

DNMI DET NORSKE METEOROLOGISKE INSTITUTT

klima

**DATABASE/MASKIN PROSJEKTET I KLIMAAVD. 1990 - 1991
INFORMASJONSMODELL, FLAGGING OG KONTROLLER
STATUS PR. 30.06.91**

M. Moe, K. Iden, P. O. Kjensli, S. Kristiansen,
S. L. Lystad, B. Nordin, Å. M. Vidal, T. Aasen.
RAPPORT NR. 32/91



DET NORSKE METEOROLOGISKE INSTITUTT
POSTBOKS 43 BLINDERN 0313 OSLO 3

TELEFON: (02) 96 30 00

ISBN

RAPPORT NR.

32/91 KLIMA

DATO

15.07.1991

TITTEL

DATABASE/MASKIN PROSJEKTET I KLIMAAVD. 1990 - 1991
INFORMASJONSMODELL, FLAGGING OG KONTROLLER
STATUS PR. 30.06.91

UTARBEIDET AV

M. Moe, K. Iden, P. O. Kjensli, S. Kristiansen,
S. L. Lystad, B. Nordin, Å. M. Vidal, T. Aasen.

OPPDRAGSGIVER

DNMI

SAMMENDRAG

To datamodeller, en for et informasjonsarkiv og en svært forenklet modell for meteorologiske data er utarbeidet.

Informasjonsarkivet innbefatter et stasjonsarkiv for alle dataområder samt annen informasjon uten direkte tilknytning til en stasjon.

Flere alternative datastrukturer for de meteorologiske dataene skisseres sammen med fordeler og ulemper med hver enkelt. En løsning med et arbeidslager og et hovedlager, hvor lagringsstrukturen i de to lagrene er forskjellig, foreslås.

Behov for flagging av data diskuteres sammen med forslag til hvordan dette kan implementeres for ulike datastrukturer. Omfattende flagging foreslås for data i arbeidslageret, enklere flagging av data i hovedlageret.

UNDERSKRIFT

.....
Margareth Moe

PROSJEKTLEDER

.....
Bjørn Aune

FAGSJEF

INNHALDSFORTEGNELSE:

1.0	INNLEDNING	1
1.1	Hensikt og begreper	1
1.1.1	Informasjonsmodellen.	1
1.1.2	Flagging	2
1.1.3	Begreper	2
2.0	ORGANISERING AV PROSJEKTET.	3
2.1	Delprosjekt 1, oppfølging og status.	3
2.1.1	Aktivitet 1A.	3
2.1.2	Aktivitet 1B.	4
2.2	Delprosjekt 2, oppfølging og status.	4
3.0	DELPROSJEKT 1, ARBEIDSMETODE.	5
3.1	Aktivitet 1A.	5
3.2	AKTIVITET 1B.	6
4.0	RESULTATER	7
4.1	Delprosjekt 1, informasjonsmodeller.	7
4.1.1	Aktivitet 1A: meteorologiske data.	7
4.1.1.1	Alternative datastrukturer.	8
4.1.1.2	Forslag til datastruktur.	12
4.1.2	Aktivitet 1B: Informasjonsarkivet.	17
4.2.	Aktivitet 2: flagging / arbeidslager.	21
5.0	VIDERE ARBEID.	23
6.0	LITTERATUR	24

1.0 INNLEDNING

Denne rapporten vil ta for seg arbeidet i database/maskin prosjektet - Klimaavdelingen høsten 1990 og våren 1991.

Kapittel 2 vil beskrive organiseringen av, og status pr. 30.06.91 for prosjektet, mens kapittel 3 vil beskrive de metoder som er benyttet i de forskjellige delprosjektene. I kapittel 4 presenteres resultatene fra de ulike delprosjektene. Kapittel 5 forsøker å skissere videre arbeid.

1.1 Hensikt og begreper

Med bakgrunn i kjøp av ny database/maskin var hensikten med prosjektet å forberede omlegging av dagens data-strukturer til en relasjonsdatabase og utarbeide et informasjonsarkiv (delprosjekt 1: Informasjonsmodell) samt klargjøre hvordan flagging av data kan utføres på forskjellige nivåer (delprosjekt 2).

1.1.1 Informasjonsmodellen.

En informasjonsmodell skal vise det totale informasjonsbehov. Elementer som inngår her er både datamodellen og databasekonstruksjon.

Hensikten med datamodellen er å få en lettfattelig oversikt over hva databasen skal inneholde/hvilken informasjon avdelingen vil ha behov for i fremtiden, og som når tiden er inne, vil bli implementert i databasen.

Databasekonstruksjon er "oversetting" av datamodellen til en datastruktur tilpasset databasen (i prosjektet; til en relasjonsdatabase).

Hovedbegreper innen datamodellering/databasekonstruksjon er:

- ENTITET: - en entitet er noe som en kan tenke på og kan beskrives (informasjon som naturlig hører sammen).
- ATTRIBUTT: - et attributt er en egenskap / et faktum om en entitet, som kan beskrive en verdi. En entitet består gjerne av flere attributter.
- RELASJON: - en relasjon er forholdet mellom entiteter. Relasjoner kan deles inn i tre typer; en-til-en, en-til-mange og mange-til-mange relasjoner.
- TABELL: - Logisk lagringsstruktur for dataene (en entitet vil utgjøre en eller flere tabeller hvor attributtene fremstår som tabellens/tabellenes kolonner).

1.1.2 Flagging

Flagging av data gjøres for å merke data. Flagget vil gi informasjon om status for dataene, hvilken fase i behandlingen de befinner seg i eller indikere kvalitet. Flagget vil være en kode som lagres sammen med dataene for å gi viktig informasjon under behandlingen av disse og for at man i ettertid skal kunne vurdere dataenes kvalitet.

1.1.3 Begreper

- DATAOMRÅDE:** Dataområde benyttes i denne rapporten for å beskrive data, som i det vesentlige stammer fra en bestemt stasjonstype (Data fra f.eks. Aanderaa stasjoner og Klimastasjoner betegnes som to dataområder). Et dataområde er imidlertid ikke alltid identisk med data fra en stasjonstype et eksempel på dette er Hindcast-arkivet som også blir betegnet som et eget dataområde.
- OMRÅDE-VIS LAGRING:** Data fra samme dataområde lagres sammen. Dette er slik dataene lagres i dag (Aanderaa data lagres for seg i en struktur, Klimadata i en annen).
- PARAMETER:** Målt / observert / beregnet meteorologisk størrelse.
- AKSESS AV DATA:** Bruk av data i databasen (innlegging av data, oppdatering av data, uthenting / spørring mot data).

2.0 ORGANISERING AV PROSJEKTET.

I Klimaavdelingen ble det høsten 1990 definert 2 prosjektgrupper og 4 delprosjekter for å forberede omlegging til en ny database/maskin. De 4 delprosjektene ble senere samlet i to delprosjekter som skulle ta for seg 4 aktiviteter:

Delprosjekt 1 (ref. prosjektplan rev.01) skulle ta for seg:

- Aktivitet 1A: Utforme en informasjonsmodell for "avdelingens meteorologiske data" (datamodell og databasedesign)
- 1B: Utforme en informasjonsmodell for et informasjonsarkiv (datamodell og databasedesign).
- 1C: Spesifisere nødvendige rutiner og utstyr.

Delprosjekt 2 skulle ta for seg:

Klargjøre nivå/spesifisere flagging - behov for arbeidslager (flagging og data-kontroll).

Bjørn Nordin, Tom Aasen, Åse Moen Vidal, Per Ove Kjensli, Margareth Moe, Sofus Lystad, Knut Iden og Stein Kristiansen har arbeidet med disse aktivitetene.

Ved behov har de ulike aktivitetene vært diskutert på felles prosjektmøter for delprosjekt 1 og 2. Det har dessuten vært avholdt ett informasjonsmøte for hele avdelingen.

Aktivitet 1C under delprosjekt 1 er ikke utført (ref. prosjektplan rev.02).

2.1 Delprosjekt 1, oppfølging og status.

Delprosjekt 1 utarbeidet i september 1990 en detaljert prosjektplan (rev. 01) med beskrivelse av oppgaver for de ulike aktivitetene, estimert arbeidsmengde og fremdrifts og bemannings-plan. På grunnlag av status i januar 1990 ble prosjektplanen revidert (rev. 02) vesentlig m.h.t. fremdrift, men også estimert arbeidsmengde og aktivitetenes innhold ble noe endret.

2.1.1 Aktivitet 1A.

Pr. 30.06.91 (uke 26) er det for aktivitet 1A utarbeidet en svært forenklet informasjonsmodell. De ulike entitetene er beskrevet og flere mulige kombinasjoner er skissert. To alternative datastrukturer er nærmere beskrevet. En kartlegging av parametre for de ulike dataområdene pågår. Det gjenstår imidlertid mye arbeid før endelige attributt-lister for de enkelte entitetene kan settes opp.

Videre diskusjoner med "database eksperter" anbefales for å få en mest mulig optimal og hensiktsmessig data-struktur. Diskusjoner / vurderinger og utprøving av ulike strukturer vil være nødvendig før vi kan komme frem til en "endelig" datastruktur her.

Pr. 30.06.91 var 80% (ca. 480 timer) av totalt estimert tid (600 timer) benyttet. Uten at vi konkret har vurdert gjenstående arbeidsmengde for aktivitet 1A, synes det som om denne aktiviteten i prosjektplanen (rev. 02) er klart underestimert.

2.1.2 Aktivitet 1B.

Pr. 30.06.91 er det utarbeidet en informasjonsmodell (datamodell med entiteter og attributter) for et informasjonsarkiv og videre arbeid med en mere detaljert datamodell, som kan danne grunnlaget for databasekonstruksjon, er kommet godt i gang.

Pr. 30.06.91 var 78 % (ca. 290 timer) av totalt estimert tid (370 timer) benyttet.

Pga. øvrige arbeidsoppgaver i avdelingen og ut i fra tilgjengelige ressurser, har det vært foretatt en utskiftning av prosjektdeltakere på aktivitet 1B. Dette har kanskje ført til noe dårligere kontinuitet i arbeidet og dermed høyere tidsforbruk enn hva som ellers kunne vært oppnådd.

2.2 Delprosjekt 2, oppfølging og status.

Det ble her ikke utarbeidet noen detaljert prosjektplan, men delprosjektet har vurdert flagging og utarbeidet et forslag til hvordan dette kan angis for ulike data (kontrollerte / ikke-kontrollerte) ref. kap. 4. Under arbeidet har delprosjektet hatt behov for også å se på de datakontroller som utføres. Det er nå tatt kontakt med EDB-avdelingen for å få oversendt kontroll-algoritmer som benyttes i England og delprosjektet har tatt kontakt med VA for samarbeid om kontroll / lagring / flagging av sanntids data. Det er nødvendig at arbeidet med vurdering av, og samarbeid om, datakontroller fortsetter.

3.0 DELPROSJEKT 1, ARBEIDSMETODE.

Selv om både aktivitet 1A og 1B skulle utarbeide en informasjonsmodell og utføre databasedesign, var utgangspunktet for de to aktivitetene vesentlig forskjellig. Noe ulike arbeidsmetoder ble derfor benyttet.

3.1 Aktivitet 1A.

For aktivitet 1A var det enkelt å sette opp de ulike dataområdene i modellen da dette er data som allerede lagres i dag. P.g.a. at ulike personer har oversikt over forskjellige dataområder, var det imidlertid en betydelig jobb å lage en fullstendig oversikt over alle de ulike parametrene innenfor hvert dataområde. De samme personene i avdelingen ble kontaktet og opplysninger om de ulike parametrene samlet inn og systematisert (fig. 1) (bla. v.h.a. en database) ved en del stikkord:

Kodeform	: angir dataområde (pt. stasjonstypevis)
Beskrivelse	: entydig beskrivelse av parameter
Dataformat	: datatype (lagringsformat) og ant. siffer
Enhet	: fysisk enhet eller kode
Verdiintervall	: lovlige grenser / koder for parameter
Anmerkning	: evt. kommentarer

fig. 1 Eks. på innsamlet informasjon om meteorologiske parametrene.

Kodeform	: Automatstasjon type EDAS
Beskrivelse	: Temperatur, luft, 2m. nivå, pt500 føler
Dataformat	: heltall, 6 siffer
Enhet	: °C
Verdiintervall	: -50.0 til 50.0
Anmerkning	: føler plassert i MI-hytte

Kodeform	: Værstasjon
Beskrivelse	: Temperatur, luft, 2m. nivå, kvikksølv
Dataformat	: heltall, 6 siffer
Enhet	: °C
Verdiintervall	: -40.0 til 50.0
Anmerkning	: føler plassert i MI-hytte

Kodeform	: Værstasjon / Nedbørstasjon / Spesialstasjon
Beskrivelse	: Fordampning, fri vannflate, Thorsrud 2500-evaporimeter
Dataformat	: heltall, 4 siffer
Enhet	: mm/dag
Verdiintervall	: 0.0 til 15.0
Anmerkning	: målesesong fom. 1/5 tom. 1/10

Utformingen av en databasestruktur for de meteorologiske data bød imidlertid på problemer da flere alternative lagringsstrukturer er mulig. Mye tid på felles prosjektmøter (delprosjekt 1 og 2) har gått med til å diskutere hvilken lagringsstruktur som er mest hensiktsmessig. En ekstern konsulent med erfaring fra relasjonsdatabaser ble også hentet inn. Ut i fra hva som kom frem her, er det ikke enkelt å sette opp en endelig databasestruktur for de meteorologiske dataene. Antakelig vil bare testing og overvåking av databasen gi svar på om vi har valgt den mest hensiktsmessige datastrukturen. Flere mulige strukturer blir imidlertid skissert og én av disse anbefalt "som et første forsøk" ved implementasjon (ref. kap. 4.1.1.2).

3.2 AKTIVITET 1B.

Aktivitet 1B skulle utforme en datamodell for data som bare delvis EDB-lagres i dag (informasjon om dataene, stasjonene mm.). Fra Klimaavdelingen ble Liv Fossheim og Gustav Bjørbæk forespurt og modellen ble diskutert og kommentert i felles prosjektmøter (delprosjekt 1 og 2). En del tid har derfor gått med til presentasjon / diskusjon av modellen samt opparbeidelse av en felles forståelse i prosjektgruppen, for de ulike deler av informasjonsarkivet og hvordan disse henger sammen (hvilke relasjoner som eksisterer).

Oversetting av modellen til databasedesign synes relativt enkelt, da dataene og applikasjonene mot disse er av mere "konvensjonell" karakter. Strukturering m.h.p. optimalisering er kanskje ikke like avgjørende som for de meteorologiske dataene.

4.0 RESULTATER

4.1 Delprosjekt 1, informasjonsmodeller.

4.1.1 Aktivitet 1A: meteorologiske data.

De meteorologiske data består her av parametre fra alle dataområder som avdelingen i dag lagrer. Andre (nye) dataområder eller parametre er ikke vurdert, da strukturen i en relasjon database er dynamisk slik at nye parametre og områder enkelt kan komme til senere.

Dataene lagres i dag områdevis. De forskjellige dataområdene vi har i dag er; Edas, Aanderaa, Metar, Klima/Synop, Stråling, Sonde, Plumatic, Strøm, Bølger, E-data, Hindcast-arkiv, Isgrense-arkiv, Nedbør og Ship (data fra enkelte prosjekter / oppdrag er ikke med). Et slikt lagringsmønster er ikke nødvendigvis det vi vil få i den nye databasen.

Mens en oversikt over de ulike dataområdene var enkel å sette opp, er det atskillig vanskeligere å få en systematisk oversikt over alle parametre som de ulike stasjonene observerer (ref. kap. 3.1).

Det som betyr mest for brukerne av dataene er at disse er korrekte samt raskt og enkelt tilgjengelige. Hvor raskt dataene hentes frem fra databasen vil i stor grad være avhengig av datastrukturen og hvordan denne aksesseres. Dette innebærer at en og samme datastruktur ikke gir like god respons (rask tilgjengelighet) fra forskjellige program (applikasjoner) som aksesserer ulike kombinasjoner av dataene.

Data fra de ovenforgitte dataområdene aksesseres på mange måter; fra én eller noen parametre til fulle observasjoner for en eller flere stasjoner.

Et sterkt ønske er at data fra ulike områder skal være enklere å bruke sammen enn hva som er tilfellet i dag.

4.1.1.1 Alternative datastrukturer.

Ut i fra dette er det to typer datastrukturer som vil være aktuelle. En struktur som den vi har i dag (område-vis) vil antakelig gi svært god respons ved uthenting av enkelt-parametre eller hele observasjoner for ett dataområde, men dårligere respons ved aksess av data fra flere dataområder. For denne type bruk av data vil antakelig en "parameter-gruppevis" struktur (der f.eks. temperaturen for flere forskjellige dataområder er lagret sammen) gi best respons.

De meteorologiske dataene aksesseres svært forskjellig avhengig av om de er ferdig / ikke ferdig kontrollert og rettet. Mens ikke-kontrollerte data, som hele tiden kommer inn, bearbeides grundig (kontrolleres, konverteres, rettes), foregår det kun spørring mot (uthenting av) ferdig kontrollerte data. Denne differensierte måten å aksessere data på sammen med ulike krav til flagging av kontrollerte / ikke kontrollerte data (ref. kap. 4.2), synes å underbygge behovet for å ha et arbeidslager (for de ikke-kontrollerte dataene) og et historisk lager (for de kontrollerte dataene). Hvis en slik nivådeling velges, kan en datastruktur (område-vis lagring) implementeres i arbeidslageret mens en annen (parametergruppevis lagring) velges i det historiske lageret.

Flere kombinasjoner av de to datastrukturene kan være aktuelle. Disse er kort beskrevet nedenfor, sammen med en kommentar som peker på sterke og svake sider ved hver av dem, mens en kombinasjon beskrives mere detaljert i kap. 4.1.1.2. Dette er også det alternativet som først anbefales implementert da vi tror det vil gi best respons fra forskjellige typer applikasjoner.

Den utvalgte kombinasjonen av datastrukturer må bare betraktes som et forslag til implementasjon i "første forsøk". Bare virkelig aksess av dataene i en ny datastruktur og overvåking av responstidene vil fortelle oss om ytterligere optimalisering (andre datastrukturer) er nødvendig.

FORDELER VED 2):

AL/HL, en struktur -

Data som hyppig oppdateres ligger for seg selv i arbeidslageret (i egne tabeller, evt. også på egen disk). Dette vil bedre sikkerheten både ved at ikke-kontrollerte data kan ha andre (hyppigere) backuprutiner enn de kontrollerte dataene og ved enklere autorisasjonskontroll for ulike brukere.

- Applikasjoner som både skal aksessere kontrollerte og ikke-kontrollerte data kan gjøre dette "på samme måte" både i arbeidslageret og i hovedlageret.

ULEMPER VED 2):

AL/HL, en struktur -

Vi får ingen gevinst ved "applikasjonstilpassede" tabeller. Enkelte søk/oppdateringer/kontroller kan komme til å gå sent (lang responstid).

FORDELER VED 3):

AL/HL, to strukturer -

Data som hyppig oppdateres ligger for seg selv i arbeidslageret (i egne tabeller, evt. også på egen disk). Dette vil bedre sikkerheten både ved at ikke-kontrollerte data kan ha andre (hyppigere) backuprutiner enn de kontrollerte dataene og ved enklere autorisasjonskontroll for ulike brukere.

- Arbeidslageret og hovedlageret kan designes optimale i forhold til de ulike typer applikasjoner som vesentlig benyttes mot kontrollerte og ikke-kontrollerte data.

ULEMPER VED 3):

AL/HL, to strukturer -

Mere kompleks struktur (flere ulike tabeller).

- Hvis data bare skal lagres ett sted, enten i arbeidslageret eller i hovedlageret, må enkelte applikasjoner, som både aksesserer kontrollerte (i HL) og ikke-kontrollerte (i AL) data, referere forskjellige datastrukturer.

- FORDELER VED a):
Områdevis lagring - Dette er slik dataene "kommer inn" og bearbeides i den første fasen (ikke-kontrollerte data) og vil antakelig gi best respons for applikasjoner som henter frem mange/alle parametre fra stasjoner innenfor samme dataområde.
- ULEMPER VED a):
Områdevis lagring - Strukturen er kanskje dårligere for kontrollerte data hvor enkelte parametre ofte er mere aktuelle enn hele observasjonen og hvor det også er naturlig å benytte ulike dataområder sammen. Områdevis lagring vil da antakelig medføre at data må hentes fra flere tabeller som vil forårsake lengre responstider.
- FORDELER VED b):
Parametergruppevis lagring - For de kontrollerte dataene er det ofte enkelte parametre eller "noen parametre" som hentes ad gangen, men fra ulike dataområder. En parametergruppevis lagring vil kunne tilpasses de mest brukte og tidskritiske applikasjonene for å gi optimal respons.
- ULEMPER VED b):
Parametergruppevis lagring - For de ikke-kontrollerte dataene er ofte hele observasjonen (alle parametre) interessant (f.eks. ved kontroll) og da vil en parametergruppevis-lagring medføre at data må hentes fra flere tabeller som vil forårsake lengre responstid.

4.1.1.2 Forslag til datastruktur.

Ut i fra de viste kombinasjoner av datastrukturer og etter å ha overveid fordeler og ulemper har vi valgt ut en struktur som synes mest hensiktsmessig ut i fra krav om korte responstider for de fleste arbeidsoppgaver. Vi mener alternativ 3 ovenfor i hovedtrekk, best vil tilfredsstillende dette kravet. Noen dataområder tror vi likevel bør lagres områdevis også i hovedlageret da disse er av en noe spesiell natur eller sjelden benyttes sammen med parametre fra andre dataområder.

Pga. at dette valget gir en forholdsvis kompleks struktur (to forskjellige datastrukturer i de to lagrene), tror vi det er en fordel at en applikasjon enten "jobber mot" arbeidslageret (innlegging/kontroll/retting av data) eller det historiske lageret (forespørsler/statistikk mm.). For at de ikke-kontrollerte dataene da skal være tilgjengelige også for applikasjoner som går mot det historiske lageret, må dataene kopieres inn hit samtidig som de legges inn i arbeidslageret. Videre oppdatering vil i dette tilfellet skje like hyppig i hovedlageret som i arbeidslageret. Når det ikke lenger er behov for å beholde data i arbeidslageret vil disse bli endelig overført til hovedlageret og slettet i arbeidslageret.

Forutsatt at man greier å ha oversikt over en slik datastruktur, vil dette antakelig gi best respons for flest applikasjoner (og ved interaktiv aksess mot dataene). Ulempen er at det medfører noe større forbruk av diskplass og at dataene i hovedlageret kanskje ikke er kontrollert, men dette vil være flagget.

I fig. 2 til 5 har vi forsøkt å vise hvordan strukturen for de meteorologiske dataene i en database kan bli. Vi vil presisere at dette bare er et forslag "å starte med", og at vi etter all sannsynlighet må "prøve oss frem" med ulike kombinasjoner av datastrukturer.

I en områdevis datastruktur (fig. 3) vil attributtene være de parametre som hører inn under hvert dataområde. F.eks. vil attributtene for Aanderaa bestå av alle parametre som observeres / beregnes for Aanderaa stasjoner.

I en parametergruppevis struktur (fig. 4) vil attributtene i hver entitet bestå av parametre fra samme gruppe eller parametre som ofte brukes sammen. Entiteten Temperatur vil f.eks. ha ulike temperaturer som attributter (temp. ved observasjonstiden, maks. temp, min. temp osv.) samt fuktighet (brukes ofte sammen med temperaturen).

I en parametergruppevis struktur må det foretas en nøye vurdering av sammenstillingen av attributter for de enkelte entitetene, da denne vil være avgjørende for responsen ved aksess av dataene.

fig. 2 Skjematisk bilde av de ulike dataområder (inf.arkiv er detaljert beskrevet i (4.1.2)).

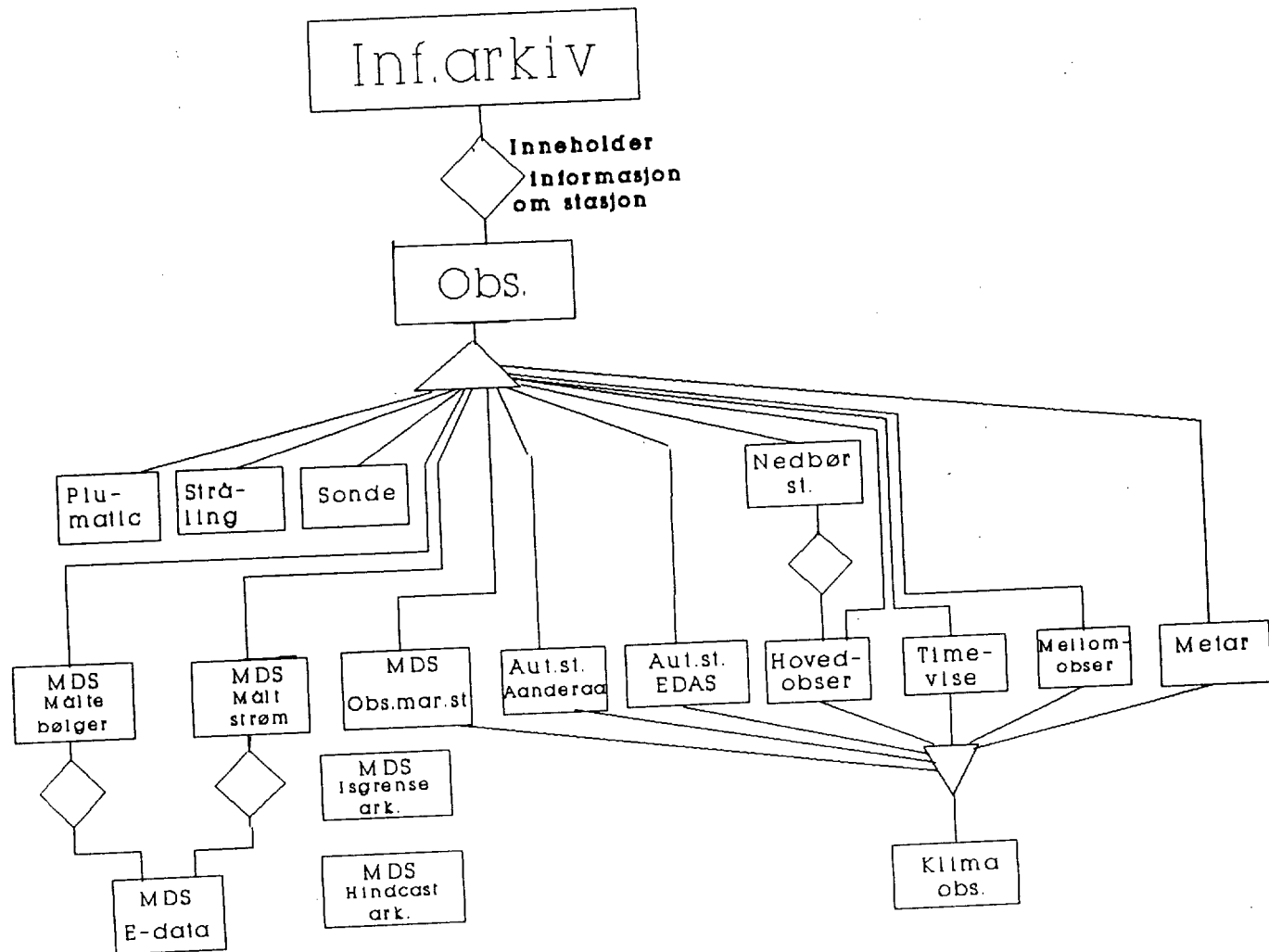


fig. 3 De forskjellige entitetene i en område-vis struktur. Soltimer og Fordampning er innbefattet i entiteten Klima nedenfor.

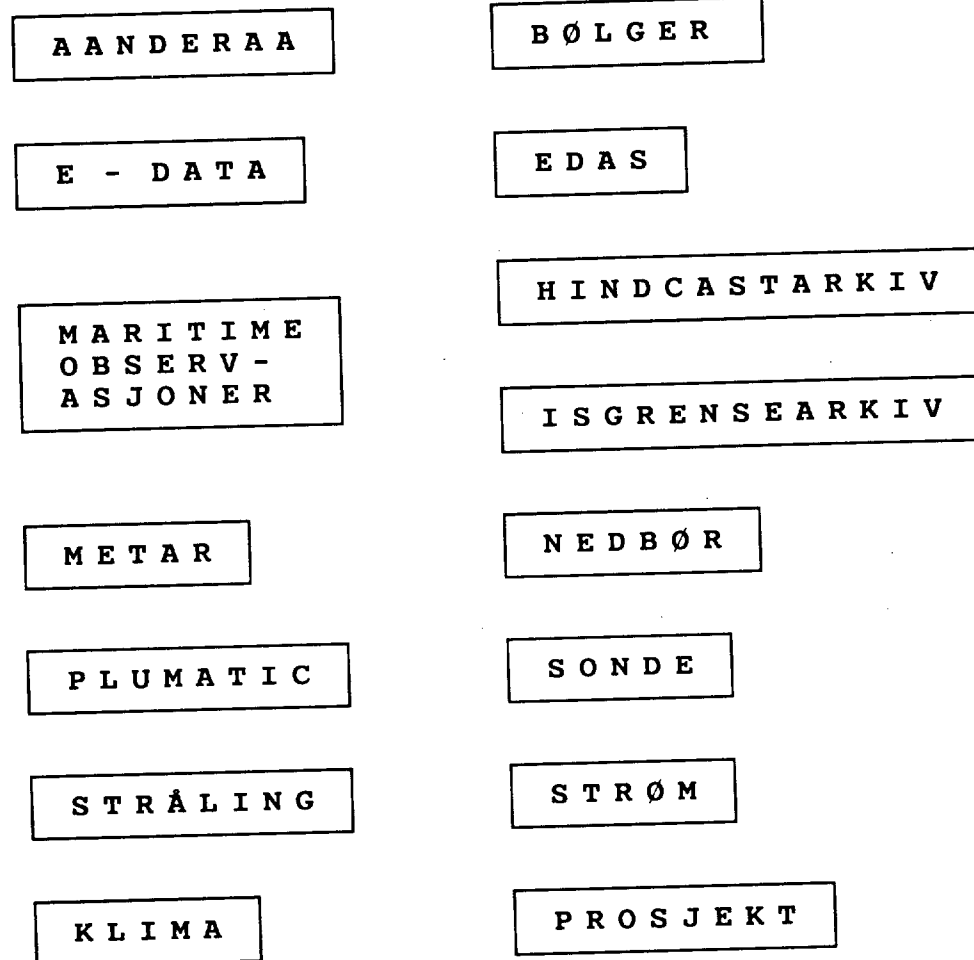


Fig. 4 De forskjellige entitetene i en kombinert områdevis / parametergruppevis struktur.

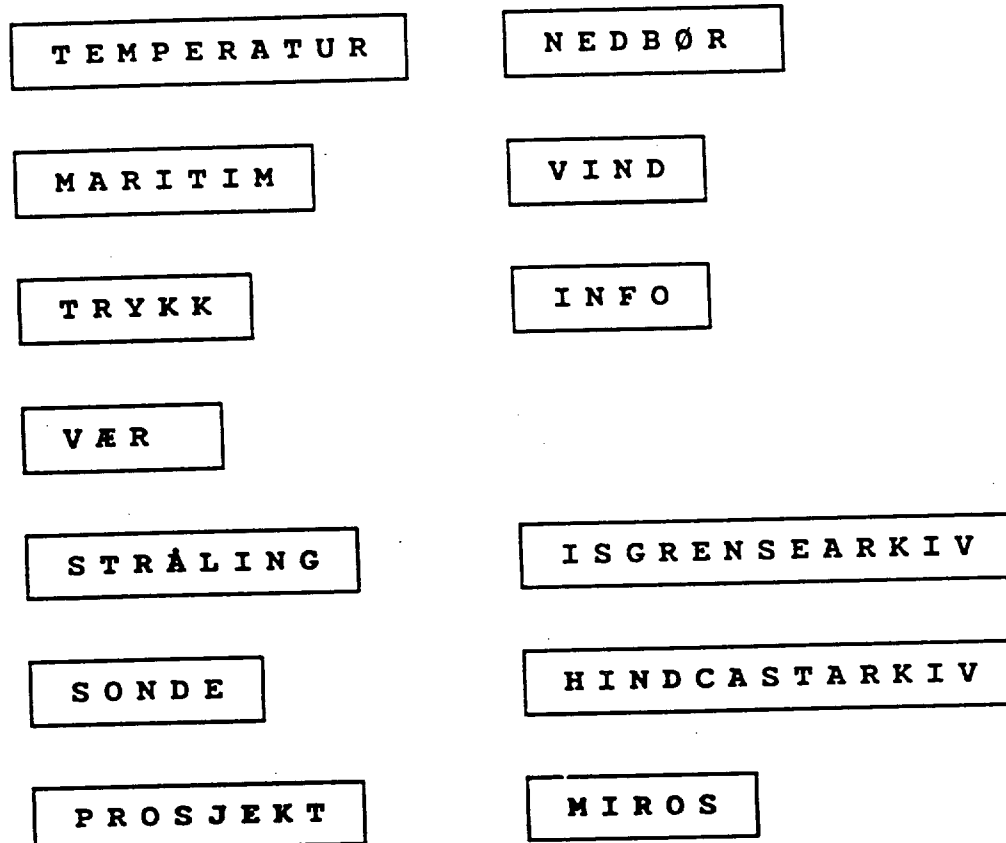
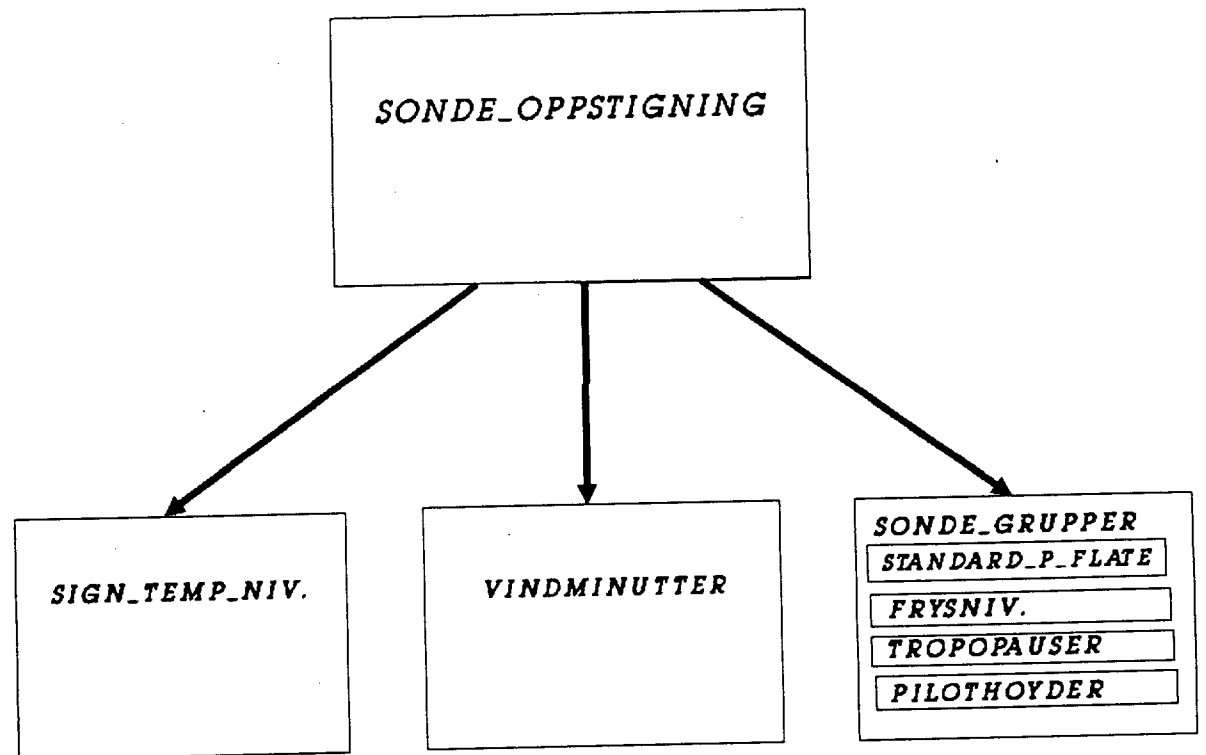


fig. 5 Detaljering av sonde-entiteten fra fig. 4.



4.1.2 Aktivitet 1B: Informasjonsarkivet.

Det har lenge vært ønske om ett felles EDB-basert stasjonsarkiv for alle dataområder. Også annen informasjon uten direkte tilknytning til stasjonene ønskes tilgjengelige i en database. Dette delsystemet har vi valgt å kalle informasjonsarkiv og delprosjekt 1B har utarbeidet en datamodell for dette.

For å få en full oversikt forsøker modellen innbefatte alle forhold omkring de meteorologiske dataene som man vil ha behov for å ha tilgjengelige (f.eks. ved kontroll av observasjonene, bruk av de meteorologiske dataene, administrasjon, inspeksjon mv.).

Foruten å kartlegge de enkelte delene (entitetene) av et informasjonsarkiv, måtte også forholdet mellom dem (relasjonene) klargjøres.

Datamodellen for informasjonsarkivet er vist i fig. 6. Attributtene til entitetene er gjengitt på listeform på de følgende sidene.

Før modellen presenteres vil vi imidlertid gi et eksempel på hvordan denne skal leses:

eks.

ENTITET: Stasjon

ATTRIBUTT: Stasjonsnummer
 Stasjonsnavn
 Synopnummer
 Historisk klimanummer
 Adresse
 Tlf
 Matrikelnummer (gård/bruksnummer)

RELASJON: STASJON har et mange-til-mange forhold (relasjonen) til f.eks. ANSATT. Dette forholdet vil føre til relasjonen; INSPEKSJON, som vil angi når inspektøren var på stasjonen (INSPEKSJON vil sannsynligvis koples mot UTSTYR også - slik at relasjonen foruten å vise når en stasjon har blitt inspisert og av hvilken inspektør, også vil vise hva som ble utført (når det gjelder utstyr) under inspeksjonen.

Relasjonene mellom entitetene er gjengitt som piler i fig. 6.

en-til-en relasjon: _____
 en-til-mange relasjon: _____>
 mange-til-mange relasjon: <_____>

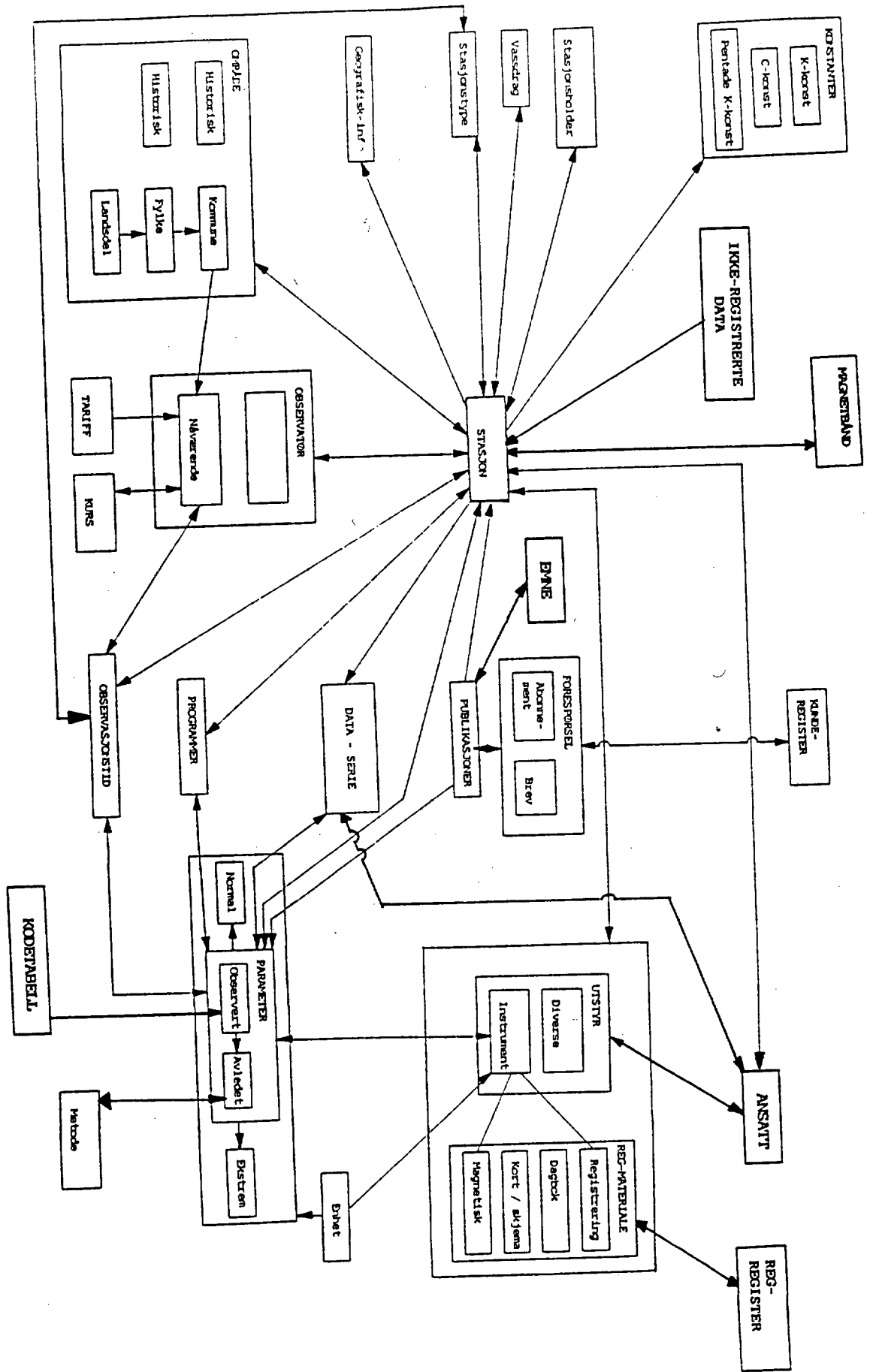


Fig. 6 Datamodel for informasjonarkivet.

ENTITETER med tilhørende ATTRIBUTTER

STASJON

- * nr
- * navn
- * synopnr
- * historisknr
- * adresse
- * godsadresse
- * tlf
- * matrikelnr (gårds/bruksnr)
- * gårdsnavn

STASJONSTYPE

- * type
- * beskrivelse

STASJONSHOLDER

- * navn
- * adresse
- * tlf
- * kontaktperson
- * avdeling/kontor
- * kommentarfelt

GEOGRAFISK-INFO

- geografisk info kan variere over tid uten at stnr forandres

- * stnr
- * bredde
- * lengde
- * pos (posisjon 1, 2 osv - flytting innen et lite område)
- * stasjonshøyde (hs foten av nedbør-stolpe)
- * høyde over havet ved foten av hytta
- * kartserienr topografisk kart
- * kartserienr økonomisk kart
- * UTM koordinater
- * periode
- * kommentarfelt

FYLKE

- konferere m/ Kartverket/Kommunaldep.

KOMMUNE

- konferere m/ Kartverket/Kommunaldep.

LAJDSDEL

- * navn

VASSDRAG

- konferere m/NVE

PARAMETER

- * navn
- * kommentarfelt

INSTRUMENT

- konferere m/ Instrumentavdelingen

METODE

- * navn
- * formel
- * kommentarfelt

DATA-SERIE

- * navn
- * (gjelder) f.o.m dato
- * (gjelder) t.o.m dato
- * forfatter
- * kommentarfelt

PUBLIKASJONER

- * navn
- * dato (første utgivelsesdato)

FORESPØRSEL

- * kundenr
- * sak
- * innkommet dato
- * besvart dato
- * saksbehandler

EMNE

- * emne

KUNDEREGISTER

- * kundenr
- * firmanavn
- * adresse
- * tlf
- * type firma (eks. forsikring)

ANSATT

- ansatt brukes som forfatter/saksbehandler i denne sammenheng

- * navn
- * avdeling/kontor

REG-REGISTER

- * navn
- * kommentarfelt

ENHET

- * benevning
- * standard omregningsfaktor

KODETABELL

- * navn

PROGRAM

- * navn
- * hoved-/subprogram
- * utviklet dato
- * forfatter/ansvarlig

OBSERVATOR

- slektskap mellom observatorer vil i rimlig grad bli registrert

- * personnr
- * navn
- * adresse
- * tlf
- * yrke
- * kategori (kvalitet på observasjonene)
- * skatteprosent
- * diplom (etter 20 år)
- * medalje (etter 40 år)
- * alderstillegg / bonus

TARIF

- brukes under lønnsberegning for observatorer
- * tarif

KURS

- * kursnavn
- * dato (kurset ble holdt første gang)

NORMAL

- * stasjonsnr
- * parameter
- * normalverdi
- * gjelder f.o.m år
- * gjelder t.o.m år
- * kommentarfelt

EKSTREM

- * stnr
- * parameter
- * ekstremverdi
- * (gjelder) tidspunkt/periode
- * (gjelder) f.o.m dato
- * (gjelder) t.o.m dato
- * kommentarfelt

K-KONSTANT

- * stnr
- * konstantverdi
- * (gjelder) obs-tid
- * (gjelder) f.o.m dat
- * (gjelder) t.o.m dat
- * kommentarfelt

C-KONSTANT

- * stnr
- * konstantverdi
- * (gjelder) obs-tid
- * (gjelder) f.o.m dato
- * (gjelder) t.o.m dato
- * kommentarfelt

PENTADE-KONSTANT

- * stnr
- * konstantverdi
- * (gjelder) obs-tid
- * (gjelder) f.o.m dato
- * (gjelder) t.o.m dato
- * kommentarfelt

OBSERVASJONSTID

- * observasjonstidspunkt
- * kommentarfelt

IKKE-REGISTRERTE-DATA

- * stnr
- * parameter
- * f.o.m dato
- * t.o.m dato
- * (hvor befinner data seg) sted/publikasjon
- * kommentarfelt

MAGNETBÅND

- * båndnr
- * eier/avdeling
- * datatype
- * dato (1.gangs innlesing av gjeldende data)
- * kommentarfelt

4.2. Aktivitet 2: flagging / arbeidslager.

Flagging av data har som formål å kvalitets-stemple data. Fordelen er at man får ekstra data-informasjon som det evt. kan tas hensyn til ved behandlingen / bruken av dataene. Ulempen er økt lagerbehov.

I dag flagges data for bare noen av data-områdene. Det er ønskelig at alle data (hver enkelt parameter) i en ny datastruktur skal flagges.

Både eksterne og interne brukere av dataene vil ha nytte av flagg-informasjonen. Foruten å gi informasjon om at en parameter virkelig er målt, kontrollert og korrekt eller f.eks. interpolert vil også flagg-informasjonen kunne gi signal om en eller annen aksjon (f.eks. igangsetting av kontroller, utelatelse fra statistikker m.v.) i forhold til bestemte parametre.

Ved bruk av dataene antas det tilstrekkelig at flagget for hver parameter inneholder to verdier:

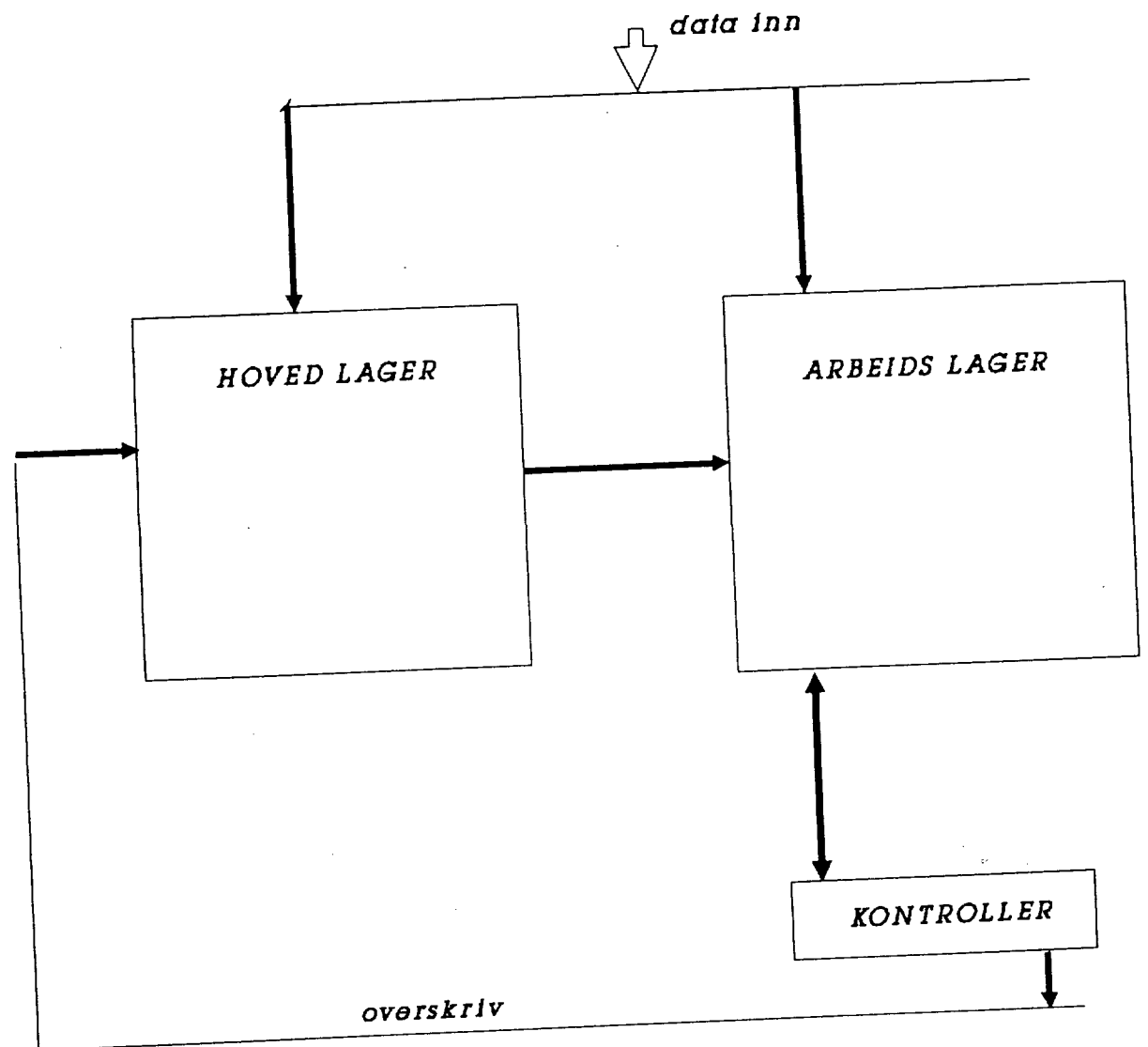
- 1 : Interpolert.
- 2 : Ikke kontrollert.

To verdier vil imidlertid ikke være nok for den løpende datastrøm inn og aktualiserer et arbeidslager hvor mer omfattende flagging kan inngå (flere flagg-verdier og kombinasjoner).

En fullstendig liste for aktuelle verdier her er ikke utarbeidet, men statuser som: Galt - logisk kontroll, Galt - rom-kontroll, Suspekt - logisk kontroll og Suspekt - rom-kontroll m.v., vil være nødvendige.

For flagging av data vil ett datalager kanskje innebære at flaggingen tar forholdsvis mye plass mens fordelene jo vil være lik aksessmetode til alle data (kun én struktur). To datalagre, ett arbeidslager og ett hovedlager, vil ha den fordel at mens en enkel flaggstruktur kan implementeres i hovedlageret, kan en mere omfattende struktur legges inn i arbeidslageret (flere flagg pr. parameter eller egne flagg-tabeller). Ulempen vil være at enkelte applikasjoner får to ulike strukturer å forholde seg til. Delprosjektet foreslår derfor ett arbeidslager og ett hovedlager hvor alle data først legges inn i begge, kontrolleres, rettes og flagges i arbeidslageret før de overleses til hovedlageret med endelig flagg-verdi (fig. 7). Data-strukturen i de to lagrene kan være som skissert i 4.1.1.2.

fig. 7 Forslag til lager-struktur med et arbeidslager og et hovedlager samt data-flyt. Flaggingen av dataene i arbeidslageret vil være mere omfattende enn i hovedlageret.



5.0VIDERE ARBEID.

For aktivitet 1A gjenstår en detaljspesifisering av alle parametre (dette arbeidet er nå påbegynt), samt en gruppering av disse i tabeller.

Videre diskusjoner med andre (personer med databaseerfaring, konsulenter ol.) av datastrukturen vil også være nyttig.

Kontakt med Meteorologisk inst. i Finland vil være svært nyttig da de nylig har tatt i bruk relasjonsdatabase for "sine" meteorologiske data.

Når ny database og maskin foreligger og implementasjonen av de meteorologiske dataene starter, må vi regne med en del "prøving" av ulike datastrukturer som vi kan teste ut mht. ulike applikasjoner og responstider.

En detaljering av datamodellen for informasjonsarkivet er nå i gang. Denne vil danne grunnlaget for videre databasekonstruksjon.

Når det gjelder administrativ informasjon og informasjon om instrumenter, bør disse vurderes nærmere i samarbeid med Administrativ divisjon og Instrumentavdelingen. Det er tatt kontakt med Instrumentavdelingen, men ingen aktiviteter er igangsatt).

Delprosjekt 2 vil se næyere på datakontroll. Hva andre som det er naturlig å sammenligne seg med benytter, vil bli undersøkt (er i gang) samt om det er standarder eller anbefalinger å ta hensyn til. Kontroll av synop-data må ses på i samarbeid med VA (kontakt er opprettet, men ingen aktiviteter igangsatt).

I det videre arbeidet ser vi på et samarbeide med andre avdelinger ved DNMI som svært viktig:

- VA: Innsamling av data og datakontroller / bruk av data.
- IN: Opprettelse og vedlikehold og bruk av informasjon om instrumenter og målinger i informasjonsarkivet.
- Adm. div. Opprettelse og vedlikehold og bruk av administrativ / økonomisk karakter i informasjonsarkivet.
- FOU: Utarbeiding av / opplegg for data-kontrollrutiner.
- EDB: Datainnsamling og ellers på alle EDB tekniske områder.

6.0LITTERATUR

"Prosjektplan database/maskin - delprosjekt 1" rev. 01 av
07.09.90.

"Prosjektplan database/maskin - delprosjekt 1" rev. 02 av
09.01.91.

Aas G., Risnes O. Datamodeller og databasekonstruksjon. SINTEF
- Runit rapport STF14 A84029, 1984.