

Samlinger på nett – Collections online: status og veien videre

Hensikten med notatet

Hva er Samlinger på nett - Collections online?

Strategisk forankring

Grunnlaget

Status for publiserte samlinger og funksjonalitet

Status for bruk av infrastrukturen

Krav til data – datakvalitet

Teknologisk løsning

Ressurser og kompetanse

Veien videre

Vedlikehold og feilretting – restanseliste

Hensikten med notatet

Hensikten med dette notatet er å sammenfatte status for Samlinger på nett – Collections online tre år etter lanseringen av versjon 1.0. Videre ta opp viktige tema om veien videre, blant annet knyttet til fagmiljøenes ønsker og behov for videreutvikling, samt forespørsler fra andre institusjoner om å ta i bruk løsningen.

Hva er Samlinger på nett - Collections online?

Collections online ble etablert ved NTNU Vitenskapsmuseet i 2016-2017, primært som en forskningsinfrastruktur med internasjonal orientering. Ett av målene var å gjøre museet, dets samlinger og fagmiljøer mere synlige og attraktive for internasjonale forskningssamarbeid gjennom å gi en engelskspråklig tilgang til kulturhistoriske gjenstandsdata. En ville samtidig gi brukerne en ny mulighet til å interagere med samlingene blant annet gjennom verktøy for å sortere, lage egne utvalg av objekter («album»), dele og eksportere data om samlingene. Andre løsninger fram til da hadde i prinsippet bare vært «søk-og-vis»-løsninger, uten at en hadde mulighet for å bearbeide eller viderebruke og dele resultatene på en hensiktsmessig måte.

Løsningen er nyttig for forskere, studenter, samlingsforvaltning, offentlig forvaltning og den interesserte allmennhet, spesielt etter at løsningen også ble tilgjengelig på norsk.

«Collections online» ble utviklingsprosjektets tittel da det i første fase kun fantes som engelsk versjon. I dag omtales løsningen som «Samlinger på nett» på norsk og «Collections online» på engelsk.

Strategisk forankring

De vitenskapelige samlingene er dynamiske samlinger av høy kvalitet. De er bygget rundt den kultur- og naturhistoriske virksomheten ved museet og er vår viktigste forskningsinfrastruktur. Samlingene utvikles planmessig for å realisere nasjonale og internasjonale forskningsambisjoner, og er også en infrastruktur for å ivareta museets lovpålagte oppgaver etter kulturminneloven. Det er en overordnet målsetning at samlingene og tilhørende data skal være digitalt tilgjengelige for forskning, forvaltning og publikum.

NTNU Vitenskapsmuseets mål mot 2025:

- I 2025 har vi styrket og aktivert samlingene som viktig kunnskapsgrunnlag og forskningsinfrastruktur.
- Samlingenes anvendelighet økes gjennom faglig tilrettelegging, dokumentasjon og digital deling.
- Vi skal tilrettelegge for flere samlingsbaserte forskningsinitiativ og videreutvikle en offensiv politikk for tilgjengeliggjøring og deling av samlinger og data.

I NTNU Vitenskapsmuseets strategi har vi tatt mål av oss om å være «nasjonalt førende på digital deling». Dette for å underbygge NTNUs visjon om "Kunnskap for en bedre verden", der det slås fast at "Kunnskap utviklet ved NTNU skal komme fellesskapet til gode".

Samlinger på nett har vært finansiert av NTNUs strategiske midler, museets strategiske midler og interne personalressurser innenfor faglig IT, kommunikasjon og samlingsforvaltning. Satsingen hadde ikke vært mulig uten at NTNUs ledelse så på dette som strategisk riktig og relevant.

Grunnlaget

NTNU, Vitenskapsmuseet og Kunnskapsdepartementet har de siste 25 årene gjort omfattende investeringer i digitalisering av de vitenskapelige samlingene. Per 31.12.2019 er drøyt 70 % av gjenstandene i de kulturhistoriske samlingene digitalisert, mens 85 % av de naturhistoriske objektene er digitalisert¹.

De underliggende databaseløsningene for museets samlingsforvaltning er i hovedsak utviklet i samarbeid med universitetene i Bergen, Oslo og Tromsø. De driftes og forvaltes gjennom samarbeidstiltaket MUSIT - universitetsmuseenes IT-organisasjon. Datagrunnlaget foreligger på norsk.

Utgangspunktet for satsingen på Collections online var å dra nytte av det omfattende og gode arbeidet med digitalisering, kvalitetssikring og normering utført ved museet gjennom blant annet Museumsprosjektet (1998-2006) og Revita-prosjektet (2006-2015). *Formålet med Collections online var å tilrettelegge materiale for et internasjonalt publikum, samtidig som en skulle gi brukerne en større mulighet til å interagere med funnene og bruke løsningen som verktøy for å sortere, samle og eksportere data om samlingene.*

¹ Kunnskapsdepartementets definisjon av digitalisert: Registrering i database mht. faglig beskrevet gjenstandspresentasjon

Vitenskapsmuseet ønsket å tilgjengeliggjøre samlingene på et annet nivå enn den eksisterende visningsportalen «Unimus» gjennom å:

- Tilby informasjonen på en engelsk språklig plattform
- Gi mulighet for brukerne til å lage egne utvalg / samlinger / album av materialet
- Gi brukerne nye og forbedrede muligheter til å dele og eksportere materiale
- Unngå Flash (som ikke fungerte på Ipads etc.)

Status for publiserte samlinger og funksjonalitet

I første fase ble datasett fra kulturhistorie publisert, i hovedsak knyttet til de arkeologiske samlingene. Denne versjonen var kun på engelsk, da målgruppen for denne forskningsinfrastrukturen var internasjonale forskere. Ettersom de opprinnelige datasettene forelå på norsk, ble det gjort et vesentlig arbeid for å lage oversettelser av grunnleggende terminologi knyttet til gjenstandstyper, materiale, periode og funnkontekster.

Den neste fasen ble utvidet med norske versjoner (nynorsk og bokmål) og den numismatiske samlingen ble lagt til. Dette gjorde at brukerne for første gang kunne hente gjenstander fra ulike baser i samme søk og legge i samme album. I samme fase ble også kartet utviklet videre fra å være en visning til et søk. Brukerne fikk mulighet til å zoome inn i kartet og finne funn, i første versjon viste kartet kun funnene man allerede hadde lagt til. Den tekniske løsningen for å koble gjenstandspresentasjoner til litteratur og digital dokumentasjon ble utviklet. En gjennomførte videreutvikling basert på tilbakemeldinger fra brukerne, som f.eks. filtrering på materialtype og ned til gårds- og bruksnummer. Man la også til funksjonalitet som gir brukeren mulighet til å laste ned store mengder data, i kommaseparert format. Denne funksjonaliteten krever innlogging.

For å dra nytte av at herbarieark hadde blitt scannet i høy oppløsning gjennom et annet prosjekt, brukte man Samlinger på nett-infrastrukturen til å få tilgjengeliggjort bildene. Samtidig som man fikk tilgjengeliggjort de botaniske samlingene, fikk man vist at man også har en infrastruktur som kan tilgjengeliggjøre data på tvers av natur- og kulturhistoriske samlinger. Både den norske og engelske visningen for botanikk, inneholder per i dag kun vitenskapelige navn.

Gjennom disse tiltakene er status for publiserte samlinger i løsningen som følger:

- 298 891 arkeologiske objekter er publisert, hvorav 52 774 med bilder.
- 15 063 mynter er publisert, hvorav 10 247 med bilder.
- 516 859 botaniske objekter er publisert, hvorav 231 416 med bilder.

Bilder deles under *Creative Commons Attribution-ShareAlike 4.0 International* (CC BY-SA 4.0). Dette medfører at alle som vil kan gjenbruke publiserte bilder i løsningen så lenge opphavsperson og institusjon krediteres.

Status for bruk av infrastrukturen

Fra Google analytics har vi informasjon om følgende bruk av Collections online:

Antall sidevisninger

2017: 15 721	2018: 63 049	2019: 48 729	2020 (jan.-feb.): 12 141
--------------	--------------	--------------	--------------------------

I 2018 kom 68,3% av sidevisningene fra Norge. De fleste fra utlandet kom fra USA og Frankrike.

Samlinger på nett har etterhvert blitt en integrert del av den løpende forvaltningen av de arkeologiske samlingene. Viktige bruksområder i dag er:

- Publisering av ferdig katalogiserte funn fra samlingene. Publisering er en integrert del av innordningsprosessen, og i korrespondansen med private finnere henviser en til Samlinger på nett når funnet er ferdig innordnet. Private finnere kan da raskt gjenfinne og gjenbruke informasjon om gjenstandene de har levert og tilhørende opplysninger.
- Løsningen brukes som referansepunkt for forskere fra inn- og utland som har henvendelser om samlingene. Løsningen brukes blant annet for å lage lister/album for utenlandske forskere som er på utkikk etter bestemte gjenstandsgrupper, eller de kan også henvises til å bruke søkefunksjonene i løsningen.
- Løsningen har også blitt brukt i undervisning og presentasjoner, f.eks. når en skal undervise om bestemte gjenstandsgrupper. Den er svært nyttig i andre situasjoner når en skal arbeide med en utvalgt gruppe funn og sette disse sammen i andre presentasjonsløsninger.
- Vi ser også at løsningens kartfunksjoner er svært anvendelige når en trenger rask informasjon om utbredelsen av en bestemt funngruppe. Dette gir oss rask tilgang på informasjon som en tidligere måtte bruke betydelig innsats for å frembringe. Dette gjør det langt enklere å kontekstualisere funn og f.eks. fremhente tidligere observasjoner som er gjort i eldre kataloger på en effektiv måte.
- En litt mere utilsiktet bruk er at også metallsøkere har brukt løsningen for å lage album over «sine» funn. Dette var opprinnelig en funksjonalitet tenkt for forskere, men har da fått en videre utbredelse enn forutsett.
- Vi ser svært positive effekter ved at en integrerer publiseringsløsninger med de interne samlingsforvaltningssystem, spesielt fordi dette er en lett måte å identifisere prioriterte kvalitetshevingstiltak for datasettene det gjelder.

Krav til data – datakvalitet

Datakvalitet for materialet har svært stor innvirkning på presentasjonen av materialet i Samlinger på nett. Presentasjonen er bygget rundt foto og kartvisning, noe som krever at en har grunnlagsmateriale for dette på plass. Samtidig krever filtreringsfunksjoner for periode og materiale at en har oppdaterte og normerte datasett på disse områdene.

Samlingene omfatter til dels store og komplekse datasett, og kvalitetsheving er egentlig en kontinuerlig virksomhet. Noe av dette har en arbeidet med i tidligere prosjekter, bl.a. når det gjelder kobling av gjenstandsfoto, stedsopplysninger, kartfestinger av funnsteder mm.

Noen mere spesifikke tiltak som ble gjort opp mot lanseringen av Collections online var:

- Normering av termer for gjenstander, materiale og periode
- Oversettelse av systematiske (normerte) termlister for ovennevnte til engelsk

Normeringsprosjekter har bestått i å gjennomgå lister over alle unike termer som er i bruk, tilordne normerte termer til disse og deretter foreta en maskinell oppdatering på bakgrunn av disse tabellene i samarbeid med MUSIT. Dvs. at en normerer termer uten å gjøre individuelle nybestemmelser. Her vil en kunne rydde i 85 – 90 % av datasettene uten å vurdere objekter individuelt.

En vil også kunne gjøre vesentlige kvalitetsheving uten at en nødvendigvis velger høyeste presisjonsnivå. Vi har f.eks. ikke gjort detaljerte kronologiske analyser av alle udaterte funn (noe som ikke er gjennomførbart med begrensede ressurser), men nøyd oss med å angi den mest presise dateringen som er mulig å angi ut ifra eksisterende data.

Vi har arbeidet tilsvarende med oversettelser, dvs. at distinkte termer for gjenstand, materiale og periode har blitt systematisk oversatt. Dette vil også dekke store deler av de øvrige museenes materiale, gitt at disse ønsker å gjenbruke oversettelsene som er foretatt.

Fagmiljøet har brukt ca. 1 månedsverk på oversettelsesarbeid og ca. 3 månedsverk på normeringsprosjektene. Dette omfatter et materiale på ca. 550 000 poster. I tillegg til dette kommer arbeidsinnsats gjennom tidligere prosjekter for stedsnavn, foto og kartfesting, noe en ikke har summert omfanget av.

Teknologisk løsning

Collections Online er utviklet med åpen kildekode og åpen API. Det også lagt stor vekt på universell utforming, og så langt det er mulig oppfyller løsningen AAA-kravene til tilgjengelighet. Løsningen er bygd slik at den skal kunne skaleres både til å håndtere flere museer og samlinger. Infrastrukturen er utviklet og skalert slik at den kan gjenbrukes.

Collections Online er en Single Page Application (SPA) som finnes i både engelsk, bokmål og nynorsk utgave. SPA frontend er bygget på Googles Angular og backend API er bygget på Ruby Sinatra. For å gjøre brukeropplevelsen rask og sømløs, lastes hele applikasjonen ned ved første kontakt. Deretter kommuniserer den med backend bak kulissene. Backend-tjeneren håndterer kontakt med andre tjenerere; en MongoDB dokument-orientert database og en Elasticsearch søkemotor. MongoDB inneholder alle databasetabeller mens Elasticsearch er en indeksert søkemotor som gjør navigering, søk og levering av datasett hurtig og responsiv. Alle tjenere kjøres på én virtualisert Ubuntu server. Om ønskelig kan disse kjøres på flere adskilte servere.

Collections online oppdateres daglig gjennom nedlasting av de offentlig tilgjengelige MUSIT-datasettene innen arkeologi, numismatikk og botanikk. Ved import av data rapporteres det

om avvik, manglende oversettelser, o.l. Hver natt tas det backup av hele serveren og fil-baserte datadump fra både MongoDB og ElasticSearch. Drift, feilretting og videreutvikling utføres av faglig IT ved NTNU Vitenskapsmuseet.

Ressurser og kompetanse

Samlinger på nett har vært finansiert av NTNUs strategiske midler, museets strategiske midler samt bruk av interne personalressurser innenfor faglig IT, kommunikasjon og samlingsforvaltning. De største direkte kostnadene har vært kjøp av tjenester knyttet til systemutvikling og prosjektering. Dette ble gjennomført basert på NTNUS rammeavtaler for kjøp av IKT-konsulenttjenester. Bouvet vant konkurransen om første fase, mens Experis

Veien videre

NTNU Vitenskapsmuseet har ambisjoner om å videreutvikle Samlinger på nett gjennom å inkludere flere av museets kulturhistoriske samlinger, øke kvaliteten med flere foto samt legge til/utbedre funksjonalitet basert på fagmiljøets behov. Feilretting, drift samt utviklingsoppgaver av mindre omfang utføres for tiden av lokal faglig IT. Omfattende videreutvikling av funksjonalitet, flere språk samt eventuell utvidelse til flere museer og flere samlinger fordrer kjøp av drifts- og utviklingstjenester.

NTNU Vitenskapsmuseet ønsker ikke å inkludere datasamlinger som har for lav kvalitet til å kunne oversettes eller kartfestes i en felles visning gjennom på Samlinger på nett. Samlingene må også være delvis fotografert. Det er ikke interessant for Vitenskapsmuseet å drive en ren visningsportal.

NTNU Vitenskapsmuseet har oppnådd målet med å publisere botanisk samling: Vi har fått tilgjengeliggjort scans og testet at publisering av naturhistoriske samlingsdata fra MUSIT-basen var mulig i løsningen. Basert på råd fra fagmiljøet og en vurdering av kost- nytte-verdi, ønsker vi ikke å gå videre innenfor Samlinger for nett med naturhistoriske samlinger. De botaniske og zoologiske samlingene har gode og utbredte portaler på plass allerede både nasjonalt og internasjonalt. En parallell løsning innenfor Samlinger på nett vil ikke medføre en merverdi med mindre samlinger på nett videreutvikles med funksjonalitet som en ikke finner i hhv. Artskart og GBiF. Samlinger på nett inneholder i dag ikke trivielle navn, noe som gjør den lite relevant for andre enn forskningsmiljøene. Norske navn ligger imidlertid i datagrunnlaget, og dette bør kunne løses uten for store kostnader.

En rendyrking av kulturhistoriske data med avpublisering av de botaniske samlingsdataene vil kunne gi en enklere brukeropplevelse og mindre komplisert filtrering/venstremeny. Fagmiljøene ser ikke en vesentlig merverdi i å presentere disse dataene sammen.

Museet har mye dokumentasjon som omhandler gjenstandene, både fra funn-/utgravingsfase, men også fra konservering. Integrering av funnpresentasjonen med digitaliserte arkivalia og/eller pekere til utgravingsrapporter, røntgenrapporter, vitenskapelige publikasjoner eller annen digital informasjon vil gi en betydelig økning i verdien av Samlinger på nett. Den underliggende tekniske løsningen for kobling til litteratur

er på plass, men det krever betydelig personalinnsats å koble sammen informasjonen. På sikt ser en for seg at etablering av digitale løsninger for arkivalia i form av en eller flere «repositories» vil gjøre det mulig å få på plass en effektiv og automatisert integrering mellom objektsamlinger og tilhørende dokumentasjon.

Siden oversettelsesmotoren er på plass, er det teknisk relativt enkelt å legge til flere språk. Det bør vurderes å få på plass nord-samisk, i alle fall om løsningen tas i bruk av flere museer.

Arkeologer er som regel opptatt av sluttede funn, dvs. det komplette, avgrensede utvalget av gjenstander som hører til f.eks. et gravfunn. Selv om det finnes underliggende grupperingsvariabler i datasettet, er det i dag vanskelig å få frem dette på en enkel måte i portalen.

En bør vurdere å skifte rammeverk for bilde-/ fotovisning som er bedre tilpasset større funnmengder. Dette vil gjøre det mere effektivt å vise og zoome i store bildefiler.

Det må sikres god dialog og oppfølging med dataforvalter, gjerne en SLA, når det gjelder endringer i datadumpen. Eksempelvis gjelder dette når kommuner og fylkeskommuner slås sammen og endrer navn.

Vedlikehold og feilretting – restanseliste

- Løsningens skalerbarhet. I dag begrenses visningen til 60 gjenstander, eller eksport av 10 000 poster. Dette er krevende å forholde seg til når en f.eks. skal se på store funn med mange gjenstander.
- «Gallerifunksjonen» er delvis ustabil, og viser ikke alltid det fulle utvalget av funn som er lagt inn i et enkelt galleri.
- Tilbakemeldingsfunksjonen kan utbedres
- Det er problemer med eksport av detaljerte datasett fra brukernes gallerier/samlinger.
- Muligheter for eksport av mere enn 10 000 poster i enklere formater som CSV eller TAB-filer