen eller i egen virksomhet eller sektor. Bruken kan være av en type vi ikke kunne forutse for 5–10 år siden, og ennå ikke klarer å beskrive dekkende og presist. Det er i dag en økende forventning til at offentlige data er åpent tilgjengelig. Likevel er både tilbud og etterspørsel foreløpig på et svært moderat nivå i kultursektoren. Det er rimelig å anta at bildet er et annet om ytterligere fem år.

Det er skapt økt bevissthet og aktivitet i sektoren, bl.a. gjennom samarbeid som Kultur- og naturreise / K-lab (Norsk kulturråd, Arkivverket, Riksantikvaren og Kartverket). Aktiviteter som *hackatons* (utviklerseminarer) har både bidratt til tilgjengeliggjøring av relevante data, lisensiering av data, økt oppmerksomhet i offentligheten og konkrete tjenester. Det er etablert god dialog og til dels samarbeid med relevante miljøer internasjonalt (bl.a. Digisam i Sverige og Europeana i EU). Økt tilgjengeliggjøring av data har også gitt mer tverrsektoriell bruk både blant offentlige og kommersielle aktører. Et eksempel er VG presentasjon av databasen Våre falne; <http://www.vg.no/spesial/2015/vaare_falne/>).

Sammenstilling av data fra ulike kilder har tydeliggjort behovet for felles vokabularer og autoritetsregistre. Selv om det fremdeles gjenstår mye arbeid er initiativer til både infrastruktur og konkrete verktøy igangsatt. Nasjonalbibliotekets arbeid med åpne autoritetsregistre for steds-, person- og korporasjonsnavn, samt KulturITs tjeneste KulturNav er positive eksempler i denne retning.

Praksis har beveget seg fra å tilgjengeliggjøre statiske datasett (CSV, XML) til utvikling av API-er mot løpende oppdaterte databaser. Det er eksperimentert med Linked Open Data (LOD), men det har ennå ikke gitt noe gjennombrudd når det gjelder tjenester og bruk.

Pågående arbeid i sektoren, nasjonalt og internasjonalt

Kultur- og naturreise / K-lab

*Kultur- og naturreise* var et tverretatlig samarbeidsprosjekt (fra 2012 til 2015) mellom Kartverket, Norsk kulturråd, Riksantikvaren og Arkivverket (Miljødirektoratet deltok 2012-2014). Prosjektet la til rette for at innhold i fagdatabaser hos partnere kunne åpnes opp og gjøres tilgjengelig for bruk og videreutvikling av andre til nye tjenester; av nærings- og reiseliv, i friluftsliv og undervisning. Målet var å øke kunnskap om og skape engasjement for kulturminner og naturverdier. Prosjektet er dokumentert på <http://knreise.org/>.

*K-lab* er et utviklingssamarbeid i forlengelsen av *Kultur- og naturreise* for å gjøre åpne kultur- og kulturminnedata og relaterte geodata mer tilgjengelige. K-lab skal legge til rette for økt bruk gjennom bedre tilgjengeliggjøring, sammenstilling og formidling. I K-lab deltar Kartverket, Riksantikvaren, Arkivverket og Norsk kulturråd. En viktig effekt av K-lab er økt bevissthet rundt relevante problemstillinger og nødvendige tiltak i de samarbeidende organisasjonene. Samarbeidet brukes også som plattform for internasjonale kontakter og samarbeid. Det er bl.a. forsøk på nordiske samarbeid knyttet til hackatons og Linked Open Data (#nordLOD), primært med Digisam i Sverige.

Prosjekter i Norsk kulturråds digitale utviklingsprogram

I perioden 2015-2017 er digital utvikling ett av fire programområder i Norsk kulturråds utviklingsprogram for museumssektoren. Blant prosjektene som har arbeidet med åpne data er *Designmodellar frå Figgjo i KulturNAV*, Jærmuseet, *Videreutvikle museenes digitale økosystem*, Lillehammer museum og Norsk folkemuseum (KulturIT). *Farleia – kystens livsnerve* (museumsnettverket for fiskerihistorie og kystkultur), *Samlinger på nett* (Nasjonalmuseet for kunst, arkitektur og design) og *Digital reiselivshistorisk formidling* (Stiftinga Musea i Sogn og Fjordane). Prosjektene arbeider på ulike måter med bruk av åpne data i samlingsforvaltning og formidling.

Felles vokabularer, autoritetsregistre og standarder

Et av behovene som er tydelig adressert i K-lab samarbeidet er autoriserte felles vokabularerog autoritetsregistre. Slike behov går på tvers av etater, sektorer og departement, og krever samarbeid og koordinering for å sikre god og effektiv forvaltning. Det pågår viktig arbeid i flere av etatene på dette området.

Norsk kulturråd har bidratt til etableringen av nettstedet KulturNav.org, som er KulturIT[[20]](#footnote-20) sitt verktøy for å skape, forvalte og dele felles åpen terminologi og autoritetsregistremed særlig fokus på behovene til museer og andre kulturarvsinstitusjoner. Listene publiseres som lenkede åpne data (LOD) i standardisert format. Eksempler på innhold og prosjekter i KulturNav per oktober 2016 er Maritim kulturarv, Poststeder i Norge, Fotografregisteret, Terminologi for forvaltning, drift og vedlikehold av kulturhistoriske bygninger, Terminologi for konservering, og Feltkatalogen 2015 (Feltkatalog for kunst- og kulturhistoriske museer). Andre viktige registre for kulturdata er Bibsys autoritetsregister over personer, Nasjonalbibliotekets autoritetsregister for navn (person- og korporasjon) og nasjonalt (bibliografisk) verksregister som skal være på plass innen 2018, og Kartverkets stedsnavnsregistre.

Nasjonalbiblioteket inngår i en rekke internasjonale samarbeid på biblioteksrelaterte områder som også kan ha interesse utover sektoren. CLARIN er et felles europeisk initiativ som samler språkressurser for å tilrettelegge for utvikling av språkteknologi. IIIF er et internasjonalt samarbeid for å utforme et standardformat for utveksling av bilder. IIPC er et internasjonalt samarbeid for utvikling av standarder for bevaring av internett. Nasjonalbiblioteket har videre startet et prosjekt som skal utrede muligheter for automatisert metadatafangst. Prosjektet startet i januar 2016 og skal ferdigstilles januar 2019.

Opphavsrettslige rammeavtaler

Kulturrådet ga i 2014 tilskudd til BONO (Billedkunst opphavsrett Norge) for å utrede mulighetene for et felles avtaleverk (avtalelisens) for kulturhistorisk fotografi i norske museer. Prosjektet følges nå opp av Norges museumsforbund i dialog med Kulturrådet. Dersom en slik avtale kommer på plass, vil det kunne være relevant for andre i kultursektoren som forvalter kulturhistorisk fotografi og eventuelt også billedkunst, å kunne koble seg på avtalen.

Internasjonale trender og prosesser

Arts Council England har fokus på å tilgjengeliggjøre sine data i tråd med *The Freedom of Information Act[[21]](#footnote-21)*. De har publisert «Arts Council England data principles», hvor viktige punkter går på transparens, at man skal publisere så mye som mulig så åpent som mulig, og at åpne og lenkede data (LOD) er et mål. Tilgjengeliggjorte data inkluderer tildelinger i form av excel-filer og museumsstatistikk og rapportering som CSV[[22]](#footnote-22). Det er derimot publisert med lite dokumentasjon og uten API-er, og er dermed å anse som en positiv tendens mer enn åpne data.

Tunge internasjonale museumsaktører forholder seg til CIDOC-CRM[[23]](#footnote-23) og tilbyr data som RDF[[24]](#footnote-24) via SPARQL endepunkter. Andre publiserer til dels åpne data, men med mindre spesifikke modeller/standarder (EDM, Dublin Core) og mindre dokumentasjon.

*Digisam* er et samordningssekretariat for digitalisering, digital bevaring og digital tilgjengeliggjøring av kulturarvsmateriale i Sverige, etablert av regjeringen/kulturdepartementet i 2011, og med ansvar lagt til Riksarkivet. En lang rekke institusjoner medvirker til Digisams arbeid, bl.a. Kungliga biblioteket (nasjonalbiblioteket) og alle de store museene. Videreføring av sekretariatet er nylig besluttet lagt til Riksantikvarieämbetet. Kunnskapsinnsamling og kompetanseoppbygging på feltet er hovedmålet. Digisam arbeider bl.a. for åpning av offentlige kulturdata ved å utgi anbefalinger, arrangere seminarer, hackatons m.m. En oversikt over deres skriftlige anbefalinger og rapporter finnes her: <http://www.digisam.se/leveranser/leveranser-alla/>.

Åpne datasett og API-er bekjentgjøres og dokumenteres på [*oppnadata.se*](http://oppnadata.se/), tilsvarende det norske *data.norge.no*. Det finnes også et alternativt nettsted, *opendata.se.*

I tråd med den svenske regjeringens målsetting om en «åpnere forvaltning som bidrar til innovasjon og vekst», har Riksarkivet fått i oppdrag å forenkle situasjonen for dem som vil viderebruke åpne data. Oppdraget omfatter bl.a. å gi hjelp til den enkelte borger med å finne ut hvor i statsforvaltningen åpne data finnes og bistå i kontakten med den aktuelle myndigheten.[[25]](#footnote-25)

*Europeana* er EUs store satsing for å etablere en felles europeisk tilgang til kunst- og kulturdata. Søkeportalen [Europeana.eu](http://www.europeana.eu/portal/) gir tilgang til over 50 mill. bøker, billedkunst, filmer, gjenstander og arkivmateriale fra ca. 40 land. Mer enn 2 500 europeiske institusjoner har bidratt til Europeana, alt fra store internasjonale navn som Rijksmuseum i Amsterdam, British Library og Louvre museet, til regionale arkiver og lokale museer. Mona Lisa av Leonardo da Vinci, arbeidene til Charles Darwin og Isaac Newton og musikken til Wolfgang Amadeus Mozart er noen av høydepunktene på Europeana.eu. De mange delprosjektene i Europeana har vært viktige for utvikling av beste praksis, retningslinjer og anbefalinger, samt europeiske fagnettverk på området. [[26]](#footnote-26)

Modenheten for økt tilgjengeliggjøring i sektoren

Det har gjennom mange år vært ulike initiativer med en felles tilnærming til digitalisering og tilgjengeliggjøring i ABM-sektoren på tvers av institusjons- og til dels sektorgrenser, både med tanke på felles prioriteringer, økt datakvalitet og forbedrede publikumstjenester. Internasjonale prosesser, standarder og samarbeid som Europeana og CIDOC (standardisering av utvekslingsformater og felles datamodell) har også vært viktige drivere i denne retning, gjennom et behov for bearbeiding av data, effektiv utveksling av data m.m.

Det finnes lite sammenliknbare tall for åpne kulturdata i Norge og kultursektoren i for eksempel Europa. Likevel indikerer de tall vi har at situasjonen på relaterte områder er vel så god i Norge og Norden enn i Europa for øvrig. Blant annet er Norge den tredje største bidragsyteren til Europeana (regnet i totalt antall etter Nederland og Tyskland) og den desidert største bidragsyteren regnet per capita. Det Europeana-eide prosjektet Enumerate leverte tall i sin rapport «Survey Report on Digitisation in European Cultural Heritage Institutions 2014» som viser at kulturarvssektoren også på andre relaterte områder står seg godt i en sammenlikning med Europa.

Sammenliknet med Sverige derimot kan organiseringen av arbeidet ha bidratt til visse modningsforskjeller. På flere områder knyttet til standardisering, koordinering og veiledning har Sverige kommet lenger enn Norge, antakelig mye grunnet Digisams aktiviteter.

Selv om norske forhold stort sett tåler en sammenlikning med europeiske, er det mye som gjenstår også i sektorens modenhet.

Nasjonalbiblioteket har konkrete strategiske målsetninger på området. De bruker den åpne lisensen CC0[[27]](#footnote-27) aktivt, bl.a. i nye nettsider, og vil fra 2017 ha bruk av åpne lisenser som internt styringsparameter. De har dessuten en målsetning om å bruke egne APIer internt for å avdekke og utbedre svakheter raskt.

Arkivverket og Norsk kulturråd følger ikke langt etter. Praktisk tilrettelegging av data i tjenester som Norvegiana, Digitalarkivet og Arkivportalen, samt aktiviteter som hackatons, har gitt verdifulle erfaringer i åpning og tilgjengeliggjøring av data. Fellesnevneren her er deltakelse i Kultur og natur-reise og K-lab som har gitt økt bevissthet i organisasjonene. Bevisstheten har derimot ennå ikke materialisert seg i konkrete målsetninger for åpning av egne data.

Når det gjelder museenes samlingsdata er bruken av åpne lisenser avgjørende for videre tilgjengeliggjøring og viderebruk. Her er status at under 2 pst. av materialet på DigitaltMuseum er merket med åpne lisenser, uten at det er åpenbare teknologiske eller juridiske grunner til det lave tallet.

Kjente utfordringer og muligheter

Flere utfordringer for videre tilgjengeliggjøring og bruk av data er også berørt tidligere i dokumentet. Blant de sentrale utfordringene er jus, kompetanse og teknologi.

Opphavsrettigheter

Klarering og bruk av rettighetsbelagt materiale kan være krevende. Det kan også være uklare rettighetsforhold knyttet til materialet som gjør prosessen med rettighetsklarering ekstra krevende.

For at nyere materiale skal kunne åpnes kreves et avtaleverk og forvaltningsordninger som gjør klarering og vederlagsbetaling enkelt. Et godt avtaleverk kan bidra positivt til at mere data blir gjort åpent tilgjengelig for viderebruk.

En avtale bør også dekke materiale av ukjente eller ikke-organiserte rettighetshavere, noe som forutsetter en avtalelisens. Slike avtaler kan være spesielt viktig for materialtyper som fotografi og billedkunst. Dette gjelder selv om avtalelisensene ikke tar opp i seg fri viderebruk.

Vi har viktige fortrinn i Norge og Norden, gjennom et godt organisert kunst- og kulturliv, og lange erfaringer med ulike former for kollektivt forvaltede rettigheter og vederlagsordninger.

Økt bruk av rettighetsbelagt materiale vil nødvendigvis gi økte kostnader til vederlag. Kostnadsnivået er derimot umulig å beregne så lenge verken volum, bruksformer eller prismodell er på plass.

Taushetbelagt materiale

Det kan i tillegg til de opphavsrettslige utfordringene også være personvernmessige utfordringer knyttet til tilgjengeliggjøring av åpne data. Sett med kulturrelatert data kan inneholde mange personopplysninger. Eksempel på dette er datasett fra Arkivverket.

Slike problemstillinger vil kreve aktsomhet og god dialog med Datatilsynet og andre relevante myndigheter. Vurderingen av om informasjon som inneholder personopplysninger eller som er underlagt taushetsplikt kan tilgjengeliggjøres, må gjøres på bakgrunn av bestemmelsene i personopplysningsloven og de lover og avtaler som regulerer taushetsplikten knyttet til de aktuelle opplysningene.

Kompetanse og ressurser

Selv om det allerede er gått noen år siden begrepet «åpne data» ble lansert i fagmiljøene, og også noen år siden 2011, da åpen tilgjengeliggjøring av data var overordnet føring i de statlige etatenes tilskuddsbrev, er åpne data som begrep og konsept i begrenset grad implementert i etater og virksomheter. En offensiv iverksetting av føringene i stortingsmelding 27 (2015-2016) *Digital agenda for Norge* vil være utfordrende for dagens administrative og IKT-faglige ressurser. En betydelig større mengde offentlige kulturdata åpent tilgjengelig vil kreve styrkede ressurser til bl.a. teknisk støtte og nødvendig oppdatering av datasett. Deler av dagens teknologiske infrastruktur er trolig ikke rustet for å gi tilfredsstillende responstid på de datamengdene det kan være snakk om.

Felles autoritetsregistre

Behovet for felles vokabularer og autoriserte registre og lister over personer, steder (nåværende og historiske), materialtyper, prosesser etc. er allerede omtalt. Det er en utfordring i dag å definere hvilke autoritetsregistre sektoren har behov for, hvilke som skal brukes, hvem som skal ha ansvar for å etablere og/eller ivareta disse registrene over tid, og hvordan de skal gjøres tilgjengelige. Slikt ansvar og slike funksjoner bør forankres i miljøer og infrastruktur som kan sikre nødvendig stabilitet og langsiktighet.

Arkivverket – spesifikke problemstillinger[[28]](#footnote-28)

Forvaltningsdata fra offentlige etater skal etter arkivloven avleveres til Arkivverket som arkivmateriale, og blir der å betrakte som samlingsdata. Selv om arkivmaterialet typisk vil være statisk og 5-25 år gammelt (historisk), mens etatenes API-er mot egne forvaltningsdata typisk vil operere mot dynamisk oppdaterte nåtidsdata, vil det trenges avtaler for å forhindre dobbelt åpning av de samme dataene over tid.

Økt tilgjengeliggjøring og bruk – fortrinn og styrker

Det knyttes store forventninger til samfunnsnyttig verdiskaping gjennom økt bruk av kulturdata. Som det sies i Stortingsmelding 27 (2015-2016) Digital agenda for Norge er kultursektorens data *«verdifulle for kulturbaserte næringer»,* og de vil være «*lette å formidle og ha bred appell til publikum».* Dette må anses som et godt utgangspunkt for økt tilgjengeliggjøring og bruk som åpne data. Likevel kan man ikke forvente at aktuelle brukere finner dataene og ser potensialet i dem uten aktiv formidling fra sektorens side.

Selv om kultursektoren skiller seg fra de andre sektorene i strategien ved å inneholde rettighetsbelagt materiale, er det likevel slik at mye av innholdet bør kunne gjenbrukes fritt uten klarering eller vederlag. Dels fordi det er eldre materiale hvor rettighetene har utløpt, eller fordi det er offentlig skapt/eid/finansiert og i stor grad ikke-kommersielt.

Når det gjelder forvaltningsdata bør hovedregelen være at alt kan deles fritt. Unntakene vil kreve rutiner for å fanges opp og administreres, men bør ikke være til hinder for at hovedtrekkene i offentlighetsloven praktiseres også for data.

1. <https://www.regjeringen.no/no/dokumenter/digitaliseringsrundskrivet/id2569983/> [↑](#footnote-ref-1)
2. ####  Åpne data er (pr. defifinisjon jf. data.norge.no) data som er gjort tilgjengelige slik at de kan leses og tolkes av både maskin og menneske. Dataene må også ha en åpen lisens slik at de enkelt kan viderebrukes av alle som vil.

 [↑](#footnote-ref-2)
3. Norsk kulturråd, Arkivverket og Nasjonalbiblioteket vil i fellesskap få en sentral rolle i oppfølgingen av strategien. [↑](#footnote-ref-3)
4. Jf. Difis forslag til definisjon per oktober 2016. [↑](#footnote-ref-4)
5. Omfattende sammenstilling av offentlige data kan være i konflikt med loven. Tjenesteutviklere henvises til Datatilsynet og andre myndigheter med tilsyns- og reguleringsansvar. [↑](#footnote-ref-5)
6. Søketjenesten Oriaer en felles portal til det samlede materialet som finnes ved de fleste norske fag- og forskningsbibliotek. [↑](#footnote-ref-6)
7. API er et applikasjonsprogrammeringsgrensesnitt (en kontaktflate) i en programvare som er åpen, slik at andre programmer kan kontakte og kjøre deler av programvaren i egne tjenester. [↑](#footnote-ref-7)
8. <https://norvegianablog.wordpress.com/> [↑](#footnote-ref-8)
9. Statistikken publiseres på <http://www.kulturradet.no/museum/statistikk>. Noen av årgangene (2012-2014) er tilgjengelige som åpne data, og dokumentasjon finnes på <http://osd.abmd.no/api.html> og <https://kulturognaturreise.wordpress.com/2014/01/07/hack4no-arkiv-og-museumsstatistikk/>. [↑](#footnote-ref-9)
10. Kilde: Museumsstatistikk for 2014 [↑](#footnote-ref-10)
11. Universitetsmuseenes samlingsportaler. **Unimus**-sidene utvikles og vedlikeholdes av Universitetsmuseenes IT-organisasjon, MUSIT, som er et samarbeid mellom universitetsmuseene i Norge. [↑](#footnote-ref-11)
12. Creative Commons (CC) er en internasjonal, [ideell organisasjon](https://no.wikipedia.org/wiki/Ideell_organisasjon) som utvikler og stiller til rådighet et sett med kvalitetssikrede [lisenser](https://no.wikipedia.org/wiki/Lisens) for [skapere av åndsverk](https://no.wikipedia.org/wiki/Opphavsmann), produsenter og utøvere. [↑](#footnote-ref-12)
13. Her telles museene avdelingsvis, ikke som konsoliderte enheter. [↑](#footnote-ref-13)
14. EUs satsing for digital tilgjengeliggjøring av kulturarv. [↑](#footnote-ref-14)
15. Informasjon om datasettene er publisert på tilsynets nettside [www.medietilsynet.no](http://www.medietilsynet.no). Data er publisert som PDF eller regneark (Excel). Filmdatabasen <http://www.medietilsynet.no/filmdatabasen/> er de dataene som blir brukt mest. [↑](#footnote-ref-15)
16. Offisiell versjon her:

<https://www.regjeringen.no/no/dokumenter/retningslinjer-ved-tilgjengeliggjoring-av-offentlige-data/id2536870/>. [↑](#footnote-ref-16)
17. <http://www.nb.no/sprakbanken/repositorium#ticketsfrom?lang=nb&collection=sbr>. [↑](#footnote-ref-17)
18. Med tradisjonell skanning får du et *bilde* av den skannede teksten – du kan ikke automatisk markere ord eller avsnitt i teksten. Derfor kan du heller ikke redigere den eller få den lest opp med f.eks CD-ORD. Med OCR-skanning (Optical Character Recognition) ‘låses teksten opp’ slik at den kan markeres.

 [↑](#footnote-ref-18)
19. JSON (JavaScriptObjectNotation) er en enkel tekstbasedrt standard for datautveksling. [↑](#footnote-ref-19)
20. **KulturIT** er et selskap som utvikler, drifter og vedlikeholder felles it-systemer for forvaltning, formidling, utvikling av samlinger ved museer og andre kulturinstitusjoner og for tilrettelegging av samlingene for forskning. [↑](#footnote-ref-20)
21. Om implementering av den engelske offentlighetsloven i forvaltningen i Storbritannia. [↑](#footnote-ref-21)
22. CSV: Comma-separated values, filformat brukt for å lagre tabellarisk data. [↑](#footnote-ref-22)
23. Ontologi for kulturarvs- og museumsdokumentasjon. [↑](#footnote-ref-23)
24. RDF: Resource Description Framework, en gruppe av spesifikasjoner som beskriver en modell for å representere metadata (data om data). [↑](#footnote-ref-24)
25. Pressemelding fra juli 2016: <http://www.regeringen.se/pressmeddelanden/2016/07/regeringen-oppnar-dorren-for-mer-oppen-data/> [↑](#footnote-ref-25)
26. Mer informasjon finnes på: Europeana Labs - <http://labs.europeana.eu/>, Fagblogg - <http://labs.europeana.eu/blog>, Europeana Pro - <http://pro.europeana.eu/>. [↑](#footnote-ref-26)
27. CC0 er et alternativ som betyr "no copyright reserved" i Creative Common systemet. [↑](#footnote-ref-27)
28. Problemstillingen adresserer også behovet for systemer som ivaretar dataproveniens, f.eks. gjennom automatisk tildeling av unike ID-er/URI-er ved avlevering av materiale til Arkivverket. En annen problemstilling er om åpne datasett publisert på data.norge.no også skal avleveres som arkivmateriale til Arkivverket fra KMD/Difi. [↑](#footnote-ref-28)