

Saksdokument

Porteføljestyre for forskning og kunnskapsressurser

Til	Porteføljestyret for forskning og kunnskapsressurser
Dato	28.08.23
Saksnummer	15/23
Saksarkivnr.	23/00000
Sakstype	O/D
Saksansvarlig	Vigdis Namtvedt Kvalheim
Saksbehandler	Ørnulf Risnes

Microdata.no

Forslag til vedtak

Porteføljestyret tar saken til orientering og ber om å få framlagt forslag til prismodell.

Saksfremstilling

Kort om microdata.no

Microdata.no er datainfrastruktur hvor forskere, analytikere og studenter får umiddelbar og fleksibel tilgang til kobling og analyse av store mengder registerdata som er tilgjengeliggjort der.

Dataprodusenter som Statistisk sentralbyrå, Kreftregisteret og andre offentlige og private dataprodusenter og -eiere kan selvbetjent dele sine registerdata i microdata.no, til bruk for forskning og analyse. Foreløpig er SSB eneste dataeier med data på plattformen, men et tett og langvarig samarbeid med de sentrale helseregistre har som mål at helseregistre også skal kunne dele sine registerdata i microdata.no, bl.a. for kobling med demografiske og sosio-økonomiske data fra SSB.

Løsningen kontrollerer all databruk gjennom et automatisert anonymiserende grensesnitt, som hindrer at sluttbrukerne får ser eller kan laste ned/printe ut personopplysninger. Fordi sluttbrukere i **juridisk forstand** ikke selv behandler personopplysninger, kan de få tilgang på dagen og dermed også adgang til å analysere detaljerte individdata uten noen form for søknad eller dokumentasjon av behandlingsgrunnlag. Tilgangen gjelder for personer som er ansatte eller studenter ved en institusjon som har tegnet avtale om bruk av microdata.no.

Microdata.no eies, utvikles og driftes av Sikt og SSB i fellesskap. Utviklingen har i hovedsak vært finansiert gjennom Forskningsrådets infrastruktursatsning i to omganger. Tjenesten ble lansert våren 2018, inneholder i dag ca. 440 variabler¹ med data om hele befolkningen på områder som demografi, inntekt, gjeld, utdanning, migrasjon, sykefravær og trygdeytelser tilbake til 1960-tallet. Så langt har data fra microdata.no vært benyttet i over 170 publikasjoner², fra nivå 2-publikasjoner til masteroppgaver, rapporter og NOU-er.



Brukerne (p.t. ca. 200 aktive brukere pr. mnd, ca. 500 aktive pr. år) kommer fra 105 institusjoner i mange ulike sektorer og segmenter; universiteter og høyskoler, forskningsinstitutter, departementer, direktorater, andre organer i statsforvaltningen, kommuner/fylkeskommuner samt konsultantselskaper med oppdrag for offentlig sektor³. Det forventes økt interesse fra helseforskningssiden når helsedata blir tilgjengelige.

Hvorfor microdata.no?

Norge har som ett av få land store mengder koblingsbare registerdata tilgjengelig. Slike data har høy verdi for forskning og analyse. Begrepet «registerdata» handler typisk om data som har blitt registrert om individer som ledd i en administrativ prosess forbindelse med en hendelse eller faste rapporteringer (f.eks. fødsel, flytting, migrasjon, selvangivelse, gjennomført utdanning, innleggelse på sykehus, trygdeytelser, etc.)

Dette er normalt data som er registrert på individnivå. Begrepet «mikrodata» brukes om slike data for å skille dem fra aggregerte data og statistikk. Koblingsbare mikrodata kan belyse langt flere problemstillinger enn det som er mulig med aggregerte data.

SSB har hovedansvaret for produksjon av offisiell statistikk. Statistikkloven gir SSB anledning til å avkreve data fra ulike organer som NAV, Skatteetaten, Lånekassen, kommunene, m.fl. for å kunne produsere offisiell statistikk. Statistikkloven sier samtidig at SSB skal stille disse dataene til disposisjon for forskning og analyseformål når det er mulig.

SSB har betjent norske forskningsmiljøer med registerdata i en årrekke, bl.a. i samarbeid med NSD som nå er en del av Sikt. SSBs registerdatatilbud startet med data knyttet til velferd/trygd, og har senere blitt utvidet med andre typer demografiske og sosio-økonomiske data. SSBs tilbud samt det faktum at Norge som ett av få land globalt har fødselsnummer tilordnet hver person, har gjort det mulig for forskningsmiljøer å få utlevert koblete mikrodata i relativt stor skala, og gjennom det studere samfunnsutviklingen over tid, mellom regioner, m.m. Slike muligheter er fraværende i de fleste land utenom Norden, grunnet mangel på fødselsnummer og/eller mangel på sentraliserte registre.

Situasjonen i Norge er likevel langt fra problemfri, og i praksis har registerdatatilfanget vært dramatisk underutnyttet også i Norge. Både EU-kommisjonen, OECD og ulike norske regjeringer har over lang tid pekt på viktigheten av å utnytte denne type data bedre i forskning og analyse.

Det er mange årsaker til at registerdata har vært underutnyttet i Norge. , Hovedproblemet sett fra et forskerståsted har vært at søknads- og utleveringsprosessen har vært så tid- og ressurskrevende at den i praksis har ekskludert store brukergrupper fra å få tilgang.

Det er hensynet til både hensynet til personvernet og behovet for tilrettelegging av data for hver enkelt bestilling, som gjør at den tradisjonelle søknads- og utleveringsprosessen er tid- og ressurskrevende. Hvert forskningsprosjekt som ønsker å bruke registerdata må spesifisere nøyaktig hvilke data de trenger ut fra tekstlige beskrivelser av data, for hvilken delpopulasjon og for hvilket tidsrom. De må legitimere behandlingsgrunnlag og forskningsmessige behov for data i forhandling med SSB. Når partene er enige, produserer SSB datafiler ihht bestillingen og leverer ut til forskningsprosjektet. SSB tar ikke betalt for data, men de tar betalt for arbeidet med skreddersydd datatilrettelegging og saksbehandling ifm utlevering.

I praksis kan slike søknads- og utleveringsprosesser ta måneder og i verste fall år, og medføre store kostnader for forskningsprosjektet. Mens enkelte store og ressurssterke forskningsmiljøer hadde anledning til å gjennomføre slike søknadsprosesser mer eller mindre rutinemessig, og har ressurser til å bygge datainfrastruktur og sterke registerforskningsmiljøer lokalt, har praksisen virket ekskluderende for mange potensielle brukermiljøer.

Tjenesten microdata.no ble utviklet for å øke bruken av registerdata i norsk forskning. Virkemiddelet var å utvikle en løsning og etablere en tjeneste som demokratiser tilgangen til registerdata og gjøre slike data tilgjengelig for langt flere på en kostnads- og tidsmessig effektiv måte.



Samarbeidet mellom SSB og Sikt (tidligere NSD)

Sikt og SSB eier, videreutvikler og drifter altså microdata.no i fellesskap. Samarbeidet er regulert i en eierskapsavtale og tilhørende samarbeidsavtale. Hva er historien bak dette samarbeidet?

Langvarig samarbeid

Samarbeidet mellom SSB og NSD startet tidlig på 1970-tallet. NSD fikk ansvaret for formidling av SSB-data til forskningsmiljøene. I første omgang handlet samarbeidet mest om data fra utvalgsdataene (spørreundersøkelser), men omfanget ble senere utvidet til kommunedata og registerdata. NSDs nasjonale rolle som forskningsdataarkiv og tette bånd til Norges forskningsråd og forskningsmiljøene og kunnskap om forskningens behov var bakgrunnen for arbeidsdelingen.

Etter et innledende samarbeid mellom NSD, SSB, Rikstrygdeverket m.fl. om formidling av registerdata fra trygdefeltet (KIRUT) på 90-tallet, etablerte SSB selv et tilbud for formidling av registerdata til forskning med utgangspunkt i tryggedata (FD-Trygd) som senere altså ble utvidet med data fra andre sektorer/aktiviteter. Etter hvert fikk NSD oppdraget med å formidle SSB-registerdata til studenter og forskere fra et forenklet og nedskalert registerdatatilfang (FD-Trygd 20%-utvalget).

NSD som teknologiutvikler og standardiseringspådriver

Fra 80-tallet og fremover etablerte NSD seg som global programvareleverandør av statistikk- og analyseverktøy (NSDstat og Nesstar). Parallelt med dette var NSD aktive pådrivere i internasjonalt arbeid rundt standardisering av metadata, og etter hvert maskinlesbare metadata som kunne drive selvbetjente datatilgangsløsninger for forskere. Statistikkprogramvareløsninger fra NSD var i perioder relativt utbredt i Europa og Nord-Amerika (primært hos universiteter og forskningsdataarkiver).

Første felles søknad til Forskningsrådet infrastruktursatsing i 2011

I 2009 etablerte Forskningsrådet ordningen «Nasjonal satsing på forskningsinfrastruktur». I forbindelse med utlysningen i 2011 ble NSD og SSB oppfordret til å skrive en felles søknad om midler til utvikling av en forbedret infrastruktur for registerdatatilgang. Det ble pekt på de respektive institusjonenes komparative fortrinn, herunder NSDs kompetanse om forskningens behov samt teknologiutviklingserfaring rundt formidling og bruk av data. Oppfordringen ble tatt til følge, og NSD og SSB leverte en felles søknad til 2011-utlysningen. Søknaden nådde opp, og 35 MNOK ble bevilget til et 5-årig utviklingsprosjekt under navnet RAIRD (Remote Access Infrastructure for Research Data)⁴. Prosjektet munnet ut i tjenesten microdata.no som altså ble lansert våren 2018.

Konseptet i microdata.no

Visjonen for RAIRD-prosjektet var å kunne tilby forskere registerdatatilgang på dagen og uten noen form for søknad. Visjonen la noen begrensninger, men åpnet samtidig mange muligheter for infrastrukturen som skulle utvikles.

Begrensninger

For å eliminere søknadsprosess og dokumentasjon av prosjekt- og behandlingsgrunnlag var det nødvendig å utvikle et anonymiserende grensesnitt hvor brukerne ikke kunne se eller laste ned data, men likevel kunne koble og analysere data. Nøkkelen var å unngå at sluttbruker behandler personopplysninger i juridisk forstand.

Det er viktig å påpeke at det ikke er *dataene* som anonymiseres i microdata.no; isteden er det analyseresultatene (statistikk, grafikk og annen output skapt av bruker gjennom systemet) som anonymiseres. Data-anonymisering er kostbart og reduserer informasjonsverdien av data, og benyttes ikke. Registerdataene i løsningen ligger inne i full detaljrikdom. Men det anonymiserende grensesnittet fungerer altså slik at sluttbrukerne ikke ser eller får ut personopplysninger av systemet.



Det australske statistikkbyrået (ABS) hadde på dette tidspunktet (rundt 2012) allerede pilotert statistikk- og analyseløsninger med et relativt simpelt anonymiserende grensesnitt-konsept. Tidlig pilotering og testing i RAIRD-prosjektet demonstrerte at et slikt konsept var mulig. Piloteringen avdekket også at brukere som var vant til å kunne se direkte på data ville trenge kompenserende tiltak for å kunne orientere seg godt i dataene gjennom et slikt anonymiserende grensesnitt. Det ble også klart at anonymiserende tiltak ville måtte designes og utvikles løpende for de statistiske metodene infrastrukturen skulle støtte, og at løsningen derfor bare ville kunne dekke en delmengde av alle analysebehov – og bare en relativt liten del ved lansering.

Muligheter

Konseptet åpnet samtidig helt nye muligheter for selvbetjent bruk av data, hvor forskerne interaktivt ville kunne koble variabler fra hele bredden av tilgjengelig datagrunnlag samt løpende definere egne analysepopulasjoner. Det var en kjent sak at forskere ikke sjelden ønsket seg flere variabler eller andre populasjoner etter at data var mottatt/utlevert i tradisjonell utlevering på slutten av en lang og smertefull søknadsprosess. Fleksibiliteten i microdata.no-konseptet og selvbetjeningsløsningene der helt eliminerte denne type problemer.

Muligheter for rik og integrert datadokumentasjon (definisjoner, koding, avvik, etc) i verktøyet åpnet seg også. Dette var viktig av flere grunner, ikke minst for å kompensere for bortfallet av muligheten til å innsipere selve dataene som konseptet innebar. Men også for å gjøre det enklere for brukerne å orientere seg best mulig om definisjoner, endringer, forbehold o.l. i datagrunnlaget. Registerdata er ikke designet for forskning, og det kan være krevende å forstå definisjoner og endringer uten rike og tilgjengelige metadata.

Styrker og svakheter

Konseptet har noen sentrale avveininger. På den ene siden ble det mulig med rask, fleksibel og bred tilgang til registerdata. På den andre siden manglet mulighet til å se data direkte. Funksjonalitet og metode var samtidig begrenset til analyseverktøyene som til enhver tid kunne gjøres tilgjengelige i det anonymiserende grensesnittet.

Fordi hovedmålet med satsingen nettopp var å senke terskelen for forskertilgang til registerdata ble det besluttet å satse på konseptet. Samtidig ble det besluttet å jobbe kontinuerlig og målrettet med implementasjon av mest mulig nødvendig funksjonalitet i samarbeid med brukermiljøene – samt å jobbe for at microdata.no kunne komplementere og støtte best mulig opp rundt tradisjonell søknadsbasert utlevering av registerdata.

Samspill med andre tilgangsformer

Microdata.no dekker mange registerdatabehov, men ikke alle. Det er fortsatt behov for tilbudet om tradisjonell søknadsbasert utlevering av data. Miljøer som driver metodeutvikling, skal koble med data som ikke ligger i microdata.no, som trenger spesialprogramvare eller metoder som ennå ikke er på plass i løsningen er avhengige av å få data utlevert på tradisjonell måte. Det betyr søknad med dokumentasjon av behandlingsgrunnlag, saksbehandling/forhandling om dataminimering, SSB-tilrettelegging av datasett, etc.

Microdata.no har blitt videreutviklet med en eksportfunksjon⁵ slik at løsningen kan strømlinjeforme og forenkle deler av den tradisjonelle søknadsprosessen.

Eksportfunksjonen gjør microdata.no til en avansert variabelutforsker der brukerne kan skape og teste seg fram til datasettene de trenger, og gjennom det får oversikt over det reelle databehovet, slik at det blir lettere å begrunne det i søknaden.

Dette fjerner en kilde til misforståelser og gir mindre arbeid for SSB fordi brukerne selv har laget/tilrettelagt datasettet. SSB henter ut de ferdige filene som brukerne selv har laget. Det er bra for både leveringstid og pris.



Videreutvikling i samarbeid med helseregistre

SSBs registerdatatilfang inneholder hovedsakelig demografiske og sosio-økonomiske data, mens helseregisterdata ligger i egne helseregistre.

De sentrale helseregistrene er også etablert for å ivareta landsomfattende oppgaver. De brukes først og fremst til helseanalyser, statistikk, kvalitetsforbedring av helsetjenester, forskning, administrasjon, styring og beredskap. Helseregistrene er opprettet med hjemmel i helseregisterloven og forskrifter. Register med direkte personidentifiserbare opplysninger som ikke er basert på samtykke, er opprettet etter behandling i Stortinget. Fra 1.1.2024 samles alle de sentrale helseregistrene i FHI.

Søknad til Forskningsrådet infrastruktursatsing i 2018 med Kreftregisteret som partner

NSD og SSB inviterte Kreftregisteret til å være partner i en ny søknad til Forskningsrådets infrastruktursatsing i 2018. Kreftregisterets rolle skulle være å delta i utviklingen av et teknisk og juridisk opplegg som kunne gjøre det mulig for andre enn SSB å dele koblingsbare registerdata via microdata.no.

Prosjektet fikk navnet Microdata 2⁶ og ble tildelt 42.5 MNOK av Forskningsrådet i 2019. Prosjektoppstart var 2020, og prosjektet løper ut 2023.

Prosjektets hovedmål var altså å legge til rette for at flere dataeiere kunne dele data via plattformen. Dette arbeidet har krevd mye teknisk utvikling og en rekke juridiske avklaringer, og målsetningen i prosjektet oppnås i den forstand at infrastrukturen høsten 2023 blir klar for at andre dataeiere kan dele sine data der. Andre delmål handlet om å øke mengden SSB-data, videreutvikle analysefunksjonalitet, pilotere mulighet for å koble forskerinnsamlede surveydata med registerdata.

Helseanalyseplattformen og gjenbrukssporet

Parallelt med utviklingen av microdata.no kjørte Direktoratet for e-helse Helsedataprogrammet. En sentral komponent her skulle være Helseanalyseplattformen (HAP). Tanken med HAP var å bygge en teknisk løsning for mer effektiv deling av helseregisterdata, samt en organisasjon (Helsedataservice/HDS) for å sikre «én vei inn» og sentralisert søknadsbehandling for forskere med behov for helseregisterdata.

Organisasjonen HDS ble etablert i 2021. Men den tekniske løsningen HAP ble skrotet⁷ i 2023. Dette gjorde at Direktoratet for e-helse måtte se seg om etter eksisterende nasjonale løsninger som kunne bidra til å oppfylle flest mulig av målene med HAP. Det såkalte «gjenbrukssporet» ble åpnet, og direktoratet etablerte dialog med etablerte tekniske løsninger (TSD/UiO, SAFE/UiB, HUNT Cloud/NTNU) samt microdata.no (Sikt og SSB). Intensjonen bak gjenbrukssporet å både å gjenbruke delresultater fra Helsedataprogrammet, men ikke minst å utnytte etablerte og modne løsninger for tilgang til og bruk av sensitive data på norsk jord. Gjenbruksaktørene (herunder Sikt/SSB) har pekt på at det er et stort potensial i de tjenestene som eksisterer, men at det krever ressurser å utnytte potentialet og få tjenestene til å spille sammen og gå opp i en høyere enhet sammen med HDS og SSB som hver seg forvalter registergrunnlaget.

Gjenbrukssporet og ulike aktiviteter igangsatt av Direktoratet for e-helse har gitt microdata.no økt synlighet i helseregister- og helseforskningsfeltet. Dette har skapt arenaer for dialog og grobunn for videre samarbeid mellom microdata.no og sentrale helseregistre samt andre nøkkelaktører i sektoren.

Våren 2023 ble det besluttet i revidert nasjonalbudsjett at Direktoratet for e-helse legges ned i sin nåværende form og slås sammen med Helsedirektoratet fra 1.1.2024. Samtidig ble det som nevnt også besluttet at de sentrale helseregistrene Kreftregisteret, Norsk pasientregister/Kommunalt pasient- og brukerregister (NPR/KPR) samt FHIs registre samles under FHI fra 1.1.2024.

Helseforvaltningen og ikke minst helseregisterlandskapet forandres altså i 2024. I praksis betyr dette at FHI med alle sine helseregistre sammen med SSB blir de to største aktørene på registerdata fremover.



Søknad til Forskningsrådet infrastruktursatsing i 2023 med FHI som partner

Søknadsfristen for årets utlysning er 15.11.2023. For å kunne sende søknad stilte Forskningsrådet krav om obligatorisk innsending skisse som grunnlag for søknad om forskningsinfrastrukturmidler innen 21.06.2023. Sikt, SSB og FHI sendte inn en felles skisse med arbeidstittel «Microdata.no Cross-FAIR. FAIR access to microdata across sectors and disciplines». Hovedmålsetningen for søknaden er å etablere mest mulig koblingsbare helseregisterdata på infrastrukturen, samt å fortsette å styrke løsningens funksjonalitet og relevans for både samfunns- og helseforskning.

Veien videre

Inneværende prosjekt Microdata 2 løper ut 2023. Innspurten i prosjektet handler om å slutføre arbeidet med etableringen av microdata.no som plattform for deling av data for Kreftregisteret og andre relevante dataeiere. Planen er å tilgjengeliggjøre data fra Kreftregisteret på infrastrukturen før året er omme. Det gjenstående arbeidet vil fokusere på sikkerhetsarbeid, slutføring av ROS-analyse samt endelig tilrettelegging av data fra Kreftregisterets side.

Utforming og forankring av søknad til Forskningsrådets infrastruktursatsing 15.11.2023 er høyt prioritert. Sikt er vertsorganisasjon for søknaden.

Finansiering etter 2023

Kunnskapsdepartementet (KD) satte høsten 2021 ned et utvalg som skulle vurdere og foreslå løsninger til videreutviklingen av datainfrastruktur i Norge. Dette infrastrukturutvalget overleverte en rapport til KD 10. mai 2022. Finansieringsmodellen for microdata.no fremover utformes i tråd med utvalgets anbefaling om en tredeling:

1. **Langsiktig finansiering** – stabil og forutsigbar finansiering.

Finansieringen kan for eksempel komme i form av grunnfinansiering fra ett relevant departement eller fra flere departementer etter en fordelingsnøkkel basert på hvilke forvaltningsområder infrastrukturen leverer forskning og forvaltning til. Slik finansiering kan også komme i form av spesifikke forvaltningsoppdrag eller gjennom forpliktende medlemskap eller partnerskap fra institusjoner som bruker infrastrukturen.

Langsiktig finansiering skal dekke kontinuerlig videreutvikling i dialog med brukermiljøene for å sikre relevans (adaptivt vedlikehold), herunder nødvendig videreutvikling av funksjonalitet i tråd med teknologiske endringer, og endrete brukerbehov. Utnytte ny teknologi og metoder for kontinuerlig å styrke analysekapabilitet, teknisk kapasitet og informasjonssikkerhet kombinert.

Sikt og SSB har bedt Kunnskapsdepartementet og Finansdepartementet, respektive, om langsiktig finansiering fra 2024. Departementenes beslutning kommer i statsbudsjettet for 2024.

2. **Konkurransesatt finansiering** – finansiering som et resultat av konkurranse mellom flere aktører som vurderes etter gitte kriterier, for eksempel Forskningsrådets infrastrukturutlysning eller EUs rammeprogrammer.

Konkurransesatt finansiering (ofte omtalt som prosjektf finansiering) gjelder typisk større løft/innovasjoner som påvirkes av endringer i teknologi-, data- eller brukerlandskapet. Eksempler kan være nye typer data, fødererte analyser med internasjonale datapartnere, støtte for sanntidsdata, maskinlæring etc.

Her deltar Sikt/SSB/FHI altså med en felles søknad knyttet til microdata.no til årets infrastrukturutlysning fra Forskningsrådet.

1. **Brukerfinansiering eller brukerbetaling** – finansiering som kommer fra de som bruker tjenestene. Dette kan enten være i form av abonnement, stykkpris eller



forbruksfinansiering, eller en kombinasjon av disse.

Brukerbetalingen skal dekke kostnadene bruken påfører infrastruktureierne. Noen kostnader påføres av enkeltbrukere, andre av samlet bruk. Det er hensiktsmessig å dele bildet i 3:

Korrektivt vedlikehold

Feilretting, håndtering av feilsituasjoner skapt gjennom bruk, gjenoppretting.

Ytelsesrelatert vedlikehold

For å håndtere økt bruk, økte datamengder, tyngre analysemetoder, effektiv og rettferdig fordeling av beregningsressurser, oppskalering av hardwareressurser etc.

Rene brukerrelaterte driftskostnader (support, opplæring og rapportering)

Kostnader knyttet til direkte brukersupport, dokumentasjon, on-boarding, opplæring, tilbakerapportering av bruken av data til dataeier og til brukerorganisasjoner.

Sikt og SSB jobber med å ferdigstille en pris- og betalingsmodell for bruk av tjenesten.

Prismodellen for brukerbetaling av microdata.no skal dekke brukerrelaterte kostnader, og er utformet med utgangspunkt i følgende kriterier:

Prismodellen skal:

- Være enkel å forstå og forklare
- Være transparent og mest mulig forutsigbar
- Ikke påføre institusjonene unødig høye kostnader
- Ikke påføre institusjonene unødig merarbeid og administrasjon
- Dekke brukerrelaterte kostnader for infrastruktureierne
- Bidra til god kapasitetsplanlegging for infrastruktureierne
- Gi infrastruktureierne incitament til å øke bruk og nytteverdi av infrastrukturen

Pris- og betalingsmodell sirkuleres til de 105 eksisterende brukerinstitusjonene så snart den er klar. Betaling innføres fra 2024.