

Liten, stor, større - Digitalarkivet i utvikling gjennom 20 år

Av Anette Skogseth Clausen, fagdirektør i Tilgjengeliggjøring i Arkivverket.

Det er en stor glede for meg å stå her i dag 20 år etter oppstarten av Digitalarkivet. Selv ble mitt møte med Digitalarkivet året etter, da jeg ble ansatt ved Statsarkivet i Bergen for å ta meg av Digitalarkivet. Dette skulle bare være en del av min stilling, men innhold og bruken vokste raskt og krevde mer, så det gikk ikke lang tid før jeg måtte bruke all min tid på Digitalarkivet. Jeg har i dag tenkt å snakke om hvordan veksten har vært for Digitalarkivet gjennom disse 20 årene og hvilke konsekvenser dette har fått for utviklingen, og de strategier og valg vi har måttet ta underveis. Jeg har ikke tenkt å bli veldig teknisk, der har vi Espen Tønnesen til å ta seg av det. Mens Kristian Hunskaar vil komme nærmere inn på det som gjelder det skannede materialet.

Digitalarkivet sitt fundament er selvsagt innholdet, det digitaliserte arkivmaterialet. Samtidig er suksessfaktoren for Digitalarkivet at tjenesten faktisk blir brukt. Uten brukere er det ikke noe grunn til å tilgjengeliggjøre det digitaliserte arkivmaterialet. Samtidig, for å få til god bruk av innholdet, forutsettes det gode løsninger for søking, framfinning og visning. Så når vi skal se på Digitalarkivets utvikling er det nettopp disse tre elementene jeg vil komme inn på, innhold, bruk og tekniske løsninger.

Innhold

Innholdsmessig hadde Digitalarkivet en litt beskjeden start. Jan Oldervoll og Historisk institutt hadde sørget for at opplysningene i 1801-tellingen var transkriberte og søkbare på nett (siden 1995) og sammen med noe transkriberte data fra arkiver ved Statsarkivet i Bergen og Hamar utgjorde dette oppstarten på Digitalarkivet.

Men dataene økte raskt. Avtalen om å overta folketellingsdataene fra Teleslekt og videreføringen av registreringsenheter i Stavanger og Voss sørget for dette. Og ansatte i Arkivverket, da særlig her i Bergen, transkriberte også i stor stil for å gjøre innholdet i arkivmaterialet søkbart i Digitalarkivet.

Ganske raskt meldte også frivillige seg. De ønsket å bidra med å transkribere kirkebøker og andre kilder. Så allerede et år etter oppstarten var Digitalpensjonatet på plass. Prinsippene var like enkle da som nå: send oss dataene du har transkribert fra arkivmateriale og vi publiserer dem i Digitalarkivet. Du beholder rettighetene til dataene og er dermed også ansvarlig for innholdet, mens vi har et redaktøransvar for det hele.

Det vi så rimelig raskt var et behov hos oss, men også et ønske fra brukerne, til noen regler om hvordan data skulle transkriberes. Det var behov for en oppskrift. For folketellingene var HISTFORM-standard på plass allerede i 1995, men for transkribering av kirkebøker måtte vi utarbeide et regelsett. KYRRE Standard for kyrkjebokregistrering i Noreg ble lansert i 2005. Det ga en forutsigbarhet som betydde arbeidsbesparelser og økt effektivitet i tilrettelegging og publisering. Samtidig åpnet enhetlig struktur på dataene for nye muligheter for søking, som jeg vil komme tilbake til.

Og i 2001 ble det gjort avtale mellom Slekt og data og Riksarkivaren om transkribering av kirkebøker. Slekt og data sørget for å mobilisere sine medlemmer til å transkribere og påtok seg koordinering av innsatsen. De stod også for utviklingen av et eget verktøy for transkribering av kirkebøker, Augustus. Samarbeidet med Slekt og Data og utviklingen av Augustus har hatt svært positiv utslag på datamengden i Digitalarkivet og den strukturerte måten det har kommet inn på.

Vi er alltid utrolig glad for alle bidrag vi mottar. I frykt for å skremme vekk bidragsytere har vi vært forsiktige med å styre innsatsen for mye. De frivillige skulle få skrive av det de selv ønsket. Likevel har vi sett behov for å ta noe regi på hva som transkriberes og sette opp noen dugnader for at vi skal bli ferdig på et område. En slik vellykket dugnad var transkriberingen av kirkebøker fra perioden 1801-1814. Det pågikk over flere år og pga eldre håndskrift og krevende oppsett i kirkebøkene var dette ikke noe alle kunne bli med på. Likevel, prosjektet kom i havn takket være stor innsats fra frivillige i samarbeid med Arkivverkets egeninnsats. Og vi lærte at vi gjerne kan ha flere slike dugnader. Dugnadsånden er stor blant bidragsyterne. Nå er litt over halvparten av dataene transkribert av frivillige. Og pågangen fra de frivillige er såpass stor at denne andelen vil nok øke framover!

Parallelt med det transkriberte innholdet har vi det skannede. I begynnelsen var det heller beskjedent. Noe skanning ble gjort ved statsarkivene og materialet ble lagt ut i Bokhylla – dette var før NB var i gang med sin bokhyllen.no. Vi så at med skanning kunne vi få ut mer arkivmateriale raskere enn hvis innholdet skulle transkriberes. Til tross for iherdig innsats, ga ikke dette de store volumene. Taktskiftet kom i 2004 da Riksarkivet satte i gang digitaliseringen av mikrofilmene. Nå snakker vi masseproduksjon! Jeg regner med Kristian Hunskaar vil gi oss gode beretninger herfra. Det er ikke tvil om at den satsingen som ble gjort da, sammen med den videre satsingen på transkribering av arkivmaterialet, la et godt fundament for Digitalarkivets suksess.

Med skanning i stor stil er det fremdeles viktig å transkribere innholdet. Det vil gi søkbare innganger til det skannede materialet og dermed en betydelig forenkling i framfinningen i materialet. Skanningen av folketellingen fra 1920 pågår nå for fullt. Og jeg håper og tror at ved hjelp av frivillige

skal vi klare å ha søkbare innganger til den skannede tellingen når sperringen oppheves i desember 2020.

Teknisk løsning

Etterhvert som innholdet i Digitalarkivet har vokst har også behovene for tekniske tilpasninger økt. I begynnelsen var mengden håndterbar og brukerne kunne lett holde oversikt over hva som var tilgjengelig i Digitalarkivet. Søkingen foregikk i hvert av registrene. Du måtte ha en ide om hvor i landet og i hva slags arkivmateriale du skulle lete for å finne opplysningene du var på jakt etter. Etterhvert som innholdet vokste ble det derfor økt behov for en oversikt over hva som var tilgjengelig, samtidig som et ønske fra brukerne meldte seg om å kunne lete gjennom alle dataene uten noen særlige forkunnskaper om geografi eller arkiv.

Et slikt globalt søk skulle vise seg vanskeligere å realisere enn ventet. Det forutsatte en mer homogen datastruktur enn det vi hadde. Den frie transkriberingen som vi hadde praktisert fram til Kyrre-standard kom, hadde gitt oss mye data, men med svært forskjellig oppsett. Vi gjorde et forsøk på å sette opp et såkalt globalt søk, men vi så raskt at dette ikke ville fungere i lengden. Vi måtte vente helt til 2011 før vi kunne oppfylle brukernes ønske. I "nye Digitalarkivet" som var blitt introdusert ved lanseringen av 1910-tellingen året før lå det søkeboks på framsiden og en datastruktur i bunn som skulle gi brukerne det etterlengtede søket gjennom alle data. Nå begynte en stor jobb med å omstrukturere alle dataene fra "gamle Digitalarkivet" slik at de kunne passe inn i den nye strukturen. Det arbeidet pågår faktisk fremdeles idag. Langt på vei er de fleste data vellykket overført til nye Digitalarkivet og brukerne kan i dag søke gjennom 32 millioner navn.

Det har likevel sin pris å sette opp ett søk gjennom alle data, f.eks hvilke opplysninger kan det søkes etter, her må man velge de mest brukte/vanlige; og hvordan det kan søkes, med store datamengder og høy bruk kan ikke søkene bli for tunge og kompliserte. Brukerne har måttet lære seg å "trunkere" og bruke "eller"-streker. Og ikke alle har tatt i mot dette med like stor entusiasme. Vi lytter til brukerne våre og forsøker å tilfredsstille deres behov der det er mulig. Derfor var det også for å forenkle søkingen for brukerne at vi introduserte «variant»-søket i forbindelse med lanseringen av "nyenye Digitalarkivet" her i juni i fjor. Dette gjør det mulig å søke etter Kristoffer uten å trunkere, og likevel få treff på alle stavevarianter av Kristoffer. Dette vil Espen Tønnesen komme nærmere inn på.

Men det er ikke bare dataene som har krevd større gjennomgang. Det å finne fram til arkivmaterialet i Digitalarkivet krever også gode systemer. For hvordan organiserer man det digitaliserte arkivmaterialet på en slik måte at alle finner fram når mengdene vokser i stort tempo? Vi har nå i underkant av 72 000 kilder i Digitalarkivet og vi ønsker at alle skal finne fram til akkurat det de er på jakt etter.

Det er viktig for oss at det er så enkelt som mulig å ta i bruk Digitalarkivet for hvem som helst. Derfor har vi valgt å presentere det med ord og oppbygging som er mer intuitive for folk flest. I Digitalarkivet ble derfor inngangene til det digitaliserte arkivmaterialet kildekategorier som folketellinger, kirkebøker, skatt og regnskap, eiendom osv. Det resulterte i at en del arkivmateriale havnet i samlekategorien "diverse". Det er materiale som ikke har en tydelig fellesnevner. Samtidig kunne kildene bare være knyttet til en kildekategori og det var ikke alltid godt nok. I forsøket på å både kunne skille arkivmateriale fra hverandre i større grad og for å kunne samle dem på tvers av kildekategorier, begynte vi dermed med tema-inndelinger av innholdet i tillegg til kildekategori, men også her så vi at det var noe begrensende. I stedet har vi nå landet på emneknagger. Her kan vi legge på flere emner om gangen og oppnå både det å skille ut materiale og det å samle dem på tvers av arkiv og kildekategori.

Mens disse forbedringene har vært store og viktige for brukerne som sådan, har vi også gjort noen betydelige forbedringer som nok ikke brukerne oppdager. I begynnelsen av skanningen, var det fokus på det mest etterspurte arkivmaterialet på lesesalene i Arkivverket. Dette skulle gjøre brukerne selvbetjente og gi dem en døgnåpen digital lesesal. Samtidig så vi at det ikke var alt som kunne publiseres på nett pga personvern hensyn. Det gjorde det tungvint for saksbehandlere som ikke kunne finne alt tilgjengelig i Digitalarkivet selv om det ikke nødvendigvis var sperret for bruk. For å gjøre en lang historie kort: tekniske utbedringer ble gjort og i 2016 kunne vi lansere ny løsning for skannet arkivmateriale, nå med tilgangsstyring.

Dette har introdusert en helt ny bruk av Digitalarkivet. Nå kan vi legge ut skannet arkivmateriale som må være sperret på nett, og så gir vi tilganger til dette materialet for Arkivverkets saksbehandlere. Men, ikke bare dem, også saksbehandlere i forvaltningen som NAV, Kartverket og domstolene, får tilgang til sitt eget arkivmateriale gjennom Digitalarkivet.

Digitalarkivet har beveget seg et skritt videre: fra en døgnåpen digital lesesal, til et døgnåpent digitalt magasin med tilhørende digital lesesal.

Bruken

Til nå har jeg snakket om innholdet i Digitalarkivet, som har vokst og også skapt nye muligheter, samt litt om de tekniske løsningene som skal ha gjort det enklere for brukerne å finne fram og gitt mer effektive søk. Hvordan har det påvirket bruken av tjenesten?

I oppstartsårene var det selvsagt få som hadde hørt om Digitalarkivet. Bruken var preget av hvilke landsdeler vi hadde innhold fra. Med mye transkriberte data fra Vestlandet var det naturlig nok her

det var størst bruk. Foreninger som fikk øynene opp for Digitalarkivet og også bidro med data i Digitalpensjonatet, sørget også for å gjøre tjenesten mer kjent i sine kretser.

Samtidig ble det en mer demokratisk tilgang på arkivmateriale da masseproduksjonen av skannede dokumenter tok fatt. Nå ble arkivmateriale fra hele landet med i Digitalarkivet og tjenesten ble straks mer interessant for et bredere publikum, både i inn – og utland. Etterhvert skapte litt medieomtale blest om innholdet i Digitalarkivet. Noen ganger for mye av det gode. Vi satt på nåler når vi visste Digitalarkivet skulle bli omtalt på Dagsrevyen eller Tore på sporet. Det ga umiddelbare utslag og serverne klarte ikke så mange samtidige brukere. Denne type plutselige økninger i bruk har vi måttet ta høyde for og nå takler vi også den økte bruken som kommer av f.eks. tv-serien Hvem tror du at du er? Denne serien har hatt en voldsom betydning for besøkstrafikken i Digitalarkivet. Den vekker en nysgjerrighet til å leite etter egen slekt og da havner man raskt i Digitalarkivet. Og så er det gjort!

Jeg overlater til Espen å ta oss gjennom grafer for statistikken, men det er ingen tvil om at Digitalarkivet når ut til folk.

Så hvem er disse brukerne våre? Fra antakelser og tidligere brukerundersøkelser har vi visst at det var en viss mannlig dominans og at de var godt voksne. Statistikken nå viser at brukerne er faktisk hakket eldre enn vi trodde. Mens 20 % av brukerne er i aldersgruppen 55-64, så er majoriteten, 45 % i gruppen 65+. Dette krever helt andre vurderinger for nettsidene, design og tilpasninger enn det vil for en nettside som har et betydelig yngre publikum. Det er noe vi er veldig bevisst på. Samtidig ser vi at den mannlige dominansen er på hell. Det er liten overvekt av kvinnelige brukere i Digitalarkivet. Tidligere nevnte jeg at det var flest brukere fra Vestlandet i begynnelsen. Mens nå er Hordaland er det eneste vestlandsfylket blant ti på topp og befinner seg på 3. plass. Oslo med 21 % og Akershus med 10 % av brukerne dominerer.

Framover vil nok Digitalarkivet-brukeren bli noe mer differensiert. Vi er bare i startfasen av å tilrettelegge for saksbehandlere i forvaltningen og vi forventer økt bruk der. Samtidig opplever vi økt interesse for Digitalarkivet fra undervisningsinstitusjoner. Digitalarkivet er en forholdsvis enkel måte å introdusere kildebruk i undervisningen og å få illustrert norgeshistorien med eksempler. Så vi håper det vil gi oss flere yngre brukere.

Når vi utvider innholdet i Digitalarkivet, så øker vi også potensialet for flere brukere og nye brukergrupper. Som før nevnt: Arkivverket ønsker nå å gjøre Digitalarkivet til en felles løsning for alle depotinstitusjoner. Dette betyr et bredere omfang av arkivmateriale tilgjengelig i Digitalarkivet. Brukerne får én inngang inn til arkivmaterialet, og søking, framfinning og visning foregår på samme måte for alt materialet. Og et slikt samarbeid om fellesløsning vil også kunne gi bedre vilkår for å kunne forbedre og utvikle nye løsninger.

Det har vært inspirerende å arbeide med Digitalarkivet. Jeg er utrolig glad for å ha vært med på denne reisen og utrolig stolt over alt vi har fått til sammen med og for brukerne. Jeg ser fram til å fortsette arbeidet og er veldig spent på hvor vi er om ti år.

Takk for oppmerksomheten.

23.01.2018