

Veileder for beskrivelse av kvalitet på datasett – kvantifiserbar kvalitet

Publisert: 2020-02-20



Innmelding av feil og mangler:

Dersom du finner feil eller mangler i dokumentet, ber vi om at dette meldes inn på [Github Issues](#). Dersom du ikke allerede har bruker på Github kan du opprette bruker gratis.

Innholdsfortegnelse

Innledning	4
Formål, omfang og avgrensninger	4
Målgruppe	4
Sammenheng mellom relevante standarder, spesifikasjoner og denne veilederen	4
Hensyn man bør ta når man beskriver kvantifiserbar kvalitet på datasett	6
Kvalitet kan måles på ulike nivåer	6
Predefinerte kvalitetsmål bruker negativt ladede ord	6
Fritekst kan brukes som supplerende forklaring	6
Hvilke kvalitetsmål som erfaringsmessig er viktige for brukerne å vite	7
Oversikt over predefinerte kvalitetsdimensjoner, kvalitetsdeldimensjoner og kvalitetsmål	8
Kvalitetsdimensjonen «fullstendighet»	9
Kvalitetsdimensjonen «aktualitet»	10
Kvalitetsdimensjonen «konsistens»	11
Kvalitetsdimensjonen «nøyaktighet»	12
Definisjoner av kvantifiserbare kvalitetsmål	13
Formål med definisjonene	13
Avgrensninger	13
Definisjoner av noen sentrale begreper	14
Imputering	14
Enhet	15
Egenskap	15
Element	15
Kvalitetsdimensjon «fullstendighet»	16
Kvalitetsdeldimensjon «underdekning»	16
Kvalitetsmål «manglende enheter»	16
Kvalitetsmål «antall manglende enheter»	17
Kvalitetsmål «andel manglende enheter»	17
Kvalitetsmål «antall enheter med manglende verdi for en gitt egenskap»	17
Kvalitetsmål «andel enheter med manglende verdi for en gitt egenskap»	18
Kvalitetsdeldimensjon «overdekning»	18
Kvalitetsmål «overflødige enheter»	18
Kvalitetsmål «antall overflødige enheter»	19
Kvalitetsmål «andel overflødige enheter»	19
Kvalitetsdeldimensjon «imputering»	20
Kvalitetsmål «antall enheter med imputert verdi for en gitt egenskap»	20
Kvalitetsmål «andel enheter med imputert verdi for en gitt egenskap»	20
Kvalitetsdimensjon «aktualitet»	21
Kvalitetsdeldimensjon «tidsdifferanse»	21

Kvalitetsmål «samlet tidsdifferanse»	21
Kvalitetsdimensjon «konsistens»	22
Kvalitetsdeldimensjon «konsistens innad i datasett»	22
Kvalitetsmål «andel enheter med inkonsistente egenskaper»	22
Kvalitetsmål «andel enheter med inkonsistens mellom gitte egenskaper»	23
Kvalitetsdimensjon «nøyaktighet»	23
Kvalitetsdeldimensjon «identifikatorriktighet»	24
Kvalitetsmål «antall enheter med identifikatorfeil»	24
Kvalitetsmål «andel enheter med identifikatorfeil»	24
Kvalitetsdeldimensjon «klassifikasjonsriktighet»	25
Kvalitetsmål «antall feilklassifiserte enheter for en gitt egenskap»	25
Kvalitetsmål «andel feilklassifiserte enheter for en gitt egenskap»	25

Innledning

Formål, omfang og avgrensninger

Formålet med dette dokumentet er å gi veiledning i felles definisjoner på og måleverktøy for kvantifiserbar kvalitet på datasett. Et felles sett med definisjoner og måleverktøy vil skape en unison beskrivelse og forståelse av datakvalitet på tvers av virksomheter. Hensikten er ikke å utelukke andre måter å beskrive datakvalitet på enn disse presentert i dette dokumentet. Man står fritt til å supplere med andre typer kvalitetsmål i kvalitetsbeskrivelsen.

I henhold til [Spesifikasjon for beskrivelse av kvalitet på datasett](#), er det flere ulike måter å beskrive kvalitet på et datasett på:

- kvantifiserbar kvalitet
- ikke-kvantifiserbar kvalitet
- kvalitet i samsvar med gitt(e) standard(er)/spesifikasjon(er)
- brukertilbakemeldinger knyttet til kvalitet

Spesifikasjonen foreslo også å predefinere kvalitetsdimensjoner, kvalitetsdeldimensjoner og kvalitetsmål. Se forøvrig predefinisjoner i [Definisjoner av kvantifiserbare kvalitetsmål](#).

Dette er ikke en veileder for datakvalitetsarbeidet generelt. Veilederen forklarer hvordan datakvalitet beskrives, med fokus på å bruke predefinerte kvantifiserbare kvalitetsmål. Ved behov vil det bli utarbeidet tilsvarende veiledere for de andre måter å beskrive kvalitet på. Disse vil inngå i [Rammeverk for informasjonsforvaltning](#).

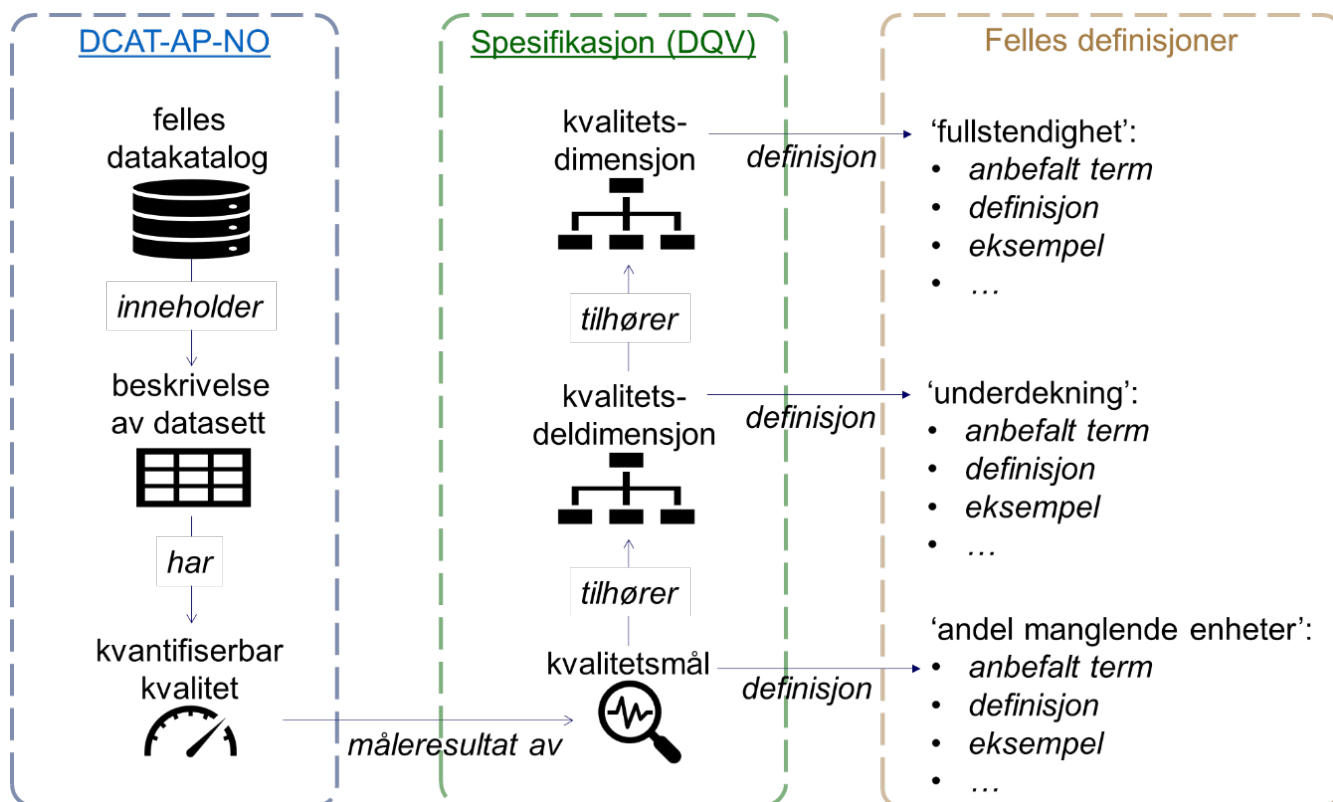
Målgruppe

Målgruppen for veilederen er:

- de som skal beskrive datasett og derunder kvantifiserbar kvalitet på datasett (primær målgruppe)
- de som leverer løsninger og verktøy for registrering og visning av denne type kvalitetsmål (primær målgruppe)
- de som skal forstå beskrivelse av kvantifiserbar kvalitet på datasett (sekundær målgruppe)

Sammenheng mellom relevante standarder, spesifikasjoner og denne veilederen

Veilederen må ses i sammenheng med [Spesifikasjon for beskrivelse av kvalitet på datasett](#).



Tegningen ovenfor illustrerer sammenhengen mellom ulike standarder og spesifikasjoner:

- <https://data.norge.no/specification/dcat-ap-no/>, standard for beskrivelse av datakataloger og datasett: Denne spesifiserer bl.a. hvordan datasett beskrives i en datakatalog, deriblant hvordan en beskrivelse av kvantifiserbar kvalitet knyttes til beskrivelsen av et datasett.



DCAT-AP-NO er i skrivende stund under revisjon, bl.a. for å kunne ta hensyn til beskrivelse av kvalitet.

- [Spesifikasjon for beskrivelse av kvalitet på datasett](#) (basert på DQV): Denne spesifiserer hvordan kvalitet på datasett beskrives, bl.a. å knytte kvalitetsbeskrivelse til kvalitetsdimensjon.



Spesifikasjonen vil bli justert etter revisjon av DCAT-AP-NO.

- Felles definisjoner: Hensikten med felles definisjoner er å skape et felles vokabular for kvalitetsbeskrivelse, slik at beskrivelsene forstås likt. Dette handler om felles definisjoner og forståelse av kvalitetsdimensjoner og kvalitetsdeldimensjoner, og også kvalitetsmål når det gjelder kvantifiserbar kvalitet.



Se ovennevnte spesifikasjonen for forklaring på kvalitetsdimensjon, kvalitetsdeldimensjon og kvalitetsmål.

Denne veilederen forklarer predefinerte kvalitetsmål for kvantifiserbar kvalitet, samt de kvalitetsdeldimensjoner og kvalitetsdimensjoner disse kvalitetsmålene tilhører.

Hensyn man bør ta når man beskriver kvantifiserbar kvalitet på datasett

Kvalitet kan måles på ulike nivåer

Datakvalitet kan måles på både enhetsnivå og egenskapsnivå. I eksempeldatasettet under vil hver bygning være en enhet (markert med rød ramme), mens opplysningene som er knyttet til hver bygning er en egenskap (eksemplifisert med blå ramme for egenskapen «bruksareal»; «eier», «byggeår», «kommune» og «bruttoareal» er også egenskaper). Se for øvrig definisjon på enhet og egenskap i [Definisjoner av kvantifiserbare kvalitetsmål](#).

Der det er relevant, finnes det predefinerte kvalitetsmål for både enhet- og egenskapsnivå. Et eksempel er kvalitetsmålene «antall manglende enheter» og «antall enheter med manglende verdi for en gitt egenskap» under kvalitetsdimensjonen «fullstendighet» som går på henholdsvis enhetsnivå og egenskapsnivå. I eksempeldatasettet under mangler det verdi for egenskapen «bruksareal» for Bygning nr. B00015.

Eksempeldatasett: Bygninger

Identifikator	Eier	Byggeår	Kommune	Bruksareal	Bruttoareal
B00013	Rex Eiendom	1974	Halden	510 m ²	650 m ²
B00014	Oslo kommune	2003	Oslo	2300 m ²	3000 m ²
B00015	Eiendomsutvikling AS	1995	Bergen		1200 m ²

Figur 1. Eksempeldatasett: Bygninger

Predefinerte kvalitetsmål bruker negativt ladede ord

I predefineringen av kvalitetsmål er det valgt å bruke såkalte negativt ladede ord for flere av kvalitetsmålene. Negativt ladede ord tydeliggjør feil og mangler i datakvaliteten.

Et eksempel er kvalitetsmålet «andel manglende enheter» som handler om mangel. Det vil være for eksempel 2 % *mangel* (negativt ladet) istedenfor 98 % *fullstendig* (positivt ladet) som oppgis. Det er viktig å være klar over dette, både ved angivelse av verdier til kvalitetsmål og ved visning av verdiene i et sluttbrukergrensesnitt. I et konkret sluttbrukergrensesnitt kan man godt presentere det positivt (f.eks. regne om «2 % mangel» til «98 % fullstendig» og presentere resultatet positivt).

Fritekst kan brukes som supplerende forklaring

[Spesifikasjon for beskrivelse av kvalitet på datasett](#) tillater bruk av fritekst-kommentarer som supplerende forklaringer til et kvantifiserbart kvalitetsmål. For eksempel, til vårt eksempeldatasett «Bygninger», hvis resultatet på kvalitetsmålet «enheter med manglende verdi for en gitt egenskap» er «2 %», kan man i fritekst-kommentaren spesifisere hvilken egenskap mangelen gjelder, for eksempel: «Dette gjelder egenskap 'byggeår'».

For de aller fleste brukstilfeller antok arbeidsgruppen som utarbeidet disse definisjonene, at det

burde holde med supplerende fritekst-kommentarer. For avanserte kvalitetsbeskrivelser, for eksempel der det er behov for å avgi resultater for hver enkelt egenskap, er det i henhold til DQV mulig å oppgi slike resultater som egne datasett ([dqv:QualityMeasurementDataset](#)). For eksempel en «tabell» som sier «2 % mangel» for egenskap «byggeår», «3 % mangel» for egenskap «bruksareal» og «0 % mangel» for alle de andre egenskapene:

Egenskap	Eier	Byggeår	Kommune	Bruksareal	Bruttoareal
Andel enheter med manglende verdi for en gitt egenskap	0%	2%	0%	3%	0%

Hvilke kvalitetsmål som erfaringsmessig er viktige for brukerne å vite

Ikke alle de predefinerte kvalitetsmålene er relevant å måle i enhver sammenheng. Man står fritt til å velge ut de kvalitetsmålene som er aktuelle for datasettet. I noen tilfeller er det «nøyaktighet» som er viktigst for brukerne av datasettet, i andre tilfeller kan det være «konsistens», eller begge. I mange tilfeller må man også ta høyde for at det er flere ulike typer brukere. Man bør derfor velge de kvalitetsmålene som erfaringsmessig er viktige for mange brukere av det aktuelle datasettet.

Samme prinsipp gjelder for kvalitetsmålets verditype. Flere av kvalitetsmålene kan måles på inntil tre forskjellige måter: boolsk (ja/nei), heltall (antall) og prosent (andel). Dette er for å legge til rette for nivået av innsikt i kvaliteten på datasettet. For eksempel kan det være at man vet at datasettet mangler noen enheter, men ikke hvor mange eller hvor stor andel, bl.a. fordi man i utgangspunktet ikke vet hvor mange og hvilke som skal være med. I slike tilfeller benytter man seg av den boolske verditypen («ja, noen enheter mangler»). I de tilfellene der man vet hvor mange enheter som mangler, benytter man seg av heltall («fire enheter mangler») og/eller prosent («5% av enhetene mangler») hvis man også vet hvor mange enheter som skulle vært med i datasettet.

Når man beskriver et datasett som gjøres tilgjengelig for andre, beskriver man som regel kvalitet ut fra sin egen brukskontekst. Om kvaliteten som er beskrevet fra datatilbyderens ståsted er god (nok) eller ikke for brukerne av datasettet, er avhengig av brukskontekst og bruksformål. «2 % mangel» kan være bra for noen og ikke bra nok for andre.

Når man beskriver kvalitet på datasett som er løpende oppdatert (f.eks. direkte oppslag i et register som løpende ajourholdes), vil det være umulig på forhånd å vite nøyaktig hvilken kvalitet datasettet kommer til å ha. Man vil derfor som regel basere seg på erfaringene man har med datasettet, f.eks. «Statistisk sett er det 2% etterregistrering» som fritekstkommentar til «2% mangel».

Oversikt over predefinerte kvalitetsdimensjoner, kvalitetsdeldimensjoner og kvalitetsmål

Denne veilederen forklarer bruk av predefinerte kvalitetsdimensjoner, kvalitetsdeldimensjoner og kvalitetsmål. Der det er hensiktsmessig kan man referere til andre definisjoner som ligger til grunn for kvalitetsmål.

Tabellen gir en oversikt over de kvalitetsdimensjoner, kvalitetsdeldimensjoner og kvalitetsmål som så langt er predefinerte.

Tabell 1. Oversikt over vokabularet for kvalitetsdimensjoner, kvalitetsdeldimensjoner og kvantifiserbare kvalitetsmål.

Kvalitetsdimensjon	Kvalitetsdeldimensjoner	Kvalitetsmål
Fullstendighet	Underdekning	Manglende enheter
		Antall manglende enheter
		Andel manglende enheter
		Antall enheter med manglende verdi for en gitt egenskap
		Andel enheter med manglende verdi for en gitt egenskap
	Overdekning	Overflødige enheter
		Antall overflødige enheter
		Andel overflødige enheter
	Imputering	Antall enheter med imputert verdi for en gitt egenskap
		Andel enheter med med imputert verdi for en gitt egenskap
Aktualitet	Tidsdifferanse	Samlet tidsdifferanse
Konsistent	Konsistens innad i datasettet	Andel enheter med inkonsistente egenskaper
		Andel enheter med inkonsistens mellom gitte egenskaper
Nøyaktighet	Identifikatorriktighet	Antall enheter med identifikatorfeil
		Andel enheter med identifikatorfeil
	Klassifikasjonsriktighet	Antall feilklassifiserte enheter for en gitt egenskap
		Andel feilklassifiserte enheter for en gitt egenskap

Kvalitetsdimensjonen «fullstendighet»

Fullstendighet handler både om mangel på elementer i datasettet (kvalitetsdeldimensjon «underdekning») og overflødige elementer i datasettet (kvalitetsdeldimensjon «overdekning»). Disse har kvalitetsmål på både enhets- og egenskapsnivå. Videre kan fullstendighet måles ut fra tre ulike verdityper, boolsk (ja/nei) på enhetsnivå, antall og andel på både enhets- og egenskapsnivå. Fritekst-feltet kan brukes til å opplyse om hvilken gitt egenskap som det mangler verdier for (for eksempel «bruksareal»).

Fullstendighet handler også om imputering. Imputering er å fylle inn verdi for en gitt egenskap der verdien mangler eller er ubrukbar (se definisjon i [Definisjoner av kvantifiserbare kvalitetsmål](#)). Dette gjøres for å håndtere manglende egenskaper (tomme celler) i et datasett der disse manglende verdiene skaper problemer for, blant annet, analysen av dataene. Imputerte verdier som kvalitetsmål gir datatilbyderen mulighet til å informere brukerne av datasettet at det er egenskaper i datasettet som ikke er hentet fra virkeligheten. Eksempel: antatt «byggeår» for eldre bygninger.

Kvalitetsdimensjonen «aktualitet»

Det er predefinert ett kvalitetsmål i kvalitetsdimensjonen «aktualitet» – «samlet tidsdifferanse».

For eksempeldatasettet «Bygninger» vet man at det statistisk sett tar ca. 25 dager fra en bygning kontraktsmessig skifter eier til eierskiftet blir meldt inn. Medregnet intern saksbehandlingstid setter datatilbydren «30 dager» som «samlet tidsdifferanse».

Kvalitetsdimensjonen «konsistens»

Kvalitetsdimensjonen konsistens gjelder konsistens innad i ett og samme datasett, og ikke konsistens mellom datasett. Om datasettet er i samsvar med gitte standarder og krav er ikke definert på nytt som et eget kvalitetsmål ettersom dette er dekket av DCAT-AP-NO ([Datasett: i samsvar med](#)). Et eksempel på slik innbyrdes inkonsistens er når bruksareal er større enn bruttoareal for en bygning.

Kvalitetsdimensjonen konsistens kan i mange tilfeller lett forveksles med kvalitetsdimensjonen nøyaktighet. Det som bl.a. skiller nøyaktighet og konsistens er at når det gjelder konsistens *vet man ut fra vurdering av flere egenskaper at det er feil, men ikke hvilken eller hvilke egenskaper som er feil* i datasettet. I eksemplet over er det ikke mulig å avgjøre om det er bruksareal eller bruttoareal (eller begge) som er feil. Når det gjelder nøyaktighet, *vet man hvilken egenskap som er feil* (for eksempel feil identifikator).

Det første kvalitetsmålet i konsistens («Andel enheter med inkonsistente verdier for gitte egenskaper») måles på enhetsnivå. Her måles andel enheter som har en form for inkonsistens knyttet til seg. Det andre kvalitetsmålet («Andel enheter med inkonsistens mellom verdier for gitte egenskaper») går mer i dybden og brukes der man har innsikt i hva inkonsistensen gjelder på egenskapsnivå. Kvalitetsmålene oppgis i prosentandel; fritekstfeltet kan brukes til å forklare for hvilke egenskaper inkonsistensen gjelder.

Kvalitetsdimensjonen «nøyaktighet»

I kvalitetsdimensjonen nøyaktighet måles i hvilken grad dataene korrekt representerer virkeligheten.

Nøyaktighet av en dataverdi er ofte avhengig av type data, og kvalitetsmål for nøyaktighet blir fort svært fag- og sektorspesifikke. De mest generelle nøyaktighetsmålene er derfor plukket ut i denne sammenheng: identifikatorriktighet som går på identifikasjonsnøkler, og klassifikasjonsriktighet som går på bruk av klassifikasjoner og kodeverk.

Definisjoner av kvantifiserbare kvalitetsmål



Redaksjonell merknad 1: Definisjonene som denne veilederen er utarbeidet for å gi veiledning for, er samlet sammen i dette vedlegg. Disse definisjonene vil bli publisert uavhengig av veilederen, og også i maskinlesbart format senere. Den endelige versjonen av veilederen vil derfor ikke inneholde definisjonene, men peke til.



Redaksjonell merknad 2: Gjennom arbeidet med å definere kvalitetsdimensjoner, kvalitetsdeldimensjoner og kvalitetsmål, har man funnet det mer hensiktsmessig å bruke fullstendighet som anbefalt term for kvalitetsdimensjon fullstendighet, enn kompletthet som var foreslått i [Spesifikasjon for beskrivelse av kvalitet på datasett](#). Spesifikasjonen vil bli oppdatert senere (tentativt etter at DCAT-AP-NO er ferdig revidert).



Redaksjonell merknad 3: Fra kommentarrunden på veilederen, fikk vi kommentar på at man ikke bør bruke termen kvalitetsdeldimensjon som arbeidsgruppen i 2017 valgte å bruke. Arbeidsgruppen i denne fasen av arbeidet klarte ikke å finne en bedre term som dekker behovet. Behovet er å ha et nivå mellom kvalitetsdimensjon og kvalitetsmål. Inntil en bedre term kommer på plass, brukes kvalitetsdeldimensjon.

Formål med definisjonene

[Spesifikasjon for beskrivelse av kvalitet på datasett](#) foreslo å etablere en felles oversikt over definisjoner av kvalitetsmål, som et kontrollert vokabular. Dette gjør det mulig å kunne referere til predefinerte kvalitetsmål, istedenfor at alle og enhver definerer disse hver gang, noe som også kan føre til at man definerer ulikt i tillegg til unødvendig dobbeltarbeid. Formålet med foreliggende definisjoner er derfor å skape et felles vokabular og måleverktøy for kvantifiserbar kvalitet på datasett. Dette vil skape en unison beskrivelse og forståelse av datakvalitet på tvers av virksomheter.

I henhold til ovennevnte spesifikasjon, som er basert på [DQV fra W3C](#), skal et gitt kvalitetsmål referere til en kvalitetsdimensjon og kvalitetsdeldimensjon. Det er derfor hensiktsmessig også å predefinere kvalitetsdimensjoner og -deldimensjoner. Dette dokumentet inneholder derfor definisjoner på kvalitetsdimensjoner, tilhørende kvalitetsdeldimensjoner og deres kvalitetsmål.

Se [Spesifikasjon for beskrivelse av kvalitet på datasett](#) for definisjon av begrepene kvalitetsdimensjon, kvalitetsdeldimensjon og kvalitetsmål.

Avgrensninger

For å forstå anvendelsesområdet for de foreliggende definisjonene, er det viktig å være klar over følgende avgrensninger:

- Det er kvantifiserbare kvalitetsmål som defineres.

- Begrunnelse: Dette var utgangspunktet og mandatet til arbeidsgruppen som utarbeidet definisjoner for *kvantifiserbare* kvalitetsmål.
- Kvalitetsmål som er relevante kun i produksjonsfasen er ikke inkludert.
 - Begrunnelse: konteksten gjelder primært for datasett som skal kunne brukes av andre.
 - Eksempler på kvalitetsmål som ikke er inkludert: (manglende) punktlighet osv.
- Kvalitetsmål som allerede er dekket av etablerte vokabularer defineres ikke på nytt.
 - Begrunnelse: Som et generelt prinsipp skal denne type definisjonsarbeid baseres på eksisterende, relevante standarder og spesifikasjoner der slike finnes.
 - Eksempler: «oppdateringsfrekvens» som finnes i [DCAT-AP-NO/DCAT-AP](#); «romlig oppløsning» som finnes i [geoDCAT-AP](#).
- Sektor-/fagspesifikke kvalitetsmål er ikke inkludert.
 - Begrunnelse: Det er viktig å ha fokus på definisjoner av sektor-/faguavhengige kvalitetsmål. Det er første steg mot et felles vokabular og måleverktøy for datakvalitet. I tillegg vil arbeidsgruppen ikke ha hatt mulighet til å bli ferdig med sitt arbeid om alle fagområder skulle vært dekket.
 - Det oppfordres til at sektor-/fagspesifikke kvalitetsmål predefineres (av sektor/fagansvarlige organer) og tilgjengeliggjøres på tilsvarende måte for gjenbruk innenfor samme sektor.
- Det er *iboende* datakvalitet (jf. "inherent data quality" i ISO 25012) som er definert. Det betyr at kvaliteten måles ut fra datasettet i seg selv og ikke eksterne faktorer. Dette gjør at andre kvalitetsaspekter, også nevnt i [Spesifikasjon for beskrivelse av kvalitet på datasett](#), faller utenfor, som for eksempel *tilgjengelighet*, *forståelighet*, *relevans*, *nyttighet*, *tilstrekkelighet*, *samsvar* osv.

Selv om det er avgrensninger ved dette arbeidet, er det på ingen måte ment å utelukke andre måter å beskrive datakvalitet på. Man står fritt til å supplere med andre typer kvalitetsmål i kvalitetsbeskrivelsen.

Der det er hensiktsmessig kan man også referere til andre definisjoner som ligger til grunn for kvalitetsmål som brukes (istedenfor å referere til predefinisjonene omtalt her). Referanse til definisjon bør oppgis.

Definisjoner av noen sentrale begreper

Imputering

Anbefalt term	imputering (bokmål), imputation (engelsk)
Definisjon (bokmål)	det å sette inn verdi for en egenskap hvis den mangler eller er ubruktbar
Kilde til definisjon (bokmål)	basert på « Glossary of Terms on Statistical Data Editing », OECD

Definisjon (engelsk)	entering a value for a specific data item where the response is missing or unusable
Kilde til definisjon (engelsk)	based on " Glossary of Terms on Statistical Data Editing ", OECD

Enhet

Anbefalt term	enhet (bokmål), object (engelsk)
Tillatt term	objekt (bokmål), feature (engelsk)
Definisjon (bokmål)	avbildning av et fenomen i den virkelige verden
Kilde til definisjon (norsk)	basert på ISO 19157
Definisjon (engelsk)	abstraction of real world phenomena
Kilde til definisjon (engelsk)	ISO 19157

Egenskap

Anbefalt term	egenskap (bokmål), property (engelsk)
Tillatt term	variabel (bokmål), attributt (bokmål), kjennemerke (bokmål), variable (engelsk), attribute (engelsk), characteristic (engelsk)
Definisjon (bokmål)	navngitt kjennetegn eller karakteristikk av en enhet
Kilde til definisjon (bokmål)	Geodatakvalitet
Eksempler (bokmål)	inntekt, alder, vekt, yrke, bransje, sykdom
Definisjon (engelsk)	named characteristic of an object
Eksempler (engelsk)	income, age, weight, occupation, industry, disease
Kilde til definisjon (engelsk)	based on Geodatakvalitet

Element

Anbefalt term	element (bokmål), item (engelsk)
Tillatt term	entity (engelsk)
Definisjon (bokmål)	noe som kan beskrives og vurderes separat
Kilde til definisjon (bokmål)	ISO 19157
Merknad (bokmål)	et element er en del av et datasett og kan være enhet eller egenskap
Definisjon (engelsk)	anything that can be described and considered separately
Merknad (engelsk)	an item can be any part of a dataset, such as an object or a property

Kilde definisjon (engelsk)	ISO 19157
----------------------------	-----------

Kvalitetsdimensjon «fullstendighet»

Anbefalt term	fullstendighet (bokmål), completeness (engelsk)
Tillatt term	kompletthet (bokmål)
Definisjon (bokmål)	graden av at datasettet inneholder forventede elementer for en spesifikk brukskontekst
Kilde til definisjon (bokmål)	basert på ISO 25012
Definisjon (engelsk)	the degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use
Kilde til definisjon (engelsk)	ISO 25012

Kvalitetsdeldimensjon «underdekning»

Anbefalt term	underdekning (bokmål), under-coverage (engelsk)
Tillatt term	omission (engelsk)
Definisjon (bokmål)	data som mangler i et datasett
Kilde til definisjon (bokmål)	ISO 19157
Definisjon (engelsk)	data absent from a data set
Kilde til definisjon (engelsk)	ISO 19157

Kvalitetsmål «manglende enheter»

Anbefalt term	manglende enheter (bokmål), missing objects (engelsk)
Definisjon (bokmål)	hvorvidt det mangler enheter i datasettet
Kilde til definisjon (bokmål)	basert på ISO 19157
Definisjon (engelsk)	whether objects are missing in the dataset
Kilde til definisjon (engelsk)	based on ISO 19157
Datatype	boolsk
Eksempel	Nei (datasettet inneholder alle bygninger)

Kvalitetsmål «antall manglende enheter»

Anbefalt term	antall manglende enheter (bokmål), number of missing objects (engelsk)
Definisjon (bokmål)	antall enheter som ikke er i datasettet men som forventes å være med
Kilde til definisjon (bokmål)	basert på ISO 19157
Definisjon (engelsk)	number of objects that are not present in the dataset but are expected to be
Kilde til definisjon (engelsk)	based on ISO 19157
Datatype	heltall
Eksempel	2 (i virkeligheten finnes det 10 bygninger, men datasettet dekker kun 8)

Kvalitetsmål «andel manglende enheter»

Anbefalt term	andel manglende enheter (bokmål), rate of missing objects (engelsk)
Definisjon (bokmål)	antall enheter som mangler i forhold til antall enheter som skulle være med i datasettet
Kilde til definisjon (bokmål)	basert på ISO 19157
Definisjon (engelsk)	number of missing objects in relation to the number of objects that should be present in the dataset
Kilde til definisjon (engelsk)	based on ISO 19157
Datatype	prosent
Eksempel	0.02% (datasettet dekker 0.02% færre bygninger enn det som eksisterer i virkeligheten)

Kvalitetsmål «antall enheter med manglende verdi for en gitt egenskap»

Anbefalt term	antall enheter med manglende verdi for en gitt egenskap (bokmål), number of objects with missing value for a given property (engelsk)
Definisjon (bokmål)	antall enheter i datasettet som mangler verdi for en gitt egenskap
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	number of objects in the data set with missing value for a given property

Kilde til definisjon (engelsk)	own defintion
Datatype	heltall
Eksempel	2 (to bygninger mangler verdi for «bruksareal»)

Kvalitetsmål «andel enheter med manglende verdi for en gitt egenskap»

Anbefalt term	andel enheter med manglende verdi for en gitt egenskap (bokmål), rate of objects with missing value for av given property (engelsk)
Definisjon (bokmål)	antall enheter med manglende verdi for en gitt egenskap i forhold til antall enheter i datasettet
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	number of objects with missing value for a given property in relation to the number of objects in the dataset
Kilde til definisjon (engelsk)	own definition
Datatype	prosent
Eksempel	0.02% (0.02% av verdiene for egenskapen «bruksareal» mangler i datasettet)

Kvalitetsdeldimensjon «overdekning»

Anbefalt term	overdekning (bokmål), over-coverage (engelsk)
Tillatt term	commission (engelsk)
Definisjon (bokmål)	data som er med men som ikke skulle være med i et datasett
Kilde til definisjon (bokmål)	ISO 19157
Definisjon (engelsk)	excess data present in a data set
Kilde til definisjon (engelsk)	ISO 19157

Kvalitetsmål «overflødige enheter»

Anbefalt term	overflødige enheter (bokmål), excess objects (engelsk)
Tillatt term	overflødige objekter (bokmål)
Definisjon (bokmål)	hvorvidt det finnes overflødige enheter i datasettet
Kilde til definisjon (bokmål)	basert på ISO 19157

Definisjon (engelsk)	whether there are objects incorrectly present in the dataset
Kilde til definisjon (engelsk)	based on ISO 19157
Datatype	boolsk
Eksempel	Ja (noen bygninger er overflødige)

Kvalitetsmål «antall overflødige enheter»

Anbefalt term	antall overflødige enheter (bokmål), number of excess objects (engelsk)
Tillatt term	antall overflødige objekter (bokmål)
Definisjon (bokmål)	antall enheter som er i datasettet, men som ikke forventes å være med
Kilde til definisjon (bokmål)	basert på ISO 19175
Definisjon (engelsk)	number of objects within the data set or sample that should not have been present
Kilde til definisjon (engelsk)	based on ISO 19175
Datatype	heltall
Eksempel	3 (i virkeligheten finnes det 15 bygninger, men datasettet dekker 18)

Kvalitetsmål «andel overflødige enheter»

Anbefalt term	andel overflødige enheter (bokmål), rate of excess objects (engelsk)
Tillatt term	andel overflødige objekter (bokmål)
Definisjon (bokmål)	antall overflødige enheter i forhold til antall enheter som skulle være med i datasettet
Kilde til definisjon (bokmål)	basert på ISO 19157
Definisjon (engelsk)	number of excess objects in the data set in relation to the number of objects that should have been present
Kilde til definisjon (engelsk)	based on ISO 19157
Datatype	prosent
Eksempel	0,03% (0,03% av bygningene i datasettet burde ikke være representert)

Kvalitetsdeldimensjon «imputering»

Anbefalt term	imputering (bokmål), imputation (engelsk)
Definisjon (bokmål)	å sette inn verdi for en egenskap hvis den mangler eller er ubrukbar
Kilde til definisjon (bokmål)	basert på « Glossary of Terms on Statistical Data Editing », OECD
Definisjon (engelsk)	entering a value for a specific data item where the value is missing or unusable
Kilde til definisjon (engelsk)	based on " Glossary of Terms on Statistical Data Editing ", OECD

Kvalitetsmål «antall enheter med imputert verdi for en gitt egenskap»

Anbefalt term	antall enheter med imputert verdi for en gitt egenskap (bokmål), number of objects with imputed value for a given property (engelsk)
Tillatt term	antall objekter med imputert verdi for en gitt egenskap (bokmål)
Definisjon (bokmål)	antall enheter i datasettet med imputert verdi for en gitt egenskap
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	number of objects in the data set with imputed value for a given property
Kilde til definisjon (engelsk)	own definition
Datatype	heltall
Eksempel	4 (fire bygninger har fått antatt verdi for «byggeår»)

Kvalitetsmål «andel enheter med imputert verdi for en gitt egenskap»

Anbefalt term	andel enheter med imputert verdi for en gitt egenskap (bokmål), rate of objects with imputed value for a given property (engelsk)
Tillatt term	andel objekter med imputert verdi for en gitt egenskap (bokmål)
Definisjon (bokmål)	antall enheter med imputert verdi for en gitt egenskap i forhold til antall enheter i datasettet
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	number of objects with imputed value for a given property in relation to the number of objects in the dataset
Kilde til definisjon (engelsk)	own definition
Datatype	prosent

Eksempel	0.04% (0.04% av bygningene har fått antatt verdi for «byggeår»)
----------	-----------------------------------------------------------------

Kvalitetsdimensjon «aktualitet»

Anbefalt term	aktualitet (bokmål), currentness (engelsk)
Tillatt term	timeliness (engelsk)
Definisjon (bokmål)	graden av «ferskhet» av datasettet, for en spesifikk brukskontekst
Kilde til definisjon (bokmål)	basert på ISO 25012
Definisjon (engelsk)	the degree to which data has attributes that are of the right age in a specific context of use
Kilde til definisjon (engelsk)	ISO 25012

Kvalitetsdeldimensjon «tidsdifferanse»

Anbefalt term	tidsdifferanse (bokmål), delay (engelsk)
Definisjon (bokmål)	ferskhet av data uttrykt som differansen mellom to tidspunkter
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	age of the dataset described as the difference between two points in time
Kilde til definisjon (engelsk)	own definition

Kvalitetsmål «samlet tidsdifferanse»

Anbefalt term	samlet tidsdifferanse (bokmål), overall time difference (engelsk)
Definisjon (bokmål)	tid mellom når datasettet kan tas i bruk og den hendelsen eller fenomenet datasettet beskriver inntreffer
Kilde til definisjon (bokmål)	Eurostats begrepsdatabase RAMON, European Union, Regulation (EC) No 223/2009
Definisjon (engelsk)	length of time between data availability and the event or phenomenon they describe
Kilde til definisjon (engelsk)	Eurostat RAMON, European Union, Regulation (EC) No 223/2009
Datatype	duration
Merknad (bokmål)	Tillatte måleenheter for duration som er hentet fra xsd, er sekunder, minutter, dager, måneder eller år, dvs. ikke uker.

Eksempel	“24 dager” (det tar i gjennomsnitt 24 dager fra en bygning står ferdig eller er revet til den er innlemmet i eller tatt ut fra datasettet)
----------	--------------------------------------------------------------------------------------------------------------------------------------------

Kvalitetsdimensjon «konsistens»

Anbefalt term	konsistens (bokmål), consistency (engelsk)
Definisjon (bokmål)	graden av at dataene har egenskaper som ikke er motsigende og som samsvarer med andre egenskaper, for en spesifikk brukskontekst
Kilde til definisjon (bokmål)	ISO 25012
Merknad (bokmål)	Konsistens kan gjelde én eller flere sammenlignbare enheter i datasettet.
Definisjon (engelsk)	the degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use
Kilde til definisjon (engelsk)	ISO 25012
Merknad (engelsk)	It can be either or both among data regarding one entity and across similar data for comparable entities.

Kvalitetsdeldimensjon «konsistens innad i datasett»

Anbefalt term	konsistens innad i datasett (bokmål), consistency within the dataset (engelsk)
Definisjon (bokmål)	graden av konsistens mellom egenskapene i datasettet
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	the degree to which there is consistency between the properties in the dataset
Kilde til definisjon (engelsk)	own definition

Kvalitetsmål «andel enheter med inkonsistente egenskaper»

Anbefalt term	andel enheter med inkonsistente egenskaper (bokmål), rate of objects with inconsistent properties (engelsk)
Tillatt term	andel objekter med inkonsistente egenskaper (bokmål)
Definisjon (bokmål)	antall enheter med inkonsistente egenskaper i forhold til antall enheter i datasettet
Kilde til definisjon (bokmål)	egendefinert

Definisjon (engelsk)	number of objects with inconsistent properties in relation to the number of objects in the data set
Kilde til definisjon (engelsk)	own definition
Datatype	prosent
Eksempel	0.03% (av bygningene har inkonsistens innbyrdes mellom noen av egenskapene)

Kvalitetsmål «andel enheter med inkonsistens mellom gitte egenskaper»

Anbefalt term	andel enheter med inkonsistens mellom gitte egenskaper (bokmål), rate of objects with inconsistency between given properties (engelsk)
Tillatt term	andel objekter med inkonsistens mellom gitte egenskaper (bokmål)
Definisjon (bokmål)	antall enheter med inkonsistens mellom gitte egenskaper i forhold til antall enheter i datasettet
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	number of objects with inconsistency between given properties in relation to the number of objects in the data set
Kilde til definisjon (engelsk)	own definition
Datatype	prosent
Eksempel	<ul style="list-style-type: none"> • 0,03% (av bygningene i datasettet står oppført med bruksareal som er høyere enn bruttoareal) • 0,4% (av ansatte i datasettet står oppført med startdato på arbeidsforhold som er før fødselsdato) • 0,2% (av personene i datasettet står oppført som utvandret, men er likevel registrert med norsk bostedsadresse)

Kvalitetsdimensjon «nøyaktighet»

Anbefalt term	nøyaktighet (bokmål), accuracy (engelsk)
Definisjon (bokmål)	graden av at dataene korrekt representerer virkeligheten, for en spesifikk brukskontekst
Kilde til definisjon (bokmål)	basert på ISO 25012
Definisjon (engelsk)	the degree to which data has attributes that correctly represent the true value of the intended attribute of a concept or event in a specific context of use

Kilde til definisjon (engelsk)	ISO 25012
--------------------------------	-----------

Kvalitetsdeldimensjon «identifikatorriktighet»

Anbefalt term	identifikatorriktighet (bokmål), identifier correctness (engelsk)
Definisjon (bokmål)	graden av at enhetene i datasettet har riktige identifikatorer
Kilde til definisjon (bokmål)	basert på BLUE-ETS
Definisjon (engelsk)	the degree to which the objects in the dataset have the correct identifiers
Kilde til definisjon (engelsk)	based on BLUE-ETS

Kvalitetsmål «antall enheter med identifikatorfeil»

Anbefalt term	antall enheter med identifikatorfeil (bokmål), number of objects with incorrect identifiers (engelsk)
Tillatt term	antall objekter med identifikatorfeil (bokmål)
Definisjon (bokmål)	antall enheter i datasettet med feil identifikatorer
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	number of objects in the data set with incorrect identifiers
Kilde til definisjon (engelsk)	own definition
Datatype	heltall
Eksempel	207 (207 personer uten f-nummer/d-nummer men en utenlandsk id som ikke kvalitetssikres)

Kvalitetsmål «andel enheter med identifikatorfeil»

Anbefalt term	andel enheter med identifikatorfeil (bokmål), rate of objects with incorrect identifiers (engelsk)
Tillatt term	andel objekter med identifikatorfeil (bokmål)
Definisjon (bokmål)	antall enheter med feil identifikatorer i forhold til antall enheter i datasettet
Kilde til definisjon (bokmål)	egendefinert
Definisjon (engelsk)	number of objects with incorrect identifiers in relation to the number of objects in the data set

Kilde til definisjon (engelsk)	own definition
Datatype	prosent
Eksempel	0,01% (0,01% av personene i datasettet har gått fra midlertidig tilknytning til permanent oppholdstillatelse og står oppført med d-nummer som identifikator istedenfor f-nummer)

Kvalitetsdeldimensjon «klassifikasjonsriktighet»

Anbefalt term	klassifikasjonsriktighet (bokmål), classification correctness (engelsk)
Definisjon (bokmål)	riktigheten til klassifiseringen av enheter eller deres egenskaper sammenlignet med sanne verdier
Kilde til definisjon (bokmål)	basert på Geodatakvalitet
Definisjon (engelsk)	comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference data)
Kilde til definisjon (engelsk)	ISO 19157

Kvalitetsmål «antall feilklassifiserte enheter for en gitt egenskap»

Anbefalt term	antall feilklassifiserte enheter for en gitt egenskap (bokmål), number of incorrectly classified objects for a given property (engelsk)
Tillatt term	antall feilklassifiserte objekter for en gitt egenskap (bokmål)
Definisjon (bokmål)	antall enheter i datasettet med feil klassifisering for en gitt egenskap
Kilde til definisjon (bokmål)	basert på ISO 19157
Definisjon (engelsk)	number of objects in the dataset that are incorrectly classified for a given property
Kilde til definisjon (engelsk)	based on ISO 19157
Datatype	heltall
Eksempel	97 (97 enheter er oppført med feil næringskode i datasettet)

Kvalitetsmål «andel feilklassifiserte enheter for en gitt egenskap»

Anbefalt term	andel feilklassifiserte enheter for en gitt egenskap (bokmål), rate of incorrectly classified objects for a given property (engelsk)
----------------------	----------------------------------------------------------------------------------------------------------------------------------------------------

Tillatt term	andel feilklassifiserte objekter for en gitt egenskap (bokmål), misclassification rate (engelsk)
Definisjon (bokmål)	antall feilklassifiserte enheter for en gitt egenskap i forhold til antall enheter i datasettet
Kilde til definisjon (bokmål)	basert på ISO 19157
Definisjon (engelsk)	number of objects that are incorrectly classified for a given property in relation to the number of objects in the dataset
Kilde til definisjon (engelsk)	based on ISO 19157
Datatype	prosent
Eksempel	0,4% (0,4% av enhetene har feil kommunenummer)