

Kartlegging av materiale på nettet

[Kartlegging av materiale på nettet](#)

[Introduksjon](#)

[Hvor stort er det norske internettet?](#)

[Kartlegging og arkivering](#)

[Nettarkivet i Danmark](#)

[Andre oversikter](#)

[Trafikkdata](#)

[Automatiske sjangerinndelinger](#)

[«Folksonomier»](#)

[Kvalitative sjangerinndelinger](#)

[Ulike kart for ulike formål](#)

[Folksonomier, hva folk mener/synes](#)

[Etter formål](#)

[Popularitetsindeks](#)

[Sektor, avsender](#)

[Oppsummering](#)

Introduksjon

Bakgrunnen for denne rapporten er behovet for å revidere pliktavleveringsloven. Loven skal sikre kildemateriale for kunnskap om fortiden for ettertiden, og den digitale utviklingen siden lovens seneste revisjon i 1990 har medført at loven med tilhørende forskrift må oppdateres. Datatilsynet, Personvernemnda og Personverkommisjonen peker spesielt på personvernrettslige problemstillinger som må utredes i forbindelse med avlevering av elektroniske dokumenter fra nettet. Spørsmål om opphavsrett, bevaringsverdi og teknologiske utfordringer knyttet til arkivering av slikt materiale vil likeledes utgjøre en vesentlig del av utredningsarbeidet. Nærværende rapport vil ikke behandle disse kompliserte problemstillingene i særlig grad, og hovedvekten vil ligge på en overordnet beskrivelse av hvilke typer materiale som finnes på nettet. Det er enten ved at materialet ligger under .no-domenet eller kan klassifiseres som Norvegica, det vil si materiale som har norsk tilhørighet. I denne oversiktsfasen er kategorisering, struktureringsprinsipper og eventuelle tilgangsbegrensninger sentrale nøkkelord og oppgaver. Rapporten skisserer mengden og variasjonen av innhold på nettet, og viser til andre nasjoners innsats for å bevare nettinnhold. Rapporten viser dessuten til hvordan andre aktører og institusjoner forsøker å strukturere innhold på nettet. Den største og mest sentrale delen av rapporten er forslagene til en kartlegging som kan danne grunnlag for det videre arbeidet med å vurdere den forskningsmessige bevaringsverdien.

Hvor stort er det norske internettet?

Vi er bedt om å gi en oversikt over hva nettet inneholder. Oppgaven er selvsagt umulig i særlig detalj. 25. juli 2008 rapporterte Google at søkemotoren hadde indeksert en billion ulike sider.¹ Google gir ikke oppdaterte tall over innholdet i sin indeks, men det er all grunn til å tro at veksten har fortsatt. Googles indeks er trolig den beste oversikten som eksisterer over nettet, men den er en kopi av nettet, ikke et kart. Formålet er å finne en håndfull sider som kan svare på et spørsmål, og ikke å kategorisere helheten. Googles oversikt er et kart i målestokken 1:1, som er «upraktisk», som Umberto Eco har uttrykt det i et humoristisk essay.

En tilsvarende oversikt finnes ikke over norske sider (selv om de aller fleste nok er inkludert i Googles indeks). Man kan få en indikasjon over størrelsen ved å telle opp hvor mange domener som er registrert, men da må man huske at noen domener bare har en håndfull sider, mens domener som *VG.no* eller *Norge.no* har tusenvis av undersider. Norske virksomheter har registrert 527 847 domenenavn som ender med “.no”², mens 247 858 domenenavn er registrert i USA av virksomheter og personer med adresse i Norge³.

I det danske Netarkivet.dk som beskrives nærmere under er det til sammenligning registrert 1.1 millioner domener som ender med “.dk”, og det høstes i tillegg innhold fra 44 000 såkalte generiske *top level domains*⁴ som tilsvarer omtrent en tidel av domenenavn registrert i USA av danske virksomheter og personer⁵. I tillegg finnes det en lang rekke nasjonale domener som er åpne for registrering uten at man behøver særskilt tilknytning til det pågjeldende landet. For å få bruksrett til et .no-domene er det et krav om norsk postadresse og registrering i enhetsregisteret, mens for eksempel .dk-domenet er fritt tilgjengelig. Det gjelder også nettstedet som er registrert på domener som .as, .cc, .nu, eller .tv, og det er neppe mulig å fastslå det totale antallet nettsteder som kan klassifiseres som norvegica.

Det er liten tvil om at en stor del av den norske befolkningens bruk av nettet foregår på nettsteder som ligger utenfor .no-domenet. Det finnes ingen offisielle målinger av nordmenns totale nettbruk på samme måte som det finnes for norske nettsteder, men lister fra Alexa.com og Googles Ad planner sammenholdt med spørreundersøkelser viser at ikke-norske nettsteder som Facebook, YouTube og Google utgjør en vesentlig del av nordmenns nettvaner. Mange blogger som regnes som viktige stemmer i samfunnsdebatten befinner seg på filtenere hos *Blogger.com* (eid av Google) eller *WordPress.com*, og mange unge deler sin hverdag og sine tanker gjennom blogger på den norske portalen *blogg.no*. I tillegg til de store nettstedene kommer det tilnærmet uendelig antallet små nettsteder som tilsammen utgjør en betydelig andel av den totale bruken. For nærværende rapport er det kun realistisk å belyse denne problemstillingen med tanke på identifisering, vurdering av relevans for bevaring samt tekniske og opphavsrettslige utfordringer knyttet til selve innsamlingen og arkiveringen av innhold.

¹<http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>

²<http://www.norid.no/statistikk/>

³http://www.webhosting.info/registries/country_stats/NO

⁴http://netarkivet.dk/nyheder/Newsletter_Netarchive_dk_august2011.pdf

⁵http://www.webhosting.info/registries/country_stats/DK

Rapporten vil derfor primært kartlegge de overordnede kategoriene for innhold og bruksområder og ikke berøre de faktiske nettstedene i særlig grad.

Kartlegging og arkivering

Intensjoner om arkivering av materiale fra nettet har en relativt lang tradisjon både i de nordiske landene og globalt. Med utgangspunkt i webarkivering, det vil si arkivering av innhold publisert på *world wide web* som ble lansert på starten av 1990-tallet, har det siden andre halvdel av samme tiår eksistert offentlige, ideelle og private initiativer for å ta vare på publisert materiale. Webarkivering er dessuten et eget forskningsfelt, beskrevet blant annet i Niels Brüggers bøker *Archiving Websites*⁶ og *Web History*⁷. Internasjonalt er *Internet Archive*⁸ det mest omfattende og lettest tilgjengelige arkivet, mens i nordisk sammenheng har det svenske prosjektet Kulturarw3⁹, samarbeidsprosjektet Nordic Web Archive og det allerede omtalte danske netarkivet.dk samlet store mengder innhold. I Norge har Nasjonalbiblioteket foretatt større innhøstinger siden 2002, men etter 2009 er innhøstingen mer sporadisk i påvente av en fornyet konsesjon fra Datatilsynet. De fleste nasjonale initiativer samarbeider og utveksler erfaringer i regi av *Internet Preservation consortium (IIPC)* som er et internasjonalt samarbeidsforum for bevaring av nettdokumenter, samtidig som private selskaper tilbyr lignende tjenester på kommersiell basis. En gjennomgang av samtlige initiativer og den vitenskapelige bakgrunnen ligger utenfor rapportens intensjoner. Netarkivet.dk trekkes under frem som et eksempel på en kombinasjon av innhøstingsstrategier, og litteraturlisten inneholder den mest sentrale litteraturen innen webarkivering.

Nettarkivet i Danmark

I Danmark startet innsamlingen av nettdokumenter i 1997, men kun med arkivering av statiske, avsluttede dokumenter. Behovet for en ny arkiveringsstrategi var tydelig, og pilotprosjektet netarkivet.dk som startet i 2001 og Kulturministeriets rapport i forbindelse med "Udredning om bevaring af kulturarven" i 2003 resulterte i en endring av pliktavleveringsloven og loven om opphavsrett. Heretter skulle alle nettsteder med tilknytning til Danmark arkiveres for å sikre et tilstrekkelig og representativt utvalg til så vel forsknings- som kulturbevaringsformål. Arkiveringen ble underlagt det virtuelle senteret Netarkivet.dk, et samarbeide mellom Det Kongelige Bibliotek i København og Statsbiblioteket i Århus.

Enhver arkivering av internettet blir et øyeblikksbilde av hvordan internettet så ut på det pågjeldende tidspunktet. Bildet kan være et forsøk på å beskrive bredden av all tilgjengelig informasjon, eller det kan gå i dybden for å dokumentere en sides innhold. De tre strategier som er valgt i forbindelse med netarkivet.dk kombinerer disse overordnede metoder gjennom tverrsnittshøsting, selektiv høsting og begivenhetshøsting. Tverrsnittshøstingen foregår i utgangspunktet fire ganger årlig av alle .dk-domener, mens den selektive høstingen skjer daglig eller flere ganger daglig for 80-100 utvalgte nettsteder. Begivenhetshøstingen utføres to til tre ganger hvert år avhengig av hvilke begivenheter som antas å ha særlig nasjonal betydning.

⁶Brügger, Niels, 2005: *Archiving Websites*. CFI, Aarhus.

⁷Brügger, Niels (ed), 2010: *Web History*. Peter Lang, New York.

⁸<http://www.archive.org/>

⁹<http://www.kb.se/om/projekt/Svenska-webbsidor---Kulturarw3/>

Tverrsnittshøstningen tegner et bredt og til en viss grad detaljert bilde av alle domener, men mangler utviklingen over tid som blir oppfanget av den selektive høstingen. Nettavisene er et godteksempel på ofte oppdaterte nettstedet som følges med opptil flere daglige høstinger. Begivenhetshøstningen følger større nasjonale eller internasjonale begivenheter tett og over tid, for eksempel valg, katastrofer, sportsarrangementer eller kongelige bryllup. Fremveksten av nye nettsteder vies spesiell oppmerksomhet, i den grad de lar seg spore ut fra eksisterende nettsteder og annen omtale.

Den danske løsningen er dermed en kombinasjon av en automatisert arkivering av bredden og dybden av alt innhold og en kuratert arkivering av innhold som antas å ha høy bevaringsmessig verdi. Med dette som bakgrunn skal vi i det følgende se på hva vi vet om hva «det norske internettet» inneholder.

Andre oversikter

Som vi har antydnet ovenfor er det nær umulig å gi en oversikt over alle verdensvevens sider. Likevel finnes noen grove oversikter. Slike lister over nettsider og nettsjanger faller gjerne innenfor fire kategorier:

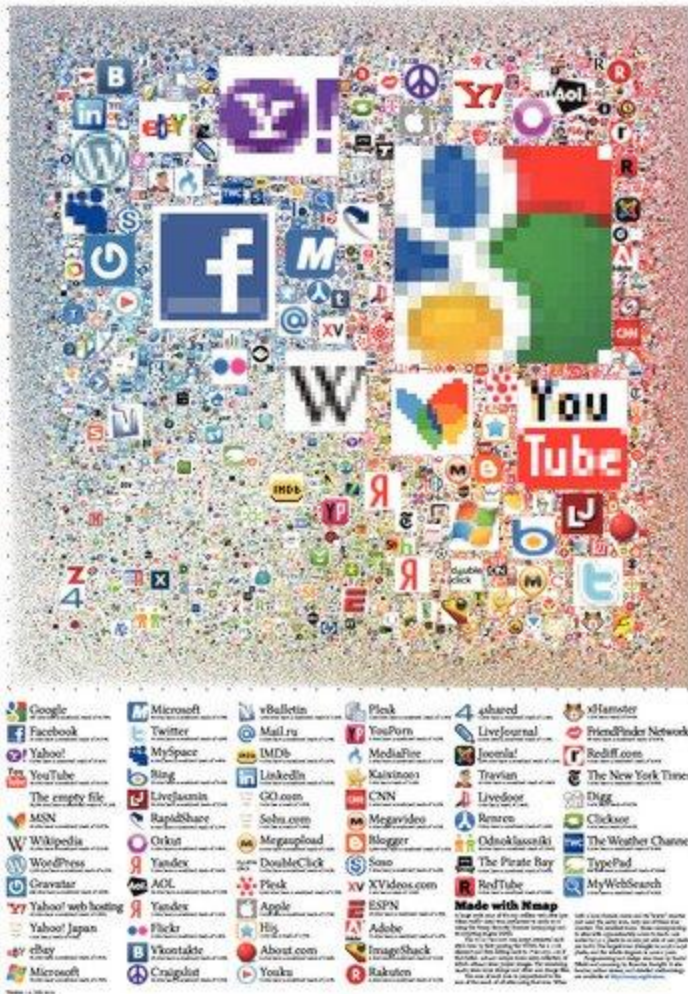
- Oversikter over trafikkdata
- «Folksonomier»
- Automatiske genererte sjangerklassifiseringer
- Kvalitative sjangeroversikter

Trafikkdata

Flere visualiseringer av World Wide Web basert på trafikkdata er publisert på nettet. Ett interessant eksempel er «Icons of the Web», som tok utgangspunkt i *Alexas* rangering av nettrafikk. Icons of the Web tok de én million mest besøkte nettstedene, og viste ikonene deres i et bilde der ikonets størrelse er proporsjonalt med nettstedets trafikk.¹⁰ Forfatterne anerkjenner at datagrunnlaget har flere feil, blant annet responderte ikke *Amazon.com*, verdens største nettbutikk, på søket deres, slik at nettstedet ikke er avbildet i oversikten under.

¹⁰<http://nmap.org/favicon/>

Icons of the Web



Trafikkdataene dette kartet bygger på, *Alexas* rangering, kan også gi en beregning av hvilke nettstedene som er mest besøkt av publikum i Norge.¹¹ De 100 nettstedene mest besøkt av brukere i Norge i september 2011 er listet i et vedlegg. Vi mener slike trafikktoversikter er en kilde til å identifisere nettsteder som er bevaringsverdige.

Automatiske sjangerinndelinger

Forskere som arbeider for å forbedre resultater fra søkemotorer har forsøkt å lære datamaskiner å kjenne igjen nettsjangre automatisk, slik at resultatene fra et søk kan sorteres

¹¹<http://www.alexa.com/topsites/countries/NO>. Oversikten er en kombinasjon av trafikktdata fra Alexas respondenter og søk foretatt i søkemotorer. Den norske oversikten er trolig over respondenter med maskiner med nettilknytning via en norsk nettilbyder. Vi vet ikke hvor representativ den er, men antar den er representativ for de største nettstedene.

etter sjanger.¹²Serge Sharoff har skrevet en algoritme som treffsikkert sorterer nettsider i sju store sjangre:

1. discussion – all texts expressing positions and discussing a state of affairs
2. information – catalogues, glossaries, other lists (mostly containing incomplete sentences)
3. instruction – how-tos, FAQs, tutorials
4. propaganda – adverts, political pamphlets
5. recreation – fiction and popular lore
6. regulations – laws, small print, rules
7. reporting – newswires and informative broadcasts, police reports¹³

Andre forskere har brukt lignende inndelinger, men ingen av dem forteller hvordan de har kommet fram til sine sett av sju-åtte sjangre. Trolig har de forsøkt seg fram ved å finne klasser av tekst med lett gjenkjennelige språktrekk.

En oversikt med bare sju kategorier er alt for grov for det formålet denne rapporten er skrevet for. Vi forlater derfor dette forsøket, men har sjekket at disse sju kategoriene er dekket av de oversiktene vi presenterer nedenfor.

«Folksonomier»

Folksonomi er en sammentrekning av *folk* og *taksonomi*, og beskriver en taksonomi basert på brukeres beskrivelse av et meningsområde. Man samler inn de termene folk bruker for å beskrive deler av området, og setter dem sammen til en helhetlig beskrivelse, nødvendigvis med stor grad av overlapp. Et problem med slike oversikter er at de ofte beskriver fenomener som ikke er varige. Sharoff nevner «digitale postkort» som et eksempel.¹⁴ På 1990-tallet var det vanlig å sende gratulasjonshilsener på e-post, enten med et pent bilde som vedlegg, eller med en lenke til en nettside som inneholdt en dekorert hilsen. Ti år senere er dette knapt i bruk. Et annet problematisk eksempel er «sosiale medier», et begrep som brukes svært mye i 2011, men som er vanskelig å avgrense. På den ene siden viser begrepets popularitet at det beskriver en viktig del av mediebildet. På den andre siden er begrepet såpass utflytende at det er vanskelig å avgjøre om et gitt nettsted er et «sosialt medium» eller ikke, og ulike personer vil trolig være uenige om tolkningen. Det eneste som synes sikkert er at *Facebook* og *Twitter* er sosiale medier.

Vår vurdering er at populære sjangerbetegnelser er viktige kilder når man leter etter bevaringsverdige nettsider. Språket er fleksibelt, og vi tar i bruk de ordene vi behøver. Når en ny, populær sjangerbetegnelse dukker opp, er det en indikasjon på at et nytt nettfenomen fortjener vår oppmerksomhet.

¹²Santini, Marina, Alexander Mehler, and Serge Sharoff. "Riding the Rough Waves of Genre on the Web: Concepts and Research Questions." In *Genres on the Web: Computational Models*

¹³Sharoff, Serge. "In the Garden and in the Jungle: Comparing Genres in the Bnc and Internet." In *Genres on the Web: Computational Models and Empirical Studies*, edited by Alexander Mehler, Serge Sharoff and Marina Santini, 153. Dordrecht: Springer, 2010.

¹⁴ibid.

Kvalitative sjangerinndelinger

Det vi her skal kalle *kvalitative sjangerinndelinger* er det som er det normale innenfor kunst- og kulturhistoriske fag som litteraturvitenskap, musikkvitenskap, kunsthistorie og filmhistorie. Basert på sitt kjennskap til store mengder verker deler eksperter dem inne i klasser og sjangre for å skape oversikt (særlig over verkene ekspertene er enige om at er gode). Det er svært mye litteratur om Web-sjangre som blogger, hjemmesider, sosiale nettverksteder eller nettnyheter¹⁵, men knapt noen forsøk på å dele alle nettets sider inn i forståelige sjangre. En av forfatterne av denne rapporten, Anders Fagerjord, har beskrevet 15 populære web-sjangre i boka *Web-medier*, uten at det er ment som en fullstendig liste¹⁶:

- Blogg
- Brukergenererte tidsskrifter
- Fotosamlinger
- Hypertekstlitteratur
- Firmahjemmesider
- Interaktiv grafikk
- Nettavis
- Nettmagasiner
- Personlige hjemmesider
- Portaler
- Søkemotorer
- Web-TV og nettradio
- Wikier

Nedenfor skal vi presentere flere slike kvalitative sjangeroversikter. Som vi har sett gir ulike tilnærminger svært ulike lister over sider på nettet. Vår vurdering er at en slik heterogenitet ikke er en ulempe, men en styrke. For å fange et noen hundre millioner sider må man trolig forsøke flere ganger med ulike metoder, noe vi skal utdype i neste del.

Ulike kart for ulike formål

Vi har ikke klart å finne én oversikt over det norske nettet som gir et pålitelig overblikk, til det er nettet for stort og mangslungent etter vår mening. I stedet foreslår vi fem ulike slags oversikter, som vi vil sammenligne med fem ulike kart over et terreng. Et godt verdensatlas vil i det minste inneholde politiske kart og topografiske kart, men gjerne også kart over arealbruk (jordbruk, byer, ørken, og så videre), befolkningstetthet og viktige næringer. Slik ulike kart trekker fram

¹⁵Se f.eks.: boyd, danah m., and Nicole B. Ellison. "Social Network Sites: Definition, History, and Scholarship." *Journal of Computer Mediated Communication* 13, no. 1 (2007). Engebretsen, Martin. *Digitale Diskurser: Nettavisen Som Kommunikativ Flerbruksarena*. Kristiansand: Høyskoleforlaget, 2007. Rettberg, Jill Walker. *Blogging*. Cambridge: Polity, 2008. Serfaty, Viviane. *The Mirror and the Veil: An Overview of American Online Diaries and Blogs*. Amsterdam: Rodopi, 2004.

¹⁶Fagerjord, Anders. *Web-medier: Introduksjon Til Sjangre Og Uttrykksformer På Nettet*. Andre utgave. Oslo: Universitetsforlaget, 2008. Legg merke til at verken «nettsamfunn» eller «sosiale medier» er på lista, de var ikke tydelige populære sjanger i 2007, da utgaven ble laget.

ulike aspekter ved et geografisk område, foreslår vi fem oversikter som gir ulik prioritering, og får ulike typer nettstedet til å tre frem som tydelige:

1. Folksonomier
2. Kommunikasjonsform
3. Formål
4. Popularitetsindeks
5. Samfunnsektor

Vi skal gi eksempler på disse nedenfor. Disse eksemplene er lette å kritisere for å være mangelfulle og skjevt sammensatt. Vi har ikke hatt tid til å utvikle og kvalitetssikre dem i særlig grad innenfor den korte fristen vi har fått for denne rapporten. De er derfor å betrakte som eksempler.

Folksonomier, hva folk mener/synes

Som nevnt over, tror vi undersøkelser av folks språkbruk alltid vil være en nyttig kilde for å se hvordan nettet blir brukt på et gitt tidspunkt. Å gjøre en grundig undersøkelse av hvilke sjangre den norske almenheten opplever at finnes på nettet i 2011 faller utenfor rammene for denne undersøkelsen, men vi vil foreslå at Nasjonalbiblioteket jevnlig gjør slike undersøkelser, for eksempel gjennom fokusgruppeintervjuer eller analyser av språkbruk i nyhetsmediene. Vi har forsøkt å ta inn dette aspektet ved å bruke populære betegnelser på sjangre i oversiktene nedenfor.

Kommunikasjonsform

En svært intuitiv inndeling vil for mange være å dele inn etter den primære uttrykksformen. Da vil man kunne operere med i hvert fall disse klassene:

- Levende bilder
- Radio
- Musikk
- Fotografier
- Tegninger
- Kart
- Skrift

Hver av disse kan deles inn etter hvem som er avsender:

- Profesjonelle redaksjoner
- Profesjonelle enkeltpersoner
- Amatører

På samme måte kan man skille mellom ulike målgrupper.

Man kan også se på formen for informasjonsutveksling:

- Offentlig, alment tilgjengelig:

- En-til-mange (kringkasting, massemedier, informasjon fra enkeltpersoner og virksomheter)
- Dialog-til-mange (debattfora)
- En-til-noen (blogg, nettsamfunn)
- Mange-til-mange (fildeling, dugnadstjenester som *Wikipedia*)
- Åpne kataloger/«system-til-en» (offentlige registre, prisoversikter, kataloger)
- Lukket
 - En-til-en (dialog)
 - Lukkede fora
 - En-til-system (katalogtjenester, billettbestilling, automatisk registrering og overføring som nettbanks eller selvangivelse)
 - Transaksjoner mellom brukere (som nettbanks og betalingstjenester, men også kontaktannonser)
 - Helt privat (nettdagbøker, private filtenere, dokumenttjenester som *Google Docs*)

Det er vanskelig å lage brukbare oversikter fra disse inndelingene, men de kan være nyttige sjekklister for andre oversikter, for å se om man har fått med det viktigste.

Etter formål

En første grovsortering av nettet kan gjøres etter hvilket formål sidene har. Her kan man liste opp store oversjangre som svært mange nettsider vil høre inn under.

- Kunnskap og informasjon: Forskningsartikler, lærebøker, slik-gjør-du-oppskrifter, oppslagsverker, osv.
- Informasjon om virksomheter til deres målgruppe: Firmahjemmesider, offentlige etater
- Nyheter og underholdning — nettaviser og kringkasting, nettmagasiner, sport
- Diskusjonsfora
- Terminaler for store transaksjonssystemer: Banker, billettbestilling, bibliotek
- Personlige uttrykk: Hjemmesider, blogger, mikroblogg
- Sosialt fellesskap: Nettsamfunn
- Deling av budskap: Bilde- og videodelingstjenester som Flickr og YouTube, fildelingstjenester som Pirate Bay, kunstnettsteder.
- Spill

Popularitetsindeks

Lister over de mest trafikkerte nettstedene viser hva store grupper av nordmenn er opptatt av, og vil være uunngåelige sjekklister for den som vil bevare de viktigste delene av nettets norvegica. Som et eksempel har vi lagt ved de hundre nettstedene som *Alexa* registrerte mest trafikk til fra Norge i september 2011.¹⁷ Vi har gått gjennom listen, og gruppert nettsider vi mener har

¹⁷ Vi har basert oss på *Alexa* av makelighetshensyn: Dataene er lett tilgjengelige. I Norge samler også TNS Gallup, og kanskje også andre, trafikkdata, som ideelt sett burde sammenstilles med *Alexa*s for å få et mer korrekt bilde.

fellestrekk under populære navn. Dette har resultert i de følgende tjue kategoriene, som vi vil påstå er de mest brukte slagene nettsider i Norge. Antallet medlemmer i hver kategori er oppgitt i parentes. Summen blir ikke hundre, da noen nettstedet er kodet i flere kategorier, og fire nettadresser ble ikke kategorisert, da de ikke ble vurdert som egne nettsteder.¹⁸

Nettavis: Et redaksjonelt styrt nettsted som bringer nyheter og underholdning, og oppdateres flere ganger i uka (22)

Virksomhetsinformasjon: Hjemmesider for firmaer og offentlige virksomheter. Ofte kombinert med nettbutikk og bestillingstjenester (11)

Deling: nettsteder hvor hvem som helst kan legge ut tekst, bilde, video eller annet materiale og dele det med andre (9)

Nettbutikk: Postordresalg via en katalog på nettet (9)

Søkemotor: Nettside hvor man kan taste inn et eller flere ord, og få opp lister over nettsider som inneholder ordet eller informasjon om det (6)

Bloggplattform: Et nettsted hvor registrerte brukere kan publisere sin egen blogg. Her har vi inkludert *Twitter*, som vi ser på som en blogg for korte innlegg, en «mikroblogg», men som også oppfattes som et nettsamfunn (6)

Nettleksikon: Oppslagsverker på nett, både generelle (som *Wikipedia*) eller tematiske (som *Internet Movie Database*) (4)

Offentlig informasjon: nettsteder som publiserer informasjon fra offentlige etater. Her har vi, under tvil, inkludert et nettsted som søker i skattelistene (4)

Pornografi: nettsteder med pornografiske bilder og filmer. Noen er også kodet som delingstjenester (4)

Portal: En nettside designet for å være et sted brukeren besøker for å komme videre til andre nettsteder. Inneholder gjerne søkemotor, tematisk katalog over nettsider, nyheter og lenker til anbefalte nettsteder (4)

Telefonkatalog: Kataloger over personer og virksomheter med telefonnummer, adresse, og annen informasjon (4)

Diskusjonsforum: Arena for offentlig debatt i form av skriftlige innlegg under et tema, og kommentarer til tidligere innlegg(3)

¹⁸ Blant de 100 mest besøkte domenene fant vi også hjelpetjenester for andre tjenester. Dette er tjenere som leverer deler av sidene til andre nettsteder. Vi fant en trafikklogg, en annonseformidler, en automatisk lenkeforkorter, og en tjener for publisering til mobiltelefoner. Disse adressene har neppe særlig mange brukere besøkt bevisst, men de er blitt kalt opp automatisk når brukere har besøkt andre nettsteder.

Nettbank: Selvbetjent bank på nettet (3)

TV-selskap: Dette er nettstedene til kjente kringkastingsselskaper. De inneholder både nettviss, Web-TV og virksomhetsinformasjon, og kunne vært kodet i de kategoriene, men vi antar de oppleves som en egen sjanger hos publikum (3)

Markedsplass: Nettsteder som formidler annonser, kontakt mellom selgere og kjøpere, og betalingstjenester (2)

Nettsamfunn: Et nettsted der en bruker kan opprette en personlig side, liste hvilke brukere han eller hun har et særlig forhold til, og studere andres lister over slike nære brukere. Vi har inkludert *Facebook* og *LinkedIn*. Mange andre nettsteder har også disse egenskapene (som *Twitter*, *Tumblr*, *YouTube* og *Flickr*), men når de overveiende blir brukt til deling eller blogg har vi kodet dem i de kategoriene (2)

Betalingstjeneste: Nettsted som formidler betaling mellom en kjøper og en selger (1)

Datingtjeneste: Kontaktannonser på nett (1)

Øversettelse: Tjeneste hvor brukere kan skrive inn en tekst og få den automatisk oversatt (1)

Værmelding (1)

Web-TV: Redaksjonelt produserte videoer som kun kringkastes på nettet (1)

En innsamling av nettsteder fra disse tjue kategoriene ville fange en stor del av det norske nettet, men samtidig ville en betydelig andel falle utenfor. Det ser vi om vi vurderer andre inndelinger.

Sektor, avsender

De fleste virksomheter og organisasjoner av en viss størrelse har i dag egne nettsider. Det er derfor mulig å gå gjennom ulike samfunnssektorer og typer organisasjoner for å finne informasjon. Vi kjenner til store mengder informasjon og tjenestetilbud på i det minste disse områdene:

- Offentlig
 - Helse og omsorg
 - Skoler
 - Universiteter
 - Forskning
 - Forsvar
 - Finans
 - Skatt
 - Politi
 - Domstoler

- Lover
- Demokratiske institusjoner
- Kongehus
- Privat
 - Ideelle organisasjoner og interesseorganisasjoner
 - Sport og idrett
 - Kulturliv
 - Værmelding
 - Produktinformasjon
 - Varehandel og nettbutikker
 - Kataloger
 - Informasjon om åpningstider osv.
 - Markeds plasser (Finn osv.)
 - Finans
 - Nettbanker
 - Aksjehandel
 - Finanspresse
 - Finansblogger
 - Kursoversikter
 - Håndverkstjenester
 - Kontakt, åpningstider, osv.
 - Bestillingstjenester
 - Medier
 - Forlag
 - Presse
 - Kringkasting
 - Blogger
 - Private medier
 - Sosiale medier
 - Én-til-en-medier
 - Delingstjenester
 - Private (lukkede) lagringstjenester («sky»)
 - Nettkataloger og søketjenester

Oppsummering

En kartlegging av hva som finnes av materiale på nettet er en oppgave som ikke kan besvares entydig i en kort rapport. Samtidig er det lite sannsynlig at en besvarelse som strekker seg over flere år og bind ville være fullt ut dekkende og samtidig fremtidssikret, for terrenget som beskrives er i hurtig og konstant endring. Oversiktsrapporten som er presentert her er ment som en inngang til feltet og åpner for videre diskusjoner og utredninger om bevaringsverdi, teknologiske arkiveringsutfordringer og ikke minst spørsmål om personvern.