

Noen sentrale begreper knyttet til metadata – til bruk i SSBs felles metadatasystemer

av Anne Gro Hustoft og Hans Viggo Sæbø

Innhold

Bakgrunn	2
Om definisjonene	2
Statistikk	4
Statistisk mål	4
Indikator	4
Indeks	4
Kvalitet	5
Kvalitet i statistikken	5
Metadata	6
Statistikkmetadata	6
Statistikkmetadatasystem	6
Register	6
Basisregister	7
Administrativt register	7
Populasjon	8
Statistisk enhet	9
Identifikator	9
Variabel	10
Måleenhet	11
Klassifikasjon	11
Kodeliste	12
Referanser	13

Bakgrunn

I arbeidet med metadatastrategi for SSB ble det klart at selv om statistikken langt på vei bygger på et felles begrepsapparat, mangler vi en samlet oversikt over sentrale begreper knyttet til metadata. Dette kan føre til at personer med ulike fagbakgrunn og knyttet til ulike arbeidsoppgaver i SSB, snakker forbi hverandre. Dette kan igjen lede til misforståelser og effektivitets- og kvalitetstap ved utveksling av data og metadata, og ved organisering og innlegging av metadata i SSBs felles metadata- eller metadatatrevne systemer som Statistikkbanken, Stabas og Vardok.

Etablering og dokumentasjon av begreper knyttet til metadata, er en viktig del av arbeidet med oppfølging av metadatastrategien. Dette notatet er et første innspill i dette arbeidet. Det er utarbeidet i nær kontakt med metode, IT, formidling og fagstatistikere, og revidert etter kommentarer fra de fleste av avdelingene i SSB, innhentet i en høringsrunde. Notatet er også diskutert i Metadataforum, i Styringsgruppa for metadatastrategien og i Standardutvalget.

En nyttig videreutvikling vil være en drøfting med anbefaling av hvordan disse begrepene bør tolkes i det praktiske arbeidet med systemer som er sentrale i denne sammenhengen.

På et mer detaljert nivå trengs det presise definisjoner også for kommunikasjon mellom maskiner. Her vil det bli utført et definisjonsarbeid i prosjektet "Tjenestebibliotek for metadatasystemer". Dette skal henge sammen med det mer overordnede arbeidet med begreper.

En av grunnene til at folk med ulike arbeidsoppgaver kan snakke forbi hverandre, er at tolkningen av sentrale begreper avhenger av formål, ståsted i organisasjonen og detaljeringsnivå. Statistikken publiseres tradisjonelt i form av tabeller, og deler av begrepsapparatet er utviklet for å betegne innholdet i tabeller (se publiseringshåndboka, håndbok 64). De samme begrepene kan ha eller gis et litt annet innhold på mikronivå, dvs. når vi karakteriserer innsamlede data som legges i datafiler. En kilde til misforståelser kan også være at noen begreper som er nært knyttet til og utviklet innen hhv. personstatistikk og makroøkonomisk statistikk (som nasjonalregnskap), tolkes ulikt innen disse fagdisiplinene. Noen begreper bør på bakgrunn av disse problemene ikke inngå i et *felles* begrepsapparat, for andre bør bruksområdet spesifiseres. Men for de fleste er det mer snakk om å presisere definisjon og tolkning enn å endre begrepene. Dette er forsøkt gjort i det følgende.

Om definisjonene

I arbeidet har en tatt utgangspunkt i internasjonale definisjoner og sedvane i SSB og ellers. Mange av definisjonene er samlet av SDMX-initiativet (Standard Data and Metadata eXchange) som er et samarbeid mellom Eurostat, den europeiske sentralbanken, BIS, OECD, IMF, FN og Verdensbanken, se SDMX i referanselista. Definisjonene her er ofte hentet fra andre kilder. Der det er relevant, er originalkilden oppgitt sammen med SDMX som kilde, mens andre kilder som gjengir samme definisjon, ikke er oppgitt.

Alternativt kunne en ta utgangspunkt i SSBs tradisjonelle definisjoner som er dokumentert i "Innføring i praktisk statistikk" (håndbok fra 1980). Noen av definisjonene herfra er gjengitt i publiseringshåndboka. Disse er som regel ikke gitt som klart formulerte definisjoner, men framgår implisitt i teksten. *Det er imidlertid ikke noe her som er i motstrid med de internasjonale definisjonene.* De eksisterende håndbøkene gir en god bakgrunn for norsk oversettelse av internasjonale begreper. Ofte finnes det synonymer, og disse er angitt i notatet.

Begrepene som defineres i notatet kan grovt sett grupperes som begreper knyttet til overordnede termer som statistikk, metadata, kvalitet og registre, og som begreper knyttet til de enkelte enhetene statistikken bygger på, og deres egenskaper. De overordnede begrepene er definert først. Der det er aktuelt med krysshenvisninger, er dette løst ved lenking i det elektroniske notatet. På papir er lenkene understreket.

De fleste begrepene er angitt i en overskrift med uthevet skrift (som det første begrepet "statistikk"). Noen begreper kan lettest defineres i sammenheng med andre. I en slik sammenheng kan det være lett å

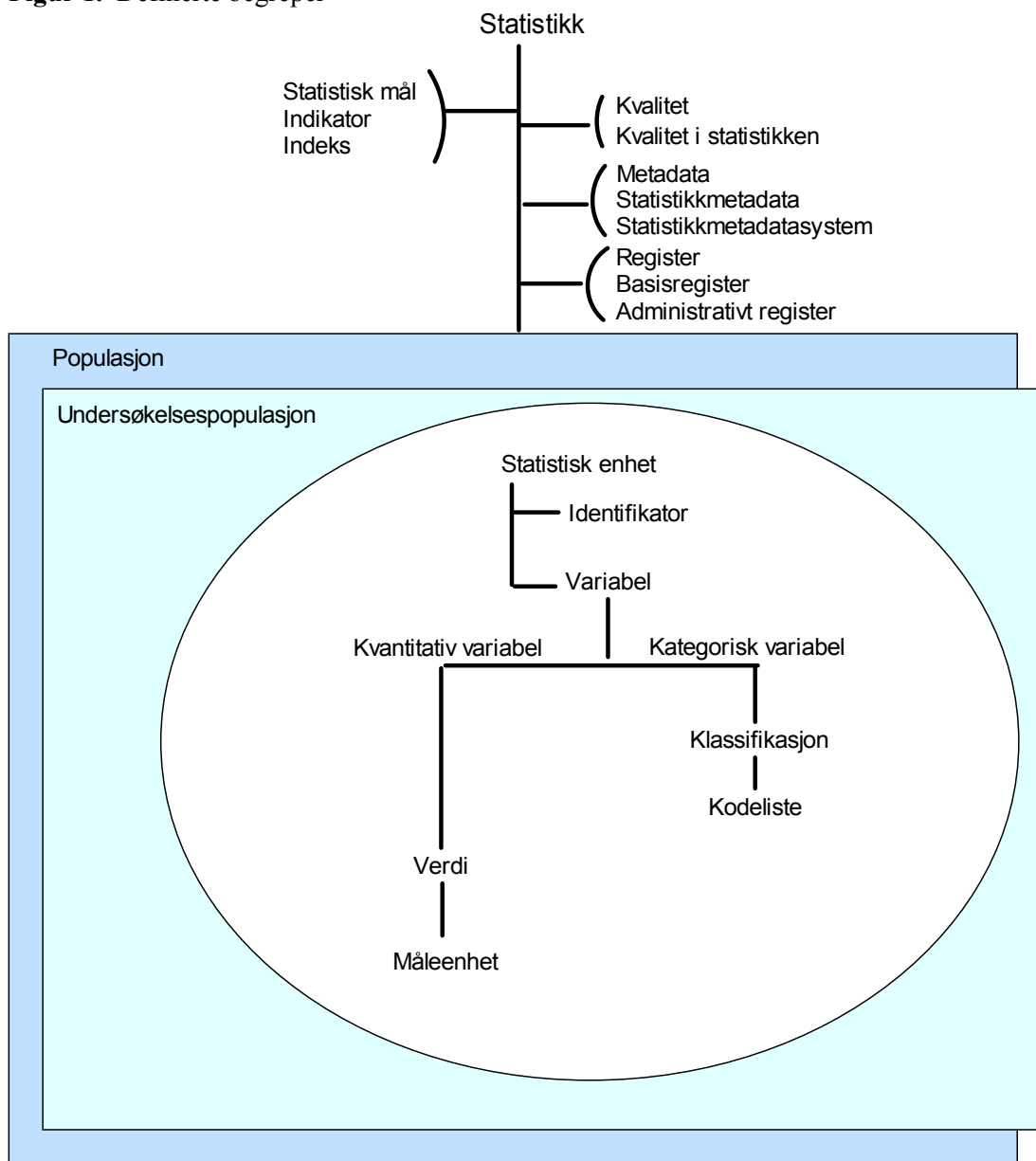
forstå hva som menes med et begrep, mens det er vanskelig å definere det presist under en egen overskrift (eksempel på et slikt begrep er "data" som bare er omtalt i avsnittet om "statistikk").

Definisjonene med kommentarer dekker to spalter. Venstre spalte inneholder selve definisjonen, på norsk og, hvis den finnes i internasjonale kilder, også på engelsk. Kilden er også angitt. I venstre spalte finner en ellers korte eksempler og synonymer. Høyre spalte er en kommentarspalte. Her nevnes f.eks. tilgrensende begreper som nyttes på spesielle områder. Noen begreper er, som nevnt i forrige avsnitt, bare definert gjennom omtale her.

Begrepene som er tatt med, er de hvor det erfaringsmessig kan oppstå misforståelser i forhold til hva og hvordan metadata skal legges inn i SSBs sentrale metadatasystemer, samt noen generelle definisjoner knyttet til statistikk og metadata. Mange andre begreper kunne vært definert. I denne omgang vil vi vise til SDMX som i Metadata Common Vocabulary (MCV) gir over 300 definisjoner på engelsk. Kilder i SSB med generelle begrepsdefinisjoner omfatter i tillegg til de nevnte håndbøkene om praktisk statistikk og publisering, den nye revisjonshåndboka (håndbok 84). I tillegg kommer mer fagspesifikke kilder som manualer for nasjonalregnskap, definisjoner i kredittmarkedsstatistikken og begrepsdefinisjonene som er gitt i NOS publikasjoner. Disse omfatter i hovedsak selve metadataene, mens dette notatet bare omhandler generelle begreper for å karakterisere disse ("meta metadata").

Figuren viser begrepene i den rekkefølgen de er definerte i dette notatet.

Figur 1. Definerte begreper



DEFINISJONER	KOMMENTARER
Statistikk <i>Statistikk er tallfestede opplysninger om en gruppe eller et fenomen, som fremkommer ved sammenstilling og bearbeiding av opplysninger om de enkelte enhetene i gruppen eller et utvalg av disse enhetene, eller ved systematisk observasjon av fenomenet.</i> Kilde: Statistikkloven	Statistikk baserer seg på, og innebærer normalt en sammenfatning av, <i>data</i>. I en statistikk der f.eks. den <u>statistiske enheten</u> er foretak, vil dataene være verdiene som innhentes for hver <u>variabel</u> (f.eks. antall ansatte) for hvert enkelt foretak. Disse dataene, som ofte kalles mikrodata, lagres i filer eller registre, og bearbeides videre til statistikk.
Statistisk mål <i>Et statistisk mål er en sammenfatning av de (kvantitative) variabelverdier de <u>statistiske enhetene</u> har.</i> Kilde: SDMX: <i>A summary (means, mode, total, index etc.) of the individual quantitative variable values for the statistical units in a specific group (study domains).</i> Eksempler: Gjennomsnitt, medianverdi, totaltall, feilmargin, <u>indeks</u>.	Statistikken uttrykkes gjennom statistiske mål.
Indikator <i>En indikator er et mål som er avledet av data og/eller statistikk og indikerer status eller utvikling på et angitt område.</i> Eksempler: Indikatorer for bærekraftig utvikling, sosiale indikatorer, konjunkturindikatorer, strukturindikatorer (i Europa)	En indikator vil ofte være basert på statistikk og kan være et tall fra en statistisk tabell eller et <u>statistisk mål</u>. Indikatorverdien må ofte sammenliknes med noe (som er angitt på forhånd) for å ha mening. Informasjonsnivået knyttet til en indikator skal være høyt. Det vil si at en indikator, i større grad enn statistikken den baserer seg på, kan tolkes som kunnskap. Det er viktig at området eller temaet indikatoren skal belyse, er angitt presist.
Indeks <i>En indeks er et tall som viser endringer i en størrelse over tid eller rom.</i> Kilde: SDMX og ISI (2003): <i>A quantity that shows by its variations the changes of a magnitude over time or space</i> Eksempler: Prisindeks (som konsumprisindeksen), produksjonsindeks, likestillingsindeks	En indeks kan være av ulike typer. Når en konstruerer en indeks, er det viktig å definere hva den skal dekke, basisperiode, vekting og metode for gjennomsnittsberegning. Verdien av en indeks i basisperioden, eventuelt referanseområdet, settes ofte lik 100.

DEFINISJONER	KOMMENTARER
Kvalitet <i>Kvalitet er helheten av egenskaper ved et produkt eller en tjeneste som vedrører dets evne til å tilfredsstille krav eller behov.</i> Kilde: ISO8402 og Norges standardiseringsforbund (NS) SDMX: <i>Quality is defined as the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs</i>	Det fins flere måter å formulere en definisjon av kvalitet på, men felles for alle er at de tar <i>utgangspunkt i brukerbehov</i> for de aktuelle produktene eller tjenestene. Men det er mange brukere med ulike behov, og produsenten av en vare eller tjeneste må ta hensyn til og balansere disse mot hverandre. Ofte setter produsenten opp et sett med krav. Når det gjelder kvalitet i statistikken, har de fleste statistikkbyråer laget en liste med faktorer som også tar hensyn til helheten i statistikken (se kvalitet i statistikken).
Kvalitet i statistikken SSB definerer <i>kvalitet i statistikken</i> ved å vise til ulike kriterier/dimensjoner: ➤ <i>Relevans</i> ➤ <i>Nøyaktighet</i> ➤ <i>Aktualitet og punktlighet</i> ➤ <i>Sammenliknbarhet og sammenheng</i> ➤ <i>Tilgjengelighet og klarhet</i> Kilde: SDMX og Eurostat (2003): ➤ <i>Relevance</i> ➤ <i>Accuracy</i> ➤ <i>Timeliness and punctuality</i> ➤ <i>Comparability and coherence</i> ➤ <i>Accessibility and clarity</i>	Disse dimensjonene svarer til ulike egenskaper i den generelle kvalitetsdefinisjonen. Helhet/bredde i statistikken dekkes av relevans. Sammenliknbarhet går på tid og rom, mens sammenheng går på ulike statistikker som må være basert på felles standarder for å kunne sammenliknes og passe sammen i systemer som nasjonalregnskapet. Sammenheng refererer også til sammenhengen mellom foreløpig og endelig statistikk. Klarhet går på at det skal være lett å forstå og kunne bruke statistikken, altså at statistikken følges av god dokumentasjon eller metadata. Denne lista omfatter kvalitetskrav til <i>produktet</i> offisiell statistikk. Disse kravene inngår i Eurostats ”Code of Practice” for offisiell statistikk og statistikkbyråer. I tillegg fins det krav til <i>prosessene</i> og til <i>institusjonene</i> som produserer statistikken. Viktige krav til prosessene er: • Statistikken skal bygge på de beste metoder og gode rutiner • Kostnadseffektivitet • Lav oppgavebyrde Krav til institusjonelle forhold omfatter bla.: • Offisiell statistikk skal utarbeides på et uavhengig grunnlag. • Datasikkerhet og personvern Faktorene som er knyttet til statistikkproduktene og definert i venstre spalte, omtales ofte som <i>produktkvalitet</i>. Vi snakker ellers om <i>prosesskvalitet</i> som dekker kravene til prosessene, og <i>strukturkvalitet</i> som dekker institusjonelle forhold. IMF, FN og OECD opererer med litt ulike kvalitetsdimensjoner, men hovedelementene er de samme som hos Eurostat.

DEFINISJONER	KOMMENTARER
Metadata <i>Metadata er data om data.</i> Kilde: SDMX og ISO/IEC FCD 11179-1: <i>Metadata is data that defines and describes other data.</i> Synonymer: Dokumentasjon, informasjon (om data)	
Statistikkmetadata <i>Statistikkmetadata er strukturert informasjon om statistikk. Dette omfatter informasjon som brukes for å produsere, formidle, forstå, finne eller (gjen)bruke statistikk.</i>	I forhold til den generelle ISO-definisjonen av metadata har vi for statistikken inkludert et krav om <i>struktur</i>. Struktur er nødvendig for bruk av informasjon i et IT-system, men også en forutsetning for effektiv bruk av metadata, f.eks. for å finne statistikk på Internett blant store informasjonsmengder, og forstå den.
Statistikkmetadatasystem <i>Et statistikkmetadatasystem er et databehandlingssystem som bruker, lagrer og produserer statistikkmetadata.</i> Kilde: SDMX og UNECE (2000): <i>A statistical metadata system is a data processing system that uses, stores and produces statistical metadata.</i>	
Register <i>Et register gir en (ideelt sett) fullstendig fortegnelse over de <u>statistiske enhetene</u> innenfor en bestemt <u>populasjon</u>, og beskriver disse vha. ulike <u>variabler</u>. Alle de statistiske enhetene i et register har en identifikator som gjør det mulig å oppdatere registeret med nye opplysninger om de statistiske enhetene.</i> Kilde: En kombinasjon av UNECE (2000) og definisjon fra SCB (2004) I Sikkerhetshåndboka står følgende definisjon: <i>Et register består av fortegnelser, lister, oversikter m.v. der informasjonen er lagret systematisk slik at opplysninger enkelt kan finnes igjen.</i>	Statistisk sentralbyrås registre skal meldes til sikkerhetsråd giver og filene dokumenteres i Datadok. For dette formålet brukes registerdefinisjonen i Sikkerhetshåndboka. Datatilsynet og Brønnøysundregistrene bruker også denne definisjonen. De to definisjonene vi her har referert, avviker, men står ikke i motsetning til hverandre. En spesiell type registre er <i>forløpsdatabaser</i>, der data er strukturert og tilrettelagt for analyse av sammenhenger over tid (flyttinger, sivilstandsendringer etc.).

DEFINISJONER	KOMMENTARER
Basisregister <i>Et basisregister er et register som definerer og identifiserer grunnleggende statistiske enheter.</i> I Strategi 2002- står det: <i>Det finnes i dag tre basisregistre:</i> • <i>Det sentrale folkeregistret i Skattedirektoratet, DSF (omfatter personer og familier)</i> • <i>Enhetsregisteret i Brønnøysund (juridiske enheter/foretak og bedrifter)</i> • <i>Registeret over Grunneiendom- Adresse- og Bygning, GAB, i Statens kartverk</i>	De tilsvarende <i>statistiske registre</i> eller <i>populasjonsregistre</i> vi bruker i SSB er: • BeReg (DSF pluss noen SSB-spesifikke tilleggsenheter og variabler) • BoF: Det sentrale bedrifts-og foretaksregisteret (Enhetsregisteret pluss noen SSB-spesifikke tilleggheter og variabler) • SSB-GAB (GAB-registeret pluss hovedsakelig FoB-variabler) Se for øvrig omtale på Byrånettet under Faglig arbeid/Registre
Administrativt register <i>Et register som er opprettet av et forvaltningsorgan for bruk i organets virksomhet, kalles et administrativt register eller datasystem.</i> Definisjonen bygger på den mer omfattende definisjonen som står i forskriften knyttet til statistikkloven: <i>Med et administrativt datasystem forstås opplysninger som organer i statsforvaltningen og landsomfattende kommunale organisasjoner samler inn og oppbevarer på en slik måte at opplysningene kan hentes frem til bruk i organets eller organisasjonens virksomhet.</i>	Denne typen register vil normalt eies av en annen offentlig institusjon enn SSB, men vi kan etter statistikkloven utnytte slike datasystemer som grunnlag for offisiell statistikk. Ofte lager SSB sin egen variant av slike registre ved uttrekk og/eller tillegg som grunnlag for statistikkproduksjonen. Vi snakker da om <i>statistiske registre</i> (som SSBs versjoner av basisregistre beskrevet over).

DEFINISJONER**KOMMENTARER****Populasjon**

En populasjon er samlingen av de statistiske enheter statistikken skal gi informasjon om

Det kan skilles mellom to typer populasjoner:

- *Målpopulasjonen er den populasjonen en ønsker å lage statistikk for*
- *Undersøkelsespopulasjonen er den populasjonen informasjonen kan skaffes for i undersøkelsen*

I publiseringshåndboka tar en også med at populasjonen er tid- og stedfestet.

Kilde: SDMX og FNs statistiske kontor:

Population is the total membership or population or "universe" of a defined class of people, objects or events.

- *Target population is the population outlined in the survey objects about which information is to be sought*
- *Survey population is the population from which information can be obtained in the survey*

Eksempler på målpopulasjoner: Alle bedrifter i Norge i 2000, alle adressene og eiendommene i Norge i 2001, alle personer i Norge i aldersgruppen 16-74 år i 2005

Tilhørende undersøkelsespopulasjoner: Alle aktive bedrifter i BoF i 2000, alle adresser og eiendommer i GAB i 2001, alle personer i aldersgruppen 16-74 år i BeReg i 2005.

Synonymer: Masse, univers

Eventuelle avvik mellom målpopulasjon og undersøkelsespopulasjon gir opphav til såkalt *dekningsfeil* i statistikken. Disse avvikene kan f.eks. skyldes at noen elementer mangler pga. forsinket oppdatering i et register.

Fra undersøkelsespopulasjonen vil en i utvalgsundersøkelser trekke et *utvalg* der de statistiske enhetene velges ut fra en prosess basert på trekk sannsynligheter.

En liste eller en annen spesifisering av de statistiske enhetene som tilhører populasjonen som skal telles, eller trekkes utvalg fra, kalles *rammen* for undersøkelsen. Rammen *spesifiserer* undersøkelsespopulasjonen, men har et litt annet begrepsinnhold enn denne, som er *samlingen* av de statistiske enhetene. Forskjellen er semantisk, og en ser ofte at begrepene undersøkelsespopulasjon og ramme brukes synonymt. Men rammen kan i noen sammenhenger også inkludere annen informasjon som er nødvendig for å trekke et optimalt utvalg (som enhetenes adresser og avlede transportkostnader for å nå dem).

DEFINISJONER	KOMMENTARER
Statistisk enhet <i>En statistisk enhet er bærer av de statistiske egenskapene i en statistisk undersøkelse eller et <u>register</u></i> Kilde: SDMX og UNECE (2000): <i>An object of statistical survey and the bearer of statistical characteristics. The statistical unit is the basic unit of statistical observation within a statistical survey.</i> Eksempler: Person, husholdning, bedrift, foretak og veitrafikkulykke En kan skille mellom tre typer statistiske enheter: • <i>Observasjonsenheten er den enheten som informasjon mottas for</i> • <i>Rapporteringsenheten er den enheten som innrapporterer data</i> • <i>Analyseenheten er en virkelig eller konstruert enhet som statistikken er utarbeidet eller formidlet for.</i> Kilde: SDMX og UNECE (2000): • <i>Observation unit are those entities on which information is received and statistics are compiled.</i> • <i>Reporting unit is a unit that supplies the data for a given survey instance</i> • <i>Analytical units represent real or artificially constructed units, for which statistics are compiled.</i> Synonym for statistisk enhet: Telleenhet	En statistisk enhet skal være unik og identifiserbar. I Nasjonalregnskapets terminologi defineres <i>enhet</i>, <i>økonomisk objekt</i> og <i>hendelse</i> som egne begreper. <i>Enhet</i> er i Nasjonalregnskapet definert som en økonomiske aktør (person, selskap etc.), <i>økonomisk objekt</i> er en verdi (vare, tjeneste, finansobjekt etc.), mens <i>hendelse</i> er en begivenhet som normalt fordrer en handling utført av enheten (produksjon, ulike transaksjoner etc.). En observasjonsenhet er ikke nødvendigvis "observert" av en som samler inn data i eller for SSB. Informasjonen ligger ofte i et <u>administrativt register</u>, eller kommer via andre, f.eks. kommune, fiskemottak eller formann i et elgvald. Vi kan da snakke om en <i>rapporteringsenhet</i> som kan være forskjellig fra observasjonsenheten. Et annet eksempel er et foretak (rapporteringsenhet) som gir data for sine bedrifter (observasjonsenheter). Det kan være flere analyseenheter knyttet til ett datagrunnlag, og observasjonsenheten er ofte også analyseenhet. En analyseenhet kan imidlertid også bestå i aggregeringer eller (beregnete) oppdelinger av observasjonsenheter. En statistisk enhet må ikke forveksles med en <u>måleenhet</u> (kg, meter, osv.).
Identifikator <i>En identifikator gir en entydig identifikasjon av de <u>statistiske enhetene</u> statistikken omfatter.</i> Eksempler: Fødselsnummer og organisasjonsnummer Synonym: Ident, identifikatorvariabel	En identifikator kalles noen ganger for identifikatorvariabel. Den er imidlertid ikke en <u>variabel</u> i vanlig forstand, fordi poenget ikke er å beskrive en "naturlig" variasjon mellom statistiske enheter, eller over tid for én statistisk enhet (slik en f.eks. gjør vha. alder), men å gi den statistiske enheten en verdi som skal skille denne enheten fra alle andre enheter av samme type.

DEFINISJONER**KOMMENTARER****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 klassifikasjon (kategorisk variabel)

Kilde: SDMX og FNs statistiske kontor:

A variable is a characteristic of a unit being observed that may assume more than one of a set of values to which a numerical measure or a category from a classification can be assigned (e.g. income, age, weight, etc. and "occupation", "industry", "disease" etc).

Eksempler på variabler: Inntekt, alder, mengde olje eller fisk, yrke, næring, dødsårsak.

Synonymer: kjennemerke, egenskap, attributt

I forhold til den engelske definisjonen har vi fjernet kravet om at den statistiske enheten nødvendigvis observeres ("unit being observed"). Vi bruker i tillegg "verdi" i vid forstand av ordet, slik at f.eks. spørsmål i skjema med avkryssingbokser kan ha verdien 1 dersom boksen er fylt ut, og verdien "ingenting" dersom den ikke er fylt ut.

Begrepene *statistikkvariabel* og *klassifikasjonsvariabel* brukes i SSB i Statistikkbanken for å strukturere tabellene.

- *Statistikkvariabler antar kvantitative verdier (f.eks. Inntekt)*
- *Klassifikasjonsvariabler har verdier knyttet til en klassifikasjon (f.eks. Yrke)*

Verdiene en klassifikasjonsvariabel kan anta, beskrives vha. en klassifikasjon, som f.eks. Standard for kommuneinndeling, eller bare ved et sett koder. Dersom en oppgir inntekt ved å bruke en klassifikasjon, f.eks. høy inntekt, middels inntekt og lav inntekt, kan inntekt gå over fra å være en statistikkvariabel til å bli en klassifikasjonsvariabel. Slik kan en variabel skifte mellom "rollene" som statistikkvariabel og klassifikasjonsvariabel alt etter hvilken framstilling som er relevant i en gitt sammenheng.

Som oftest vil tallene i en tabell vise verdiene for en statistikkvariabel, mens forspalten og/eller tabellhodet angir en eller flere klassifikasjonsvariabler.

Når verdien av en variabel er et tall, er det ofte en måleenhet knyttet til verdien. En verdi i form av en kategori er en betegnelse *eller* en kode som svarer til denne. Når det gjelder verdien for kategoriske variabler (klassifikasjoner), antar vi at denne kan angis både som betegnelse (som "mann") eller som kode ("1" i dette tilfellet).

Eksempler på verdier: "300 000 kroner"(inntekt), "1000 tonn" (oljeproduksjon), "42 år" (alder), "mann" med kodeverdien 1 (kjønn).

Mange variabler finnes også som *tidsserier*. Tidsserier framkommer ved at ulike fenomener (variabler) observeres over tid.

En spesiell type variabler er dato- eller gyldighetsvariabler. De står i sammenheng med en annen variabel (f.eks. næring) og angir den dato variabelens verdi er gyldig fra.

DEFINISJONER	KOMMENTARER
Måleenhet <i>En måleenhet er det en <u>variabel</u> angis i.</i> Eksempler på måleenheter: 1000 NOK, Euro, tonn, liter	Ofte har en også en faktor inkludert i måleenheten slik at måleenheten kan være enten NOK eller 1000 NOK. Måleenheter kan grupperes i ulike typer som vekt, valuta, lengde etc., og en måleenhet kan regnes om til en annen måleenhet så lenge de tilhører same type.
Klassifikasjon <i>En klassifikasjon er ett sett av diskrete, uttømmende og gjensidig utelukkende kategorier som en <u>variabel</u> kan anta.</i> Kilde: SDMX og FNs statistiske kontor: <i>A classification is a set of discrete, exhaustive and mutually exclusive observations, which can be assigned to one or more variables to be measured in the collation and/or presentation of data. The terms "classification" and "nomenclature" are often used interchangeably, despite the definition of a "nomenclature" being narrower than that of a "classification".</i> Eksempler på klassifikasjoner: Standard for næringsgruppering (SN), Standard for yrkesklassifisering (STYRK), Norsk standard for utdanningsgruppering (NUS), Standard for familietype	I en klassifikasjon vil kategoriene ofte være identifisert ved en kode og alltid ved en betegnelse. (Se egen definisjon av <u>kodeliste</u>). En klassifikasjon kan bestå av flere nivåer (hierarkisk). I praksis er det <i>klassifikasjonsversjoner</i> vi bruker i vår statistikk. En klassifikasjonsversjon er gyldig i et bestemt tidsrom, og erstattes, når det er nødvendig, av en ny versjon. Det er viktig med nøkler (ofte kalt korrespondansetabeller) mellom ulike versjoner slik at en kan se hvordan en kategori i en versjon har endret seg i neste versjon (f.eks. at "29.100 Produksjon av jern og stål" i SN94 endret seg til "29.100 Produksjon av jern og stål samt ferrolegeringer" i SN2002). Det er ulik praksis når det gjelder å utarbeide nye versjoner av klassifikasjoner. For noen av klassifikasjonene vil det være tillatt å gjøre mindre endringer (f.eks. legge til elementer) uten at en får en ny versjon. Det er bare større revisjoner som berører selve strukturen i klassifikasjonen, som utløser en ny versjon. Dette gjelder f.eks. Norsk standard for utdanningsgruppering (NUS). I andre klassifikasjoner vil alle endringer (bortsett fra korrigering av skrivefeil o.l.) utløse en ny versjon, f.eks. fikk vi en ny versjon av Kommuneinndelingen da 1842 Skjerstad forsvant fra inndelingen ved at den ble innlemmet i 1804 Bodø i 2005. En klassifikasjon kan være en formell standard klassifikasjon (som SN2002), men kan også representere en gruppering av en kvantitativ <u>variabel</u>, som f.eks. alder (selv om vi prøver å standardisere på visse grupperinger) og jordbruksareal. Disse grupperingene lages ofte for å forenkle presentasjonen eller unngå problemer med konfidensialitet Begrepet <i>standard</i> har et videre innhold, men i SSB bruker vi ofte <i>standard</i> som forkortelse for standard klassifikasjon. Alle

	klassifikasjonseksempler i venstre kolonne er standard klassifikasjoner. Disse er vedtatt av administrerende direktør, bygger ofte på internasjonale klassifikasjoner og skal bidra til å gjøre SSBs statistikk sammenlignbar på tvers av statistikkområder.
Kodeliste <i>En kodeliste er en forhåndsdefinert liste som en <u>variabel</u> henter sine verdier fra.</i> Kilde: SDMX: <i>A code list is a predefined list from which some statistical concepts (coded concepts) take their values.</i> Eksempel på kodeliste (brukt i Mediebrukundersøkelsen): 1 Ja 2 Nei 8 Vil ikke svare 9 Vet ikke	Hver kategori i en kodeliste har en unik kode og en unik betegnelse. Det er ikke krav om, som for <u>klassifikasjon</u>, at kategoriene skal være uttømmende og gjensidig utelukkende. En kodeliste i SSB kan være en liste over kategorier som ikke er uttømmende for et område, men som dekker kategorier som er aktuelle for en spesiell undersøkelse. Disse kategoriene kan være gjensidig utelukkende, men trenger ikke være det. Vi har også i våre spørreundersøkelser en del kodelister av typen "smørbrødlister" der kategoriene ikke er gjensidig utelukkende, men uttømmende. I dagligtale bruker en ofte kodeliste om (korte) klassifikasjoner der kodene er tatt med, f.eks. kodelisten for kjønn (1=mann, 2=kvinne).

Referanser

Datarevisjon, Statistisk sentralbyrås håndbøker, nr. 84, 2005

Eurostat (2003): *Assessment of Quality in Statistics: Definition of quality in statistics*, WG, Luxembourg, October 2003

Eurostat (2005): *European Statistics Code of Practice*. Adopted by the SPC 24 February 2005

FNs statistiske kontor: *United Nations Glossary of Classification Terms*, publisert på http://unstats.un.org/unsd/class/family/glossary_short.htm.

Innføring i praktisk statistikk, Statistisk sentralbyrås håndbøker 44, 1980

ISI (2003): *The Oxford Dictionary of Statistical Terms* - Oxford University Press 2003

Publiseringshåndboka, Statistisk sentralbyrås håndbøker, nr. 64, 1997

SCB: Report 2004:2- Registerstatistikk - administrativa data för statistiska syften.

SDMX: *Metadata Common Vocabulary*. Edited by Marco Pellgrino (Eurostat) and Denis Ward (OECD), siste versjon oktober 2005

Sikkerhetshåndboka, Statistisk sentralbyrås håndbøker, nr. 45, 1998

SSBs Metadatastrategi - Sæbø, Andersen, Hoel, Hustoft, Linnerud, Torvbråten: SSB-rapport 2005/2

Statistikkloven av 16. juni 1989, Statistisk sentralbyrås håndbøker 54, 1989

Strategi 2002-, Planer og meldinger 2002/18 fra Statistisk sentralbyrå

UNECE (2000): *Terminology on Statistical Metadata*, Conference of European Statisticians - Statistical Standards and Studies No 53.