

---

# TECHNICAL REPORT

---

## SEMICOLON: SSB CASE

MIKRODATA TIL FORSKERE



REPORT No 2008-0997

REVISION No 1.1A

DET NORSKE VERITAS

# TECHNICAL REPORT

|                                                        |                                                                     |                                                                                                                                                |
|--------------------------------------------------------|---------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| Date of first issue:<br><b>2008-10-20</b>              | Project No:<br><b>913A0304 Semicolon</b>                            | DET NORSKE VERITAS AS<br><br>1322 Høvik<br>Norway<br>Tel:<br>Fax:<br><a href="http://www.dnv.com">http://www.dnv.com</a><br>NO 945 748 931 MVA |
| Approved by:<br><b>Jon Ølnes</b><br>Principal Engineer | Organisational unit:<br><b>Information Processes and Technology</b> |                                                                                                                                                |
| Client:<br><b>Semicolon: SSB case</b>                  | Client ref.:                                                        |                                                                                                                                                |

Summary:

During the last ten years there has been an increase in research based on micro data (data about individuals like persons or companies). SSB has established a set of repositories of such data to solve and improve their production of statistical analysis. External research institutes have also shown great interest in using these micro data in their research. This report describes how the collaboration between SSB and these research institutes is done as is. Some areas identified as challenging are: (i) time spent on establishing necessary application forms and licenses to use these micro data, (ii) pricing of public data, (iii) balancing SSB roles as data repository owner and research competitor, and (iv) getting a good overview of which data exists and make sure these data are properly understood. The report suggests some activities and directions on how to improve current situation.

|                                                                     |                             |                               |
|---------------------------------------------------------------------|-----------------------------|-------------------------------|
| Report No:<br><b>2008-0997</b>                                      | Subject Group:              |                               |
| Report title:<br><b>Mikrodata til forskere</b>                      |                             |                               |
| Work carried out by:<br><b>Per Myrseth og Kjell Tore Guttormsen</b> |                             |                               |
| Work verified by:<br><b>Terje Grimstad</b>                          |                             |                               |
| Date of this revision:<br><b>2009-05-16</b>                         | Revision No:<br><b>1.1a</b> | Number of pages:<br><b>32</b> |

|                                                                                                                                                                                                                                                                                             |                                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------|
| <b>Indexing terms</b>                                                                                                                                                                                                                                                                       |                                                  |
| Keywords<br><b>organisational interoperability, semantic interoperability, statistics, micro data, research</b>                                                                                                                                                                             | Service Area<br><b>Interoperability services</b> |
|                                                                                                                                                                                                                                                                                             | Market Sector<br><b>Public sector</b>            |
| <input checked="" type="checkbox"/> Unrestricted distribution (internal and external)<br><input type="checkbox"/> Unrestricted distribution within DNV<br><input type="checkbox"/> Limited distribution within DNV after 3 years<br><input type="checkbox"/> No distribution (confidential) |                                                  |

© 2002 Det Norske Veritas AS  
All rights reserved. This publication or parts thereof may not be reproduced or transmitted in any form or by any means, including photocopying or recording, without the prior written consent of Det Norske Veritas AS.

# TECHNICAL REPORT

## Table of Contents

|                                                                                | <i>Page</i> |
|--------------------------------------------------------------------------------|-------------|
| 1 SAMMENDRAG .....                                                             | 1           |
| 2 INTRODUKSJON .....                                                           | 1           |
| 2.1 Om Semicolonprosjektet .....                                               | 1           |
| 2.2 Bakgrunn og målsetninger for caset Mikrodata til forskere .....            | 2           |
| 2.3 Målgruppe for rapporten .....                                              | 3           |
| 2.4 Leseveiledning, avgrensning og dette dokumentets rolle .....               | 3           |
| 2.5 Veien videre .....                                                         | 4           |
| 2.6 Intervjuer og kilder .....                                                 | 4           |
| 3 GENERELT OM MIKRODATA TIL FORSKERE .....                                     | 5           |
| 3.1 Innledning .....                                                           | 5           |
| 3.2 Interne, relaterte prosjekter .....                                        | 5           |
| 4 NÅSITUASJON, MIKRODATA TIL FORSKERE .....                                    | 6           |
| 4.1 Samhandling og samarbeid sett fra et forskningsinstitutt .....             | 7           |
| 4.1.1 Tilbudsfasen .....                                                       | 7           |
| 4.1.2 Prosjektgjennomføringsfasen .....                                        | 9           |
| 4.2 Kompetanse og holdninger .....                                             | 12          |
| 4.3 Meldeplikt, konsesjoner og jus .....                                       | 13          |
| 4.4 Organisering av offentlig sektor .....                                     | 15          |
| 4.5 Samarbeid mellom SSB og NSD med betydning for forskningsinstituttene ..... | 15          |
| 4.6 Online tilgang og analyser av mikrodata .....                              | 16          |
| 4.7 SSBs informasjon om metadata .....                                         | 19          |
| 4.8 SSBs informasjon om mikrodata .....                                        | 20          |
| 4.9 Nasjonal oversikt over mikrodata .....                                     | 20          |
| 4.10 SSB som underleverandør; leveransepunktligheit og prising .....           | 21          |
| 4.11 Survey koblet med mikrodata .....                                         | 21          |
| 4.12 Egenskaper ved mikrodata som utleveres .....                              | 22          |
| 4.13 Begreper, virkelighetsforståelse, semantikk .....                         | 22          |
| 4.14 Tidsperspektiv og ressursbruk .....                                       | 24          |
| 4.15 Prising og rettigheter til betalte uttrekk av mikrodata .....             | 24          |
| 4.16 Samarbeid om forskning og konkurranseforhold .....                        | 25          |
| 5 ØNSKET SITUASJON .....                                                       | 25          |
| 6 ”BESTE PRAKSIS” OG ”STATE OF THE ART” .....                                  | 26          |
| 7 KONKLUSJON .....                                                             | 27          |
| 8 REFERANSER, AKRONYMER OG TERMINOLOGI .....                                   | 28          |
| 9 APPENDIKS: SEMICOLON FORSKNINGSMETODE FOR CASE .....                         | 31          |

---

## TECHNICAL REPORT

---

### 1 SAMMENDRAG

Det siste tiår har det vært en økning i forskning basert på mikrodata. Økning i bruk skyldes flere momenter, men viktigst er kanskje at SSBs tilbud knyttet til mikrodata har blitt utvidet. De mikrodata SSB har, er i hovedsak data om de enkelte personer, husholdninger og bedrifter. Som offer for egen suksess får SSB utfordringer knyttet til økt etterspørsel og levering av nye tjenester spesielt når bruk, utveksling og lagring av disse data er underlagt en rekke krav bl.a. knyttet til personvernlovgivningen.

Hensikten med denne rapporten er å bedre SSBs forståelse for hvordan forskningsinstitutter jobber med mikrodata, og hvordan de opplever SSB som leverandør av tjenester knyttet til mikrodata. Rapporten vil bli brukt som input i SSBs forbedringsprosjekter knyttet til utlevering av mikrodata til forskere.

Årlig blir det gjennomført mange forskningsprosjekter hos forskningsinstitutter med vellykkede resultater, og hvor samarbeidet med SSB fungerer bra. Vi har i arbeidet med denne rapporten søkt å få beskrevet områder hvor det er rom for forbedring og rapporten vil derfor oppleves som problemorientert. Veien videre er at beskrivelsen av problemene skal brukes til konstruktiv videreutvikling av SSBs forvaltning og utlevering av mikrodata. Rapportens anbefalinger er rettet mot forbedringer, og noen viktige problemstillinger sett fra eksterne forskeres ståsted er:

- Forskningsprosjekter som benytter mikrodata, bruker mye tid på å få på plass nødvendige godkjenninger og konsesjoner. Med tid menes her både person-timer og lang tidsperiode.
- Pris på utlevering av mikrodata oppleves for noen forskningsformål å være for høy.
- SSBs roller som både forvalter av nasjonalt arkiv for flere typer mikrodata og forskningsinstitutt som konkurrerer i markedet, er utfordrende.
- Det er i mange tilfeller vanskelig å forstå både hvilke data som kan bestilles, og hvordan mikrodata og tilhørende variabler skal tolkes.

Denne rapporten baserer seg på møter med SSB-ansatte, litteraturstudie av i hovedsak SSB-interne dokumenter, og noen intervjuer med eksterne forskere. Noen poeng vil være knyttet til kun en respondent og formen på rapporten er slik at vi angir om kilden til et moment er ekstern forsker, fra SSB ansatte/ SSB dokumenter, eller anbefaling fra forfatterne av dokumentet.

### 2 INTRODUKSJON

#### 2.1 Om Semicolonprosjektet<sup>1</sup>

Semicolon er et brukerstyrt innovasjonsprosjekt, med delfinansiering fra Norges Forskningsråds VERDIKT-program.

Semicolon adresserer utfordringer i organisatorisk og semantisk interoperabilitet innen offentlig sektor. Problemstillingene innen semantisk interoperabilitet handler bl.a. om å etablere kompatible begrepsapparater og modeller for kartlegging av informasjon som det offentlige bruker i sin tjenesteproduksjon. Mye av informasjonen offentlig sektor benytter i sin tjenesteproduksjon, oppstår i et samspill med borgere og næringsliv.

---

<sup>1</sup> Se også [www.semicolon.no](http://www.semicolon.no)

## TECHNICAL REPORT

Prosjektets hovedmål er å utvikle og utprøve IKT-baserte metoder, verktøy og metrikker for hurtigere og billigere å oppnå semantisk og organisatorisk interoperabilitet innen offentlig sektor og mellom offentlig sektor og både næringsliv og borgere.

Semicolonprosjektets delmål er å:

- Identifisere hindringer for interoperabilitet og strategier/løsninger for å overkomme disse.
- Utvikle metoder og verktøy for å etablere effektiv interoperabilitet, samt metrikker for å kunne måle effektene (inkludert bruk av ontologier).
- Anvende metoder og verktøy for å utvikle ontologier og modeller til bruk i konkrete eksempelprosjekter i etatene, og måle effekter ved hjelp av metrikken.
- Spre resultater og erfaringer til offentlig sektor, IKT-industri og akademia.
- Lage en arena med toneangivende etater, kompetente bedrifter og forskningsmiljøer, som kan være en pådriver for semantisk og organisatorisk interoperabilitet i offentlig sektor.

Prosjektet har en varighet på 42 måneder, med formell oppstart i oktober 2007. Hovedarbeidet i prosjektet vil vare i 36 måneder, mens de siste 6 månedene vil brukes til å ferdigstille en doktorgrad. Prosjektet arbeider hovedsakelig ut fra en case-orientering, men har også definert seks faglige og/eller administrative arbeidspakker.

### Deltakerne

- Fem store offentlige organisasjoner
  - Skattedirektoratet
  - Brønnøysundregistrene
  - Helsedirektoratet
  - Statistisk sentralbyrå
  - Kommunenes sentralforbund samt Asker og Bærum kommuner
- Private og offentlige selskaper
  - Det Norske Veritas - DNV
  - Karde AS
  - eKor AS
  - Kompetansesenter for IT i helse- og sosialsektoren – KITH
- Universiteter
  - Handelshøyskolen BI (organisatorisk interoperabilitet)
  - Universitetet i Oslo (semantisk interoperabilitet, objektorientering)
  - University of Oxford (semantisk teknologi)
  - University of Aberdeen (sammenheng mellom modelleringspråk og semantikk)

## 2.2 Bakgrunn og målsetninger for caset Mikrodata til forskere

SSB har en stor og interessant samling av data om personer og bedrifter. Dette er data som SSB benytter i sin statistikkproduksjon, samtidig som disse data i økende grad gjenbrukes til forskning både av SSBs egne forskere og av eksterne forskere. Et vanlig scenario for en ekstern forsker er gjennomføringen av et forskningsprosjekt som innhenter noen mikrodata ved surveys, og noen data fra eksisterende kilder hos SSB. SSB bruker en del persontid på prosessen med å finne ut hvilke data de eksterne forskerne trenger og å få disse data ut på en for mottaker hensiktsmessig måte. Eksterne forskere og SSB har over tid forbedret rutiner og muligheter, men ser at det er rom for ytterligere forbedringer i SSBs kapasitet og rutiner knyttet til utlevering av mikrodata til forskere.

---

## TECHNICAL REPORT

---

Sett fra SSB er forskningsinstitutter gjenbrukere av SSBs mikrodata. SSB har startet interne prosjekter (bl.a. FOSS-prosjektet) som har til målsetning å forbedre og standardisere rutinene i statistikkproduksjon, herunder også løsninger knyttet til gjenbruk av mikrodata. SSB har stor kunnskap om hvilke mikrodata som finnes og hvordan disse data kan benyttes og tolkes. Denne kunnskapen er det viktig å få formidlet til eksterne forskere.

SSB forvalter et viktig nasjonalt mikrodataarkiv og ønsker bedre innsikt i hvordan eksterne forskere benytter disse mikrodata og hvordan samhandlingsprosessen med SSB oppleves sett fra forskningsinstituttene. Det er et mål å enklere tilgjengeliggjøre mikrodata til eksterne forskere, men det må gjøres innenfor praktiske, tekniske og juridiske rammer.

Ved oppstart av prosjektet hadde SSB noen tanker om mulige tiltak som kan forbedre gjenbruken av mikrodata for SSB og / eller eksterne forskere utover dagens praksis rundt utlevering av data til forskningsformål. Dette er:

- Etablering av remote access/online tilgang til mikrodata som ligger fysisk hos SSB.
- Public use files, som er anonymiserte og klargjorte filer med mikrodata, og som ikke krever konsesjoner eller godkjenninger for å kunne brukes.
- Felles søknadsskjema for konsesjoner og godkjenninger.

Punktene om online tilgang og Public use files, er ment som et forbedret alternativ og supplement til dagens utleverings praksis.

SSB har tidligere gjennomført prosjekter for å bedre forvaltning og utlevering av mikrodata til eksterne forskere. Et slik prosjekt var knyttet til et oppdrag fra Husbanken, som ønsket å få analysert ulike forhold ved boliger. Husbankens initiativ kom samtidig med at SSB hadde pågående interne prosjekter, og disse prosjektene ble samkjørt med suksess. I dette samkjørte prosjektet ble eksterne forskere involvert i prosjektets referansegruppe.

Semicolon er bedt av SSB om å intervju eksterne forskere for å belyse deres ønsker og behov knyttet til bruk av mikrodata fra SSB. Basert på disse intervjuene vil Semicolon etablere første steg i sin forskningsmetode som er å kartlegge dagens praksis, rutiner og systemer.

### 2.3 Målgruppe for rapporten

De primære målgruppene for rapporten er

1. SSB og deres interne prosjekter knyttet til forvaltning og utlevering av mikrodata.
2. Semicolonprosjektet med partnere og forskere.

### 2.4 Leseveiledning, avgrensning og dette dokumentets rolle

Dette dokumentet beskriver resultatene fra undersøkelser som har hatt til hensikt å kartlegge nåsituasjon knyttet til mikrodata. Vi har i forbindelse med dette arbeidet kun intervjuet eksterne forskere som benytter mikrodata om person, og ikke forskere som benytter mikrodata om bedrifter. Denne rapporten baserer seg på et lite antall intervjuer med eksterne forskere, og noen poeng vil være knyttet til kun en respondent.

Dette dokumentet tar ikke stilling til om enkeltforskernes vurderinger skal legges til grunn for endringer av SSBs mikrodataforvaltning og utleveringstjenester.

---

## TECHNICAL REPORT

---

Dokumentet kommer med en rekke anbefalinger som er forfatterens anbefalinger. Avsnitt og punkter merket med "Ønsket situasjon" beskriver hva eksterne forskere og SSB internt har formidlet om ønsket situasjon. Noen steder har vi merket poeng som vi tror er viktig for SSB med innrykk og "Funn: x"

En del steder i rapporten har vi beskrevet hvordan ulike aktører (både interne og eksterne) oppfatter SSB som organisasjon. Eksterne aktører kan ha et syn på SSB som ikke er sammenfallende med SSBs syn på seg selv. Disse forskjellene har vi bevisst latt stå for å illustrere at eksterne personers oppfattelse kan avvike fra SSBs eget syn, samt at personer internt i SSB kan ha forskjellig forståelse for hvordan utlevering av mikrodata foregår. Ulike syn illustrerer SSBs pedagogiske utfordring med å forklare SSBs verdikjede knyttet til mikrodataforvaltning og utlevering både for SSB interne ansatte og overfor eksterne forskere.

Dette dokumentet beskriver primært nåsituasjonen for SSB og er starten på en forskningsprosess. Dette dokumentet påvirker omfanget for Semicolons leveranser av "Beste Praksis" og "State Of The Art" dokumentene som skal utarbeides. Semicolons forskningsprosess er beskrevet i eget appendiks i dette dokumentet.

Det er foreløpig ikke avgjort om det blir et separat dokument som beskriver SSBs ønskete situasjon, eller om ønsket situasjon beskrives i en oppdatert og seinere versjon av dette dokumentet. En oppdatert versjon vil i så fall beskrive både nåsituasjon og ønsket situasjon.

### 2.5 Veien videre

Denne rapporten er første leveranse til SSB fra Semicolonprosjektet og skal bidra med kartlegging, og øke SSBs kunnskap om eksterne forskeres samhandling med byrået.

Semicolons leveranser av "State Of The Art"- og "Beste Praksis"-dokumenter er planlagt senhøsten 2008. Detaljer for disse leveransene finnes i Semicolons prosjektplan.

Fremdrift i caset "mikrodata til forskere" blir påvirket av SSBs FOSS-prosjekter og Semicolons videre dialog og avtaler med SSB.

### 2.6 Intervjuer og kilder

Intervjuer med følgende personer er foretatt i perioden juni-august 2008 listet i den rekkefølge de er blitt gjennomført:

1. Ole Sandvik, Seniorrådgiver, Avdeling for økonomi, energi og miljø, seksjon Arbeidsmarkedsstatistikk, SSB Oslo.
2. Rune Gløersen, Fagdirektør, Avdeling for IT og datafangst, SSB Oslo.
3. Lars Gulbrandsen, NOVA, Norsk institutt for forskning om oppvekst, velferd og aldring.
4. Thorleiv Valen, Jurist, Seniorrådgiver, SSB Kongsvinger.
5. Johan-Kristian Tønder, Fagdirektør, Avdeling for personstatistikk, SSB Oslo.
6. Hanne Cecilie Kavli, Forskningsleder for Arbeid, inkludering og kompetanse. FAFO, Institutt for arbeidslivs- og velferdsforskning.
7. Paul Inge Severeide, Seksjonssjef, Seksjon for befolkningsstatistikk, SSB Kongsvinger.
8. Arild Mellesdal, Spesialrådgiver, NSD, Norsk samfunnsvitenskaplige datatjeneste i Bergen.

Personene har uttalt seg som enkeltpersoner og deres erfaringer og uttalelser reflekterer nødvendigvis ikke SSBs eller forskningsinstituttet totale erfaring og syn på samarbeide med SSB.

## TECHNICAL REPORT

Vi har fått tilgang til en rekke relevante dokumenter og deltatt i flere av FOSS-prosjektenes møter. Dette har vært viktig input for å forstå hvordan SSB fungerer sett fra innsiden. Innledningsvis i prosjektet hadde vi møter med Jenny Linnerud, Sunniva Maria Solstrand og Rune Gløersen.

### 3 GENERELT OM MIKRODATA TIL FORSKERE

#### 3.1 Innledning

Det case-arbeidet som beskrives i denne rapporten skal kartlegge samhandlingen mellom SSB og forskningsinstitutter med deres forskere relatert til utveksling og bruk av mikrodata.

Med termen mikrodata menes data om enkeltpersoner, husholdninger, bedrifter/virksomheter etc. SSB får og innhenter slike data fra nasjonale offentlige registre (registerdata), pluss at SSB gjennomfører egne undersøkelser. SSB kan på forespørsel utlevere mikrodata til forskere som arbeider i forskningsinstitutter som er godkjent av SSB. SSB har muligheten til å flette mikrodata fra flere kilder, noe som gjør at SSBs mikrodata og utleveringer er bedre egnet til de fleste forskningsformål enn mikrodata hentet direkte fra enkelte etater. For at SSB skal kunne levere ut mikrodata må mottaker etablere nødvendige godkjenninger og konsesjoner.

For å illustrere hva *forskning på mikrodata* innebærer velger vi her å ta med to eksempler på mikrodata-forskningsprosjekter hos NOVA:

- Analysere hvordan det gikk med de personer som i henhold til folke- og boligtellinger leide bolig av kommuner i 2001. Basert på tilgjengelige mikrodata er det mulig å følge disse personene frem til 1. januar 2005. Analysen skulle se på mønstre for flytting, hva de flytter til etc.
- Sammenlikne borgere som var i utvalget av personer med 10% laveste inntekt i 2001 og 2004 for å se hvilke grupper av personer som beveget seg ut og inn av utvalget.

SSBs innhenting, bruk og utlevering av mikrodata er regulert av bl.a. statistikkloven og personopplysningsloven.

#### 3.2 Interne, relaterte prosjekter

Direktørmøtet i SSB har etablert et prosjekt for forbedring og standardisering av arbeidsprosessene i statistikkproduksjonen (FOSS prosjektet). Under denne paraplyen inngår også delprosjekter for å forbedre verdikjeden relatert til mikrodata. De relevant delprosjektene er

- Virksomhetsmodell. Utarbeiding av en virksomhetsmodell som beskriver standardiserte arbeidsprosesser i statistikkproduksjonen.
- Kartleggingsprosjektet. Skal kartlegge:
  - o SSBs nåværende rutiner for utlevering av mikrodata til forskere.
  - o Behov for bestillings- og uttakssystem.
  - o Forskernes behov for arkivering, gjenbruk av og tilgang til data, basert på erfaringer internt i SSB.
- Arkivering. Prosjekt som skal bl.a. kartlegge forskernes behov for arkivering og gjenbruk av data, samt vurdere behov for og utarbeide opplegg for teknisk løsning for arkivering av utleverte mikrodata i SSB.
- Tilgang. Kartlegge rutiner for utlevering av mikrodata i andre institusjoner og vurdere alternative metoder for utlevering fra SSB.



---

## TECHNICAL REPORT

---

- FORUM. Det skal utvikles et system for uttak av forløpsdata (internt i SSB), og lages en beskrivelse av mulighetene for videre utnytting av systemet i SSB.
- Felles inngang til SSBs mastersystemer for metadata (metadata portal).

### 4 NÅSITUASJON, MIKRODATA TIL FORSKERE,

Dette kapittelet baserer seg på informasjon innhentet fra SSB internt, SSBs web-sider, SSB-interne møter og en serie intervjuer med eksterne og SSB-interne personer. Se også referanselisten i dette dokumentet.

Etterspørselen etter mikrodata til forskningsformål ut over SSBs egen bruk har vært økende. Et grovt overslag er at SSB bruker til sammen ca. 10 årsverk på utlevering av mikrodata til forskere. Disse årsverkene omfatter: dialog med eksterne forskere, design av uttrekk, bistand til tolkning av data, konsesjonsrådgivning, gjennomføring av uttrekk, aktiviteter knyttet til arkivering og sletting av utleverte data. Antall årsverk knyttet til innhenting, kvalitetssikring, forvaltning m.m. av mikrodata er en vesentlig del av SSB totale årsverk. SSB er i underkant av 1000 ansatte totalt i Oslo og Kongsvinger[2].

Noen typer mikrodata finnes kun hos SSB. SSB har opparbeidet stor og anerkjent kompetanse på hvordan mikrodata og tilhørende variabler skal tolkes og hvordan disse data kan benyttes for analyser. I motsetning til meningsmålingsinstitutter har SSB mulighet til å koble data mellom registre, og mellom registre og surveys. I den grad noen datakilder skal kobles med mikrodata fra SSB, foretas koblingen av SSB, og resultatet avidentifiseres/anonymiseres ved utlevering.

Sett fra forskningsinstituttene har SSB en monopolsituasjon som dataleverandør på en rekke typer mikrodata. SSB har en liste med godkjente forskningsinstitutter som de kan levere data til. Med godkjent her menes at forskningsinstituttens prosesser for forvaltning og oppbevaring av mikrodata er vurdert og funnet tilstrekkelige. Videre må forskningsinstituttene før hver utlevering av mikrodata fremvise dokumentasjon på formål, og nødvendige godkjenninger og konsesjoner må være på plass før selve utleveringen kan finne sted. SSBs kapasitet til utleveringer av mikrodata varierer, og det er eksempler på leveringer som tar 6 måneder.

SSB er underleverandør inn i et forskningsprosjekt hos de eksterne forskningsinstituttene. Slike forskningsprosjekt er ofte finansiert på fast pris og med stramme tidsfrister. SSB er tydelig i sin dialog med eksterne at leveringsfrister til tider kan være lange, og det bestrebes at det som estimeres og avtales blir levert.

Forskningsinstituttene har flere kategorier prosjekter / situasjoner som benytter mikrodata fra SSB. Eksempler på slike er:

- Ekstern oppdragsgiver utlyser et anbud/ ber om tilbud fra forskningsinstituttet på leveranse av forskningsoppdrag. (anbuds og tilbudsprosess).
- Doktorgradsarbeide som krever egne mikrodatasett.
- Forskningsinstituttet tar selv initiativ til og søker om prosjektmidler fra eksempel Norges forskningsråd.

I det videre arbeidet har vi fokusert mest på situasjonen hvor forskningsinstituttet er i en konkurranseutsatt situasjon og gjør et oppdrag basert på en anbudsutlysning/konkurranse.

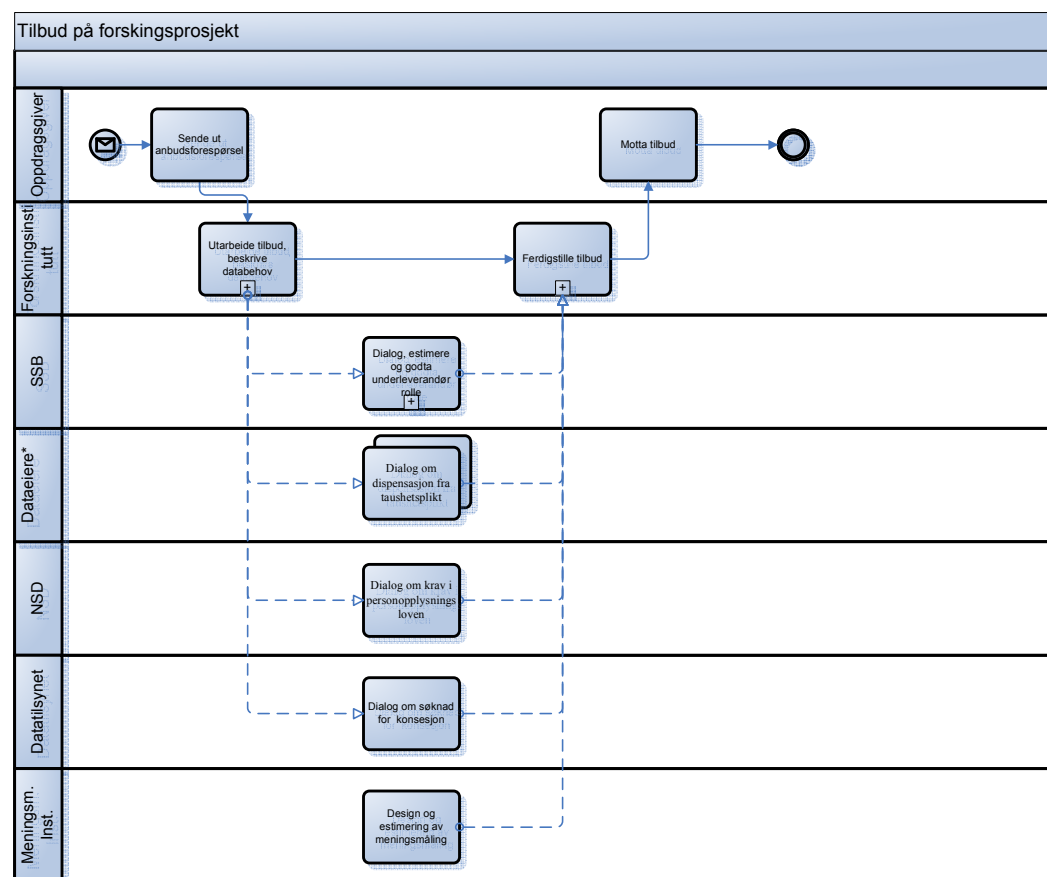
## TECHNICAL REPORT

### 4.1 Samhandling og samarbeid sett fra et forskningsinstitutt

Når et forskningsinstitutt utarbeider et tilbud eller planlegger et prosjekt, vil ofte SSB bli forespurt om å estimere kostnad og leveringstid for mikrodata samt å gi bistand med å velge hvilke variabler som skal benyttes i gjennomføringen av forskningsprosjektet. Sett fra forskningsinstituttet blir SSB en underleverandør. Et viktig tiltak fra SSBs side for å håndtere denne underleverandørrollen er at det tildeles én kontaktperson i SSB som følger opp utleveringsprosjekter. Dette synes spesielt viktig når mikrodata fra flere kilder skal kobles.

Samhandlingen mellom forskningsinstituttet og SSB som beskrives under, er delt opp i to hoveddeler, tilbudsfasen og prosjektgjennomføringsfasen.

#### 4.1.1 Tilbudsfasen



Figur 1, Tilbud på forskningsprosjekt. (\*Dataeiere er i flertall fordi mikrodata som flettes kan kreve dialog og godkjenning fra flere av involverte dataeierne.)

Forklaring til symboler i figuren:

- Tegnet "\*" knyttet til dataeier er brukt for å illustrere at flere dataeiere vanligvis vil være involvert.
- Heltrukket linje betyr kontrollflyt, brukt i diagrammet for flyt mellom aktiviteter for banen til forskningsinstituttet.
- Stiplet linje betyr utveksling av melding/data.

Figuren over er en grov beskrivelse av aktiviteter forskningsinstituttet går gjennom når en oppdragsgiver utlyser et anbud, eller når forskningsinstituttet selv designer et prosjekt. Om

---

## TECHNICAL REPORT

---

forskningsinstituttet vinner anbudskonkurransen eller på annen måte har finansiering og er i posisjon til å starte, vil estimater og beskrivelser fra tilbudsfasen bli benyttet i den videre dialog med SSB. For forskningsinstituttet er det tidkrevende å håndtere godkjenninger, vurdering om et prosjekt er meldepliktig eller konsesjonspliktig, og søknad om konsesjon.

Det er også eksempler på at SSBs forskere er en forskingspartner i prosjekter ledet av eksterne institutter, men for anledningen anser vi dette som et unntak fra den vanlige underleverandørrollen.

Mikrodata som benyttes i forskningsprosjekter kan komme fra:

1. Forskningsinstituttet selv/ tidligere forskningsprosjekter (kan kreve at de har konsesjon til gjenbruk, eller skaffer seg dette).
2. NSD (Norsk samfunnsvitenskaplige datatjeneste).
3. Etater/Dataeiere; ofte i form av registerdata.
4. SSB.
5. Meningsmålinger/ Surveys.

I tillegg benyttes åpne statistikkilder og SSBs publiserte statistikker. Mikrodata fra disse kildene hentes fra utvalgsundersøkelser, i form av tverrsnittdata eller forløpsdata. Tilbakemelding fra forskerne som er intervjuet er at SSB oppfyller ønsker og behov når det gjelder format på de data som leveres.

For å få gjennomført et forskningsprosjekt som baserer seg på mikrodata må et forskningsinstitutt gjennom en rekke aktiviteter. Aktivitetene under tar som utgangspunkt at mikrodata fra mer enn en kilde benyttes og at disse kildene må kobles sammen. Hver av aktivitetene som er listet under, har samhandling (input/output) med andre virksomheter. Rekkefølgen er delvis sekvensiell, men med en del parallelle aktiviteter:

- Utarbeide tilbud og levere svar på anbud/ evt. etablere et internt prosjekt ved forskningsinstituttet.
  - Omfatter dialog med SSB om hvilke variabler som trengs.
  - Pris og leveringsestimat fra SSB.
- Initielt design og beskrivelse av behov for mikrodata, bruk av web hos SSB og NSD, epost og telefoner med fagpersoner hos SSB og NSD evt. andre.
- Identifisere nødvendige datakilder hos dataeiere, NSD, SSB, egne tidligere datasett etc.
- Identifiser eierforhold til ulike datakilder, saksbehandler og personvernombud som er nødvendig for å etablere konsesjoner og uttrekk fra de enkelte kilder.
- Avklare om prosjektet er meldepliktig eller trenger konsesjon.
- Innhente tilbud på mikrodata.

Aktivitetene over inngår i tilbudsfasen for å kunne levere et best mulig tilbud. Aktivitetene vil ofte helt eller delvis gjentas hvis prosjektet blir finansiert og startet.

Anbefaling:

- Konkretisere hva slags bistand som tilbys fra SSB i tilbudsfasen
- Beskrive eventuelle kriterier som må være oppfylt for at SSB skal kunne gi et fastpristilbud

---

## TECHNICAL REPORT

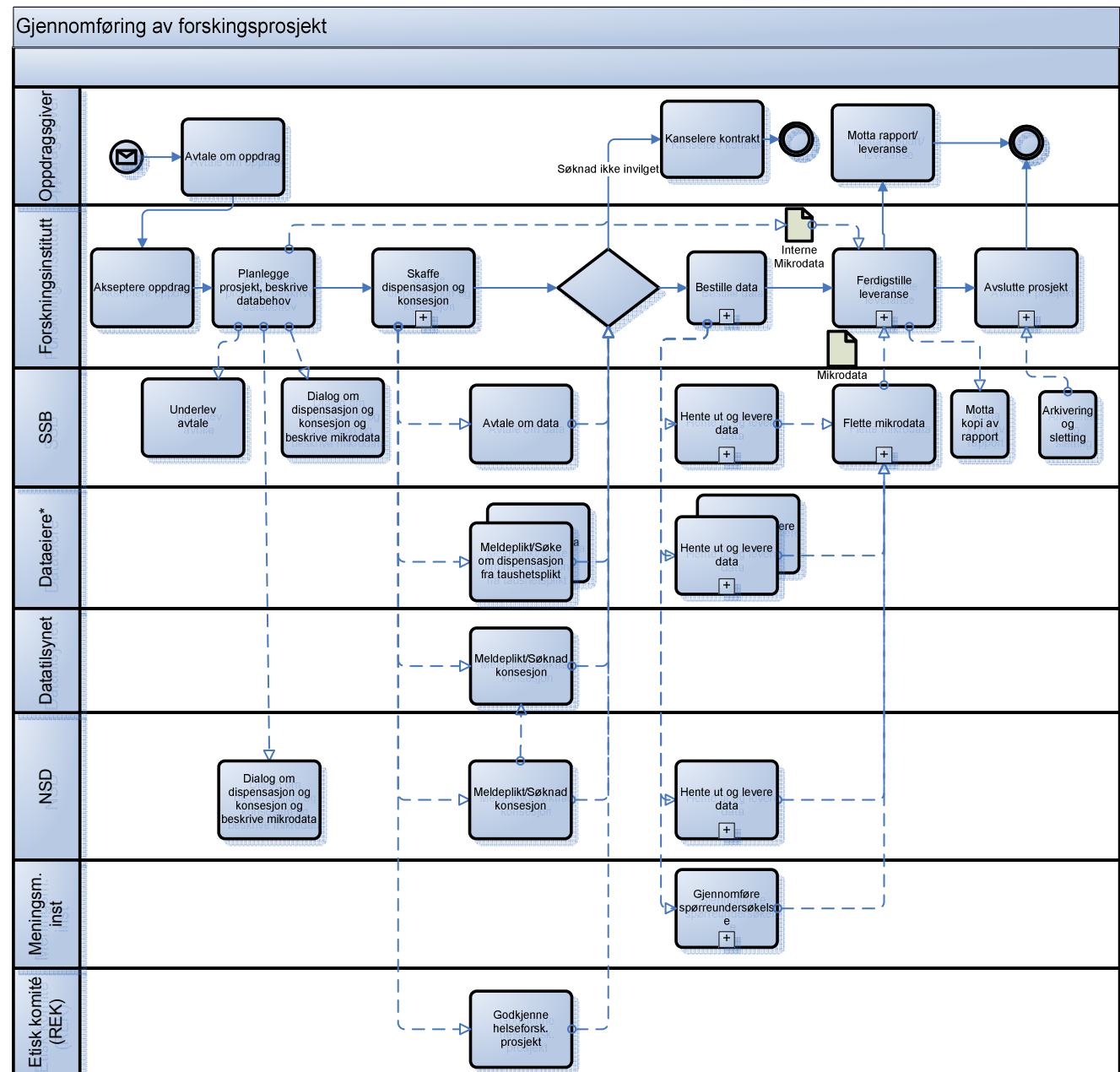
---

### 4.1.2 Prosjektgjennomføringsfasen

Aktivitetene under er forbeholdt prosjektgjennomføringsfasen:

- Akseptere oppdraget, omfatter avtaleinngåelse med oppdragsgiver.
- Planlegge prosjektet i detalj og bestemme hvilke datakilder og variabler som ønskes.
- Sende søknad om godkjenning, konsesjon; evt. å utføre meldeplikter.
- Inngå avtaler om leveranser av mikrodata.
- Informere om gyldige konsesjoner / kun meldepliktig, hvem som er ansvarlig for data og hvem som er brukere av data.
- Innhente og koble mikrodata.
- Ferdigstille forskningsresultater.
- Publisering av resultater.

## TECHNICAL REPORT



**Figur 2, Gjennomføring av forskningsprosjekt**

Figur 2 viser hovedløpet for forskningsprosjekter som benytter konsesjonsbelagte og sammensatte mikrodata. Aktiviteter som er tatt med er de som har ekstern kommunikasjon sett fra forskningsinstituttet. Noen av aktivitetene i diagrammet gjelder kun for spesielle situasjoner. Regional etisk komité for medisinsk forskning (REK) skal vurderer prosjekter som har en helsefaglig vinkling og mikrodata inneholder helseopplysninger. For andre typer prosjekter er ikke REK involvert. Det vil si at figuren viser kun noen valg av prosess-steg, og få av de mulige unntak og spesialtilfeller som kan oppstå er tatt med.

Hvis data fra et pseudonymt register skal benyttes, må TPF også involveres i prosessen (tiltrodde pseudonymforvalter). Dette er ikke vist i figuren over. Pseudonymt her betyr at registeret er avidentifisert. Eksempel på slikt pseudonymt register er IPLOS. IPLOS er betegnelsen på et sentralt

## TECHNICAL REPORT

helseregister som skal danne grunnlag for nasjonal statistikk for pleie- og omsorgssektoren. For mer informasjon se: <http://www.shdir.no/iplos>.

Når man har fått prosjektet, tas det nærmere kontakt med SSB. I denne fasen trengs bistand til å skrive en bestilling av variabler, mikrodatakilder, utvalg og koblinger som er så presis som mulig. Kontaktperson i SSB hjelper forskeren med å velge de beste kildene som er mest aktuelle i forhold til forskerens problemstillinger.

Når SSB vet hvilke variabler som trengs, finner de relevante mikrodatakilder og dermed også hvilke registreiere forskeren må henvende seg til. SSB kan ikke levere ut data uten at nødvendige godkjenninger og konsesjon foreligger evt. at det er beskjed fra Datatilsynet om at de godtar at prosjektet kun er meldepliktig.

For noen registeropplysninger har SSB fullmakt til selv å avgjøre dispensasjon fra taushetsplikt, men for eksempel ved bruk av folkeregisteret må sentralkontoret for folkeregistrering (SKD) godkjenne at SSB kan utgi ikke-anonyme data. Skal forskere ha tilgang til data fra Folkehelsen eller Kreftregisteret må de ta kontakt direkte med registreierne.

De aller fleste forskningsinstitusjoner benytter NSD som personvernombud, men ikke alle. Det vil si at vanligvis må forskere gå til NSD for å få avklart om et prosjekt er konsesjonspliktig eller ikke. Basert på prosjektdokumentasjon vurderer NSD prosjektet. Når vurderingen er gjort, tar NSD kontakt med forskningsinstituttet og sender evt. en søknad til Datatilsynet om tilrådning/ konsesjon. Om prosjektet ikke er konsesjonspliktig kan NSD gi godkjenning til at prosjektet starter.

I en del forskningsprosjekter vil det vært nyttig å enkelt kunne forlenge godkjenninger og konsesjon til å benytte data. Praksis i dag er at aktivitetene med godkjenninger og konsesjon kjøres en ny runde. Om denne runden er mindre tidkrevende har vi ikke kunne konkludere med basert på den input vi har per i dag.

Hvis mikrodata fra flere kilder skal kobles og minst en av disse kildene forvaltes av SSB, vil denne koblingen gjøres hos SSB.

### **Kommentarer til prosjektgjennomføringsfasen:**

Forutsigbarhet i forhold til SSBs leveringer og kostnader er viktig. I en tilbuds- og prosjektgjennomføringsfase som vist i figurene over er det viktig for forskningsinstituttet at de kan utarbeide tilbud med forvisning om at tid- og krone-estimer fra SSB blir som avtalt.

En sårbar og viktig fase for et vellykket forskningsprosjekt er dialogen mellom forskeren og SSB. Denne fasen leder fram til valg av datakilder, utvalg og variabler. Sett fra SSB kan det av ressurs hensyn være ønskelig å effektivisere denne type personkontakt. Tilbakemelding fra eksterne er at personkontakt trolig kan reduseres når (i) seniorforskere etterspør data som de har stor kunnskap om, (ii) eller når forskningsinstituttet etterspør supplerende tidsserier.

Funn: Hvor mye ressurser SSB bruker i disse tilfellene og hvor mye som kan reduseres, er ikke målt, og mulig gevinst er ikke estimert.

Tilbudsskriving er ofte en aktivitet med tidspress og usikkerheter knyttet til om innsatsen med tilbudsarbeidet blir "betalt" med å vinne prosjektet. Dermed er det ønskelig å minimere ressursbruk. Basis for SSBs estimer er beskrivelsen av hvilke variabler som inngår, og første versjon av beskrivelsen blir laget i tilbudsfasen. Når prosjektet starter opp blir disse beskrivelsene ofte bearbeidet videre, og hvis forskningsinstituttet gjør større endring i beskrivelsen av hvilke variabler som inngår i

## TECHNICAL REPORT

uttrekket, er det også forståelse for at kostnad og tidsestimater fra SSB må kunne endres. Forskningsinstitutter har en del prosjekter med fastpris overfor sine oppdragsgivere, og en eventuell kostnadsøkning fra SSB som underleverandør vil kunne gjøre et ellers økonomisk vellykket prosjekt om til et underskuddsforetakende. En annen situasjon er om bestillingen fra tilbudsfasen ikke endres, men SSB får en kostnadssprekk i forhold til sitt eget estimat.

### Anbefaling:

- Konkretisere hva som inngår i den betalte bistand fra SSB som tilbys i prosjektfasen
- Konkretisere hva som er forutsetningene for SSBs estimer, og hva som er konsekvenser ved ulike typer endringer
- Avklare hvem som må bære ulike typer kostnadsendringer og hvorfor

## 4.2 Kompetanse og holdninger

Det er vanlig at nyutdannede i et forskningsinstitutt får den nødvendige opplæringen knyttet til bruk og bestilling av mikrodata av en senior forsker. Opplæringen omfatter hva slags data som finnes hos SSB, og hvordan en går frem for å formulere en søknad om konsesjon.

Ideelt sett burde forskningsinstitutter trenge lite kompetanse om SSB-interne forhold for å kunne fungere godt. I dialogen med eksterne forskere nevnes det at i samhandling med SSB og bruk av mikrodata bør en god forsker ha kompetanse om:

- Hvilke mikrodata som finnes og hvor
- Hva ulike variabler betyr
- SSBs ansatte og organisasjon
- Søknad og konsesjonsprosessen for de ulike typer mikrodata

Fra utsiden oppfattes mikrodata fra SSB som en stor samling data som henger mer eller mindre sammen. Mikrodata forvaltes i flere SSB-interne systemer og etter ulike prinsipper. Denne kompleksiteten bør eksterne forskere slippe å forholde seg til. De begrensninger som finnes i SSBs interne systemer, koblingsmuligheter mellom datakilder og utfordringer med hvordan koblede mikrodatasett skal kunne tolkes, er et komplekst område og forbeholdt spesielt interesserte.

For å øke sin kompetanse om hvilke data som finnes hos SSB og hvordan disse kan benyttes til forskning, deltar forskningsinstitutter på SSB-seminarer og brukergrupper.

Har forskningsinstituttet god tid vil dialogen være en kombinasjon av telefon og e-post. Møter benyttes for kompliserte uttrekk og hvis koblinger av data er spesielt vanskelig. Noen ganger tar SSB initiativ til slike møter.

Er den eksterne forskeren skolert i SSBs mikrodata, sender forskeren over forslag til variabelliste for noen tidsperioder. Det er imidlertid sjeldent forskeren kjenner SSBs mikrodata så godt at dette lar seg gjennomføre. Dessuten kan det være endringer på både hvilke variable som finnes i ulike mikrodatasett og hva ulike variabler betyr, så som oftest trenger man hjelp fra fagpersoner hos SSB for å sikre at en får de best mulige mikrodata i forhold til forskningsprosjektets målsetninger.

I samarbeid med SSB er det stor forskjell på å være en ny forsker versus det å være en forsker med ti års erfaring. Nyansatte følges tett opp internt hos forskningsinstituttet og gis maler, og alt som sendes ut av huset blir kvalitetssikret. Opplæringen av nyansatte ønsker forskningsinstituttene å gjennomføre selv, men det er nyttig å se sjekklister på sidene til SSB.

---

## TECHNICAL REPORT

---

### 4.3 Meldeplikt, konsesjoner og jus

NSD (Norsk samfunnsvitenskaplige datatjeneste) er personvernombud for de fleste forskningsinstitutter. Prosjektbeskrivelsen sendes fra forskningsinstituttet til NSD, og der vurderes prosjektet for å finne ut om ønsket mikrodatasett er personsensitive og dermed om prosjektet er meldepliktig eller konsesjonspliktig.

Vurdering tar for seg de personvernmessige forhold knyttet til forskningsprosjekt, om de har hjemmelsgrunnlag i personopplysningsloven til å gjennomføre det, om ev. samtykket er vidt nok, om det kreves konsesjon, og om unntaket fra konsesjonsplikten vil gjelde. Dersom prosjektet er konsesjonspliktig, oversendes saken til Datatilsynet med NSDs vurdering av saken. Om et prosjekt ikke får konsesjon til å benytte sensitive personopplysning, kan det ankes til Personvernemnda som kan overprøve Datatilsynets vurdering.

Sensitiv personopplysning er definert i personopplysningslovens paragraf 2 nr 8, og kort dreier det seg om opplysninger knyttet til helse, kriminalitet, etnisitet, osv.

De som ikke benytter NSD, har enten eget personvernombud eller et eksternt ombud - eller de har ikke ombud og forholder seg direkte til Datatilsynet.

Forsker må ha dispensasjon fra taushetsplikt for å få tilgang til ikke-anonyme registeropplysninger. Ofte vil det være slik at det er overordnet organ og ikke dataeier selv som skal behandle søknad om dispensasjon fra taushetsplikt i henhold til forvaltningsloven § 13 d. Noen ganger er dette delegert til dataeier selv, hovedregel er at overordnet organ skal behandle saken. Ved kobling av ulike datakilder vil det ofte være behov for dispensasjon fra flere dataeiere, noe som kan medføre mer tidsbruk enn planlagt.

Forskningsinstitutter opplever at utfallet av NSDs vurdering kan være vanskelig å forutsi. Hvis konsesjon er nødvendig må Datatilsynet gi konsesjon, og det tar noen uker ekstra.

Når avklaring om meldeplikt evt. konsesjoner er på plass går dialogen med SSB raskt. Kontakten fra tilbudsfasen tas opp igjen og SSB blir bedt om å utlevere data.

SSB har hjemmel i statistikkloven til å forske på mikrodata som er innhentet i henhold til lovens formål. Slik forskning er meldepliktig til SSBs eget personvernombud. Hvis SSB-forsker henter inn eksterne data som faller utenfor statistikklovens ramme, og disse data utløser konsesjonsplikt, må SSBs forskere innhente konsesjon. Dette gjelder i første rekke taushetsbelagte helseopplysninger, herunder data fra de sentrale helseregistrene (f.eks. dødsårsaksregisteret).

SSB har etablert en liste over godkjente forskningsinstitutter som kan søke om å få utlevert mikrodata. Utlevering til andre aktører forutsetter at søker kommer inn under unntakene i reglene SSB har satt opp knyttet til utlevering. Den enkelte fagseksjon i SSB er delegert eierskap til egne data, så det er den enkelte fagseksjon som må vurdere om dataene kan leveres ut til en ekstern forsker. SSB er kjent med tilfeller hvor en seksjon har lovet data på vegne av andre seksjoner. Dette har ført til at eksterne forskere opplever først å få ja til fletting og utlevering for deretter å få en ny diskusjon. Interne prosedyrer i SSB for utlevering skal være tilpasset denne distribuerte godkjennings- og eier-modellen.

Statistikkloven sikrer at SSB har tilgang til veldig mye mikrodata. Oppfattelsen blant eksterne forskere er at SSB ifølge loven plikter å gjøre dataene tilgjengelig for forskning. Loven har ikke vært endret siden 1989, men SSB sin fortolkning og praktisering av loven har endret seg. Eksempel på endring i



## TECHNICAL REPORT

praktisering er at det nå kun lånes ut anonyme data til offentlig planlegging. Gratis utlevering av data fra FD-Trygd var en ordning som var ment å kun gjelde for en periode, SSB tar seg nå betalt for arbeid knyttet til selve klargjøring av datauttrekk og utleveringen av også denne type mikrodata.

SSBs jurister har et økende tidsforbruk knyttet til utøvelse av personvernrollen og bistand til SSBs fagseksjoner ved utlevering av mikrodata til forskere. Dette på grunn av mer og mer komplekse forskningsprosjekter, og at det ved kobling av registre er blitt strengere regelverk på en del registerkilder, eksempelvis de sentrale helseregistrene.

Et arbeid med et felles søknadsskjema for tilgang til data og søknad om dispensasjon fra taushetsplikt er påbegynt mellom SSB og NAV. En videreutvikling av samordning mellom flere etater kunne for eksempel være å benytte Altinn som portal for felles søknadsskjema, og dermed gi forsker muligheten til å følge hvor hver søknad er i prosess. SSB og NAV har arbeidet for et felles søknadsskjema i et år. Hvis andre som for eksempel Helsedirektoratet, Folkehelseinstituttet, Datatilsynet også skal inn i samordningsarbeidet kan det ta en del tid, men SSB har begynt å luften ideen sammen med disse andre etatene.

SSB har interne diskusjoner på om det er ønskelig å prøve å få til en tilsvarende løsning som Riksarkivet har når det gjelder å avgjøre dispensasjon fra taushetsplikt. Riksarkivet har fullmakt til å håndtere dispensasjon fra taushetsplikt ved utlevering til forskning for alle data som er avlevert til dem. Det vil si at forskere som søker om å få data utlevert fra Riksarkivet, ikke trenger forholde seg til andre instanser med hensyn til godkjenninger og dispensasjon fra taushetsplikten. For SSB ville slike fullmakter gitt et mye større ansvar, men ovenfor forskeren ville det vært en fordel om dette var gjennomført. Et SSB-internt notat som utreder saken, er under utarbeidelse.

Et eksempel på utfordring med kobling av registre er knyttet til Kreftregisteret. SSB har hatt en dialog med Kreftregisteret om kobling av mikrodata, fordi organisasjonene har hatt ulikt syn på hvordan data fra Kreftregisteret skal kobles med data fra SSB.

### Anbefaling:

- SSB bør ta initiativ til at det blir etablert en samlet oversikt over hvem som er kontaktperson for ulike datakilder/registre og kriterier for søknader.
- SSB bør i dialog med NSD bedre forutsigbarheten knyttet til om et prosjekt er konsesjonspliktig eller ikke. Ved for eksempel å etablere felles sjekklister eller annen formalisering og veiledning.
- SSB bør begrunne bedre overfor eksterne aktører hvorfor tolkning og praktisering av statistikkloven har endret seg. Gå i dialog med forskningsinstituttene og informer dem om endringer i de krav Datatilsynet stiller og som begrenser SSBs utlevering av mikrodata.
- SSB bør vurdere om det i godkjennings- og søknadsprosessen er mulig å få aksept for noen årsaker som automatisk fører til forlengelse av godkjenninger og konsesjoner, slik at forskningsinstituttet ved behov kun trenger å informere om endring/utvidelse uten på nytt å søke.
- Vurdere hvordan konsesjoner kan løsrives fra konkrete prosjekter og evt. knyttes mot forskningsinstituttens generelle arbeid.

---

## TECHNICAL REPORT

---

### 4.4 Organisering av offentlig sektor

Ved omorganisering av offentlig sektor flyttes ansvar for registre. Det oppstår fra tid til annen usikkerhet hos gammel og ny registereier om saksbehandling og konsesjoner. Forskeren må få godkjenning av registereier og gjeldende personvernombud. For eksempel har NAV-reformen ført til denne type problemer.

Det er også blitt nevnt at det pågår en diskusjon mellom Skattedirektoratet (SKD), EDB Business Partner og SSB. Problemstillingen er om SSB kan utlevere data som kommer fra Det sentrale folkeregisteret (DSF). Bakgrunn for diskusjonen er at EDB Business Partner har eksklusiv avtale med SKD om å formidle data fra DSF, samtidig som SSB i kraft av statistikkloven mottar daglige endringer i DSF direkte fra SKD. Sett fra forskningsinstitutter stilles det spørsmålsteget ved om EDB Business Partner har de samme data som SSB, og om de forstår mikrodata godt nok til å kunne overta SSBs utleveringsrolle. Videre må en sikre at dette ikke får konsekvenser for kobling av registre.

En annen utfordring er at når offentlige etater gjør endringer i sine registre/systemer, endres ofte grunnlaget (variabler og utvalg) for de mikrodata som avleveres til SSB. Dette fører til utfordringer med tolkning av data over tid og sammenlikning av dataserier.

Anbefaling:

- SSB bør systematisere hvordan de informere registereiere om konsekvenser ved systemendringer og organisatoriske endringer, da disse endringene kan påvirke både variabler og utvalg.

### 4.5 Samarbeid mellom SSB og NSD med betydning for forskningsinstituttene

På basis av intervjuet med NSD kommer det frem at NSD har et veldig godt samarbeid med SSB på mange ulike plan. I arbeidet med å identifisere forbedringer har vi i arbeidet med denne rapporten tatt ut en del momenter som er kommet frem fra summen av intervjuene.

-----  
NSD ble etablert i 1971 og tilbyr infrastrukturtjenester for norsk forskning. NSD prøver å tilrettelegge best mulig for at forskningen skal få gratis tilgang på mikrodata.

Om NSD, fra årsmeldingen 2007 [2]:

*NSD er et av verdens største arkiv for forskningsdata og formidler av data og datatjenester overfor forskningssektoren. NSD er også et kompetansesenter som veileder forskere og studenter i forhold til datainnsamling, dataanalyse, metode, personvern og forskningsetikk. Hovedmålsettingen er å forbedre muligheter og arbeidsvilkår for empirisk forskning som først og fremst er avhengig av tilgang til data. Dette gjøres ved å bygge ned økonomiske, tekniske, kunnskapsmessige, juridiske og administrative barrierer mellom brukere og dataressurser.*

*NSD er personvernombud for forsker- og studentprosjekt som gjennomføres ved alle universiteter, statlige, vitenskapelige og private høyskoler, en rekke helseforetak og andre forskningsinstitusjoner. Kort fortalt betyr det at forskere og studenter som skal melde behandling av personopplysninger, har meldeplikt til NSD som tilrår prosjektet i henhold til personopplysningsloven og helseregisterloven.*

*NSDs finansieringsstruktur er den samme som tidligere. Hovedbevilgningene kommer fra Norges forskningsråd, departementene, universitet og høgskolesektoren og EU.*

## TECHNICAL REPORT

En viktig oppgave for NSD er å være et dataarkiv for offentlig produserte data og andre data som forskningen selv genererer. NSD er et aksjeselskap som er heleid av Kunnskapsdepartementet. NSD driver også med utvikling av programvare som benyttes av forskere, og de er norsk representant i noen internasjonale organisasjoner. NSD har arkivert mye av dataene som SSB har produsert gjennom tidene. NSD innhenter data fra forskere, foretar ulike kvalitetssjekker på dem og vedlikeholder dataene slik at de er tilgjengelige på egnede formater for forskerne over lang tid.

NSD har over tid og i tråd med teknisk utvikling utviklet forskjellige tilbud som har gjort tilgang til mikrodata enklere og gratis. NSD mottar mikrodata fra registreiere, og disse data er egnet for å kobles.

NSD har mikrodata fra fastlegeregisteret, regionaldatabasen og kontantstøttereisteret. Disse leveres ut gratis til forskning. De sendes også til SSB, siden SSB har krav på å få alle data som blir produsert. Eksterne forskere opplyser at noen av dataleveringene fra NSD kan ha veldig lang leveringstid.

Forskningsinstituttene må nå betale for data som tidligere var gratis tilgjengelig via NSD. For forskningsinstituttene, og spesielt studenter som har begrenset finansiell ramme, er dette et tilbakesteg. SSB survey-undersøkelser kan NSD fremdeles levere ut gratis med mindre de inneholder registerdata.

Forskningsinstitutter diskuterer konsesjonsspørsmålet med både NSD og SSB. NSD håndterer søknader om konsesjon på vegne av Datatilsynet. Prosessen med å få tilgang til registre og konsesjon har blitt bedre de senere år. NSD opererer med seks uker saksbehandlingstid. Utlevering fra NSD tar varierende tid. Når man trenger ekstra sensitive data kan det bli en ekstra vurdering hos SSB, og kalendertiden kan bli ytterligere forlenget.

NSD har ofte etablert tjenester på områder hvor SSB har vist liten interesse. En overordnet interessedeling kan oppfattes å være at NSD har et internasjonalt fokus på komparative data, driver forskning på forvaltning og det politiske system, mens SSB ønsker å være sterke på formidling av regionale data og registerdata/mikrodata.

Sett fra SSB så er det historiske grunner til at de tidligere ikke har regnet oversendelse av anonyme data til NSD for å være noe problem. Nyere juridiske vurderinger gjør at SSB er føre var og er mer restriktiv til oversendelse av denne type data til NSD, men SSB arbeider med å få samhandlingen med NSD inn i en ny avtaleform. Dette avtale-arbeidet er planlagt i løpet av de nærmeste to - tre årene. Avdeling for personstatistikk har ansvaret for SSB sin avtale med NSD.

Anbefaling:

- SSB bør ta initiativ til å avklare og forbedre samordning, samhandling og rollefordeling med NSD for å sikre best mulig arbeidsvilkår for eksterne forskere.

### 4.6 Online tilgang og analyser av mikrodata

I intervjuer har vi spurt forskere om de i tillegg til, evt. alternativt til, dagens utleveringsrutiner, ønsker online tilgang til mikrodata som ligger hos SSB. En slik løsning kan medføre raskere tilgang til data og enklere håndtering av meldeplikt og konsesjoner.

Dagens praksis fra SSB er at forskere får mikrodata på for eksempel CD. Dagens arbeidsprosess med mikrodata er at forskere gjennomgår datamaterialet de har fått, prøvekjør ulike analyser med verktøy de kjenner, for etter hvert å komme frem til de analyser som gir de beste resultatene. I intervjuer responderte eksterne forskere med følgende refleksjoner rundt online tilgang til data:

---

## TECHNICAL REPORT

---

- En ønsker å kunne benytte mikrodata lokalt i egne systemer og rutiner. Samt at de ønsker å kunne bruke egne data sammen med mikrodata fra SSB.
- Forskning er en prosess hvor en ofte basert på initielle analyser kommer frem til funn som trigger avanserte kjøringer. Da er både tilgang til data og kjennskap til de verktøy som benyttes viktig.
- En ønsker å kunne se på selve innholdet i datamatriksen for å kunne se hvordan data er organisert.
- Viktig å få sikkerhet i hvordan for eksempel "ikke besvart" er representert i datamaterialet, hvilke tegn har de etterlatt seg i datamatriksen. Har man "ikke besvart" i basis utvalget eller om man bare baserer seg på de som har besvart, slike detaljer påvirker sluttresultatet.
- Krav til kompetanse ved bruk av systemet er vesentlig å få avklart, samt at om bestilling av variable skal skje uten dialog med fagpersoner hos SSB så vil trolig risiko for mangelfulle evt. feilbestillinger øke.

Funn: På spørsmål om forskerne ønsker å bestille sine variabler direkte i et online-system, uten dialog med SSB forskere, er tilbakemeldingen at de ønsker dialogen for å bli trygg på at de best mulige variabler er valgt og en sammen kan formalisere bestillingen. Det er derfor viktig å holde åpen muligheten for personlig kontakt og dialog mellom eksterne forskere og SSB ressurspersoner.

Funn: Punktene over kan tolkes som et behov for at eksterne forskere bør bli tatt med i prosessen med å velge ønsket funksjonalitet i ny online-løsning. SSBs etablering av online tilgang og bruk av public use files, er ment som et forbedret alternativ og supplement til dagens utleverings praksis. Basert på overstående punktliste så vil online tilgang kunne forbedre situasjonen for flere av punktene.

SSBs har en egen løsning for online tilgang til makrodata, Statistikkbanken<sup>2</sup>, som gir muligheter for å trekke ut data online til egne tabeller, tilpasse visninger, søke på de variabler som er predefinert for fremvisningen etc. Statistikkbanken gir ikke tilgang til statistiske kalkulasjoner eller logiske avgrensninger i datagrunnlaget ut over et sett predefinert kombinasjoner av valg. Avdeling for personstatistikk har etablert en online-løsning til bruk for Oslo kommune basert på produktet SuperCross. Egenskapene i denne løsningen har vi dessverre ikke hatt tid til å sette oss inn i. Erfaring fra Oslo kommune, andre land og fra NSDs bruk av NESSTAR (se forklaring senere i avsnittet) er at forskere benytter online-verktøy med suksess.

SSB vurderer en overgang til online tilgang dit hen at forskeren som får tilgang til dataene uansett må ha konsesjon ved bruk av personsensitive data selv om han/hun ikke fysisk har dataene liggende hos seg. Men ut fra personvern hensyn er det en god løsning at SSB slipper å sende CD-ene til forskerne. Dermed har SSB mer kontroll på hvor data faktisk befinner seg, hvem som benytter dem, og at data ikke kommer på avveie.

SSBs notat om "Online tilgang til data for offentlig planlegging" [1] beskriver behovet og løsninger for online tilgang som følger:

*"Innskjerping fra Datatilsynet har gjort det vanskeligere for planleggere i offentlig virksomhet å få tilgang til detaljert ikke-anonym statistikk fra SSB. Problemet ble spesielt merkbart ved*

---

<sup>2</sup> <http://statbank.ssb.no/statistikkbanken/>

## TECHNICAL REPORT

*den årlige leveranse av plandata til Oslo kommune hvor mange tabeller ikke lenger kunne leveres på grunnkrets uten undertrykking. En alternativ løsning er at kommunene får tilgang til statistikkdata via Internett. Dette er tenkt løst ved at det etableres en database med individdata på dedikert server hvor eksterne brukere får tilgang via sikker webløsning. Brukeren har ikke tilgang til individdataene, men kan ved hjelp av SuperCross selv generere tabeller ut fra databasen - og tabellene vil automatisk avrundes for verdier 1 og 2. En slik løsning medfører at kommunene får tilgang til den detaljerte statistikken de trenger uten at store mengder detaljerte statistikkfiler bringes utenfor SSB som i dag. .... Tabellene som genereres er mer detaljerte enn hva som kan tilbys i statistikkbanken, og det vil fortsatt være restriksjoner på den videre bruk av tabellene som tas ut.”*

I forbindelse med forarbeidet til denne rapporten har vi ikke gjort noen sammenlikning av NESSTAR og SSBs initiativ overfor Oslo kommune.

Funn: SSB har en egeninteresse i etablere alternativ utleveringsmåte basert på en online løsning bl.a. for å bedre oversikt over hvilke data som befinner seg hvor, sikre at data ikke kommer på avveie, og at tilgang/sletting av data håndteres i henhold til konsesjoner og avtaler mellom partene.

NESSTAR er et system for internettbasert tilgang til forskningsdata og er utviklet over 10 år bl.a. av NSD. NESSTAR bistår forskere med å finne data. Data kan lastes opp/ publiseres inn i NESSTAR-portalen og vil så være tilgjengelig for søking og bearbeiding. Resultater av bearbeidingene kan lastes ned lokalt. Hvis mikrodata er anonyme, kan en søke NSD om tilgang til mikrodata uten å søke om videre godkjenning. NESSTAR brukes av mange store dataprodusenter i England og Tyskland for å gjøre datasettene tilgjengelig for forskning. Det er begrenset hva NSD kan legge ut av data siden SSB har eiendomsrettighetene til en stor del av norske surveydata. Alt det NSD kan legge ut, blir lagt ut.

NSD oppfatter at SSB ikke er interessert i å benytte NESSTAR-løsningen I NESSTAR kan man analysere dataene og kjøre statistiske analyser og regresjonsanalyser på dataene online. Studenter bruker denne funksjonaliteten aktivt. Man kan ikke kjøre avanserte online analyser som faktoranalyse og diskriminantanalyse. Det er i følge NSD enkelt å publisere data inn i løsningen og det er lett å få data ut fra løsningen. Løsningen kan bl.a. importere SPSS filer. Dokumentasjon finnes på nesstar.com. NESSTAR er i bruk på NSDs web-sider.

Intervjuer og kilder hos SSB har informert om at SSB har vurdert og testet NESSTAR ved flere anledninger og påpekt svakheter i nåværende versjon.

Anbefaling:

- Datatilsynet gjorde i 2006 en innskjerping i vedtaket etter statistikkloven § 2-5 som åpner for bruk av SSBs data til andre formål enn statistikk, som innebærer at kun *anonyme* mikrodata kan lånes ut til offentlig planlegging. Det er viktig at dette kommuniseres til berørte parter, herunder også forskningsinstitutter og NSD. Dette for å unngå spekulasjoner om SSBs motivasjon for å endre praksis vedrørende utlevering av mikrodata.
- Datatilsynet har gjort en innskjerping i tolkning av personopplysningsloven og dermed gjort det vanskeligere for SSB å utlevere ikke-anonyme data og statistikk. Det er viktig at dette kommuniseres til berørte parter som forskningsinstitutter og NSD. Dette for å unngå spekulasjoner om SSBs motivasjon for å endre praksis vedrørende utlevering av mikrodata.

## TECHNICAL REPORT

- Online tilgang vil være et alternativ til, men også en endring fra dagens arbeidsprosess både for SSB og eksterne forskere. Eksterne forskere bør være involvert i prosessen med å etablere krav funksjonalitet og egenskaper i ny løsning.
- SSB bør ta initiativ til å avklare og forbedre samordning, samhandling og rollefordeling med NSD.
- SSB bør bidra til at aktører som forskningsinstitutter og NSD, som i større eller mindre grad baserer sine tjenester på data fra SSB, har mulighet for å planlegge langsiktig og strategisk.
- Det er løsninger utviklet i SSB og i NSD som kan ha overlappende funksjonalitet. SSB bør ta initiativ til å se på mulighetene for å samordne utvikling og anvendelse av løsningene.

### 4.7 SSBs informasjon om metadata

I de to følgende avsnitt har vi prøvd å skille eksterne forskeres bruk av SSBs websider om henholdsvis metadata og mikrodata i separate avsnitt.

I dialog med erfarne forskere har vi blitt orientert om at de bl.a. bruker SSB websider om metadata<sup>3</sup> som kilde før og etter dialog med fagpersoner i SSB. Forskerne oppfatter det som viktig å ha tilgang til fagpersoner i SSB for å kunne forstå mest mulig angående hvilke data som finnes, hvilke begrensninger som er knyttet til datasett og hvordan mikrodata kan settes sammen.

Funn: Hvordan metadata endres over tid er nevnt som en utfordring. Dette er vanskelig for eksterne forskere å forstå. Eksempel er SSBs metadata web som per variabel har metadata som ”gyldig fra” og ”gyldig til”.

Det har i intervjuer med SSB interne kommet frem at det er behov for å kunne håndtere flere språk på metadata, eksempelvis definisjoner, forklaringer og koderlister etc.

Anbefaling:

- Det burde vært mulig å få informasjon om hva en endring i en variabel faktisk er, evt. oppgi både gammel og ny definisjon og evt. endring i utvalg. Enda bedre er om det også er mulig å si noe om hvilke typer statistiske analyser som blir påvirket av endringen.
- Utviklingen i de enkelte variablene over tid kan være av stor betydning for hvordan mikrodata skal tolkes, og disse endringene bør publiseres.
- Hvilke variabler som beskriver mikrodata, og hvilke som beskriver aggregater av mikrodata, bør fremkomme.
- Hvilke ulike typer relasjoner det er mellom de ulike variablene bør fremkomme. Med relasjoner menes her bl.a. arv, komposisjon etc.
- SSB bør tilby eksterne interessenter å laste ned variabel-definisjoner og -strukturer på et egnet format for utveksling av metadata. Da kan mottaker selv velge hvordan de vil søke, vise, navigere og implementere variabelinformasjon i lokale systemer. Eksempler på utvekslings- og strukturerings-formater kan være
  - a. World Wide Web Consortium (W3C) sitt initiativ Web Ontology Language (OWL<sup>4</sup>) eller

<sup>3</sup> <http://www.ssb.no/metadata/>

<sup>4</sup> <http://www.w3.org/2004/OWL/>

---

## TECHNICAL REPORT

---

- b. Object Management Group (OMG) sitt initiativ XML Metadata interchange (XMI<sup>5</sup>).
- Etablere støtte for flere språk i SSBs metadata systemer. Poenget er bl.a. nevnt i [13].

### 4.8 SSBs informasjon om mikrodata

Mikrodata web-en er sammen med metadata web-sidene viktige formidlere av hva som finnes av data hos SSB og hvordan forskere kan få tilgang til dem. For å avlaste SSBs personell i arbeidet med mikrodata må den informasjon som finnes her, være oppdatert og av høy kvalitet. Målgruppen for disse web-sidene er et begrenset antall forskere, så riktig prioritering av vedlikehold og videreutvikling er svært viktig. Ellers vil gevinsten med web-sidene, sett fra SSBs side, ikke stå i forhold til kostnaden med etablering og vedlikehold. Det kan allikevel være at eksterne forskere får en gevinst selv om SSBs interne ROI er negativ.

Forskere oppgir at de benytter SSBs nettsider mye for generell orientering, og de temabaserte nettsidene oppfattes som veldig gode. Selv om Statistikkbanken er en makrodatabase, kan den også benyttes for å gi et innblikk i hvilke mikrodata SSB har tilgjengelig., samtidig som den også kan benyttes til uttrekk av data for analyse uten å måtte bestille utlevering av mikrodata.

Anbefaling:

- SSB bør i dialog med forskere finne ut hva som er hensiktsmessig kvalitet og hvilke typer informasjon som forskere vil gjøre seg nytte av.
- SSB bør publisere flere typer visninger av koblingen mellom variable, mikrodatakilder og allerede produsert statistikk. Dette vil gjøre det enklere for forskere å forstå hvilke tidsserier og forløp som finnes samt hvor de er benyttet.
- SSB bør publisere hvilke typer sammenstilling av data som allerede er gjort og som kan prises lavt eller gis ut gratis.

### 4.9 Nasjonal oversikt over mikrodata

En utfordring for forskere innen mikrodata er at SSB ikke er den eneste kilden til mikrodata, og forskere burde hatt enklere tilgang til en nasjonal oversikt over hvilke mikrodata som faktisk finnes, og hvilke prosedyrer som gjelder for å få konsesjoner, uttrekk, fletting etc.

På SSB mikrodata-websider henvises det bl.a. til NSD, dette bidrar til å skape en bedre oversikt, men gir på ingen måte en komplett oversikt. SSB har videre en oversikt over norske nettsteder som publiserer statistikk<sup>6</sup>.

Anbefaling:

- SSB bør ta initiativ til at det blir etablert en oversikt over alle nasjonale mikrodatakilder. Mikrodatas "gule sider" eller kalt et registry, som er en indeks til de faktiske repository/kildene.

---

<sup>5</sup> [http://www.omg.org/technology/documents/modeling\\_spec\\_catalog.htm#XMI](http://www.omg.org/technology/documents/modeling_spec_catalog.htm#XMI)

<sup>6</sup> <http://www.ssb.no/lenker/nostat.html>

---

## TECHNICAL REPORT

---

### 4.10 SSB som underleverandør; leveransepunktligheit og prising

Sett fra forskningsinstituttet er SSB en underleverandør. Som vist i prosessmodellen tidligere i dette kapitlet går forskningsinstituttet gjennom en rekke aktiviteter der mikrodata fra SSB er avgjørende for at forskningsinstituttet skal kunne levere sine avtalte resultater.

For forskningsinstituttet vil både medgått tid fra bestilling av mikrodata til utlevering og prisen på levering av mikrodata være kritisk for at forskningsinstituttet skal kunne levere sine forpliktelser til rett tid, til rett kostnad og med god kvalitet. Hendelser som lange leveringstider, forsinkelser eller endring av kostnad i forhold til estimat vil sette forskningsinstituttets egen produksjon i vanskeligheter og SSBs rolle som underleverandør blir presset.

Fra intervjuer med eksterne forskere er det fremkommet at når for eksempel et forskningsinstitutt skriver et tilbud og arbeidet er under tidspress, oppleves det at forståelsen for denne situasjonen hos SSB er personavhengig. Da kontakter gjerne forskere enkeltpersoner i SSB som en fra tidligere vet har erfaring og som gjerne er i en formell posisjon som seksjonsledere. Når slik kontakt opprettes er det forståelse fra begge sider om at estimat og overslag som gis ikke er forpliktende. Nyansatte eller personer lenger ned i systemet hos SSB har større problemer med å gi uforpliktende prisoverslag pr. telefon.

I forbindelse med arbeidet med denne rapporten har kilder internt i SSB informert om at evt. budsjettsprekker på levering av mikrodata ikke belastes oppdragsgiver. I dialog med forskningsinstitutter har vi blitt orientert om at en evt. budsjettsprekke på SSBs estimater blir fakturert forskningsinstituttet.

Funn: Når det gjelder SSBs håndtering av budsjettsprekker har vi fått ulike forklaringer på et og samme fenomen avhengig av kilde.

Eksempel er bl.a. et prosjekt med et initielt SSB-estimat på over 500 000 kr. Det ble ikke gjort endringer i hvilke variabler som skulle leveres. Leveransen endte opp med å bli ca 10 prosent dyrere. I et annet eksempel ble det mer enn 10 ukers forsinkelse på datalevering. SSB var ikke villig til å redusere prisen pga. for sein leveranse. For forskningsinstitutter vil situasjoner som nevnt i eksemplene over skape problemer. Forskningsinstituttets lønnsomhet og renommé kan bli utfordret på grunn av SSB som underleverandør.

Et prosjekt hos et forskningsinstitutt fikk forsinket levering fordi kommunikasjonen internt hos SSB var uklar. Forskningsinstituttet opplevde at etter at data var bestilt og bestillingen godtatt av SSB, ble det fra SSB-ansatte stilt spørsmål om denne utleveringen var mulig/tillatt. Det ble deretter laget et notat til ledergruppen, og notatet måtte behandles før arbeidet med uthenting av data ble fullført.

Anbefaling:

- SSB må tydeligere beskrive:
  - a. Forutsetninger for estimater.
  - b. Hva som leder til kostnadsendringer for forskningsinstituttet i forhold til estimat.
  - c. Hva som kan lede til endringer i leveringstidspunkt.

### 4.11 Survey koblet med mikrodata

I noen prosjekter velger eksterne forskere å benytte SSBs seksjon for "Intervjuundersøkelser" under avdeling for "IT og datafangst" til å gjennomføre surveys. Det oppleves at det er enklere å få koblet



## TECHNICAL REPORT

survey-data med registerdata når undersøkelsene gjøres av SSB selv. I denne sammenheng vil også forskningsinstitutter som bruker seksjon for "Intervjuundersøkelser" trolig sammenlikne tjenestetilbudet og kvaliteten i denne avdelingen med tilsvarende tjenester i markedet.

### Anbefaling:

- Det er viktig for SSBs omdømme at både utlevering av mikrodata og i den grad tjenester fra seksjon for "Intervjuundersøkelser" kan oppleves som, eller som tilnærmet, monopoltjenester har høy tjenestekvalitet. SSB bør ha rutiner for å sikre at disse tjenestene har et målbart og ønsket kvalitetsnivå.

### 4.12 Egenskaper ved mikrodata som utleveres

Forskningsinstitutter oppfatter datakvaliteten på mikrodata som noe variabel. Om dette skyldes SSB selv, eller de kilder / registre SSB får data fra, er ikke studert i denne sammenheng. En annen utfordring oppstår når ulike datakilder kobles sammen og koblingsnøkkelen som er mulig å benytte mellom ulike datasett er av lav kvalitet. Sluttresultatet blir da mikrodata med usikker kvalitet.

#### Eksempel på et slikt problem:

Når et sett av mikrodata som omhandler husholdninger ønskes koblet med mikrodata som er identifisert på enkeltperson, må det lages koblinger mellom adresser på husholdning og adresser på personer. Disse koblingene vil ha feilmarginer. En utfordring knyttet til nevnte kobling er å sikre tilstrekkelig kvalitet på hvilke personer som er knyttet til hvilke adresser og dermed til hvilke husholdninger. Ved bruk av flyttemeldinger til folkeregisteret vedlikeholdes denne koblingen. Kvalitet på flyttemeldinger er spesielt problematisk for Oslo. Også studenter og deres knytting til husholdninger skaper utfordringer.

Eksterne forskere jobber vanligvis med usikkerhetsmarginer i sine undersøkelser, og til tross for at de oppfatter at mikrodata fra SSB har noen kvalitetsproblemer er disse mikrodata allikevel bedre enn det som vanligvis kan fremskaffes på andre måter med fornuftig ressursbruk.

For å forstå de data som mottas er det vanlig at forskere tar utgangspunkt i mikrodatamaterialet fra SSB, og i den grad de trenger forklaringer til variabler kan de benytte metadatatjenestene hos SSB.

Angående format på data så leverer SSB ofte på string-variabler. Ved kategoriserte data der verdi = "1" betyr mann og verdi = "2" betyr kvinne så mister man ved omdanning ofte betydningen av kodene. Eksterne forskere oppfatter at representasjonsformatet og modellen dette formatet er basert på, er årsak til at en mister semantiske detaljer som er viktig i de videre analysene.

### Anbefaling:

- SSB må bidra til at mikrodata tolkes korrekt og at semantikk ikke endres ved konvertering av representasjonsformater.
- SSB må beskrive tydelig eventuelle svakheter ved kobling og koblingsnøkler i de datasett som utleveres.

### 4.13 Begreper, virkelighetsforståelse, semantikk

Eksterne forskere oppfatter det som utfordrende å komme med konkrete bestillinger uten å ha en dialog med fagpersoner hos SSB. De ser på SSBs kompetanse om mikrodata innen ulike domener som avgjørende for at de beste variabler, utvalg og mikrodatakilder blir valgt som input til forskningsprosjekter.

## TECHNICAL REPORT

---

I bestillingsfasen benytter forskere SSBs nettsider i varierende grad, mest for generell orientering. Også statistikkbanken (som tilgjengeliggjør makrodata) benyttes for å etablere forståelse for hva som finnes av mikrodata og statistikk. Når det kommer til selve bestillingen av mikrodata må forskerne snakke med personer hos SSB som har fagkompetanse på registrene.

Forskere oppfatter at dokumentasjon om hva som kan bestilles og informasjon om hvilke variabler fra hvilke datasett som er tilgjengelig i stor grad er lagt ut på SSB mikrodata og metadata web. En av utfordringene for eksterne forskere er at metadata endres over tid, samt at kontekst og metode for innsamling av data endres. Kunnskap om disse endringene er vesentlig å få formidlet til eksterne forskere. SSBs måte å dekke dette formidlingsbehovet er å legge ut omtale av de ulike statistikkene, kalt "om statistikken". På SSB mikrodatawebsidene ligger det informasjon om de ulike mikrodatakildene, for eksempel websiden "Om Bedrifts- og foretaksregisteret (BoF)"<sup>7</sup>. Denne siden beskriver en rekke endringer og hendelser som påvirker hvordan mikrodata fra ulike perioder skal tolkes.

I intervjuer med ansatte hos SSB kom det frem at når etater endrer i sine systemer så forekommer det endringer på mikrodata som gjøre det vanskelig å sammenlikne tidsserier av data. Dette er et eksempel på hendelser og endringer som påvirker kvalitet på mikrodata, men hvor styring av hendelsen er utenfor SSBs kontroll. Det er ofte dialog mellom SSB og gjeldende etater i slike tilfeller, og i varierende grad får SSB påvirket endringene slik at mikrodata for ulike tidspunkt og perioder er sammenliknbare.

Eksterne forskere får bistand fra SSB i å forstå hvilke begrensninger som ligger i bruk av mikrodata. Dette er fundamental kunnskap for å kunne komme frem til riktige konklusjoner i de statistiske analysene. For å kunne besvare forskningsspørsmålet best mulig er følgende punkter viktig:

- Forskeren må enkelt få oversikt over og forstå innholdet i de datakilder og statistikker SSB faktisk har
- Utvalget av variabler fra de valgte datasett må være optimale i forhold til forskningsspørsmålet
- Beskrivelse av begrensninger og ulikheter i tidsserier i et datasett er vesentlig for å sikre korrekt bruk av mikrodata
- Begrensninger knyttet til kobling av data må klargjøres for forskerne

Funn: Fra intervjuene er vår oppfattelse at variabler og mikrodata fra FD Trygd og Nasjonal Utdanningsdatabase er vanskelig å forstå.

Når det gjelder registerdata er det vanskelig å få gode variabelbeskrivelser. Det oppfattes som om variabelbeskrivelser er på "stammespråket". Når det gjelder surveydata-variablene er disse lettere å forstå, det oppfattes som om disse er tilpasset internasjonal terminologi, og en utdannet samfunnsforsker vil lettere kjenne igjen denne terminologi.

Semantikken i representasjonsformatet skaper rom for tolkning. Det er for eksempel uklart for forskerne hvordan detaljer i dataformatet skal tolkes. Dette gjelder for eksempel å kunne sondere mellom "ikke besvart", null eller undertrykt feltverdi i representasjonsformatet når et felt er tomt, etc.

For å forstå endringer i betydningen av mikrodata over tid, oppfatter forskere at de må snakke med fagpersoner i SSB.

---

<sup>7</sup> <http://www.ssb.no/mikrodata/datasamling/bof>

---

## TECHNICAL REPORT

---

### Anbefaling:

- For å bedre utenforståendes innsikt i hvilke mikrodata som er tilgjengelig og når disse blir tilgjengelig, bør SSB publisere også når nye mikrodatasett er planlagt
- For å gjøre det enklere å bruke mikrodata riktig bør SSB bedre kvaliteten på metadata om variabler, spesielt innen områdene FD Trygd og Nasjonal Utdanningsdatabase.

### 4.14 Tidsperspektiv og ressursbruk

Forskningsinstitutter opplever at SSB har en leveringskapasitet som er alt for lav, særlig hvis man må koble flere registre. Forskningsinstitutter prøver å kompensere ved god planlegging, men i noen prosjekter kan man ikke belage seg på å benytte SSB fordi leveringstiden er for lang. Det er enkeltpersoner hos SSB som strekker seg langt for å tilfredsstille forskningsinstitutters behov, men dette er ikke nok til å løse ressursproblemet.

Funn: Et av forskningsinstituttene nevner at de i alle tilbud tar forbehold om SSBs leveringstid.

Underbemanning virker å være den vanligste årsaken til leveringsproblemene. I tillegg er kobling av data mellom flere seksjoner en kompliserende faktor, men dette er også personavhengig og basert på erfaring. SSBs forklaring til forskningsinstitutter på lang leveringstid eller forsinkelser er at de har for mye å gjøre, f.eks. uforutsett fravær som reduserer kapasiteten. Eksternt oppfattes det som SSB generelt er flinke til å estimere hvor lang tid en leveranse tar.

- Funn: Forskningsinstitutter oppfatter at det vil være veldig arbeidsbesparende om de kunne beholde mikrodata etter prosjektslutt og bare koble på nye årganger slik som de oppfatter at Frisch-senteret gjør. Det ville vært arbeidsbesparende både for SSB og forskningsinstituttene.

### Anbefaling:

- SSB bør bidra til å avklare forskningsinstitutters mulighet til å beholde mikrodata ut over prosjektperioder.

### 4.15 Prising og rettigheter til betalte uttrekk av mikrodata

Ut fra den informasjon som har fremkommet i forbindelse med utarbeidelsen av denne rapporten, tar SSB betalt av forskningsinstitutter for å ta ut mikrodata, mens NSD tilbyr deler av de samme mikrodata gratis. Dette er en komplisert situasjon for eksterne forskere og bør være gjenstand for klargjøring og mulig endring.

Videre må SSB være tydelig på hvem som har hvilke rettigheter til materialet som leveres i tilknytting til betalte oppdrag.

### Anbefaling:

SSB bør klargjøre overfor forskningsmiljøer:

- Kan SSB ta seg betalt flere ganger for å ta ut de samme data?
- Har betalende forskningsinstitutter enerett på datasettet, i så fall under hvilke betingelser?
- Om betalende forskningsinstitutter ikke har enerett, er disse mikrodata da å anse som gratis tilgjengelig fra SSB?

I dialog med eksterne ble det tatt opp at det er en utfordring at mikrodata i større grad nå enn tidligere bare blir tilgjengelig for dem som har betalingsevne til å dekke SSB kostnader. Dette er blant annet en

## TECHNICAL REPORT

utfordring for master- og doktorgradstudenter som innen noen fagfelt baserer sine oppgaver på mikrodata. Er studenter upresise med bestillingen får de nødvendigvis ikke nye uttrekk av mikrodata uten å måtte vente ny periode og kanskje også betale for et nytt uttrekk, dette oppleves som et problem. Det er også nevnt som en demokratisk utfordring at de med best betalingsevne får de beste mulighetene til å forske på viktige samfunnsspørsmål.

Per i dag finansierer NSD og SSB sin ressursbruk ved utlevering av mikrodata på to vesentlig ulike måter. Til tross for dette blir de begge i stor grad finansiert over statsbudsjettet via ulike direkte og indirekte kanaler. Det bør derfor være mulig på sikt å arbeide frem en bedre finansieringsmodell enn dagens for utlevering av mikrodata til forskere.

### Anbefaling:

- SSB og NSD bør ta initiativ til å få gjennomført en samfunnsøkonomisk betraktning på prising av utlevering av mikrodata til forskning.
- SSB bør bidra til at kostnaden på utlevering av mikrodata holdes nede evt. fjernes.
  - a. gi for eksempel NSD en tydelig rolle knyttet til utlevering av gratis mikrodata til forskning.
  - b. sørge for at SSB kommer i samme finansielle posisjon som NSD knyttet til utlevering av mikrodata.

### 4.16 Samarbeid om forskning og konkurranseforhold

Forskningsinstitutter har direkte samarbeid med forskere og forskningsavdelingen til SSB. Sett fra forskningsinstituttet kan dette være strategisk på flere måter, bl.a. for å sikre leveranse innenfor tidsrammene og at SSB har mange kompetente ansatte som arbeider med rapportskrivning, faglige artikler og datafangst.

SSB kan også levere tilbud på de samme anbudene som andre forskningsinstitutter leverer tilbud på. Det vil si at noen ganger er man konkurrenter og andre ganger samarbeider man og utfyller hverandre.

Konkurransesituasjonen mellom SSBs forskningsavdeling og forskningsinstitutter har relevans for samhandlingsproblematikken som ligger til grunn for Semicolon prosjektet, men prosjektet gjør ingen videre vurdering på dette feltet.

## 5 ØNSKET SITUASJON

I forbindelse med intervjuene er det fremkommet en rekke ønsker om fremtidig situasjon når det gjelder utlevering, bruk og forvaltning av mikrodata. Disse punktene er:

- Bedring av SSBs kapasitet til å levere mikrodata raskere.
- Forenkling, redusert tidsbruk og kalendertid knyttet til prosessen med å få avklart behov for konsesjon, og etablering av nødvendige godkjenninger, konsesjoner og dispensasjoner. Spesielt gjelder dette når data fra flere kilder benyttes.
- Mulighet til å kunne beholde mikrodata fra tidligere forskningsprosjekter og bare be om oppdateringer.
- Etablere og vedlikeholde en god oppdatert oversikt over hvem som er registereiere, gjerne med saksbehandler, ville bidratt til at søknader kommer raskere til riktig saksbehandler.
- Etablere et samleskjema/søknadsblankett som kan benyttes mot flere registereiere og godkjenningsinstanser.

---

## TECHNICAL REPORT

---

- Redusere prising av supplerende uttak, som for eksempel nye tidsserier/årganger. Prisingen bør være lavere enn kostnaden av initialt uttak.
- Gjenbruk av gamle uttrekk bør være enklere, evt. hente uttrekk fra et SSB-arkiv.

I intervjuene spurte vi eksterne forskere om kjøring av statistikkanalyser kan gjøres mens data ligger hos SSB. Tilbakemeldingen er at forskere ofte gjør en del innledende undersøkelser og analyser på mikrodata før de velger den endelige statistiske metoden og vinklingen som er best egnet på det gjeldende datamaterialet. Til dette arbeidet ønsker forskere å benytte verktøy de er kjent med. I tillegg ønsker de å kunne benytte egne data sammen med SSBs data.

Forskningsinstituttene foretrekker å snakke med fagpersoner hos SSB om rådgivning og mener at de ikke vil være komfortable med bare å benytte en systemstøttet bestillingsprosess av variablene. En slik løsning vil overføre risiko ved variabelvalg til forskeren. Dvs. et forskningsinstitutt kan risikere å bestille feil/ ikke de optimale variabler og evt. ikke benytte de beste mikrodatakildene i forhold til det forskningsproblemet som skal løses. Respons fra eksterne forskere er at det vil ta tid og ressurser å bygge opp kompetanse på bruk av et online system. Det er viktig at forskningsinstituttens innsats med å lære seg nye systemer står i forhold til fremtidig gevinst.

Noen forskningsinstitutter bestiller for mange og overlappende uttrekk hos SSB. Det er kostnaddrivende og arbeidskrevende for SSB og forskningsinstituttet. Et forskningsinstitutt bør koordinere behovene sine og samle forespørslene. Dette er en konklusjon som har vokst fram over tid i dialog med SSB, men som også kan påvirke hvordan konsesjoner tildeles.

Ved oppstart av FOSS-prosjektene som Semicolons arbeid inngår i, har SSB noen tanker om mulige tiltak som kan forbedre sekundærbruken av mikrodata. Bl.a. etablering av remote access, public use files og felles søknadsskjema for godkjenninger. Responsen fra eksterne forskere er at de ser en egenytte i public use files og et felles søknadsskjema, men mer usikker på gevinster ved bruk av en online-løsning. Denne responsen kan ha fremkommet som resultat av for lite/ for dårlig informasjon i intervjusituasjonen om hva en alternativ og supplerende online-løsning kan innebære av endringer og nye muligheter.

### Anbefaling:

- SSB må sikre at god informasjon gjøres tilgjengelig om hva en online-løsning kan gi av endringer og muligheter.
- Eksterne forskere bør delta i utformingen av en online-løsning.

## 6 "BESTE PRAKSIS" OG "STATE OF THE ART"

Basert på dette dokumentet vil Semicolon og SSB diskutere seg frem til avgrensning og ambisjonsnivå for "State Of The Art" og "Beste Praksis". Innholdet i dette kapittelet er derfor kun stikkord som er ment som input til de neste fasene av arbeidet med mikrodata til forskere.

I tillegg til anbefalinger og funn i rapporten listes her noen supplerende stikkord til videre arbeid:

- Hvordan organiseres verdikjeder knyttet til mikrodata andre steder, inklusive utlevering.
- Hvordan er rollefordeling mellom statistikkbyråer, arkivtjenester, formidlere og brukere av mikrodata andre steder.
- Hva er muligheten for et (evt. noen få) søknadsskjema for godkjenninger og konsesjoner for mange typer registerdata.

## TECHNICAL REPORT

- Hvordan måle kvalitet på mikrodata og metadata.
- Hvordan bør vi måle effektiviteten og samfunnsgevinstene av SSB samhandling med forskningsinstituttene om utlevering av mikrodata.
- Hvilke kategorier av standarder bør SSB basere sine tjenester og arkitekturer for utveksling av mikrodata på, og innen hver type, hvilke faktiske standarder bør benyttes, mulige eksempel.
  - Klassifisering for modellering av variabler.
    - Neuchâtel Terminology Modell.
    - SERES.
    - ISO 11179.
    - Hva av arbeidet innen UN/CEFACT, OASIS, OAIS passer?
    - Hvordan knytte valgt metode til ulike retninger innen semantiske teknologier.
      - Deskriptiv logikk.
      - Objekt orientert.
  - Modell for og representasjon av metadata for mikrodata.
    - DDI (data document initiative).
    - OWL.
    - XMI.
  - Arkitekturer for interoperabilitet
    - IDABC og EIF.
  - Transformasjoner og sammenlikninger av mikrodatasett.
    - OMGs Common Data Warehouse model.
    - DDI tilnærming.
    - SERES.
  - Presentasjon av statistikk.
- Hvilke typer programvare benyttes andre steder.
  - Online tilgang til mikrodata i verktøy som NESSTAR og SuperCross.
  - Presentasjon av forskningsresultater med verktøy som eks. Gapminder.
  - Hvordan anvendes semantiske teknologier og metoder og verktøy innen verdikjeder for statistikk produksjon basert på mikrodata.
- Trender og initiativ som bør vurderes.
  - Open data foundation<sup>8</sup>.
  - Web 2.0 og Web 3.0.

## 7 KONKLUSJON

SSB har allerede ved flere anledninger gjort forbedringer i sine tjenester knyttet til utlevering av mikrodata, men er inneforstått med at det fremdeles er rom for forbedring. Intervju med eksterne forskere og SSB ansatte har bekreftet SSBs anerkjennelse innen statistikkfaget, men også påpekt utfordringer ved SSBs verdikjede knyttet til mikrodata-fangst, forvaltning, bearbeiding, utlevering etc.

For det enkelte forskningsprosjekt vil det ut over å gjøre selve forskningen brukes mye tid på: (i) håndtering av godkjenninger, vurdering om et prosjekt er meldepliktig eller konsesjonspliktig, og søknad om konsesjon, og (ii) etablere en bestilling av utvalg og variabler som er så presise som mulig og som er mest aktuelle i forhold til forskerens problemstillinger.

Både i tilbudsfasen og når man har fått prosjektet, tas det nærmere kontakt med SSB. I disse fasene trengs bistand til å skrive en bestilling av utvalg og variabler som er så presis som mulig.

<sup>8</sup> <http://www.opendatafoundation.org/>

## TECHNICAL REPORT

Kontaktperson i SSB hjelper forskeren med å velge de beste mikrodatakilder, utvalg, variabler og koblingsmuligheter.

Personopplysningsloven har en viktig funksjon, men samtidig fører oppfyllelse av den til tidsbruk, kostnader og hindringer for forskning som baserer seg på bruk av mikrodata om personer. Uten å bryte med lovens intensjon bør en vurdere på kort sikt om det er mulig å praktisere loven slik at det passer bedre med forskningsinstitutters behov for tilgang til mikrodata. Forskningsinstituttene ønsker å kunne beholde mikrodata fra tidligere forskningsprosjekter og bare be om oppdateringer. Om dette er mulig, har vi ikke svart på som del av denne rapporten, men bør utredes. Det vil også være av interesse å avklare om utvalgte forskningsinstitutter kan få konsesjoner på årsbasis i stedet for prosjektbasis. Å konkludere om det er SSB eller forskningsinstituttene selv som bør ta initiativ til å starte et slik arbeid vil være å gå ut over mandatet til denne rapporten.

Av de punktene som er kommet frem i denne rapporten, er flere av forbedringene knyttet til tema som konkurranseforhold og rolleavklaringer. SSB har et mikrodataarkiv som er av stor interesse for eksterne forskere, og SSBs egne forskere benytter det samme materialet. Et eget kapittel er viet problemstillinger rundt SSB som tilbyder på anbud i markedet.

Rolledelingen mellom SSB og NSD er endret over tid. For forskere oppleves det som om mikrodata som tidligere var gratis fra NSD, nå må kjøpes av SSB. Merknad JKT: Dette er som tidligere nevnt helt feil. Hvem kan ha skapt dette inntrykket, tro?

Ved oppstart av FOSS-prosjektene har SSB gjort seg noen tanker om mulige tiltak som kan forbedre gjenbruken av mikrodata. Bl.a. etablering av remote access, public use files, og felles søknadsskjema for godkjenninger. Eksterne forskere responderer i intervjuene at de ønsker å kunne benytte mikrodata lokalt i egne systemer og rutiner og var lunken til en online-løsning. Denne responsen kan ha fremkommet som resultat av for lite/ for dårlig informasjon i intervjusituasjonen om hva en supplerende online-løsning kan innebære av endringer og nye muligheter. I tillegg er det ikke sikkert at det ble godt nok formidlet hvorfor SSB har behov for endring i forvaltning og utlevering av mikrodata. Om SSB endrer sin praksis vedrørende utlevering av mikrodata må det formidles hvorfor endringen gjennomføres og hvilke konsekvenser endringene får for de involverte parter.

SSB bruker termen ”sekundærbruk av mikrodata” om eksterne forskningsmiljøers gjenbruk av mikrodata. Denne terminologibruken forsterker trolig et inntrykk av at utlevering av mikrodata er noe SSB gjør, men som ikke har høy prioritet.

En del steder i rapporten har vi beskrevet hvordan eksterne miljøer og SSBs egne ansatte oppfatter og forstår SSBs organisasjon, interne ansvarsforhold eller systemportefølje. På samme måte som SSB ansatte har begrenset kunnskap om interne forhold hos forskningsinstituttene har eksterne forskere begrenset forståelse om SSB interne forhold. Hvis effektiv samhandling mellom forskningsinstitutt og SSB krever stor forståelse om hverandre bør en enten øke forståelsen eller endre samhandlingsprosessen. Mikrodata og metadata er komplekse områder, og en må nok akseptere at nye forskere nødvendigvis må bruke en del tid før de er blitt produktive forskere på mikrodata.

Det er gledelig at SSB arbeider for at forskere skal få en bedre og mer effektiv hverdag, og at arbeid som denne rapporten lar seg finansiere og produsere.

## 8 REFERANSER, AKRONYMER OG TERMINOLOGI

## TECHNICAL REPORT

### Refererte dokumenter

|   |                                                                                                                                                                                                                                                        |
|---|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | SSB internt notat. 10. februar 2008. Online tilgang til data for offentlig planlegging. Til Øystein Olsen, Fra: Johan Kristian Tønder.                                                                                                                 |
| 2 | Årsmelding NSD 2007. <a href="http://www.nsd.uib.no/nsd/doc/nsd_annualreport2007.pdf">http://www.nsd.uib.no/nsd/doc/nsd_annualreport2007.pdf</a>                                                                                                       |
| 3 | FAOS, Felles IKT-arkitektur i offentlig sektor, desember 2007.                                                                                                                                                                                         |
| 4 | Noen sentrale begreper knyttet til metadata – til bruk i SSBs felles metadatasystemer. Av Anne Gro Hustoft og Hans Viggo Sæbø. <a href="http://www.ssb.no/metadata/definisjoner/begreper.pdf">http://www.ssb.no/metadata/definisjoner/begreper.pdf</a> |

### Bakgrunnsliteratur

|    |                                                                                                                                                                                                                               |
|----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5  | Planlegging av hensiktsmessig lagringsstruktur for SSBs dataarkiv. 6. November 2007. Jostein Ryssevik                                                                                                                         |
| 6  | Mandat til utredning om framtidig infrastruktur for mikrodata til forskere. Torbjørn Hægeland. Utkast, 12. November 2007.                                                                                                     |
| 7  | Oppfølging av notat [6], Johan-Kristian Tønder. 14. Januar 2008.                                                                                                                                                              |
| 8  | Draft document as basis for EIF 2.0. European Interoperability Framework for pan-European eGovernment services. <a href="http://ec.europa.eu/idabc/servlets/Doc?id=31597">http://ec.europa.eu/idabc/servlets/Doc?id=31597</a> |
| 9  | FOSS-prosjekt dokumentasjon fra bl.a. arbeidet med virksomhetsmodell for SSB, kartlegging, arkivering, tilgang, FORUM m.fl.                                                                                                   |
| 10 | SSBs websider om mikrodata. <a href="http://www.ssb.no/mikrodata/">http://www.ssb.no/mikrodata/</a>                                                                                                                           |
| 11 | SSBs websider om metadata. <a href="http://www.ssb.no/metadata/">http://www.ssb.no/metadata/</a>                                                                                                                              |
| 12 | Kurs og kursdokumentasjon i DDI, Data Document Initiative. <a href="http://www.ddialliance.org/">http://www.ddialliance.org/</a>                                                                                              |
| 13 | Møtereferater fra Arbeidsgruppe Data til Forskning v/Anders Grøndahl av 18 okt 2007.                                                                                                                                          |

### Akronymer

|      |                                                     |
|------|-----------------------------------------------------|
| SKD  | Skattedirektoratet                                  |
| NSD  | Norsk samfunnsvitenskaplige datatjeneste (i Bergen) |
| SPSS | Programvare for bl.a. statistiske analyser          |
| SAS  | Programvare for bl.a. statistiske analyser          |
| ROI  | Return on investment                                |
| DDI  | Data Document Initiative                            |

### Terminologi

|               |                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Beste Praksis | <p>Med Beste Praksis menes de metoder, verktøy, prosesser, insentiver de mest suksessrike / effektive organisasjoner benytter.</p> <p>(Wikipedia: <b>Best Practice</b> is an idea that asserts that there is a <a href="#">technique</a>, method, process, activity, incentive or reward that is more effective at delivering a particular outcome than any other technique, method, process, etc.)</p> |
|---------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|



## TECHNICAL REPORT

|                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| State Of The Art             | <p>Med State Of The Art menes en beskrivelse av det fremste akademika har av kunnskap og løsninger som de beste metoder, tilnærminger og løsninger innen en teknologi og/eller et fagområde. Løsninger og metoder kan være rent teoretiske og trenger ikke være tatt i bruk av organisasjoner.</p> <p>Wikipedia: The State Of The Art is the highest level of development, as of a device, technique, or scientific field, achieved at a particular time. It also applies to the level of development (as of a device, procedure, process, technique, or science) reached at any particular time usually as a result of modern methods.</p> <p>Webster: the level of development (as of a device, procedure, process, technique, or science) reached at any particular time usually as a result of modern methods</p> |
| Mikrodata                    | Mikrodata er en generell term for data på individnivå om enkeltpersoner, husholdninger, bedrifter/virksomheter etc.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Identifiserbare opplysninger | Opplysninger der det er mogleg å identifisere fysiske eller juridiske personar ved hjelp av identifikasjonsnøkkel (fødselsnr., organisasjonsnr., namn, eksakt adresse) (kilde <a href="http://www.ssb.no/mikrodata/">http://www.ssb.no/mikrodata/</a> )                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Aidentifiserte opplysninger  | Opplysninger der data som blir brukte som identifikasjonsnøkkel er fjerna eller skjult. Ei fil der SSB har kryptert fødselsnummer, vil vere aidentifisert. I forhold til personopplyvningsloven vil både aidentifiserte og krypterte opplysninger vere personopplyvnings, og dermed underlagt det same regelverket som identifiserbare opplysninger. (kilde <a href="http://www.ssb.no/mikrodata/">http://www.ssb.no/mikrodata/</a> )                                                                                                                                                                                                                                                                                                                                                                                 |
| Anonyme opplysninger         | Opplysninger der så mange kjennemerke/kategoriar er fjerna at ein med rimelege middel ikkje direkte eller indirekte kan identifisere fysiske eller juridiske personar. (kilde <a href="http://www.ssb.no/mikrodata/">http://www.ssb.no/mikrodata/</a> )                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Public use data files        | Public use data files are data files prepared by investigators or data suppliers with the intent of making them available for public use. The data available to the public are not individually identified or maintained in a readily identifiable form. <a href="http://www.hhs.gov/ohrp/nhrpac/documents/dataltr.pdf">http://www.hhs.gov/ohrp/nhrpac/documents/dataltr.pdf</a>                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Variabel                     | En variabel er en egenskap ved en statistisk enhet, og den kan anta mer enn én verdi, i form av en numerisk verdi (kvantitativ variabel) eller en kategori fra en klassifisering (kategorisk variabel). Kilde [4]                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Interoperability             | Interoperability is the ability of disparate and diverse organisations, to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organizations via the business processes they support, by means of the exchange of data between their respective information and communication technology (ICT) systems. [4]                                                                                                                                                                                                                                                                                                                                                                                                                                  |

---

## TECHNICAL REPORT

---

### 9 APPENDIKS: SEMICOLON FORSKNINGSMETODE FOR CASE

Casene er på mange måter bærebjelken i Semicolon. De danner det empiriske grunnlaget og står derfor sentralt i produksjonen av resultater og funn. Det er viktig at casene behandles på en metodisk måte, og at metoden brukes gjennomført slik at casene kan sammenlignes.

Metoden for case-studiene i Semicolon ligner på mange måter metoder for systemutvikling. I begynnelsen av prosjektet skal fokuset være på analyse. Etter hvert beveger man seg mot design og konstruksjon, og nye prosesser og løsninger testes og evalueres. På basis av dette vil Semicolons bidrag til "Beste Praksis" og "State Of The Art" kunne beskrives og publiseres. I hver av casene er det så langt mulig også fire tematiske områder som skal behandles:

1. Organisatorisk interoperabilitet.
2. Semantisk interoperabilitet.
3. Informasjonsmodellering.
4. Verktøy og metodestøtte.

Case-studiene gjennomfører kvalitative studier ved bruk av for eksempel semistrukturerte informantintervjuer og fokusgrupper. Det anbefales at intervjuguider justeres etter hvert som case-deltakerne får økt forståelse for nåsituasjonen og dens problemstillinger. Videre foreslås det at funnene systematiseres i

- Kritiske suksessfaktorer som identifiseres og som enten er på plass i nåsituasjon eller som må på plass i ønsket situasjon.
- En problemanalyse som gjerne visualiseres som såkalt fiskeben (delproblemer innenfor temaområder som leder til det grunnleggende problemet).

Resultater fra intervjuer og eventuelt fokusgrupper kan gi input til design av en survey blant offentlige aktører. En slik survey kan da kjøres som en web-basert spørreundersøkelse.

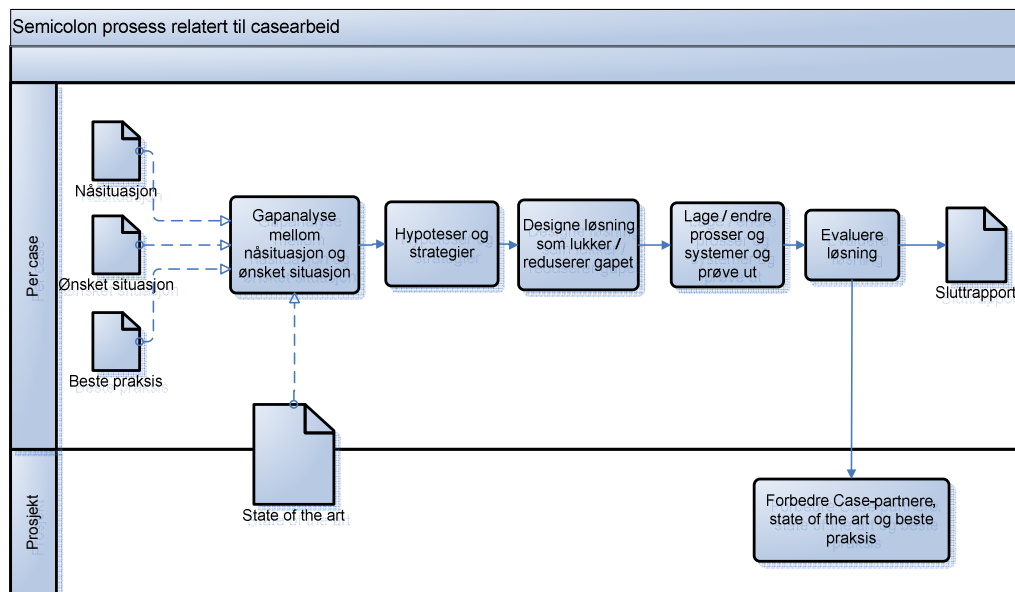
Når analysefasen av case-studier er gjennomført, rettes blikket mot ønsket situasjon, de framtidige løsninger, mulighetsrommet som beskrives i Beste Praksis og State Of The Art dokumentasjonen. For å fortsette med kvalitative metoder foreslås bruken av scenarier, brukscase, tankekart for informasjonsmodellering og evt. illustrasjoner av mulige brukergrensesnitt. Slik kan brukerne fortsatt delta i tidlige faser av designoppgaver, før man setter i gang med:

- Metodevalg og bruk av formelle metoder.
- Modellering av prosesser, ansvar, roller, tilstander, regelverk, avtaler, avhengigheter etc.
- Identifisere og beskrive hindringer og design av nødvendige tiltak.
- Definere informasjonsbehov og tilhørende kvalitetskriterier.
- Skissere utvekslingsscenarier for informasjon.
- Designe ulike typer arkitekturer.
- Sette krav til ulike typer verktøystøtte.
- Beskrive ulike typer og nivåer av interoperabilitet.
- Med mer.

For hver case vil det bli utarbeidet beskrivelser av nåsituasjon og ønsket situasjon. Semicolon prosjektet vil bl.a. basert på kunnskap fra casene utarbeide en beskrivelse av Beste Praksis og State Of The Art. Avgrensning av Beste Praksis og State Of The Art vil bli påvirket av casenes problemområder og de tematiske faglige områdene som Semicolon fokuserer på. Det lages en gap-analyse som er input til design og kravbeskrivelser for etatens fremtidige prosesser og systemer, eller til deler

## TECHNICAL REPORT

av disse prosesser og systemer. Etaten, sammen med Semicolon, kan etablere en pilot, men etaten er ansvarlig for utvikling og tilgang til evt. programvare, test og driftsmiljø. Tilsvarende vil Semicolon-prosjektet kunne bidra inn i aktiviteter som skal lede frem mot produksjonsrutiner og systemer, men etaten vil ha ansvaret for gjennomføringen. Løsninger evalueres og etatens og Semicolons bidrag til Beste Praksis og State Of The Art publiseres. Videre vil resultatene spilles inn til de partnere som er potensielle brukere av resultatene. Hver case vil ha en evaluering og sluttrapport, denne kan være i form av en artikkel.



Figur 3, Skisse av Semicolon forskningsmetode for case

Semicolon forskningsbidrag vil være knyttet til forbedringer og endringer hos partnerne forbedringer av State Of The Art og endringer i Beste Praksis. Bidragene skal gi prosjektpartnerne gevinster på kort og eller lang sikt.

- o0o -