

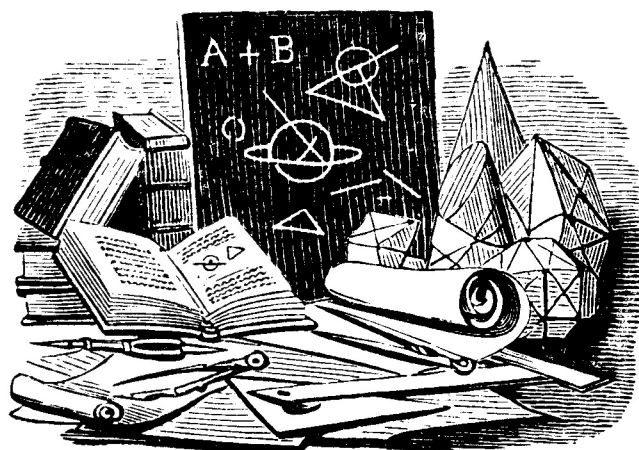
f f a k t a

f ø r s t

Viderebruk av data- kilder i offentlig sektor: potensial og hindringer

*Prosjektrapport,
Institutt for informasjons-
og medievitenskap*

*Universitetet i Bergen
Januar 2010*



Innhold

1. Oppsummering/Summary
2. Åpne data: Fra nerdetema til valgvinner
3. Kartlegging av offentlige datakilder i Norge: resultater
 - 3.1. Nettsteder
 - 3.2. Begynnelsen på et norsk data.gov
 - 3.3. Spørreundersøkelse
 - 3.4. Kommune og fylkeskommune: Bergen og Hordaland
 - 3.5. Oppsummering av kartleggingsarbeidet
4. Fra data til innsikt: muligheter og motforestillinger
5. Sikker og effektiv frigivelse av data: forslag til tiltak
6. Analyser og "bruksanvisninger"
 - 6.1. Nick Diakopoulos: From Civic Data to Civic Insight
 - 6.2. Eirik Stavelin: Kan mine data gjenbrukes?
 - 6.3. Visuell ABC for data-byråkrater
7. Dokumentasjon

Liste over figurer:

1. Ikke forsidemateriale?
2. Finnbare datakilde-opplysninger
3. Departemental datamangel
4. Koordinerte fylkesmenn
5. Har romstert i skattkisten
6. Det er mer på kistebunnen
7. Mer data til viderebrukerfolket
8. Det meste gratis
9. Ingen budsjettredder
10. Data best på hjemmebane?
11. Det koster å være åpen
12. Varierende betingelser

Opphavsrett: Denne rapporten er publisert under lisensen Creative Commons Navngivelse 3.0 Norge. Det betyr at alle har lov til å dele - å kopiere, distribuere og spre verket. Lisensen gir også lov til å remikse - å bearbeide - verket videre. Mer informasjon om lisensen finnes her:

<http://creativecommons.org/licenses/by/3.0/no/legalcode>

1. Oppsummering

Offentlig sektor samler inn og genererer store mengder data av ulik type. De senere årene har interessen vært økende blant samfunnsaktører utenfor offentlig sektor for å bruke offentlige data i nye sammenhenger (forutsatt at det ikke er snakk om persondata). Både medier, frivillige organisasjoner, bedrifter og privatpersoner kan skape nye tjenester, ny innsikt og økonomiske verdier ved å bruke datakilder fra det offentlige som råmateriale. Effektiv viderebruk av data forutsetter at virksomhetene i offentlig sektor informerer om datakildene de forvalter og gjør data tilgjengelig i relevante formater.

Praksis varierer sterkt mellom offentlige virksomheter på ulike fagfelt og forvaltningsnivåer, viser kartleggingsarbeidet utført i dette prosjektet i perioden august-desember 2009. Enkelte virksomheter informerer grundig om sine datakilder og har gjort data tilgjengelig for nedlasting. En stor andel av de undersøkte virksomhetene tilbyr imidlertid mangelfull eller ingen informasjon om datakilder på sine hjemmesider. Hos disse virksomhetene svikter en grunnleggende forutsetning for viderebruk av data.

Inntrykket av varierende interesse og et ubrukt potensial forsterker seg når en ser resultatene fra en spørreundersøkelse blant statlige virksomheter fordelt på ulike fagfelt. To tredeler sier deres virksomhet har data med potensial for viderebruk som i dag ikke blir utnyttet. Undersøkelsen gir også inntrykk av at det arbeides med problemstillingen i mange virksomheter: Over seks av ti sier de vil gjøre flere data tilgjengelig for viderebruk i løpet av det kommende året. Økte kostnader og faren for at eksterne aktører kan misforstå dataene og spre villedende informasjon er de største hindrene mot at mer data blir frigitt, ifølge undersøkelsen.

Intervjuer med representanter for virksomheter viser også at selve problemstillingen om å gjøre data tilgjengelig er uvant for mange. En sammenligning med initiativer og debatter om offentlige data i et utvalg andre land (Storbritannia, Danmark, Nederland, USA) indikerer at oppmerksomheten om temaet er størst der hvor det har blitt satt på dagsordenen på øverste regjeringnivå.

Rapporten avsluttes med forslag om en rekke tiltak som kan settes i verk hurtig for å øke antall datakilder tilgjengelig for viderebruk. Et nettsted der offentlige datakilder samles, inspirert av USAs data.gov, vil være et åpenbart effektivt tiltak, særlig hvis det parallelt utarbeides klare prinsipper, regler og "bruksanvisninger" for hvordan datakilder gjøres tilgjengelig på sikker og brukervennlig måte. Rapporten peker også på behovet for en løpende debatt om kriterier for konstruktiv viderebruk av data. Mediene bør, i samarbeid med sitt publikum, ta en lederrolle i å arbeide fram eksempler på god viderebruk av åpne data.

English Summary

The public sector collects and generates vast amounts of data. In recent years the interest in re-using public, non-personal data has been increasing among citizens, groups and companies outside the public sector. The media, civil society groups, businesses and private citizens can use public data as "raw material" to create new services, new insight and economic value. Efficient re-use of public data requires that public sector agencies inform about their data sources and make data available in relevant formats.

Practice varies strongly between Norwegian public sector agencies in different subject areas and across administrative levels (state/regional/local), this fact finding project from August to December 2009 has revealed. Some agencies offer detailed information about their data sources and have made data available for download. However, a major part of the agencies assessed offer insufficient or no information about data sources on the homepage of their websites. Here a fundamental requirement for the re-use of data is missing. The impression of varying interest and unused potential is amplified by the results of a survey among state agencies. Two thirds of respondents say their agency possesses data with potential for re-use that is not utilized today. The survey on the other hand suggests that the subject of open data is on the agenda in many agencies; more than six out of ten say they plan to make more data available for re-use

during the coming year. The survey shows that increased costs and the concern that external groups will misunderstand the data and misinform the public are cited as the two greatest obstacles against more data being made available. In addition, interviews with public sector agency employees suggest that the topic of making data available is new to some agencies.

A comparison with initiatives and debates about open public data in a selection of other countries (Britain, Denmark, Netherlands, USA) show that the attention the topic receives is greatest when it is placed on the agenda at the highest political level. The report recommends a number of concrete measures that it is assumed would quickly increase the selection of data sets made available for re-use. A website that collects public data sources, inspired by the US government's data.gov, would be an obviously efficient initiative, especially when accompanied by a set of clear principles and rules and an "instruction manual" that describes how to make data available in a secure and user-friendly way. The report also points out the need for a parallel, ongoing debate about criteria for the constructive re-use of data. The media should, in cooperation with the public, play a leading role by producing examples of best practices in re-using open data.

Takk

Prosjektteamet har bestått av Gudrun Tokle Grene, Line Thams Reiersen og Olav Anders Øvrebø. Håvard Legreid har illustrert rapportens forside og laget figurene. En stor takk til alle som har hjulpet oss i arbeidet: De som har bidratt med å legge inn datakilder i vårt provisoriske "data.gov"-regneark; Nick Diakopoulos, Eirik Stavelin og Georg Apenes for glimrende artikler i Vox Publica; alle som har tatt seg tid til å svare på våre spørsmål i Bergen kommune, Hordaland fylkeskommune og Kristiansand kommune; alle som har kommentert våre blogginnlegg på "Fakta først"-bloggen og tipset oss via Twitter og andre kilder; Sondre Bjellås for å ha startet datakilder.no på eget initiativ; Dag Digernes, Harald Groven, Kristine Løwe og Bente Kalsnes for tips og inspirasjon; alle som stilte på NONA-møtet i Oslo og ga verdifulle innspill i starten av arbeidet; Ton Zijlstra og James Burke for plakater og opplysninger fra Nederland; Cathrine Lippert for informasjon om digitaliser.dk; og sist, men ikke minst, instituttleder Dag Elgesem for å ha satt prosjektet i gang. Listen kunne sikkert vært gjort lengre, vi beklager sterkt hvis vi har glemte noen!

2. Åpne data: Fra nerdetema til valgvinner

"Alle kan krevje innsyn i ei samanstilling av opplysningar som er elektronisk lagra i databasane til organet dersom samstillinga kan gjerast med enkle framgangsmåtar." Slik er ordlyden i offentlighetslovens nye paragraf 9, som trådte i kraft 1. januar 2009. Bortsett fra det generelt positive ved mer åpenhet i offentlig forvaltning, har enkelte ekstra god grunn til å glede seg over lovfestingen av innsynsrett i databaseopplysninger. Det er begynt å gå opp for et blandet, løst fellesskap av journalister, aktivister, tjenesteutviklere og rådgivere at tilgang til datakildene som produseres i offentlig sektor er en svært viktig forutsetning for nyskaping på flere samfunnsområder.

Fra fakta til data

Innenfor journalistikken har pionerer som Adrian Holovaty vist vei med flere prosjekter delvis basert på bearbeiding av offentlig sektors data. Det mest kjente av disse, EveryBlock, setter sammen en rekke ulike datasett til en aktuell, lokal informasjonsportal som nå dekker 15 amerikanske byer. Data om byggesaker, bedriftsetableringer, kriminalitet, restaurantinspeksjoner med mer settes bokstavelig talt på kartet, og summen blir en "nyhetsstrøm for ditt kvartal".

Bak slike tjenester ligger en ny forståelse av det evige råstoffet for journalistikk - de aktuelle faktaopplysningene. Holovaty sier at journalister og aviser må ta et oppgjør med sitt "story-centric worldview" (Holovaty 2006). Redaksjonene samler inn fakta som er perfekt egnet for strukturert framstilling, som tidspunkt, steder, personnavn, handlingstyper. Men istedenfor å bruke potensialet i disse dataene, strør de dem isteden ustrukturert utover i nyhetsartiklene (Nærland 2009). Å forlate fortellingen som framstillingsform er neppe ønskelig, men analyse- og opplysningsmulighetene i strukturerte data kan tilføre journalistikken en ny dimensjon.

James T. Hamilton og Fred Turner (2009) foreslår en foreløpig definisjon på den framvoksende journalistiske retningen "computational journalism":

[W]e define computational journalism as the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism.

Mediene gjennomlever for tiden en krise. Overgangen til digital produksjon og distribusjon presser de etablerte forretningsmodellene, og finanskrisen har akselerert endringstakten. Krisen har tvunget fram oppsigelser av journalister (og i USA, nedleggelse av store dagsaviser). Mange frykter at det ikke vil bli mulig å finansiere produksjon av god journalistikk i framtiden, i hvert fall ikke i samme omfang som tidligere.

I en slik situasjon får den nye offentlighetsloven og "computational journalism" potensielt stor betydning. Hvis denne nye formen for journalistikk virkelig kan bidra til å styrke de redaksjonelle mediens klassiske funksjon som demokratiets vaktbikkje, blir det i dagens mediekriser desto viktigere å gi den gode vekstvilkår. Et sterkt virkemiddel er tilgjengeliggjøring av dataene offentlig sektor produserer og forvalter.

Erfaringene både Holovaty og norske pionerer på feltet har gjort, synes å underbygge dette synspunktet. Journalister som vil lage nyskapende tjenester basert på analyse og framstilling av data må ofte bruke uforholdsmessig mye tid og krefter på å grave fram dataene. Dette innebærer ofte å "skrape" informasjonen fra de offentlige virksomhetenes nettsider. Enda vanskeligere er det hvis virksomheter ikke engang opplyser om hvilke data de disponerer. Ved å informere grundig om hvilke data de har og legge aktivt til rette for viderebruk, kan offentlige virksomheter øke sjansene for at samfunnet får mer av den formen for journalistikk "alle" etterlyser.

Aktivisme og verdiskaping

Offentlig sektors data er av stor interesse også for andre enn redaksjonelle medier. Debatter om frigivelse av data har versert lenger i andre miljøer enn de journalistiske, spesielt slike som ser kommersielle muligheter i viderebruk av data. Private virksomheter og miljøer som benytter seg av geodata har for eksempel lenge ønsket andre prisingsmodeller, eventuelt full frigivelse, av dataene som produseres av Statens kartverk, kommuner og andre virksomheter innenfor samarbeidet Norge digitalt. I de senere årene har også journalistiske miljøer sluttet seg til dette ønsket. Fra 1. desember 2009 har Kartverket frigitt sine visningstjenester for viderebruk, men dette inkluderer ikke selve kartdataene.

De som ønsker å gjøre offentlige data lettere tilgjengelig for bruk av andre aktører, har drahjelp fra EUs viderebruksdirektiv. Det ble vedtatt i 2003 og er tatt inn i norsk lov, blant annet via den nevnte endringen i offentlighetsloven.

Hovedhensikten med direktivet var å stimulere til økt verdiskaping gjennom å legge til rette for nye tjenester. Resonnementet bak er at staten vil ta inn mye mer i økte skatteinntekter som følge av ny næringsvirksomhet enn den taper ved at offentlige virksomheters adgang til å ta betalt for utlevering av data reduseres. De foreløpige erfaringene med frigivelse av data i Avinor kan belyse potensialet (se ramme).

Det er imidlertid enda en potensielt betydningsfull side ved frigivelse av offentlig sektors data. Data fra ulike fagfelt kan også styrke frivillige organisasjoner og borgerinitiativers mulighet til å kontrollere den utøvende makt og kjempe for sine hjertesaker.

Ofte kan det være snakk om at organisasjonene samarbeider med virksomheter og etater i offentlig sektor om å forbedre og videreutvikle det offentlige tjenestetilbudet. Slike ambisjoner går igjen i begrunnelsen for flere initiativer i ulike land de siste årene, der myndighetsorganer forsøker å oppmuntre til at borgere og organisasjoner bidrar til å identifisere uløste samfunnsoppgaver og til å finne nye, kreative løsninger på dem.

De ulike aktørene som er beskrevet her kan selvsagt samarbeide på mange ulike måter. Det er ingen grunn til å stille opp skarpe skiller mellom mediens og andre aktørers viderebruk av offentlige data; ofte vil for eksempel frivillige organisasjoner drive journalistikk-lignende virksomhet. Det er også fullt mulig å tenke seg konkrete samarbeidsprosjekter mellom mediebedrifter og mer eller mindre organiserte grupper. Slik kan offentlig sektors data også bidra til å virkeliggjøre noen av visjonene om borgerjournalistikk som er formulert det siste tiåret.

"Open Data": ikke lenger bare for nerdene

Statsminister Gordon Brown presenterte 7. desember 2009 den britiske regjeringens nye plan for reform av offentlig sektor under tittelen "Smarter government". Både i Browns tale og handlingsplanen legges det stor vekt på frigivelse av offentlig sektors datasett. Regjeringen lovet en "radikal åpning" av datakilder og

Avinors frislipp av flydata

Sommeren 2009 gjorde Avinor deler av sine data om trafikkbevegelser (flytider, flyplassnavn, selskaper) tilgjengelig for viderebruk. Eneste betingelse er kreditering og at dataene ikke blir brukt i sammenhenger som bryter norsk lov. Etter et halvt år oppsummerte Anders Christensen i konsultantselskapet Bekk erfaringene slik (Bekk har oppdrag for Avinor):

1. Økt datakvalitet og tjenesteorientering: Systemutviklere og "flyspottere" (flyentusiaster) ga Avinor tilbakemeldinger om mangler i datasettene, som så ble rettet opp. Åpenheten har også bidratt til å gjøre organisasjonen mer tjenesteorientert.
2. Kollektiv innovasjon: Kort tid etter lanseringen hadde eksterne utviklere laget de første tjenestene basert på dataene, både mobile (f.eks. iPhone, Android) og på nettstedet.
3. Omdømmeeffekt: Avinor har fått positiv medieomtale som følge av data-frigivelsen.

(Kilder: avinor.no, Christensen 2009).

en "sterk satsing på mer transparens". En egen liste med prinsipper for offentlig sektors håndtering av data ble presentert (se ramme). Her loves det blant annet at offentlige data skal publiseres i maskinlesbar form, tilrettelagt for viderebruk. Det skal bli lett å finne fram til datasett ved å gå til ett sentralt nettsted (www.data.gov.uk). Data skal offentliggjøres under en lisens som legger til rette for viderebruk, også kommersiell.

Den høye profilen Browns regjering gir offentlig sektors data - "open data" - henger til dels sammen med den aktuelle politiske situasjonen i Storbritannia. Etter skandalen med parlamentsrepresentantenes frynsegoder er økt åpenhet et politisk tema som engasjerer sterkt. Dessuten er britene særlig hardt rammet av finanskrisen og må snu alle steiner for å redusere statens utgifter. Likevel er vektleggingen av data-frigivelse oppsiktsvekkende og en indikasjon på hvordan dette temaet har rykket oppover på den politiske dagsordenen -- og blitt interessant for langt større grupper enn den begrensede kretsen som har forsøkt å skape debatt om det de siste årene.

Allerede før Brown gjorde offentlige data til en kampsak (og utnevnte Sir Tim Berners-Lee til frontfigur for satsingen), hadde britiske initiativer på feltet vakt oppmerksomhet internasjonalt. Til dels kan dette skyldes at språket gjør dem lettere for omverdenen å snappe opp enn for eksempel nederlandske eksempler. Det britiske sivilsamfunnet har stått for flere av de interessante initiativene. TheyWorkForYou.com, en database med informasjon om blant annet folkevalgte stemmegivning, drives av mySociety, der en frivillig organisasjon står bak.

Britiske medier har også fulgt opp. The Guardian har i lengre tid drevet kampanjen "Free our data", og har i senere tid fulgt opp med stadig flere eksempler på kreativ journalistisk utnyttelse av datakilder. Disse samles og publiseres fortløpende på en egen blogg på avisens nettsted, Datablog, med den forpliktende undertittelen "Facts are sacred".

I tillegg til Storbritannia skal her kort nevnes noen tendenser i USA og et lite utvalg europeiske land.

USA: Obamas åpenhetsideologi

Barack Obamas regjering har gjort mer åpenhet i offentlig forvaltning til en prioritert satsing gjennom sitt Open Government Initiative. Åpning av offentlig sektors datakilder er et av de viktigste virkemidlene.

Nettstedet data.gov ble lansert i mai 2009 for å gjøre flere datakilder tilgjengelig (og der har selvsagt den britiske regjeringen "stjålet" navnet til sin satsing). I tillegg til "rådata" i ulike formater kan interesserte her finne en katalog over rapporter og analyser av datasettene, og det er en egen oversikt over geodata. Antall datakilder var begrenset i begynnelsen, men nye blir kontinuerlig føyd til.

Britiske prinsipper for åpne offentlige data

I handlingsplanen "Putting the Frontline First: Smarter Government", presentert i desember 2009, definerer den britiske regjeringen et sett av prinsipper som skal styre håndteringen av offentlig sektors data.

Definisjon: 'Public data' are 'government-held non-personal data that are collected or generated in the course of public service delivery'.

Prinsippene:

- Public data will be published in reusable, machine-readable form
- Public data will be available and easy to find through a single easy to use online access point (www.data.gov.uk)
- Public data will be published using open standards and following the recommendations of the World Wide Web Consortium
- Any 'raw' dataset will be represented in linked data form
- More public data will be released under an open licence which enables free reuse, including commercial reuse
- Data underlying the Government's own websites will be published in reusable form for others to use
- Personal, classified, commercially sensitive and third-party data will continue to be protected.

Flere andre nettsteder med anvendelse av offentlige datakilder ble også lansert. Recovery.gov gir oversikt over pengebruken i de offentlige redningspakkene under finanskrisen, mens usaspending.gov skal gi publikum bedre oversikt over hvordan skattebetalernes midler brukes generelt.

I desember 2009 fulgte regjeringen opp med et eget direktiv der offentlige virksomheter instrueres til å oppfylle en rekke nye krav til transparens. Om datakilder het det at hver virksomhet innen 45 dager skulle finne fram til og publisere minst tre datasett av "høy verdi" på nettet, i et åpent format, og registrere datasettene på data.gov (hva som utgjør "høy verdi" defineres også nærmere i direktivet). Datasettene må ikke ha vært tilgjengelig på nett før.

Også i USA finnes etter hvert mange eksempler på både mediers og organisasjoners viderebruk av data. Sunlight-stiftelsen, som arbeider for mer åpenhet i amerikansk politikk og forvaltning, har startet tankesmien Sunlight Labs. Her driver man kampanjer for frigivelse av flere datakilder og ønsker å vise fram hvordan data kan viderebrukes til samfunnets beste. Idekonkurransen "Apps for America" ble avholdt to ganger i 2009, og sikter mot å få fram kreative eksempler på nyttig viderebruk av data.

Det er mange eksempler på journalistisk viderebruk av data. EveryBlock er nevnt; mer tradisjonell er ProPublicas oversikt over pengebruken i de offentlige stimuleringspakkene. Her har journalister arbeidet videre med data regjeringen allerede har publisert, for å skape mer oversiktlige og detaljerte framstillinger.

Nederland: Bruksanvisning for offentlige virksomheter

Grunnleggende kompetansebygging og kunnskapsdeling er nødvendig for å komme videre i arbeidet med frigivelse av offentlig sektors data, konkluderer konsulentene Ton Zijlstra og James Burke. I første halvdel av 2009 gjennomførte de et prosjekt for det nederlandske innenriksdepartementet, der målet var å flytte temaet høyere opp i bevisstheten hos medarbeidere i offentlig forvaltning.

Zijlstra og Burke endte opp med å utarbeide en nettbasert manual eller bruksanvisning for offentlig ansatte. Den er tilgjengelig (på nederlandsk) på nettstedet vrijedata.nl. I dette oppslagsverket kan man finne gjennomganger av praktiske spørsmål -- teknologiske, juridiske, organisatoriske -- som må besvares før data kan frigis. Prosessen er også visualisert i en plakatt, som vi har tatt med i rapporten (se del 6.3.). Videre ble det opprettet en wiki der tilgjengelige datakilder listes opp.

I det nederlandske prosjektet legges det stor vekt på nettverksbygging mellom medarbeidere i ulike deler av offentlig sektor. De som er opptatt av spørsmålet er foreløpig få og isolerte, så det er viktig å gjøre dem oppmerksom på hverandre og legge til rette for kunnskapsdeling, framholder Zijlstra (2009):

[S]trengthening the network and creating the conditions for forming a community of practice around those interested in opening up government data (civil servants, citizens, organizations alike) is an important aspect of bringing practice in line with the law, and making sure it becomes integrated in the cultural fabric of our government organisations.

De nederlandske erfaringene er utvilsomt relevante for miljøer som arbeider med de samme spørsmålene i andre land. Zijlstra og Burke har tatt konsekvensen av dette og startet nettstedet ourdata.eu, som skal samle erfaringer og gode eksempler fra hele Europa.

Danmark: Samarbeid om datakilder

IT- og Telestyrelsen i Danmark har satt i gang flere initiativer for å sette søkelyset på potensialet i viderebruk av offentlige data. Initiativene er samlet på portalen digitaliser.dk, og inkluderer en katalog over datakilder. I samarbeid med konsultentselskapet Devoteam samlet IT- og Telestyrelsen informasjon om 900 datakilder, som var tilgjengelig da katalogen ble lansert i september 2009. Oversikten skal oppdateres fortløpende, og det er åpnet for at alle interesserte kan føye til datakilder, ikke bare

virksomhetene selv.

For å skape mer oppmerksomhet om utnyttelse av offentlige data i nye sammenhenger, utlyste IT- og Telestyrelsen høsten 2009 idekonkurransen "Offentlige Data I Spil". Vinnerne får økonomisk støtte til å videreutvikle sine ideer til viderebruk av offentlig sektors data.

Blant interessante eksempler på viderebruk av data i Danmark er nettstedet "Hvem stemmer hvad". Her samles det inn voteringsdata fra Folketinget og informasjon om medieomtale, som så kobles sammen med biografisk informasjon om den enkelte representant. Slik kan velgerne få bedre oversikt over representantens innsats.

Begynne med begynnelsen

Selv et slikt raskt sveip over noen tendenser i andre land viser umiddelbart at mange av de samme problemstillingene er gjengangere i debatter og initiativer. Diskusjonene har en tendens til å vende tilbake til de helt grunnleggende spørsmålene: Hvor er datakildene, hva må man tenke på før de kan frigis, hva er motforestillingene, hva hindrer frigivelse?

I uformelle møter og samtaler med norske kilder i forbindelse med dette prosjektet har vi gjentatte ganger mottatt det samme budskapet -- både fra journalister og offentlig ansatte. Før vyer om bedre journalistikk, mer åpen og effektiv forvaltning og styrket verdiskaping kan brynes mot virkeligheten, må disse spørsmålene besvares og avklares.

Denne enkle erkjennelsen har også styrt prioriteringene bak kartleggingsprosjektet "Fakta først". Vi har valgt å starte med en kartlegging av de grunnleggende spørsmålene, og har valgt flere innfallsvinkler: Registrering av datakilder, en spørreundersøkelse, gjennomgang av hvordan offentlige virksomheter informerer om datakilder, intervjuer hos et utvalg virksomheter. Resultatene presenteres i neste kapittel.

3. Kartlegging av offentlige datakilder i Norge: resultater

Hvilke datakilder forvalter offentlige virksomheter, og hva hindrer at mer av disse dataene gjøres tilgjengelig for viderebruk?

Disse spørsmålene var utgangspunktet for vårt kartleggingsprosjekt. Målet var å grave fram nye fakta og stimulere til debatt og økt bevissthet om problemstillingene både blant produsenter og brukere av "innovasjonsråstoffet" offentlig sektors data. I praksis, og gitt tidsrammen på noen få måneder, valgte vi flere parallelle tilnæringsmåter.

- Informasjonsnivå: Kartlegging av hvordan statlige virksomheter informerer om datakilder på eget nettsted
- Datakilde-samling: Registrering av informasjon om så mange datakilder som mulig på en provisorisk nettside vi selv satte opp, der alle interesserte ble invitert til å bidra
- Spørreundersøkelse: Blant et utvalg statlige virksomheter om praksis og holdninger til publisering og viderebruk av data
- Kommune og fylke: Kartlegging av datakilder og holdninger i én utvalgt kommune (Bergen) og én fylkeskommune (Hordaland)

I det følgende presenterer vi resultatene fra dette kartleggingsarbeidet.

3.1 Nettsteder

Det første vi gjorde da prosjektet ble satt i gang i august 2009 var å ha et åpent, uformelt møte med medlemmer av The Norwegian Online News Association (NONA) og andre interesserte i Oslo. Her fikk vi bekreftet de første antakelsene: Det er stort behov for helt grunnleggende informasjon om hva slags datakilder som finnes rundt om i ulike offentlige virksomheter og etater. Det inspirerte til ideen om en enkel kartlegging av hvordan offentlige virksomheter informerer om datakilder på sine egne nettsteder.

En avgrensning var nødvendig. Vi tok utgangspunkt i Norge.nos liste over våre styresmakter (app.norge.no/styresmakter) og valgte å gå gjennom alle nettsteder til virksomheter innen den utøvende statsmakt: Departementene, direktorater og tilsyn, ombud, fylkesmenn, råd og utvalg. Totalt utgjør disse 125 virksomheter, som altså er fordelt på ulike nivåer og fagområder i statsforvaltningen og av svært ulik størrelse og innretning.

Informasjonen vi registrerte om nettstedene er tilgjengelig på nett (se del 7). De viktige kategoriene var disse:

- Opplyser virksomheten om egne datakilder på egen forside/hjemmeside?
- Opplyser virksomheten om egne datakilder på eget nettsted (delvis eller fullstendig)?

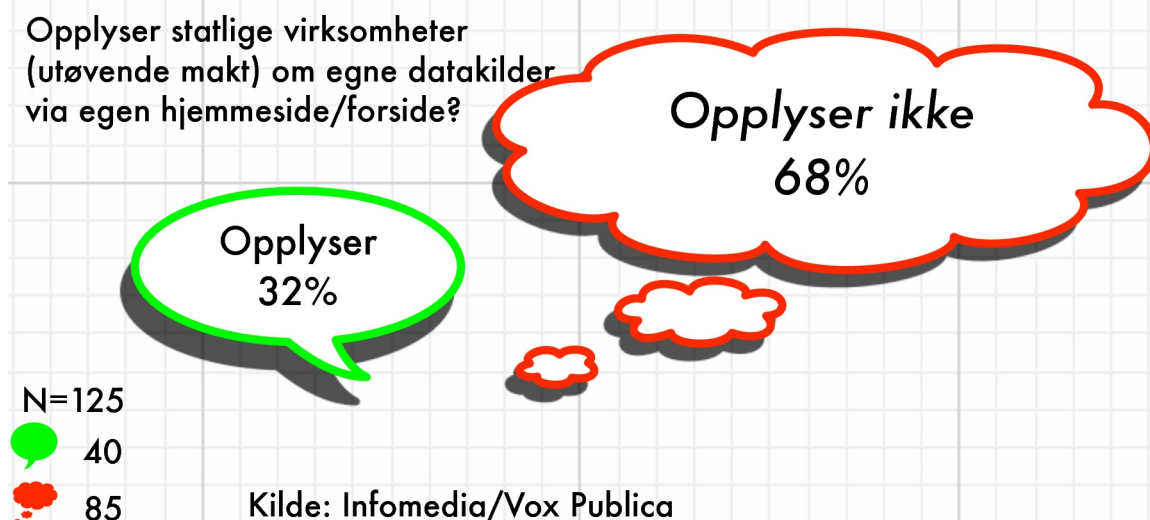
For å finne svaret på det første spørsmålet, lette vi etter bestemte nøkkelord på virksomhetens forside. Hvis ordene statistikk, data, registre, tall forekom i en eller annen kombinasjon på et fast sted på forsiden, betydde det som regel at virksomheten hadde en eller annen form for informasjon om datakilder direkte tilgjengelig fra forsiden. Vi gikk også inn på siden det ble lenket til og sjekket at det faktisk var relevant informasjon om datakilder der.

En tredjedel av virksomhetene - 40 virksomheter - informerte om datakilder fra forsiden, to tredjedeler - 85 virksomheter - gjorde det ikke (se figur 1).

Ikke forsidemateriale?

fig. 1

Opplyser statlige virksomheter (utøvende makt) om egne datakilder via egen hjemmeside/forside?



N=125

40

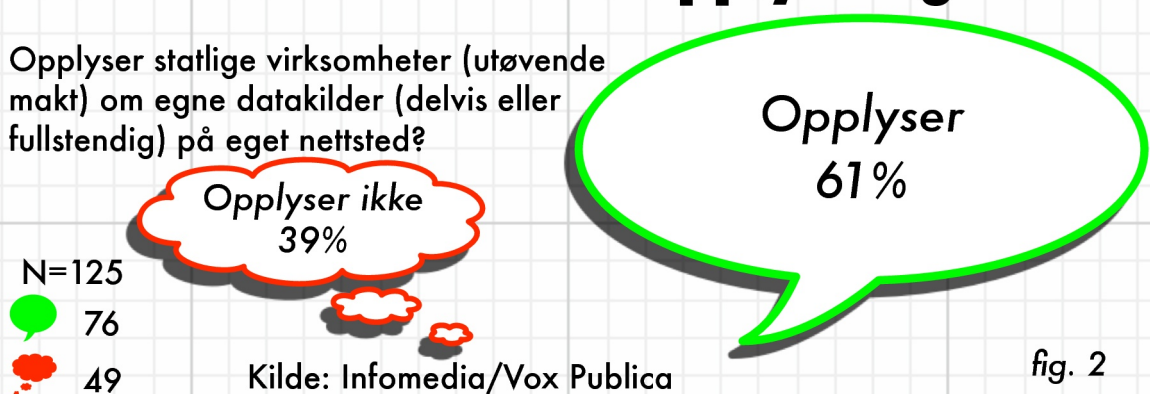
85

Kilde: Infomedia/Vox Publica

Hvis man mener at det er viktig å opplyse om datakilder, burde informasjonen være tilgjengelig fra forsiden. Men hvis virksomheter informerer på en annen side på eget nettsted, bør dette også registreres (eksempelvis finner mange brukere fram til informasjon via søkemotorer, og besøker kanskje ikke forsiden). For å undersøke om datakilder ble omtalt på nettstedet i det hele tatt, brukte vi som regel nettstedskartet som de fleste tilbyr. Her ga kartleggingen som resultat at seks av ti (76 virksomheter) på en eller annen måte informerer om datakilder, mens vi hos fire av ti (49 virksomheter) ikke kunne finne noe informasjon (se figur 2).

Finnbare datakilde-opplysninger

Opplyser statlige virksomheter (utøvende makt) om egne datakilder (delvis eller fullstendig) på eget nettsted?



N=125

76

49

Kilde: Infomedia/Vox Publica

fig. 2

Det er grunn til å bemerke at vi har vært "snille" i vurderingene. Selv om vi visste eller ante at en virksomhet har atskillig mer data enn den opplyser om, har den likevel fått "ja" hvis den hadde noe informasjon om datakilder i det hele tatt. For eksempel har mange en enkel oppstilling av statistikk fra sitt fagfelt eller ansvarsområde. Det er jo nettopp noe av poenget og problemet at så lenge virksomhetene ikke tilbyr fullstendig informasjon om sine datakilder, hindres effektiv viderebruk av data. Vi har i denne delen av kartleggingen ikke vurdert spesielt om virksomhetene tilbyr nedlasting eller på annen måte gjør datasett lett tilgjengelig.

Når resultatene fordeles på type virksomhet, viser det seg at departementene er dårligst til å informere om datakilder (vi har her holdt departementenes informasjon om statsbudsjettet utenfor). Kun tre av 19 departementer informerer om egne datakilder på forsiden av sine nettsted (Statsministerens kontor og Regjeringsadvokaten er her for enkelthets skyld regnet til departementene, selv om de i Norge.no-oversikten står under "Regjeringen"). Direktorat og tilsyn gjør det her best; hos nær halvparten av disse totalt 51 virksomhetene kan man finne informasjon om datakilder på forsiden (se figur 3).

Når vi leter etter informasjon om datakilder på hele nettstedet, kommer fylkesmannsembetene best ut. Samtlige har noe informasjon om datakilder. Det ser ut som dette skyldes en gjennomført linje om å informere om nettstedet Miljøstatus i Norge (miljostatus.no), som har miljøinformasjon og -data om hvert enkelt fylke. Alle lenker til dette nettstedet. For øvrig er det også fylkesmennene imellom stor variasjon i iveren etter å informere om datakilder. Fylkesmannen og fylkeskommunen i Østfold har for eksempel samarbeidet om nettstedet Østfold analyse, der en finner statistikk og data fra ulike kilder om fylket.

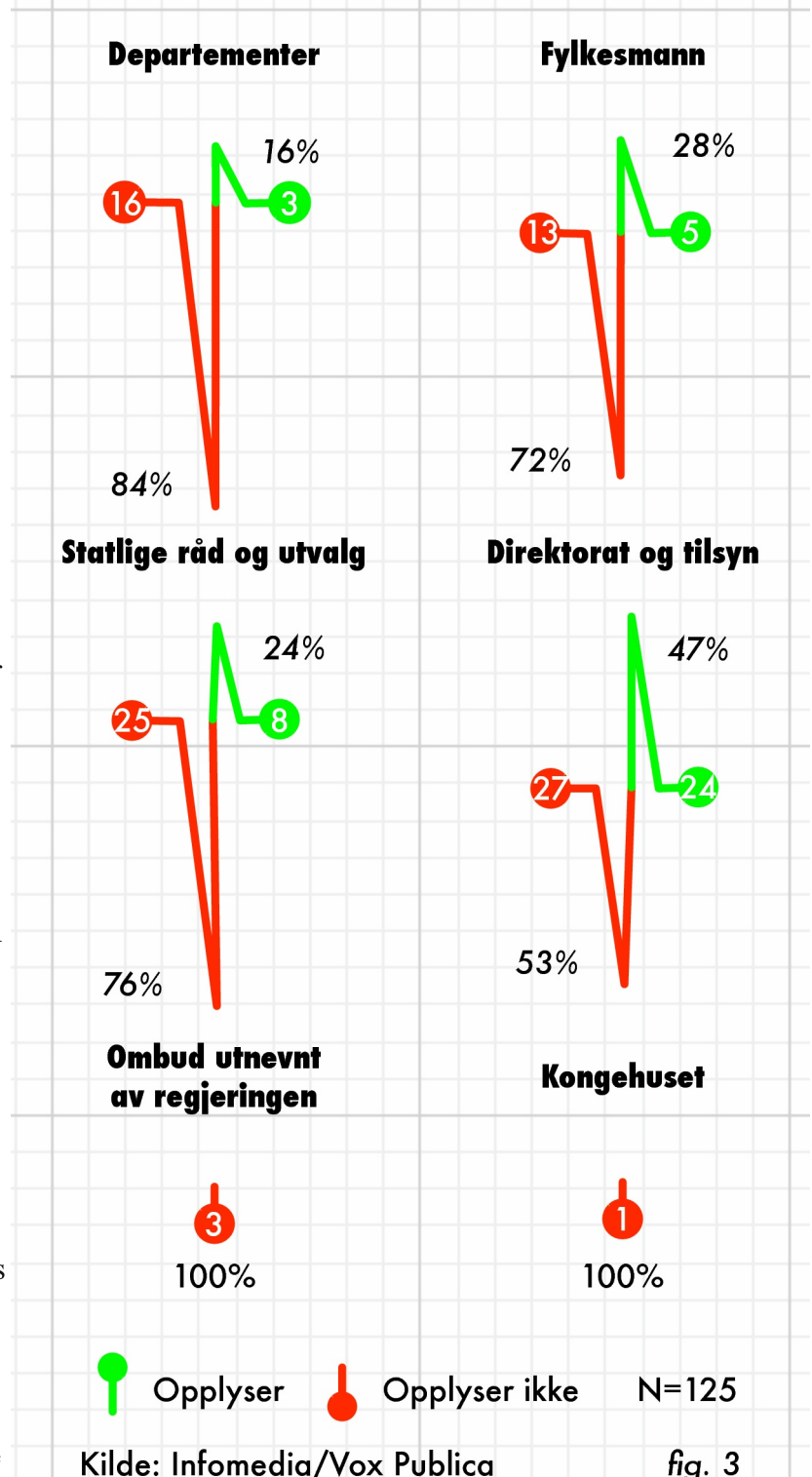
På dette spørsmålet kommer også direktorat og tilsyn bra ut, 84 prosent opplyser om datakilder på et eller annet vis (se figur 4).

Virksomhetene er som nevnt høyst ulike med hensyn til oppgaver, størrelse og nivå i statsforvaltningen. Det er å forvente at de faglig orienterte virksomhetene blant direktoratene samler inn og disponerer mye data, og at departementene i liten grad gjør dette. Det kan forklare noen av forskjellene, men samtidig er det ingen grunn til at departementene ikke skal informere om datakilder innen sektoren de styrer. Hvordan dette enkelt kan gjøres, viser Kommunal- og regionaldepartementet (KRD). På en egen side, som er lenket fra departementets forsiden, listes opp databaser og registre som KRD har ansvaret for.

Det kanskje beste eksemplet vi kom over i kartleggingsarbeidet er Statens forurensningstilsyns katalog for

Departemental datamangel

Opplyser statlige virksomheter (utøvende makt) om egne datakilder via egen hjemmeside/forside?



miljødata. Her er de enkelte datakildene registrert på egne sider. Hvert datasett er beskrevet grundig og det går fram hvordan dataene produseres og hvem som er eier. Sentralt i vår sammenheng er at mange av datasettene -- men ikke alle -- er tilgjengelig for nedlasting i ulike formater. Dette gjør jobben atskillig enklere for den som vil bruke dataene videre. Hvis SFT skal brukes som eksempel på god praksis, bør det imidlertid også påpekes at det ikke tilbys noe API (et programmeringsgrensesnitt som muliggjør kommunikasjon mellom programvare). Med et API kan data raskt, presist og enkelt utveksles (se mer om dette i del 6.2).

Andre virksomheter som utmerker seg er Utdanningsdirektoratet, Helsetilsynet og Statistisk sentralbyrå (SSB). Utdanningsdirektoratet viser direkte fra sin forside til ulike datakilder for skolen. Helsetilsynet har satt sammen en egen side med oversikt og beskrivelse av norske og internasjonale datakilder. SSB er i en særstilling som produsent og leverandør av data fra norsk forvaltning. Byrået arbeider med ny nettløsning som blant annet vil inkludere et åpent API.

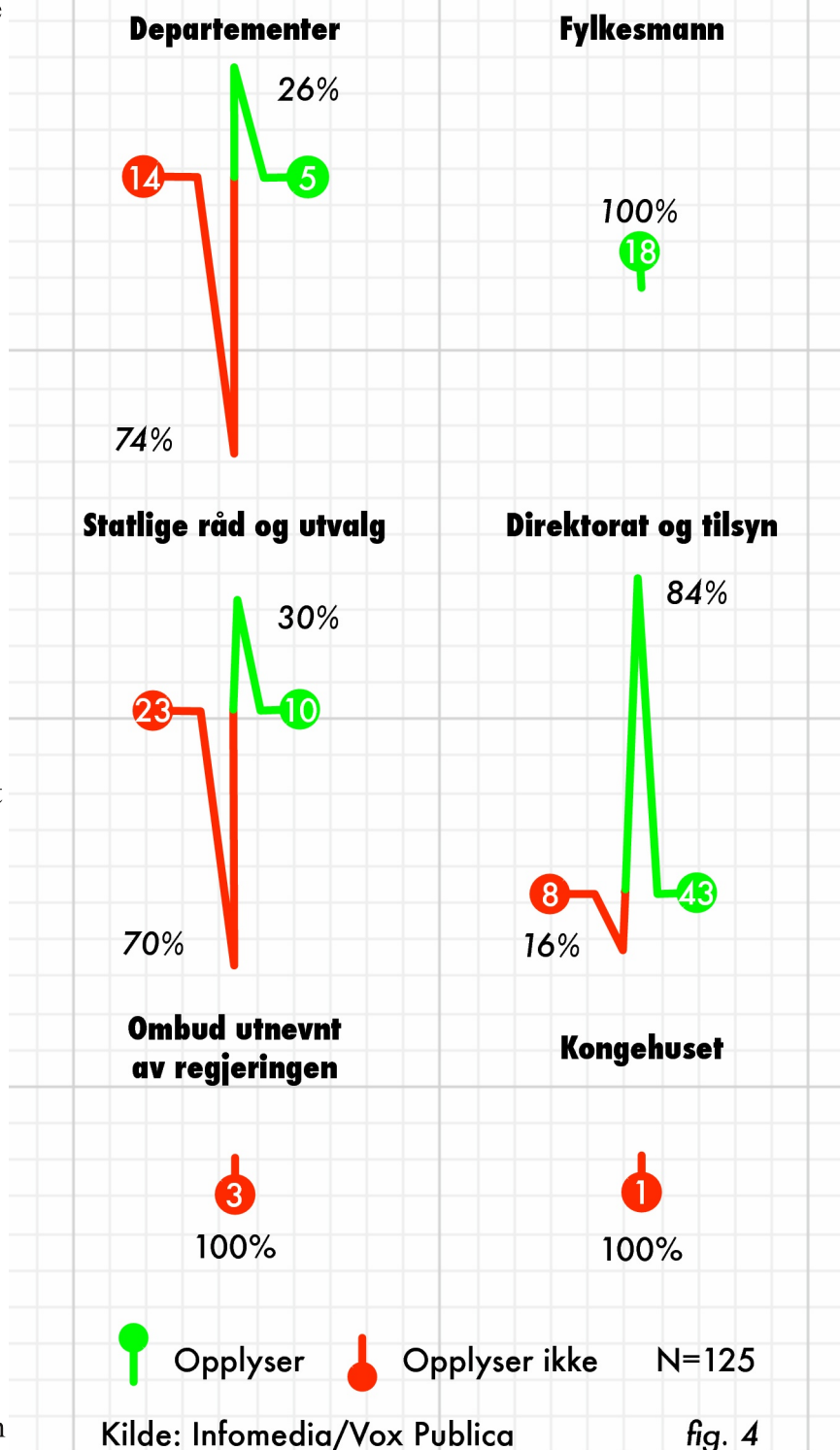
Resultatene av denne delen av kartleggingen var omtrent i tråd med antakelser vi hadde på forhånd:

- Det er ingen enhetlig praksis for å informere om datakilder
- Mange virksomheter informerer ikke i det hele tatt
- På mange nettsteder er informasjonen godt gjemt
- Det er få virkelig gode eksempler på forbilledlig informasjon om datakilder

Nettstedene innen den utøvende makt ble undersøkt med henblikk på svært grunnleggende informasjon. Resultatene viser at det er stort forbedringsbehov for de aller fleste. For eksempel burde KRDs informasjonsside om datakilder være et minstekrav på departementsnivå, men selv dette er det altså bare ett departement som kan tilby.

Koordinerte fylkesmenn

Opplyser statlige virksomheter (utøvende makt) om egne datakilder (delvis eller fullstendig) på eget nettsted?



3.2 Begynnelsen på et norsk data.gov

"Data som vi alt har betalt for", lød overskriften på et innlegg på bloggen NRKbeta 3. juli 2009. Her gikk Øyvind Solstad gjennom noen nyheter om frigivelse av datakilder i norsk forvaltning -- Avinor, Statistisk sentralbyrå og Meteorologisk institutt, blant annet -- og etterlyste data fra Statens kartverk. Solstad avsluttet med å spørre leserne:

For hvem har vi glemt? Hvilke andre data er det som noen har samlet inn men som er låst ned i en database et eller annet sted? Bruk kommentarfeltet!

I løpet av få timer strømmet det på med kommentarer og forslag, etter hvert så mange at det begynte å bli vanskelig å holde oversikten i et kronologisk ordnet kommentarfelt. Solstad nevnte også det amerikanske data.gov i innlegget, og til sammen ga dette oss viktige impulser i den første fasen av kartleggingsarbeidet. Behovet for et nettsted som data.gov er åpenbart sterkt i Norge, og for så vidt i alle land. Vi hadde ikke ressurser til å starte et helt nettsted av denne typen, og uansett må dette være en statlig oppgave (i hvert fall koordineringen av arbeidet). Men vi ønsket heller ikke å sitte og vente. I kartleggingsarbeidets ulike deler ville vi komme over datakilder, og disse ønsket vi å registrere et sted.

Google docs Kartlegging av offentlig sektors data

| | A | B | C | D | E |
|----|---------------------------|--|--|---|---------------|
| | | KARTLEGGING AV OFFENTLIGE DATA | BIDRA VED Å SKRIVE INN DATAKILDER DU KJENNER TIL OG IDEER TIL BRUK. | KONTAKT: olav.ovrebo@infomedia.uib.no | |
| 2 | Unik identifikator | Offentlig virksomhet (dataeier, datasettprodusent) | Type data (tittel) | Tilgjengelig? | Format |
| 3 | Sosialhjelp | Arbeids- og inkluderingsdepartementet / NAV, Drift og utvikling(NDU) / SSB | Sosialhjelp | | db |
| 4 | Avinor_flytider | Avinor | Data for flyavganger- og ankomster | Ja | XML |
| 5 | Barnvern | Barne- og likestillingsdepartementet / SSB | Barnvern | | db |
| 6 | Bergen_kommune_møter | Bergen kommune | Møteoversikt, saksoversikt, spørsmålsregister | Ja | html / doc |
| 7 | Bergen_kommune_politikere | Bergen kommune | Politikerregister m/verv og økonomiske interesser | Ja | html |
| 8 | Bibsys_katalogdata | Bibsys | Katalogdata | Ja | XML (?) |
| 9 | Partiregisteret | Brønnøysundregistrene | Partiregisteret | Nei (kun papir, må bestilles) | |
| 10 | | Byrådsavdeling for byutvikling, næring og klima, Bergen kommune | Planarkiv (Bergen kommunes arealplaner med bestemmelser) | Ja | html, pdf |
| 11 | Fiskeridirektoratet | Fiskeridirektoratet | Akvakulturtillatelser, Fartøyregisteret, Fiskermanntallet, Merkeregisteret, Kjøperregisteret | ja | html |
| 12 | Medisinsk_fødselsregister | Folkehelseinstituttet / HelseDirektoratet / Medisinsk fødselsregister | Medisinsk fødselsregister | kun aggregerte tall | db |
| 13 | | Fornyings- og administrasjonsdepartementet/SSB | Støtte til politiske partier | Ja, inkl. nedlasting | html, xls |
| 14 | | Fylkesarkivet, Hordaland fylkeskommune | Klientmapper institusjonsoppbygg (barnevern, Vestlandsheimen, Bergen mødre hjem) | Nei, kun ved henvendelse (doc) | |

Den enklest mulige løsningen ble valgt: Et regneark ble opprettet under Google Docs i begynnelsen av september 2009 (se ill.). Vi definerte kategorier og valgte så å gi alle som ønsket det redigeringstilgang til regnearket. Enkelte vi visste var interessert ble informert, og ellers har vi i løpet av høsten forsøkt å informere om regnearket i ulike sammenhenger -- i nettdebatter, på Twitter, og naturligvis på vår egen prosjektblogg i Vox Publica. Innen utgangen av året var ca. 130 datakilder registrert, og 20 personer hadde bidratt på ulike måter.

Et symptom på behovet for et data.gov -- og mer debatt om data-åpenhet -- er at det dukket opp andre, beslektede initiativer i løpet av vår korte prosjektperiode høsten 2009:

- Wiki: Sondre Bjellås startet i oktober wikien datakilder.no, med samme motivasjon som oss. Han har også modifisert vårt Google-regneark slik at informasjon derfra kan importeres til datakilder.no.
- Debattsoner: I oktober kom enda et initiativ. Bjørn Tennøe og Kjersti Corneliussen opprettet siden ("sonen") "Offentlig infrastruktur og søk" på Origo.no. Målet her var å skape en nettverksside "for folk som vil utvikle bedre måter å bruke offentlige data på." Blant annet ba initiativtakerne også her om tips til offentlige datakilder og hvilke data deltakerne mener burde bli gjort tilgjengelig, og mange av de samme, mest kjente datakildene gikk igjen.
- Ekspertgruppe: Teknologirådet satte i høst ned en ekspertgruppe som skal utarbeide en rapport om åpne offentlige data.
- Del av forskningsprosjekt: Forskningsprosjektet Semicolon har ambisjoner om å bidra til bedre tilgang til offentlige data. I høst utlyste de et mastergradsprosjekt der oppdraget er å utvikle en portalarkitektur "der brukere får tilgang til åpne offentlige data gjennom et webbasert grensesnitt." Forskningsprosjektet slår også fast dagens lite tilfredsstillende situasjon: "Det finnes ikke engang en systematisk oversikt over hvem som har hvilke data, langt mindre en oversikt som lenker sammen relatert informasjon fra ulike etater."

Disse ulike initiativene kan forhåpentlig bidra til å legge press på relevante myndigheter -- Fornyings- og administrasjonsdepartementet (FAD), Direktoratet for forvaltning og IKT (DIFI) -- slik at planer for et norsk data.gov hurtig blir lagt og satt ut i livet. Her begynner en heller ikke på bar bakke. Eksempelvis har Brønnøysundregistrene allerede en omfattende oversikt over "registre knyttet til oppgaveplikter", som inneholder beskrivelser og dataidentifikatorer.

Det er vanskelig å tenke seg at potensialet i viderebruk av offentlige data kan hentes ut uten at denne helt grunnleggende basisinformasjonen er på plass. En felles inngang til datakilder vil i seg selv øke bevisstheten om temaet og dermed bidra til mer fart i arbeidet med å opplyse om og frigi datakilder i de ulike delene av offentlig sektor.

3.3 Spørreundersøkelse

Hva er praksis i offentlig sektor på feltet frigivelse av datakilder, hvilke planer har virksomhetene, hva hindrer at mer data blir gjort tilgjengelig? Dette var spørsmål vi ønsket å få belyst bedre ved hjelp av en spørreundersøkelse. Vi valgte å henvende oss til et utvalg statlige virksomheter innenfor ulike temaområder. Etter noen innledende sonderinger bestemte vi oss for å holde forvaltningsnivåene fylkeskommune og kommune utenfor denne spørreundersøkelsen, og istedenfor foreta en grundigere kartlegging av én utvalgt kommune og én fylkeskommune (se punkt 3.4).

I spørreundersøkelsen ønsket vi å dekke spekteret av fagområder i statsforvaltningen. Vi anså det som viktig at alle relevante fagfelt var inkludert, siden vi antok at vurderinger og praksis kan være ulike på ulike fagområder. Samtidig sørget vi for at forskjellige typer virksomheter var representert i utvalget (direktorat, tilsyn, forskningsinstitusjon, registre, statsbedrifter). Inndelingen i fagfelt voldt en del hodebry. Trolig kommer man ikke utenom overlappende kategorier, uansett valg av inndeling. Etter å ha vurdert inndelingen gjort i en rapport fra Statskonsult (2004) -- det eneste

Datakilder på 11 fagområder

Utvalget av virksomheter til spørreundersøkelsen ble gjort med utgangspunkt i disse fagområdene:

- Geografisk informasjon
- Miljø- og ressursinformasjon
- Næringsliv og økonomi
- Informasjon om rettslige handlinger (rettsstiftelser)
- Helse og sosial
- Meteorologi
- Samfunnsdata
- Politi, kriminalitet, sikkerhet
- Forskning
- Samferdsel og kommunikasjon
- Kultur

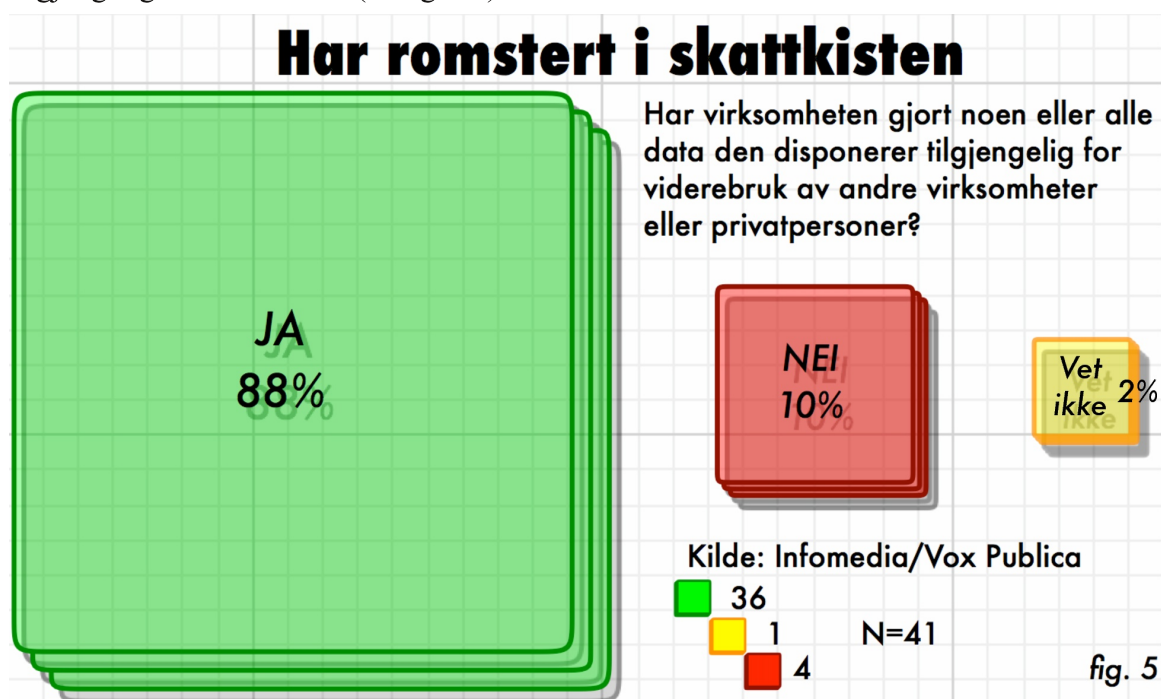
større kartleggingsarbeidet på feltet av nyere dato -- og fått verdifulle innspill fra DIFI, falt vi ned på 11 fagområder (se presentasjon i egen ramme).

Undersøkelsen ble utført via et web-basert spørreskjema i perioden 15. oktober-15. november 2009. Vi la en del arbeid i henvendelsen til virksomhetene for å øke sjansene for å få godt informerte svar. Vi tok kontakt, forklarte bakgrunnen for prosjektet og lot det være opp til virksomheten selv å avgjøre hvilken medarbeider som var best egnet til å svare på spørsmålene. Totalt ble 67 spørreskjemaer sendt ut. Vi mottok 41 svar, en svarprosent på 61.

Frigivelse av data: praksis og planer

Vi ønsket å skaffe oss kunnskap om hvordan virksomhetene arbeider med å frigi datakilder i dag, og hvilke planer de har i nærmeste framtid.

En svært høy andel av de som svarte sier at virksomheten deres gjør noen eller alle data den disponerer tilgjengelig for viderebruk (se figur 5).



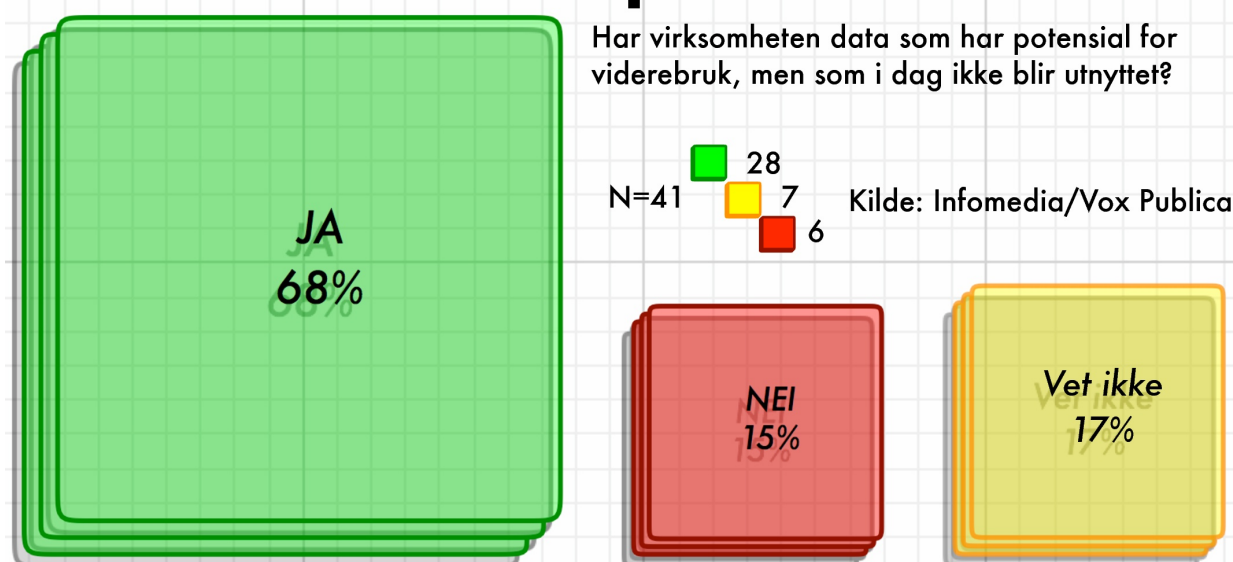
Det er imidlertid verdt å merke seg her at svaralternativene ikke differensierer mellom frigivelse av data i ulike formater. Slik sett kan dette gi et for positivt inntrykk. Svarene her kan dekke tilgjengeliggjøring i mange varianter som i svært ulik grad er nyttige for dem som vil bruke dem. Det er ikke det samme om dataene presenteres i søkbare databaser, enkle talloppstillinger i pdf-dokumenter, nedlasting i ulike formater eller via tilrettelegging med et API. Dessuten kan virksomhetene selvsagt ha mye interessante data som de ikke har gjort tilgjengelig selv om de svarte ja her.

Statskonsult spurte i sin undersøkelse fra 2004 om virksomheten hadde "strukturert registerinformasjon som har potensial for gjenbruk, men som i dag ikke blir utnyttet." To av tre svarte bekreftende på dette (Statskonsult 2004, s. 16). Vi gjentok dette spørsmålet i vår undersøkelse med omtrent samme formulering. Selv om de to undersøkelsene ikke kan sammenlignes direkte, er det i det minste interessant å notere at vi fikk nøyaktig samme svarfordeling (se figur 6).

Dette synes å støtte vurderingen av svarene gjengitt i figur 5 ovenfor. Selv om man har gjort noe data tilgjengelig, er det svært sannsynlig at virksomheten har mer data med potensial for viderebruk. Her kom det også inn mange fritekstsvar som gir utdypende informasjon om hvordan virksomhetene tenker (se presentasjon i egen ramme).

Det er mer på kistebunnen

fig. 6



For å få et inntrykk av planene framover, spurte vi om virksomhetene vil gjøre mer data tilgjengelig det kommende året (se figur 7).

Det synes altså som om mange virksomheter allerede arbeider med planer om å gjøre flere datakilder tilgjengelig. Interessant i så måte var Statens kartverks frigivelse av karttjenester fra 1. desember 2009. Denne nyheten kom samtidig med at vi avsluttet spørreundersøkelsen.

Verdiøkning, kapasitet, innsynsrett...

Har virksomheten data som har potensial for viderebruk, men som i dag ikke blir utnyttet? Utvalg av utdypende svar avgitt i fritekst:

"Dataene har i dag et stort verdiøkende potensiale som dels er utnyttet."

"Det meste av dataene er tilgjengelig, men de er ikke alltid like godt tilrettelagt for folk uten geofaglig bakgrunn."

"Det er et potensiale for mer utnyttelse av etablerte tjenester, selv om utnyttelsen og viderebruken er meget omfattende."

"Instituttet sitter på store samlinger av historiske data og lange tidsserier. Tilgjengeligheten til disse er ofte begrenset fordi vi ikke har hatt tilstrekkelig kapasitet til å organisere og dokumentere datasettene. Dette er et prosjekt under arbeid."

"Vi ønsker å gjøre mer ut av dataene våre selv, skaffe bedre tilgang via våre hjemmesider."

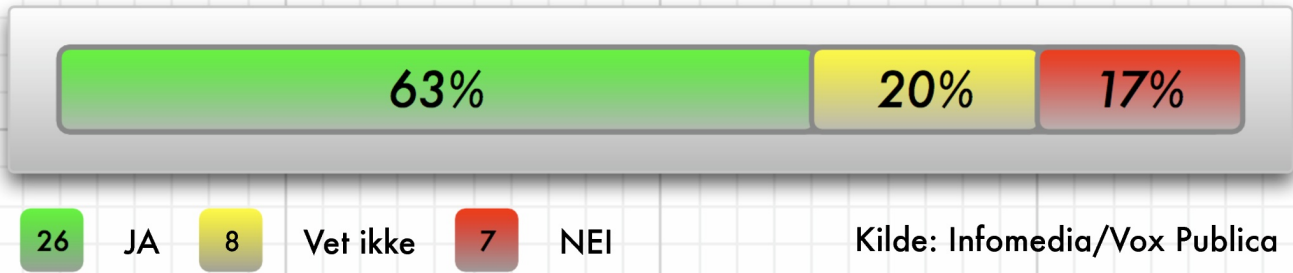
"Alle våre data er frie. Det er sikkert flere muligheter for å utnytte dem."

"Aktører bruker innsynsretten for å skaffe seg data som publiseres kommersielt."

Mer data til viderebrukerfolket

fig. 7

Planlegger din virksomhet å gjøre flere data tilgjengelig for viderebruk i løpet av det neste året?



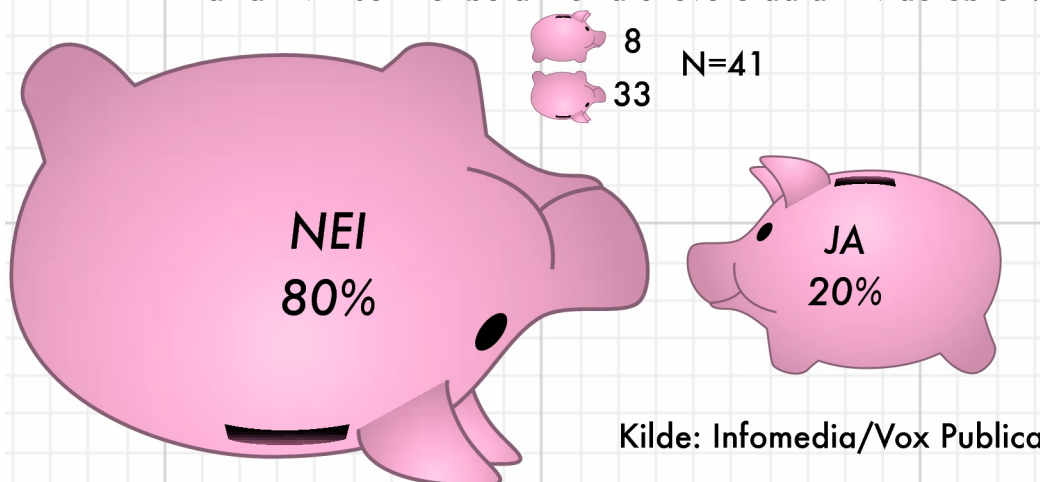
Spørsmålet om offentlige virksomheter skal ta seg betalt for utlevering av data, er en gjenganger i debattene på feltet. For miljøer som vil bruke data til å skape nye tjenester er dette ofte det viktigste hinderet og en hyppig kilde til strid. Sentrale data-innsamlere og -eiere som Statens kartverk (direkte og ved videresalg ved Norsk Eiendomsinformasjon) og Brønnøysundregistrene har omstridte betalingsmodeller. Derfor var det interessant for oss å undersøke virksomhetenes praksis og planer (se figur 8 og 9).

Betaling for data er altså et ikke-tema for det store flertallet av virksomheter i vår undersøkelse i dag, og det er også få som planlegger å ta betalt. Vi kan ikke ut fra dette konkludere at det er en trend bort fra betaling for utlevering av data. Det er imidlertid fristende å se svarene i sammenheng med økningen i interesse for potensielle positive sider ved å frigi data. Meteorologisk Institutt er kåret til statsetaten med best omdømme fire år på rad, og vant i år også Fyrlyktprisen, som deles ut av Statens dataforum. Juryen begrunnet valget blant annet med at yr.no går foran i arbeidet med å gjøre offentlige data tilgjengelig.

Det meste gratis

fig. 8

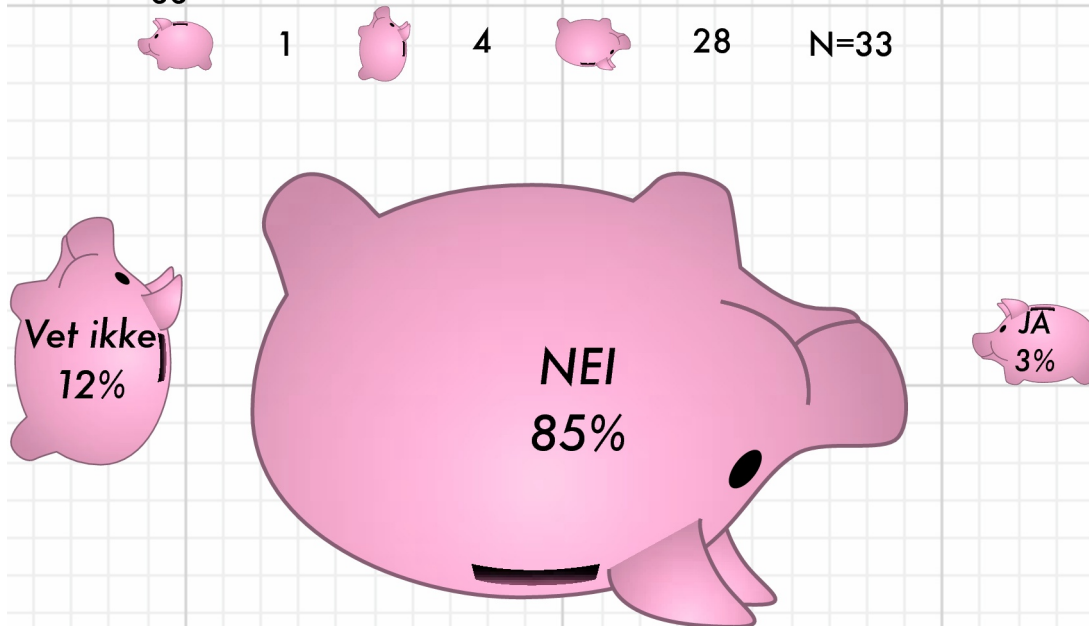
Tar din virksomhet betalt for å utlevere data til viderebruk?



Ingen budsjettredder

fig. 9

Planlegger din virksomhet å ta betalt for å utlevere data?

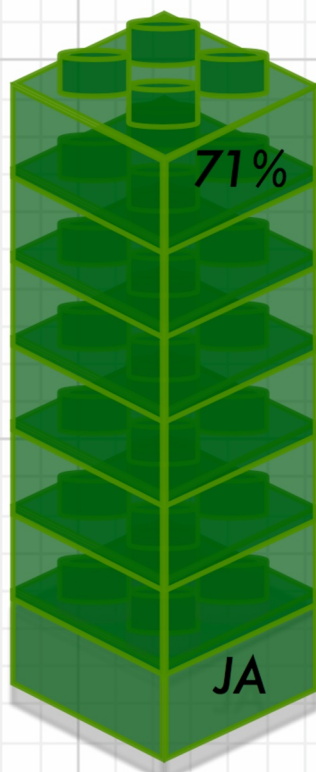


Kilde: Infomedia/Vox Publica

Det er ikke umulig å tenke seg at andre offentlige virksomheter trekker lærdom av meteorologenes suksess. Men hvilken lærdom? Meteorologisk Institutt har gjort to ting samtidig: Frigitt data og lansert en populær ny publikumstjeneste sammen med NRK. Det er en viss bekymring å spore blant journalister og andre for at offentlige virksomheter først og fremst vil lære at det gjelder å bruke ressursene på å utvikle egne tjenester rettet mot publikum, ikke frigi selve datagrunnlaget. Det var bakgrunnen for dette spørsmålet (se figur 10).

En kan ikke trekke for mye ut av dette svaret. Det er vanskelig å stille spørsmålet slik at det presist fanger opp resonnetet i avsnittet over. Et oppfølgingsspørsmål om virksomhetene har planer om å utvikle egne nett-tjenester ble besvart av så få at det ikke ga nyttige indikasjoner. Problemstillingen må uansett diskuteres når frigivelse av data skal vurderes på politisk hold og i virksomhetene framover.

Data best på hjemmebane?



Har din virksomhet utviklet egne nett-tjenester for å presentere egne data for offentligheten?



Kilde: Infomedia/Vox Publica

fig. 10

Frigivelse av data: hindringer

Den nye offentlighetsloven gjør det lettere for allmennheten å kreve data utlevert, men hvis alle data må hales ut fra virksomhetene, vil det gå lang tid før de ønskede positive effektene av viderebruk viser seg. Aktiv tilrettelegging fra offentlige virksomheters side er en bedre løsning.

Frigivelse av data krever kompetanse og handling fra virksomhetenes side. Først må interessen for å gjøre noe på feltet være til stede, så må datasett kvalitetssikres, man må finne hensiktsmessige tekniske løsninger -- blant annet. Innledende samtaler med aktører både i ulike virksomheter og blant potensielle "viderebrukere" ga oss noen indikasjoner på hvor de viktigste hindringene er, og disse ville vi teste i spørreundersøkelsen (se figur 11 og påfølgende side).

Det koster å være åpen

fig. 11

Nedenfor har vi listet opp noen mulige hindre for at virksomheter i offentlig sektor kan gjøre data tilgjengelig for viderebruk. Vennligst kryss av de du mener er relevante (flere kryss mulig) og/eller fyll inn eget svar.

Tilgjengeliggjøring vil føre til tap av inntekter

4

14%

Private aktører bør ikke tjene penger på data offentlige virksomheter produserer

4

14%

Private aktører kan misbruke dataene til å spre sensitive personopplysninger

8

29%

Private aktører kan misforstå dataene og slik spre villredende informasjon

12

43%

Det koster for mye å gjøre dataene tilgjengelig

18

64%

Annet: Vennligst skriv inn din kommentar

19

68%

N= 28

Kilde: Infomedia/Vox Publica

Hindre og muligheter

Fritekst-svar på spørsmål (se figur 11) om hindre mot at virksomheter i offentlig sektor gjør data tilgjengelig for viderebruk. Enkelte av svarene er noe redigert/forkortet for å bevare respondentenes anonymitet.

"Generelt mener jeg at offentlige data bør være tilgjengelige for viderebruk. Vi ser ingen hindringer hos oss. Inntektstapet for oss var lite."

"Taushetsbelagt informasjon, samarbeid med private aktører og forretningshemmeligheter."

"Informasjon har alltid en fare ved seg at den kan brukes i utilsiktet sammenheng"

"- manglende endringsevne (eks ressurser, kompetanse, osv)
- manglende mål og krav til videreutvikling av tjenester"

"[Direktoratets] data er ikke egnet for viderebruk utenfor etaten."

"Det er politisk vedtatt at vi kan ta betalt for våre data."

"På de fagområdene [virksomheten] arbeider bør ikke de foran nevnte argumentene være så tungtveiende at de hindrer viderebruk av [...] data."

"Ingen av ovenstående gjelder. De fleste punktene kan inverteres. Vi anser det å gjøre dataene tilgjengelige for viderebruk som en effektiv distribusjonsform og en viktig del av vårt totale distribusjonsapparat. Effektiv distribusjon via andre sikrer at aktørene har riktig og oppdatert informasjon. Noen tjenester er gratis og andre ikke. Vurdering av betaling eller ikke gjøres ut fra en totalvurdering av effektivitet i distribusjon, finansiering av distribusjonstjenester og bidrag til statskassen for dekning av selvkost for [...] virksomheten. I den grad informasjon som kan betraktes som sensitiv distribueres gjøres dette via aktører som oppfyller visse vilkår i hht egne avtaler og offentlig regelverk."

"Delvis aktuell. Det må fremgå av årsplan og VP, dvs budsjetteres. Det koster å produsere sikre data som må kvalitetssikres før de gis ut av hensyn til liv og helse."

"Angående betaling for videreformidling av data: I en del tilfeller gjøres dette, i samsvar med føringer fra vårt eierdepartement. [...] Forskningsdata tar vi ikke betalt for, men her kan det være begrensninger på tilgjengeligheten som springer ut av forskerrettigheter. "Det koster for mye" betyr ikke at kostnadene nødvendigvis forhindrer tilgjengeliggjøring, men heller at det tar lengre tid enn ønskelig."

"Kostnadsspørsmål er det viktigste hinderet. I tillegg må vi passe på at vi ikke overtrer lovverk, eks. personvern. Vi er omfattet av Lov om rett til miljøinformasjon og deltakelse i offentlige beslutningsprosesser av betydning for miljøet (miljøinformasjonsloven). Det medfører at vi ikke har anledning til å ta betalt for utlevering av data. Vi ser det slik at evt. opplysninger som kan misbrukes til å krenke f.eks. personvern, er opplysninger vi ikke har anledning til å publisere fritt i utgangspunktet."

"Politiske beslutninger om at data ikke skal tilgjengeliggjøres for viderebruk. Eventuelle lovverk som stopper tilgjengeliggjøring."

Kostnadene ved å frigi data trekkes altså fram av over seks av ti virksomheter som har besvart dette spørsmålet. Dette resultatet bør inspirere til nærmere undersøkelser og initiativer fra f.eks. DIFI og FAD. Hvor reell er bekymringen? Enkelte kan frykte økte utgifter til servere og annet IT-utstyr hvis datakildene de frigir blir svært populære, men med de sterkt fallende prisene på lagringskapasitet burde ikke dette være en uoverkommelig tilleggsutgift for særlig mange. Hvis økte kostnader gjelder antatt behov for å styrke kompetansen internt, kan dette kanskje avhjelpes ved bedre tilrettelagt informasjon om hva som faktisk kreves av forarbeid og vedlikehold når datakilder gjøres tilgjengelig. I et nederlandsk prosjekt om offentlige data ble nettopp kompetanse- og nettverksbygging blant dem som jobber med temaet i offentlig sektor vektlagt sterkest.

Et av svarene under "Annet" her nyanserer nettopp kostnadsspørsmålet. Når en velger svaralternativet "Det koster for mye" betyr ikke det at kostnadene i seg selv forhindrer å gjøre data tilgjengelig, "men heller at det tar lengre tid enn ønskelig," skriver denne virksomhetens representant.

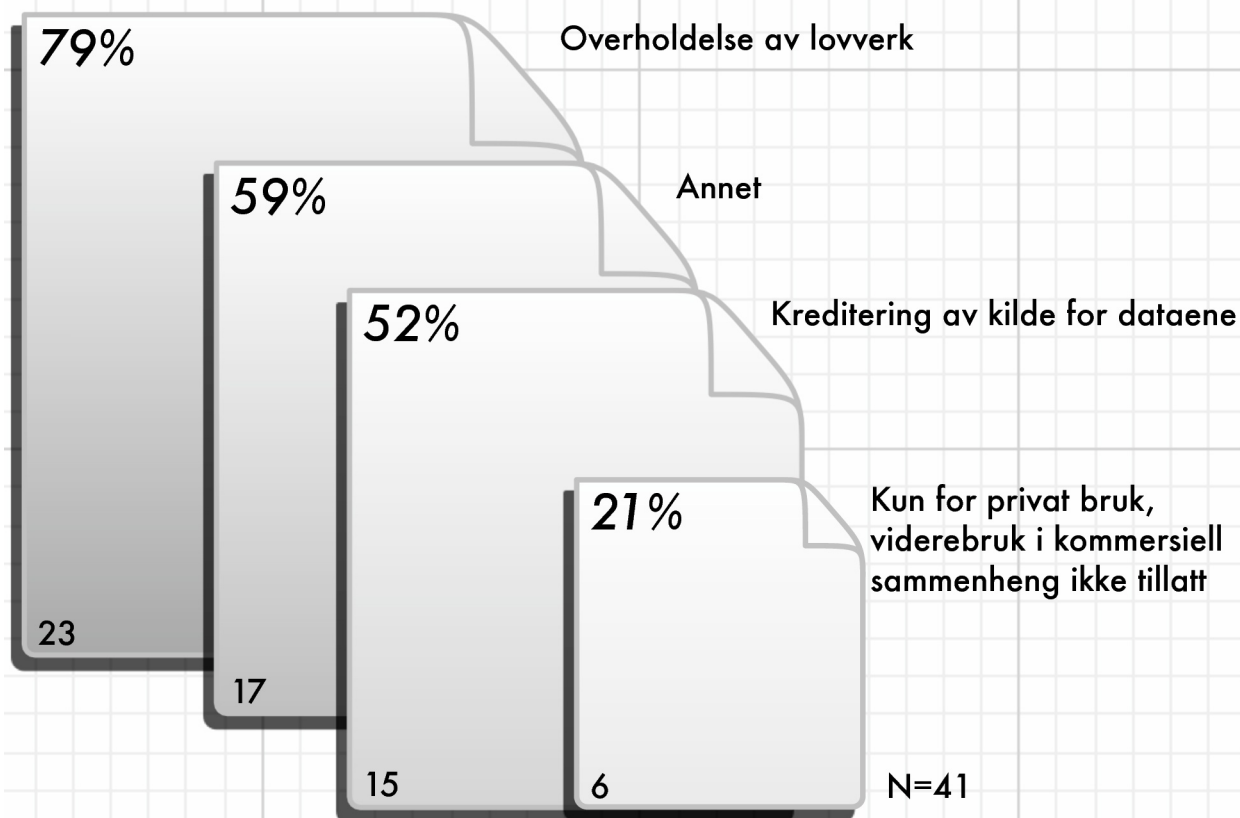
En hindring av en helt annen karakter er det nest mest valgte alternativet om faren for at private aktører misforstår dataene og slik sprer villedende informasjon (varianter av dette nevnes også i fritekst-svarene). Også i samtaler vi har hatt trekkes dette fram som en vanlig innvending. Her venter en viktig og prinsipiell debatt som alle interesserte aktører bør engasjere seg i. Et par synspunkter kan her kort skisseres: På den ene siden er det åpenbart riktig at datakilder kan komme til å bli feiltolket og satt inn i urimelige sammenhenger når data frigis. Ja, dette vil helt sikkert skje. På den annen side er det et vanskelig prinsipielt standpunkt å innta i dag at offentligheten ikke er moden for å behandle offentlig sektors datakilder på en seriøs måte (vi vender tilbake til denne diskusjonen i rapportens del 4).

Hensynet til personopplysninger - personvern - velges her av relativt få respondenter. Dette kan ha å gjøre med at hinderet er opplagt. Det er lovfestet, og det er neppe mange som mener at det bør avskaffes.

Varierende betingelser

fig. 12

Hvilke eventuelle betingelser er knyttet til viderebruk av dataene?



Kilde: Infomedia/Vox Publica

I kartleggingsarbeidet har vi ellers fått et inntrykk av at det er en del usikkerhet rundt de begrensninger personvernkrav faktisk legger på frigivelse av data.

Et til av spørsmålene i undersøkelsen kan bidra til å belyse hindre mot frigivelse av data. Vi spurte om hvilke betingelser virksomhetene eventuelt knytter til viderebruk av data (se figur 12).

Delvis basert på observasjon av hva virksomheter opplyser om på egne nettsteder, hadde vi en antakelse om at praksis varierer sterkt. Dette synes å bli bekreftet av svarene her. Eksplicit krav om overholdelse av lovverk er altså det som velges hyppigst blant våre forhåndsformulerte svaralternativer. Noen krever kreditering av kilde for data, andre setter sperrer for kommersiell viderebruk. Øvrige varianter sitert fra fritekstsvarene under "Annet":

- "Data til forskning har ytterligere krav iht. avtale"
- "Offentlig tilgjengelig på samme vilkår som annen offentlig informasjon."
- "Ovenstående gjelder for en del tjenester ssom krever egen avtale, men det er også flere tjenester hvor slike betingelser ikke gjelder."
- "Informasjon og data fra [datakilde] er enerettslig beskyttet etter Åndsverkloven § 43 første ledd."
- "Det forventes nok at kilde oppgis, men vet ikke om det er noe "system" på dette."

Svar-mangfoldet indikerer behov for en opprydning. Generelt er det selvsagt ønskelig med enkle, klare og mest mulig enhetlige betingelser. Større private bedrifter som har behov for viderebruk av data kan nok forventes å ha ressurser til å vurdere ulike betingelser for viderebruk. Men utviklingen i dag går mot at kunnskapsrike privatpersoner, frivillige organisasjoner og løse interessesammenslutninger også vil bli viktige viderebrukere av data. I denne situasjonen vil sterkt varierende betingelser begrense viderebruken unødige.

Oppsummering

Spørreundersøkelsen gir et differensiert inntrykk. Totalt sett virker det som mange av virksomhetene som har svart har et gjennomtenkt forhold til temaet frigivelse av data og at de aktivt vurderer å frigi mer data. Det er mulig at undersøkelsen blir litt skjev her, fordi mange faglig tunge virksomheter er spurt. Det er grunn til å tro at disse i større grad har vurdert spørsmål om data-forvaltning og -frigivelse enn andre typer virksomheter. Likevel indikerer undersøkelsen at det er interesse for temaet og mye kompetanse å bygge videre på ute i virksomhetene. Nettverksbygging for å dra nytte av denne kompetansen på tvers av virksomhetene samt oppklaring av juridiske forhold knyttet til betingelser for viderebruk av data, peker seg ut som effektive tiltak (se anbefalinger i rapportens del 5).

3.4 Kommune og fylkeskommune: Bergen og Hordaland

De kommunale og fylkeskommunale forvaltningsnivåene er meget interessante som kilder for offentlige data. På begge nivåer samles det inn mye data av svært forskjellig karakter. Siden det er her mye av de offentlige tjenestene produseres, genereres det data som kan være relevante for viderebruk.

Innledende samtaler gjorde det klart at en grundig kartlegging av kommunale og fylkeskommunale datakilder er en meget stor oppgave. Gitt tiden og ressursene vi hadde til rådighet, var det klart at en kartlegging av alle eller et stort utvalg kommuner og fylkeskommuner var uaktuelt. Kommuner og fylkeskommuner spiller på et vis det statlige nivået når det gjelder mangelen på sentral oversikt over datakilder. Innen hver kommune er det lite sannsynlig at man finner ett kontor som kan levere ut en oversikt over kommunens datakilder. En spørreundersøkelse ville fort strande allerede her.

På den annen side er kommuner og fylkeskommuner strukturelt like og har gjerne de samme typene datakilder - de små kommunene vil skille seg ut ved å ha færre datakilder. Vi bestemte oss derfor for å gå i dybden på ett "eksemplar" på hvert forvaltningsnivå -- og det burde være store enheter. Vi antok at vi

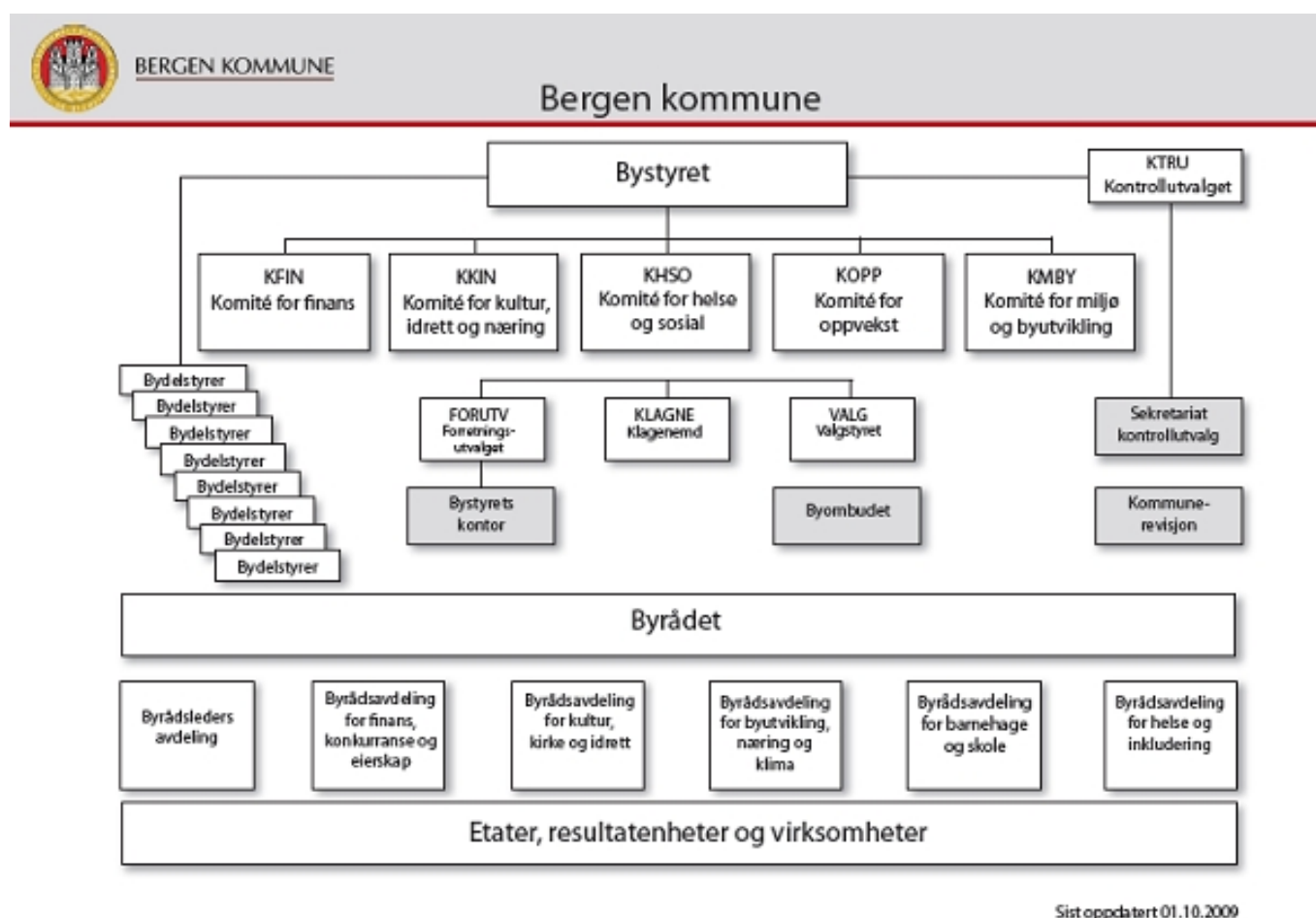
derved ville få fram mange problemstillinger som vil være felles for kommunal og fylkeskommunal sektor som helhet. Geografisk nærhet gjorde det naturlig å velge Bergen kommune og Hordaland fylkeskommune.

Vi søkte informasjon om disse hovedpunktene:

- Etatens datakilder
- Vurderinger av tilgjengeligjering av data, eventuelle hindre
- Etatens eventuelle planer om å gjøre mer data tilgjengelig

Bergen kommune

Bergen kommune er organisert etter en parlamentarisk modell. Bystyrerepresentantene velges ved kommunevalg. Byrådet på sin side er kommunens ”regjering”, ansvarlig overfor bystyret. Byrådet er organisert i seks byrådsavdelinger, som igjen består av egne etater og avdelinger med ansvar for sine respektive fagområder (se ramme med organisasjonskart).



Kommunens etater og virksomheter er organisert i seks byrådsavdelinger.

I tillegg til oppgavene som er lagt til byrådsavdelingene er enkelte oppgaver og funksjoner organisert i kommunale foretak (KF) eller i kommunale/interkommunale selskaper. Enkelte av disse igjen er organisert som aksjeselskaper.

Bergen kommune har rundt 15 000 ansatte, noe som gjør kommunen til byens største arbeidsgiver. Byrådets budsjettforslag for 2010 gir kommunen ca. 14,1 milliarder kroner til drift. I tillegg kommer et eget investeringsbudsjett på 2,25 milliarder kroner.

Kartleggingsarbeidet

Innledende samtaler hadde vi med representanter fra Byarkivet og Portaltjenesten. Møtet med Byarkivet begynte å gi oss en pekepinn på hvilke mengder mulige datakilder det var snakk om, med utdrag fra en lang liste over noen av datasystemene som har vært brukt, og brukes, i kommunen. Dette inntrykket ble bekreftet av Portaltjenesten, som kunne fortelle at forrige gjennomgang av kommunens fagsystemer viste at det fins rundt 400 av disse i Bergen kommune.

Dessverre fikk vi ikke tilgang til denne oversikten. Dermed måtte vi henvende oss direkte til etater, avdelinger og virksomheter. Bergen kommunes størrelse tatt i betraktning, konsentrerte vi oss om et utvalg potensielt interessante etater og avdelinger. I løpet av kartleggingsarbeidet var vi i kontakt med flere etater og avdelinger enn de som beskrives under, men disse er utelatt fordi vi av ulike årsaker ikke fikk den informasjonen vi var ute etter. Vi har fått materiale fra etater som sorterer under tre av de seks byrådsavdelingene, og i tillegg fra et av bystyrets organer og et interkommunalt selskap.

Bystyrets kontor

Bystyrets kontor er et av fire kontorer i bystyrets administrasjon. De øvrige er Byombudet, Sekretariat for kontrollutvalget og Kommunerevisjonen. Alle er direkte underlagt ordføreren og bystyret.

Bystyrets kontor har ansvar for tilrettelegging og avvikling av møter i bystyret, komiteer, bydelsstyrer og andre utvalg som hører inn under bystyret. Saksflyt og kommunikasjon mellom byrådet og bystyrets organer er også blant kontorets hovedoppgaver. Videre har det rådgivnings- og informasjonansvar overfor folkevalgte, administrasjon, presse og publikum.

Datakilder

Bystyret har per i dag to registre som allerede ligger ute på nettsidene: Et politikerregister og et register for politiske saker og spørsmål. Begge disse registrene har eksistert siden 2002/2003.

Politiske saker og spørsmålsregister

I dette registeret finnes referater fra møter i bystyret, saker, vedtak som gjøres og oversikt over spørsmål fra representanter. Det fins også noe voteringsdata, men disse er ikke spesifisert på representanter og registreres ikke systematisk. Protokollene og saksdokumentene finnes i pdf-format.

I saksregisteret kan man kun søke på sakstitler, og det eksisterer så langt ingen muligheter for fritekstsøk inne i selve dokumentet. Dette gjør det for eksempel vrient å finne ut raskt hvilke spørsmål én enkelt representant har stilt på et bestemt bystyremøte. Problematikken rundt manglende fritekstsøk er noe som diskuteres på Bystyrets kontor. Det har blant annet vært snakket om at for eksempel sakstitlene kan formuleres bedre, og slik øke søkbarheten. Hvis man går inn på en enkeltsak framfor protokollen fra et bystyremøte, får man imidlertid opp mer informasjon. I bystyreprotokollen refereres kun det som foregår i bystyret, men hvis en sak har blitt behandlet av en komité, kan man enten finne det utvidete referatet via å søke opp komitémøtet, eller ved å gå inn på den enkelte saken.

Samtidig protokollføres alle spørsmål som stilles i egne dokumenter, slik at disse er søkbare på politikernavn i spørsmålsregisteret. Ordvekslinger, det vil si innlegg og motinnlegg under bystyremøtene som ikke betegnes som spørsmål, protokollføres ikke.

Politikerregisteret

Bergen kommunes politikerregister inneholder informasjon om folkevalgte i Bergen (byrådet, bystyret, bydelsstyrer, utvalg og komiteer). Hver politiker har en egen side hvor følgende informasjon er registrert: Nåværende og tidligere kommunale verv, økonomiske interesser (eierinteresser, styreverv), kontaktopplysninger. Enkelte opplysninger registreres kun for medlemmer i byrådet og bystyret.

Dette registeret viser kun politikere med aktive kommunale verv. Det vil si at for eksempel Herman Friele fortsatt er i registeret, mens Henning Warloe og andre som nettopp har blitt valgt inn på Stortinget har blitt fjernet fra registeret.

Henvendelser angående dette registeret går veldig ofte på epost-lister, fra folk som ønsker å få utlevert fullstendige lister over politikernes epost-adresser. Man kan klikke seg inn på hver enkelt politiker for å finne epost-adresse, men for å få utlevert samlede lister må man henvende seg til Bystyrets kontor som vurderer forespørselen.

Andre, ikke offentlige registre

Mye av informasjonen Bystyret sitter på finnes allerede i de to nevnte registrene. Det ser heller ut til at utfordringene ligger i presentasjonen eller struktureringen av disse, slik at man lettere kan utnytte informasjonen i dokumentene.

Samtidig opererer administrasjonen med egne, private såkalte ”utvalg”, som kodes på en spesiell måte. Her registreres hvem som melder eller ber om frafall/fritak og permisjon fra bystyremøtene. Denne registreringen fungerer som en slags kladdebok for konsulentene, med informasjon som de bruker igjen til årsmeldingen for bystyrets organer. Dataene offentliggjøres ikke, først og fremst av hensyn til personvern, samtidig som de ville blitt frigitt ved en eventuell forespørsel – for eksempel fra en journalist.

Bystyrets kontor bruker også et system fra firmaet ErgoGroup i forbindelse med kommunevalg. Rådataene herfra sendes imidlertid direkte til Statistisk sentralbyrå som behandler dem, før statistikken sendes tilbake til kommunen.

Vurderinger og planer

Det har vært snakket om å integrere private forslag og innbyggerinitiativ i spørsmålsregisteret, eller eventuelt å opprette et eget register for disse. Per i dag registreres disse kun ved at ordene brukes i sakstittel. Det jobbes også med å få i gang både et voteringsanlegg som gjør at man kan se hvem som stemmer hvordan i de enkelte sakene, og videooverføring av bystyremøtene.

Etat for plan og geodata

Denne etaten er organisert under byrådsavdeling for byutvikling, næring og klima. To hovedfelt utgjør ansvarsområdene til etaten: På den ene siden dreier det seg om byplanlegging i Bergen. Under her faller utarbeiding av kommuneplanens arealdel, kommunedelplaner, offentlige reguleringsplaner og noe utredningsarbeid i forhold til transport og arealbruk. I tillegg dekker dette ansvarsområdet Bergensprogrammet for transport, byutvikling og miljø. Det andre store ansvarsfeltet handler om produksjon og vedlikehold av kommunens grunnkart, planarkiv og register over eiendommer.

Datakilder

Ifølge fagsjefen i etaten ligger mesteparten av dataene de produserer ute på visningstjenesten bergenskart.no. Denne tjenesten er beregnet på privatpersoner, og det er knyttet betingelser til bruken (nedlastede dokumenter er kun til bruk for privatpersoner). Profesjonelle brukere henvises til Infoland, en kommersiell løsning drevet av Norsk Eiendomsinformasjon. Kommunale og statlige virksomheter kan hente ut data via samarbeidsprosjektet Norge digitalt.

Noe av etatens grunntanke er at alle datasettene de produserer skal gjøres tilgjengelige for gjenbruk. Fagsjefen forstår gjenbruk særlig som de tilfellene hvor etaten har produsert et datasett for et bestemt formål, og så viser det seg i ettertid at dette kan også andre gjøre seg bruk av, på måter man kanskje ikke hadde tenkt på i utgangspunktet.

Vurderinger og planer

Likevel produserer etaten også spesialkart som foreløpig ikke ligger ute, som aktsomhetskart for skred, vannstand og lignende. Bruken av disse er først og fremst tiltenkt kommunens saksbehandling av plan- og byggesaker. Samtidig er det ingen spesiell grunn til at disse ikke ligger ute ennå, utover det at det bare ikke har blitt gjort. Fagsjefen framhevet også at det ikke ville være noe problem å få lagt dem ut, hvis det skulle være folk utenfor "huset" som ville ha tilgang. All informasjonen som ligger i kartene ligger også ute i rapportform på kommunens sider om risiko og sårbarhet.

Boligetaten

Denne etaten sogner til byrådsavdeling for byutvikling, næring og klima. Dens hovedoppgave er å tilby hjelp innenfor boligfinansiering, bostøtte og kommunale leieboliger til boligmarkedets vanskeligstilte.

Datakilder

Boligetaten registrer informasjon om alle leiligheter som eies av kommunen. Denne informasjonen dreier seg om boligene generelt, beboerne, husleien, boligdriften, i tillegg til meldinger for hvilke boliger som er ledige. Alle søknader, vedtaksbrev og mangelbrev registreres i et system som heter Boinfo.

Ingen av disse dataene er gjort tilgjengelige for allmennheten. Boligetaten har gitt noen aktører pålogging med begrenset innsyn, som vil si at de kun kan se den biten som er relatert til deres fagområde. Eksempler på disse er sosialkontorene eller kemneren, som gis tilgang til de delene av dataene de trenger i forhold til sine arbeidsoppgaver. Dataene gis ikke ut til andre enn de funksjonene i kommunen som bruker dem i arbeidet sitt. Hovedgrunnen til dette er personvernrelatert, men det har heller ikke vært interesse fra folk for å få tak i disse dataene.

Vurderinger og planer

Seksjonslederen avviser ikke at dataene i framtiden en gang kan frigis som statistikk, forutsatt at de var kvalitetssikret og at aggregerte data ble presentert. Foreløpig har de ingen planer om tilgjengeliggjøring i overskuelig framtid.

Etat for byggesak og private planer

Denne etaten tar for seg private planforslag, fradelingssaker, byggesaker, seksjoneringsaker og kart- og delingsforretninger. Etaten er organisert under byrådsavdeling for byutvikling, næring og klima.

Datakilder

Etaten bruker saksbehandlingssystemet DocuLive. Dette er et system som brukes av mange kommuner til både byggesaker og plansaker. Fra dette systemet tar kommunen ut statistikker til eget bruk, hvor de tar utgangspunkt i saksgangen. Målet er å se for eksempel hvor lang tid man har brukt på en enkel sak, antall saker av ulike typer, hvor mange saker som har kommet inn, hvor mange som er behandlet; og så settes disse opp mot målsettinger internt i kommunen. Etaten har også et faktureringsystem hvor de tar ut statistikk om gebyrer til eget bruk.

Disse dataene er ikke tilgjengelige slik at utenforstående kan finne informasjon selv, men det gis i noen tilfeller innsyn hvis det etterspørres. Når det gis innsyn og ikke, er avhengig av hva slags saker det er snakk om. I noen tilfeller kan det være deler av eller hele saker som kommunen vil unnta offentlighet. Det kan også være enkelte momenter i sakene som kommunen ikke vil gjøre tilgjengelig, og det kan være interne dokumenter som er unntatt offentlighet.

Vurderinger og planer

Tilgjengeliggjøring har vært diskutert i etaten. Oslo kommune, som bruker samme system, legger ut saksinnsyn på nett, mens Bergen har en såkalt ”lett” variant hvor dette ikke er mulig. Gitt de like systemene kan Bergen få en praksis som ligner Oslos på sikt, men det forutsetter at det gjøres mer arbeid med dataene i forkant. Slik det er i dag, har ikke etaten kommet så langt at tilgjengeliggjøring kan gjøres forsvarlig. Dette begrunnes med at dagens system ikke har nok muligheter til å registrere og behandle dataene i forkant av en eventuell publisering, og at man dermed ikke har nok kontroll over dataene til å kunne gjøre dem tilgjengelig.

Grønn etat

Grønn etat har forvaltningsansvar overfor kommunens parker, byrom, skog, natur- og friluftsområder. Landskapsplanlegging, veiledning og næringsutvikling i landbruket, sikring av areal til friluftsmål, miljøsikring og tilrettelegging for friluftsliv hører også til Grønn etats ansvarsområde. Etaten er organisert under byrådsavdeling for byutvikling, næring og klima.

Datakilder

Grønn etat har ansvar for de såkalte grønne dataene i kommunen; data over biologisk mangfold, turveier, friluftsområder m.v. Av datasettene til Grønn etat er det særlig to kartleggingsprosjekter som anses for de mest komplette. Disse datasettene hører til dataene over biologisk mangfold, og gjelder et kartleggingsprosjekt av naturtype og et prosjekt hvor viltbestanden kartlegges. Grønn etat forvalter også et datasett som heter AR5, som er et digitalt arealressurskart. Det var opprinnelig tenkt til bruk innenfor jordbruk og skogbruk, og viser hva som er skog, hva som er dyrket, og hva som er beitemark. Dette ligger ute på nettet, og forvaltes på nasjonalt plan av Norsk institutt for skog og landskap. Det finnes også tilgjengelig på bergenskart.no, i tillegg til på nasjonale kartløsninger som publikum selv kan gå inn på.

Ellers har etaten samlet en del ulike datasett, men den største utfordringen er store forskjeller i kvalitet. Enkelte av datasettene anses for å være av for lav kvalitet for viderebruk.

Grønn etats datakilder tilgjengeliggjøres hovedsakelig på to måter, i tillegg til at noen datasett fortsatt ikke er tilgjengelig for det brede publikum. En god del av dataene forvaltes av Etat for plan og geodata gjennom nettsiden bergenskart.no. Mange datasett legges også inn i Naturbase, Direktoratet for naturforvaltnings database over vilt, naturtyper, arter, fri- og friluftsområder m.v.

Det er per i dag ingen betingelser knyttet til viderebruk av Grønn etats datakilder. De ser på dataene de forvalter som til fri bruk. Samtidig tilgjengeliggjøres dataene blant annet på bergenskart.no, og som vi så med Etat for plan og geodata er det her knyttet enkelte betingelser til viderebruk.

Vurderinger og planer

Grønn etat jobber kontinuerlig med å legge til rette for tilgjengeliggjøring av nye data. De sitter på en del data, og er i gang med å legge ut mer. Hovedgrunnen til at ikke alt legges ut, er dårlig kvalitet på noen av datasettene. Disse er i stor grad uferdige, og må kvalitetssikres før de kan legges ut. Etaten har planer om å utnytte disse på sikt. Representanten påpeker at Bergen er med i Norge digitalt, hvor alle offentlige institusjoner gjør kartdata tilgjengelig for hverandre via en portal (prosjektet ledes av Statens kartverk). Derfra har det kommet nye signaler om at nye data også bør tilrettelegges for offentlig bruk, men foreløpig er det mest visningstjenester, ikke fri nedlasting av selve dataene.

Bergen kommunale bygg

Ansvarsområdet til Bergen Kommunale Bygg (BKB) er kommunens behov i forhold til lokaler til kommunal virksomhet. Dette behovet dekker BKB gjennom leie, bygging og forvaltning. Etaten er organisert under byrådsavdeling for finans, konkurranse og eierskap.

Datakilder

Tidligere har BKB hatt oversikt over eiendommer og grunneiendommer. Per i dag er det litt opphold fordi de bytter ut systemer, hvor de blant annet bygger opp et nytt system som går på intern leie. I det gamle systemet var det bare forvaltningsavdelingen som hadde et datasystem i forhold til innkreving av leie fra leietakere og lignende. Nå kommer avdelingen til å få et FDV-system: for forvaltning, drift og vedlikehold. En viktig forandring i det nye systemet er nemlig at dette nå skal utvides slik at både driftsavdelingen og de som har med byggingen å gjøre skal kunne gå inn på det samme systemet. Da får man oversikt over bygninger og arealer, hvilket utstyr som finnes, hvem som er leietakere, hva de betaler, meldinger fra leietakere om ting som er galt, hva som er ledig og hva som ikke er ledig. Det skal også gjøres rapportering av byggekostnader, for så å beregne hva internleien vil bli. Ellers tas det ut rapporter på intern leie og forvaltning.

Tidligere har dette systemet vært et internt system, slik at leietakere for eksempel har vært nødt til å ta kontakt på telefon for å få se tegninger over bygget og lignende. Med det nye systemet skal dette gjøres "frontend", med at man etter hvert skal kunne gå inn på sitt eget bygg og se branntegninger, egne kontrakter på bygget, hva de betaler i internleie m.v.

Vurderinger og planer

Informasjonen i BKBs systemer har det kun vært internt behov for så langt, og BKB ser heller ikke hva som kunne vært av allmenn interesse av disse dataene. I tillegg er det snakk om personvern – "de é jo enkelte ting som ikkje ska spres på byen". BKBs representant tror heller ikke at informasjonen vil ha potensial for viderebruk senere.

Portalorganisasjonen

Arbeidet til Portalorganisasjonen er organisert under Seksjon for konkurranse og utvikling, tilhørende byrådsavdeling for finans, konkurranse og eierskap. Ansvarsområdet til Portalorganisasjonen dekker drift og utvikling av kommuneportalen. Gjennom denne portalen skal kommunen drive formidling av informasjon og tjenester, i tillegg til at den skal fungere som en kommunikasjonsplattform overfor innbyggere, samarbeidspartnere og ansatte.

Datakilder

For å lage tjenestene ut mot innbyggerne benytter Portalorganisasjonen seg av offentlige registre, som de binder sammen i sine systemer. Dette brukes for eksempel på en slik måte at hvis man som innbygger i kommunen logger seg på portalen, får man opp informasjon tilknyttet eiendommer som står i ens navn. Det er viktig for kommunen å koble informasjon fra Folkeregisteret om personer sammen med eiendomsdata fra matrikkelsystemet – altså å knytte person/bedrift til eiendomsdata, for slik å kunne sikre kommunens inntektsgrunnlag.

Portalorganisasjonen samler, strukturerer, likestiller, kvalitetssikrer og tilgjengeliggjør virksomhetsinformasjon ut fra konseptet Master Data Management. De har altså ingen egne datakilder, men jobber med å koble sammen ulike datasett.

Byarkivet

Bergen byarkivs ansvarsområde er å ta imot, oppbevare og betjene arkivmateriale fra kommunal virksomhet, bedrifter, organisasjoner og enkeltpersoner i Bergen. Byarkivet er kommunens arkivinstitusjon

og –myndighet og er organisert under byrådsavdeling for kultur, kirke og idrett.

I forhold til fagsystemene i Bergen kommune fører Byarkivet oversikt over datasystemer i bruk, og systemer som er ute av bruk. Dette gjelder alle systemer som produserer informasjon tilknyttet saksbehandling, drift m.v. I Bergen kommune har man ett stort saksbehandlingssystem kalt BKSak, og det er dette systemet som produserer den offentlige journalen.

Det er særlig eiendomsinformasjon som etterspørres fra Byarkivet, og de er i gang med et stort prosjekt der dokumentene fra gamle byggesaker digitaliseres.

Bergen og Omland havnevesen

Bergen og Omland havnevesen har blant annet ansvar for planlegging, bygging, drift og vedlikehold av alle offentlige trafikkhavneanlegg i samarbeidsregionen. Havnevesenet er en interkommunal bedrift som dekker kommunene Askøy, Austrheim, Bergen, Fedje, Fjell, Lindås, Meland, Os, Radøy, Sund, Øygarden og Hordaland fylkeskommune. Havnevesenet styres av et havneråd og et havnestyre.

Datakilder

Havnevesenet registrerer data tilknyttet all aktivitet i havnen. Disse dataene er hovedsakelig anløpsdata og data om gods. Det registreres informasjon om anløp av skip, hva skipet fører med seg av last og containere, biler om bord i skipene og passasjerer.

Dataene brukes i statistikker som legges ut på Havnevesenets nettsted under to kategorier:

Anløpsstatistikker

Statistikkene på disse sidene gjelder hovedsakelig tre områder: forventede skipsanløp, antall skipsanløp, og skipenes bruttotonnasje. De forventede skipsanløpene kan lastes ned i en mer detaljert versjon, men som pdf-fil. Statistikken som gjelder antall skipsanløp og skipenes bruttotonnasje kan lastes ned i mer detaljert form som excel-dokument. Her kan man bla seg fem år tilbake, i tillegg til at statistikkene er spesifisert på type skip og innenriks/utenriks i tillegg til indre havn/øvrig distrikt.

Godsstatistikker

To typer data registreres og gjøres til statistikk i forbindelse med gods: godsmengde og antall containere. Begge disse kan også lastes ned i detaljert versjon som excel-dokumenter, med statistikk for de siste fem årene. Statistikkene deles i begge tilfeller også inn i indre havn/øvrig distrikter, innenriks/utenriks, i tillegg til lastet/losset. Når det kommer til godsmengdestatistikken, er det ti forskjellige kategorier for hva slags type gods som losses og lastes, og denne oversikten er dermed ikke så detaljert – tre av kategoriene er for eksempel ulike typer containere. Containerstatistikken viser altså oversikt over antall containere som losses og lastes.

Vurderinger og planer

Havnevesenet gjør foreløpig bare informasjon som har med sjøaktivitet å gjøre tilgjengelig på nettstedet sitt. Havnevesenet registrerer imidlertid også aktivitet på land i havneområdet, og dette gjøres i et eget system. Her registreres kjøretøyene som beveger seg inn på kaiområdene, data om gods som losses og lastes i forbindelse med de ulike kjøretøyene og lignende.

Det er to årsaker til at denne informasjonen ikke er lagt ut på nett: Først og fremst personvernproblematikk. Det registreres en del personopplysninger og bilnummer fordi det trengs adgangskort for å komme inn på kaiområdet. Likevel er holdningen til representanten fra Havnevesenet at det ikke er noe i veien for å filtrere vekk personopplysningene, slik at man kunne lagd den samme typen statistikker for biler og last som for sjøtransporten. Den andre grunnen er at den etterspørselen som Havnevesenet opplever fra medier, publikum og andre organisasjoner som har med skipstrafikk å gjøre,

ikke handler om det som foregår på land, kun skipstrafikken og sjøtransporten. Siden de fleste henvendelsene går på dette, fokuserer også Havnevesenet på denne statistikken.

Det er ikke knyttet noen betingelser til viderebruk av dataene, fordi de – som det understrekes – er offentlige.

Oppsummering

Samtalene med dette utvalget av etater og virksomheter i Bergen kommune gir et innblikk i sentrale utfordringer kommuner står overfor når de skal forvalte sine datakilder med tanke på viderebruk. Ledelsen i kommunene må takle både teknisk tilrettelegging og lovverkets begrensninger; de må motivere organisasjonen om at mer transparenser er ønskelig, og øke kompetansen internt:

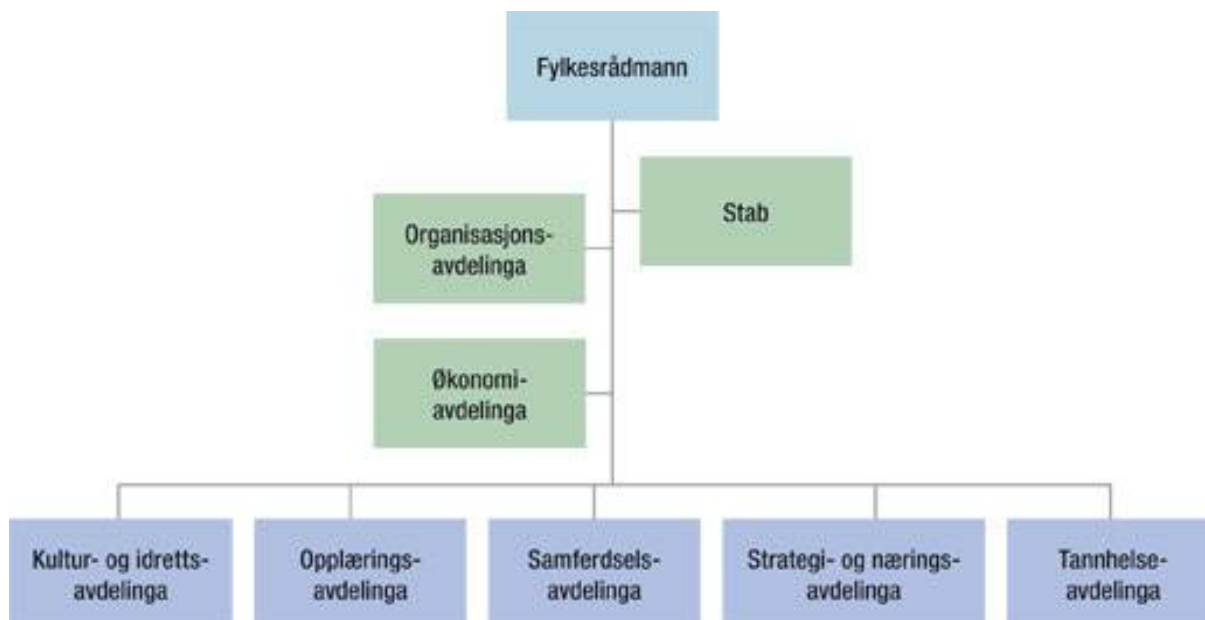
- **Personvern:** Hensynet til personvern er en av de største utfordringene som flere etater trekker fram i forhold til mulig frigivelse av data, da særlig i forhold til innbyggere og klienter. Antakelig finnes det her et behov for mer kunnskap: Hvilket aggregeringsnivå gir for eksempel et tilstrekkelig personvern? Det kan tenkes at virksomheter holder tilbake mer data enn nødvendig for å være på den sikre siden. Et mulig eksempel er Bergen kommunes politikerregister: Også her ble personvern pekt på som en grunn til å holde tilbake data, i det tilfellet permisjons- og fraværdata fra de enkelte politikernes sider - selv om det vel kan tenkes at politikeres fravær er av interesse når velgerne skal evaluere folkevalgtes innsats.

- **Systemer og konsesjoner:** Flere peker også på utfordringer i forhold til systemer, verktøy og programvare. Dette ble trukket fram som et stort hinder for å få gjennomført endringer, fordi man innenfor kommunesystemet er prisgitt de datasystemene som allerede eksisterer. I flere tilfeller mangler også økonomien for å få til endringer. Flere nevner at konsesjons- og distribusjonsavtaler for ulike systemer, i tillegg til lisenser tilknyttet disse, gjorde informasjonsflyt tungvint innad i kommunen. For eksempel er hvert sosialkontor nødt til å ha en egen konsesjon for å kunne innhente informasjon fra Folkeregisteret. Dessuten er det mange ulike systemleverandører som legger premisser. Dette skaper problemer i forhold til semantisk interoperabilitet; altså en felles forståelse for dataelementene. Som en påpekte: I den digitaliserte kommunen ”flyter det” nå informasjon i datasystemer i stedet for i papirer. Et eksempel på denne flyten som ble trukket fram, gjaldt i forhold til politiske endringer og hvilke systemutfordringer de skaper. Etter sist valg ble nemlig Brannvesenet flyttet til en ny avdeling. Det innebar at systemstøtten også måtte endres, noe som igjen skapte større uoversiktighet over de forskjellige fagsystemene og hva de inneholder. Videre er en generell utfordring for kommunen at all informasjonen de har til publisering er gratis, mens publiseringen i seg selv koster penger gjennom konsesjoner og lignende.

- **Holdninger og kompetanse:** I et av møtene ble konservative holdninger innad i kommunen trukket fram som et hinder for videreutvikling av allerede eksisterende registre på nett, og det ble etterlyst innovativ tenkning på disse områdene i kommunesystemet. Flere av spørsmålene som ble stilt på dette kartleggingsprosjektets vegne, gjaldt forhold som ikke var diskutert eller reflektert over i det hele tatt i enkelte avdelinger. Det er i hvert fall en svært tydelig indikasjon på at bevisstheten om viderebruk av data er lav i enkelte fagmiljøer. Noen har også problemer med å se at datasettene de sitter på skulle ha allmennhetens interesse. Andre forvalter data som de mener er av for dårlig kvalitet til å frigis.

Hordaland fylkeskommune

Hordaland fylkeskommune har ansvaret for videregående opplæring, kultur, samferdsel, tannhelse og regional utvikling. Fylkeskommunens øverste politiske organ er fylkestinget. Fylkesordføreren er øverste politiske leder, mens fylkesrådmannen leder administrasjonen med stabs- og sektoravdelinger (se organisasjonskart).



Hordaland fylkeskommune er en av fylkets største arbeidsgivere med over 4000 ansatte. Totalbudsjettet for 2010 er på 8,3 milliarder kroner. 5,3 milliarder går med til drift av de ulike tjenesteområdene og 2,8 milliarder er satt av til investeringer.

Kartleggingsarbeidet

Kartleggingsprosessen startet med et møte med leder for enheten Analyse, utgreiing og dokumentasjon, og vi fikk der bedre oversikt over hvem vi burde kontakte. Etter e-post-utsendelse til diverse avdelinger med presentasjon av vårt mål og ønske med prosjektet, samt påfølgende telefoner, fikk vi i havn en del møter. Likevel lot det seg dessverre ikke gjøre å få møter med alle relevante avdelinger og etater. Vi har fått materiale fra seksjoner og enheter under fire av de fem sektoravdelingene.

Opplæringsavdelingen

Videregående opplæring er fylkeskommunens største tjenesteområde og omfatter all kompetansegivende opplæring mellom grunnskolen og høyere utdanning. I tillegg til Hordalands 46 videregående skoler inkluderer dette fagopplæring i bedrift, oppfølgings- og pedagogisk-psykologisk tjeneste (OT/PPT), og voksenopplæring i regi av den enkelte skole.

Opplæringsavdelingen bruker fagsystemet Extens, som benyttes av totalt ti fylkeskommuner. Her registreres skoleadministrative data: om elever og elevforhold (karakterer, om de må ta fag om igjen osv.), pedagogisk personale (ansettelsesforhold), skoler, fag.

Opplæringsavdelingen mottar data om skolesøknader fra VIGO, en portal der elever søker om plass på videregående skole. Ellers registrerer skolene egne data inn i systemet. Det er kun intern tilgang til fagsystemet. Data rapporteres herfra til Statistisk sentralbyrå (KOSTRA), til nasjonal vitnemålsdatabase (NVB) og til Utdanningsdirektoratet (Skoleporten).

Kontaktpersonen i opplæringsavdelingen påpeker at de ofte får henvendelser angående forskjellige typer data, sammenstillinger av disse etc., og at de i henhold til offentlighetsloven er pålagt å gi et svar. I 90 prosent av tilfellene henvises til SSB, KOSTRA og Skoleporten. I visse tilfeller er det en skjønnsvurdering. Det kan stilles spørsmål ved hva dataene skal brukes til; særlig om det er "hensiktsmessige" sammenligninger som skal utføres. Det ble understreket at det er en fare for at data tas ut av sin korrekte sammenheng, og at dette kan bidra til at det fremstilles feilaktige bilder av virkeligheten.

Vurderinger og planer

Avdelingen kan gi ut statistikk direkte hentet fra Extens, men det frigis naturligvis ikke data som kan knyttes til personer, ei heller til spesifikke skoleklasser.

Det ble uttrykt skepsis til ønskeligheten av å legge mer til rette for å tilgjengeliggjøre aidentifiserte data. Faren for at data blir presentert utenfor sin opprinnelige sammenheng var hovedgrunnen til skepsisen.

Oppfølgings- og pedagogisk-psykologisk tjeneste (OT/PPT)

Oppfølgingstjenesten ligger under opplæringsavdelingen. Tjenesten skal kartlegge, ha oversikt over og gi tilbud til unge (16-24) som hopper av videregående opplæring.

Tjenesten mottar data fra opplæringsavdelingen gjennom VIGO, og kontakter ut fra denne informasjonen de aktuelle elevene. I løpet av året skal også alle elever som slutter rapporteres fra skolene til oppfølgingstjenesten. Data som tjenesten selv produserer er kun tiltakskoder (hva som tilbys de forskjellige - kurs, praksisplass etc.).

Oppfølgingstjenesten er pålagt å levere rapport til Utdanningsdirektoratet to ganger i året. Det trekkes ut tall som anses som viktige: statistikk over hvor mange som hopper av videregående skole i løpet av året, hvor mange som får tilbud om hjelp, benytter seg av den og hva som tilbys.

Dette er ganske nytt, og har tvunget avdelingen til å ha bedre oversikt over data. Det arbeides nå med å utvikle et system som bedrer kvaliteten på statistikken. Hordaland fylkeskommune er den eneste som har slått sammen oppfølgingstjenesten og pedagogisk-psykologisk tjeneste. Saksbehandlingssystemet er i hovedsak tilpasset den pedagogisk-psykologiske tjenesten. Man ønsker et bedre system slik at oppfølgingstjenesten fungerer optimalt mot eksempelvis VIGO.

Pedagogisk-psykologisk tjeneste er ikke pålagt å avlevere noe data. Ingen informasjon tilknyttet denne tjenesten ligger ute for andre enn saksbehandlere.

Vurderinger og planer

Som forventet er det ikke mulig å få tilgang til noe av dataene som OT/PPT produserer; disse er kun til internt bruk grunnet sensitive personopplysninger. På spørsmål om data som eksempelvis går på hvilke tilbud fra OT/PPT som de forespurte benytter seg av kan tilgjengeliggjøres, ble det svart at det er mulig – men at det ikke er planer om å gjøre det. Det ble henvist til rapporten som leveres til Utdanningsdirektoratet. Her skal man finne det som er mest relevant. Dersom det ønskes data for viderebruk, må de hentes herfra.

Kultur- og idrettsavdelingen

Kultur- og idrettsavdelingen har ansvar for utvikling innen områdene kunst og kultur, kulturminnevern og museum, arkiv, idrett og friluftsliv, organisasjonsliv og bibliotek. Det ble kun ett møte i denne seksjonen, med Fylkesarkivet. Av de øvrige ble vi henvist til informasjon på nettsiden hordaland.no.

Fylkesarkivet

Arkivet håndterer eldre arkivsaker fra sentraladministrasjon og institusjoner i fylkeskommunen. Data i arkivet hentes fra alle fylkeskommunale enheter og virksomheter, som personalmapper, saksdokumenter, møtereferater osv. Dette inkluderer all aktivitet i fylkestinget (voteringsdata, møtereferater osv). Fylkesarkivet oppdaterer også forandring i drift av offentlige virksomheter, eksempelvis bytte av navn, lokalisering m.v.

I tillegg finnes det arkivsaker fra de tidligere institusjonene for psykisk utviklingshemmede (før HVPU-reformen), og barnevernsinstitusjoner, og det tas inn private virksomheters arkiv.

Per i dag ligger alle data i word-dokumenter som er søkbare på arkivnivå, ikke dokumentnivå. Det er kun intern tilgang, men filene vil kunne oversendes ved forespørsel. Ellers må personlig henvendelse/oppmøte til for å få innsyn. Vår kontaktperson uttrykte stor misnøye med dette systemet; det var generelt en positiv holdning til at data skal benyttes, men det lite brukervennlige systemet kompliserer dette.

Vurderinger og planer

Fylkesarkivet har som mål å få alle data inn i arkivinformasjonssystemet Asta, en felles nasjonal standard for registrering, distribusjon og fremfinning av arkivdata, men dette vil ta tid. Det forventes at det i hvert fall vil være mulig å søke i denne basen på serienivå, slik at det kommer opp hvilke typer dokumenter som finnes (foto, personalarkiv osv). Vår kontaktperson var svært positiv til et felles, nasjonalt system; det vil i tilfelle gi en langt bredere oversikt over hvor, og på hvilket nivå, arkivsaker finnes. Det ble påpekt at dette må komme på initiativ fra staten.

Samferdselsavdelingen

Samferdselsavdelingen håndterer fylkeskommunens ansvar for blant annet kollektivtransport, fylkesveinett, prioritering av riksveimidler og den overordnede planleggingen av transport i fylket.

I denne avdelingen hadde vi møte med selskapet Skyss, som er eid av fylkeskommunen. Skyss har ansvaret for ruteplanlegging, takster og rabatter, salg og billettering, markedsføring og informasjon i ruteområdene. I en egen melding om kollektivtransporten tydeliggjøres mål, status og strategier i sektoren.

Skyss understreker at det er viktig at alt ved kollektivtransporten gjøres så enkelt og tilgjengelig som mulig, slik at befolkningen ønsker å reise kollektivt. Ett tiltak er et sanntidsinformasjonssystem som gjør det mulig å spore opp hvor på ruten et kjøretøy befinner seg, og gjøre denne informasjonen tilgjengelig for passasjerene. Skyss skal teste sanntidsinformasjon på en busslinje i Bergen i begynnelsen av 2010, for deretter å videreutvikle dette dersom det er vellykket.

Vurderinger og planer

Vi fikk ikke inntrykk av at det er planer om at sanntidsdataene skal tilgjengeliggjøres for viderebruk slik at andre har mulighet til å utvikle tjenester med dataene som grunnlag.

Tannhelseavdelingen

Den offentlige tannhelsetjenesten, som denne avdelingen styrer, gir tilbud om behandling til barn og ungdom i alderen 0-18 år, psykisk utviklingshemmede, eldre uføre og langtidssyke i institusjon og hjemmesykepleie, ungdom 19-20 år og andre grupper som blir prioritert av fylkeskommunen. Tannhelsetjenesten tar også imot voksne, betalende pasienter.

Tannhelsetjenesten har data på hvem som har blitt tilbudt tjenestene, og på hvem som har benyttet seg av hvilke tilbud. Men fylkeskommunen har ikke tilgang på kommunens registre over de nevnte målgruppene, da denne informasjonsflyten er beskyttet i tråd med pasientvern og personvern. Eksempelvis vil data på

personer som benytter seg av hjemmesykepleie være utilgjengelig for fylkeskommunen frem til det rapporteres inn av kommunen hvem som ønsker å benytte seg av tannhelsetilbudet. For å sikre rapporteringen inngår fylkeskommunen og kommunen samarbeidsavtaler spesifikt utarbeidet for hver målgruppe.

Det brukes felles journalsystem OPUS og røntgensystemet SIDEXIS. Det er nødvendig for tannhelsetjenesten å ha tilgang på pasienters helseopplysninger, medikamentbruk o.l grunnet behandlingen.

Tallene på hvem som har fått tilbud og hvor mange som benytter seg av tilbudet rapporteres fra fylkeskommunen til KOSTRA.

Vurderinger og planer

Det gis ikke tilgang til disse dataene utover det som rapporteres til KOSTRA og i avdelingens årsmelding. Det er heller ikke, ut fra informasjonen vi fikk, planer om å gjøre mer data tilgjengelig. Dette begrunnes med at datakildene inneholder sensitive personopplysninger, og at det vil være for komplisert og tidkrevende å tilrettelegge for bruk av data som utelater disse.

Oppsummering

Samtlige av avdelingene vi kontaktet i Hordaland fylkeskommune henviste oss videre til nettsiden hordaland.no/statistikk, en sammenstilling av data fra SSB. Her kan det hentes ut statistikk på områder som befolkning; næring og arbeid (arbeidsplasser, bedrifter); næringer (bransjer); kompetanse (utdanningsnivå mm); folkebibliotek (utlån); samferdsel (kjøretøy, ulykker mm); areal, bosted og miljø; folkehelse – på hele fylket, en region eller kommunenivå. Dataene kan bl.a. sorteres, brukeren kan generere grafikk og laste ned data i ulike formater.

Vi fikk opplyst at intensjonen med tjenesten da den ble lansert (2006) var at de ulike avdelingene skulle legge inn egne data. Det er til nå kun fylkesbiblioteket som har gjort dette.

Fylkeskommunens nettside hordaland.no ble trukket fram som en god kilde for data fylkeskommunen kan tilby av våre kontaktpersoner. Det ble også henvist til nettsiden polsak.ivist.no, en database over alle politiske saker som er behandlet i fylkestinget og de ulike underliggende utvalgene.

I våre samtaler med representanter for fylkeskommunen var disse vurderingene fremtredende når det gjaldt spørsmålet om tilrettelegging for viderebruk på andre måter enn det som gjøres i dag:

- **Personvern:** Ingen av de spurte virksomhetene hadde rådata tilgjengelig for nedlasting. Samtlige av de forespurte oppgav ”inneholder sensitive personopplysninger” som hovedgrunnen til at data ikke tilgjengeliggjøres.
- **Kostnader og relevans:** Flere svarte også at det ville være tidkrevende å gjøre data tilgjengelig - altså er også dette et spørsmål om ressurser. Et annet svar som gikk igjen om tilgjengeliggjøring av data var ”oppfattes ikke som nødvendig”. Inntrykket er at grunnet krav om rapportering til KOSTRA og Utdanningsdirektoratet, årsmeldinger og referat oppleves tilgjengeliggjøring på egne nettsider som overflødig.
- **Ikke prioritert:** Samtidig er det generelt lite bevissthet rundt temaet. Det var få av de vi snakket med, både via telefon, e-post og ansikt til ansikt, som hadde reflektert over problemstillingen utenfor sitt ansvarsområde. Unntaket var de som jobbet tett på systemutvikling, men blant disse var det ingen som så det som ønskelig at data skulle frigis i større grad enn i dag.

3.5 Oppsummering av kartleggingsarbeidet

Sett i sammenheng utgjør resultatene fra kartleggingsarbeidet en skisse av tilstanden på feltet offentlig sektors data i Norge. Bevissthetsnivået, kompetansen og de faktiske initiativene som er tatt varierer betydelig, både mellom forvaltningsnivåer og fagområder, men også mellom ulike etater innen samme virksomhet. Det er eksempler både på virksomheter som informerer bredt og gjør data tilgjengelig for viderebruk på mest brukervennlige måte; og på virksomheter der man til nå i liten grad har reflektert over temaet.

Vår skisse bør ideelt sett bearbeides videre til et mer detaljert bilde. Når det gjelder oversikt over selve datakildene, er det som nevnt vanskelig å se at dette kan løses på noen annen måte enn at virksomhetene selv registrerer sine datakilder i et norsk "data.gov". Datakildene i kommune og fylkeskommune bør definitivt bli gjenstand for en bredere og mer fullstendig kartlegging. Det er antakelig en god ide at KS bidrar til å koordinere et slikt prosjekt. Det kan avlaste kommunene, som uansett ville måtte sette av noe ressurser til en slik utvidet kartlegging.

Våre hovedfunn kan oppsummeres slik:

- **Informasjonstørke:** Statlige virksomheter har ingen enhetlig praksis for å informere om hvilke datakilder de forvalter. Mange virksomheter informerer ikke om dette i det hele tatt på sine nettsider, og hos andre er informasjonen godt gjemt. Eksempler på utmerket informasjonspraksis finnes, men det er ikke mange av dem. De aller fleste virksomheter har et stort forbedringsbehov.
- **Manglende oversikt:** Det finnes ingen sentral oversikt over datakilder i offentlig sektor. Fragmenter til en slik oversikt eksisterer, men det er et sterkt behov for at det opprettes en slik sentral side der virksomheter kan legge inn informasjon om sine datakilder. Flere ferske initiativer, fra dette prosjektet og andre, er et tegn på at behovet for en slik oversikt er økende (se også våre anbefalinger i del 5).
- **Stort potensial:** To av tre statlige virksomheter som deltok i vår undersøkelse sier de har data som har potensial for viderebruk, men som i dag ikke blir utnyttet. Seks av ti sier de vil gjøre mer data tilgjengelig i løpet av et år. Undersøkelsen formidler et bilde av at mange fagvirksomheter i staten er kjent med problemstillinger rundt viderebruk av data og arbeider aktivt med tilgjengeliggjøring. Det virker som det hos mange er en positiv interesse for å arbeide med viderebruk. Nye initiativer for data-frigivelse lansert bare de siste månedene peker i samme retning.
- **Stort opplysningsbehov:** Interessen for og kompetansen på temaet viderebruk av data er svært ujevnt fordelt i virksomhetene. Mye kunne gjøres med en koordinert innsats for å opplyse om hvordan datakilder kan gjøres tilgjengelig for viderebruk på best mulig og korrekt vis. I del 5 foreslår vi blant annet at det utarbeides et sett av prinsipper for data-frigivelse og en håndbok som medarbeiderne i virksomhetene kan benytte.

4. Fra data til innsikt: Muligheter og motforestillinger

Sterke internasjonale trender forplanter seg hurtig, og det er all grunn til å tro at vi vil få se flere gode eksempler på gjennomtenkt frigivelse av offentlige data i Norge i månedene og årene som kommer. Selv om temaet er nytt for mange, både i offentlig sektor og blant de potensielle "viderebrukerne", er det viktig å begynne en debatt om prinsipper for konstruktiv bruk av data allerede nå. Debatten bør gå parallelt med at det arbeides med utforming av en mer enhetlig praksis i offentlig sektor. Den bør fokusere på mulighetene som åpner seg ved frigivelse av offentlige datakilder, men også ta motforestillinger alvorlig.

Frigivelse av offentlige data og utnyttelsen av dem henger sammen med de etter hvert velkjente større strukturelle endringene av betingelsene for produksjon og distribusjon av nyheter, og informasjon og kunnskap generelt. Med internett har effektive verktøy for å skape og dele informasjon og ytre seg offentlig blitt tilgjengelige for alle. Verktøyene blir stadig billigere, stadig mer kraftfulle og stadig enklere å bruke.

Erkjennelsen om at informasjon kan produseres og spres av hvem som helst, har inspirert til mange svært optimistiske vyer om demokratiserende effekter av disse strukturelle endringene. Spesielt med tanke på land med mer eller mindre sterke innskrenkninger i friheter og rettigheter er det blitt stilt forventninger til de positive virkningene av at aktivister får tilgang på nye kommunikasjonsverktøy. Det virker logisk; autoritære regimer verner alltid om sitt informasjonsmonopol.

Den internasjonale debatten om nettets demokratiserende effekter føres etter hvert i en mer nøktern tone. Eksempelvis har det vist seg at undertrykkende regimer kan leve godt med en ganske betydelig ytringsfrihet på nett, så lenge borgerne holder seg til "godkjente" temaområder og styrer unna politikk. En annen variant er når regimet hyrer inn egne folk til å oversvømme nettdebatter med regjeringshyllende ytringer. Analytikeren Evgeny Morozov (2009) er en av dem som minner om at makthavere kvikt lærer seg hvordan de kan utnytte de nye nettverktøyene, som sosiale medier, til sin fordel:

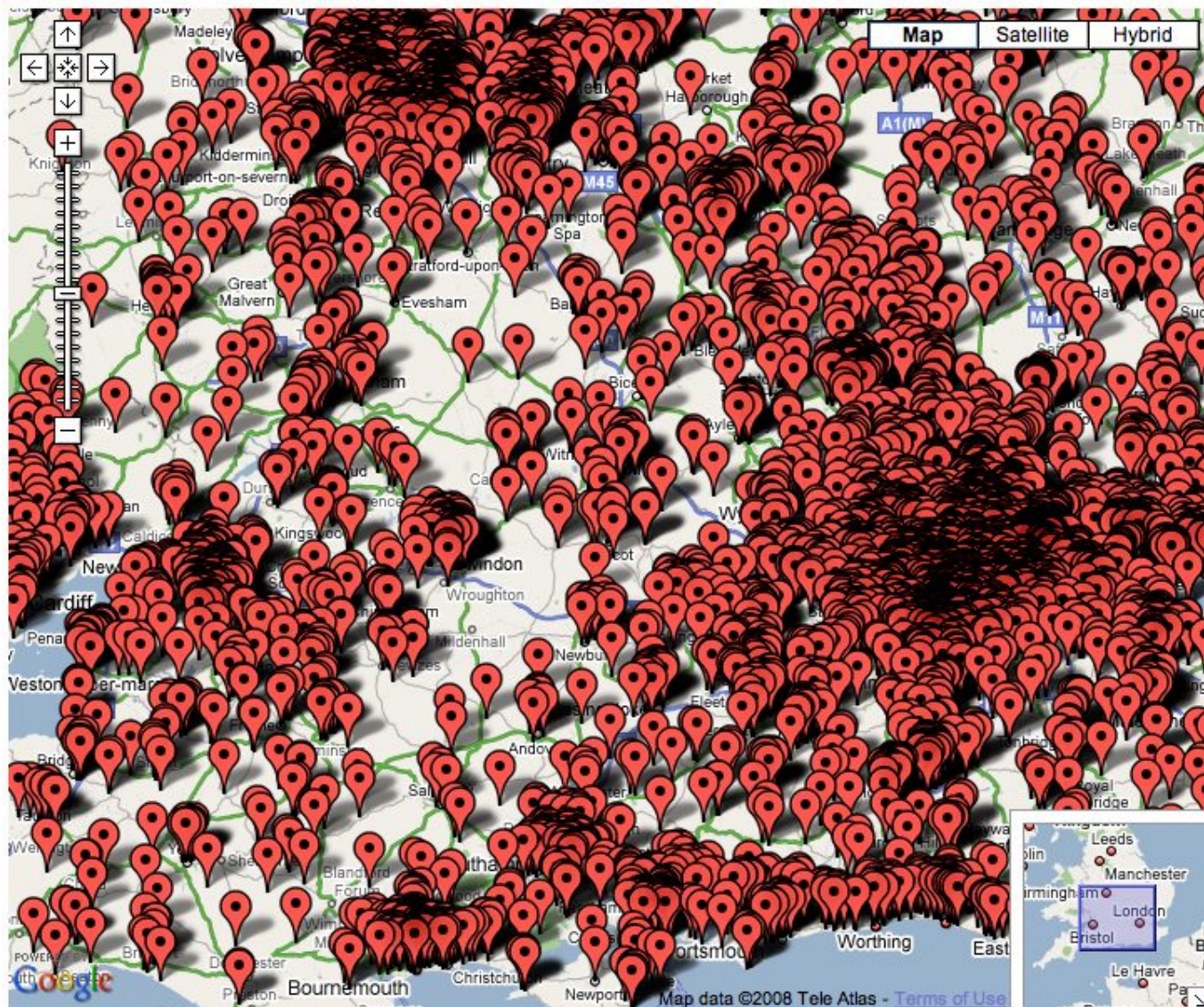
Despite what digital enthusiasts tell you, the emergence of new digital spaces for dissent also lead to new ways of tracking it. Analogue activism was pretty safe: if one node in a protest network got busted, the rest of the group was probably OK. But getting access to an activist's inbox puts all their interlocutors in the frame, too. The result is a cat-and-mouse game in which protesters try to hide from the authorities by carving out unconventional niches.

Hva kan vi så vente oss når mer av offentlig sektors data blir tilgjengelig for alle -- og ulike datakilder sammenstilles med hverandre? De nye produksjons- og distribusjonsmåtene skaper muligheter for aktører av mange slag. Vi bør være forberedt på at det neppe blir bare nyttige tjenester og opplysende erkjennelser. To eksempler kan illustrere.

Den høyreekstreme russiske organisasjonen DPNI ("Bevegelsen mot ulovlig innvandring") har omfavnet "mashups", altså sammenstilling av data fra ulike kilder. I 2008 rapporterte Paul Goble om dette på bloggen Window on Eurasia. DPNI's avdeling i Volgograd publiserte kart som viste hvor i byen mennesker fra ikke-russiske minoritetsgrupper bodde. Kartene var ledsaget av oppildnende kommentarer om hvordan minoritetsgruppene angivelig fortrenger russere fra bydeler. DPNI hadde her tatt i bruk informasjon som Volgograds byadministrasjon hadde publisert med de beste hensikter, for å informere om byens økende mangfold og skape forståelse mellom befolkningsgruppene. Som Morozov påpeker i sin omtale av saken: Slike mashups er også en velprøvd teknikk blant frivillige organisasjoner og aktivister som arbeider for saker som miljøvern og mot etnisk diskriminering.

British National Party har tiltrukket seg mye oppmerksomhet de siste årene. I juni 2009 ble to av det høyreekstreme partiets representanter valgt inn i Europaparlamentet. Ved flere anledninger har partiets medlemslister blitt lekket og publisert på nettet. I november 2008 kastet aktivister og andre interesserte seg

umiddelbart over listene da de ble gjort tilgjengelig. IT-mannen Ben Charlton brukte Google Maps til å lage et kart over hvor i Storbritannia BNP-medlemmene bodde. Kartet ble umiddelbart kritisert for å skape et unøyaktig inntrykk (Butcher 2008). Charlton brukte Google Maps' vanlige design med knappenålshoder, og dette fikk det til å se ut som det lille partiet hadde medlemmer tett i tett over hele landet:



Kritikk ble også ytret mot å gjengi BNP-medlemmenes adresser -- et potensielt farlig brudd på personvernet til medlemmene i et lovlig politisk parti. Impliserte kartet at det nå var fritt fram for å avlegge BNP-medlemmer et uvennlig besøk? Charlton sa han med tanke på slike konsekvenser bevisst hadde gjort adressedetaljene mindre presise, men dette ble også kritisert. Det kunne føre til at folk ble blinket ut som slett ikke var BNP-medlemmer. I tillegg kom usikkerheten om listene i det hele tatt var oppdaterte og genuine. Charlton tok etter hvert kritikken til følge og fjernet kartet han hadde laget. På MySocietys epost-liste ble det diskutert om man skulle bruke dataene, men organisasjonens direktør Tom Steinberg gikk sterkt imot (Steinberg 2008):

[T]he moment you sacrifice the values and compromises that hold together liberal democracies (such as a presumption of innocence and a right to privacy for people who've not actually been convicted of crimes) for the sake of humiliating your political opponents, you're starting on a path far more likely to result in ruination for us all than a bunch of marginal wing nuts.

Dataene i BNP-saken kommer ikke fra offentlig sektor, de var stjålet, men håndteringen og debatten er uansett relevante. Hvordan kan det sikres at data ikke brukes på måter som setter personvernet og andre sentrale demokratiske verdier over styr? Hvordan kan offentlige virksomheter bidra, og hva kan medier og frivillige organisasjoner gjøre? Til syvende og sist er det kanskje offentlighetens demokratiske styrke og kvalitet som avgjør om data viderebrukes på måter som er demokratisk og etisk holdbare, for ikke å si opplysende. Det kan argumenteres for at kritikken av Ben Charltons mashup nettopp er et eksempel på en fungerende offentlighet. Misvisende fremstilling og fortolkning av data ble påpekt og korrigeret.

Hvordan skape mening ut av data?

Spørsmålet om edruelighet i omgang med datakilder kan også stilles på mindre dramatisk vis. Er den nødvendige kompetansen til stede i offentligheten, hos journalister og andre, til å kunne forstå og framstille sammenhenger på holdbart vis? Datatilsynets direktør Georg Apenes formulerte noen motforestillinger i en artikkel i Vox Publica i november 2009:

Det er (...) fullt mulig å holde seg personvernlovgivningen etterrettelig og unngå overtramp i forhold til personopplysningsloven, men likevel komme galt av sted fordi man regner med at alt som karakteriserer den definerte gruppen automatisk også gjelder den enkelte medlem av den samme gruppen. Når for noen år siden en avis i California samkjørte listen over innehavere av kjørekort med listen over mottagere av offentlige bidrag til blinde og svaksynte, fant den ut at det kjørte 18 blinde omkring på statens veier. Den slags blir det selvsagt fengende overskrifter av, men for at sammenstillingen av de to basene skal kunne gi edruelig mening, var det nødvendig for avisens medarbeidere å grave seg bakover i tallene.

Neste runde i debatten om viderebruk av offentlig sektors data må med andre ord dreie seg om hvordan mashups og nye tjenester kan legges til rette slik at dataene gir mening og har praktisk nytte for publikum. Et sentralt argument for frigivelse av data er at det vil styrke deltakelsen i demokratiske prosesser. Borgerne får bedre mulighet til å sette seg inn i hvordan de styres. I neste omgang kan de bruke denne informasjonen til å påvirke folkevalgte og offentlig forvaltning.

Nick Diakopoulos har i en artikkel drøftet utviklingen i USA, og påvist at det fortsatt er et stykke fram til at alle rådataene som nå "frigjøres" omsettes til innsikt (Diakopoulos 2009 - se artikkelen i sin helhet under punkt 6.1). Til nå har forsøkene på sammenstilling av datakilder ofte endt med temmelig meningsløse kart som i liten grad bidrar til innsikt om datakildenes betydning, skriver han.

Data.gov and indeed many of the applications built around it somehow fall short of the mark in terms of helping people share and build on the insights of others – to produce information. It's not simply that we need interfaces to data, we also need ways to collaboratively make sense of that data.

Journalister har lang erfaring med å trekke essensen ut av komplekse saksforhold. Dermed burde de ha en god mulighet til å spille en viktig rolle i arbeidet med fornuftig viderebruk av datakilder, mener Diakopoulos. Han foreslår følgende tiltak og prinsipper:

- Applikasjoner basert på offentlige data (feks mashups, tjenester) skal lages med tanke på å være informative og gi innsikt, ikke bare som enkel sammenstilling av data fra ulike baser
- Journalister skal, i samarbeid med publikum, ta en lederrolle i å skape mening ut av data
- Applikasjonene må være laget slik at de samler opp brukernes fortolkninger og kommentarer og presenterer disse i tilknytning til presentasjonen av data
- Myndighetenes data-portaler (som data.gov) skal samle inn og presentere lenker til alle applikasjoner som bruker data fra de ulike offentlige datakildene

Det er naturligvis et åpent spørsmål om medier vil sette av ressurser til nybrottsarbeid som dette i en økonomisk vanskelig fase. Her vil de fleste redaksjoner trenge påfyll av kompetanse og investere mer tid enn det som brukes i dag. De eksemplene som er kommet på bruk av datakilder i norsk journalistikk de

siste årene, gir imidlertid grunn til forsiktig optimisme. NRK Brennpunkt har fått mye oppmerksomhet for Maktbasen, som blant annet samler data om folkevalgtes engasjement i næringsvirksomhet. TV 2 har i flere nyhetssaker sammenstilt data fra ulike offentlige kilder og slik gitt en ny dimensjon til nyhetsdekningen (eksempler inkluderer tall fra de nasjonale prøvene i skolen og politiets førerkortbeslag). Redaksjonene i TV 2 og NRK har i arbeidet med disse prosjektene støtt på betydelige praktiske problemer med å få tilgang til dataene som behøves for å lage databaser og sammenstillinger. I enkelte tilfeller får man likevel tilgang, til dels ved kreativ egeninnsats, men i andre tilfeller er det rett og slett ikke mulig å utføre det journalistiske prosjektet slik man ønsker fordi de nødvendige dataene ikke er tilgjengelig elektronisk.

Det journalistiske potensialet i åpne data: Hva stemte Kari Kommunepolitiker?

Høsten 2009 presenterte Drammens Tidende (dt.no) og Edda Media en side om Buskerudbenken på Stortinget. Her samles data om Buskerud-politikernes aktivitet på Stortinget (saksordførerverv, spørsmål og forslag de stiller m.v), og dette sammenstilles med biografiske opplysninger og nyheter om den enkelte politiker. Initiativet indikerer at norske medier er interessert i å bruke ressurser på data-basert journalistikk. Enda bedre vil denne tjenesten bli når representantenes voteringer i hver enkelt sak kan gjengis.

Som ledd i kartleggingsprosjektet har vi sett nærmere på potensialet for slike tjenester på det lokalpolitiske nivået. Det burde være stort. Lokale beslutninger påvirker ofte velgerne på en mer direkte merkbar måte enn de nasjonale. Informasjonstilbudet til velgerne er imidlertid ikke godt nok i dag. Voteringer i fylkesting og kommunestyre er en god indikasjon. Hvorfor er det så vanskelig å finne ut hva Kari Kommunepolitiker faktisk har stemt i ulike saker? For velgerne er det viktig å ha god informasjon om de løpende sakene kommunestyret behandler. Dels for å holde oppsikt med bestemte saker som til og med kan berøre en direkte, dels for å kunne gjøre seg opp en veloverveid mening om hvem man skal stemme på ved neste valg. Hva kan være bedre for en velger enn enkelt å kunne følge med på hva enkeltrepresentanter, og partigruppene samlet, stemmer i konkrete saker?

En enkel gjennomgang av de fem største bykommunene viser at bare Kristiansand kommune tilbyr velgerne å sjekke stemmegivning i enkeltsaker. Kristiansands tjeneste kunne vært atskillig mer brukervennlig, men er en god start og selvsagt mye bedre enn ingenting. Det er fristende å sette opp en lang ønskeliste for forbedringer: Brukeren kan lage en side for en enkeltrepresentant via søket, men siden har ingen permanent nettside. Siden er ikke koblet til annen informasjon om politikeren, og det er ikke mulig å abonnere på nytt om representantens stemmeatferd via e-post eller RSS.

Dette kan vi i dag gjøre mye enklere. Prosjektene TheyWorkForYou.com og Hvem stemmer hvad viser hvordan. Det britiske prosjektet er spesielt fint, her kan den interesserte se hele stemmegivningshistorien til enkeltrepresentanter og abonnere på nytt om dem. Begge disse eksemplene gir oversikt over representanter på nasjonalt nivå, men hvorfor skal ikke det samme kunne gjøres lokalt?

På lokalt og regionalt nivå kan det være verdt å prøve et samarbeid mellom medier og de folkevalgte forsamlingene. Mediene har en ekspertise i å tilrettelegge informasjon som kommunen mangler; resultatet kunne bli lokale portaler der velgerne fikk full oversikt over de lokale folkevalgtes innsats. Dette krever imidlertid at de grunnleggende dataene er på plass. I svært mange kommuner registreres ikke en gang hver enkelt representants stemmegivning, isteden noteres bare resultatet av avstemningen (antall stemmer for/mot). Kristiansand har kunnet lage sin tjeneste fordi kommunen har et elektronisk voteringssystem. Alle politikere i bystyresalen er utstyrt med en datamaskin, og stemmer via denne.

Det ligger mye viktig informasjon i voteringsdata, som i mange andre typer offentlige data. Eksemplet viser igjen det grunnleggende problemet: Tilgangen på data. Medier og journalister må fortsatt bevise at de evner å omsette data til innsikt, men først må de få en reell sjanse.

5. Sikker og effektiv frigivelse av data: forslag til tiltak

På bakgrunn av kartleggingen utført i dette prosjektet og sammenligningene med praksis i andre land, vil vi i denne delen foreslå noen konkrete tiltak som kan bidra til økt oppmerksomhet om og frigivelse av datakilder i norsk offentlig sektor.

Utgangspunktet må være et enkelt definert mål: Flere offentlige datakilder må frigis for viderebruk. Dette må selvsagt gjøres på en måte som ikke truer personvernet.

Initiativet ligger først og fremst i offentlig sektor, og spesielt hos Fornyings- og administrasjonsdepartementet (FAD). Men som kartleggingen har vist, kompetansen og interessen er ujevnt fordelt på fag- og forvaltningsnivåer, og det er mange virksomheter som ligger langt framme og kan bidra med avgjørende erfaringer som vil komme godt med når et slikt mål skal virkeliggjøres.

Tiltakene vi foreslår knytter seg til kunnskap og kompetanse, regler og rammeverk, insentiver og virkemidler.

1. Opprett et norsk data.gov: Dette bør defineres som et hasteprojekt. Erfaringene fra spesielt USA viser at en portal der datakilder fra alle deler av offentlig sektor presenteres under ett, er et effektivt tiltak for å fokusere og øke oppmerksomheten om verdien av offentlig sektors data. En raskt arbeidende prosjektgruppe burde kunne få opp et slikt nettsted i løpet av kort tid (i hvert fall innen første halvår 2010). Den amerikanske og den kommende britiske portalen kan tas som utgangspunkt og de foreløpige erfaringene derfra studeres for å se om det er behov for spesifikt norske justeringer. Selvsagt bør man også bygge på erfaringene til de norske virksomhetene som har gjort et arbeid på feltet allerede, som SFT, SSB og Avinor.

2. Oppmuntre til regionale og lokale datakildesamlinger: Kommuner og fylkeskommuner bør i samarbeid med KS, FAD og andre lage enkle maler for lokale og regionale samlinger av datakilder. Spesielt større byer, men også mindre kommuner har data som vil være av stor verdi for innbyggere, næringsliv og medier å få tilgang til (her kan man også trekke på erfaringene til regionale dataportaler som San Franciscos på datasf.org).

3. Definer prinsipper og lisenser: Det er stor variasjon i hvordan data gjøres tilgjengelig for viderebruk (ulike formater osv.) og hvilke betingelser som knyttes til dataene (fra fri viderebruk til kun ikke-kommersiell m.fl). FAD bør definere et sett klare prinsipper for offentlig sektors data. Her kan man la seg inspirere av den britiske regjeringens liste (gjengitt i del 2). Prinsippene bør inneholde en definisjon av hva slags lisenser datasettene skal utstyres med. I dag brukes altfor mye energi på å finne ut om og hvordan data kan viderebrukes. Et enhetlig regelverk vil være et av de viktigste enkelttiltakene.

4. Gi personvernet spesialbehandling: Både i virksomheter i offentlig sektor og i befolkningen er det bekymring for at personvernet kan svekkes dersom mer offentlige data frigis. I prinsippene for frigivelse av offentlige data bør spørsmålet om sikkerhet for persondata gis inngående behandling, slik at virksomhetene har klare kriterier å forholde seg til når de skal vurdere frigivelse av datakilder.

5. Gjør data-frigivelse til kvalitetskriterium: Den årlige kvalitetsvurderingen av offentlig sektors nettsteder tiltrekker seg mye oppmerksomhet. Det er gjevt å komme høyt opp her og virksomhetene setter av ressurser for å forbedre seg. I 2010-vurderingen bør nye kriterier tas inn for å måle hvor godt virksomhetene informerer om datakilder på nettstedene sine, og hvor dyktige de er til å frigi datakilder på korrekt måte. Mange av prinsippene vi foreslår blir definert (under forrige punkt) bør her brukes mer eller mindre direkte som målepunkter. Dette bør gi resultater allerede i løpet av det kommende året.

6. Definer pilotprosjekter: Plukk ut bestemte typer datakilder som i dag ikke er tilgjengelige og som har særlig stort potensial. Definer pilotprosjekter for å utvikle verktøy og definere standarder for frigivelse av disse datakildene, og ha som del av prosjektet at data faktisk frigis. Inviter næringsliv, utviklere og medier til å konkurrere om pilotprosjektene. Eksempel: Voteringsdata fra Stortinget, fylkestinget og kommunestyrene er i rapporten pekt på som en type data med stort potensial for viderebruk i sammenhenger som kan styrke interessen for demokrati og valg.

7. Skriv en håndbok: Samle alle regler, prinsipper, lisenser og eksempler på beste praksis for frigivelse av data i en nettbasert håndbok som oppdateres fortløpende (en wiki egner seg meget godt til dette). Her kan man la seg direkte inspirere av det nederlandske prosjektet omtalt i denne rapporten. Som del av det prosjektet ble det også produsert en plakat som illustrerer stegene fram mot frigivelse av data (se del 6.3). Som en begynnelse på et slikt prosjekt har vi publisert en artikkel om god praksis for frigivelse av data (se del 6.2)

8. Driv nettverksbygging: Ansatte som arbeider med datakilder i ulike etater og nivåer i offentlig sektor bør møtes til workshops -- fysisk og nettbasert -- for å utveksle erfaringer. Denne nettverks- og kompetansebyggingen bør være åpen for alle interesserte også fra næringsliv, journalistikk og utviklermiljøer. Også her kan man trekke direkte på de nederlandske erfaringene referert til i denne rapporten.

9. Utlys ekstramidler til data-frigivelse: I offentlige virksomheter er mange opptatt av direkte eller indirekte kostnader som påløper hvis de skal arbeide mer med å klarere og frigi datakilder. Dette kostnadsspørsmålet bør tas alvorlig: Ett tiltak kan være å utlyse en konkurranse der friske midler øremerkes dette arbeidet. En del av ekstrakostnadene vil være knyttet til arbeid som må gjøres én gang.

10. Opprett en "utrykningsenhet": Kostnadsspørsmålet kan også takles ved at det opprettes en prosjektgruppe eller "task force" som arbeider på tvers av etater og virksomheter. Gruppen bør utvikle løsninger som hjelper virksomhetene med å komme hurtigere i mål med å legge data til rette for viderebruk.

Alle disse forslagene retter seg mot de relevante politiske myndighetene og fagorganene, altså særlig FAD og DIFI. Flere av tiltakene kan imidlertid med fordel settes ut i livet i samarbeid med organisasjoner, utviklere, medier, alminnelige brukere. I tillegg er det selvsagt mye de potensielle "viderebrukerne" kan gjøre. Medier og journalister har antakelig bare så vidt begynt å prøve ut mulighetene offentlighetslovens nye paragraf 9 (og andre lovhjemler) gir for å få utlevert data (denne muligheten er som kjent også åpen for alle, ikke bare journalister, selv om de sikkert er storbrukerne). Der hvor datakilden er kjent, men ikke gjort tilgjengelig for nedlasting, kan medier og andre i mange tilfeller "skrape" dataene ut av offentlige nettsider -- en praksis som allerede med hell er prøvd av enkelte. Slik "skraping" kan nok brukes som pressmiddel mot enkelte motvillige virksomheter.

Hvis de mener saken er viktig nok, kan mediernes organisasjoner også ta initiativet til en kampanje for å frigi offentlige data. I Storbritannia har The Guardian gått i bresjen for en slik "Free our data"-kampanje. Skulle man starte noe slikt nå, ville det være naturlig at mediene samarbeider med næringslivsaktører, organisasjoner og andre interesserte.

Konkrete tiltak kan bidra mye til økt bevissthet om offentlig sektors data og mulighetene frigivelse åpner. Men et virkelig gjennombrudd krever antakelig at det formuleres og kommuniseres en klar politisk målsetting om å frigi data for å øke verdiskaping og innovasjonstakt og åpne opp demokratiske prosesser. Det burde her være tilstrekkelig å vise til hvilken vekt den britiske og den amerikanske regjeringen har lagt på dette temaet det siste året. I dag er det i stor grad opp til hver enkelt virksomhet å definere sitt eget forhold til åpne data. En slik fragmentert situasjon er ingen tjent med.

6. Analyser og "bruksanvisninger"

I arbeidet med dette prosjektet har vi rapportert fortløpende i nettmagasinet Vox Publica om nyheter på feltet fra inn- og utland, og om framskritt og utfordringer med vårt eget prosjekt. For dette formålet opprettet vi bloggen Fakta først (se alle innlegg og kommentarer på voxpublica.no).

Vi har også fått to analytikere med spesialkompetanse til å skrive om sentrale aspekter ved frigivelse og viderebruk av data. I denne delen av rapporten gjengir vi deres artikler i sin helhet. I tillegg gjengir vi en visuell framstilling av prosessen med å gjøre data tilgjengelig for viderebruk, laget av Ton Zijlstra og James Burke for et prosjekt om datakilder i Nederland.

6.1 Nicholas Diakopoulos: From Civic Data to Civic Insight

Journalists can help the public make sense of the growing amount of government data now being made available.

Earlier this year the water began to recede on government data in the United States with President Obama's announcement of an unprecedented push toward further transparency in the federal government. But with the rush of new data comes the challenge of making sense of it all — something admittedly still in its formative stages.

By June of 2009 the nation's Chief Information Officer, Vivek Kundra, had overseen the launch of data.gov with the goal of increasing public access to machine readable datasets produced by the federal government. Other government sites such as usaspending.gov and recovery.gov have since been launched to provide even more focused data on how the U.S. spends its taxpayers' dollars.

A promise of increased participation

The promise of data.gov and of many of these other civic data collections is in allowing citizens to participate in the scrutiny of their government and of society at large, opening vast stores of data for examination by anyone with the interest and patience to do so. Open source civic data analysis, if you will.

The array of data available on data.gov is still sparse in some areas but has steadily grown to include things ranging from residential energy consumption, to patent applications, to national water quality data among others.

And the data transparency movement isn't just federal anymore: State and local municipalities such as California and New York City are following suit with pledges to make more civic data available.

The benefits of the open data movement are also starting to be recognized throughout Europe. The U.K. has called on Sir Tim Berners-Lee, the inventor of the world wide web, to lead a similar government data transparency effort there, which should soon result in a data.gov analogue. And indications of movements are beginning to stir in Germany.

Beyond Data Scraping

In the U.S. various government data resources have been available in some form or another online for years now. Programmers could scrape these online data sources by writing custom parsers to scan webpages and create their own databases. And many journalist-programmers working in today's modern newsrooms still do. But it's messy, it doesn't scale or extend well, it's brittle, and ultimately the data that results may not interoperate well with other data.

Having government buy-in to the publication of organized and structured data lowers the barriers substantially for developers and others to get involved with analyzing that data. It also means that structured formats, such as those that conform to semantic Web standards can interoperate more easily and be utilized to build ever more complex applications on top of the data.

Data.gov ≠ Insight.gov

So now that the U.S. government is publishing all kinds of data online, society will be better, right? Well – maybe. Let’s not forget that data has a long way to go before it becomes the information and knowledge that can ultimately impact back on policy.

Some non-governmental organizations are pushing data to become information by incentivizing contests with big prizes. For instance, the Apps for America 2 contest, coordinated by Sunlight Labs, awarded a total of \$25,000 to the top application submissions which made data.gov data more transparent and accessible for citizens.

These efforts at coordinating developers and stimulating application development around government data are vital, no doubt. The applications which result typically involve polished interfaces and visuals which make it much easier for people to search, browse, and mashup the data.

Take for example the Apps for America 2 winner, DataMasher, which lets users create national heat maps by crossing two datasets (either adding, subtracting, dividing, or multiplying values). These operations, however, can’t show correlation, and at best they can only show outliers. As one anonymous commenter put it:

I don't get it. It shows violent crime times poverty. So these are either poor, or violent, or both? I don't think multiplying the two factors is very enlightening.

What we end up with is that many of the possible combinations of datasets lead to downright pointless maps which add little if any information to a discourse about those datasets.

Data.gov and indeed many of the applications built around it somehow fall short of the mark in terms of helping people share and build on the insights of others – to produce information. It’s not simply that we need interfaces to data, we also need ways to collaboratively make sense of that data.

The Minnesota Employment Explorer was an early foray into helping people collaboratively make sense of government data. It not only visualizes employment information but also allows people to ask questions and build on the insights of others looking at the visuals in order to make sense of the data. In the long run it’s these kinds of sensemaking tools that will really unlock to potential of the datasets published by the government.

What’s Next?

With a long tradition of making sense of the complex, there’s a unique opportunity for the institution of journalism to play a leadership role here. Journalists can leverage their experience and expertise with storytelling to provide structured and comprehensive explorations of datasets as well as context for the interpretation of data via these applications. Moreover, journalists can focus the efforts and attention of interested citizens to channel the sensemaking process.

I’ll suggest four explicit ways forward here:

(1) that data-based applications be built with an understanding of trying to promote information and insight rather than simply be database widgets,

- (2) that journalists should be leaders (but still collaborators with the public) in this sensemaking enterprise,
- (3) that these applications incorporate the ability to aggregate insights around whatever visual interface is being presented, and
- (4) that data.gov or other governmental data portals should collect and show trackback links to all applications pulling from its various datasets.

And finally, after we all figure out how to make sense of all this great new data, lies the question of whether government is even “listening” to these applications. Is the federal government prepared to accept or adopt the insight of its constituents’ data analysis into policy?

6.2. Eirik Stavelin: Kan mine data gjenbrukes?

Hva er viktig å tenke på når data skal frigis for gjenbruk? Vi har laget et utkast til bruksanvisning.

Prosjektet Fakta først arbeider med å kartlegge offentlige datakilder som bl.a. kan brukes til journalistiske formål. Første innlegg i bloggen fikk overskriften Data er fakta vi kan bruke. Lesere har oppfordret oss til å ta et skritt tilbake og forklare hvordan det rent teknisk går til når en virksomhet frigir data for viderebruk av andre. Hva kreves av tilrettelegging, hvilke formater er å foretrekke — det er noen av spørsmålene vi forsøker å svare på her.

I informasjonsvitenskapen beskrives data som “bærer av informasjon”, “en samling symboler som når de settes sammen etter bestemte regler, kan gi informasjon”[1]. Før vi har lagt på disse reglene omtales ofte datasettene som “rådata”. Det er så reglene vi bruker til å behandle rådataene som avgjør hvilken informasjon vi kan få ut av dem. Å dele data åpner for alternative måter å behandle dataene på, alternative regelsett til å oversette rådata til informasjon. Hvis dette gjøres på metodisk holdbare måter blir dataene omgjort til informasjon vi kan bruke. I beste fall hjelper dette oss til å basere vår forståelse av verden på fakta.

Hva skal til for at data skal kunne gjenbrukes?

Enkelt sagt bør dataene være plattformuavhengige, maskinlesbare og ha en tydelig struktur. Det betyr at de kan brukes på en hvilken som helst datamaskin, og må være mulig å håndtere med programkode uten at et menneskelig øye og tolkning må til for å forstå innholdet. Struktur og metadata (data om data) hjelper oss å gjøre dette mulig.

Intensjonen bak frislipp av offentlige data er ikke bare målet om gjennomsiktighet, men også innovasjon. Hvilke produkter og tjenester som bygges på grunnlag av data som deles aner vi lite om før dataene er tilgjengelig. Når data som tidligere har vært skjult bak kommunegrå vegger åpnes, åpnes også et potensial for friske øyne til å se nye måter å skrive kode som omdanner rådata til informasjon.

Litt enkelt: hvis frislipp ikke innebærer betydelig ulempe for andre (personvern, konkurransehensyn, med mer) og er betalt av skattebetalerne, så bør dataene frislippes.

Hva slags format skal jeg dele i?

Viktigst av alt, formater likesinnede kan bruke. Er du usikker, del det du tror andre kan bruke. Men dette er ikke alltid et enkelt spørsmål å besvare, det kommer i stor grad an på hva slags data du har.

Det er vanskelig å kategorisere datakilder, offentlige så vel som andre. Hvor ofte datasettene endres og hvor store de er kan gi en indikasjon på hvordan de bør deles.

Enkelte datasett er små og oppdateres sjeldent, slik som navn på fylker og kommuner, postnummer, flyplasser, politi-, helse og sosialdistrikt, osv. Slike datasett kan gjerne være tilgjengelig som flate filer. Med “flate filer” menes datafiler som er strippet for programspesifikke koder og notasjoner, slik at dataene

kan leses av ulik programvare. Erik Bolstad deler forbilledlig postnummerregisteret etter en opprydningsdugnad i sommer.

Data som sjeldent eller aldri blir oppdatert bør deles åpent på nett, slik at søkemotorer kan hjelpe folk å finne ut at disse dataene eksisterer. Dette trenger ikke være mer komplisert enn å legge filene med beskrivelse tilgjengelig på nett.

Dette er på ingen måte noen «best practice» for deling av data, men det er en metode med lav terskel for å komme i gang. Flate filer krever at de som skal gjenbruke dataene må reorganisere dataenes logiske struktur, en potensielt tidkrevende og vanskelig oppgave, som kan resultere i uriktige tolkninger av dataene.

Data som ofte oppdateres kan også legges på nett som filer, men dette skaper fort mye manuelt arbeid for alle parter. Med store datasett likeså.

Gi oss et API!

Et API er et programmeringsgrensesnitt som muliggjør kommunikasjon mellom programvare. I mange tilfeller vil det som er mest aktuelt være et web-API, altså et grensesnitt en programmerer kan nå over http (protokollen weben er bygget på).

Litt enkelt igjen: når du peker nettleseren din til www.uib.no, så svarer webserveren som denne websiden er lagret på med å servere deg html som nettleseren viser som en webside. Den kunne (og kan) likeledes svare med andre formater, som XML, JSON, RSS, kommaseparerte verdier, SVG, et bilde, osv. “Svaret”, altså de dataene som sendes tilbake etter at et kall er utført, kan behandles videre i den programvaren som utførte kallet. Ved å definere et sett med regler for hvilke ressurser som kan nås med ulike metoder og parametre, kan data raskt, presist og enkelt utveksles.

Når data deles på denne måten kalles tjenesten for en web service, som er et av temaene vi legger i ideen om den semantiske veven [2].

Yr.no har fått mye fortjent kudos for sitt frislipp av værdata, og ved å lese dokumentasjonen deres får vi et inntrykk av hvordan dette kan gjøres. NRK grubler på hvordan de kan gjøre noe liknende, det samme gjør SSB.no under prosjektnavnet Nye ssb.no (se eget blogginnlegg om SSBs nettplaner).

Dokumentasjon er et nøkkelord her. For at en slik tjeneste skal være virkelig nyttig bør dokumentasjonen være brukervennlig. Google kan trekkes fram som et godt eksempel. Deres tjenester som har et offentlig API er som regel dokumentert i detalj og med gode eksempler med kode. Google Maps API er et eksempel for kart. Se også Flickr for bilder, Last.fm for nettsamfunn/musikkdata, osv. Nye tjenester som dukker opp og blir populære på nett har ofte et dokumentert API som gjør at data lett kan gjenbrukes.

RSS er din venn. Uansett hva slags informasjonskanal du deler på nett er det noen som er interessert i å vite når du publiserer noe nytt. Hvis du sitter på data som uregelmessig produseres og er uforutsigbare av natur (kalenderdata den ene dagen, informasjonsskriv, kart eller bilder den neste) så er det bedre at vi får vite at noe nytt er på gang framfor å måtte tråle weben kontinuerlig.

Men viktigst: Vi må vite hva vi ser etter

Vi trenger også en norsk utgave av data.gov, en samlet oversikt over rådata på ett sted. Danskene har det allerede i digitaliser.dk, og en britisk versjon er rett rundt hjørnet. Kartleggingsprosjektet vi informerer om på denne bloggen har begynt på en slik oversikt. Det er i det hele tatt god grunn til å tro at det bare er et spørsmål om tid før et norsk data.gov kommer. Forskningsprosjektet Semicolon har slike planer. Det er også private initiativer; nylig har Sondre Bjellås startet wikien datakilder.no.

Som Nick Diakopoulos påpeker bør en slik sentral samling av datakilder inneholde automatisk sporing

(f.eks. ved trackbacks) av prosjekter som bygger på data herfra. Slik kan vi bygge på hverandres erfaringer og kode, og unngå dobbeltarbeid.

Det største problemet for kartlegging av offentlige data sett fra en prosumenters perspektiv, er at vi ikke vet hva vi ser etter. Først når vi vet hvilke rådata som finnes, kan vi begynne å gruble over hvordan disse dataene kan omformes til samfunnsnyttig informasjon.

Referanser

1. Trond R. Braadland, Innføring i informasjonsbehandling (Fagbokforl., 2002), books.google.com.
2. Thomas B. Passin, Explorer's Guide to the Semantic Web (Manning Publications, 2004).

6.3. Visuell ABC for databyråkrater

Til et nederlandsk prosjekt om offentlig sektors data laget Ton Zijlstra og James Burke en visuell framstilling av prosessen fra vurdering til frigivelse av datakilder. Plakaten gjengis på de neste sidene under en Creative Commons- lisens (Attribution-Noncommercial-Share Alike 2.0 Generic).

I'm a civil servant and i want to open some data

step 1: find your data

A

Do you have the data you need?

yes no

Is the data already public?

no yes

Is your organisation owner of this data?

its ours belongs to others

find the data

talk more about possibilities

talk with the owner/owners

aggregate the data to the level that privacy is no longer a problem

make a simple release (data dump)

Go to Step 4: Publish your data

step 2: judicial check

B

Copyright rules

authorship rights

pass no negotiate

Safety

pass no negotiate

Contractual

pass no negotiate

step 3: process your data

C

User permission

pass no process data at aggregated levels

Is the right aggregation-level already available?

no yes

Your data can now be opened :-)

step 4: publish your data

Publish data according to open gov. data principles

If you are analysing the data or presenting it on a website, then, publish your analysis and presentation material (maps, graphics, etc) too for reuse

D
Is it real-time data?

yes

no

Make an API

Publish a complete data dump in different formats

Add it to your gov. data catalog, a wiki, add a unique page for it on your URL to document it correctly

www.

E
My Published Data

Choose right license

documentation

F
Congrats, u have published your data :-)

Keep an eye on who and how your data is used

follow up

DATA PUBLICATION

7. Dokumentasjon

Bakgrunnsdata fra kartleggingen vil bli publisert på voxpublica.no. Vår samling av datakilder utvides stadig og er å finne på: <http://tinyurl.com/offdata>.

Litteratur, kilder

Apenes, Georg (2009): "Ny bruk av "gamle" persondata", Vox Publica 25. november (<http://voxpublica.no/2009/11/ny-bruk-av-gamle-persondata/>)

Butcher, Mike (2008): "Updated: BNP member list mashed with Google maps creates a sea of red dots, but dangerously inaccurate", TechCrunch 19. november (<http://eu.techcrunch.com/2008/11/19/bnp-member-list-mashed-with-google-maps-creates-a-sea-of-red-dots/>)

Christensen, Anders (2009): "Hva skjer når man gir bort flydata gratis?", blogginnlegg på open.bekk.no 13. desember (<http://open.bekk.no/2009/12/13/hva-skjer-nar-man-gir-bort-flydata-gratis/>).

Diakopoulos, Nicholas (2009): "From Civic Data to Civic Insight", Vox Publica 8. oktober (<http://voxpublica.no/2009/10/from-civic-data-to-civic-insight/>)

Goble, Paul (2008): "Window on Eurasia: Russian Nationalists Now Mapping Location of Immigrants in Major Cities", blogginnlegg 5. juni. (<http://windowoneurasia.blogspot.com/2008/06/window-on-eurasia-russian-nationalists.html>)

Hamilton, James T. og Fred Turner (2009): Accountability Through Algorithm: Developing the Field of Computational Journalism. Rapport. (http://dewitt.sanford.duke.edu/images/uploads/About_3_Research_B_cj_1_finalreport.pdf)

Holovaty, Adrian (2006): "A fundamental way newspaper sites need to change", holovaty.com 6. september (<http://www.holovaty.com/writing/fundamental-change/>).

Lie, Håkon Wium (2009): "Offentlig informasjon skal være offentlig", i Grande Røys (red.): Delte meninger, Oslo. (http://deltemeninger.no/-/page/show/2809_offentlig-informasjon-skal-vaere-offentlig)

Morozov, Evgeny (2009): "The Internet: A room of our own?", Dissent Magazine. (<http://www.dissentmagazine.org/article/?article=1939>)

Nærland, Torgeir Uberg (2009): "Bakstreverske holdninger i redaksjonene", Vox Publica 18. mai (<http://voxpublica.no/2009/05/bakstreverske-holdninger-i-redaksjonene/>).

Open Government Directive (2009): Direktiv fra Det hvite hus, 8. desember (<http://www.whitehouse.gov/open/documents/open-government-directive>).

Putting the Frontline First: Smarter Government (2009): Handlingsplan fra den britiske regjeringen (<http://www.hmg.gov.uk/frontlinefirst.aspx>)

Solstad, Øyvind (2009): "Data som vi alt har betalt for", blogginnlegg NRKbeta 3. juli (<http://nrkbeta.no/2009/07/03/data-som-vi-alt-har-betalt-for/>)

Statskonsult (2004): Tilgang til offentlige data i offentlige virksomheter – status på feltet i Norge i dag, statusrapport. (http://www.regjeringen.no/upload/kilde/aad/hdk/2004/0013/ddd/pdfv/220763-statskonsult_sluttrapport.pdf)

Stavelin, Eirik (2009): "Kan mine data gjenbrukes", Vox Publica 2. november
(<http://voxpública.no/2009/11/kan-mine-data-gjenbrukes/>)

Steinberg, Tom (2008): Innlegg i diskusjonsgruppe på mysociety.com
(<https://secure.mysociety.org/admin/lists/pipermail/developers-public/2008-November/003076.html>)

Zijlstra, Ton (2009): "Open Government Data Netherlands Update", blogginnlegg Ton's Interdependent Thoughts 16. juni (http://www.zijlstra.org/blog/archives/2009/06/open_government_2.html)

(Samtlige nettkilder besøkt 5. januar 2010).

Se for øvrig alle blogginnlegg på prosjektbloggen Fakta først, som inneholder flere lenker og henvisninger til nettsteder og andre kilder brukt i arbeidet med prosjektet.
(<http://voxpública.no/seksjon/allmenningen/fakta-foerst/>)

Ophavsrett:

Denne rapporten er publisert under lisensen Creative Commons Navngivelse 3.0 Norge. Det betyr at alle har lov til å dele - å kopiere, distribuere og spre verket. Lisensen gir også lov til å remikse - å bearbeide - verket videre. Mer informasjon om lisensen finnes her:
<http://creativecommons.org/licenses/by/3.0/no/legalcode>