

Prosjekt igangsatt av NTNFs  
program for kart og geodata

REFERANSESYSTEM FOR  
STEDFESTEDE DATA

S.Høseggen, Norges geologiske undersøkelse  
H.Munthe-Kaas, Norsk institutt for vann-  
forskning

NGU Des. 1983



# Norges geologiske undersøkelse

Leiv Eirikssons vei 39, Postboks 3006, 7001 Trondheim - Tlf. (07) 92 16 11

Oslokontor, Drammensveien 230, Oslo 2, - Tlf. (02) 55 31 65

Rapport nr.	84.056	ISSN 0800-3416	Åpen/Forfattet
Tittel:  Referansesystem for stedfestede data			
Forfatter: S. Høseggen, Norges geol. unders. H. Munthe-Kaas, Norsk institutt for vannforskning		Oppdragsgiver: NTNFs Program for kart og geodata	
Fylke:		Kommune:	
Kartbladnavn (M. 1:250 000)		Kartbladnr. og -navn (M. 1:50 000)	
Forekomstens navn og koordinater:		Sidetall: 187	Pris: kr. 200,-
		Kartbilag:	
Feltarbeid utført:	Rapportdato: 31.12.1983	Prosjektnr.: 21.9.34	Prosjektleder: S. Høseggen
Sammendrag: <p>En referanse er korte og oftest standardiserte opplysninger om data. Rapporten tar for seg referanser til data innen områdene naturmiljø og naturressurser. Disse dataene er vanligvis stedfestet.</p> <p>Rapporten understreker behovet for et referansesystem til nytte for produsentene og brukere av denne typen data. Det er ikke vanskelig å påvise samfunnsøkonomiske besparelser.</p> <p>Rapporten beskriver en modell for oppbyggingen av et referansesystem. Modellen er både sentralisert og distribuert. Dataprodusentene opprettholder sin frihet til å organisere sine data på egne data-maskiner, mens brukerne også tilbys et opplegg med sentral lagring og søking på referansene. Datakommunikasjon og dataoverføring mellom sentral og desentrale maskiner er en viktig del av systemet.</p> <p>Mange av de ideer som framkom under prosjektet blir forsøkt realisert ved NGU som arbeider med en prototyp for et slikt edb-basert referansearkiv innen NGUs fagområde.</p>			
Emneord	EDB	Naturressursdata	
	Datakommunikasjon	Miljødata	

## FORORD

Dette er den avsluttende rapporten utgitt av det NTNf-støttede prosjektet "Referansesystem for stedfestede data".

Prosjektet har bestått av

- praktiske forsøk
- teoretisk modellering

De praktiske forsøkene er blitt gjennomført med hjelp fra Kommunedata Vestlandet (KDV), Teledirektoratet og Norges geologiske undersøkelse, som med stor interesse har fulgt forsøkene. Jeg er takknemlig for at forholdene ble lagt tilrette for å kunne gjennomføre dem.

Nærværende rapport omhandler i hovedsak den teoretiske modellen av referansesystemet. Endel av de praktiske forsøkene er også beskrevet samtidig som resultatene fra disse har påvirket den teoretiske modellen. Her har jeg fått verdifull støtte fra flere enkeltpersoner. Særlig vil jeg takke H. Munthe-Kaas, NIVA. Han har bidratt med flere av rapportens kapitler. Dessuten har han vært en god og kritisk diskusjonspartner i spesifikasjonsarbeidet.

Andre fortjener også takk idet de har bidratt med ideer, brukerkrav og kritikk, spesielt gjelder det T. Vogt SSB, J.M. Larsen MD, B. Follestad NGU og A. Rein NGU.

Norges geologiske undersøkelse, des. 1983  
S. Høseggen

NTNF-PROSJEKT: "REFERANSESYSTEM FOR STEDFESTEDE DATA "

I N N H O L D S F O R T E G N E L S E

-----

0.	Sammendrag	7
1.	Bakgrunn og formål	10
1.1	Kort om NTNF-prosjektet	10
1.2	Forklaring på noen begrep	12
2.	Faglig grunnlag	14
2.1	Gjenbruk av data - situasjonen frem til i dag	14
2.2	Situasjonen pr. i dag	15
2.3	Behovet for gjenbruk av data	16
2.4	Gjenbrukens praktiske forutsetninger	17
2.5	Tiltak for å gjøre gjenbruk praktisk mulig	18
3.	Generelt om referansesystemer for miljødata	20
3.1	Referansesystem-prinsipper	20
3.2	Referansesystemets dekningsområder	22
3.3	Referansesystemets detaljeringsgrad	23
3.4	En dataseriereferanses innhold	24
3.5	Datasikkerhet	25
3.6	Hvem har bruk for referansesystemet	26
4.	Lagring, søking og håndtering av data - i dag og i fremtiden	32
4.1	Situasjonen i dag	33
4.2	Situasjonen i fremtiden	35
4.2.1	Den overordnede organisering av miljø- datahåndteringen	36
4.2.2	Prinsippene for den lokale datalagring	39

5.	Kravspesifikasjon til referansesystem	45
5.1	Generelle krav	45
5.2	Funksjonelle krav	48
5.2.1	Funksjonell hovedstruktur	48
5.2.2	Den nasjonale referansesentralen	51
5.2.3	Det offentlige datanettet	55
5.2.4	Fagpersonell	57
5.2.5	Fagrettede informasjonssentraler	58
5.3	Referansens datainnhold	60
5.3.1	Referansens identifikasjon	60
5.3.2	Stedfesting	62
5.3.3	Tid	63
5.3.4	Emneord	64
5.3.5	Produsent	66
5.3.6	Arkiv	67
5.4	Informasjonsbaser	68
5.4.1	Referansebasen	68
5.4.2	Oppslagsregister	69
5.4.3	Meldingsformidling	72
5.5	Søkeprogram	73
5.5.1	Søking i referansebasen	73
5.5.2	Søking med oppslagsregistre	75
5.5.3	Formidling av søkeord	75
5.5.4	Brukers navn og adresse	76
6.	Bruk av koder	77
6.1	Kodebegrepet	77
6.2	Situasjonen frem til i dag	79
6.3	Kravene til fremtidens kodeopplegg	80
6.4	Tiltak for å få kravene oppfylt	81
6.5	Prinsipper og begreper	83
6.5.1	Nøytralkodeprinsippet	83
6.5.2	Begrepet kodeopplegg	85
6.5.3	Utforming av kodeopplegg	86
6.5.4	Koder og oppslagsregistre	88

7.	Bruk av oppslagsregistre	93
7.1	Behovet for generelle oppslagsregistre på nasjonalt nivå	94
7.2	De enkelte registergruppene	97
8.	Fagrettede informasjonssentraler	99
8.1	Hvordan etablere et edb-basert referansesystem	100
8.1.1	Definisjoner	100
8.1.2	Forutsetninger	101
8.1.3	Dataarkivet skal være absolutt førende	102
8.1.4	Funksjonell oversikt	103
8.1.5	Referansebasen	104
8.1.5.1	Prinsipiell todeling av referansebasen	104
8.1.5.2	Realisering av referansebasen	105
8.1.6	Registrering av referanser	107
8.1.6.1	Fra EDB-arkiv til "korrekte" referanser	107
8.1.6.2	Fra EDB-arkiv til "ukorrekte" referanser	108
8.1.6.3	Referanser fra ikke edb-baserte arkiver	108
8.1.7	Kataloger	110
8.1.8	Overføring av referanser til referansesentralen	110
8.1.9	Søking i referansebasen	112
8.1.10	Søking fra referansesentralen	114
8.1.11	Bruk av oppslagsregistre	116
8.2	Eksempel på løsning	117
8.3	Valg av utstyr. Noen råd	117
8.3.1	Ikke noe datautstyr	119
8.3.2	Dataskjermer	119
8.3.3	Datakommunikasjon	121
8.3.4	Mikrodatamaskiner	125
8.3.5	Mycron	127
8.3.6	Norsk Data	128
8.3.7	Andre datamaskiner	130

9. Videreføring	131
-----------------	-----

Litteraturreferanser	139
----------------------	-----

Vedlegg I	Referansenes innhold
Vedlegg II	Eksempel på referanser
Vedlegg III	Foreløpige informasjoner og oppfatninger om aktuelle generelle oppslagsregistre på nasjonalt nivå
Vedlegg IV	Prosedyre for registrering av biologiske data ved direkte kodegenerering
Vedlegg V	NGUs edb-baserte referansearkiv

## 0. SAMMENDRAG

Denne rapporten beskriver et forslag til et referansesystem for stedfestede data.

En referanse er en samling opplysninger (informasjon) om data. I denne sammenhengen begrenser vi oss til referanser til data innen miljødata- og ressursdatavirksomheten. Ofte forkorter vi det til miljødata.

Referansesystemet er en modell som i en stor grad baserer seg på edb-teknologi, og det skal være et informasjonssystem som i stor grad skal fungere ut fra brukernes premisser. Dette gjelder både faglige og brukermessige forhold.

De faglige premisser blir behandlet i flere kapitler (2,3,4,6,7) hvor det er lagt spesielt vekt på hvordan man skal løse presisjonsproblemet.

Den tekniske modellen med brukerfunksjoner beskrives teoretisk som en kravspesifikasjon (kap. 5). Men deler av modellen (fagrettete informasjonssentral) er også blitt realisert under prosjektet. Den fungerer som et lokalt system ved Norges geologiske undersøkelse (NGU).

Kapittel 1 gir noe av bakgrunnen for at NTN-prosjektet kom igang. Dessuten forklares (defineres) noen viktige begrep som benyttes i rapporten.

I kap. 2 diskuteres nytten av og behovet for gjenbruk av allerede innsamlede miljødata. Gjenbruk av data er en problemstilling som det har vært arbeidet med i mange sammenhenger som det også refereres til. Konklusjonen er at en viktig forutsetning til mer gjenbruk av data er å etablere et referansesystem som detaljert kan fortelle

"hvem har data om hva fra hvor og når".



I kap. 3 går vi videre inn på de generelle prinsipper som må legges til grunn for oppbyggingen av et referansesystem sett med brukerens øyne. Her diskuteres også referansenes dekningsområde og detaljeringsgrad. Behovet for et referansesystem belyses ved eksempler.

Før vi lager en modell av et referansesystem for miljødata, beskrives i kap. 4 hvordan strukturen i miljødata ofte er og hvordan dataene brukes. EDB-systemer som behandler miljødata stilles overfor mange krav som også vil gjelde for referansesystemet. Videre i kapitlet vises til den store spredning i måten miljødata behandles på i dag. Spredningen er også stor mht. bruk av edb-teknikk og spredningen er stor når det gjelder bruk av et presist faglig begrepsapparat.

Et presist og lett håndterbart begrepsapparat er en viktig forutsetning for presis databehandling. Kapitlene 6 og 7 går nærmere inn på dette temaet.

Kap. 5 beskriver et forslag (en modell) av et referansesystem for stedfestede data skrevet i kravspesifikasjons form. Modellen inneholder de to hovedkomponenter i miljødatavirksomheten.

- dataprodusentene
- databrukerne

koplet sammen i et opplegg med en referansesentral og fagpersonell som skal bistå i framhenting av referanser. Med utgangspunkt i modellen beskrives de funksjoner som må være tilstede for at systemet skal fungere.

For å få et best mulig presist begrepsapparat for de opplysninger som lagres av miljødata, er det behov for standardisering i en eller annen form. For å gjøre begrepsapparatet praktisk håndterbart, trenger man koder. I kap. 6 diskuteres nødvendigheten av koding og kravene til et kodeopplegg. Nøytralkodeprinsippet lanseres i kap. 6 som et forslag til kodeopplegg. Dette opplegget løser flere problemer som oppstår i mange andre kodeopplegg. Nøytralkodeprinsippet muliggjør at enkeltinstitusjoner kan

operere med egne lokale kodeopplegg, men knyttes likevel til et nasjonalt entydig opplegg.

Kap. 7 omhandler bruk av oppslagsregistre. Et oppslagsregister er et edb-register som inneholder de faglige kodene med ekstrainformasjon. Det vises til at standardiserte oppslagsregistre vil være nødvendig for å oppnå god håndterbarhet og høy presisjon i bruk av faglige begrep både i et referansearkiv og ved lagring av miljødata generelt. Det presiseres flere steder i rapporten at det er svært nødvendig at oppslagsregistre etableres innen de forskjellige fagområder og tas i bruk.

Kap. 8 omhandler det vi har definert som fagrettede informasjonssentraler. Mer konkret kan dette være en sentral faginstusjon som selv har bygd opp et referansesystem for de data som den selv har lagret. Slike informasjonssentraler er viktige brikker i modellen av referansesystemet. Under arbeidet med dette NTN-prosjektet ble NGUs referansearkiv utviklet. De erfaringer som man fikk fra dette utviklingsarbeidet beskrives i kap. 8. Beskrivelsen vil være nyttig for andre dataprodusenter som skal igang med tilsvarende systemer. Kapitlet har en generell del som omhandler prinsipper og forskjellige problemstillinger. NGUs referansearkiv er beskrevet spesielt (vedlegg V), og det gis også noen konkrete råd i valg av utstyr og systemer.

Kap. 9 oppsummerer de viktigste resultatene fra prosjektet og foreslår en strategi for det videre arbeidet.

## 1. BAKGRUNN OG FORMÅL

### 1.1. Kort om NTNf-prosjektet

Dette NTNf-prosjektet startet ved at NGU i 1981 bestemte seg for å bygge opp et informasjonssystem som kunne gi oversikt over eksisterende geologiske data. Fagpersonell ved NGU har erfart flere ganger at

- penger og tid kunne vært spart om man hadde visst om eksistensen av data,
- NGU kunne gi bedre informasjon utad dersom oversikten var bedre.

Et slikt informasjonssystem skulle selvsagt være edb-basert, men tidlig i utviklingsarbeidet så man visse farer i at man laget et skreddersydd system for NGU uten å samordne seg med andre. For det første har andre institusjoner enn NGU også geologiske data. For det andre er brukere av dataene ikke bare interessert i geologi. Deres forespørsler til NGU er ofte en del av en tverrfaglig problemstilling.

Informasjonssystemet måtte nødvendigvis utvikles ettersom man ikke kunne finne noe tilsvarende ved andre institusjoner. SSBs referansearkiv, [6] var for NGUs behov for lite detaljert.

NGU foreslo derfor et prosjekt overfor kartprogramkomiteen i NTNf. Målet med prosjektet var å komme fram til en modell for et referansesystem, der brukere kunne få tverrfaglig informasjon, samtidig som de enkelte dataprodusentene kunne få rammebetingelsene og retningslinjer for håndtering av sine data.

Forslaget var en naturlig fortsettelse av Miljøverndepartementets tidligere arbeid i Geodatametoden, [22], om et informasjonssystem for ressurs- og forurensningsdata. MD og Kartprogramkomiteen var følgelig interessert i prosjektet og NGU ble forespurt om å utføre prosjektet.

Prosjektets mål ble formulert slik:

"Å spesifisere et informasjonssystem der referanser til stedfestede data som er lagret ved forskjellige institusjoner/etater blir gjort tilgjengelig for brukere. Dette tenkes gjort ved en datanettbasert kommunikasjon mellom produsenter og brukere av stedfestede data.

Hovedformålet er utredning av tekniske løsninger for etablering av institusjonelle referansearkiver.

Disse arkivene skal tjene både institusjonene selv og forvaltningen.

De store datamengder som etter hvert blir samlet inn og lagret krever en fornuftig organisering for effektivt å kunne gjenfinnes. I tillegg vil forvaltningen få et øket behov for på en rask måte å kunne finne frem til forekomsten av relevante data om et spesielt tema. Det er derfor vesentlig at det utvikles standardiserte løsninger for referansearkiver.

De fleste tekniske oppgavene på datanettsiden i tilknytning til et slikt prosjekt er allerede avklart og tildels implementert i eksisterende systemer. Prosjektet bør derfor konsentrere oppmerksomheten om utvikling av en prototyp for dernest å høste erfaring med denne."

Oppgaven ble delt i en teoretisk del og en praktisk del. De praktiske oppgavene var ment å skulle utprøve noen av de ideene som ble lansert i den teoretiske delen.

Følgende praktiske oppgaver ble utført

- (i) Deler av SSBs referansearkiv, [6], ble implementert som dataregistre på KDVs IBM-maskin. Omkring informasjonskilderegisteret og emneregisteret ble det laget programmer som gjorde det mulig for brukere og gjøre oppslag fra

dataskjerm. Funksjonene er tilgjengelig for brukere (kommuner) tilkoblet Kommunedatasentralen Vestlandet.

(ii) NGUs referansearkiv ble utviklet og tilpasset den struktur som er beskrevet i denne rapporten.

(iii) Kommunikasjon mellom IBM og en mikromaskin (ZILOG) ble uttestet basert på asynkron utgang på mikromaskinen og en adapter med IBM2780-protokoll. Teledirektoratets datautstyr ble her benyttet.

Nærværende rapport inneholder den teoretiske modellen som forfatterne mener systemet bør ta utgangspunkt i. I tillegg er noen av de praktiske forsøkene beskrevet for å støtte opp under den teoretiske beskrivelsen.

## 1.2. Forklaring av noen begrep

Stedfestede data er data som er knyttet til geografisk sted. Måten stedfestingen er gjort på, kan være forskjellig.

### Miljødata

I denne rapporten benyttes begrepet miljødata som en forkortelse for stedfestede data fra miljø- og naturressursvirksomhetene i Norge.

### Dataserie

Dataserie er en samling med data som det er naturlig å se som en enhet og kan refereres til ved hjelp av én referanse.

Dataseriens tilstand kan være en av følgende

- primærdata
- bearbeidede data
- dokumentasjon om data

### Referanse

Samling opplysninger (informasjon) om data.

### Referansearkiv

SSBs rapport "Referansearkiv for naturressurs- og forurensningsdata", [6], blir i mange sammenhenger forkortet som "referansearkivet". Flere dataprodusenter lagrer referanser på edb-registre, som de kaller et referansearkiv. I denne rapporten er referansearkiv blitt brukt i den siste betydningen.

### Database

Data lagret på edb-registre er en database.

### Referansebase

Referanser lagret på edb-registre har vi valgt å kalle en referansebase. Grunnen er at en referanse er informasjon og ikke selve dataene. Derfor skapes det lett forvirring å si at referansene lagres i en database.

## 2. FAGLIG GRUNNLAG

### 2.1 Gjenbruk av data - situasjonen frem til i dag

Det fins i Norge et hundretalls større institusjoner og etater som samler inn miljødata. Om også de mindre regnes med, blir det mange hundre. De fordeler seg over mange fagområder, har forskjellige oppgaver i samfunnet og forskjellige tilknytningsformer til samfunnsmaskineriet. Men de typer av miljødata som ulike institusjoner har bruk for, er ofte de samme.

Den totale mengde av data som samles inn, er stor og mangfoldig, og den øker fra år til år både i omfang og bredde.

Før i tiden var det ofte slik at miljødata som ble samlet inn for ett konkret formål, ble brukt til dette ene formål - og deretter lagt på lager og aldri tatt frem igjen. I løpet av de siste år er det etter hvert blitt noe mer vanlig å ta frem igjen eldre data fra egen institusjon og å be om data fra andre institusjoner. Men fremdeles foregår gjenbruk av data langt sjeldnere enn hva den nytten man ville ha av slik gjenbruk, skulle tilsi.

Det er flere grunner til at data gjenbrukes og utveksles i langt mindre grad enn nytten skulle tilsi. Blant disse er følgende antagelig de viktigste:

- Den enkelte dataserie eksistens er ikke alment kunngjort. Det vil si at den enkelte institusjon ikke kan vite hva de andre har. Til og med har det vist seg at mange heller ikke vet hva de selv har i sitt lager.
- En del eldre bevaringsverdige data på lager er etter hvert blitt kastet eller gått tapt på annen måte.
- En del data er lite brukbare - fordi nødvendige opplysninger om sted, analysemetoder etc. mangler.

Konsekvensene for samfunnet av denne situasjon er store:

- Data som egentlig fins, blir ofte ikke utnyttet i samfunnets beslutningsprosesser og forskningsvirksomhet.
- Dataserier som egentlig fins, blir av og til samlet inn på nytt igjen.
- Ny-innsamlinger av dataserier blir ofte planlagt uten at tilpasning til alle aktuelle eksisterende dataserier foretas - og uten at erfaringer fra tidligere dataserier utnyttes.

## 2.2 Situasjonen pr. i dag

Selv om denne situasjon har vart lenge og dens uheldige konsekvenser stadig er blitt større, har den først i løpet av de siste få år fått noen større oppmerksomhet. Temaet ble første gang tatt opp i 1972 av et nordisk kontaktorgan for vannforskningsinstitusjoner opprettet i regi av Nordforsk. I 1973 arrangerte denne gruppen et Nordisk symposium om vanndata-samordning [20]. I midten av 70-årene ble det nordiske samarbeidsansvaret for dette temaet flyttet over fra Nordforsk til Nordisk Ministerråd. Ministerrådet utvidet rammen til å omfatte alle typer av miljødata og har siden fulgt opp med en rekke tema-avgrensede prosjekter, ref. [1], [2], [3], [4], og en sammenfattende utredning i 1982, ref. [5]. Den sistnevnte utredningen (FOSAMINO-rapporten) er siden fulgt opp med et nytt "Nordisk kontaktorgan for miljødata" som tildels driver egne prosjekter og tildels tilråder og samordner nasjonale prosjekter.

I Norge har Miljøverndepartementet vært en aktiv deltager i Ministerrådets ovennevnte virksomhet. Dessuten har departementet gjort to større nasjonale fremstøt på miljødatasiden:

- o Opprettelse av et nasjonalt referansearkiv i 1978 med Statistisk Sentralbyrå som hovedaktør [6].



- o Etablering av et nasjonalt standardformat for utveksling av miljødata (SOSI-formatet) i 1983 med Norsk Regnesentral som hovedaktør [7].

### 2.3 Behovet for gjenbruk av data

I avsnitt 2.1 ovenfor er det hevdet at datagjenbruk er nyttig. Denne påstand kan riktignok ikke underbygges med noen systematisk gjennomført undersøkelse - og den kan følgelig heller ikke kvantifiseres. Når den ikke er forsøkt underbygget, er det fordi den i seg selv er selvvinnlysene. I den prosjektbestilling som fins bak nærværende rapport, er det forutsatt at påstanden gjelder. Også Nordisk Ministerråd finner, på basis av FOSAMINO-rapportens uttalelser om dette, at påstanden holder.

I FOSAMINO-rapporten er behovet for gjenbruk av data for sikkerhets skyld formulert på to måter

- o Det er viktig at alle miljødata som eksisterer, blir tatt vare på og gjort tilgjengelige på en slik måte at de kan utnyttes for kjente og ukjente formål i fremtiden.

og

- o Man må sikre at dagens og fremtidens miljøforvaltere og -forskere får det best mulige grunnlag for å utnytte dagens og fortidens miljødata.

Indirekte påpeker FOSAMINO-rapporten videre at gjenbruksdata kan fås fra to typer av kilder -

- o eldre data fra institusjonens egne arkiver
- o data fra andre institusjoner

- og at det å sikre disse to kildetyper tilgjengelighet tildels krever separate tiltak.

## 2.4 Gjenbrukens praktiske forutsetninger

To typer av forutsetninger må være oppfylt for at gjenbruk av data skal kunne foregå på en hensiktsmessig måte - tekniske og organisatoriske.

- Pr. i dag er de tekniske muligheter for gjenbruk av egne og andres data blitt så gode at man fra den siden neppe vil møte noen vesentlige problemer. Disse mulighetene omtales ikke nærmere i dette kapitlet.
- På det organisatoriske plan mangler det en rekke forutsetninger. Disse forutsetninger er blitt grundig vurdert i FOSAMINO-rapporten.

De organisatoriske forutsetninger er i FOSAMINO-rapporten presentert i form av 7 krav til forvaltningens fremtidige miljødata-administrasjon.

- A Dataene må være entydig tolkbare og lett håndterbare.
- B Dataenes eksistens må være alment kunngjort.
- C Dataene må være lagret slik at de er lett tilgjengelige og lett søkbare ut fra mange synsvinkler.
- D Dataene må være lett overførbare.
- E Dataene må være beskyttet mot tap, skade og uautorisert bruk.
- F Regler og/eller retningslinjer må finnes for vurdering av datas bevaringsverdighet og almene tilgjengelighet.
- G Ikke bare data som produseres i dag, men også eldre data, må tas vare på.

Utover dette stiller rapporten opp et krav som skal bidra til å sikre at man, ved planlegging av dagens dataproduksjon, i tillegg til å sikte mot dagens behov også tar rimelig hensyn til fremtidens antatte behov:

- H Retningslinjer bør finnes m.h.t. det å sikre fremtidens behov for miljødata og miljøprøver fra i dag.

Videre stiller den et krav som har sammenheng med det forhold at data produsert på basis av standardiserte eller harmoniserte metoder vil ha større anvendelighet enn andre data for fremtidige problemstillinger:

- I Retningslinjer bør finnes for anvendelse av standardiserte eller harmoniserte metoder (for observasjon, prøvetagning og analyse) når data skal produseres.

Til sist stiller den et krav som, til forskjell fra de andre ni, har som hovedsiktepunkt å spare ressurser.

- J Et opplegg bør finnes som gjør det enklere å utveksle og å samordne utviklingen av EDB-metoder for bearbeidelse av miljødata.

Med disse 10 kravene som basis foreslo så FOSAMINO-prosjektet 9 forskjellige tiltak som skulle gjennomføres. Av disse burde 3 helst gjennomføres i Nordisk regi og de øvrige 6 i nasjonal regi.

## 2.5 Tiltak for å gjøre gjenbruk praktisk mulig

De 9 foreslåtte tiltakene var disse:

- a 1 Videreføring av arbeidet med "Håndbok for karakterisering av miljødata", [3].
- a 2 Servicesenter for oppslagsregistre
- b 1 Referansesystem for dataserier
- c 1 Samordningsvirksomhet m.h.t. ! a) - lagring av data  
metoder for ! b) - utveksling av data
- c 2 Samordningssenter for EDB-programmer

- d           Retningslinjer for           ! 1) - datas bevarings-  
  !       verdighet og  
  !       tilgjengelighet  
  !  
  ! 2) - sikring av frem-  
  !       tidens databehov  
  !  
  ! 3) - anvendelse av stan-  
  !       dardiserte og har-  
  !       moniserte metoder

e           Aksjon for sikring av eldre data

Pr. i dag (oktober 1983) er følgende skjedd m.h.t. disse forslagene

a 1 og a 2 o   tas hånd om av Nordisk Ministerråd med prosjekt-  
                  start i 1984.

b 1           o   er tema for nærværende rapport

c 1 b         o   er i gang i Miljøverndepartementets regi med  
              Norsk Regnesentral som hovedmedarbeider (SOSI-  
              prosjektet).

Med de øvrige er intet skjedd så langt.

### 3. GENERELT OM REFERANSESYSTEMER FOR MILJØDATA

#### 3.1 Referansesystem-prinsipper

Forskjellige prinsipielle opplegg kan tenkes for organisering av miljødatalagringen i et land. Som f.eks.

- 1 En sentral, nasjonal databank som inneholder alle landets miljødata.
- 2 Et sett av regionale og/eller bransjevisse databanker.  
(Eksempel: Geologisk databank for Norge; oceanografisk databank for Skagerrak).
- 3 Den regel at hver institusjon tar vare på alle data som den selv produserer.

Også andre prinsipper kan tenkes - og et stort antall kombinasjoner av 2 eller flere av disse prinsipper. I Norge har vi i dag en blanding av prinsippene 2 og 3 med prinsipp 3 som det dominerende.

Etter forfatterens mening er dette en god situasjon som det ikke er noen grunn til å forandre på.

For at kravene B, C og D i avsnitt 2.4 ovenfor (om at dataenes eksistens skal være alment kunngjort og at dataene skal være lett tilgjengelige, søkbare og overførbare), må den lagringsteknikk som brukes, ha visse egenskaper innebygget. En av disse egenskapene er av spesiell betydning for referanseaspektet og skal omtales her:

For at data i praksis skal være lettvint søkbare (og også av andre grunner) - må dataene være organisert i dataserier. Organiseringen må innebære at ethvert måledata-element skal tilhøre en eller annen dataserie. (Dataseriene selv kan enten alle være likestilte m.h.t. hierarkisk nivå i systemet

- eller de kan organiseres i en egen hierarkisk sidegren i databasen).

Ut i fra den situasjon at alle data er organisert i et dataseriemønster, kan man så tenke seg at det for hver dataserie lages en referanse som kortfattet forteller hva dataserien inneholder. (Om man har et godt gjennomtenkt lagringssystem for dataene, vil datareferansen kunne produseres automatisk av EDB-systemet.) Datareferansen selv må være systematisk oppbygget, slik at den uten omforming kan inngå i et EDB-basert referansearkiv.

I det krav som her skal oppfylles (dataenes eksistens må være "alment" og "allsidig" kunngjort) ligger følgende

- \* Alle mennesker skal ha adgang til informasjon om hvilke dataserier som fins.
- \* Det skal være mulighet til lettvint å søke seg frem i referansesenterets informasjonsmasse til en hvilken som helst dataserie-referanse. Eksempler på dette kan være:
  - Hvilke institusjoner har kjemiske data fra sedimentene i Kragerøfjorden fra før 1980.
  - Referanser ønskes til alle dataserier som inneholder data fra innsjøer i Sør-Norge beliggende i høydebeltet 800-1200 m over havflaten som gjelder innhold av tungmetaller i vann, sediment eller fiskeorganer).

Det er viktig at utvalget av mulige typer av søkenøkler er allsidig og godt gjennomtenkt.

For å oppfylle dette krav, foreslo Nordforsk symposiet, [20], følgende:

- A Det opprettes en referansesentral i hvert land som dekker alle typer av miljødata. Disse sentrene skal basere sin drift på databaseteknikk.

- B Alle institusjoner innen landet som har ansvaret for et miljødatalager, blir registrert av referansesenteret som "miljødatalager".
- C Alle miljødatalagre skal regelmessig (f.eks. 1 gang pr. år) sende inn referanser for alle nye dataserier som er kommet inn siden sist - og suppleringer til tidligere referanser for dataserier som er utvidet siden sist.
- D Referansesenteret registrerer disse referanser i sitt databasesystem.
- E Alle som ønsker informasjon av typen "Hvem har data om hva fra hvor og når" eller lignende typer kan få svar ved henvendelse til referansesenteret.

Ofte vil det være slik at samme dataserie finnes i mer enn ett datalager. (Fordi en institusjon som har fått en serie fra en produsent, kanskje også tar vare på den). I den forbindelse er det viktig å sette den regel at det skal være én - og bare én - institusjon som har det offisielle lagringsansvar for serien. (Som regel vil det være produsentinstitusjonen, men det behøver ikke å være den). En slik regel trengs av følgende grunn: Av og til oppdages det datafeil som må rettes. Det er viktig at disse feil blir rettet i det offisielle lageret. Hvis dataserien fantes i flere offisielle lagre, vil man trenge nokså strenge og tunge prosedyrer for å sikre at alle rettelser var kommet med i alle lagre.

### 3.2 Referansesystemets dekningsområde

Miljødata er et vidt begrep. Den mulighet finnes at ett felles referansearkiv skulle dekke det hele. Eller den motsatte ytterlighet at et referansearkiv blir etablert for hver miljøbransje.

Efter forfatterens mening vil det være mest hensiktsmessig at referansesystemet blir oppdelt i et fåtall store og brede bransje-arkiver, men slik at det også etableres en sterk overordnet samordning og styring av hele systemet, - en referansesentral.

Med hensyn til en slik oppdeling, har forfatterne ikke vurdert nærmere hvordan den bør være. Som et utgangspunkt for diskusjonen foreslås følgende deler:

- Måledata fra naturen (vann, geologi, biologi, botanikk osv.)
- Miljøprøver
- Offentlige arealplaner
- Forurensningskilder og utslippsdata
- Data om vern og kulturminner

### 3.3 Referansesystemets detaljeringsgrad

I høringsrunden etter 1. utgave av det eksisterende norske referansesystemet [6] kom det frem ulike meninger om hvilket detaljeringsnivå et slikt system burde legges på.

Dagens norske system har en relativt lav detaljeringsgrad. Det gir først og fremst informasjon om hvilke datakategorier den enkelte institusjon dekker - og omvendt hvilke institusjoner som dekker de enkelte datakategorier. På dette nivå gir systemet god informasjon. Egentlig er innholdet tildels noe mer detaljert enn dette. Fordi systemet ikke er edb-basert, men bare manuelt basert, får man imidlertid liten nytte av disse detaljene.

Etter forfatterens mening bør detaljeringsgraden være meget høyere - slik at man som nevnt ovenfor (avsnitt 3.1) f.eks. kan spørre om "hvilke institusjoner som har data fra innsjøer i Sør-Norge beliggende i høydebeltet 800-1200 m over havet som gjelder innhold av tungmetaller i vann, sediment eller fiskeorganer". Videre bør systemet benytte database-teknologi. Og



det bør være hierarkisk organisert på en slik måte at en informasjon på detaljnivå også inkluderes når man spør på et høyere nivå. (Eksempel: Hvis man under søkeordene "småvilt" og "Eggedal kommune" finner en dataserie om hare i Eggedal, må man også finne den også om man spør om "småvilt" i "Buskerud fylke").

Som grunnlag for diskusjonen foreslås at det norske referansesystemet skal ha dette detaljnivå:

Stedfesting: "Statistikk-områder" (se nedenfor) eller  
XY-koordinat  
Tid : Måned, År  
Medium : Inntil videre hovedgrupper som f.eks.  
"løsmasser" og "ferskvann".  
Parameter : Inntil videre hovedgrupper som f.eks.  
"fysisk-kjemiske" og "botaniske".  
Opprinnelse: Institusjon - (og også prosjekt når det fins),  
samt person

(Et "statistikk-område" er den minste gruppering av grunnkretser i et område som kan lages på en slik måte at den faller sammen med et nedbørfelt eller delnedbørfelt. Med en slik gruppering oppnår man den fordel å kunne aggregere både langs linjen "kommuner - fylke", linjen "delnedbørfelt - nedbørfelt", og linjen koordinat-kartblad).

### 3.4 En referanses innhold

I en dataseriereferanse må det alltid finnes et minimumssett av opplysninger. Hvilke opplysninger dette gjelder, varierer til en viss grad av hvilket fagområde den hører til. Følgende informasjon må alltid inngå:

Geografisk avgrensning  
Tidspunkt eller tidsavgrensning

Medium

Parametergruppe

Andre sentrale emneord

Produsent-institusjon

Lagringsansvarlig institusjon

Lagringsansvarlig institusjons fagavdeling/kontaktperson

Systemet bør lages slik at det senere, når endel erfaring er innvunnet, kan føyes til andre informasjoner i tillegg til disse.

### 3.5 Datasikkerhet

Mot ideen om et nasjonalt referansesystem for miljødata er det synspunkt av og til kommet frem - fra naturvitenskapelige kolleger - at de ikke våger å gi fra seg sine data. Deres begrunnelse var denne:

- o Bare jeg (produsenten) kjenner alle omstendigheter og forbehold omkring tilblivelsen. Andre, som ikke kjenner disse omstendigheter og forbehold, vil lett kunne bruke dataene galt.

Til dette syn vil forfatterne anføre følgende:

- o Det å bekjentgjøre en dataserie's eksistens, gjennom å sende inn en referanse, betyr ikke at man må gi dataene fra seg.
- o Hvis ikke spesielle grunner tilsier noe annet, bør man dog helst være villig til å gi dem fra seg. Fordi de som regel skal brukes til noe som gagnar samfunnet. Men man kan samtidig sikre seg ganske godt mot feilaktig bruk ved f.eks. å:
  - o sende med gode kringinformasjoner [3]
  - o orientere personlig om dataene
  - o diskutere rekvirentens hensikt med og bruksplan for

dataene.

- o be om å få se resultatet av databruken før dette publiseres eller anvendes for beslutningsformål.

Det fins en del nødvendige og aktverdige grunner for å nekte utlevering av dataserier. Det gjelder f.eks. visse data for militært bruk og visse data innsamlet i industrisammenheng. Dessuten bør f.eks. en doktorgradskandidat eller hovedfagsstudent kunne holde igjen sine data inntil avhandlingen er ferdig.

Nesten i alle slike tilfeller vil det imidlertid være forsvarlig å bekjentgjøre seriens eksistens gjennom å sende inn selve referansen - og i den gjøre oppmerksom på de foreliggende sperringer. Også referanser i seg selv kan være nyttig i visse sammenhenger. Dessuten behøver en sperring ikke gjelde for alle brukerkategorier.

Videre vil de forhold som betinger tilbakeholdelse som regel forsvinne etter en tid. Når det skjer, bør det kunngjøres via referansen.

### 3.6 Hva referansesystemet kan brukes til

De mest vanlige spørsmål som referansesystemet kommer til å få, vil antagelig bli enkle spørsmål av typen:

Hvem har data om hva fra hvor og når.

Spørsmåleksemplet i avsnitt 3.5 ovenfor om data fra Kragerøfjorden er av denne typen.

Det andre spørsmåleksemplet i avsnitt 3.5, vedrørende et nærmere avgrenset utvalg av innsjøer, er noe mer komplisert. Også slike vil det bli en del av.

Etter hvert vil mange hundre institusjoner begynne å bruke referansesystemet. Institusjoner som fordeler seg over mange fagområder, har forskjellige oppgaver i samfunnet og forskjellige tilknytningsformer til samfunnsmaskineriet. De vil komme til å ønske å kunne stille spørsmål av mange ulike typer.

Det skal stor fantasi og endel innsats til for å kunne forutsi de fleste av de typer av spørsmål som institusjonene kan komme til å ønske å kunne stille systemet i fremtiden. Når forfatterne ikke har forsøkt å sette noen slik fantasi og innsats, er det fordi vi ikke anser det som nødvendig. Med dagens teknologi er det nemlig praktisk mulig å bygge referansesystemet så fleksibelt at det vil kunne hankses med omtrent alle typer spørsmål.

Noen eksempler skal imidlertid gis på antatt aktuelle typer av spørsmål, og på aktuelle typer av spørgerne.

Først noen enkle, tilfeldige spørsmål som antyder spennvidden:

- \* Hvem har data om de kjemiske forhold i innsjøen Limingen i Nord-Trøndelag.
- \* Hvor fins det data om når de forskjellige løvtrearter spretter og feller bladene i forskjellige landsdeler og høydesoner.
- \* Informasjoner ønskes om gamle ferdselsveier (fra før 1800) mellom Hallingdal og Numedal.
- \* Hva fins om diversiteten i edderkoppbestanden i Norge nord for Lyngenfjorden.
- \* Hvilke permanente eller langtidsbrukte målestasjoner fins for fysisk-kjemiske og/eller biologiske data for luft, vann og/eller sedimenter i Hardangerfjorden.
- \* Fins det opptegnelser om lokale tåkeforhold rundt Krøderen fra tiden før 1976.

- \* Hvilke data fins om utbredelse av de enkelte sopparter i Akershus fylke.
- \* Hvem har utført systematiske tellinger av måker i Østlandets innlandsstrøk.
- \* Hva fins det utført av grunnboringer innenfor UTM kartrute 1841-IV.

Bak hvert spørsmål av denne type vil det finnes en interessant (en etat, institusjon, forening eller enkeltperson som har et motiv for å spørre). Som eksempel kan vi se litt på hva bakgrunnen for det siste spørsmålet kan være:

Grunnboring er en kostbar aktivitet som utføres av både offentlige etater og private firma. Hensikten med boringene vil ofte være forskjellig, - f.eks. leting etter drikkevann, kartlegging av grusforekomster eller vurdering av rasfarlighet. Som regel vil imidlertid resultater fra en boring for et bestemt formål ha informasjon som kan benyttes i andre sammenhenger.

I dag oppbevares dataene om boringene hos den som har gjort boringene. I noen tilfelle kan det være liten vilje til å offentliggjøre selve dataene for boringen p.g.a. konkurranse mellom prospekteringsfirma. Dette konkurranseforholdet er vel også en viktig årsak til at det ikke fins noen samlende oversikt over hvor boringer er utført [11].

Om man samlet referansene til boringene i referansesentralen, ville samfunnet bli spart for mange overflødige boringer. En slik oversikt behøver ikke å røpe noe av datainnholdet. Når det er aktuelt med en ny boring, vil referansesystemet gi en oversikt over tidligere boringer i området og vise til hvem som har selve dataene. I tillegg kan systemet gi tips om andre aktiviteter som en evt. boring bør ta hensyn til, f.eks. vikinggraver, sjeldne planter o.l.

Når det gjelder grunnboringer - og kanskje også mange andre typer av stedfestede data - er situasjonen den at hver enkelt dataprodusent ikke har så store dataarkiver at det er behov for eget edb-system. Dermed har man også ut fra en slik synsvinkel behov for at en sentral-institusjon tar seg av denne tjenesten.

I dette tilfellet var formålet med å spørre referansesystemet i første omgang å få vite hvilke boredataserier som forelå. Dernest ville man velge ut de seriene som ut fra informasjonene i referansen så mest interessante ut for det egne formål. Til slutt ville man så henvende seg til dem som hadde selve dataene fra disse serier for å få eller kjøpe dem.

I det neste eksemplet opptrer det et prosjekt som skal kartlegge naturressursene i et område. Dette prosjektet vil gjennom en tidsperiode ha et løpende behov for å kunne slå opp i selve dataarkivene ved visse institusjoner. Referansesystemet vil i første omgang hjelpe til med å finne ut hvilke institusjoner og dataarkiver dette gjelder. Dernest vil den om det ønskes, også stå som fysisk koblingsledd mellom spørger og de enkelte arkiver.

Under betegnelsen "Nord-Trøndelagsplanen", [9] er igangsatt et arbeidsprogram for å kartlegge naturressursene i Nord-Trøndelag fylke. Flere interesser i samfunnet vil kunne nyttiggjøre seg av resultatene fra dette arbeidet:

- bergverksnæringen vil stå sterkere
- forvaltningen av byggeråstoffer kan bedres
- konflikter i forbindelse med disponering av arealet kan unngås.

Geologi har en sentral rolle i dette undersøkelsesprogrammet, og NGUs databaser og referansesystem vil være en sentral informasjonskilde for fylkets planlegging. Man tenker seg en løsning der fylkesadministrasjonen har direkte kobling til NGUs datasystem via datanettet. På den måten vil fylker ha tilgang til geologisk informasjon.

Som vi kjenner til, er det imidlertid flere tema enn geologi som inngår i arealplanlegging i fylkes- og kommunalsammenheng. Ser vi f.eks. på disponering av et område til industriformål; vil følgende hensyn måtte tas:

- geologiske ressurser bør ikke båndlegges
- ødeleggelse av særegen natur må unngås
- forurensningseffekter må klarlegges
- ødeleggelse av fornminner må unngås
- kostbare undersøkelser som allerede er gjort, må unngås

I arbeidet med Nord-Trøndelagsplanen har man et eksempel på behovet for en tverrfaglig oversikt over kunnskap som foreligger for et geografisk område.

Enda et eksempel skal tas med - en institusjon som har fått i oppgave å studere forurensningssituasjonen for et vassdrag samt mulighetene for å forbedre denne:

Vassdraget x er ca. 12 mil langt. Dets lengste grener har sin begynnelse oppe på høyfjellet, men mesteparten av nedbørfeltet består av skog, mark og jordbruk. Noen industribedrifter og tettbebyggelser fins underveis, og ved utløpet til havet fins det en middelstor by. Det løper igjennom i alt 6 kommuner.

Oppgaven er mangesidig. De viktigste sidene skal omtales kort.

- a) Dagens forurensningssituasjon i selve innsjøen må kartlegges. Dette er en stor, men i og for seg grei, oppgave som bare i liten grad er avhengig av andres data.
- b) Informasjoner om dagens situasjon m.h.t. forurensningstilførsler er nødvendig her. Mange kilder må oppsøkes: Statistisk sentralbyrå, de kommunale helserådene, kommunens tekniske etater, herredsagronomene, industrien samt kanskje noen universitetsinstitutter.

- c) Situasjonen før i tiden er det viktig å kjenne til, både m.h.t. selve vassdraget og tilførslene. Alle kildene under punkt b vil være aktuelle her. I tillegg kan interessegrupper av typen fiskeforeninger og naturforeninger være gode kilder.

I praksis er det i dag en meget strevsom jobb å fremskaffe data fra de kildene som er nevnt under punktene b og c. I mange tilfeller kan den være større enn punkt a. Med et nasjonalt referansesystem som har informasjon om alle eksisterende dataserier, både nye og eldre, ville punktene b og c bare bli en brøkdel av hva de er i dag.



#### 4. LAGRING, SØKNING OG HÅNTERING AV DATA - IDAG OG I FREMTIDEN

Sett fra EDB-spesialistens synsvinkel er stedfestede data av kategorien "miljø" en krevende datatype. De fins i et meget stort antall varianter, de kan lett mistolkes, de har komplekse og skiftende strukturelle sammenhenger - og de fins i store mengder.

Sett fra miljødataprodusentens og miljødatabrukerens synsvinkler er det nødvendig å stille store krav med hensyn til regler og systemer for lagring og håndtering av slike data. De viktigste krav er disse:

- A. Dataene må være entydig tolkbare og lett håndterbare.
- B. Dataene må være lagret slik at de er lett tilgjengelige og lett søkbare ut fra mange synsvinkler.
- C. Dataene må være lett overførbare.
- D. Dataene må være beskyttet mot tap, skade og uautorisert bruk.
- E. Systemene må være fleksible m.h.t. økede datavolumer, nye datatyper, endrede datastrukturer og overføring av data fra en maskintype til en annen.

Med begrepet "entydig tolkbar" menes her at en dataserie innsamlet f.eks.

- o innen ett tidsrom
- o for ett bestemt formål
- o av en institusjon
- o innen ett geografisk område

må være så godt definert at den kan være utnyttbar i fremtiden for lignende eller forskjelligartede formål, av den samme eller andre institusjoner og for samme eller sammenlignbare områder.

Med begrepet "lett håndterbar" menes at dataene skal være angitt på en slik måte at det vil være lett å håndtere dem i forbindelse med registrering, lagring, søkning, bearbeidelse, utveksling, sikring m.v.

#### 4.1 Situasjonen idag

Om man ser på situasjonen rundt om i landets mange miljødata-institusjoner - med hensyn til hva de har pr. i dag av opplegg for lagring og håndtering av data - finner man et variert bilde:

- \* Ganske mange har ennå ikke noe EDB-basert system. For disse gjelder det (unntatt for dem som bare har meget små og enkle datasamlinger) at kravene ovenfor ikke er oppfylt.
- \* Endel har enkle, EDB-baserte systemer - utformet spesielt for eget bruk og for eget EDB-utstyr. Som regel er disse laget av institusjonene selv med mer eller mindre hjelp fra EDB-spesialister utenfra. Vanligvis oppfylles slike enkle systemer kravene ovenfor i meget liten grad.
- \* Noen har mer avanserte EDB-baserte systemer som i rimelig grad oppfyller kravene om lett tilgjengelighet og lett og allsidig søkbarhet - men som ikke oppfyller alle de øvrige fleksibilitetskravene ovenfor.
- \* Meget få, hvis noen, har EDB-baserte systemer som oppfyller alle kravene ovenfor.

Denne lite tilfredsstillende situasjon er etter hvert blitt erkjent av de fleste institusjoner - og mange av dem er nå iferd med å orientere seg mot mer moderne, EDB-baserte data-lagringmetoder.

Dette er selvfølgelig bra i seg selv. Men det fins i kortene fare for en uheldig utvikling her som det er grunn til å peke på - nemlig den at systemene blir laget av (eller i regi av) de enkelte institusjoner uten noen form for koordinering eller gjensidig tilpasning med andre institusjoner. Konsekvensene av, og mulighetene for å unngå denne fare skal belyses litt nærmere:

Det å utvikle et system som oppfyller alle de krav som er stilt opp ovenfor, er en stor oppgave. Og det er ikke bare den EDB-tekniske del av oppgaven som er stor. Også den forutgående systemanalytiske del vil være krevende. Og for mange institusjoner vil kanskje den begrepsmessige opprydding/finpussing som kreves innenfor de aktuelle fagområder, kreve aller mest. Slike systemer krever nemlig et langt mer konsist og detaljert faglig begrepsapparat enn det man tidligere hadde bruk for.

Som en følge av dette vil de enkelte institusjoner hver for seg vanligvis ikke ha råd til å lage systemene så gode som de burde være - med det resultat at de får systemer som bare delvis oppfyller de stilte krav.

Om man istedet kunne få til, for denne store utviklingsoppgaven, en koordinering mellom institusjonene (nasjonalt, regionalt og/eller bransjevis), ville alle institusjoner kunne få fullgode systemer. Og den samlede omkostning på landsbasis for å lage fullgode systemer til alle ville bare bli en brøkdel av hva det vil koste om alle hver for seg lager sine halvgode systemer. Med koordinering menes det her i første rekke to tiltak: At det utarbeides et sett av retningslinjer, råd og veiledninger m.h.t. hvordan slike systemer bør bygges opp. Og at det utvikles en rekke forskjellige EDB-programmoduler som kan egne seg som byggestener til de enkelte institusjoners egne lokale systemer. Kanskje burde også et utvalg av ferdige, komplette standardsystemer utvikles som kunne

tilbys de institusjoner som ikke trenger avanserte eller spesielle løsninger.

En koordinert løsning av oppgaven vil også gi en annen gevinst av stor verdi. Nemlig den at grunnlaget for et rasjonelt datautvekslings-opplegg mellom institusjonene vil bli langt bedre. Dette har sammenheng med at også gode utvekslingssystemer er avhengig av at det faglige begrepsapparat er mest mulig likt hos alle og at de individuelle systemopplegg og EDB tekniske løsninger er best mulig tilpasset de krav som utveksling stiller.

Man er i den nasjonale miljøforvaltning klar over den fare som her fins, - og har under vurdering forskjellige tiltak som kan fremme samordning av de mange spredte utviklingstiltak som nå er i startgropen eller underveis, FOSAMINO-rapporten, [5].

#### 4.2 Situasjonen i fremtiden

I dette avsnitt har vi tegnet en skisse av hvordan vi antar at fremtidens opplegg for datalagring og generell datahåndtering for miljødata i Norge vil se ut om noen få år (f.eks. i 1987). Vi har her forutsatt at de hovedideene som ble lansert i FOSAMINO-rapporten er blitt fulgt opp, - men ikke forutsatt at dette nødvendigvis har skjedd gjennom det konkrete utvalg av tiltak som rapporten anbefalte.

I vår fremtidsskisse går vi nærmere inn på to tema som har spesiell tilknytning til nærværende rapport.

- o Den overordnede organisering av miljødatahåndteringen (avsnitt 4.2.1).
- o Prinsippene for den lokale datalagring (avsnitt 4.2.2).

#### 4.2.1 Den overordnede organisering av miljødatahåndteringen

Først gir vi en kort delbeskrivelse av hvordan vi tenker oss at miljødatahåndteringen i hovedtrekkene er organisert i Norge i den nære fremtid (f.eks. i 1987 ?). Delbeskrivelsen omfatter bare de sider av opplegget som er av interesse i forbindelse med lagring og utveksling av data:

1. Miljødata vil som hovedregel lagret bli desentralt og ved de institusjoner som produserte dem. Unntak vil finnes for visse bransjer hvor alle eller noen datatyper lagres av en sentral institusjon på regionalt eller nasjonalt nivå. (Eksempel: Visse typer oceanografiske data lagres på nasjonalt nivå ved NOD). Unntak vil også finnes for noen institusjoner som har valgt å la andre institusjoner overta lagringsansvaret for seg. (Eksempel: Innen geologien overveies det at grunnboringsdataene skal lagres på nasjonalt nivå uansett hvem som produserer dem).
2. Det vil i Norge finnes et sentralt offentlig organ som har det overordnede ansvar for at bevaringsverdige miljødata blir tatt forsvarlig vare på. Alle miljødatalagrende institusjoner er registrert hos dette organ og er underlagt et "lagringsreglement for miljødata". Dette reglement vil inneholde retningslinjer bl.a. med hensyn til:
  - o sikring av dataenes fremtidige tolkbarhet og håndterbarhet (ved hjelp av kringinformasjoner)
  - o kunnkjøring om en dataserie eksistens (gjennom melding om den til det nasjonale referansearkivet)
  - o sikring av lett tilgjengelighet og søkbarhet for dataene (gjennom minstekrav til datalagringsssystemene)
  - o sikring av lett utvekselbarhet for dataene (gjennom tilknytning til offentlig standardiserte utvekslingsformater) (Jfr. pkt. 5 nedenfor)

- o sikring av dataene mot tap, skade og uautorisert bruk
- o vurdering av en dataseriers bevaringsverdighet og almene tilgjengelighet.

I tillegg til dette reglementet vil det finnes en rådgivende håndbok vedrørende all generell håndtering av miljødata - utarbeidet av det sentrale offentlige organet.

3. Alle institusjoner vil ha EDB-baserte lagringsarkiver. Disse lagringsarkivene vil stort sett være individuelt utformet, slik at de best mulig passer til den enkelte institusjons behov. (En institusjon står ganske fritt m.h.t. hvordan den vil utforme sine lagringsarkiver.) Den som ønsker det, kan imidlertid få råd og hjelp til dette fra sentralt hold. Dette tilbudet omfatter også ferdige standardsystemer eller byggestener for slike. Mer om dette tema fins i avsnitt 4.2.2 nedenfor.
4. For datalagring vil det gjelde at det anvendes koder for de fleste av kringinformasjonene (og til dels også for selve måledataene). Ved bruk av koder trenger man som oftest oppslagsregistre for å finne ut hva en kode betyr eller hvilken kode som skal brukes for en bestemt situasjon.

Noen av disse oppslagsregistrene vil være av lokal karakter og er laget av de enkelte institusjonene selv. De fleste vil imidlertid være av mer generell karakter og er derfor laget og vedlikeholdt sentralt i regi av det nasjonale miljødataorganet. Derfra vil de enkelte institusjoner etter bestilling ha fått tilsendt på magnetbånd de registre eller registerutdrag som de har bruk for. (Mer om dette tema fins i avsnitt 4.2.2 nedenfor).

5. Utveksling av dataserier mellom institusjoner vil foregå i stort omfang. Et generelt utvekslingsformat vil foreligge som kan dekke alle typer av miljødata. I reglementet

(omtalt under punkt 2 ovenfor) sies det at alle offentlige institusjoner og alle andre institusjoner som påtar seg offentlige oppdrag, skal være istand til å sende og motta data i dette formatet. (Men de er ikke nødt til å bruke det. Selv om standardformatet er greit å bruke, kan det finnes spesielle utvekslingssituasjoner hvor det er mer praktisk å bruke raskere, skreddersydde spesialformater. Dette vil f.eks. kunne gjelde situasjoner med så store datamengder og så hyppig utveksling at raskere utveksling gir en merkbar reduksjon av maskin og linjeleie).

I det generelle standardformatet kan man overføre både den interne struktur og selve dataene i en dataserie. For en institusjon å ta i bruk dette standardformatet er enkelt - fordi det fins ferdiglagede datarutiner som automatisk kan transformere en hvilken som helst strukturert dataserie fra det lokale format til standardformatet og omvendt. Det standardformat som man legger opp til å bruke i Norge i dag, kalles SOSI-formatet.

Ved utveksling av dataserier mellom institusjonene (ved hjelp av SOSI eller skreddersydde formater), vil det offentlige datatelenettet stort sett bli benyttet. Alle institusjonene kan koble sine dataarkiver til dette. Såfremt en dataserie er digetalt registrert, blir den vanligvis overført i form av digetale tekststrenger. Hvis den ikke er det, kan den overføres i bildeform. Denne teknikken brukes bl.a. for kart og fotografiske bilder.

I avsnitt 4.2.3 nedenfor er en slik utvekslingssituasjon beskrevet mer detaljert - sett fra brukersiden.

#### 4.2.2 Prinsippene for den lokale datalagring

Ved lagring av data vil det sammen med dataene være lagret en rekke "kringinformasjoner" for at dataene skal være tolkbare (informasjon om tid, sted, parametre, medier, metoder osv.). For at dataene og kringinformasjonene samtidig skal være best mulig håndterbare (m.h.t. registrering, lagring, søkning, osv.) fins en del standarder m.h.t. koder og prosedyrer.

Under avsnitt 4.2.1 ovenfor er det underpunkt 3 fortalt at alle miljødatainstitusjoner har sine individuelt utformede dataarkiver - og at alle disse er basert på database-prinsippet. Under punkt 4 er det videre fortalt at de fleste kringinformasjonene, og til dels også selve måleresultatene, er angitt i form av koder, - og at kodene er forklart i spesielle oppslagsregistre. Noen av disse oppslagsregistrene kan gjelde lokale forhold og er laget lokalt. De fleste av dem er imidlertid av generell karakter og er kopiert i utdrag fra sentrale oppslagsregistre som er laget/vedlikeholdt på nasjonalt nivå.

I tilknytning til de nasjonale oppslagsregistre fins det også spesielle håndteringsprogrammer som gjør det mulig å foreta oppslag på en meget rasjonell måte. Disse håndteringsprogrammer kan leveres sammen med registerutdragene fra sentralt hold til de enkelte institusjoner.

I prinsippet vil datalagringsopplegget ved en norsk miljødata-institusjon i vår fremtidsskisse 1987 kunne skisseres slik som vist i fig. 4.1.

Skissen viser

- o at det ved en institusjon kan finnes mer enn ett data-arkiv. Grunnen til det er at en institusjon kan ha flere kategorier av data som det p.g.a. forskjellige inndataopplegg og datastrukturer er mest rasjonelt å holde fra hverandre. (Det at dataene fins spredt i



flere forskjellige arkiver, betyr ikke at det er vanskelig å føre dem sammen. En egen mekanisme fins for utsøking og overføring til en felles "arbeidsfil" hvor dataene fra de forskjellige arkivene kan kobles sammen for videre bearbeidelse (stiplet del av figuren).

- o at det ved en institusjon fins flere oppslagsregistre. Hvor mange vil variere fra institusjon til institusjon, men de fleste vil i det minste trenge oppslagsregistre for parametre, metoder og målestasjoner. Hvilke typer av oppslagsregistre som eksisterer, er antydnet i kapittel 7 nedenfor.
  
- o at det fra hvert av dataarkivene fins en forbindelse til det nasjonale referansearkiv via et eget kommunikasjonsprogram. (Prikket del av figuren). Denne forbindelse kan brukes av begge parter: Av institusjonen når denne skal sende inn nye dataseriereferanser eller endre noen av de bestående - og av referansearkivet når dette skal innhente mer detaljer om en dataserie enn dem som selve referansen angir. (Institusjonen vil ha anledning til å sette restriksjoner for den sistnevnte bruk av denne forbindelsen).
  
- o at oppslagsregistrene og dataarkivene er klart adskilt i en institusjons datalagringsystem. Dette er et viktig prinsipp som er basert på følgende forhold:
  - Dataarkivene er individuelt utformet for best mulig å passe til de datakategorier de skal inneholde og de formål dataene skal brukes til.
  
  - Oppslagsregistrene derimot er av to grunner stort sett likt oppbygget ved alle institusjonene og for alle dataarkivene. For det første fordi man derved har kunnet utvikle dem sentralt - slik at enkeltinstitu-

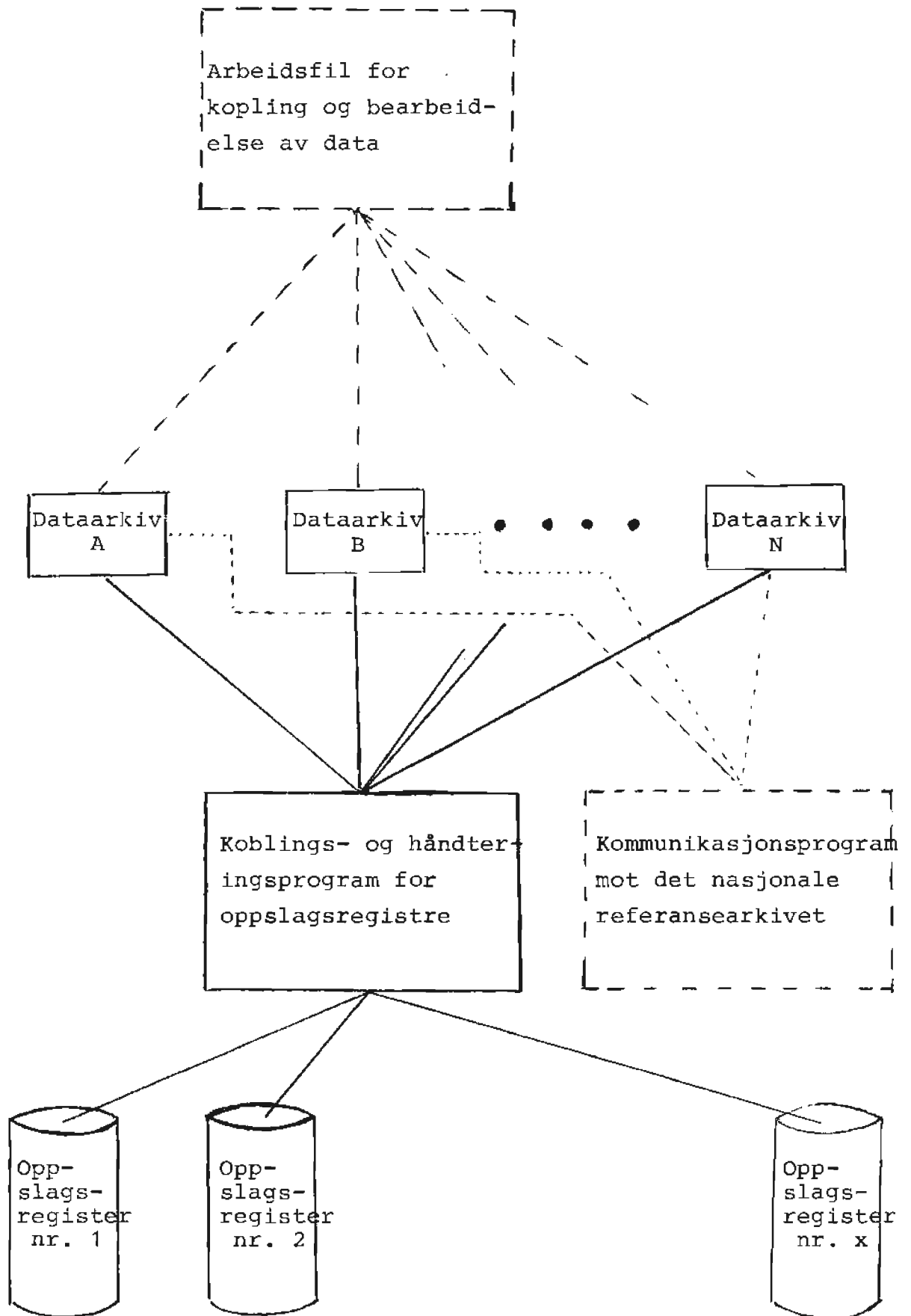


Fig. 4.1. Typisk struktur for en miljødatainstitusjons data-lagrings-system i 1987

sjonene har sluppet å bruke sine ressurser til det. (Det koster mye i tid og penger å bygge opp slike registre). For det annet innebar det at oppbyggingen ble utført i regi av en sentral instans, at det ble mulig å standardisere og eller å harmonisere det faglige begrepsapparat og bruken av koder. Dermed er det blitt langt enklere både å håndtere dataene internt ved institusjonene og utveksle dataserier og databehandlingsmetoder mellom dem.

Det skal her for sikkerhets skyld presiseres at den standardisering og harmonisering av det faglige begrepsapparat og bruken av koder som er oppnådd, med hensikt er gjennomført på en slik måte at den ikke har redusert den enkelte institusjons frihet til internt å bruke de begreper og koder den vil. (Standardiseringen ble innført med det primære formål å sikre at datautveksling mellom institusjoner skal kunne foregå på en sikker og rasjonell måte).

I praksis har det vist seg at den standardisering og harmonisering av selve det faglige begrepsapparat som ble oppnådd, stort sett blir fulgt ved de fleste institusjoner. Når det gjelder kodestandardene, er det derimot endel institusjoner som har valgt å bruke mer skreddersydde og enkle kodesett ved sin interne datavirksomhet.

- o Når man i et dataarkiv skal registrere, søke eller hente ut data, bruker man koder fra det kodesett som institusjonen har valgt for dette arkivet. Hvis den aktuelle databehandlingssituasjon tilsier at kodene må bli kontrollert og/eller tolket, vil dataarkivet automatisk kalle opp den enheten som kalles koblings- og håndteringsprogrammet (KHP).

- En slik kontroll går ut på å undersøke om koden er entydig og gyldig. Hvis den er så, meldes det straks

tilbake til dataarkivet. Hvis den ikke er, får brukere straks melding på sin skjerm om hva som er galt og hjelp til å finne den riktige koden.

- En slik tolkning går primært ut på å finne ut hva koden betyr. (F.eks. å finne ut hvilket parameternavn en parameterkode står for). Den kan imidlertid, hvis man ber om det, også omfatte mange andre opplysninger (som f.eks. den nevnte parameters navn på engelsk, dens kjente, synonyme navn og dens plassering i et eller flere kjente parameterhierarkier). Resultatet av tolkningen går straks tilbake til dataarkivet og anvendes der slik brukeren har bestemt.

KHP-programmet tar seg av denne kontrollen og tolkningen. Hvis oppgaven er enkel og gjelder ofte benyttede koder, gjør det så på basis av et lite, innebygget snarregister. Om svaret ikke fins der, finner KHP-programmet selv frem til de adekvate oppslagsregistrene og henter inn opplysninger derfra. Også kontakten med oppslagsregistrene foregår automatisk og raskt og uten at brukeren behøver å foreta seg noe i den anledning.

Selve KHP-programmet er et standard basisprogram som er utviklet i regi av - og eies av - det sentrale miljødataorganet. Det er stilt til disposisjon for institusjonene mot en lav avgift. Ved den enkelte institusjon har man måttet tilpasse programmet m.h.t. hvilke dataarkivtyper den har, hvilke oppslagsregistre den trenger og hva slags maskinutrustning den har. Denne tilpasningen har vært enkel å gjennomføre.

De fleste av de oppslagsregistrene som fins ved en institusjon er utdragskopier av komplette originale registre som er laget og vedlikeholdt i regi av det sentrale miljødataorganet. Vanligvis har de enkelte institusjonene også noen få oppslagsregistre som gjelder lokale forhold og som er utviklet lokalt. Eksempler på slike kan være prosjektregistre og personregistre.

Når det gjelder de enkelte institusjonenes dataarkivopplegg, er de vanligvis utviklet lokalt med assistanse av egne eller innleide EDB-spesialister. De aller fleste av dem er imidlertid basert på ferdige databasesystemer og andre systemmoduler som er stilt til disposisjon fra det sentrale miljødataorganet eller andre kilder. Alle dataarkivene er forutsatt å oppfylle visse minstekrav, slik som nevnt i avsnitt 4.2 pkt.2 ovenfor.

## 5. KRAVSPESIFIKASJON TIL REFERANSESYSTEM

### 5.1. GENERELLE KRAV

Et referansesystem skal være et system oppbygd av manuelle og automatiske rutiner som skal lagre, gjenfinne og formidle data om data.

Data om data

Systemet skal behandle referanser til stedfestede data som hører til miljø- og ressursvirksomheten i Norge.

Stedfestede data

Systemet skal inneholde referanser til dataserier uten å begrense seg til en bestemt tilstand på dataene. Noen referanser vil vise til primærdata, andre til bearbeidede data, og atter andre til dokumentasjon om dataene (rapporter og publikasjoner).

Referanser til  
- primærdata  
- bearbeidede data  
- dokumentasjon

Som eksempler på aktuelle datatyper kan nevnes

Eksempler

- måledata fra naturen
- miljøprøver
- geologi
- vann
- rapporter, kart og publikasjoner
- miljødataprosjekter og -planer
- forurensningskilder og utslippsdata
- data om vern og kulturminner

uten derved å ville utelukke andre.

Referansene skal lagres vhj. av edb-baserte lagringsmetoder og slik at brukerne til enhver tid kan hente fram igjen interessante referanser. Lageret med referanser, kalles videre i dette kapitlet for en referansebase.

Referansebase

Systemet skal betjene både dataprodu-  
senter og databrukere innenfor miljø-  
data og ressursdatavirksomheten. I  
utgangspunktet skal det ikke være  
restriksjoner i tilgang på referansene.

Hvem skal bruke  
systemet?

Systemet skal ikke komme som erstatning  
til eksisterende informasjonssystemer  
hos enkelte dataproducenter. På den  
ene side skal systemet ivareta data-  
produsentens frihet og ansvar til selv  
å organisere sine databaser og informa-  
sjonssystemer. Referansesystemet må  
fungere naturlig mot slike lokalsystemer.  
På den annen side må systemet kunne yte  
service til mindre dataproducenter som  
ikke har ambisjoner eller ressurser til  
å bygge opp egne informasjonssystemer.

Store dataproducenter

Systemet skal også bidra med å knytte  
kontakten nærmere mellom brukere og  
fageksperter.

Kontakt bruker/  
fagekspert

For å kunne mestre datamengden og sam-  
tidig sikre et minimumsnivå av informa-  
sjonsinnhold i referansene, er valgt å  
legge opp til et todelt informasjons-  
system.

Todelt datalagring

En bruker som "henvender" seg til  
referansesystemet skal først betjenes  
av et overordnet referansearkiv som  
forteller "hvem har data om hva fra  
hvor og når" ned til et visst detalj-  
nivå. Denne delen av systemet som vi  
kaller referansesentralen, skal være  
edb-basert, men også ha et manuelt  
servicetilbud.

Nasjonal refe-  
ransesentral

Ytterligere detaljer om dataene skal kunne tilbys. Ansvaret for denne tjenesten tenkes utført enten på manuelt vis av fagpersonell eller ved at systemet påkaller edb-baserte informasjonsentraler ved faginstitusjoner. Referansesystemet skal formidle spørsmålene som brukeren har stilt når det er behov for detaljinformasjon og når brukeren ønsker det.

Fagpersonell

Fagrettede informasjonsentraler

Formidling av spørsmål

Mange fagrettede institusjoner vil ha høy detaljeringsgrad i referansene som de lagrer i sine egne informasjonssystemer. I det sentrale referansesystemet vil informasjonen være nær begrenset, men skal omfatte

Detaljeringsgrad

- referansens tittel
- stedfesting
- tidfesting
- emneord
- produsentopplysning
- lagring opplysning

dvs. tilstrekkelig til å kunne identifisere hver enkelt dataserie.

Referansesystemets verdi vil øke med økende presisjon i de opplysninger som gis, i første rekke stedfesting og emneord. Referansesystemet må legges opp slik at presisjonen kan økes etter hvert som de enkelte fagområder får en standardisert bruk av emneord.

Presisjonsgrad



## 5.2. FUNKSJONELLE KRAV

### 5.2.1. Funksjonell hovedstruktur

Fig. 5.1 viser den prinsipielle og logiske hovedstrukturen i referansesystemet. I praksis vil den ha flere varianter og være mer infløkt.

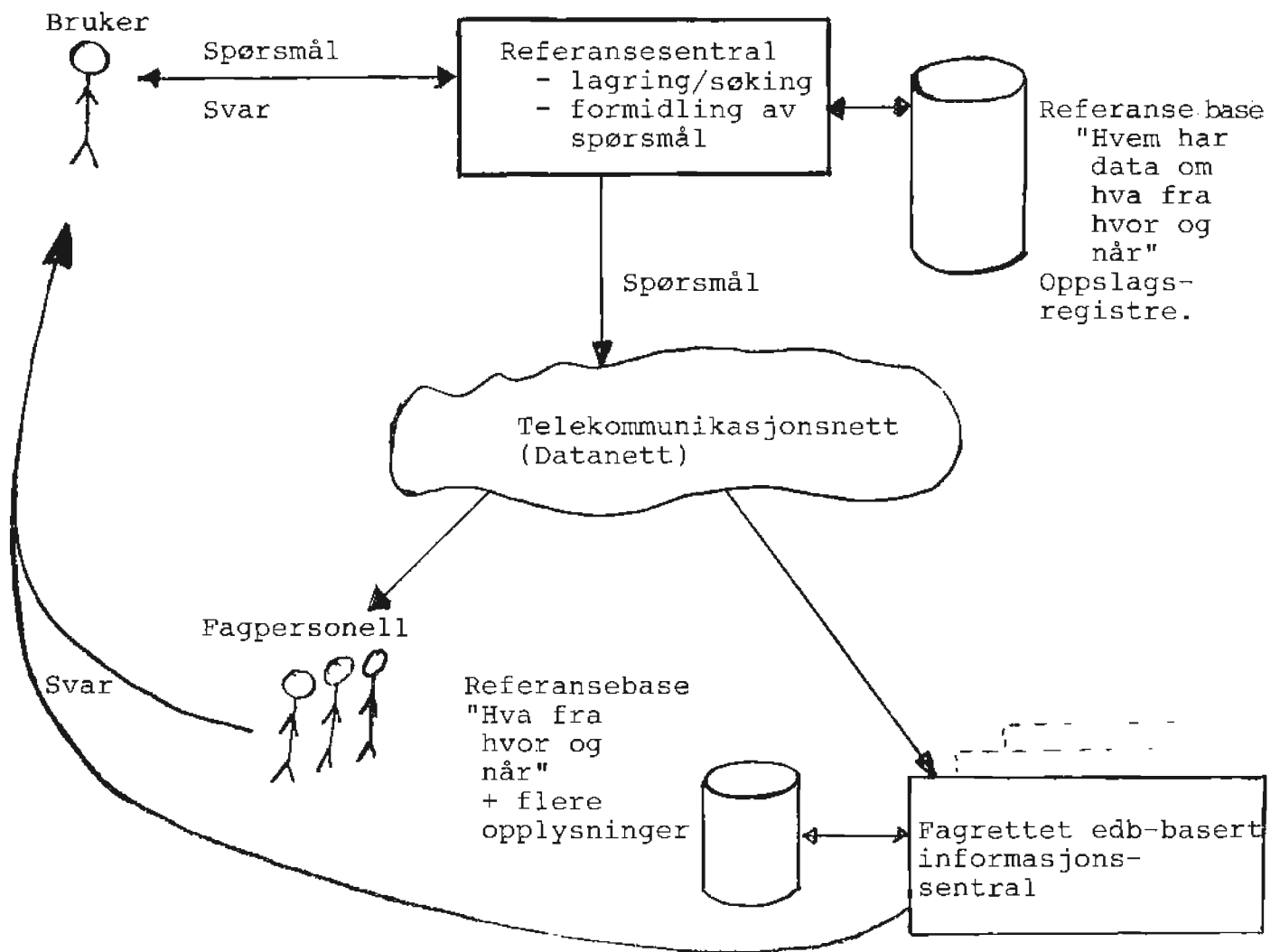


Fig.5.1. Referansesystemets hovedstruktur

Referansesystemet består av følgende hoveddeler:

- Brukerne
- Nasjonal referansesentral
- Telekommunikasjonsnett
- Fagrettede informasjonssentraler
- Fagpersonell

At brukerne hører med til systemet, er selvsagt, ellers har det ikke sin berettigelse. Brukerne er i denne sammenheng definert som de personer som søker informasjon. Brukeren er ikke en homogen enhet. Teknisk sett vil de bestå av enkeltpersoner som vil ha et av følgende grensesnitt mot systemet:

- Telefon (post)
- Dataskjerm knyttet direkte mot referansesentralen, fast eller oppringt
- Dataskjerm knyttet til referansesentralen via en datamaskin
- Teletex, teledata

Referansesentralen er en hoveddel som består av

- manuelle tjenester
- maskinelle tjenester

De manuelle tjenestene vil være nødvendig for

- drift av referansesentralens referansebase,
- ta seg av henvendelser fra brukere som ikke kan benytte de maskinelle tjenester direkte,
- ajourhold av referansebasen.

De maskinelle (edb-baserte) funksjoner vil operere mot en referansebase som inneholder data om data:

"Hvem har data om hva fra hvor og når" (Referanser)

Referansesentralens hovedfunksjoner er:

- Lagring av referanser
- Søking i referansebase
- Formidling av spørsmål

Lagring og innsamling av referanser i referansesentralen er behandlet i kapittel 5.4.

Informasjon som kan søkes ut av referansesentralens referansebase vil være begrenset. Referansenes datainnhold er diskutert i kapittel 5.3. I mange tilfelle vil brukeren ha behov for mer detaljert informasjon enn det som er funnet ved søket eller ha mer faglig støtte. Dersom brukeren ønsker det kan funksjonen, formidling av spørsmål, ta seg av dette behovet. Brukerens ønske (spørsmål) formuleres i referansesentralen til en melding. Meldingen får en eller flere adresser og sendes som "elektronisk post" til faginstitusjoner som kan gi svar.

Som alternativ til å benytte den formidlingen, kan brukeren henvende seg til dataprodusenten direkte.

Det offentlige datanettet formidler meldingene.

Meldingene er adressert til prinsipielt to typer tjenester

- Fagpersonell
- Fagrettede informasjonssentraler

Med fagpersonell tenkes her personell ute i institusjonene som er utpekt til å besvare slike henvendelser innen bestemte fagområder, men ikke som fulltidsbeskjeftigelse. Meldinger til fagpersonell lagres i referansesentralen i vedkommendes "postkasse". Med jevne mellomrom eller ved anledning, kobler fagpersonellet seg opp mot referansesentralen og får spørsmålene utskrevet.

Hvordan fagpersonellet vil besvare henvendelsene vil variere, men følgende muligheter finnes

- søke mer "fornuftig" i referansesentralens referansebase,
- svare ut fra sin faglige oversikt,
- søke i fagrettede informasjonssentraler

Fagrettede informasjonssentraler er sentraler opprettet ved enkelte faginstitusjoner. Organiseringen synes her å være varierende. Innenfor noen fagområder bygger institusjonene egne referansesystemer (f.eks. vann). Innenfor andre fagområder velger man å samarbeide om dette (f.eks. geologi).

Noen av de fagrettede referansesentralene vil være helautomatisert og vil kunne besvare meldingene fra referansesentralen direkte, mens andre vil være manuelt betjent. Meldingen som kommer fra referansesentralen, vil da bli tatt hånd om manuelt og de konkrete søkeoperasjonene blir formulert lokalt.

Informasjon tilbake til brukerne vil fortrinnsvis gå pr. post eller telefon.

#### 5.2.2. Den nasjonale referansesentralen

Den nasjonale referansesentralen er en datamaskin med en driftsorganisasjon.

De viktigste bestanddelene (egenskaper og funksjoner) i referansesentralen er

- referansebasen
- søkeprogrammet
- portnerfunksjonen
- brukertjenesten
- ajourholdfunksjonen

Referansebasen er lagringssystemet for referansene. Innholdsmessig er referansebasen tredelt

Referansebase

- lagring av referanser
- oppslagsregistre
- meldingsformidling

Se forøvrig kap. 5.4.

Selv om dataskjerm blir et mer og mer vanlig verktøy på arbeidsplassene, vil mange brukere ha behov for at andre foretar selve informasjonssøket på referansebasen. En viktig driftsfunksjon ved referansesentralen vil være å motta henvendelser fra slike brukere.

Brukertjeneste

Et eget program skal gjøre det mulig å søke informasjon ut av databasen. Det er dette programmet som brukeren kommer i kontakt med.

Søkeprogrammet er tilgjengelig fra dataskjerm og gir svar fra databasen uten nevneverdig ventetid.

Brukere som har tilgang på dataskjerm i sitt daglige arbeid skal kunne bruke søkeprogrammet selv.

Referansesentralen må derfor være tilknyttet de viktigste datamaskinene i miljødatavirksomheten:

- SDS (Statens driftsentral)
- KD (Kommunedatasentralene)
- FKK-ene (Fylkeskartkontorenes maskin)
- Faginstitusjoner

En nærmere beskrivelse av hvordan søkeprogrammet virker er beskrevet i kap. 5.5.

Referansesentralens datamaskin skal fungere som portnermaskin mot de enkelte fagmiljøer. Med det menes at brukere som først kobler seg til referansesentralen, kan - om de der

Portnerfunksjonen

ikke finner detaljer nok få - distribuert sine ønsker og spørsmål videre.

Meldingsformidlingen vil som nevnt tidligere anta forskjellig form.

Eksempler

- Meldingene kan adresseres til fagfolk. De legges i en elektronisk "postkasse" som tømmes og besvares av fagpersonell ved anledning.
- Meldingene kan sendes til en datamaskin ved et bestemt fagmiljø som har lokale informasjonssystemer. Her vil forespørselen bli tatt hånd om av bibliotekstjenesten eller bli behandlet automatisk.
- Portnerfunksjonen kan også virke slik at brukerne blir koblet direkte til informasjonssystemet i fagmiljøene. Et slikt system kan være et søkeprogram mot et fagrettet referansearkiv.

For at portnerfunksjonen skal kunne fungere, må referansesentralen lagre adresseopplysninger, såkalt rutinginformasjon.

Rutinginformasjon

Referansesentralen må ha informasjon om HVOR detaljinformasjon kan hentes og HVORDAN.

SSBs referansearkiv, ref. [6], inneholder slike opplysninger i informasjonskilderegisteret, men det vil være nødvendig med tilleggsinformasjon som

- om en dataprodusent kan gi tilleggsinformasjon,

- hvordan meldinger kan distribueres videre til kontaktpersonene.

Meldinger med spørsmål fra bruker er tenkt lagret i "postkasser". For hver fagperson eller hver lokal informasjonssentral som kan besvare detaljspørsmål eller hjelpe til i et informasjonssøk, opprettes en "postkasse", en datafil, inne på referansesentralen. Postkassene kan tømmes når det best passer lokalt, ved at man kobler seg til referansesentralen med en dataskjerm eller datamaskin og får overført meldingene.

Elektronisk  
postkasse

For mange kontaktpersoner som kan gi detaljinformasjon, vil det ikke være mulig å få tilsendt meldingen som elektronisk post. Referansesentralen må derfor ha et brevprogram som kjøres med jevne mellomrom. Meldingene skrives ut på papir og sendes kontaktpersonene.

Brev

Ajourhold av referansesentralens referansebase vil prinsipielt skje på to måter:

Ajourhold av  
referansebasen

- registrering av referanser ved referansesentralen
- automatisk overføring av referanser fra de lokale informasjonssentralene.

En viktig driftsfunksjon ved referansesentralen er innsamling, registrering og korreksjoner av referanser. En stor del

av dataprodusentene vil måtte rapportere inn referanser på manuelt vis ved utfylling av skjema. Driftspersonellets oppgave blir å

- følge opp innrapportering
- registrere referansene på skjerm
- foreta korrektur.

Når det gjelder dataprodusenter med edb-baserte informasjonssystemer, vil det være aktuelt å generere referansene automatisk og overføre dem til referansesentralen i edb-lesbar form. Noen slike dataprodusenter har påtatt seg rollen som nasjonalt eller regionalt sentralarkiv innen sitt fagområde (NGO, NOD, NGU o.fl.) og mottar referanser fra andre dataprodusenter. Mange referanser fra dataprodusenter uten eget edb-verktøy vil således komme med i den automatiske genereringen av referanser.

### 5.2.3. Det offentlige datanettet

Det offentlige datanettet er et telekommunikasjonsnett for overføring av data. Nettet er oppbygd etter en internasjonal standard (CCITT).

I referansesystemet vil datanettet inngå som en viktig enhet på 2 måter

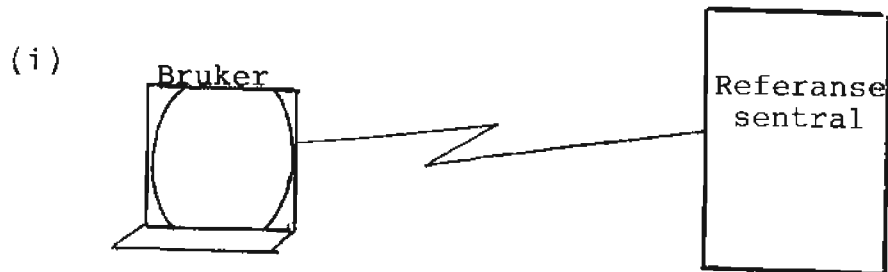
- interaktiv oppkopling til referansesentralen
- formidling av meldinger

Med interaktiv oppkopling til referansesentralen, menes den funksjon der en bruker med dataskjerm kan kjøre referansesentralens søkeprogram.



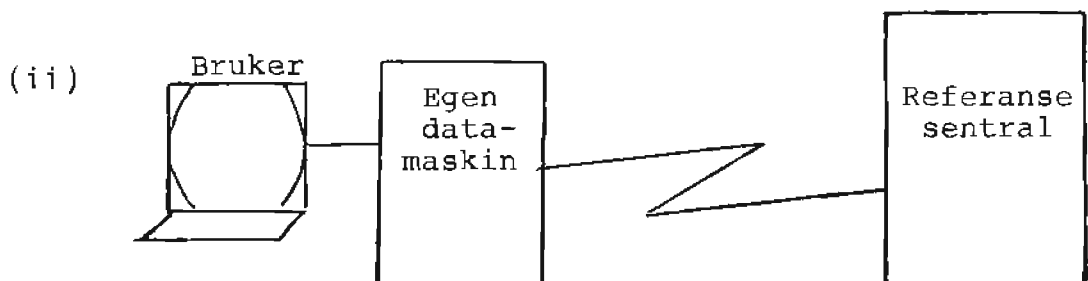
Referansesentralen må ha et utvalg av kommunikasjonsmuligheter. Etter hvert som Televerket tilbyr nye tjenester på dette området vil det fra brukersiden oppstå ønsker om at referansesentralen kan betjene nytt utstyr.

La oss se litt på noen typiske situasjoner



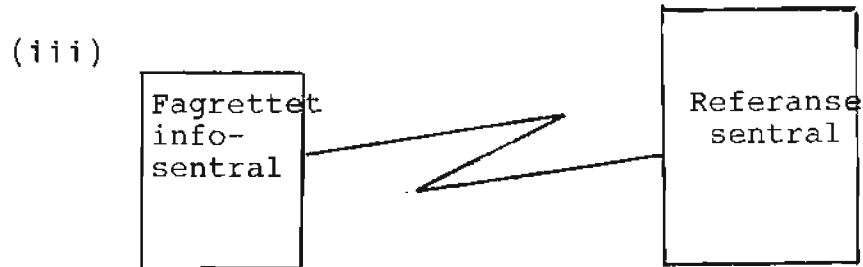
En bruker kopler seg til referansesentralen på et oppringt samband.

Referansesentralen må kunne betjene de 2 vanligste former for slik samband - Modem og Datex, asynkron forbindelse. Teletex og pakkesvitsjet datanett vil også bli aktuelt.



En bruker som vanligvis står tilkopleet egen data-maskin, må kunne koble seg videre til referansesentralen.

Mest aktuelt vil være oppringt samband, men synkron dataoverføring i tillegg til asynkron forbindelse vil måtte tilbys.



Når en fagrettet informasjonssentral skal hente meldinger fra referansesentralen, kreves samme muligheter som under (ii).

Når referansesentralen selv vil oversende meldinger til en eller flere fagrettede informasjonssentraler, må den i tillegg være utstyrt med automatisk nummervelger.

Pakkesvitsjet nett eller Teletex vil være nødvendig, men det avhenger av at også det fagrettede informasjonssystemet benytter samme tjeneste.

#### 5.2.4. Fagpersonell

Referansesystemet er tenkt ha tilknyttet to typer fagpersonell.

- Driftspersonell
- Fagekspert

Tilknyttet den daglige driften, tenkes ansatt personell som kan bistå brukerne i søk på referanser i referansesentralen. Først og fremst må personellet kjenne søkesystemet, dernest kreves en tverrfaglig oversikt slik at søkekriteriene kan formuleres fornuftig. Fagpersonellet bør ha ansvar for vedlikehold av databasene.

Fagekspertise innen bestemte fag - geologi, vann, planlegging, forurensning, fornminner osv., tenkes også være tilknyttet referansesystemet.

Disse fagpersonene har sin faste arbeidsplass ved sine respektive faginstitusjoner og vil ha sine "postkasser" i referansesentralen.

Hovedverktøy for fagekspertene vil være

- dataskjerm
- telefon

Dataskjerm og evt. telefon benyttes til å kople seg opp til referansesentralen for å tømme "postkassen" eller søke i referansedatabasen.

Dataskjermen benyttes evt. også mot lokale informasjonssentraler for å søke fram detaljinformasjon.

Telefon vil være mest naturlig for nødvendig kontakt mellom fagekspert og bruker.

#### 5.2.5. Fagrettede informasjonssentraler

Fagrettede informasjonssentraler er et servicetilbud som større faginstitusjoner har etablert eller vil komme til å etablere.

Formen for service kan variere:

- noen institusjoner har en kontaktperson som har oversikt over arkivene,
- noen institusjoner har fagbiblioteker som har som hovedoppgave å gi informasjon,
- noen institusjoner har bygd opp edb-baserte informasjonssystemer som betjenes av personell ved institusjonen,
- noen institusjoner har edb-systemer hvor eksterne brukere kan benytte systemet direkte for informasjonssøk.

Kravene til funksjoner ved de fagrettede informasjonssentralene vil variere med automasjonsgraden. Utføres servicefunksjonene manuelt, er det et minimumskrav at institusjonen er utstyrt med en dataskjerm tilkopledd datanettet. Da kan man tømme "postkassen" med jevne mellomrom. Svar til brukeren vil skje pr. telefon eller post.

Dersom institusjonen er utstyrt med mer avansert datautstyr som tar hånd om deler av informasjonsbehandlingen, kan det være aktuelt å kople den lokale datamaskinen opp mot referansesentralen. Personellet disponerer en dataskjerm mot lokalanlegget.

Overføringen av meldinger fra referansesentralen kan skje på en av følgende to måter:

i: Når en melding er klar, kopler referansesentralen seg automatisk opp mot lokalanlegget og sender meldingen dit. Her legges den i en lokal "postkasse".

ii: Det er den fagrettede informasjonssentralen som bestemmer når den vil ha overført meldinger. Den kopler seg opp mot referansesentralen og tømmer sin "postkasse" og overfører meldingene til lokalanlegget.

Referansesystemet må kunne behandle begge prinsipp for kommunikasjon med fagrettede informasjonssentraler.

Med tiden må vi være forberedt på at noen lokale informasjonssentraler kan operere helautomatisk.

Følgende krav til lokalanlegget må da kunne oppfylles:

- Referansesentralen må kunne kople seg opp mot lokalanlegget når en melding foreligger.
- Meldingen består av brukerens spørsmål som ble brukt mot referansesentralens referansebase. Lokalanleggets søkeprogram må kunne tolke meldingen, og konvertere spørsmålene til det format som det lokale søkeprogrammet krever.
- Minstekravet til informasjon tilbake til brukeren bør være en melding om hvor mange referanser som søkingen resulterte i. Meldingen sendes til brukeren pr. post/telefon. Resultatet bør kunne sendes tilbake til referansesentralen.

Brukeren kan deretter bestille utskrift av informasjonen som er søkt ut og få den tilsendt.

- Den idelle situasjon er at brukeren kan søke i de fagrettede informasjonssystemene på samme måte som på referansesentralen. I tillegg til å sende søkeordene som en melding, må referansesentralen og det lokale søkeprogrammet utveksle meldinger som tilsvarer de funksjonene som referansesentralens søkeprogram har:
  - gjenta søket med andre søkeord
  - lese igjennom resultatet av søket direkte på skjerm
  - bestille utskrift av deler av resultatet.

Kap. 8 behandler oppbygging av en fagrettet informasjonssentral.

### 5.3. REFERANSENE'S INNHOLD

Dette kapitlet beskriver datainnholdet for referanser som er lagret i referansesentralens database. Vedlegg I gir en tabellarisk oversikt over variablene som er tenkt inngå i referansene.

#### 5.3.1. Referansens identifikasjon

Referanseidentifikasjonen består av 3 variabler som identifiserer og klassifiserer referansen med referansetype, referansenummer og referansetittel.

##### Referansetype

Referansetypen angir hva slags dataserie referansen viser til. Hovedinndelingen av data som referansene viser til, er

- primærdata

- bearbeidede data
- dokumentasjon om data

Denne inndelingen kunne benyttes, men mer detaljerte angivelser er å foretrekke, f.eks.

- måledata
- miljøprøve
- forekomst
- rapport
- prosjekt
- osv.

#### Referansenummer

Referansenr er et løpenr. som er en entydig identifikasjon på referansen. Referansenummeret er oppbygd på samme måte som nummereringen i SSBs referansearkiv, ref. [6].

Format:

<institusjon>.<informasjonskildenr.>.<løpenr.>

Eks.

Referanse til et kart i NGUs kartarkiv  
1501.01.162

#### Referansens tittel

Dette feltet angir dataseriens betegnelse. Feltet er i fritekst og har egentlig ingen øvre grense i antall ord, men skal i prinsippet være kort.

Eksempler:

- "Refleksjonsseismiske undersøkelser i Drammensfjorden"
- "Berggrunnsgeologisk kart"
- "Måling av surhetsgrad i Seljordvannet"

Ord i tittelen har ikke status som stikkord i søkingen.

### 5.3.2. Stedfesting (geografi)

Stedfesting kan skje på fire måter, hvorav de tre første gir søkemulighet.

- administrativ stedfesting
- faglig stedfesting
- kartografisk stedfesting
- lokale stedsnavn

#### Administrativ stedfesting

Med administrativ stedfesting menes at det skal angis stedsnavn iflg. offentlig administrativ inndeling av landet.

- grunnkrets
- bygdelag/bydel
- kommuner
- fylker

Hovedregelen skal være at man benytter den inndeling som nærmest omslutter det området som referansen gjelder for. Dersom data-serien ligger innenfor en grunnkrets angis denne, strekker den seg over flere grunnkretser angis enten hver enkelt grunnkrets eller kommunen som er nærmest omsluttende område. Gjelder data-serien flere kommuner kan enten alle kommuner angis eller fylket.

Minst ett offisielt administrativt stedsnavn må alltid registreres. Administrative stedsnavn styres av et oppslagsregister som er hierarkisk søkbart.

#### Faglig stedfesting

Med faglig stedfesting menes en stedfesting som følger en faglig inndeling av landet.

- Eks.
- nedbørfelt
  - vegetasjonssone
  - geologi

Denne stedfestingen er frivillig og det skal være mulig å angi flere. Sammen med selve stedfestingsbegrepet må man angi hvilken

faglig inndeling som menes. Den faglige stedsfestingen styres av et oppslagsregister som også er hierarkisk søkbar.

Format:

<fag><geografisk betegn.>

Eksempler

V-401.2.1 - etter vassdragsinndelingen.

#### Kartografisk stedfesting

Med kartografisk stedfesting menes angivelse av det kartverk som dekker referansen. Bruk av målestokk vil variere fra dataproducent til dataproducent. Punktkoordinater må kunne angis. Den kartografiske stedfesting er hierarkisk søkbar.

#### Lokale stedsnavn

Lokale stedsnavn kan oppgis for å presisere, men egner seg ikke som søkeord ettersom de ikke er entydig i landsmålestokk.

Eks. Sandelva, Moen, Strømmen.

#### 5.3.3. Tid

Hver referanse skal være tidfestet.

Minimumangivelse er årstall. Månedsnr. kan angis i tillegg.

Ofte angis bare ett tidspunkt, men det vil også være behov å angi et tidsrom.

Format:

<årstall>:<mndsnr.>

eller

<AAAA>:<MM>



Eksempler:

1982

1975:09

1975:01 - 1980:01

#### 5.3.4. Emneord

For hver referanse skal det være mulig å angi et ubegrenset antall emneord som man vil ha mulighet å søke på under et informasjonssøk. Emneordene deles inn i følgende klasser:

- medium
- aktivitet
- parameter
- øvrige emneord

Hovedinndelingen av emneord er lik med den i Samlet Plan, ref. [17].

##### Medium

Medium forteller hvilken type naturressurs som referansen omhandler. Detaljeringsgraden vil variere slik at det er nødvendig med flere hierarkisk oppslagsregister i samspill med disse emneordene. Registreringen av emneordene kan således skje på det nivå som er naturlig.

##### Format

<medium>

##### Eksempler

løsmasser  
hydrologi  
topografi  
ferskvann

Emneordene som brukes må selvsagt godkjennes av systemets oppslagsregister. Minst ett emneord under medium må angis, men det skal være mulighet å angi flere.

#### Aktivitet

Her angis aktiviteten/forhold knyttet til det medium som er angitt referansen.

Format:

<aktivitet>

Eksempler:

støy

utslipp

Slike emneord må også kontrolleres av et oppslagsregister.

#### Parametergruppe

Parameter forteller hvilke naturvitenskapelige egenskaper som dataserien har informasjon om. Emneord av denne klassen skal stå i samspill med et oppslagsregister.

Format:

<parametergruppe>

Eksempler:

fysisk/kjemiske

biologisk

bakteriologisk

sedimentologisk

seismiske dybder

Parametergruppen kan utelates. Flere parametergrupper kan angis om ønskelig.

### Emneord forøvrig

Andre emneord som ikke kan klassifiseres som medium, aktivitet eller parameter, men som likevel utdyper referansenes innhold, kan angis.

Eksempelvis kan man her angi målemetode, måleprinsipp, hensikt o.l.

Format:

<emneord>

Eksempler:

- Refleksjonsseismikk
- Drikkevannsundersøkelser

### 5.3.5. Produsent

#### Institusjon

Institusjonens navn gis med dens "offentlige" forkortelse.

Eks.

NGU

NSKV

NGO

Vanligvis vil det være én institusjon pr. referanse, men det skal være mulig å angi flere.

For store institusjoner skal det være mulig å angi underavdelinger.

#### Person

Under produsent angis også hvem som hadde ansvaret for data-serien. Det kunne f.eks. være prosjektleder og/eller saksbehandler. En eller flere personer skal kunne registreres.

### 5.3.6. Arkiv

Arkivfeltene gir opplysninger om hvor dataserien er lagret, informasjonskilden.

#### Institusjon

Institusjonens offentlige forkortelse benyttes. Se pkt. 5.3.5.

#### Arkivbetegnelse

Institusjonens arkiv/database angis med navn. Kontaktperson kan være aktuelt å oppgi, men dette forandres over tid. Derfor kan det være bedre å oppgi ansvarlig avdeling.

#### Arkivnøkkel

Navnet på institusjonens egen arkivnøkkel for dataserier.

#### Lagringsmedium

Hvordan dataserien er lagret, kan angis. Samlet Plan, ref. [17], har spesifisert en inndeling som kan følges.

- EDB-lesbart (magnetbånd, diskette, hullkort)
- Mikrofilm
- Maskintabeller
- Fotografier
- Fjernfoto
- Prøver
- Publikasjoner, rapporter
- Journaler
- Kart

Denne inndelingen bør utvides med databaser som etter hvert utgjør en betydningsfull lagringsform.

#### Tilgjengelighet

Noen dataserier kan ha begrenset tilgjengelighet via avtaler eller lovgivning. Her har også Samlet Plan en oppdeling som er et godt utgangspunkt.

- Normalt INGEN tilgjengelighet for andre brukere

- Ved FORHANDLING/VURDERING i hvert enkelt tilfelle
- VARIERER for deler av dataserien
- Tilgjengelig, men RESTRIKSJONER på bruk
- AVHENGER av arbeidsbehov/ressurssituasjon
- INGEN RESTRIKSJONER

#### 5.4. REFERANSESENTRALENS REFERANSEBASE

Hoveddelene i referansesentralens lagringsopplegg er

- lagring for referanser
- oppslagsregistre
- meldingsformidling

Referansebasen inneholder referanser, og er den som brukerne søker i og som vokser etter hvert som nye referanser meldes inn til systemet.

Oppslagsregistrene og meldingsformidlingen er mer å betrakte som hjelpesystemer, men er svært viktig for at systemet skal fungere.

##### 5.4.1. Referansebasen

For å lagre referansene i referansebasen, vil det være aktuelt å benytte samme teknologi som for lagring av data i databaser. Slike systemer kalles ofte databasesystemer (DBMS). Referansebasen krever et velutviklet databasesystem.

Noen viktige krav er listet opp under. De vil ha konsekvenser for oppbygging av referansebasen og i valg av databasesystem:

- store datamengder skal behandles
- raskt søk og oppslag i referansebasen er nødvendig
- referansene vil ha variabel lengde i antall bokstaver/tall

- variabelt antall opplysninger i hvert felt
- søking på flere nøkler samtidig, der hver referanse kan ha flere søkenøkler pr. felt
- søking sammen med oppslagsregistre.

Kravene er såpass store at et generelt databasesystem synes være det eneste naturlige. Kravene vil også gjøre mange av de som eksisterer i dag uaktuelle.

I tillegg vil følgende forhold tilsi at det er fornuftig å basere seg på et generelt databasesystem:

- slike systemer har innbygd sikkerhetsfunksjoner og vedlikeholdsverktøy
- flere brukere kan benytte slike systemer samtidig
- rapportgeneratorer standard

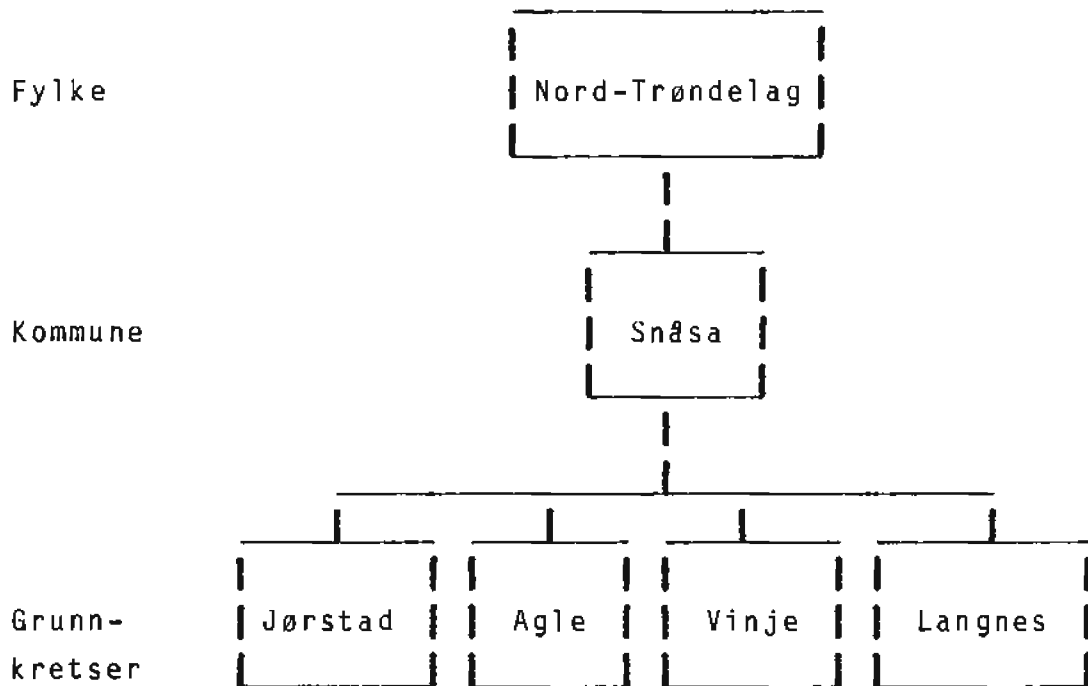
Det kan bli aktuelt å dele opp referansebasen etter karakteristikkk dvs. en for måledata, en for rapporter osv. På nåværende stadium av spesifikasjonene er det ikke nødvendig å bestemme dette.

#### 5.4.2. Oppslagsregister

Et oppslagsregister kan populært defineres som en ordliste med lovlige ord for et bestemt fagområde. I tillegg til å være en ren ordliste, kan oppslagsregisteret inneholde definisjoner og forklaring på hvert ord og dessuten ha innbygd en hierarkisk struktur. Denne informasjonen kan være veldig nyttig ved

- at systemet kan plukke fram synonymer til søkeord eller
- finne søkeord som kan være alternativer ettersom de ligger på et annet nivå i hierarkiet.

Ta som eksempel den administrative inndeling av landet. Da vil kommunen 'Snåsa' ha følgende hierarkiske struktur:



Bakgrunnen for å ha hierarkiske oppslagsregistre i referanse-systemet, er at det er uaktuelt å "dressere" dataprodusentene til å benytte et bestemt nivå f.eks. kommunenavn. Noen vil registrere på et grovere nivå, noen på et mer detaljert nivå.

Når en bruker skal søke i referansebasen, vil han benytte "sitt" nivå.

Fra referansesynspunkt er hovedhensikten med oppslagsregister å kunne plukke fram alternativer til det nøkleordet som brukeren har foreslått.

Eksempel:

Er 'Snåsa' foreslått bør systemet også automatisk ha alternativene 'Nord-Trøndelag', 'Jørstad', 'Agle', 'Vinje' og 'Langnes' klar og at disse kan bli brukt i søkingen hvis brukeren ønsker det.

Oppslagsregistrenes hovedoppgaver er:

- kontrollere om et søkeord er lovlig ved lagring av en ny referanse og hvis ikke assistere i å finne et nytt,
- foreslå alternativer til et søkeord under et søk

Referansesystemet vil ha bruk for oppslagsregistre for følgende områder:

- Administrative stedsnavn
- Faglig og kartografisk inndeling av landet
  - Nedbørfelt
  - Fjorder
  - Kartblad
  - Geologi
  - Biologi
  - osv.
- Medier
  - Luft
  - Vann
  - Mark
  - Biologisk materiale
- Parametergrupper
  - Metoder
  - Taxa (innenfor biologi)
  - Geologiske begrepslister
  - Målestasjoner
  - Institusjoner

Utviklingen av oppslagsregistrene er en oppgave som ikke er spesiell - men nødvendig - for referansesystemet. Kapittel 7 omhandler dette emnet i detalj.

Det å utvikle et komplett sett med oppslagsregistre er en stor oppgave som vil ta lang tid. Men arbeidet er igang og en del registre fins allerede. Når utbyggingen av referansesystemet kommer igang, er det viktig at man raskt skaffer seg oversikt over hvilke registre det haster mest med slik at arbeidet med disse får prioritet. Referansesystemet trenger endel av listene fra starten av, men de behøver ikke å være komplett fra begynnelsen.

De oppslagsregistre som er aktuelle, vil også være nyttig i andre informasjonssystemer og databaser for miljødata. Det synes der-



for naturlig at etableringen skjer med det siktemål at også andre systemer skal ta dem i bruk.

To viktige konsekvenser følger av dette:

- oppslagsregistrene må benytte et lagringsopplegg som kan installeres på flere datamaskintyper
- funksjonene må kodes i et standardisert programspråk.

#### 5.4.3. Meldingsformidling (portnerfunksjon)

Hensikten med meldingsformidling er at brukeren skal kunne formidle sine spørsmål videre for å

- få hjelp av fagekspertter
- detaljinformasjon fra lokale informasjonssentraler

Portnerfunksjonen krever tilgang til to datalagringsmekanismer. De er

- en database med ruting-informasjon, dvs. adresser til de lokale informasjonssentraler eller kontaktpersoner
- en database som lagrer meldinger i "postkasser" som kan tømmes med jevne mellomrom

Når en bruker har søkt i referansesentralens referansebase, kan han velge å gå videre ved å sende meldinger. Ut fra de søkeord som brukeren da benyttet for 'geografi' og 'temaord', vil systemet kunne lage en liste over aktuelle dataprodusenter. Denne listen presenteres for brukeren, og han vil få valget mellom å sende spørsmålene videre til alle, noen utvalgte eller ingen.

## 5.5. REFERANSESENTRALENS SØKEPROGRAM

Når en bruker kopler seg opp til referansesentralens datamaskin, vil det være søkeprogrammet som vil hjelpe brukeren til å få utført ønskede funksjoner.

Søkeprogrammet bør være

- skjermbildeorientert
- menystyrt

For at programmet skal kunne operere på hele skjermbildet må bruker først angi hvilken type dataskjerm som benyttes.

Deretter presenteres brukeren for en hovedmeny som gir adgang til hovedfunksjonene:

```
|           Velkommen til           |
|           referansesystemet       |
|           for                     |
|           stedfestede data        |
|                                     |
| 1. Søking i sentralens referansebase |
| 2. Søking med oppslagsregistre     |
| 3. Formidling av søkeord          |
| 4. Brukers navn og adresse        |
|           Hvilken?                |
|                                     |
```

### 5.5.1. Søking i referansebasen

Søking i referansesentralens referansebase, skjer ved at brukeren får presentert et skjema på skjermen.

Skjemaet inneholder de søkbare feltene som en referanse inneholder.



Utskriften av referansene på skjerm eller printer er i komprimert form etter ISBD(CM)-standard.

Eksempler er vist i vedlegg II.

#### 5.5.2. Søking med oppslagsregister

Søking med oppslagsregister innebærer en utvidet søkemulighet. Velger brukeren å benytte oppslagsregisteret for adm.stedsnavn og benytter et kommunenavn som stikkord, vil søkeprogrammet også godta referanser som inneholder navnet på tilhørende fylke og/eller tilhørende grunnkretser.

Brukeren kan selv bestemme hvilke oppslagsregistre som skal benyttes i søkingen.

#### 5.5.3. Formidling av søkeord

I mange situasjoner vil en bruker være i tvil om han har fått med seg all tilgjengelig informasjon. Følgende situasjoner kan forekomme:

- brukeren er usikker på om han har søkt riktig
- det kan være fornuftig å la en fagperson se på spørsmålstillingen
- brukeren ønsker flere detaljer enn det som referansesentralen kan gi.

En viktig funksjon blir å distribuere stikkordene videre.

Denne funksjonen går ut på å pakke stikkordene inn som melding/brev og påføre adresseinformasjon til mottaker og avsender. Meldingene legges deretter i tilhørende postkasser.

#### 5.5.4. Brukers navn og adresse

I de situasjoner hvor brukeren ønsker informasjon tilsendt eller tilringt, er det selvsagt nødvendig å oppgi:

- navn
- adresse
- telefon

Brukere som søker direkte på skjerm og ikke bruker utskrift eller distribusjonsfunksjonene, behøver ikke å oppgi denne informasjonen.

## 6 BRUK AV KODER

Som allerede nevnt i kapittel 4, vil det for miljødata være aktuelt å angi de fleste kringinformasjonstyper, og også noen måleverdi-typer, ved hjelp av koder.

Bruken av koder for dette formål vil, såfremt den er godt organisert, gi store fordeler i alle datahåndteringsfaser - ved registrering, lagring, utsøking, utveksling, bearbeidelse og presentasjon.

### 6.1 Kodebegrepet

Begrepet "kode" har i seg selv ingen entydig definisjon og brukes for litt forskjellige formål og i litt forskjellige betydninger. For miljødataformål har vi forsøksvis valgt følgende definisjon:

En kode er en kortfattet betegnelse på et vitenskapelig begrep. Koden er gyldig bare innenfor et avgrenset sett av fagområder og et avgrenset sett av institusjoner.

Innenfor disse avgrensninger er koden entydig. D.v.s. at en kode bare kan bety ett bestemt vitenskapelig begrep. Omvendt kan derimot et vitenskapelig begrep være representert av flere alternative koder.

Kodene kan dannes på mange forskjellige måter. Disse mange måter skal ikke gjennomgås her, men et eksempel (tidsangivelse) kan illustrere mulighetene:

#### Tidsangivelse

Et tidspunkt kan angis på mange måter - enda flere enn vist i dette eksemplet finnes i [2].

JUNE, 6TH 1977	06/06/1977
06-06-1977	6 JUNI 1977
6-VI 1977	06.06.1977
06061977	6TH OF JUNE, 1977
77231	77-06-06

Ved manuell databehandling vil man som oftest kunne gjenkjenne og tolke slike angivelser. Når EDB brukes, er det imidlertid nødvendig å deklarerer for systemet hvilken angivelsesform man velger, og derefter bruke denne konsekvent - i hvertfall innen samme programsystem.

I Nordisk Ministerråds forslag til kodestandarder, [3a] har man satset på to hovedvarianter.

TIMMO: tidssone-år-måned-dato-time-minutt  
TIMWE: tidssone-år-uke-ukedag-time-minutt

hvor tidssone varierer med lengdegradssone på jordkloden.

Det tidspunkt da denne tekst ble skrevet, vil med disse to varianter bli angitt slik

TIMMO: A 1983 07 12 12 45  
TIMWE: A 1983 28 2 12 45

Nordisk Ministerråds forslag til tidskoder inneholder mange detaljer i tillegg til dette. En av dem er eksempelvis denne: For ved dataregistrering å slippe å repetere den stabile del av tidsangivelsen (som oftest tidsrom og århundre) bruker man koden TISC. Hvis man i tidseksemplet foran deklarerer TIMMO som hovedkode og overordnet angir TISC=A19, vil man senere kunne angi tidspunktet slik: 83 07 12 12 45. Hvis man istedet setter TISC=A1983, vil tidspunktangivelsen bli kortet ned til 07121245.

Ved utveksling av data skal det - i systemer som er basert på NMR's retningslinjer - sammen med dataene gis informasjon om

hvilke kodestandarder som er benyttet. Denne informasjonen kan gis i korte henvisninger og tabeller, idet enhver institusjon lett vil kunne skaffe seg oversikt over de alternative sett av kodestandarder som fins.

Ytterligere eksempler på kodeopplegg fins senere i rapporten (i avsnitt 6.5.2 angående administrativ arealoppdeling og i avsnitt 6.5.4 angående biologiske taxoner).

Forøvrig henvises det til ref. [3a] hvor hovedprinsippene er mer inngående behandlet.

## 6.2 Situasjonen frem til i dag

De fleste institusjoner benytter seg allerede av kodeteknikk i større eller mindre grad. Stort sett er det i denne forbindelse tale om lokale kodeopplegg som er blitt utviklet ved de enkelte institusjoner. Noen oversikt over hvilke kodesystemer som eksisterer rundt om, eller hvilke erfaringer man har høstet med slike systemer, eksisterer ikke.

Ut i fra forfatterne egne institusjoner - og ut ifra andre som vi kjenner til - antar vi at erfaringene er nokså varierte. Best erfaringer har man for de enkle, lokale systemer som ingen andre har bruk for og som har stabile grunnstrukturer. Og dårligst med dem som også andre får kontakt med og som har en ustabil grunnstruktur. For de systemene som man har hatt vanskeligheter med, er det spesielt et av de to følgende forhold - eller begge - som har vært dominerende:

- o Det ene problemet er at slike opplegg opprinnelig gjerne ble laget uten noen forutgående grundig analyse av struktur og strukturavvik i det fagområde som skulle kodifiseres. Konsekvensen av det ble ofte at kodeopplegget ble for lite fleksibelt, slik at det etterhvert sprakk. Så lappet man på



det ved hjelp av tilleggsregler, slik at det holdt en stund. Så sprakk det kanskje på nytt et annet sted, og nye tilleggsregler måtte til. Og tilslutt ble opplegget så komplisert og inkonsekvent at det ikke var praktisk å bruke det lenger.

- o Det andre problemet er at de fleste av de lokale kodeopplegg som er laget, ikke har gyldighet utenfor egen institusjon. Det innebærer at man, hver gang data skal sendes til andre, enten må bruke de egentlige begreper (som det er meget upraktisk å bruke) eller forklare det benyttede kodeopplegg (som det er vanskelig å forklare p.g.a. alle tilleggsreglene og avvikene). Dessuten må jo mottageren i slike tilfeller lage en oversettelsestabell mellom senderens og egne koder.

Konsekvensen av disse to problemene har ofte vært at kodeoppleggene på den ene side har kostet mye å lage og/eller å vedlikeholde og på den annen side at de har gitt en langt mindre rasjonaliseringsgevinst enn de egentlig kunne ha hatt.

### 6.3 Kravene til fremtidens kodeopplegg

Inntil for få år siden har de fleste institusjonene kunnet greie seg på et vis med sine lokale kodeopplegg og deres svakheter. Men nå er det snart ikke mulig lenger - fordi datasituasjonen har endret seg så sterkt:

- \* Antall aktuelle datatyper har øket sterkt.
- \* Datatypenes gjensidige avhengighetsforhold er blitt langt mer komplisert enn før.
- \* Antall miljødatainstitusjoner har steget mye.
- \* Miljødatainstitusjonene trenger hverandres data mer enn før.

- \* Den samlede mengde av data har øket svært meget.
- \* EDB-alderen har gitt forbedrede måter å utnytte data på - slik at dataene blir mer brukt enn før.

Disse forhold krever at vi må få frem kodeopplegg som ikke har de nevnte svakhetene. Det vil si at

- \* de må være så logiske og fleksible at de ikke sprekker
- \* de må kommunisere med hverandre
- \* det må ikke koste for mye å lage og vedlikeholde dem
- \* det må være enkelt å bruke dem
- \* de må også kunne ta hånd om gamle oppslagsregistre.

Alle enkeltinstitusjoner som har med miljødata å gjøre - som produsenter og/eller brukere - vil i fremtiden bli avhengige av at disse krav blir oppfylt, om de skal kunne utføre sine oppgaver skikkelig. Det nasjonale organ "referansearkivet", som nærværende rapport handler om, vil ikke kunne fungere i det hele tatt uten at disse kravene blir oppfylt.

#### 6.4 Tiltak for å få kravene oppfylt

For å få til en situasjon hvor disse kravene er oppfylt, må det en koordinert innsats til. De sentrale miljømyndigheter som føler ansvaret for dette, anbefales å gå inn for følgende tiltak i denne forbindelse:

- \* Fremskaffe en oversikt over hvilke fagområder innenfor miljøforvaltning og -forskning som trenger kodeopplegg. (En foreløpig oversikt fins i kapittel 7).

- \* Spesifisere de formelle og overordnede faglige krav som skal stilles m.h.t. hvordan slike kodeopplegg skal fungere.
- \* Vurdere hvilke kodeopplegg som bør lages på h.h.v. nasjonalt, regionalt og lokalt nivå. (Et foreløpig forslag som bare omfatter nordiske og nasjonale opplegg, fins i vedlegg III).
- \* La et nasjonalt organ (bestående eller nyopprettet) få tildelt ansvaret for at de avtalte kodeopplegg blir utviklet og senere vedlikeholdt. Dette organ må sørge for å utpeke enkeltinstitusjoner eller grupper av slike som i praksis skal forestå utviklingen og vedlikehold av de enkelte kodeopplegg.
- \* Gi alle miljødatainstitusjoner mulighet til å påvirke de opplegg som lages.

Et slikt sett av tiltak må uten tvil kunne sies å være et godt tilbud til institusjonene:

- a) For alle de institusjoner som enten ikke har noe kodeopplegg fra før, eller ikke er fornøyd med det de har, kan dette sies uten forbehold.
- b) For de institusjoner som allerede har et opplegg som de er fornøyd med, kan dette i mange tilfelle være like godt eller bedre enn det nye de får tilbudt - så lenge man ser bort fra kravet om kommunikasjon med andres opplegg. Og dette spesielle krav om ekstern kommunikasjon kan oppfylles uten at det gamle opplegg skiftes ut (ved hjelp av nøytralkodeprinsippet, som er omtalt nedenfor).
- c) For endel institusjoner kan situasjonen være slik at de allerede har et kodeopplegg som fungerer, men at det ikke er så godt som det nye som tilbys. På den annen side

koster det kanskje endel å legge om. Videre kan det være intern motstand mot å forlate tradisjonen. Og endelig kan det hende at det gamle opplegget er et internasjonalt opplegg som brukes av endel kollegainstitusjoner i andre land og at det derfor er vanskelig å forlate. Også i disse tilfeller kan det være riktig å beholde det gamle opplegg, såfremt man samtidig sørger for kommunikasjon med andre institusjoner i Norge via nøytralkodeprinsippet.

## 6.5 Prinsipper og begreper

### 6.5.1 Nøytralkodeprinsippet

Av foregående avsnitt fremgår det at det neppe vil være mulig å få alle institusjoner til å bruke samme koder for samme begreper. Flere grunner til det er angitt. Det fins en viktig grunn til - som gjelder spesielt for fagområder som inneholder mange enkeltbegreper.

For et kode-opplegg gjelder det absolutt at alle enkeltkoder må være unike (entydige) innenfor dets brukerkrets. Innenfor en liten brukerkrets (som f.eks. en enkelt institusjon) greier man stort sett å få dette til, fordi fagkretsen er snever og antall enkeltbegreper som skal kodes, ikke så stort. Hvis brukerkretsen derimot utvides til f.eks. alle miljødata-institusjoner i hele landet, vil den fagkretsen som skal dekkes, bli meget bredere, og antall aktuelle koder innenfor visse fagområder bli meget stort.

For at kodene skal bli entydige innenfor et kodeopplegg med mange begreper, må kodene bli lange. Og det strider mot kravet om at de skal være enkle å bruke.

For å løse dette problemet - og samtidig gi institusjoner som av andre grunner ønsker å bruke sine lokale kodeopplegg, muligheten

til det - er nøytralkodeprinsippet lansert. Det går ut på følgende:

Når det nasjonale kodeopplegg for et fagområde utvikles, lager man først og fremst en nøytralkode for hvert enkelt begrep. Det er en kode som er helt entydig på landsbasis og innenfor det totale fagfelt som gjelder for miljødata. Denne koden kan gjerne være lang og tungvint i praktisk bruk (fordi den ikke skal brukes av mennesker, men bare av datamaskiner). Som en prøveordning foreslås at et løpenummer på 15 siffer tas i bruk for dette formål.

Den enkelte institusjon får tilsendt (på EDB-medium) en liste over alle aktuelle begreper samt tilhørende nøytralkoder for sine fagområder. Disse nøytralkoder skal brukes når institusjonen skal sende data til andre. Dermed er entydig kommunikasjon mellom institusjonene sikret.

Den enkelte institusjon kan fortsatt bruke sitt eget gamle kodeopplegg i sin interne virksomhet. Eller lage seg et nytt om den ikke har noe eller ikke er fornøyd med det den har. I alle fall må den i sitt datasystem få laget og holdt ajour en oppslagsliste som viser sammenhengen mellom selve begrepet, den ofisielle nøytralkoden og den egne lokalkoden.

Det organ som lager det nasjonale kodeopplegget, skal (i tillegg til nøytralkoden) - for hvert av de fagområder som egner seg for det - lage et sett med mnemotekniske koder som egner seg for praktisk bruk. Dette skal tilbys institusjonene som standardsett. De av institusjonene som ønsker å benytte dette standardsettet istedet for et eget lokalt sett, kan få det tilsendt på magnetbånd.

### 6.5.2 Begrepet "kodeopplegg"

Med begrepet "kodeopplegg" for et fagområde menes det prinsipp og det sett av regler som forteller hvordan visse sider av fagområdets egentlige begrepsapparat kan representeres ved hjelp av koder. Et enkelt eksempel som gjelder den administrative inndeling av Norge skal brukes for å belyse hva dette innebærer:

Norge er administrativt delt opp i fylker, kommuner, bydeler/bygdslag og grunnkretser. Et entydig hierarkisk kodenummersett er etablert som dekker alle enheter:

Enhets- type	Antall siffrer i koden	Eksempel	Enhets- kode	Entydig totalkode
Fylke	2	Oppland	05	05
Kommune	2	Østre Toten	28	05 28
Bydel/Bygdslag	2	Skreia	03	05 28 03
Grunnkrets	2	Hveen	02	05 28 03 02

Med en tallkode kan altså hvilken som helst administrativ enhet i Norge angis entydig.

Prinsippet her er altså at man har fire hierarkiske nivåer, og at man innen hvert nivå angir objektet med to siffrer.

Reglene, som er utarbeidet av Statistisk sentralbyrå og som det fins mange av, går ut på å bestemme etter hvilke kriterier grensene skal trekkes, i hvilken ende nummereringen skal begynne etc. På de høyere nivåer var, i dette eksemplet, de fleste beslutninger av denne art allerede tatt før kodene ble laget. På det laveste nivå (grunnkretser) var det derimot få tradisjoner å ta hensyn til - men til gjengjeld mange praktiske forhold.

### 6.5.3 Utformning av kodeopplegg

Det å utvikle kodeopplegg for miljødataformål er vanligvis ingen enkel oppgave - om de skal oppfylle de krav som er skissert i avsnitt 6.2 ovenfor. De vanskeligste problemene kan samles i to grupper:

For det første fins det en del problemer i tilknytning til de enkelte fagområdenes indre hierarki (struktur) eller navnsystem (nomenklatur). Viktigst blant disse problemene er følgende:

- o Strukturen er ikke komplett og/eller detaljert nok.
- o Avvikende oppfatninger fins m.h.t. hvordan strukturen egentlig er.
- o Strukturen forandrer seg med tiden.
- o Strukturen er inkonsekvent.
- o Nomenklaturen er ikke entydig, idet flere begreper har samme navn.

Noen av disse problemene kan løses ved en begrepsmessig opprydding innenfor fagområdene, men ikke alle. De øvrige må man akseptere og finne kodetekniske løsninger for.

For det annet fins det endel valgmuligheter hvor det kan være vanskelig å forutsi alle de konsekvenser som de valg man gjør kan føre til. Viktige eksempler her er følgende:

- o Skal man bygge opp informasjon inn i kodene? Om man gjør det, blir kodene mer effektive. Men risikoen for at opplegget sprekker blir større.
- o Hvilken kodelengde skal man velge. Korte koder er raskest å bruke. På den annen side er korte koder mnemoteknisk sett dårligere, samtidig som de øker risikoen for at systemet skal sprekke.
- o Skal man velge rene tallkoder, eller tillate også bokstaver (og kanskje også visse spesialtegn). Tallkoder har før vært

brukt mye, fordi de sparer datamaskinkraft i forhold til bokstavkodene. I dag er maskinkraften imidlertid billigere, slik at dette momentet betyr mindre. Bokstavkoder gir store praktiske fordeler for brukeren.

Så langt vi kjenner til, fins det ikke utgitt noen generelle veiledninger eller retningslinjer vedrørende utformning av kodeoppbygg. Et bidrag fra litteraturen fins imidlertid som er et skritt i denne retning. Det er Nordisk Ministerråds rapporter "Karakterisering av Miljødata", [2], og "Handbook for Characterization of Environmental Data (1981) [3]. I den første av disse rapporter fins det en oversikt over og vurdering av endel kodeoppbygg og informasjonssystemer rundt om i verden - samt endel tanker og ideer om hvordan fremtidens opplegg og systemer burde organiseres. I den siste rapporten er endel av disse tanker og ideer konkretisert i form av foreløpige veiledninger og retningslinjer.

Den gruppen som sto bak disse to rapportene hadde også som oppgave å prøve å finne frem til ferdige kodeoppbygg ute i verden som kunne egne seg for nordiske forhold. Den fant ingen slike som i tilstrekkelig grad oppfylte de krav som er skissert i avsnitt 6.3. ovenfor. Derimot kom den over flere ideer og tekniske del-løsninger som det kan være aktuelt å bygge inn i de kommende systemer i de nordiske land.

Gode kodeopplegg er en forutsetning for at den fremtidsskisse som er gitt i avsnitt 4.2 ovenfor, kan bli virkelighet. For å sikre at de kodeopplegg som nå tvinger seg frem, blir gode nok, er det viktig at utviklingen styres og at tilstrekkelige ressurser settes inn. Følgende momenter er viktige i denne forbindelse:

1. Det å utvikle kodeopplegg av noen kompleksitet, er som regel en vanskelig og arbeidskrevende oppgave.
2. På grunn av de store uheldige konsekvenser som dårlige kodeopplegg fører med seg, både på kort og lang sikt, er det



viktig at de man utvikler, gjennomtenkes grundig og lages så logiske og fleksible som mulig. Den ekstrainsats som kreves for å sikre oss mot uheldige løsninger idag, vil raskt tjenes inn ved mer problemfri bruk av dem imorgen.

3. Det fins allerede endel kodeopplegg i Norge (Jfr. vedlegg III tabell 1). Ingen av dem oppfyller fullt ut kravene i avsnitt 6.3. For å få dem til å gjøre det, vil noen kreve bare små justeringer mens andre vil kreve større justeringer eller full omlegging.
4. Både fordi det økonomisk vil gi store besparelser, og fordi det i høy grad vil bedre mulighetene for rasjonell datautveksling, bør alle almene kodeopplegg utvikles i regi av den sentrale miljøforvaltningen.

#### 6.5.4 Koder og oppslagsregistre

Når man skal bruke koder, må man - for å finne ut hvilke begrep en kode står for, eller omvendt hvilken kode som skal brukes for et begrep - ha et oppslagsregister. Dette gjelder når data skal registreres, lagres, søkes ut, sendes, bearbeides eller rapporteres.

Frem til i dag - i "før-EDB-alderen" - har slike oppslagsregistre stort sett vært basert på manuell teknikk i forskjellige varianter.

- a En variant er den hvor antallet av koder er så lite at man husker dem. En skrevet oppslagsliste fins som regel i en skuff et sted, men man trenger den ikke til daglig. Man har en kopi i hodet.
- b En annen variant er den hvor man - ut fra mnemotekniske eller andre logiske regler - kan lage en kode på sparket uten å slå den opp i noen liste. Også her fins det - i en skuff et sted

- gjerne en skrevet liste over kodene og/eller reglene. Men man trenger den ikke til daglig.

- c En tredje variant er den hvor det fins en manuell skrevet oppslagsliste i daglig bruk. Den kan være sortert på forskjellige måter - ut i fra alfabetet, løpenummer, geografi, hierarkisk begrepsstruktur m.v. (Når det gjelder alfabetet kan sorteringen være gjort ut fra kodene eller selve begrepene). En vesentlig svakhet her er at man som regel ikke vil ha råd til å lage og vedlikeholde mer enn en liste, slik at man må bestemme seg på forhånd for en av de mange sorteringsmåtene.

I EDB-alderen er ingen av disse tre manuelle varianter gode nok lenger. Riktignok kan de fortsatt være brukbare ved registrering og manuell søkning av data. Men når EDB-systemet skal søke, bearbeide og/eller presentere resultater på bakgrunn av data fra dataarkivet, må det ha et eller flere EDB-baserte oppslagsregistre å slå opp i for å finne ut hva kodene derfra står for.

De fremtidige EDB-baserte oppslagsregistrene (d og e) vil bygge på prinsippene i variantene b og c ovenfor:

- d For de fleste av oppslagsregistrene vil prinsippet fra c, direkte oppslag i sorterte lister, være mest aktuell. Men til forskjell fra c vil det nå være praktisk mulig å lage flere sorterte lister fra hvert oppslagsregister, slik at man fra gang til gang kan velge om det skal slås opp etter løpenummer, alfabetet, geografi, hierarki m.v. (Et eksempel fra dagliglivet kan belyse denne mulighet: Pr. i dag vil man i telefonkatalogens persondel bare kunne søke på navn (etter alfabetet). Med en EDB-basert katalog vil man etter valg kunne søke ut fra navn, yrke, adresse eller telefonnummer).

Med et slikt opplegg vil man også ha valget mellom å la EDB-systemet slå opp ved selv å spørre systemet via terminal

eller selv å slå opp i trykte oppslagslister som er produsert av systemet.

- e For noen få oppslagsregistre vil prinsippet fra b ovenfor (hvor man ut fra begrepsnavnet kan se hva koden skal være uten å slå opp i noen liste) være det beste. Det gjelder slike registre som har så mange oppslagsord at selv det å slå opp i sorterte lister vil være tungvint.

Bruk av dette prinsippet har imidlertid en viktig forutsetning: At begrepsområdet har et systematisk og entydig navnsystem - og at dette "navnsystemet" er tilstrekkelig bredt akseptert av brukerinstitutionene.

Dette prinsippet er allerede i praktisk bruk for biologiske taxa ved noen institusjoner i Sverige - og også prøvet i Norge med godt resultat. I Norge vurderer man for tiden å prøve det for de tilsvarende geologiske begreper.

Når man ved et opplegg basert på dette prinsippet skal registrere f.eks et biologisk taxon, skal registrøren, istedet for å se etter koden i en oppslagsbok, lage koden direkte ut fra taxonets latinske navn etter spesielle enkle regler. Reglene varierer litt fra nivå til nivå i det biologiske hierarki. (For artsnivået er regelen f.eks. at man bruker de 4 første bokstavene i navnets slektsdel + mellomrom + de tre første bokstaver i navnets artsdel. (Eksempel: Torsk = *Gadus morhua* får koden GADU MOR). Når koden tastes inn, blir den straks testet m.h.t. gyldighet mot et oppslagsregister. Hvis i orden, svarer systemet med å vise hele begrepet. Hvis ikke i orden, hjelper systemet til med å finne den riktige koden.

I vedlegg IV er det gitt eksempel som viser noe mer detaljert hvordan en slik registrering kan foregå.

Et oppslagsregister skal fylle to funksjoner:

- o For det første skal det (ved forespørsel fra menneske eller maskin) kunne fortelle hvilket begrep en kode står for - eller omvendt fortelle hvilken kode et begrep har:
  - Med utgangspunkt i en kode må spørgeren kunne velge om han bare vil ha det offisielle begrepsnavn som svar eller også synonymer eller trivialnavn.
  - Med utgangspunkt i begrepet må spørgeren kunne få svar, uansett om han bruker det offisielle begrepsnavn, synonym eller et trivialnavn.
  
- o For det annet skal det kunne gi endel tilleggsinformasjoner om de enkelte begrepene. Eksempler på slike kan være et begreps
  - plass i institusjonens standardhierarki for dette fagområdet
  - plass i andre hierarkier som institusjonen har i bruk
  - navn på engelsk, latin og eller andre språk
  - "lovlige" og "ulovlige" synonymer
  - vanlige brukte trivialnavn
  - nøytralkode
  - faglige standardkjennetegn (for biologiske arter f.eks. midlere form og dimensjoner samt vanlige farver).
  - kjemiske formel.

Disse tilleggsinformasjonene skal man kunne få ut ved direkte forespørsel. Dessuten skal de kunne brukes som basis for EDB-produserte resultat-tabeller. (Eksempel: En tabell som

beskriver resultater av et fugletellings-prosjekt, må etter valg kunne genereres med de latinske og/eller de norske fuglenavnene.

Når oppslagsregistre skal utvikles, er det endel generelle krav som man må ta hensyn til:

1. For hierarkiske registre gjelder det at koder må finnes på alle aktuelle nivåer. Videre må det for enhver kode være mulig å få vite hvilken overordnet kode som eier den og hvilke koder den selv eier.
2. For registre hvor kodeprinsipp e brukes (generering ved bruk), må det finnes automatisk varslings om kodekollisjoner og synonym-restriksjoner.
3. For registre hvor kodeprinsipp d brukes (direkte oppslag), må det være mulig å slå opp i alternative lister som tilsammen dekker alle rimelige sorteringskriterier.

I neste kapittel (7) og vedlegg III gis det en foreløpig oversikt om hvilke oppslagsregistre av almen karakter som det er behov for i Norge og hvilke av disse som allerede fins eller er underveis.

## 7 BRUK AV OPPSLAGSREGISTRE

For omtrent alle miljødata-etater og -institusjoner gjelder det at de har behov for ganske mange oppslagsregistre, og at de - kanskje uten at de selv vet det - egentlig bruker oppslagsteknikken nokså mye.

Når det er slik at man kanskje bruker oppslagsteknikken uten å vite det, skyldes det at man i stor grad bruker de to variantene a og b som er beskrevet i avsnitt 9.4.4 ovenfor. (Man husker kodene, eller man lager dem på sparket ut fra regler som man husker. Kodelister og koderegler fins bortgjent i skuffer og er lite formaliserte). Om man istedet bruker variant c (oppslag i manuelle lister), fins listene gjerne mer tilgjengelige (står i hyller), men de er fortsatt lite formaliserte.

Som forklart i samme avsnitt, vil situasjonen imidlertid bli en annen når man tar EDB-teknikken i bruk for alvor. Da vil det jo være naturlig å overlate mange av bearbeidings- og presentasjonsprosedyrene til EDB-systemet. Og for at EDB-systemet skal kunne greie det, må de skrevne listene som i dag fins rundt i skuffene eller hyllene bringes frem, formaliseres og legges inn i systemet.

De oppslagsregistre som miljødata-etatene og -institusjonene trenger kan inndeles i tre kategorier:

1. Generelle registre som trengs av etater/institusjoner over hele landet, og som flere enn ett fagområde har bruk for.
2. Fagspesifikke registre som trengs over hele landet, men ikke utenfor eget fagområde.
3. Generelle og fagspesifikke registre som bare trengs regionalt og/eller lokalt.

Blant disse bør kategori 1, som allerede begrunnet i avsnitt 6.2.2 ovenfor, utvikles og vedlikeholdes i regi av den sentrale

miljøforvaltning. Hva kategori 2 angår, bør den antagelig tas hånd om på nasjonalt nivå av de respektive fagområdenes kontaktorganer (som tildels fins allerede og som bør etableres hvis de ikke fins). For kategori 3 bør ansvaret ligge hos den mest sentrale instans innenfor regionen/institusjonen og det eventuelle fagområdet

I nærværende rapport vil bare kategori 1, som den sentrale miljøforvaltning bør ha ansvaret for, bli omtalt nærmere. Men det presiseres at den sentrale miljøforvaltning også bør føle et visst ansvar for de to andre kategoriene - fordi retningslinjer, rådgivning, system-moduler og koordinering fra sentralt hold også her kan bidra til lavere samlede omkostninger og bedre grunnlag for rasjonell utveksling av data.

I de følgende avsnitt vil det først (i avsnitt 7.1) bli gitt en foreløpig oversikt over hvilke registre som trengs i kategori 1 og hvordan de naturlig kan grupperes. I oversikten gis det også noen eksempler på registre av kategorien 2 og 3 og registre som faller utenom alle tre kategorier. Derneft vil det (i avsnitt 7.2 og vedlegg III) bli gitt en kort faglig/teknisk beskrivelse av de enkelte registergruppene og registrene.

### 7.1 Behovet for generelle oppslagsregistre på nasjonalt nivå

Noen oversikt over hvilke "generelle oppslagsregistre som trengs av etater/institusjoner over hele landet og som flere enn ett fagområde har bruk for", er neppe forsøkt laget tidligere. Den oversikt som presenteres her, er ment som en foreløpig utgave. Den er basert kun på en rask gjennomgåelse av miljødatavirksomheter innen noen få sentrale etater og institusjoner og vil neppe være komplett. Den foreliggende oversikt kan imidlertid antas å være fullstendig nok til å angi behovets størrelsesorden, til å avdekke de viktigste muligheter og problemer som utvikling og bruk av oppslagsregistre innebærer og til å antyde det innsatsnivå som må til for å utvikle de registre man trenger.

Oversikten omfatter 26 registre. Ikke alle disse 26 registrene er imidlertid helt selvstendige, idet mange av dem har sterk tilknytning til og er avhengig av ett eller flere av de andre. Dette innebærer at noen av registrene kunne ha vært slått sammen, mens andre kunne ha vært delt opp videre. Tallet 26 er derfor tilfeldig og kan bli forandret når arbeidet med å utvikle registrene kommer igang.

De 26 registrene er inndelt i 7 grupper. Også denne inndelingen kunne ha vært annerledes - med flere eller færre grupper og annen fordeling mellom gruppene. Hensikten med grupperingen er kun å gjøre det mulig å utnytte felles trekk og å ta hensyn til avhengigheter når registrene skal utvikles. Også den kan bli forandret underveis.

En oversikt over registrene og deres gruppering er vist i tabell 7.1.

I det neste avsnitt (7.2) er hver av de 7 gruppene nærmere omtalt.

I innledningsavsnittet under kapittel 7 ble de oppslagsregistre som miljødataetatene og -institusjonene trenger, inndelt i tre kategorier, hvorav kategori 1 er den som fins i tabell 7.1. Hvordan avgrensningen av kategori 1 er mot kategoriene 2 og 3, vil stort sett fremgå av kategori-definisjonene og innholdet i tabell 7.1. For ytterligere å demonstrere avgrensninger skal det her gis noen eksempler på registre som hører hjemme under kategori 2 og 3.:

Kategori 2 (Fagspesifikke registre som trengs over hele landet, men bare innenfor ett fagområde):

- o Parameter- og metoderegister for ornitologi.
- o Parameter- og metoderegister for avløpsteknikk og vannforsyning



Tabell 7.1 Generelle oppslagsregistre på nasjonalt nivå  
- en foreløpig oversikt.

Registergruppe	Registernavn
Generelle arealregistre	Administrative områder Nedbørfelt og vassdrag Fjorder
Faglige arealregistre av almen interesse	Geologiske enhetsområder Botaniske           " Zoologiske           " Meteorologiske      "
Taxon - og taxonlignende registre	Biologiske Geologiske
Miljøvariabel register	Medier Parametre, 1) Målemetoder, 1) Enheter Statistiske stati
Koderegistre for generell identifikasjon og beskrivelse av data og datafangst	Tidkoder Dyp/høydekoder Koordinatkoder Stasjonsgeometri-koder Prøvetagningsmetoder Prøvebevaringsmetoder Målefartøy-register
Koderegistre for generell beskrivelse av dataserier	Opprinnelig hensikt og representativitet Datatilgjengelighet Generell datakvalitet
Aktivitetsregister	Miljødata etater/-institusjoner Faste målestasjoner for miljødata

1) Biologiske, fysisk-kjemiske, geologiske, meteorologiske m.v.

Kategori 3 (Generelle og fagspesifikke registre som bare trengs regionalt og/eller lokalt):

Forurensningskilder	Alle disse eksemplene er
Utslippssteder for forurensning	aktuelle for fylkes- og
Prosjekter	kommuneforvaltninger,
Målestasjoner	forskningsinstitusjoner
Analysepersonell	m.v., som har med for-
Feltpersonell	urensning av vann og
Kontaktpersoner dataarkiv	luft å gjøre

Videre skal det gis noen eksempler på oppslagsregistre som faller utenfor kategoriene 1,2 og 3 fordi vi regner dem som datakilder og ikke som oppslagsregistre.

Nasjonale	Folketellingsdata Landbrukstellingsdata Arealbruksregistre GAB-registeret (Grunneiendom, Areal, Bygning)
Regionale	Registre over mengder forurensning i utslipp fra avløpssystemer, industrien m.v.

## 7.2 De enkelte registergruppene

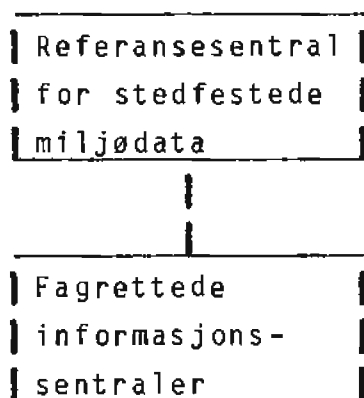
Det har ikke, innenfor rammen av nærværende rapport, vært mulig å gå noe nærmere inn på de enkelte registrene i tabell 7.1 - hverken med hensyn til å definere behovet for dem, å skissere deres struktur og omfang eller å beskrive eventuelle eksisterende utgaver.

Imidlertid har vi allerede visse informasjoner og oppfatninger vedrørende behov, strukturer, omfang og eksisterende utgaver. Selv om disse er ufullstendige og fragmentariske, har vi funnet

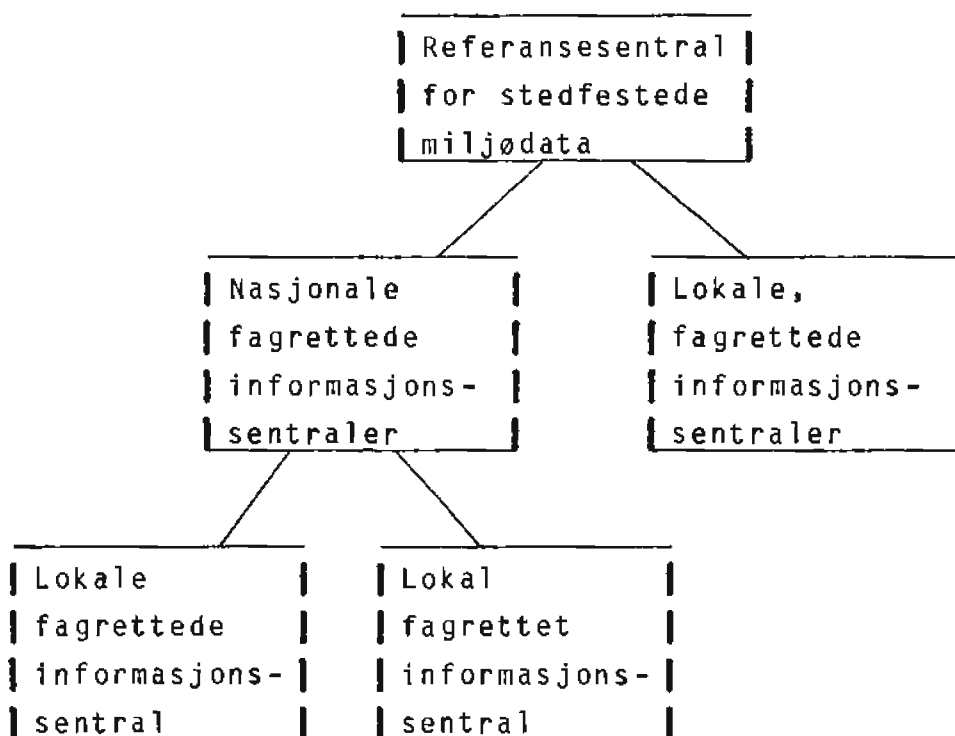
det nyttig å ta dem med som et vedlegg (vedlegg III) til rapporten. Nyttig fordi de, til tross for sin mangelfullhet, likevel antas å gi en brukbar indikasjon på omfanget av og karakteren av den oppgave man her står overfor.

## 8. FAGRETTEDE INFORMASJONSSENTRALER

Kapittel 5 beskriver et informasjonssystem for lagring og framhenting av referanser til stedfestede miljødata. Her tenker man seg en tverrfaglig referansesentral på øverste nivå, med fagrettede informasjonssentraler på neste nivå. Grunnlaget for denne oppdelingen er tatt fra [22].



Denne modellen kan synes svært grov. I virkeligheten har vi en 3-nivås struktur; som vist:



Fagrettede informasjonssentraler er prinsipielt av to typer

- lokale
- nasjonale

Lokale informasjonssentraler lagrer referanser til data som de har produsert selv (f.eks. NIVA, DVF). Nasjonale fagrettede informasjonssentraler (f.eks. NGU) er større faginstitusjoner som lagrer referanser til data som de selv har produsert og referanser som de har fått fra andre dataprodusenter innen samme fagområde.

Referansesentralen er ment å inneholde kopier av referanser fra de fagrettede informasjonssentralene. Noen lokale informasjonssentraler "rapporterer" direkte til referansesentralen, mens andre "rapporterer" via "sin" nasjonale informasjonssentral.

For oppbygging av en informasjonssentral, vil problematikken være svært lik uavhengig om man nå er lokal eller nasjonal. Derfor kaller vi det videre for fagrettede informasjonssentraler.

## 8.1. Generelt om fagrettede referansesystemer

### 8.1.1. Definisjoner

Vi gjentar for ordens skyld definisjonene av noen sentrale begrep som blir brukt i dette hovedavsnittet.

#### Referanse

En referanse er informasjon om data. I vår sammenheng gjelder det stedfestede data. Referansen forteller

"hvem har data om hva fra hvem, hvor og når".

#### Referansearkiv

SSBs rapport "Referansearkiv for naturressurs- og miljødata". Ref. [6], blir vanligvis kalt "referansearkivet". Ved mange faginstitusjoner og i denne sammenheng betegner referansearkiv et kartotek/arkiv med referanser til institusjonens data.

### Referansebase

Når referansen lagres på edb, kaller vi arkivet eller kartoteket en referansebase.

### Referansesystem

Referansebasen, edb-funksjoner og manuelle rutiner er til sammen et referansesystem.

### Referansesentral

Brukes som forkortelse for den nasjonale referansesentralen for miljødata (også foreslått kalt NOREMI) med et tverrfaglig referansesystem.

### Fagrettet informasjonssentral

Synonymt med faginstitutionens referansesystem.

### Database

Når data lagres på edb, kaller vi arkivet en database.

### Databasesystem

Et databasesystem er et EDB-verktøy (programmer) som benyttes til å lagre data i en database. Databasesystemer kan også benyttes til å lagre referanser i en referansebase.

## 8.1.2. Forutsetninger

Vi bygger den videre diskusjon på to viktige forutsetninger.

For det første antar vi at institusjonen allerede har anskaffet en datamaskin med skjermterminaler for bruk til annen databehandling.

For det andre forutsettes at institusjonen har ett eller flere dataarkiver som inneholder de fagdata som de arbeider med. Disse dataarkivene behøver ikke være edb-registre. Manuelle data-

registre som rapportarkiv, kartarkiv, arkiv med prøver osv. er viktige fagarkiver i denne sammenheng.

Selv om vi forutsetter at institusjonene allerede har anskaffet datamaskin, behandles flere problemstillinger som kan være nyttig bakgrunnsmateriale for institusjoner som tenker å anskaffe ny datamaskin eller programsystemer. For å realisere et referansesystem ved en institusjon kan den videre diskusjon nødvendiggjøre kjøp av ytterligere utstyr eller programmer.

### 8.1.3. Dataarkivet skal være absolutt førende

Dataarkivene ved faginstitutionen er selvsagt førende for innholdet i referansesystemet. Referansene skal fortelle HVA dataarkivene inneholder. Informasjonsinnholdet i en referanse vil ofte være endel av opplysningene som ligger i dataarkivet, først og fremst kringinformasjonen, [3]. I tillegg kommer bibliografiske opplysninger (arkivopplysninger) som kan være systematisert manuelt for arkivet, men i mange tilfelle vil slike opplysninger være "lagret" i hodet på de ansvarlige for dataarkivet eller i form av notater o.l.

I dette avsnittet vil vi ikke diskutere i detalj informasjonsinnholdet i referansene. Vi viser til kap. 5 som definerer et minimumsinnhold til referanser og bilag 5 som viser et eksempel på en institusjons referanse. Ut over et minimumsinnhold i referansene, bør det være opp til hver enkelt institusjon å definere referansens innhold.

Hovedkilde for å etablere referansesystemet er selve dataarkivet.

For edb-baserte dataarkiver vil man ofte oppleve at referansene kan genereres automatisk på grunnlag av dataene i dataregisteret, men i de fleste tilfelle vil det være snakk om å registrere informasjonen manuelt ved å gå igjennom dataarkivene.

#### 8.1.4. Funksjonell oversikt

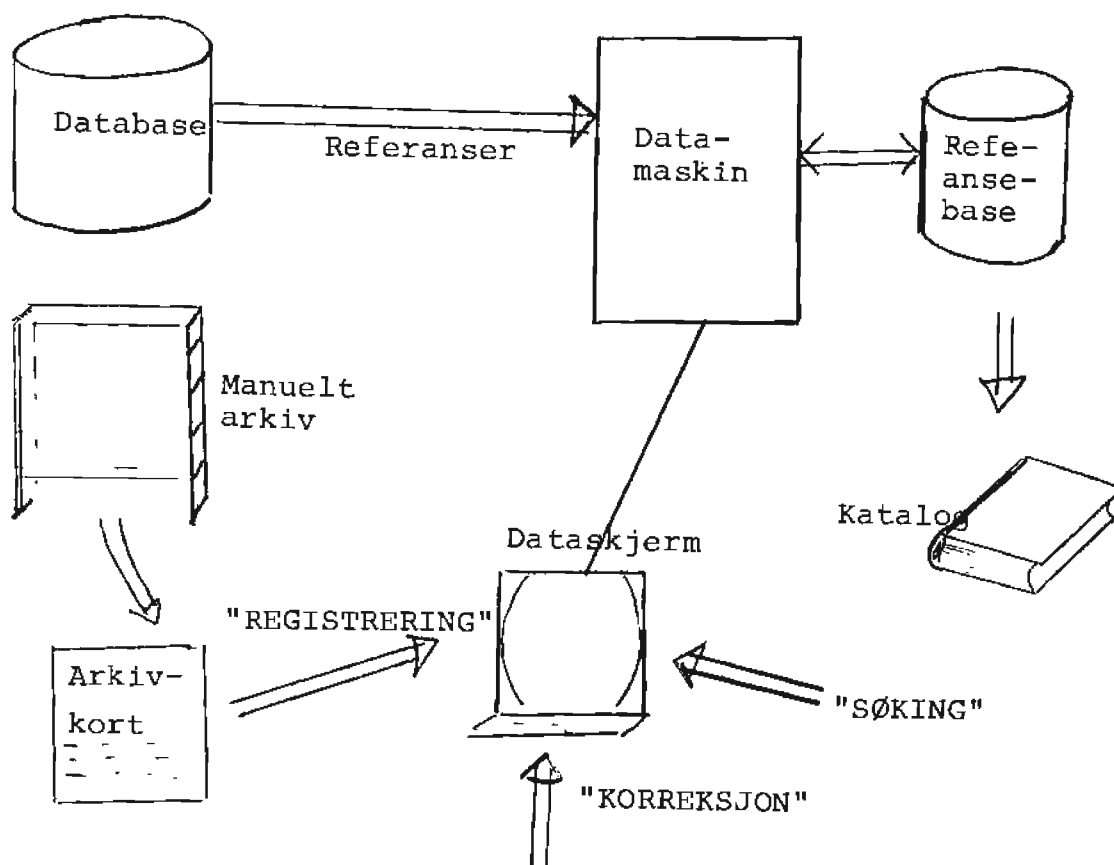


Fig. 8.1.

Fig. 8.1. er en forenklet skisse for hva et edb-basert referansesystem innebærer av manuelle rutiner og utstyr for en institusjon.

Hoveddelene i systemet er dataarkivet og datamaskinen. Omkring disse elementene vil det måtte etableres manuelle rutiner.

##### \* Registrering av referanser

Referansene vil i stor utstrekning måtte registreres manuelt ved at kartotek kort fylles ut på basis av dataarkivene. Deretter blir referansene tastet inn på en dataskjerm og lagt inn i referansebasen. For edb-registre vil referansen kunne etableres mer eller mindre automatisk.



\* Korrektur av referanser

En arbeidskrevende, men nødvendig prosess vil være å kontrollere at inntastede referanser er uten feil. Korreksjon foretas fra dataskjerm.

\* Søking i referansebasen

Hovedhensikten med referansebasen er å kunne søke i den lagrede informasjonen. Denne aktiviteten utføres fra dataskjerm ved at brukeren angir søkekriteriene. Resultatene skal kunne vises på skjerm eller fås utskrevet på papir.

\* Kataloger

Det vil ofte være behov for å ha sorterte utskrifter av referansene i katalogs form. Katalogene brukes i mer rutinemessige oppslag.

\* Kobling til referansesentralen

Referansesystemet bør ha en kobling til utenverden. For større faginstitusjoner vil det bli aktuelt å være tilknyttet det offentlige datanettet slik at brukere utenfra kan søke direkte i referansebasen fra sin dataterminal. I noen tilfelle vil det være aktuelt at referansesentralen også er i direkte kontakt med systemet, enten ved at den overfører meldinger med søkeoppdrag eller ved direkte søking. Dette er omtalt nærmere i kap. 5. For alle fagrettede referansesystemer vil det kreves at referansene skal kunne overføres til referansesentralen. Det kan gjøres via datanettet eller via disketter, magnetbånd o.l.

### 8.1.5. Referansebasen

#### 8.1.5.1. Prinsipiell todeling av referansebasen

La oss først se på lagringen av referansene i datamaskinen. Som nevnt kaller vi lagringssystemet for en referansebase. Logisk

sett bør referansebasen deles i to deler. "Korrekte" referanser er den ene delen og referanser "under bearbeidelse" den andre. Med disse betegnelsene menes at referanser som registreres, først må kontrolleres og godkjennes i en revideringssyklus før de gis status som "korrekte" referanser. Forslag til arbeidsrutine blir beskrevet senere i dette kapittel.

Søking i referansebasen vil bare skje i den "korrekte" delen, og er således referansesystemets "ansikt" utad. Det vil ta seg dårlig ut - og i mange tilfeller føre til misforståelser - å ha skrivefeil og innholdsfeil i referanser som blir søkt ut.

To-delingen er også nødvendig ut fra det at bare "korrekte" referanser skal kunne overføres til den referansesentralen. Etter overføring vil det være vanskeligere å ajourholde/korrigere informasjonsinnholdet.

#### 8.1.5.2. Realisering av referansebasen

Realisering av referansebasen vil i de aller fleste tilfelle kunne løses ved hjelp av ferdige programsystemer, database-systemer. Utvalget varierer imidlertid fra datamaskintype til datamaskintype. Dette vil påvirke valg av ferdigsystem.

Referansesystemet krever effektivitet på to måter:

- (i) - raskt oppslag og lagring av bestemte referanser enkeltvis ved registrering og korreksjon.
- (ii) - rask søking gjennom hele referansebasen ut fra vilkårlige søkeord.

Noen databasesystemer er konstruert slik at begge egenskaper er tatt hånd om, men de fleste er slik at de tilfredsstiller bare det ene kravet.

På edb-teknisk grunnlag kan vi gruppere systemene slik:

- (a) - settorienterte databasesystemer og indeks-sekvensielle filsystem gir rask tilgang på enkelt-referanser, men søketiden på tekstinformasjon vil ofte være stor,
- (b) - databasesystemer bygd på inverterte filer eller mulighet for binærsøkning i tekstinnhold vil være effektive søkesystemer. Oppdatering av data i slike systemer kan være tidkrevende,
- (c) - spesiallagde arkivsystemer for bibliotek o.l. har vanligvis effektive søkemekanismer og også gode oppdateringsfunksjoner.

Systemer av type (a) som er standardverktøy på de fleste datamaskiner, kan i mange tilfelle være egnet til begge oppgaver dersom datamengden ikke blir for stor.

Den prinsipielle todeling av referansebasen kan følgelig skje på to måter:

- fysisk todeling
- logisk todeling

I mange tilfelle vil det være aktuelt at referansesystemet bør basere seg på 2 lagringssystemer,

- ett med "data-entry"egenskaper til registrering, korreksjon og direkteoppslag
- ett til søking.

Da har vi en fysisk deling. Man behøver følgelig programfunksjoner som overfører referansene fra tilstand "under bearbeidelse" til "korrekt". Baseres systemet på ett databasesystem kan referansebasen fremdeles være fysisk todelt. Overføring av referansene vil da være så enkelt som å kopiere referansene. Alternativt kan man legge inn som en informasjon i hver referanse hva som er referansens tilstand. Da har vi en logisk deling. Søkesystemet må da kunne ta hensyn til denne informasjonen, noe som egentlig kompliserer systemet.

### 8.1.6. Registrering av referanser

#### 8.1.6.1. Fra EDB-arkiv til "korrekte" referanser

Vi har tidligere nevnt at dataarkivet er førende for opplysningene i referansesystemet. I mange edb-baserte arkiver vil informasjonen som skal inngå i referansen være lagret sammen med data-seriene, som såkalt kringinformasjon. Resten av referansens innhold slik som lagringsopplysninger o.l. vil selvsagt også være tilgjengelig ettersom arkivet selv kan gi disse opplysningene. Dette gjelder først og fremst de nyeste arkivene som er etablert etter at det er utkommet anbefalinger om hva kringinformasjonene skal bestå av, [3a].

I slike tilfelle vil det være naturlig å ha et eget dataprogram som automatisk genererer referanser ut fra arkivet. Disse referansene er av typen "korrekte" og kan legges direkte inn i den søkbare referansebasen.

Mange vil her spørre seg om det er nødvendig å lagre kringinformasjoner som dubletter i eget referansesystem. I de tilfeller hvor institusjonen har et arkiv og én type referanser på edb kan løsningen være å la søkesystemet arbeide direkte mot dataarkivet. Men kravet til kort svartid vil ved store datamengder fort gjøre denne løsningen uaktuell. Mange institusjoner har ofte flere dataarkiver som er forskjellig oppbygd. Det kan da by på problemer å tilpasse et effektivt søkesystem mot flere slike arkiver. Et annet spørsmål er hvor ønskelig det er at søkesystemet går inn i dataarkivet. Selv om kringinformasjonen er tilgjengelig for søking, kan selve dataene være konfidensielle. En effektiv hindring mot misbruk er å skille referansene og data.

#### 8.1.6.2. Fra EDB-arkiv til "ukorrekte" referanser

Selv der dataserier er lagret på edb, vil det i de aller fleste tilfelle være slik at kringinformasjonen ikke er lagret edb-lesbart eller bare delvis. Annen viktig informasjon som bør være med i referansene foreligger på annen måte. Årsakene til dette er gjerne at man ikke tenkte på dette da systemet ble laget. Andre forhold kan også ha forhindret lagring av kringinformasjon, f.eks. måledata fra instrumenter.

I slike tilfelle kan det være aktuelt å kombinere maskinell og manuell teknikk til å etablere referansene. Først kan man kjøre et program som lager et skjellett av referansen og legger inn den kringinformasjon som er lagret sammen med dataene. Disse må kompletteres manuelt via den ajourholdsrutinen som er beskrevet i neste avsnitt. Referanser av denne typen lagres først i den referansebasen som inneholder referanser "under bearbeidelse".

#### 8.1.6.3. Referanser fra ikke edb-baserte dataarkiv (manuelle arkiv)

En stor del av de dataserier som vil inngå i et referansesystem vil være lagret i manuelle arkiv. F.eks. prøver, litteratur, kart, bilder o.l.

I disse tilfellene må hele referansen registreres manuelt. Dette arbeidet vil vanligvis gå i flere faser.

- Blankettutfylling
- Dataregistrering
- Korrektur

Først bør man utarbeide en blankett eller et skjema (punchegrunnlag) som inneholder de informasjonsfeltene som referansen skal bestå av. Selve utfyllingen av referansene vil ofte skje i

flere runder. Det vil være naturlig å starte med en gjennomgang av dataseriene i arkivet eller dets kartotek hvis det finnes, og fylle ut så mye som mulig. Deretter vil det ofte være nødvendig å komplettere skjemaet med informasjon fra personer som kjenner arkivet godt eller fra fagpersonell som kjenner de enkelte data-seriene. Slik uformell informasjon vil øke referansenes verdi. Referansenes verdi vil også øke dersom man definerte regler for utfylling av bestemte felter. Dette gjelder både skrivemåte og bruk av faglige ord og uttrykk.

Når skjemaene er ferdig utfylt, begynner dataregistreringen på edb-systemet. Dette vil vanligvis foregå fra en dataskjerm. Operatøren får presentert et "tomt" skjema på skjermen og fyller inn opplysningene fra blanketten. Under registrering kan man legge opp til å kontrollere enkelte av opplysningene som tastes inn. Først og fremst gjelder dette felter som kan kontrolleres mot oppslagsregistre (se egen omtale i 8.1.11), men også andre tester kan legges, f.eks. rimelighetstester, verditester o.l.

Hver referanse må få en entydig identifikasjon, f.eks. et løpenr. Dette har man først og fremst bruk for under korreksjonarbeidet.

Alle referanser legges først inn i referansebasen "under bearbeidelse". Når registreringssesjonen avsluttes, startes et eget program som skriver de sist registrerte referanser ut på en skriver sammen med eventuelle feilmeldinger. Dersom man lagrer registreringsdato sammen med referansene, vil denne kunne benyttes til å velge ut de som skal være med på korrekturutskriften.

Det er flere årsaker til at kontroll bør foretas etter registrering av referansene på edb-systemet. Skrivefeil kan ha oppstått under inntastingen. Referansen kan også være feil innholdsmessig. Kontroll på innholdsfeil er lettest å utføre etter edb-registreringen ettersom man kan lage oversiktlige utskrifter.

Korrekturlesing må utføres av fagpersonell og helst av de som har vært opphav til referansene.

Opprettingen av feil foregår på tilsvarende måte som nyregistrering. Referansene eksisterer nå på referansebasen. Operatøren angir referansens identifikasjon (løpenr.) og får fram den tidligere registrerte referansen på skjermen. Man kan nå korrigere enkeltfelter uten å måtte taste inn alt annet på nytt. Fjerning/sletting kan også bli nødvendig dersom operatøren har tastet feil, laget dubletter e.l. Deretter skrives det ut ny kontrolliste og revidering av denne foretas.

#### 8.1.7. Kataloger

Utskrift av referanser på papir eller mikrofilm, kan være aktuelt med jevne mellomrom. I tillegg til at dette er en form for sikkerhetskopi og kan benyttes i nødstilfeller, kan de utskrevne opplysninger redigeres i katalogs form som kan utgis og distribueres til interesserte samarbeidspartnere. Ikke alle slike vil ha tilgang til referansebasen via terminal. Hvor ofte kataloger skal utgis, er avhengig av hvilke aktualitetskrav som eksisterer, hvor omfattende katalogene er, hvor kostbare de er å framstille o.l. Dersom katalogene er av litt omfang, bør de utstyres med effektive stikkordregistre og annen form for oppslagshjelpemidler.

For produksjon av kataloger vil det høyst sannsynlig bli behov for sorteringsprogrammer. En rapportgenerator vil i denne sammenheng være et nyttig hjelpemiddel. Statistikk-/beregningsprogrammer kan også trenge.

Kravet til lagringsform for kataloginnholdet bør være slik at utskriften må kunne viderebehandles av et tekstbehandlingssystem eller kunne overføres til fotosetting.

#### 8.1.8. Overføring av referanser til referansesentralen

Referansesentralen er beskrevet i kap. 5.

Referansesentralen skal inneholde kopier av referansene som ligger i de fagrettede informasjonssentraler - og skal dessuten lagre og ajourholde referanser for institusjoner som ikke har eget edb-basert referansesystem. Referansenes informasjonssinnhold i referansesentralen kan være mindre omfattende enn ved institusjonenes referansearkiv. Siden de lokale referansesystemer er under kontinuerlig ajourhold, må referansesentralens database få overført endringer. Det synes mest hensiktsmessig å foreta oppdateringene med jevne mellomrom.

En måte å gjøre oppdateringen på, er å sende alle "korrekte" referanser til referansesentralen og la ajourføringsprogrammet der avgjøre hva som skal legges inn i sentralens referansebase. F.eks. kan alle gamle referanser som tilhører institusjonen slettes og alt legges inn på nytt. Dette er en enkel metode, men kan virke lite effektiv, spesielt dersom det er store mengder referanser som overføres.

En annen måte er å benytte referansenes løpenummer, eller dato for siste oppdatering. Alle referanser med løpenummer (eller dato) større enn løpenummeret (eller dato) ved siste overføring, overføres. Ved innlegging av referanser må programmet på referansesentralen kunne skille om det er en ny referanse eller en allerede eksisterende referanse. Nye legges inn direkte, mens i det andre tilfellet skal oppdateringen erstatte den gamle referansen.

Overføringsprogrammet må kunne justere referansene til et format som referansesentralen kan tolke.

Overføring til referansesentralen tenkes foregå ved hjelp av edb-tekniske media. I dagens teknologi vil telelinje, datanett, diskett, magnetbånd være aktuelle måter.



### 8.1.9. Søking i referansebasen

Søking er den funksjonen som brukes for å bringe informasjon fram til brukeren på grunnlag av hans spørsmål.

I foregående punkt ble det nevnt at fysisk to-delning kunne være begrunnet med ønske om effektiv søking. Et eget databasesystem spesielt egnet for slikt kan være nødvendig. I et ideelt søkesystem bør alle ord i referansen være søkebegrep. Systemet bør kunne besvare uforutsatte spørsmål, slik at kombinasjoner av søkeord må være mulig å gi.

I praksis vil en finne forskjellige løsninger av disse funksjoner i søkesystemene. Vanligvis kan man grovt sett skille mellom følgende typer søkesystemer:

#### \* Query-språk

Query-språk er knyttet til databasesystemer som kjenner oppbyggingen av enkeltfeltene. Hver opplysning (variabel) er navngitt. Søkingen vil bestå i å angi variabler i kombinasjon med ønsket innhold.

Eksempler:

```
FIND KOMM IS 'SNÅSA' AND MEDIUM IS 'LØSMASSER'  
FIND (FYLKE IS 'NORD-TRØNDELAGE' AND KOMM NOT 'SNÅSA') AND  
MEDIUM IS 'LØSMASSER'
```

Her er KOMM, FYLKE og MEDIUM variabelnavn, men IS, AND, NOT, OR angir logiske relasjoner for utsøkingen.

I første eksempel søkes fram alle referanser som inneholder 'Snåsa' som kommunenavn og har 'Løsmasser' som emneordet MEDIUM. I det andre eksemplet søkes ut de referanser som ikke tilhører 'Snåsa', men likevel ligger i Nord-Trøndelag.

Slike søkesystemer betinger at brukeren også kjenner strukturen (variabelnavnene) i referansen. I tillegg kan ikke slike søk

få med seg de referanser som har søkeordet (f.eks. 'Snåsa') i et annet felt (f.eks. tittelfeltet).

\* Fritekst søking

Denne metoden ligner Query-språk, men man skiller ikke mellom hver enkel opplysning i og med at man ser på referansen som en mengde med ord, og søker fritt i denne mengden uten å tenke om ordet hører til en bestemt variabel.

Eksempel:

'SNÅSA' AND 'LØSMASSER'?

'NORD-TRØNDELAG' NOT 'SNÅSA' AND 'LØSMASSER'?

Tilsvare eksemplet for Query-språk.

Fritekst søking kan medføre at man kommer opp i tolkningskonflikter. F.eks. GRONG er en kommune, men innen geologi er GRONG-feltet et område som ikke følger kommunegrensene. Men slike konflikter vil i praksis ikke ha store ulemper.

\* Søking ved eksempel

Søking ved eksempel fungerer slik at brukeren blir presentert for et 'tomt' skjema med plass for alle de opplysninger som referansen inneholder. Han kan fylle inn med søkeord i noen av feltene. Systemet vil da søke seg fram til referanser som mest mulig tilfredsstillende eksemplet.

Søkingen vil vanlig utføres fra en dataskjerm, og et søk resulterer i første omgang i en melding som forteller hvor mange referanser som tilfredsstillende søkekriteriene.

Deretter vil søkesystemet ha funksjoner som gjør det mulig å 'bla' igjennom det som er søkt ut, eller få dem utskrevet på en skriver.

Eksterne brukere, det vil si personer (eller datasystemer) utenom institusjonen, bør tillates å kople seg til referansesystemet fra sine egne dataskjermer. Om disse skal få tilgang til alle

referanser eller bare utvalgte, er opp til institusjonen å bestemme. Det er mulig å la datasystemet betjene flere kategorier brukere. Spørsmålet om informasjon skal være tilgjengelig for alle, er det ikke mulig å gi noe generelt svar på. Hver institusjon må her komme fram til et eget opplegg. Det samme gjelder spørsmålet om informasjonstjenestene skal koste noe - og i tilfelle hvor mye.

Referanser ("informasjon om data") vil vanligvis ikke inneholde opplysninger som er hemmelige eller krever beskyttelse selv om de refererte data er det. I slike tilfeller bør referansen gi opplysning om dette.

#### 8.1.10. Søking fra referansesentralen

Referansesentralen kan formidle forespørsler til de fagrettede referansesystemer dersom den selv ikke kan gi de nødvendige opplysninger.

Den tekniske løsningen vil avhenge av tilbudet på referansesentralen, men også på den teknologiske utviklingen som forventes på datakommunikasjonssiden.

Men innenfor dagens teknologi kan følgende løsninger være aktuelle

##### \* Transparent forbindelse

En person som først har koplet seg til referansesentralen kan bli automatisk koplet opp mot et fagrettet referansesystem. Datakommunikasjonen går via referansesentralen. Et eget kommunikasjonsprogram formidler data mellom brukerterminal og lokalt referansesystem på en slik måte at det ser ut for brukeren som om han er koplet til direkte. For det lokale referansesystemet ved institusjonen, krever ikke denne løsningen mer enn hva som kreves ved en vanlig ekstern terminaltilkopling.

\* Postkasse

En bruker kan legge igjen meldinger med søkeord i en elektronisk 'postkasse' på referansesentralen.

Lokalt ved institusjonens referansesystem må det være en rutine å kople seg til referansesentralen over datanettet og tømme 'postkassen'. Behandlingen av forespørselen kan gjøres på to måter

- Manuelt søk,
- Automatisk søk.

Med manuelt søk tenker vi oss her at oppstartingen av søket gjøres av en operatør ved institusjonen.

Med automatisk søk tenkes at det eksisterer et program på lokalanlegget som kan tolke meldingen fra referansesentralen og sette opp spørsmålene til søkesystemet automatisk. For å løse denne funksjonen må det i de fleste tilfelle utvikles egne programmer på lokalanlegget.

Ved begge typer forespørsler er det naturlig at resultatene fra søket, sendes spørsmålstilleren pr. post, evt. pr. telefon. Grunnen til dette er at det vanligvis vil være et lengre tidsrom mellom når spørsmålet ble stilt og når svaret foreligger.

\* Automatisk postkasse

Man kan også tenke seg at det lokale referansesystemet er tilkopledd datanettet kontinuerlig. Forespørsler om søk kan formidles videre av referansesentralen i det øyeblikket den ønsker det. Referansesentralen kopler seg automatisk opp mot det lokale systemet og sender over en melding i et bestemt format. Lokalt kan man behandle forespørselen på 2 måter

- Manuell behandling
- Automatisk behandling

Dersom en operatør må foreta søket manuelt, må meldingen legges i en 'postkasse' lokalt. Spørsmålstilleren vil da ikke få svar før søket utføres. F.eks. pr. post/telefon.

Dersom det lokale referansesystemet har et program som kan tolke meldingen og automatisk starte et søk, kan man godt tenke seg at systemet også genererer svarmeldinger. En slik svarmelding kan f.eks. opplyse om hvor mange referanser som søket resulterte i. Om det er hensiktsmessig å overføre en utskrift av disse referansene, er et spørsmål, men teknisk sett er det mulig.

Et program som håndterer denne form for kommunikasjon vil være omfattende. Pr. i dag vil det være vanskelig å etablere en slik løsning. Teknologien ligger tilrette for det, men kommunikasjonsstandarder på dette nivået mangler.

#### 8.1.11. Bruk av oppslagsregister

Oppslagsregistre er beskrevet i tidligere kapitler (5 og 7) hvor man viser til flere bruksområder. I forbindelse med de fagrettede referansesystemene dekker oppslagsregistrene to hovedfunksjoner:

- Kontrollere begrep som registreres,
- Kunne gi synonymer eller alternative søkeord under et søk.

Ved det lokale referansesystemet vil sannsynligvis kontrollfunksjonen være den viktigste og den enkleste å implementere. Dataregistreringsprogrammene bør være utformet slik at kodene blir kontrollert av oppslagsregistre i registreringsøyeblikket. Når en slik opplysning blir tastet inn, vil programmet undersøke om koden fins i tilhørende oppslagsregister. Hvis den ikke fins, må det gis beskjed til operatøren som får sjansen til å rette det med en gang eller gi en midlertidig kode.

Når det gjelder å benytte oppslagsregistrene sammen med et søkesystem slik at man automatisk fikk bruke synonymer i søkingen, vil mulighetene for dette variere stort fra system til system. Regelen vil nok være at man finner ut at det vil by på store

problemer. Men dersom man bygde ut søkesystemet med automatisk tolkning av meldinger med søkeord, slik som det er beskrevet i kap. 8.1.10, vil det være enkelt å etablere en slik bruk av oppslagsregistrene.

## 8.2. Eksempel på løsning. NGUs referansesystem

Noen av de teoretiske beskrivelsene i denne rapporten bygger på praktiske forsøk. Først og fremst gjelder det oppbyggingen av et fagrettet referansesystem.

Beskrivelsen av NGUs referansesystem er formet som eget notat vedlagt denne rapporten. Vi velger i helhet å vise til vedlegg V for videre lesning.

Oppbygging av referansesystemet ved NGU har gått parallelt med dette prosjektet. Erfaringene derfra har bidratt med å utforme spesifikasjonene samtidig som kravene mot referansesentralen og ekstern bruk av referanser har påvirket oppbyggingen av NGUs system.

NGU har fra slutten av 1983 startet en systematisk registrering av egne referanser i sitt lokale referansearkiv. I tillegg vil dette fungere som fagarkiv ettersom både private bedrifter (f.eks. NOTEBY) og offentlige institusjoner (f.eks. fylkesgeologene) ønsker å lagre sine referanser i NGUs system.

## 8.3. Valg av utstyr og systemer

På bakgrunn av den generelle diskusjon omkring et sentralt referansesystem og fagrettede referansearkiver, er det nå naturlig at potensielle brukere av et referansesystem for sted-festede data, stiller seg følgende spørsmål:

"Hva slags utstyr trenger vi for lagring og framhenting av referanser?"

Noe entydig svar finnes ikke og vil ikke være riktig å gi ettersom vi i utgangspunktet har sagt at hver dataprodusent/bruker må selv ha frihet til å organisere sin databehandling. Samtidig er utviklingen innen datateknologien sterk og omskiftende. Utstyr blir fort gammeldags. De som har ansvar for databehandling eller informasjonsbehandling i en institusjon er ofte i et dilemma ved valg av utstyr. Dette har sin bakgrunn i følgende:

- Å etablere et informasjonssystem tar tid. Menneskene i bedriftene skal også delta i systemet. Manuelle rutiner skal etableres.
- Informasjonssystemet må bygge på et konsept som skal leve i lang tid.
- Hva er riktig utstyr å satse på for en framtidig løsning?
- Kan vi sikre en skrittvis overgang til den framtidige løsningen?
- Og ennå verre - hvordan skal vi finne ut hva som er den riktige løsningen på lang sikt?

Stilt ovenfor dette dilemma, er det mange som blir "sittende på gjerdet" og venter på utstyr som kommer, men som ikke ennå er i praktisk bruk. Når utstyret er kommet, opplever man at den framtidige løsningen man valgte kan erstattes av nye og bedre konsept med nye utstyrsalternativer. Så blir man fortsatt "sittende på gjerdet".

Men når dette er sagt, må vi tilføye at det finnes mange brukbare løsninger for å etablere et referansesystem. Valg av utstyr vil i stor grad være avhengig av

- allerede eksisterende utstyr ved institusjonen,
- datavolum,
- ambisjoner og ressurser.

Men det finnes skjær i sjøen. De mer generelle problemstillinger ble diskutert i kap. 8.1. I dette kapitlet skal vi peke på noen spesielle detaljer som kan påvirke valg av utstyr.

Det er ikke meningen å gi en komplett oversikt over utstyrsalternativer, men vi prøver heller å se på flere typiske situasjoner som vi vil finne ved arbeidsplasser som enten produserer sted-festede data eller har bruk for dem.

### 8.3.1. Ikke noe datautstyr

Hos mange dataprodusenter og brukere av stedfestede data har ikke edb-teknologien gjort sitt inntog ennå. Svært mange av disse har heller ikke ambisjoner, ressurser eller datamengder som forsvarer etablering av egne edb-registre.

For slike dataprodusenter vil den mest naturlige løsningen være å samarbeide med sin sentrale faginstitusjon eller referansesentralen. Referanser fylles ut på skjema og sendes faginstitusjonen eller referansesentralen som lagrer referansene i sitt referansesystem. Dataprodusenten får tilbake korrekturlister og om ønskelig sorterte kataloger. Innenfor f.eks. geologi er dette opplegget allerede aktuelt. Flere fylkesgeologer ønsker referanser lagt inn på NGUs referansearkiv.

Brukere av referanser uten datautstyr må selvsagt gjøre henvendelser til de enkelte fagrettede referansesystemene (eller referansesentralen) på tradisjonell måte - telefon, post o.l.

### 8.3.2. Dataskjermer

Dataskjermer eller terminaler er et mer og mer vanlig utstyr på arbeidsplassene. Også her skjer det en rivende utvikling teknologisk sett. Her er noe av utvalget:



- Svart/hvitt karakterskjermer
- Grafiske skjermer
- Fargeterminaler
- Arbeidsstasjoner
- Bærbare terminaler
- Berøringssensitive terminaler

I forbindelse med referansesystemet stilles det ikke store krav til terminalene. De enkleste karakterskjermer kan benyttes såvel som mer avanserte arbeidsstasjoner. Karakterskjermene er "uintelligente", dvs. det som skrives ut på skjermen er enten det brukeren taster inn eller det som datamaskinen som skjermen står tilknyttet, skriver ut. Arbeidsstasjonene er mer "intelligente" idet det er mulig å utføre funksjoner lokalt på skjermen, f.eks. tekstbehandling, framstilling av grafiske figurer, kommunikasjon med flere datamaskiner samtidig.

I tillegg til dataskjermen, må man ha tilkoplingsmulighet til datamaskin. Har institusjonen referansesystem på egen maskin er det snakk om en ledning, men vil man ha tilgang til referansesentralen eller fagrettede referansesystemer, må man benytte datakommunikasjonsutstyr i tillegg. Datakommunikasjon behandles separat i neste avsnitt, men kort kan vi her si at datakommunikasjonsutstyr får man leid hos Televerket. Man kan velge mellom

- det offentlige datanettet (DATEX) eller
- telenettet (MODEM).

Men det er noen viktige detaljer som man må ta hensyn til ved valg av dataskjerm. I denne sammenheng kan vi skille mellom

- skjermer med asynkron dataoverføring (TTY),
- skjermer med annet enn asynkron dataoverføring.

Dette skillet er framkommet ved at noen leverandører av datamaskiner har valgt å basere seg på en av disse måtene å overføre data mellom dataskjerm og datamaskin.

De fleste institusjoner innen miljødatavirksomheten som i dag har datamaskiner, baserer seg på asynkron dataoverføring. Det er

først og fremst de store datamaskinene (f.eks. IBM) som har den andre formen for overføring.

Dette betyr f.eks. at en IBM3270-skjerm ikke kan brukes mot flere av datamaskinene med miljødata.

Omvendt er det ikke sikkert at en asynkron skjerm uten videre kan kobles mot en stormaskin.

Med de mer "intelligente" skjermene er det mulig å kommunisere på begge måter. Dvs. man kan kjøre mot en maskin over TTY-forbindelse og mot IBM med IBM3270-forbindelse. Imidlertid kreves det da 2 forskjellige datanettabonnementer.

Hovedregelen bør være å velge en skjerm som har asynkron (TTY) dataoverføring.
--

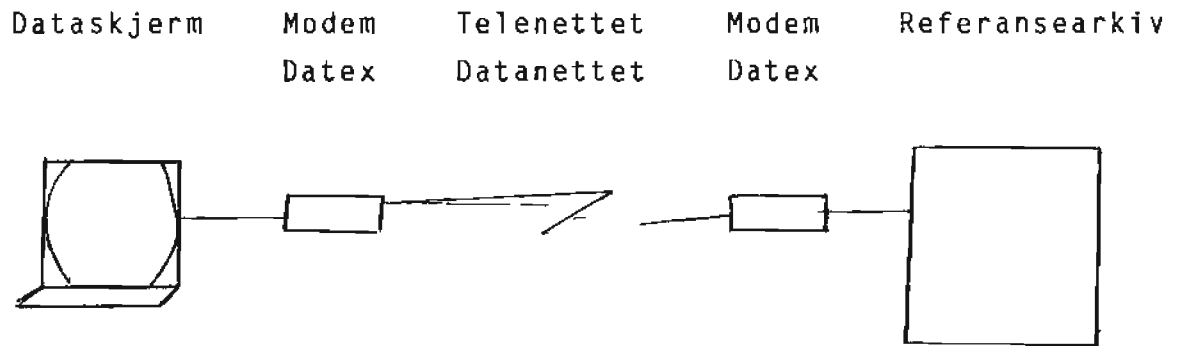
For å kjøre mot de enkelte maskiner på asynkron forbindelse, er det nødvendig at skjermen kan gis noen spesialkarakteristikker. (Hastighet, start/stopp bit, antall bit pr. karakter osv.). De fleste skjermer som i dag er på markedet kan innstilles ved at brukeren selv setter bestemte "brytere" i riktig posisjon. Andre egenskaper til skjermen bør avgjøres på grunnlag av hva man skal bruke skjermen til for øvrig.

### 8.3.3. Datakommunikasjon

Som nevnt er det Televerket som har ansvaret for dataoverføring over større avstander her i landet.

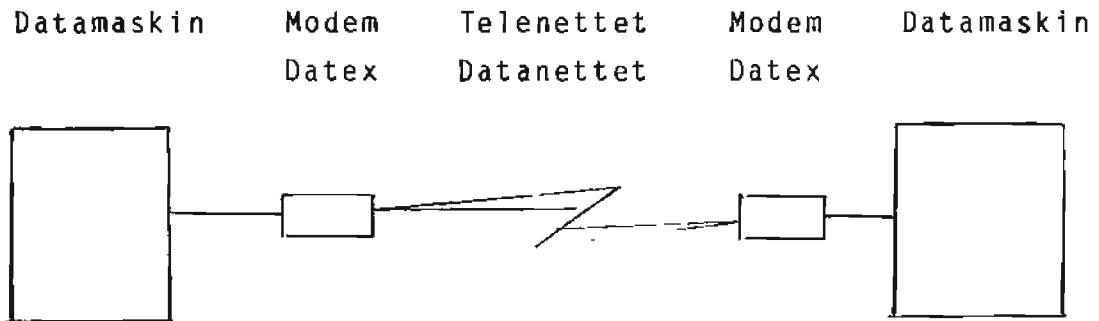
I forbindelse med referansesystemet vil man ha to former for datakommunikasjon

i) Mellom bruker (dataskjerm) og referansesystem (datamaskin)



Eks. Bruker/Referansesentralen  
Bruker/Fagrettet referansearkiv

ii) Mellom to datamaskiner



Eks. Referansesentral/Fagrettet referansearkiv

Hovedregelen for valg av kommunikasjonsutstyr er at spesifikasjonene for utstyret må være likt på begge sider.

Utstyret som kan leies av Televerket er av to hovedtyper:

- Modem
- Datex

Den tradisjonelle måten for datakommunikasjon er over telenettet med MODEM-forbindelse. Det kan igjen være to forskjellige måter:

- Fast telelinje,
- Oppringt samband.

Fast telelinje vil være forholdsvis uaktuelt i sammenheng med referansesystemet p.g.a. at oppkoplingsperiodene vil være forholdsvis kort. Med et oppringt samband benyttes en vanlig telefonlinje sammen med et MODEM, og vel og merke et asynkront modem.

Man kan i dag velge mellom to overføringshastigheter:

300 baud (ca. 30 karakterer i sekundet)  
1200 baud (" 120 " " " ).

For søking i en referansebase er 300 baud tilfredsstillende, idet det er forholdsvis lite informasjon som skal overføres, men 1200 baud er å foretrekke. Dersom en dataprodusent ønsker å registrere referanser over telenettet bør 1200 baud benyttes.

Kostnadene ved modem-leie er en fast pris pr. år pluss antall telleskritt som påløper når datakommunikasjonen er i gang.

I de siste par årene har flere og flere tatt i bruk Televerkets datanett - det offentlige datanettet (DATEX). Datanettet benytter spesielle linjer i dataoverføringen og hensikten med nettet var å skape en sikrere og mer fleksibel datakommunikasjon. Dette er en sannhet med modifikasjoner. Riktignok vil utvalgte linjer i nettet medvirke at dataene blir mindre utsatt for støy, men mellom bruker og nærmeste datanettsentral benyttes fremdeles telelinjer. For bedrifter som har faste, livsviktige dataoverføringer, er datanettet et framskritt ettersom nettet kan finne alternative linjer ved linjebrudd.

Men for en vanlig bruker med dataskjerm har fordelene først og fremst vært mindre støy på dataene og kortere oppkoblingstid, men mulighetene for å kjøre mot forskjellige datamaskiner slik vi ønsker det i referansesammenheng er faktisk blitt mer begrenset.

Dette skyldes at Televerket leverer DATEX i 2 versjoner:

- 10 bits-versjon og
- 11 bits-versjon

i tillegg til å skille mellom asynkron og synkrone brukere.

I praksis betyr det at brukeren må velge en av versjonene når han får installert DATEX og kan da bare kjøre mot datamaskiner som har samme versjon. Man burde ha lov til å kalle dette et tilbakeskritt i forhold til modem-løsningen.

Oppringt DATEX har hastighet på 2400 baud, men utnytter i praksis halvparten, 1200 baud (ca. 120 karakterer i sekundet), når den brukes mot asynkron dataskjerm.

Kostnadene med DATEX beregnes med en fast leiesum pr. år og en sats pr. minutt oppkoblingstid uavhengig av avstand.

Det offentlige datanettet er i dag en linje-svitsjet-nett. Med det menes at brukeren for tildelt en linje hele tiden mens han er oppkoblet.

Neste versjon av det offentlige datanettet blir pakke-svitsjet. Forskjellen er at brukeren bare benytter datanettet den tiden data virkelig overføres. Brukeren er frakoblet nettet mens han utfører lokale operasjoner på skjermen eller venter på data fra datamaskinen. Dette kan bety svært mange opp- og nedkoblinger under en kjøring, noe som brukeren selvsagt ikke vil merke. I et pakke-svitsjet system blir hver enkelt datamelding "pakket inn" med adresseinformasjon (adressat og avsender). Datanettet tar seg av forsendelsen.

Vi ser tydelig med bakgrunn i det vi har diskutert, at et pakke-svitsjet nett vil være den ideelle kommunikasjonsmåten i referansesystemet. Dette gjelder både kommunikasjon mellom referansesentralen og fagrettede referansearkiver, og for brukere som ønsker tilgang til flere datamaskiner.

TELETEX er en ny tjeneste som Televerket tilbyr for overføring av tekster definert av CCITT. TELETEX er den første internasjonalt standardiserte datakommunikasjonsprotokoll som følger ISOs lagdelte modell for åpne systemer. Standardene definerer nøyaktig hvordan all kommunikasjonen skal foregå. TELETEX er først og

fremst ment å erstatte dagens telextrafikk og store deler av forretningskorrespondansen. Etter som dokumentene som overføres er i maskinlesbar form, venter man også at TELETEX skal erstatte mye av den datamaskinkommunikasjon som foregår i dag. Det vil således være aktuelt at både referansesentralen og de fagrettede informasjonsentralene har mulighet å kommunisere over TELETEX.

På brukersiden er TELETEX-terminalene spesialkonstruerte idet de er 'intelligente' terminaler med et program som styrer kommunikasjonen. For eksisterende utstyr må brukeren kjøpe en såkalt adapter som er et spesielt datautstyr som man kopler inn mellom dataskjermene og telenettet.

Teledata er en annen ny teletjeneste som ennå bare er på prøve-stadiet. Systemet er basert på at en skal kunne søke fram informasjon og foreta bestillinger mot datamaskinbaserte informasjonsentraler. Teledata krever egentlig ingen dataskjerm. Fjernsynskjerm med fjernkontrolltastatur skulle være tilstrekkelig.

#### 8.3.4. Mikrodatamaskiner

Mikrodatamaskiner er blitt tatt i bruk i databehandlingen innen miljødatavirksomheten. Det er datamaskiner bygd opp omkring mikroprosessorer (8 bits eller 16 bits), utstyrt med en dataskjerm og datalagringsmedium (vanligvis disketter). Ytelse og kapasitet varierer sterkt på slike maskiner. Nedad er det vanskelig å sette grense mellom mikrodatamaskiner og de mest "intelligente" arbeidsstasjonene. Oppad er det vanskelig å skille de sterkeste mikrodatamaskinene og minimaskiner.

Hvis vi ser bort fra mikromaskinene i bruk som intelligente terminaler, kan en dataproducent bruke dem til å lage egne dataregistre og referansearkiv.

I jungelen av tilbud på mikromarkedet, er det lett å forville seg. Ettersom det er et enormt spenn i ytelse og kapasiteter på slike maskiner, hender det ofte at kjøpere blir et lett bytte for ivrige selgere, og velger feil.

Dersom en ønsker å etablere et dataregister eller et referansearkiv på en mikrodatamaskin, er det etter forfatterens mening følgende man må se etter av spesifikasjoner.

i) Operativsystem.

Maskin må kjøre ett av de "standard" operativsystemene som tilbys.

CP/M, MSDOS, UNIX, CROMIX, XENIX.

Rene Basic-maskiner anbefales det å holde seg borte fra i denne sammenheng

ii) Applikasjongsgenerator.

For å bygge opp et brukervennlig register må man ha et verktøy til å definere registerinnholdet og programmere funksjonene. Å bruke de tradisjonelle verktøyene som Basic, Fortran og delvis Cobol vil kreve uforholdsmessig stor programmeringsinnsats. Heldigvis finnes det mer moderne verktøy som bygger på de ovennevnte operativsystemene, som f.eks. dBaseII, Perfect Filer, MicroRIM o.l.

iii) Kommunikasjon.

Maskinen må kunne kommunisere med andre maskiner som en arbeidsstasjon.

Når det gjelder dataregistre på mikromaskiner, tilsier flere erfaringer at det kan være et vellykket opplegg. Men man må være forberedt på flere ting som man ikke opplever på større data-maskiner.

-> Når datamengdene øker, er det vanlig at datalageret (disketten) blir full. Å begynne å lagre samme type data på flere disketter oppleves etter hvert lite hensiktsmessig

("edb til fots"). Vær forberedt på å kjøpe større lagringsenheter (disker av Winchester-teknologi).

-> På mikrodatamaskiner må brukeren selv ta mer ansvar for sikkerhetskopiering og andre funksjoner som driftsoperatørene tar hånd om på større anlegg. Ting tar tid.

-> Sortering av registre tar lang tid på mikrodatamaskiner. Spesielt med disketter, men selv med disk er det tid.

De samme momentene overfor gjelder også for referansearkiv på mikrodatamaskin.

Kort kan vi si at anvendeligheten er slik:

- For oppbygging av en referansebase for referanser "under bearbeidelse" (se pkt. 8.1.6), vil en mikromaskin være velegnet.
- For relativt lite antall referanser kan man også etablere den "korrekte" referansebasen og søke på referanser. Men søketiden vil fort bli lang, og det vil være begrenset mulighet til søking ettersom man må benytte feltsøking.

Den mest aktuelle løsningen er derfor å vedlikeholde referansebasen "under bearbeidelse" på mikromaskinen og overføre referansene til et større referansearkiv eller referansesentralen slik at de gjøres tilgjengelig for søking via kommunikasjon.

#### 8.3.5. Mycron

Som norskbygd datamaskin er det naturlig at Mycron datamaskiner er å finne i bruk i mange norske bedrifter og institusjoner. Selv om maskinen hører under mikrodatamaskiner, tillater vi oss å gi en liten særømtale. De første 8-bits Mycron-maskiner er blitt



brukt til å lage dataregistre på, men her vil man møte de samme problemer som nevnt under mikrodatamaskiner. En vanlig erfaring [12], er at programmene blir kostbare å utvikle, p.g.a. lite utbygd programverktøy.

Når det gjelder Mycrons 16-bits maskin, er situasjonen annerledes. Her benyttes operativsystemet MP/M-86 som er en videreføring av CP/M. Dermed er det aktuelt å benytte dBaseII som verktøy for etableringen av en referansebase. Men mer aktuelt er Mycrons egen applikasjongsgenerator, INFO-2000. Dette systemet bygger på et filsystem med mulighet for etablering av indeksfiler, dvs. nøkkelfiler. Med INFO-2000 kan man med forholdsvis rimelig innsats lage en referansebase for referanser "under bearbeidelse" der hvert informasjonsfelt er identifiserbart. Noen av disse feltene kan man definere som nøkkelfelt. Systemet vil lage indeksfiler for disse nøklene. Dermed oppnår man rask søking såfremt man begrenser seg til å søke etter informasjon som ligger i nøkkelfeltene. Dette er selvsagt en begrensning, men kan være en tilfredsstillende løsning for mange som dermed kan benytte samme system for referansebase med "korrekte" referanser.

#### 8.3.6. Norsk Data

Norsk Datas datamaskiner blir brukt ved mange institusjoner som arbeider med miljødata. ND-maskinene egner seg både til teknisk og administrativ databehandling, og kan selvsagt brukes til et fagrettet referansearkiv og er også aspirant som maskin i referansesentralen.

På ND-maskinen er det flere mulige programverktøy som er aktuelle å bruke i et referansesystem. Her vil vi nevne noen.

GRAFS eller NOTIS-IR er en videreføring av NOVA\*STATUS som er et fritekst søkesystem. GRAFS [13] inneholder forbedringer mht. oppdatering av dokumenter (referanser) som skjer fra dataskjerm.

I tillegg kan søkingen skje som "søking ved eksempel", omtalt i pkt. 8.1.5. GRAFS eller NOTIR-IR er et ND-produkt.

SIFT (Søking I Fri Tekst) kan betegnes som neste generasjon av NOVA\*STATUS og GRAFS. R-direktoratet har stått for utviklingen av SIFT [14] som er bygd opp fra bunnen av. Erfaringen fra bruk av NOVA\*STATUS har påvirket oppbyggingen. SIFT vil være velegnet for både lagring av referanser og søking. Rasjonaliseringsdirektoratet leverer SIFT.

FICS blir også mye brukt for oppbygging av registre. Alle informasjonsfelt blir entydig identifisert i registret. FICS har gode funksjoner for å lage skjermorienterte oppdateringsfunksjoner, likeledes rapporter (kataloger) sortert på bestemte informasjonsfelt. FICS er derfor velegnet for etablering av referansebase for referanser "under bearbeidelse".

Søking i FICS systemet må basere seg på feltsøking, som selvsagt er en begrensning. For store datamengder vil søketiden bli lang. FICS selges av bedriften Kvam Data A.S.

MIMER er et databasesystem som går på flere datamaskiner, blant dem ND-100. MIMER har gode funksjoner for lagring, sortering og søking. De enkelte informasjonsfelter er entydig identifiserbar og søkingen er feltorientert.

Ved store mengder referanser, er det rimelig å tro at søketiden kan bli lang [15].

MIMER leveres av MIMER DATA SYSTEM AB, Sverige.

POLYDOC er et system som først og fremst er skreddersydd for bibliografiske referanser. For brukere som vil finne en slik standard brukbar er er POLYDOC et godt alternativ. POLYDOC er først og fremst kjent for gode søkemekanismer og sorteringsfunksjoner. Ankepunktet tidligere var lite brukervennlig oppdateringsfunksjoner, men dette er annonsert som bedre i de nyeste versjoner. POLYDOC er også installert på andre maskiner. Det finnes også en versjon for mikromaskiner [16].

### 8.3.7. Andre datamaskiner

Innen de forskjellige virksomheter innen miljødata vil vi finne mange andre minimaskiner og større datamaskiner i bruk. F.eks. VAX, Hewlett Packard, PRIME, Data General, IBM, Honeywell 8ell, Univac o.fl.

Det vil ha liten hensikt å gå i detalj på hva som er aktuelle verktøy for et referansesystem på disse maskinene.

Hver av dem vil ha tilbud som enten dekker

- lagring av referanser "under bearbeidelse"

og/eller

- søking i store informasjonsmengder, "korrekte" referanser.

De maskinavhengige systemene som MIMER, POLYDOC, SIFT o.l. er selvsagt også aktuelle på disse maskintypene.

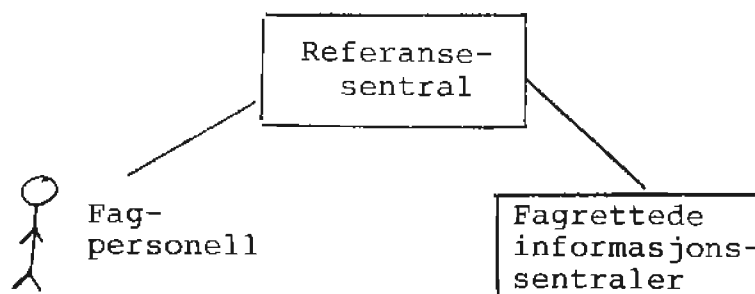
## 9. VIDEREFØRING

De viktigste resultatene av dette arbeidet med et referansesystem for stedfestede data, kan oppsummeres i følgende punkter:

- Forslag til totalsystem for lagring og bruk av referanser.
- Forslag til forbedring av presisjonen i informasjonsinnholdet i referansene ved å realisere oppslagsregistre bygd på nøytralkodeprinsippet.
- Fått erfaring i oppbygging og igangsettelse av en fagrettet informasjonssentral.
- Vise til at realisering av systemet innebærer utvikling av dataprogrammene både på brukernivå og på metodisk nivå, samt utarbeidelse av fagstandarder. Imidlertid forhindrer ikke disse forhold at realiseringen av modellen kan oppstartes med dagens teknologi.

Totalsystemet har tre hovedkomponenter

- referansesentralen
- fagrettede informasjonssentraler
- fagpersonell



Både referansesentralen og de fagrettede informasjonssentralene skal lagre referanser. Referansesentralen skal være hovedinnfallsport til informasjonssystemet fra brukersiden for framhenting av referanser. Men de skal også betjene mindre datapro- duserter med lagring av referanser.

De fagrettede informasjonssentralene finner man først og fremst ved de største dataprodusentene av miljødata, sentralinstituttene. De lagrer referanser til egne data og referanser til data innen samme fag som andre har produsert.

Fagpersonellet skal først og fremst hjelpe brukere av systemet til å hente fram interessante referanser.

Modellen for referansesystemet er beskrevet forholdsvis konseptuelt slik vi mener den kan være holdbar og riktig i lang tid framover selv om den teknologiske utvikling vil bidra med stadig nytt utstyr. Begrunnelsen for dette er at systemet skal inkludere mennesker med tradisjonelle måter å tenke på og gjøre ting på. I tillegg er det mest riktig at det faglige innholdet/bruk skal diktere systemets egenskaper.

#### Hyordan skal man begynne for realisering av modellen?

Ut fra en 'top-down'-analyse er det selvsagt fristende å starte med oppbygging av referansesentralen, men her anbefales først og fremst innsats på følgende tre arbeidsoppgaver parallelt med en forsiktig utvikling av referansesentralen:

- Utvikling av et portabelt programsystem for oppslagsregistre etter nøytralkodeprinsippet.
- Utvikling av en "Håndbok for konstruksjon av fagrettede informasjonssentraler".
- Stimulerende til oppbygging av referansebaser hos dataprodusentene.

Viktigheten med oppslagsregistre er behørig diskutert i kap. 7. Vi har også pekt på at slike oppslagsregistre kan ha kompliserte hierarkiske strukturer. Oppslagsregistrene er nødvendig for både lagring og framhenting av referanser, både ved referansesentralen og ved fagrettede referansearkiver, men selvsagt også ved lagring av miljødata i databaser. Utviklingsinnsatsen av edb-baserte

oppslagsregistre vil følgelig medføre store kostnader. Baseres oppslagsregistrene på nøytralkode-prinsippet, er det sannsynlig at mange forskjellige dataprodusenter kan benytte samme edb-løsning. Det er da et viktig poeng at programmene er portable ettersom det finnes forskjellig edb-utstyr ved de forskjellige dataprodusenter.

En av erfaringene med utviklingen av NGUs referansearkiv var at mye tid og arbeid gikk med til å komme fram til hvilken informasjon som skulle være med i referansen. Dette gjaldt ikke først og fremst det faglige innhold, men mer opplysninger som vil være felles med andre institusjoner og er dermed aktuelt i referansesentralens referanser. Usikkerheten gjaldt først og fremst hvilke opplysninger og skrivemåte.

En håndbok med krav og forslag til etablering av fagrettede informasjonssentraler ville være til stor hjelp for institusjonene ved etablering av referansearkiver. Samtidig vil man minimalisere problemene ved opprettelsen av referansesentralen ved at de fleste dataprodusenter bygger systemene ut fra samme mal.

Et hovedpoeng med fagrettede referansearkiv er at dataprodusentene selv har ansvaret for referansearkivet. Dermed vil de ha egeninteresse av å ajourholde referansene. Et ledd i realiseringen av modellen av referansesystemet vil derfor være å stimulere til etablering av referansearkiv hos dataprodusenter med eget edb-utstyr.

- Sentrale faginstitutioner bør bygge seg opp til å være fagrettede informasjonssentraler basert på edb-teknologi. Referanser lagres, framhentes og distribueres til interesserte.
- Mindre dataprodusenter bør først og fremst etablere referansebase for referanser "under bearbeidelse" (se kap. 8.1). Disse referansene kan med jevne mellomrom overføres til referansesentralen eller fagrettede informasjonssentraler.

Når vi anbefaler at referansesentralen bør bygges opp forsiktig, dvs. over et litt lengre tidsrom, begrunnes dette i følgende forhold:

<-> Som nevnt flere ganger, vil de fagrettede informasjonsentralene være store bidragsyttere med referanser til referansesentralen. Først når flere fagrettede informasjonssentraler er kommet i drift, vil referansesentralen kunne gi tverrfaglig informasjon.

<-> Modellen stiller krav til det offentlige datanettet som i dag ikke tilbys. I enkleste form kan man realisere kommunikasjon, men det blir dyrt. Derfor er det fornuftig å avvente et pakkesvitsjet datanett med en standard dataoverføring.

På den annen side er det viktig at detaljerte systemspesifikasjoner for referansesentralen blir utarbeidet. Det gjelder i første rekke grensesnittene mot de fagrettede informasjonssentralene, både faglig og edb-teknisk. Det vil være mest fornuftig å foreta en total systemspesifikasjon av referansesentralen for derved å klarlegge alle forhold. Noen av spesifikasjonene må innlemmes i den foreslåtte "Håndbok for konstruksjon av fagrettede informasjonssentraler". F.eks.

- > Referansenes innhold,
- > Overføringsformat på referansene,
- > Overføringsformat på søkeforespørsler,
- > Prosedyrer for kommunikasjon og overføring.

Neste fase vil være å bygge en driftsversjon av referansesentralen med lagrings-, framhenting- og distribusjons funksjoner. Et mellomsteg ville være å bygge opp lagrings- og framhentingsfunksjonene først og etablere en referansebase for referanser som ikke dekkes av fagrettede informasjonssentraler.

La oss se litt på kostnadsbildet for realiseringen av referansesentralen. Det vil være investeringer og driftsutgifter. Fordelingen av disse kostnadene vil avhenge av driftsform.

Driften av referansesentralen kan realiseres på to forskjellige måter:

- (i) - drift av referansesentralen utføres ved å etablere en egen driftsorganisasjon.
- (ii) - et kommersielt selskap som driver med edb-basert informasjonstjeneste får ansvar for referansesentralen på oppdragsbasis.

Med egen driftsorganisasjon kan det være aktuelt å benytte eksisterende maskinutstyr eller egen dedisert datamaskin.

Vi forestiller oss at en minimumsbemanning må bestå av 3 personer (2 fagpersonell, 1 operatør).

Ser vi bort fra investeringer i edb-utstyr, vil øvrige edb-utgiftene være strøm, leie telekommunikasjonsutstyr, edb-papir, magnetbånd, service, edb-utstyr. Disse utgiftene vil anslagsvis tilsammen ligge på tilsvarende lønnsutgift for 1 årsverk.

Dersom et kommersielt selskap får som oppdrag å drive referansesentralen vil sannsynligvis driftskostnadene bli noe høyere. Det ligger imidlertid også fordeler innbygd i dette alternativet.

- bedre kontinuitet ved at driftspersonellet er en del av et større miljø
- et profesjonelt informasjonsselskap vil måtte følge med i utstyrsutviklingen og dette vil komme referansesentralen tilgode
- ved en god utnyttelse av utstyret, kan investeringen reduseres.

Investeringene vil omfatte:

- Kjøp av edb-utstyr
- Utvikling av dataprogrammer

Hva som må investeres av edb-utstyr avhenger av driftsform, men grovt sett vil det være behov for



- datamaskin i mini/midi-klassen
- disk-kapasiteten på flere 100 Mbyte
- mulighet for flere terminaler tilknyttet samtidig
- mulighet for flere telekommunikasjonssamband samtidig.

Med dagens priser vil en minimumsløsning nærme seg en investering på 1 mill. kroner, mens en større driftsversjon vil koste mellom 1,5 og 2 mill. kroner. Velger man å benytte en allerede eksisterende stormaskin (f.eks. SDS) vil investeringene bestå i kjøp av disker og telekommunikasjonsutstyr. Men dette vil også fort komme opp i 0,5-1 mill. kroner.

De viktigste programmer som må utvikles er (kap. 5)

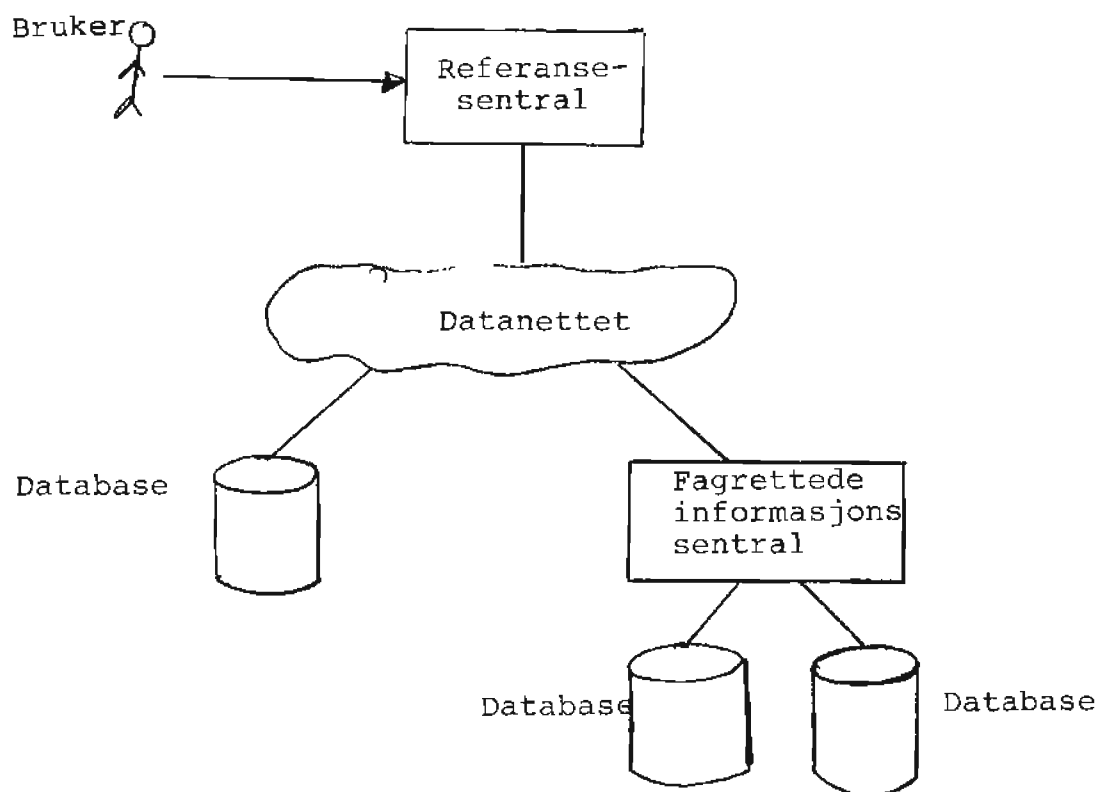
- Etablering av referansebasen
- Lagring av referanser
- Søkeprogram
- Postkasse
- Kommunikasjon på datanettet

Utviklingskostnaden vil avhenge av hvilke utgangspunkt man velger i maskinutstyr, databasesystem og driftsform. Men kravene i kap. 5 er slik at det nødvendigvis må utvikles endel nye programmer. Som et røft anslag ligger det her innsats på 1-3 årsverk.

Senere vil det være behov for vedlikehold av programmene. Det vil gjelde justeringer mot nytt utstyr, mer raffinerte funksjoner, nye behov osv. Det er ikke usannsynlig at det er behov for ca. 0,5 årsverk pr. år til vedlikehold.

Forholdet referanse/data vil etter forfatterens mening bli en interessant problemstilling når modellen etter hvert realiseres. Data som ligger i databaser vil som nevnt i kap. 8, ha nær kontakt med referansene idet de genereres automatisk.

Bildet ser egentlig slik ut:



Noen databaser lager referanser direkte for referansesentralen (små dataprodusenter), mens andre lager først referanser for de fagrettede informasjonsentralene som kopieres over til referansesentralen. Når edb-teknikken muliggjør automatisk trafikk av referanser til referansesentralen, er det ikke noe i veien for at forespørsler fra brukerne kan gå den andre veien helt ned til databasene. Disse problemstillingen må analyseres.

Forslag til videreføring:

- Fase 1:
- Utvikle programsystem for oppslagsregistre
  - "Håndbok for konstruksjon av fagrettede informasjons-sentraler"

- Stimulere dataprodusenter til å etablere referansearkiv og bruke oppslagsregistre i dataregistreringen
- Systemspesifisere referansesentralen i detalj

- Fase 2:
- Etablere referansesentralen med lagrings- og framhentesfunksjoner
  - Ajourholde referanser for dataprodusenter som ikke har tilknytning til fagrettede informasjonssentraler
  - Etablere mottaksapparat for referanser fra fagrettede informasjonssentraler
  - Tilby søking referansebasen

- Fase 3:
- Etablere distribusjon av spørsmål til fagpersonell eller fagrettede informasjonssentraler
  - Analyse av forholdet referanse/data

LITTERATUR-REFERANSER

- [1] Nordisk Ministerråd: Nordisk Miljøprovsbank - et hjelpemiddel for miljø-övervakning och forskning. (NU-B 1977:6).
- [2] Nordisk Ministerråd: Karakterisering av miljødata (NU-B 1978:10)
- [3] Nordisk Ministerråd: Characterization of Environmental Data, Handbook
- (a) Part I Introduction and Element Standardization.  
(b) Part II Data Organization and Exchange.  
(c) Part III Keywords and Codes.
- (Oslo sept. 1981).
- [4] Nordisk Ministerråd: Miljøkvalitetsövervakning i Norden (Oslo mars 1981).
- [5] Nordisk Ministerråd: Former for samordning av miljødatavirk-somhet i Norden (FOSAMINO-prosjektet) (Datert 1982-04-15).
- [6] Statistisk Sentralbyrå: Referansearkiv for naturressurs- og forurensningsdata. Rapport nr. 81/2 (Oslo 1981).
- [7] Norsk Regnesentral: Forslag til utvekslingsformat for dige-tale geodata (SOSI-formatet, versjon 1.0) (Oslo nov. 1980).
- [8] Statens Naturvårdsverk: BIN - Biologiska Inventeringsnormer. F.eks. Daggdjup. (SNU meddelande 1/1979).
- [9] Norges geologiske undersøkelse: Forslag til et samordnet geologisk undersøkelsesprogram for Nord-Trøndelag fylke. NGU-rapp. 1699/2. (Trondheim Feb/82)
- [10] Norges offentlige utredninger: Universitets- og forsknings-bibliotekene i Norge. NOU 1983:17.

- [11] H. Hugdahl: Sentralregister for grunnboringer. NGU. Trondheim (1982).
- [12] H.P. Christiansen: Aspekter ved bruk av dediserte mikrodata-maskiner med skreddersydd diskettbasert "database" til interaktiv administrativ databehandling. Hovedoppgave. Universitetet i Oslo. (Vår 1983).
- [13] Norsk Oata A.S.: GRAFS brukerbeskrivelse. Oslo. (Aug. 1980).
- [14] Statens Rasjonaliseringsdirektorat: SIFT. Searching In Free Text. A Text Retrieval System. Design Specifications. (Sep. 1980).
- [15] UDAC Uppsala Datasenter: MIMER. Concepts and Facilities. (Uppsala Nov/82).
- [16] Norsk Senter for Informatikk, NSI: Polydoc Systemdokumentasjon. Oslo (Nov. 1982).
- [17] Miljøverndepartementet: Veiledningshefte for Samlet Plan for forvaltning av vannressurser. Rapport fra Sekretariatet Samlet Plan. Oslo (mai 1982).
- [18] RUNIT: Bibsys. Brukerhåndbok. Søking. (Trondheim 1980).
- [19] Norsk Dokument Data: EDB-basert litteratursøking og dokumentetterspørse]. DD-rapport 801101. (Oslo 1980).
- [20] NORDFORSK: Strukturering og kodifisering av stedsbundne måledata fra vannmiljø. Rapport fra et nordisk symposium 31. okt. - 2. nov. 1973 (Miljøvårdsekretariatets publikasjon 1974:9)
- [21] Statens dirftssentral for administrativ databehandling: Brukerveiledning NORD-FILCOM. (Oslo juni 1980).

- [22] Miljøverndepartementet: Forslag til utvikling av informasjonssystem for ressurs- og forurensningsregistrering. Rapport 3. (Oslo Aug. 1974).
  
- [23] Tiril Vogt: Erfaringer for arbeidet med oppbygging av et referansearkiv for naturressurs- og foruresningsdata. (Oslo Feb/1977).
  
- [24] J.E. Engebretsen: Description of the UNINETT. Functional Levels. UNINETT report RUNIT. Trondheim (1976).
  
- [25] Tiril Vogt: Forstudie om referansesystem for miljødata. Statistisk Sentralbyrå. Rapport (Oslo 1980).
  
- [26] Norsk institutt for vannforskning: Retningslinjer for et felles taksonomisk basisregister ved NIVA og SFT. Rapport (Oslo Des/82).
  
- [27] D. Belsnes: UNINETT FILE MANAGEMENT SYSTEM. Runit. Trondheim (Jul 1980).

Referansenes innhold

- Referansens identifikasjon
  - Referansetype
  - Referanser
  - Referansenes tittel
  
- Stedfesting
  - Administrativ stedfesting
  - Faglig stedfesting
  - Kartografisk stedfesting
  - Lokal stedsnavn
  
- Tidfesting
  - Tidspunkt (tidsrom)
  
- Emneord/Karakteristikk
  - Type naturressurs (medium)
  - Aktivitet
  - Egenskap (parametergruppe)
  - Øvrige emneord
  
- Produsent/Opphav
  - Institusjon
  - Person
  
- Arkiv
  - Institusjon
  - Arkivets navn
  - Arkivnøkkel
  - Lagringsmedium
  - Tilgjengelighet

Referansenes innhold

Variabel gruppe	Variabler	Format	Søkbart	Antall pr. referanse	Kan utelates	Kommentar
Referanse	Referansenr.	IIII.AA.LLL	Ja	1	Nei	Iflg. SSBs referansearkiv
	Referansetype	Fast antall muligheter	Ja	1	Nei	
	Referansenr. navn/tittel	fritekst	Nei	1	Nei	
Sted-festing	Adm.stedsnavn	AA-----A	Ja	Min. 1 Ubegrenset oppad	Nei	Oppslagsreg.
	Lokalt stedsnavn	AA-----A	Nei		Ja	
	Kartografisk stedsfesting Faglig geografisk navn	FFF.AA--A	Ja	Ubegrenset oppad	Nei	Oppslagsreg. Kart el. XY. Oppslagsreg.
Tidfesting	Tidsrom					
	Fom. (ÅR:MND) Tom. (ÅR:MND)	AAAA:MM AAAA:MM	Ja Ja	1 1	Nei Ja	
Emne	Medium	n(AA----A)	Ja	Min. 1	Nei	Oppslagsreg.
	Aktivitet	n(AA----A)			Nei	
	Parametergruppe	n(AA----A)	Ja	Min. 1	Nei	Oppslagsreg.
	Begrepsliste (taxon) Emneord forøvrig	n(AAA---A)	Nei		Ja	
Opphav	Institusjon	n(AA----A)	Nei		Ja	
	Person	n(AA----A)	Nei	0-	Ja	
Arkiv	Institusjon	AA----A	Nei	1	Nei	
	Arkivbetegn.	AA----A	Nei	1	Nei	
	Arkivnøkkel	AA----A	Nei	1	Nei	
	Lagringsmedium	AA----A	Nei	1	Nei	
	Tilgjengelighet	AA----A	Nei	1	Ja	



1057.01.162

Måledata

Drammen

Løsmasser: Sedimentologi

Seismiske dybder: Refleksjonsseismikk

Refleksjonsseismiske undersøkelser i Drammensfjorden.

NGU/Bjerkli, K. 1982

NGU/Borarkiv/edb; Olsen, H., 375 prøver.

--0--

1064.03.4567

Måledata

Møre og Romsdal

Borgundfjorden

NSV: 140.3.10

Sjøvann: Fys.ocean

Værdata

Borgundfjorden fys.ocean 1967-77.

NIVA/Molvær, Jarle. 1967-1977.

NIVA/Edb 1 NIVA 0-78123 , 5000 data.

--0--

1051.05.1940

Rapport

Sogn og Fjordane

Vik

Myrkdalen

K50: 13164

Geofysikk: Magnetiske borhull

Prospektering

Magnetiske borhullsmålinger ved Framfjord i Sogn

NGU/Brandhaug, K. 1982:11

NGU/Biblioteket; Minken, T., 6 s + 2 kart.

--0--

1051.01.162

Kart

Grong, Snåsa, Namskogan

K100: SANDØLA

Berggrunnsgeologi

Berggrunnsgeologisk kart

NGU/Foslie, S. 1924

NGU/Kartarkiv/Kartfolier; Godø, A., 1 folie

### VEDLEGG III

Foreløpige informasjoner og oppfatninger om aktuelle generelle oppslagsregistre på nasjonalt nivå.

I rapportens kapittel 7 fins det en foreløpig oversikt over aktuelle, generelle oppslagsregistre på nasjonalt nivå (tabell 7.1). Tabellen omfatter 26 registre fordelt på 7 grupper.

I nærværende bilag blir det for hver av de 26 registrene gitt noen informasjoner og oppfatninger vedrørende behov, struktur, omfang og eventuelle eksisterende utgaver. Disse informasjonene er basert kun på de informasjoner og oppfatninger som vi allerede har og ikke på noen egen kartlegging eller vurdering gjennomført i forbindelse med denne rapport.

Hensikten med dette vedlegg er å gi en indikasjon på omfanget og karakteren av den oppgave man står overfor hvis den skisserte samling av oppslagsregistre skal utvikles. Selv om vedleggets innhold er fragmentarisk, upresist og lite gjennomarbeidet, antar vi at det likevel kan gi en brukbar indikasjon.

I vedleggets tabell 1 finner man registrene satt opp i samme gruppering og rekkefølge som i rapportens tabell 7.1. I tabell 1 er det i tillegg laget noen nye rubrikker som forteller om de enkelte registrenes antatte størrelse, struktur, kodeprinsipp, kodetype og eksistens pr. i dag.

Videre inneholder vedlegget kommentarer i tilknytning til denne tabellen - gitt samlet for hver registergruppe.

Tabell 1 - Informasjoner og idéer vedrørende registergrupper og registre.

Register-gruppe	Registrernavn	Register-størrelse (antall koder)	Registerstruktur	Kode-prinsipp 1)	Kodetype 2)	Eksistens - status pr. i dag	
Generelle areal-registre	Administrative områder	Ca. 15 000	Trestruktur (enkel)	d	X(N)	SSB har et internt system	
	Nedbørfelt og vassdrag	> 20 000	Trestruktur (komplisert)	d	X(N)	NVE har et internt system	
	Fjorder	noen tusen ?	Trestruktur (komplisert)	(d)	X	NOD har et grovmasket register	
Faglige areal-registre av almen interesse	Geologiske enhetsområder						
	Botaniske "	?	?	?	?	?	
	Zoologiske "						
	Meteorologiske "						
Taxonregistre	Biologiske	Meget stort	Trestruktur	(e)	X	NIVA + SFT har fremmet forslag til utvikling av et biologisk register	
	Geologiske	Meget stort	Trestruktur	(e)	X		
Miljøvariabel-registre 3)	Medier	< 1 000	Spesielt opplegg - samordnet for alle fem	(d eller e)	X	Intet opplegg fins. Et skisseforslag fins utarbeidet i regi av Nordisk Ministerråd	
	Parametre, 3)	noen tusen ?		(d eller e)	X		
	Målemetoder, 3)	mange tusen ?		(d)	X		
	Enheter	noen hundre ?		(d)	X		
	Statistiske stati	< 100		(d)	X		
Koderegistre for generell identifikasjon og beskrivelse av data og datafangst	Tidkoder	< 20 (?)	Liste	(d)	(X)	For disse fire fins forslag til slike lister. Disse listene er imidlertid ikke komplette	
	Dyp/høydekoder	< 20 (?)	Liste	(d)	(X)		
	Koordinatkoder	< 20 (?)	Liste	(d)	(X)		
	Stasjonsgeometri-koder	< 20 (?)	Liste	(d)	(X)		
	Prøvetagningsmetoder	noen tusen ?	Trestruktur ?	(d)	(X)		Fins for enkelte fagområder ?
	Prøvebevaringsmetoder	noen hundre ?	Trestruktur ?	(d)	(X)		"
	Målefartøy-register	noen tusen ?	Trestruktur ?	(d)	(X)		?
Koderegistre for generell beskrivelse av data-serier	Opprinnelig hensikt og representativitet	< 100 ?	Trestruktur ?	d	X	Intet fins	
	Datatilgjengelighet	< 100 ?	Trestruktur (?)	d	X		
	Generell datakvalitet	< 100 ?	Trestruktur (?)	d	X		
Aktivitets-registre	Miljødata etater/-instit.	noen tusen ?	Nettstruktur	d	X	SSB har et grunnlag	
	Faste målestasjoner for miljødata	noen titusener ?	Liste	d	X	Lokale lister fins	

1) Kodeprinsippene er nærmere forklart i rapportens avsnitt 6.5.4. "d" står for direkte oppslag, "e" står for generering ved bruk. Parentes viser vår anbefaling for systemer som ikke fins i dag.

3) Biologiske, fysisk-kjemiske, geologiske, meteorologiske m.v.

2) Kodetyperne betegnes slik:

N = numerisk (bare sifre)

X = alfanumerisk (sifre, bokstaver og eller spesialtyper).

Angivelse uten parentes viser hva som fins i dag.

Angivelse med parentes viser vår anbefaling.

## 1 Gruppen "Generelle arealregistre"

Det som er felles for registrene i denne gruppen, er at de alle har med arealmessig inndeling av landet å gjøre, og at inndelingene er basert på generelle (faguavhengige) kriterier. Alle er relativt stabile i den forstand at de, når de engang er ferdig laget, ikke vil forandre seg særlig meget senere (bortsett fra at man kanskje vil justere grensene eller findele arealene noe mer etterhvert). De er videre relativt uavhengige av hverandre, bortsett fra at en viss gjensidig grensetilpasning kan være eller bli foretatt.

To av registrene fins i dag

Registeret "Administrative områder" er utviklet av og brukes ved SSB. Den hierarkiske struktur er enkel og har fire nivåer. (Jfr. eksempelet i avsnitt 6.5.2 ovenfor). Systemet vil lett kunne flyttes over til andre maskiner. Kodetypen er nummerisk, har 8 sifre, og inneholder den hierarkiske informasjonen. En overgang til alfanummeriske koder vil medføre visse fordeler for dem som bruker data knyttet til denne strukturen.

Registeret "Nedbørfelt og vassdrag" er under utvikling ved NVE i samråd med en interinstitusjonell styringsgruppe. Systemprinsippene er fastlagt, og alle nedbørfelt ned til et noe varierende detaljeringsnivå er lagt inn. Systemet tillater videre detaljering. Strukturen er komplisert, slik at det skal et relativt avansert databasesystem til for å lage en programpakke som gjør det lett å benytte data knyttet til denne strukturen. Siktemålet er å lage et programsystem som kan distribueres til alle interesserte etater og institusjoner.

Kodetypen, som inneholder den hierarkiske informasjonen, er i prinsippet nummerisk og antas i fremtiden å trenge ca. 20 sifre. I tillegg er visse andre informasjoner bygget inn i kodene ved hjelp av innskutte bokstaver. Av flere grunner antas det at en alfanummerisk struktur her hadde vært langt

mer fordelaktig. En omlegging bør derfor vurderes.

En forløpig arealinndeling av Norges kystfarvann, et "fjordregister", ble foretatt i regi av NOD midt i 70-årene. Den er imidlertid for grovmasket for de krav som nu stilles til en slik inndeling fra miljødatasiden. En annen ulempe ved den er at den ikke korresponderer med inndelingen av "registeret for nedbørfelt og vassdrag" langs kontaktlinjen mellom land og vann.

Det antas derfor at det blir nødvendig å lage et helt nytt fjord-register.

## 2 Gruppen "Faglige arealregistre av almen interesse".

### 3 Gruppen "Taxonregistre"

De to registrene i denne gruppen vil bli de største i registersamlingen. Hvor store de vil bli, er det vanskelig å si, men det er ikke urimelig å vente at "biologiregisteret" etter noen år vil vokse opp til noen hundre tusen koder.

Begge kan i prinsippet settes opp som trestrukturer, men det kommer endel komplikasjoner til:

- o Deler av de to strukturene kan alternativt betraktes som nettverksstrukturer.
- o Hver av strukturene har mange nivåer (biologistrukturene har f.eks ca. 42). Brukerne trenger som regel forenklede strukturer med noen få nivåer og kortslutning av de andre. Men deres enkle strukturønsker er forskjellige, slik at et stort antall varianter må finnes.
- o det råder faglig uenighet om deler av strukturene

- o strukturene ikke er helt stabile (de endres noe av og til)

Intet brukbart system foreligger pr. i dag. Et skisseforslag til biologiske oppslagsregistre er imidlertid laget (ved NIVA i samarbeid med SFT) som ser ut til å kunne oppfylle de krav som stilles. (26). Det er også mulig at dette forslaget kan brukes for et geologisk oppslagsregister.

Det arbeides nå med å finansiere et prosjekt hvor dette forslag kan bli prøvet i praksis.

For disse to oppslagsregistrene vil det å legge inn koder, begreper og tilleggsinformasjoner være en omfattende oppgave. Rundt om i institusjonene fins det imidlertid mange fagavgrensede begrepslister som gir et godt startgrunnlag for dette arbeidet. Bl.a. fins det, ved NIVA og ved noen svenske institusjoner, ferdige kodelister for en del biologigrener som ligner meget på den type som trengs.

#### 4 Gruppen "Miljøvariable"

Denne gruppen avviker på visse måter ganske sterkt fra de andre gruppene og trenger et registeropplegg som er forskjellig fra dem som de andre bruker. Avvikene kan kort beskrives slik:

- o For denne gruppen vil det ikke, som for de andre gruppene, være mulig å lage et ferdig, avsluttet sett av oppslagsregistre som i fremtiden bare trenger noe justering og mindre utvidelser. Alle registrene i denne gruppen vil nemlig i all fremtid stadig måtte ta i bruk nye begreper, koder og strukturer.
- o Registrene i denne gruppe er sterkt knyttet til hverandre på den måte at de sammen eier et underordnet element (som vi her foreløpig kaller "det felles miljøvariabel-element").

Dette siste punktet skal forklares gjennom et eksempelsett som omfatter to institusjoner:

Institusjon A, som har kjemisk vassdragskontroll som oppgave, arbeider med et relativt lite og nokså stabilt utvalg av parametre, medier og metoder. For den er det praktisk å ha oppslagsregisteret for de fem miljøvariable som den trenger i form av en fellestabell:

Koder		Miljøvariable					
		1)		1)			2)Stat.
Lokalkode	Nøytralkode	Parameter	Medium	Metode	Enhet	status	
ORT-P/F/A	123456789012	Ortofosfat	Ferskvann	X12	mg/l	Mom.v.	
ORT-P/F/B	123456789013	"	"	X13	"	"	
ORT-P/G/A	123486780902	"	Grunnvann	X12	mg/l	"	
ORT-P/S/A	123426789012	"	Sediment	X23	mg/100 g	"	
PH/F/A	127356012395	pH	Ferskvann	Y34	-	"	
VF/F/A	261256231842	Vannføring	"	Z45	m /sek	D.middel	
VF/F/B	261256231943	"	"	Z46	m /døgn	A.middel	

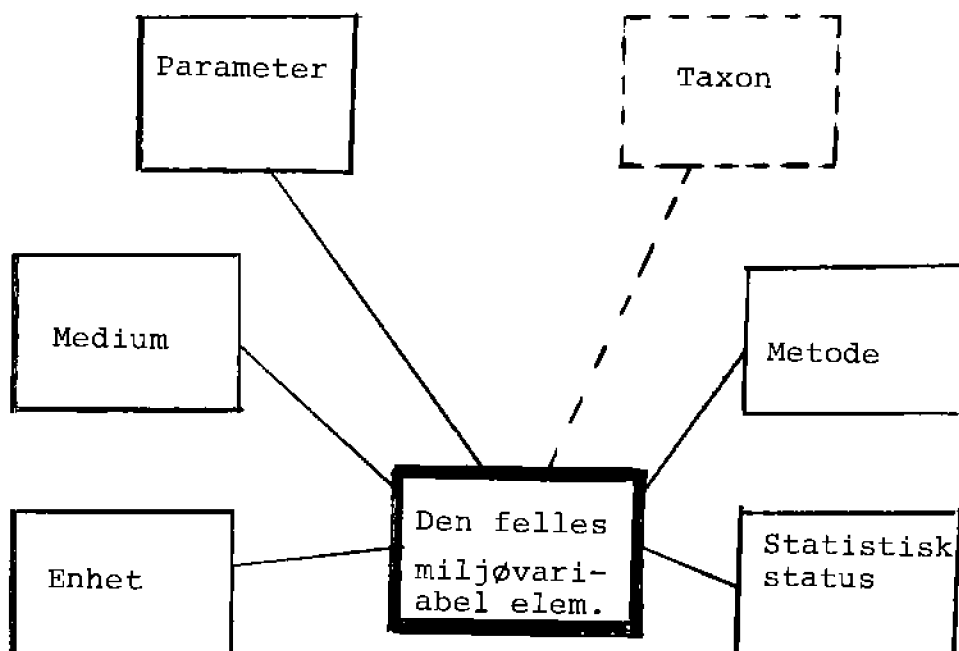
2) Mom. verdi = Momentanverdi  
 D. middel = Døgnmiddel  
 A. middel = Årsmiddel

1) For biologi, fysikk/kjemi, geologi og meteorologi.

Hver linje her representerer en aktuell kombinasjon av de fem miljøvariable. Denne tabellen vil finnes i tre varianter sortert etter h.h.v. lokalkode, paramter og medium. Den vil aldri komme til å omfatte mer enn kanskje 2-3000 linjer pr. variant og vil derfor være grei å slå opp i, både manuelt og ved hjelp av EDB.



Institusjon B har vannforskning som arbeidsoppgave og har et behov for et meget bredt og stadig voksende utvalg av parametre, medier, metoder, enheter og statistiske stati. En tilsvarende oppslagstabell for den ville for hver variant måtte inneholde kanskje 100 000 linjer (kombinasjoner) og være lite hensiktsmessig både å vedlikeholde og å slå opp i. Her er det mer naturlig å velge en databaseløsning av denne type:



I dette systemet vil det finnes et register for hver miljøvariabel (de tynnveggede heltrukne boksene). I hver av dem er hvert begrep (f.eks. Orto-fosfat eller VF i boksen parameter) oppført bare en gang - og det samlede antall oppføringer blir neppe mer enn noen få tusen. I boksen for det felles miljøvariabel-elementet blir så de kombinasjoner av varianter som faktisk er brukt (og bare de) registrert ved hjelp av en enkel teknikk. (Denne boksen inneholder ingen repetisjon av de aktuelle linjene fra de andre boksene - den bare knytter dem sammen

ved hjelp av "adresse-pekere").

I tillegg til de fem miljøvariabel-registrene som fins i eksemplet fra institusjon A er det også aktuelt å knytte taxon registeret til det felles miljøvariabel-element-registeret. (Vist med stipling i skissen)

Dette systemet vil det være raskt å søke i ved hjelp av EDB. For å skaffe seg oversikt (og mulighet for manuell søkning om det ønskes) kan så den enkelte institusjon eller gruppe ta ut en tabell (i alle tre varianter) av den type som institusjon A bruker, men som bare inneholder de grupper av linjer som man selv regner med å få bruk for.

Til ethvert element her, både i de enkelte miljøvariabelboksene og de i boksen for det felles miljøvariabel-element, skal det finnes en unik nøytralkode til bruk ved datautveksling. Dessuten kan det finnes forslag til en mnemoteknisk kode til bruk for dem som ikke foretrekker å lage sine egne.

Denne måten å bygge opp miljøvariabelregisteret på er lansert i ref (3a) (section 2.2.2 and appendix C). Dette opplegget er den eneste vi har sett som er i stand til å tåle stadig nye begreper, koder og strukturer uten å sprekke, og samtidig sikre full kommunikasjon mellom institusjonene.

Av to grunner bør et eventuelt slikt opplegg utvikles i regi av den nasjonale forvaltning (og helst under sterk nordisk samordning): For det første fordi full kommunikasjon ved datautveksling ellers ikke kan oppnås. Og for det annet fordi det her er virkelig store beløp å spare ved en samordnet innsats.

##### 5 Koderegistre for generell identifikasjon og beskrivelse av data og datafangst

De fire første koderegistrene i denne gruppen - "tid", "dyp/høyde", "koordinater", og "stasjonsgeometri" - er meget små, - med kanskje mindre enn 20 koder i hvert. Likevel må de, om man bruker moderne EDB-metoder for datahåndteringen, gis status som oppslagsregistre. I

ref. (3a) del I, fins det foreløpige og ufullstendige kodesett for dem alle.

Registeret for "prøvetagningsmetoder", må derimot bli ganske stort. I Norge fins det ikke noe samlet register av denne type. Derimot fins det antagelig endel bransjevise lister - både for landet som helhet og lokalt ved institusjonene. I Sverige har man et landsomfattende system som brukes av en del institusjoner og innenfor flere fagområder. ("BIN-systemet") Ref ~~8~~ [8]. Under hver oppslagskode i dette registeret fins det fyldig informasjon om metoder presentert på en oversiktlig måte. Dette registeret bør studeres nøye med tanke på et nasjonalt opplegg i Norge (som det haster å få etablert).

Det neste registeret, som gjelder "prøvebevaringsmetoder" er igjen et forholdsvis lite register. Ingen samlet oversikt fins over de prøvebevaringsmetoder som har vært eller er i bruk, og med mulig unntak for noen få bransjer, heller ingen offisielle kodesett. Det er rimelig å anta at et offisielt kodesett som i det minste vil dekke en del av metodene vil vokse frem etterhvert. Mens man venter på det, anbefaler vi at hver institusjon beskriver de metoder som brukes i fritekst, og/eller utarbeider lokale kodesett.

## 6 Koderegistre for generell beskrivelse av dataserier

I KMD-håndboken regner man med tre oppslagsregistre i denne gruppe

"Dataseriens opprinnelige hensikt og representativitet

"Dataseriens datatilgjengelighet

"Dataseriens generelle datakvalitet

For disse tre gjelder noe av det samme som for prøvebevaringsmetoder ovenfor. Ingen offisielle kodesett eksisterer i dag, og man er inntil videre henvist til å bruke fritekst eller lokale koder. Etter en tid vil imidlertid et offisielt kodesett antagelig komme til å vokse frem som vil dekke endel av de begreper som brukes.

## 7 Aktivitetsregistre

Bakgrunnen for å gruppere disse to siste registrene sammen er at de mer enn de andre har direkte tilknytning til de miljødataaktivitetene som foregår.

Gjennom SSB's arbeid med referansearkivet for miljødata (ref.) har vi for første gang fått en noenlunde komplett oversikt over landets miljødata- institusjoner. Dermed er man et godt stykke på vei mot et "oppslagsregister for miljødata-etater og -institusjoner". Det viktigste som må gjøres videre for å nå dit, er å lage et mnemoteknisk basert kodesett og en struktur som gjenspeiler eierforbindelsene - og å legge det hele inn i en fleksibel database. Denne database må være brukervennlig og den må kunne hankses med hyppige endringer p.g.a. nye etableringer, nedleggelse og endrede eierforbindelser.

Det fins i Norge et meget stort antall "målestasjoner for miljødata" som har vært eller er i bruk. Og nye kommer til i økende omfang. Alle disse må registreres på en slik måte at både deres eksistens og alle informasjonene om dem blir tatt vare på og gjort generelt tilgjengelig. Blant dette store antall stasjoner vil det være endel som er av interesse for mange institusjoner og/eller for forskjellige fagområder - mens de øvrige bare vil ha lokal interesse. Av denne grunn synes det hensiktsmessig å registrere stasjonene på to nivåer:

- o Alle stasjoner registreres lokalt av de institusjoner som har ansvaret for dem. Tilsammen vil det her allerede pr. i dag antagelig være tale om flere hundretusener av stasjoner.
- o Stasjoner av sentral interesse registreres også i et "nasjonalt register for målestasjoner". Av denne kategori vil det kanskje pr. i dag finnes noen titusener.

På begge nivåer må registrene være laget slik at det er mulig å finne frem til dem ut i fra forskjellige søkeprofiler.

#### VEDLEGG IV

### Prosedyre ved registrering av biologiske data ved direkte kodegenerering.

---

I rapportens avsnitt 6.5.4 er prinsippet for direkte kodegenerering kort forklart. .

I ~~nærværende~~ bilag skal det gjennom et tenkt eksempel vises hvordan bruk av dette prinsippet vil ta seg ut for en registrør av biologiske data.

Vi tenker oss en registrør som er i ferd med å legge inn fjorddata i institusjonens fysisk-kjemiske-biologiske database. De innledende trinn i prosedyren (a-c nedenfor) er allerede gjennomført.:

- a) Han satte seg ved dataskjermen, presenterte seg og kalte opp databasens innleggingsprogram. Programmet kom opp - og startet med å stille ham to serier av spørsmål.
- b) Som svar på programmets første spørsmålsserie ga han så endel overordnet dataserie-informasjoner:
  - o Prosjektreferanse
  - o Målemedium (i dette tilfelle var svaret "marin fjord")
  - o Lokalitet (fjordens ref.-nr. eller navn)
  - o Datakategori ( i dette tilfelle var svaret "hvirvelløse bunndyr")
  - o Metoder for prøvetagning, observasjon og analyse (i dette tilfelle var svaret "prosjektets vanlige standard-metoder"
  - o Ønskede parametre (i dette tilfelle var svaret at det for hvert taxon observert skulle angis antall individer funnet samt disse individers samlede vekt.
- c) Som svar på programmets annen spørsmålsserie identifiserte han så første prøve
  - o Målestasjonens referanse. (Svar "D4".)
  - o Tidspunkt prøven ble tatt på ("dato 19830618 og klokkeslett 14:20".
  - o Prøvetype (svar: "grabb")
  - o Stasjonens vandyp (målt da prøven ble tatt). (Svar 41 m).

Efter disse tre trinn fikk han følgende bilde på skjermen som sto fast så lenge han holdt på med denne prøven.

<u>Stasjon</u> :D4	<u>Dato</u> 1983-06-18	<u>Kl.</u> 1420
<u>Kategori</u> :	Hvirvelløse bunndyr	
<u>Prøvetype</u> :Grabb	<u>Dyp</u> :	41 m

- d) Under streken spør så systemet om det første taxonet som skal registreres.

<u>Stasjon</u> :D4	<u>Dato</u> 1983-06-18	<u>Kl.</u> 1420
<u>Kategori</u> :	Hvirvelløse bunndyr	
<u>Prøvetype</u> :Grabb		<u>Dyp</u> : 41 m
Taxon:		

- e) Det første taxonet som ønskes registrert er en art som heter Amphitrite cerrates. registrøren anvender den standardiserte forkortningsregelen og skriver koden inn.

<u>Stasjon</u> :D4	<u>Dato</u> 1983-06-18	<u>Kl.</u> 1420
<u>Kategori</u> :	Hvirvelløse bunndyr	
<u>Prøvetype</u> :Grabb		<u>Dyp</u> : 41 m
Taxon: AMPH CER		

- f) Systemet kontrollerer i oppslagsregisteret og finner at koden fins og er entydig. Den forteller dette ved å svare med det fulle navnet. Samtidig ber den om parameterverdiene:

<u>Stasjon</u> :D4	<u>Dato</u> 1983-06-18	<u>Kl.</u> 1420
<u>Kategori</u> :	Hvirvelløse bunndyr	
<u>Prøvetype</u> :Grabb		<u>Dyp</u> : 41 m
AMPH CER= = Amphitrite cerrates		
Antall individer:		
Samlet vekt i gram:		

Brukeren skriver parameterverdiene inn på skjermen. Systemet kvitterer med å skrive nederst at funnet er registrert - og bringer så bilde d) tilbake for å få neste taxon.

- g) Det neste taxon som registrøren skriver inn, arten Amphitritides graciles (Grube), er også kjent av systemet. Men dens "standard-kode" AMPH GRA er ikke entydig. Det fins også en annen art som gir samme forkortelse. Derfor er koden AMPH GRA sperret og to nye koder laget istedet.

I dette tilfelle kvitterer ikke systemet slik som i bilde f, men skriver istedet på skjermen

<u>Stasjon</u> :D4	<u>Dato</u> 1983-06-18	<u>Kl.</u> 1420
<u>Kategori</u> :	Hvirvelløse bunndyr	
<u>Prøvetype</u> :Grabb		<u>Dyp</u> : 41 m

AMPH GRA er flertydig. Kan være:

Amphitritides Gracilis - kode AMPH GR1  
 Amphiascus Graciloides - kode AMPH GR2  
 Angi aktuell kode:

Registrøren ser at den førstnevnte er den riktige - og skriver inn AMPH GR1. Så kommer systemet med bilde av type f - og prosedyren går videre.

- h) Etter at noen flere taxa er registrert på normal måte (sekvensen d, e, f) dukker det opp et nytt spesialtilfelle. Registrøren skal legge inn arten Cardium Edule, og skriver inn koden CARD EDU. Men dette navnet er ikke lenger det offisielt godkjente og har nu status bare som synonym. Systemet svarer slik:

<u>Stasjon</u> :D4	<u>Dato</u> 1983-06-18	<u>Kl.</u> 1420
<u>Kategori</u> :	Hvirvelløse bunndyr	
<u>Prøvetype</u> :Grabb		<u>Dyp</u> : 41 m

Navnet Cardium Edule (kode CARD EDU) er et synonym. Taxonets godkjente kode og navn er:  
 CERA EDU = Cerastoderma Edule  
 antall individer:  
 samlet vekt i gram:

Dette svaret viser at systemet kjenner synonymene, men at det ikke uten videre godtar dem. Hvis en institusjon av en eller annen grunn likevel ønsker å bruke et synonym istedet for det offisielle taxonnavn i sin lokale databehandling, vil det lett kunne ordnes. I datakommunikasjonen med andre institusjoner lager ikke det noe problem, fordi kommunikasjonsformatet jo hverken bruker navnene eller de forkortede kodene, men nøytralkoden.

- i) Senere dukker enda et nytt spesialtilfelle opp - når arten Nototropis vedlomensis skal registreres. Registrøren skriver inn koden NOTO VED - og får dette svar

<u>Stasjon</u> :D4	<u>Dato</u> 1983-06-18	<u>Kl.</u> 1420
<u>Kategori</u> :	Hvirvelløse bunndyr	
<u>Prøvetype</u> :Grabb		<u>Dyp</u> : 41 m

Koden NOTO VED fins ikke i registeret. Hvis det fins et alternativt taxonnavn; pr.: det:

Hvis ikke, lag en foreløpig kode som består av registrørens initialer, bindestrek og den koden som nettopp ble brukt.

Foreløpig kode:

Taxonets navn :

Systemet vil underrette registeransvarlig om den foreløpige koden.

Registrøren kjenner ikke til noe alternativt taxonnavn og velger den andre muligheten. Siden han har initialene ABC, skriver han inn koden ABC - NOTO VED, samt artens navn - og skjermbilde av typen f kommer frem (med den foreløpige koden i bruk). Han legger parameterverdien inn.

Inntil den registeransvarlige har vurdert det nye taxonet - og enten godkjent det og registrert det med normal kode eller forkastet det av en eller annen grunn - vil det være tilgjengelig via koden ABC - NOTO VED.

- j) Når alle taxoner er skrevet inn, gir registrøren inn et spesielt tegn, og trinn c kommer tilbake med mulighet for å annelde neste prøve.

Når alle prøver er registrert, gir registrøren et annet tegn som bringer trinn b frem igjen. Her kan han så velge mellom å presentere en ny dataserie eller avslutte datainnleggingen.



NGUs edb-baserte referansearkiv

Beskrivelse

S. Høseggen, Teknisk Data a.s.

A. Rein, Norges geologiske undersøkelse

Nov./1983

INNHOOLD

	<u>Side</u>
1. NGUs edb-baserte referansearkiv	3
1.1. Formål	3
1.2. Omfang	4
2. Referansearkiv	5
2.1. Hva er en referanse, et referansearkiv og et dataarkiv?	5
2.2. Eksempler på referanser	7
3. Valg av programsystem for referansearkivet	8
3.1. Krav til hovedfunksjoner	8
3.2. Maskinteknisk utgangspunkt	9
3.3. Systemteknisk løsning	10
3.4. Programoversikt	12
3.5. Utviklingsmuligheter	13
4. Innlegging av referanser i arkivet	13
4.1. Innleggingsskjemaet for kart, rapporter og publikasjoner	13
4.2. Andre innleggingsskjemaer	15
4.3. Kravet til presisjon ved innlegging av opplysninger i referansearkivet	15
4.4. Manuelle rutiner	20
4.5. Ajourhold fra dataskjerm	23
5. Bruken av referansearkivet	23
5.1. Søking	23
5.2. Ekstern bruk	25
6. Sikring mot ødeleggelse og misbruk av referansearkivet	26
7. Framdriftsplan	27
7.1. Dokumentasjon om data	27
7.2. Bearbeidede data	28
7.3. Primærdata	28

## 1. NGUs edb-baserte referansesystem

### 1.1. Formål

Ut fra NGUs statutter er det institusjonens oppgave å bidra til et best mulig grunnlag for forvaltningen av Norges naturressurser og til å gi almenheten kunnskaper om landets geologi. Dette gjøres ved

- Geologisk, geofysisk og geokjemisk kartlegging av landets berggrunn og løsmasser.
- Undersøkelse og kartlegging av landets ressurser av malmer, industrimineraler, mineralske byggeråstoffer og grunnvann.
- Forskning og metodeutvikling til fremme av disse fagområdene.
- Utbygging og drift av sentralarkiv for data innen institusjonens fagområder.
- Utgivelse av kart, publikasjoner og rapporter, samt rådgivning, undervisning og annen informasjonsvirksomhet.

I årenes løp er det utgitt noen tusen geologiske, geofysiske og geokjemiske kart i Norge, og mange tusen rapporter og publikasjoner om geofaglige emner knyttet til berggrunn, løsmasser og grunnvann i Norge. Utgivere er i første rekke norske offentlige geoinstitusjoner, men noe er også publisert av utenlandske institusjoner, og en god del er utgitt av private konsulentfirmaer og bergverksindustrien.

Det finnes ingen samlet oversikt over denne store datamassen slik at den kan bli nyttiggjort på en mest mulig hensiktsmessig måte av de forskjellige som måtte ha bruk for denne viten. En kan regne med en god del dobbeltarbeid og betydelige feildisponeringer på grunn av dette.

Som den sentrale institusjonen for den geologiske utforskning av landet er det en av NGUs viktigste oppgaver å sørge for at de innsamlede data om Norges geologi blir gjort tilgjengelig for de aktuelle brukergruppene. NGU har allerede en betydelig oversikt gjennom samlingen av kart, rapporter og publikasjoner i sitt

bibliotek. Problemene er knyttet til at de manuelle systemene vanskeliggjør en rask og presis framletning av den aktuelle informasjonen, og at et betydelig antall rapporter m.v. ikke finnes i NGUs bibliotek.

Et edb-basert referansearkiv vil løse problemene med framletningen uavhengig av om informasjonen er lagret i NGUs bibliotek eller ved en annen institusjon. Referansearkivet skal være et hjelpemiddel til å kunne

- søke fram,
- sortere ut,
- få oversikt,

over tilgjengelige data som omhandler en geologisk problemstilling.

Et eksempel på bruk av referansearkivet er geologen på NGU som skal kartlegge et bestemt område. Det er da naturlig for han/henne å undersøke først i referansearkivet om det tidligere er foretatt geologiske undersøkelser i dette området. På den måten vil man kunne unngå dobbeltarbeid som selvsagt er kostbart og vanligvis fullstendig overflødig.

Et annet eksempel er planleggeren i fylkeskommunen. En vei skal planlegges, og da gjelder det bl.a. å skaffe seg en oversikt over geologisk forhold der veien tenkes lagt. Referansene kan gi en indikasjon på at det kan være risiko for konflikter mellom forskjellige interesser, f.eks. grunnvannskilder, grus o.l., eller spesielle grunnforhold som det må tas hensyn til.

## 1.2. Omfang

Referansearkivet skal ikke begrense seg til referanser til ferdigprodukter. NGUs målsetning er å samle referanser til data i hele spektret

- primærdata (bergartsprøver, geokjemiske prøver o.l.)
- bearbeidede data (feltdagbøker, forekomster o.l.)
- dokumentasjon om data (rapporter, publikasjoner og kart)

I framdriftsplanen som kommer som eget avsnitt, vil man se at NGU prioriterer å legge inn referanser til 'dokumentasjon om data' (rapporter, publikasjoner og kart), men andre referanser vil følge etter hvert.

Målet med referansearkivet er å gi en klart bedre service overfor alle som har bruk for opplysninger av forskjellig slag om berggrunn, løsmasser og grunnvann i Norge.

## 2. Hva er en referanse, et referansearkiv og et dataarkiv?

Eksempel på en referanse er bibliotekenes kartotek kort.

En referanse er informasjon om data. Informasjonen i referansen forteller om eksistensen av data i manuelle arkiver og edb-baserte databaser.

Populært sagt forteller en referanse

"hvem har data om hva fra hvem, hvor og når".

Som man ser av populærdefinisjonen er referansenes hovedelementer følgende

- Dataenes tittel (OM HVA)
- Dataprodusent (FRA HVEM)
- Stedfesting (FRA HVOR)
- Tidfesting (FRA NÅR)
- Arkiv (HOS HVEM)
- Emneord, sammendrag (OM HVA)

Et referansearkiv er et system som lagrer referanser og som på en rask og systematisk måte gjør det mulig å finne fram til interessante referanser. NGUs referansearkiv er edb-basert, dvs. man nytter edb-teknologi til å lagre referanser og til å assistere i søking på interessante referanser.

Dataarkivene er i noen tilfelle edb-baserte, såkalte databaser, men mange vil være manuelle arkiver. I databasen skjer lagringen på datamaskinens disker som alltid står tilkoblet datamaskinen.

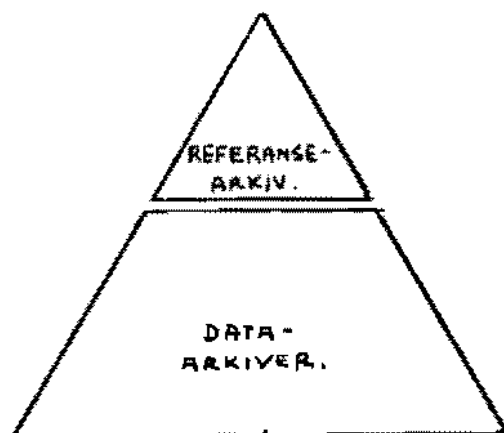
Eksempler på manuelle dataarkiver ved NGU er

- Biblioteket
- Arkivet for feltdagbøker og feltkart m.v.
- Lagre for borkjerner, bergartsprøver og geokjemiske prøvemedier.
- Magnetbånd med geofysiske målinger.

Eksempler på databaser er

- Registret over grusforekomster, i drift med Seksjon for ingeniørgeologi som ansvarlig
- Registret for grunnvannsbrønner i fjell, i drift med Seksjon for hydrogeologisk som ansvarlig
- Registret over forekomster av malmer, industrimineraler og naturstein
- Register over mutinger og utmål, i drift med Bergmesterembetene som ansvarlig
- Registret for berggrunnsobservasjoner (GEOMAP), i drift ved Berggrunnsavdelingen
- Registret for sedimentanalyser (SENGU), i drift med Seksjon for ingeniørgeologi som ansvarlig.

Kort oppsummert kan en si at referansearkivet er en overbygning over de forskjellige dataarkivene. Noen av dataarkivene kan ligge utenom NGU.



For at referansearkivet skal fungere som en tilfredsstillende kobling mellom de forskjellige dataarkivene må alle referansene ha noen felles søkemuligheter uavhengig av hvilket dataarkiv de er knyttet til. Konsistent bruk av stedfesting og geofaglige emneord er viktigst i denne sammenheng.

## 2.2. Eksempler på referanser

Referansenes hovedelementer er nevnt tidligere, beskrivelsen i referansene skal være kort og oversiktlig, men likevel skal den ut fra sitt innhold være lesbar og tolkbar. Dvs. selv om referanser er i form av stikkord, forholdsvis komprimert og variabel i presisjonsgrad, skal brukeren av arkivet lett kunne tolke hva referansen gjelder.

Først et eksempel på en referanse til en dataserie lagret i en database.

"105.03.991

Måledata. Drammen

Refleksjonsseismiske undersøkelser i Drammensfjorden.

NGU/Bjerklif, K. 1982.

NGU/Borarkiv/edb; Olsen, H., 375 prøver.

Løsmasser. Sedimentologi. Seismiske dybder.

Refleksjonsseismikk."

En referanse til en rapport i NGUs bibliotek kan se slik ut:

Magnetiske borhullsmålinger ved Framfjord i Sogn.

Brandhaug, K. 1982.

NGU-rapport nr. 1940, 6 s. + 2 kart

NGU/Biblioteket.

sogn og Fjordane. Vik. Myrkedalen 1316 4.

Geofysikk. Magnetiske borhull. Prospektering."

### 3. Valg av programsystem for referansearkivet

#### 3.1. Krav til hovedfunksjoner

Hovedfunksjonene i referansearkivet er

- lagring av referanser
- søking etter referanser

Utvalgte referanser eller deler av dem skal kunne skrives ut sortert etter bestemte regler. Dermed kan man utgi forskjellige kataloger og oversikter. Systemet bør også gi statistiske opplysninger.

Innlegging av referanser utgjør en stor del av de manuelle operasjonene knyttet til systemet. Følgelig er det lagt vekt på at arbeidet ved dataskjermen for

- registrering og
- korreksjon

av referanser skal være brukervennlig.

Lageret med referanser vil etter hvert utgjøre et stort volum på datamaskinens lagringsenhet, disk. Søkertiden vil selvsagt påvirkes av mengden av informasjon som skal gjennomføres, men likevel må søkefunksjonen være slik at brukeren får svar på skjermen innen en rimelig tid (sekunder).

Bl.a. fordi det ikke finnes noen norsk geologisk tesaurus (oversikt over tillatte geologiske faguttrykk), har NGU prioritert systemer der alle meningsbærende ord i referansen skal kunne søkes på (søking i fritekst).

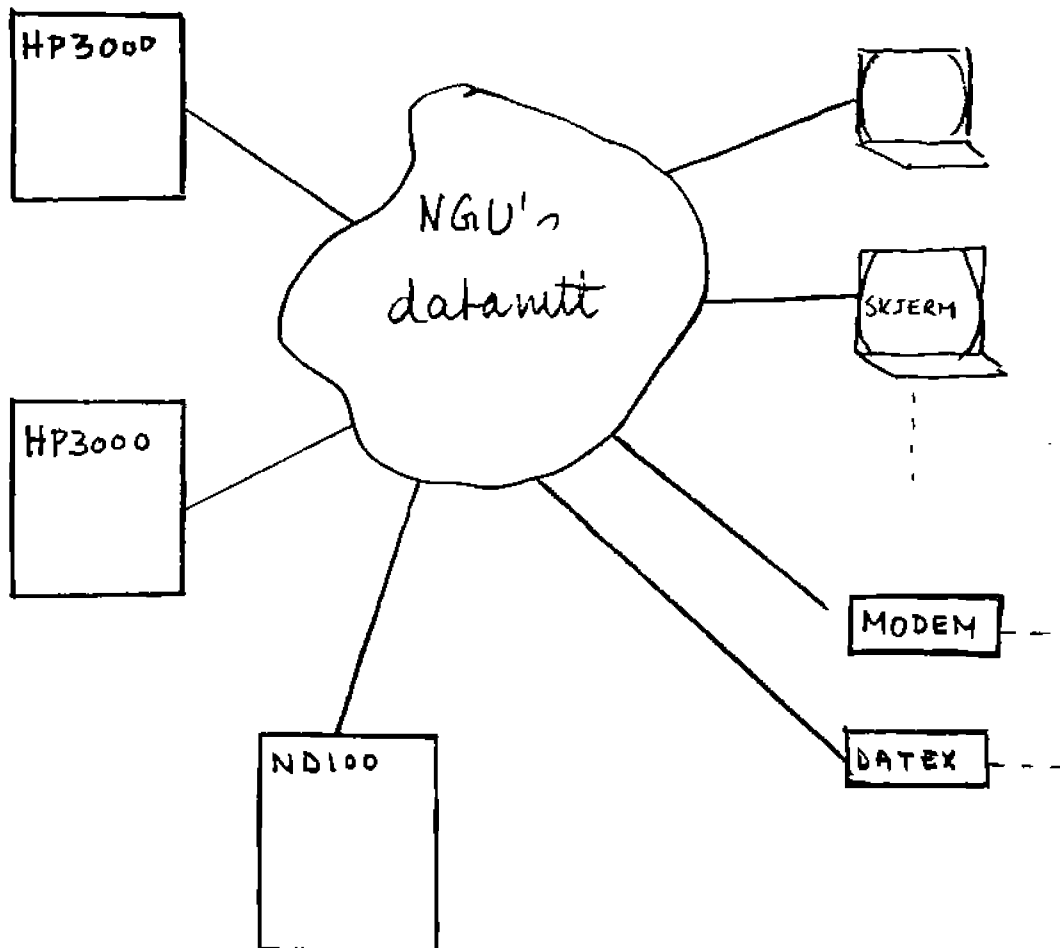


### 3.2. Maskinteknisk utgangspunkt

NGU har basert sin hovedaktivitet innen edb på Hewlett Packard-utstyr. NGU har 2 stk. HP3000 som er å betrakte som et sentralanlegg som bl.a. administrerer databasene. Begge har 1 Mbyte hukommelse og til sammen 390 Mbyte lagringskapasitet. Første halvår 1984 vil den ene HP3000-III erstattes med en HP3000/64 med 3 Mbyte hukommelse og diskkapasiteten øker med 400 Mbyte.

Maskinene er koblet til NGUs lokale datanett hvor også brukerne er tilkoblet med dataskjermer plassert ved sine arbeidsplasser.

Via det offentlige datanettet og modem kan brukere utenfor NGU koble seg til systemet.



### 3.3. Systemteknisk løsning

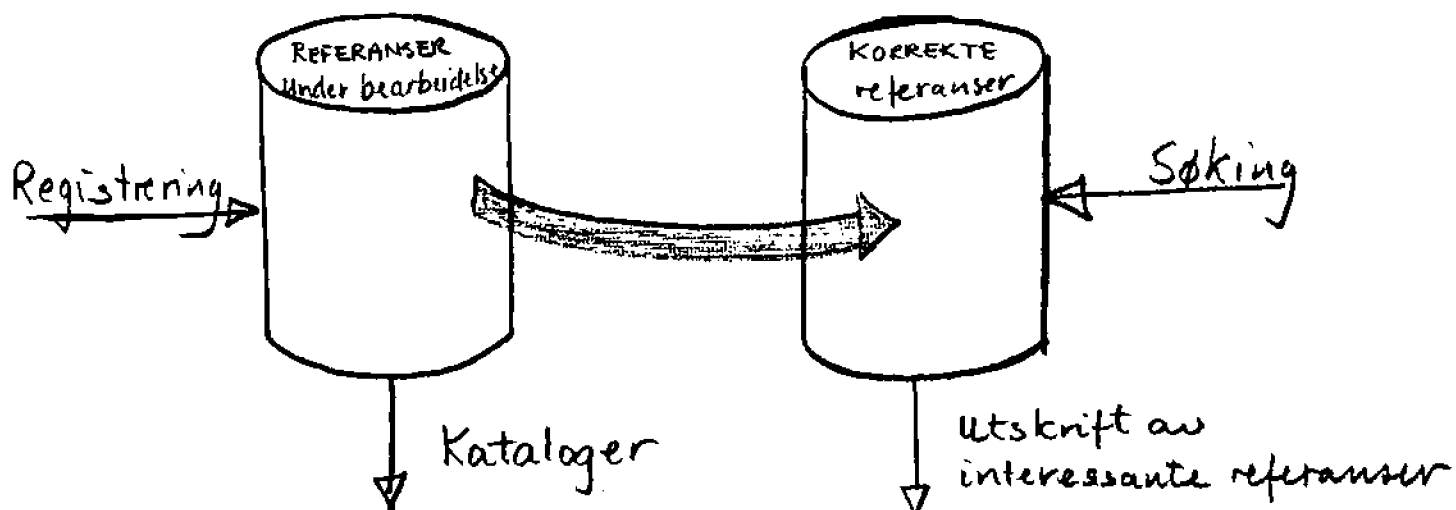
Valg og utforming av den systemtekniske løsning av referansearkivet er basert på

- kravene til hovedfunksjoner,
- maskinteknisk utgangspunkt.

Kravene til lagring med ajourholdsfunksjoner betinger at en referanse raskt må kunne gjenfinnes, skrives ut på skjerm, modifiseres og skrives ut igjen fra/til lageret. Kravene til rask søking og fritekst søking stiller spesielle krav til hvordan dataene er lagret.

For flere aktuelle systemer viste det seg at disse to hovedkravene er motstridende. Noen systemer har rask oppslagstid på enkeltreferanser når de er angitt med en eller et fåtall oppslagsnøkler, men derimot vil søketiden gjennom alle referanser ta lang tid. For andre systemer er dataene organisert slik at søketiden er kort, men systemene er tungvint å oppdatere.

NGU valgte et todelt system



I stedet for å basere seg på ett programsystem er referansene innom to systemer. Slik er det innvendig, men fra brukersiden

ser det ut som ett referansearkiv. Brukerne har ett program hvor man via menyer kommer fram til de funksjoner man skal bruke.

Likevel synes vi det er viktig å være klar over todelingen av systemet

- referanser under bearbeidelse
- korrekte referanser

Selv om vi baserte oss på ett programsystem, måtte vi logisk sett operere med to typer referanser etter samme prinsipp som todelingen over. Ellers ville brukere kunne søke seg fram til referanser som det ikke er lest korrektur på.

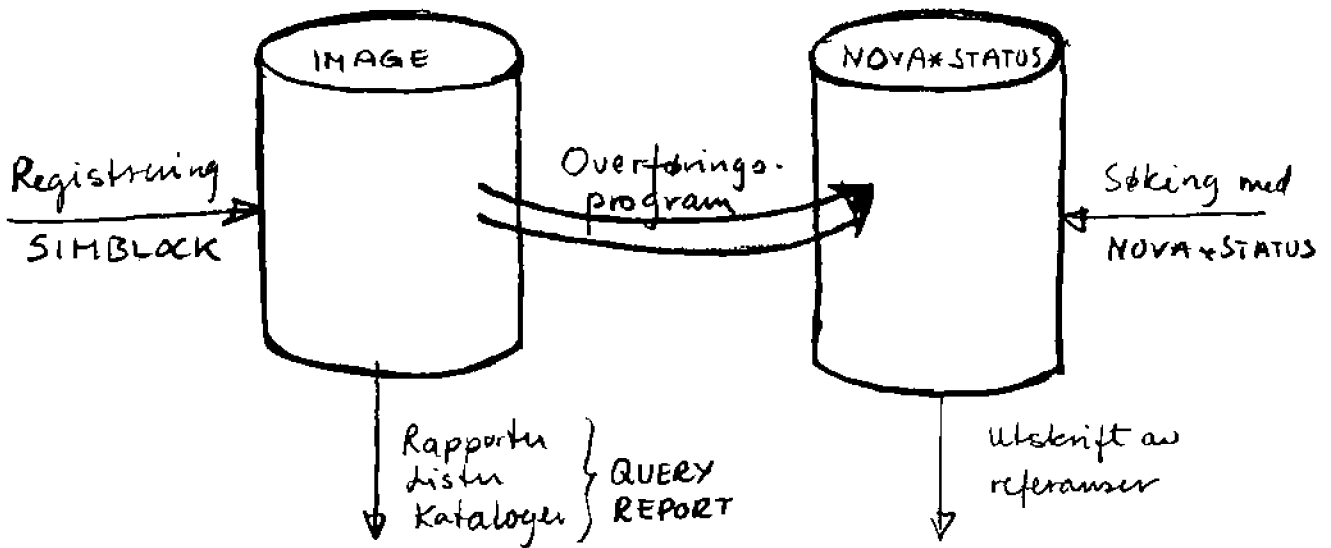
Referanser under bearbeiding lagres i ett programsystem hvor hver referanse har en entydig oppslagsnøkkel, hvert informasjonsfelt i referansen er entydig identifisert og hver referanse har fått avsatt en fast plass.

Når referansene er tastet inn, kan man produsere korrekturlister for kontroll slik at man er sikker på at de riktige opplysningene er lagt inn.

Fra dette lagringssystemet produseres også kataloger og oversikter sortert på bestemte informasjonsfelt, som f.eks. forfatter, kommuner, kartblad o.l.

Når man har avsluttet en registreringsrunde og har korrigert referansene, overføres (kopieres) de til det andre lagringssystemet. Dette er en operasjon som gjøres med jevne mellomrom. Referansene blir nå komprimert, dvs. man mister feltidentifikasjonen på hver informasjon. Men referansen vil fremdeles være lesbar og brukerne vil forstå hva hvert ord i referansen står for. Referansene blir nå lagret slik at de kan søkes ut raskt.

### 3.4. Programoversikt



Programsystemer som inngår er:

- IMAGE databasesystem tar seg av lagring av referanser under bearbeidelse. Set-orientert databasesystem.
- NOVA\*STATUS lagrer de korrekte referanser. Lagringsmetoden er inverterte filer. NOVA\*STATUS har også et raskt søkeprogram.
- SIMBLOCK er et maskin- og skjermuavhengig skjermhåndterings-system. Dette forenkler arbeidet med registrering og ajourhold av referanser fra dataskjerm.
- QUERY og REPORT er et spørrespråk og rapportgenerator for IMAGE. Dette benyttes til utskrift av kataloger og andre oversikter.
- NGU har også utviklet programmer som knytter systemene sammen. Dette gjelder først og fremst for overføring av data mellom systemene, samt menyopplegget.

### 3.5. Utviklingsmuligheter

Når NGU valgte en todelt løsning med

- IMAGE
- NOVA\*STATUS

var det ut fra en vurdering av mulige systemer som var tilgjengelig på HP3000 og som tilfredsstilte kravene.

Man ønsket å komme i gang så snart som mulig. Men samtidig burde systemet være slik at vi relativt lett kunne gå over til mer raffinerte systemer når de var tilgjengelige, og når NGU var moden for det. Dagens løsning vil være brukbar en god stund framover. Tiden vil være inne for å se på andre systemer når kapasitetshensyn, lang søketid o.l. vil kreve at noe blir gjort.

Allerede i dag finnes systemer som kan erstatte NOVA\*STATUS, f.eks.

- SIFT (Neste generasjon av NOVA\*STATUS, utviklet av Statens Rasjonaliseringsdirektorat, Norsk Data A/S og flere)
- MIMER

Det vil også sikkert komme bedre hjelpemidler på registreringssiden.

## 4. Innlegging av referanser i arkivet

### 4.1. Innleggingsskjemaet for kart, rapporter og publikasjoner

De klart viktigste kildene for geofaglig informasjon er kart, rapporter og publikasjoner. For disse er det laget ett felles skjema for innlegging av referanser i referansearkivet. Skjemaet er vist på neste side. Det vil bli utarbeidet mer detaljerte retningslinjer for bruken av skjemaet.

REGISTRERINGSSKJEMA FOR REFERANSER TIL RAPPORTER, PUBLIKASJONER OG KART

Tittel: \_\_\_\_\_  
\_\_\_\_\_

Forfatter(e): \_\_\_\_\_

Publisert i: \_\_\_\_\_

Volum/nr./side \_\_\_\_\_ Sidetall: \_\_\_\_\_

Utgiver: \_\_\_\_\_ År: \_\_\_\_\_

Fylke(r): \_\_\_\_\_

Kommuner: \_\_\_\_\_

Kartblad (1:250 000): \_\_\_\_\_

Kartblad (1: 50 000): \_\_\_\_\_

Stedsnavn: \_\_\_\_\_ Sone: \_\_\_\_\_ X: \_\_\_\_\_ Y: \_\_\_\_\_

Fysisk lagerplass: \_\_\_\_\_

Gradering: \_\_\_\_\_ UDK: \_\_\_\_\_

Emneord: \_\_\_\_\_  
\_\_\_\_\_

Sammendrag:  
\_\_\_\_\_  
\_\_\_\_\_  
<20 linjer>  
\_\_\_\_\_  
\_\_\_\_\_

#### 4.2. Andre innleggings skjemaer

Referanser til dokumentasjon om data (rapporter, publikasjoner, kart) har som nevnt første prioritet. I neste omgang kommer innlegging av referanser til

- bearbeidede data
- primærdata

Eksempler på slike data er:

- grusregistret
- borebrønner i fjell
- feltdagbøker
- forekomster av malmer, industrimineraler og naturstein
- analyseprøver på magnetbånd
- geofysiske målinger på magnetbånd

Noen av disse dataene lagres i manuelle arkiver og må derfor registreres via utfylling av skjema. Men i mange tilfelle vil man kunne avlede referansene fra data i databaser, dvs. at referansene genereres automatisk slik tilfellet er for grusregistret.

Geofysiske målinger fra helikopter og kjemiske analyser behandles også av datamaskin selv om resultatene eller primærdataene ikke lagres i databaser. Referanser til slike dataserier vil kunne genereres automatisk når dataene behandles i datamaskinen.

#### 4.3. Kravet til presisjon ved innlegging av opplysninger i referansearkivet

Et problem er presisjonen i referansenes innhold. Eller for å si det enklere: Hvor sikker kan jeg være for å få med alle aktuelle referanser i et søk? Svaret er: 100% sikker kan du aldri være, og det skyldes referansenes manglende presisjon.

Hvorfor har NGU startet etableringen av referansearkivet uten å kunne tilby 100% presisjon? Her er det mange faktorer som spiller inn, men overordnet mener vi det er bedre å kunne gi 90% av samlet informasjon enn 0. På den annen side er presisjonsfaktorene noe som vi kan forbedre med tiden uten at det forhindrer at referansearkivet oppstartes nå.

Som nevnt er det flere faktorer som berører presisjonen, og la oss se litt på de hovedretningslinjer som er lagt til grunn.

- Referanser til data av eldre årgang er ofte vanskeligst å lage med god presisjon. I slike tilfelle må man vanligvis legge ned mye arbeid for å lage referansene. I praksis må vi her akseptere at presisjonen ikke blir god nok.
- Selv med data av nyere årgang kan det noen ganger være litt av en innsats å gi tilfredsstillende presisjon for alle informasjonsfelter. I slike tilfelle kan man akseptere at man lempet litt på presisjonen. I og med at referansearkivet er edb-basert kan man om nødvendig følge opp med oversikter over referanser som senere bør oppdateres mht. presisjon.
- En standardisering av hva man har lov til å bruke av emneord og stedsangivelser vil forbedre presisjonen betydelig.

Vi skal orientere litt nærmere om de hovedlinjer som er valgt når det gjelder stedsangivelse og bruk av emneord.

#### Presisjonen ved søkingen på geografisk område

For et referansearkiv om norsk geologi er en tilfredsstillende søking på geografisk område helt avgjørende.

Vi har gjennomgående for hele referansearkivet valgt å nytte fire forskjellige og standardiserte søkemuligheter slik det går fram



av skjemaet under 4.1.

- kartblad i M 1:50 000, landet er inndelt i 725 slike kartblad
- kartblad i M 1:250 000, landet er delt i 46 slike kartblad
- kommune, landet består av 454 kommuner
- fylke, landet består av 18 fylker

Kommune og fylke er valgt fordi disse er de to viktigste politisk-administrative inndelingene av landet. Kartblad i M 1:50 000 og M 1:250 000 er tatt med fordi disse er de to viktigste landsomfattende, topografiske og geofaglige kartseriene. Av den grunn er de ofte et basismateriale for geofaglige arbeidssoppgaver og problemstillinger. Det er også lagt vekt på at det skal være en rimelig stor forskjell i geografisk utstrekning fra den ene søkemuligheten til den andre. Således er et fylke i gjennomsnitt 40 ganger større enn et kartblad i M 1:50 000.

Det er et visst problem at noen navn på kartblad endres og at det kan forekomme endringer i kommunegrensene. Men noen bedre alternativer finnes ikke.

Ut over denne gjennomgående og obligatoriske form for stedfesting, vil det der det er praktisk mulig kunne gis en mer nøyaktig stedfesting, f.eks. av forekomster av malmer, industrimineraler, naturstein, grus og grunnvann.

XY-koordinater (UTM) vil være naturlig å innlemme i referanser for slike forekomster og geologiske dataserier. Registreringen av XY-koordinater vil ikke medføre merarbeid av betydning, etter slike referanser vanligvis vil genereres automatisk fra databaser.

Som en frivillig mulighet kan referanser også inneholde lokale stedsnavn o.l. Slike opplysninger vil vanligvis bidra til å øke presisjonen.

Et spesielt problem er stedsangivelsen for dokumenter som omfatter store geografiske områder. I innleggings-skjemaet er det satt

av 3 felter for kommune, 4 felter for kartblad i M 1:50 000, 2 felter for fylke og 2 felter for kartblad i M 1:250 000. Mange dokumenter vil dekke mer enn 4 kartblad i M 1:50 000 og noen vil dekke mer enn 2 fylker.

Vanligvis vil storregionale dokumenter være av mindre interesse for alle kartblad i M 1:50 000, og man bør skjønnsmessig ta med de 4 viktigste. For dokumenter som går over flere fylker kan samme prinsippet nyttes. Men det er grunn til å nevne at systemet med søking i fritekst muliggjør en innlegging til slutt etter sammendraget av stedsnavn som det ikke er blitt plass til i de standardiserte feltene.

#### Presisjonen ved søkingen på emneord

Det ville vært ønskelig med en vedtatt emneordskatalog (oppslagsregister/tesaurus) for geologi. Men det vil være et meget stort arbeid å lage denne, og innleggingen av de store mengdene data i referansearkivet kan ikke utsettes til katalogen er utarbeidet.

En hovedinndeling på fagområder må likevel gjennomføres for å sikre en noenlunde sikker søking på hovedfagområder. NGUs basisorganisasjon er inndelt i seksjoner etter fagområder. Denne faginndelingen skal nyttes ved bruken av minst ett av de fire feltene for emneord:

1. Berggrunn (omfatter berggrunnskartlegging og generell berggrunnsgeologi)
2. Malmer (omfatter malmundersøkelser og generell malmegeologi)
3. Industrimineraler (omfatter industrimineraler og naturstein, dvs. skifer og blokkstein)
4. Løsmasser (omfatter løsmassekartlegging og generell kvartærgeologi)
5. Hydrogeologi (omfatter grunnvannsundersøkelser og generell hydrogeologi)

6. Ingeniørgeologi (omfatter undersøkelser av sand, grus, leire, pukk og generell ingeniørgeologi)
7. Geofysikk (omfatter generell geofysikk, geofysisk kartlegging og geofysisk prospektering)
8. Geokjemi (omfatter generell geokjemi, geokjemisk kartlegging og geokjemisk prospektering)
9. Kjemiske analyser
10. EDB

Noen dokumenter kan omfatte flere enn ett av de 10 fagområdene. Da skal et tilsvarende antall emneord nyttes.

### Oppslagsregistre

Oppslagsregister har klar sammenheng med presisjonsgraden som ble diskutert foran.

Det er et klart behov for å få etablert slike registre som inneholder de ord eller koder som er "lov" å bruke innen

- geografi
- fagområde.

Oppslagsregistre har to hovedfunksjoner i forbindelse med referansearkivet.

- Under dataregistrering kan oppslagsregistret si fra om et ord eller en kode er feil brukt eller feil skrevet.
- Under søking kan et oppslagsregister benyttes til å gi alternative søkenøkler. F.eks. om et kommunenavn angis som søkenøkkel, kan oppslagsregistret foreslå at fylkesnavnet, navnet på grunnkretsen, samt kartbladbetegnelsene som dekker kommunene, blir benyttet som alternative søkeord.

Enkle oppslagsregistre er forholdsvis lett å lage. Men som man ser foran, vil det være nyttig å bygge oppslagsregistrene

hierarkiske. Dette øker kompleksiteten som igjen gjør at en edb-teknisk løsning blir mye mer kostbar.

NGU følger med interesse det arbeidet som Miljøverndepartementet og Norsk institutt for vannforskning nå er i gang med for å kunne utvikle oppslagsregistre som kan brukes i flere fagområder.

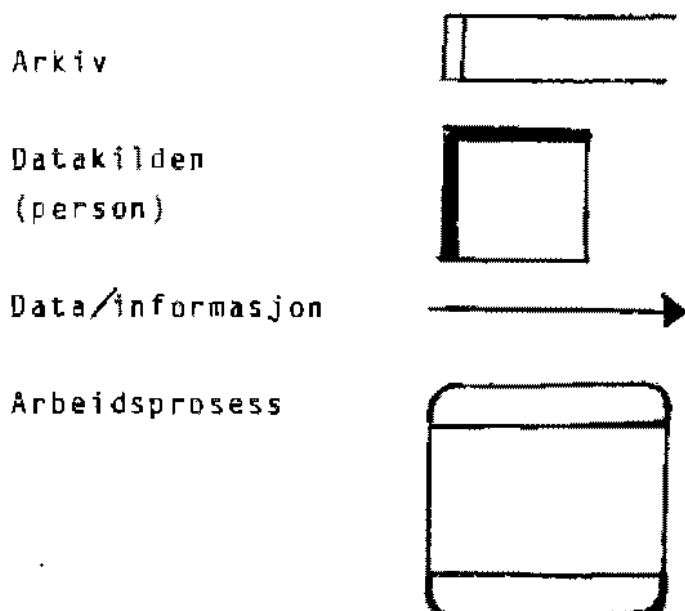
#### 4.4. Manuelle rutiner

Ansvar for engangsregistreringer er lagt til hver enkelt seksjon ved NGU. Seksjonen må avsette fagpersonell til å gå igjennom det aktuelle datamateriale og lage referanser. Et viktig poeng er at det er en faglig kompetent person som er satt til å gå igjennom referansene. På den måten vil en få god presisjon på referansene.

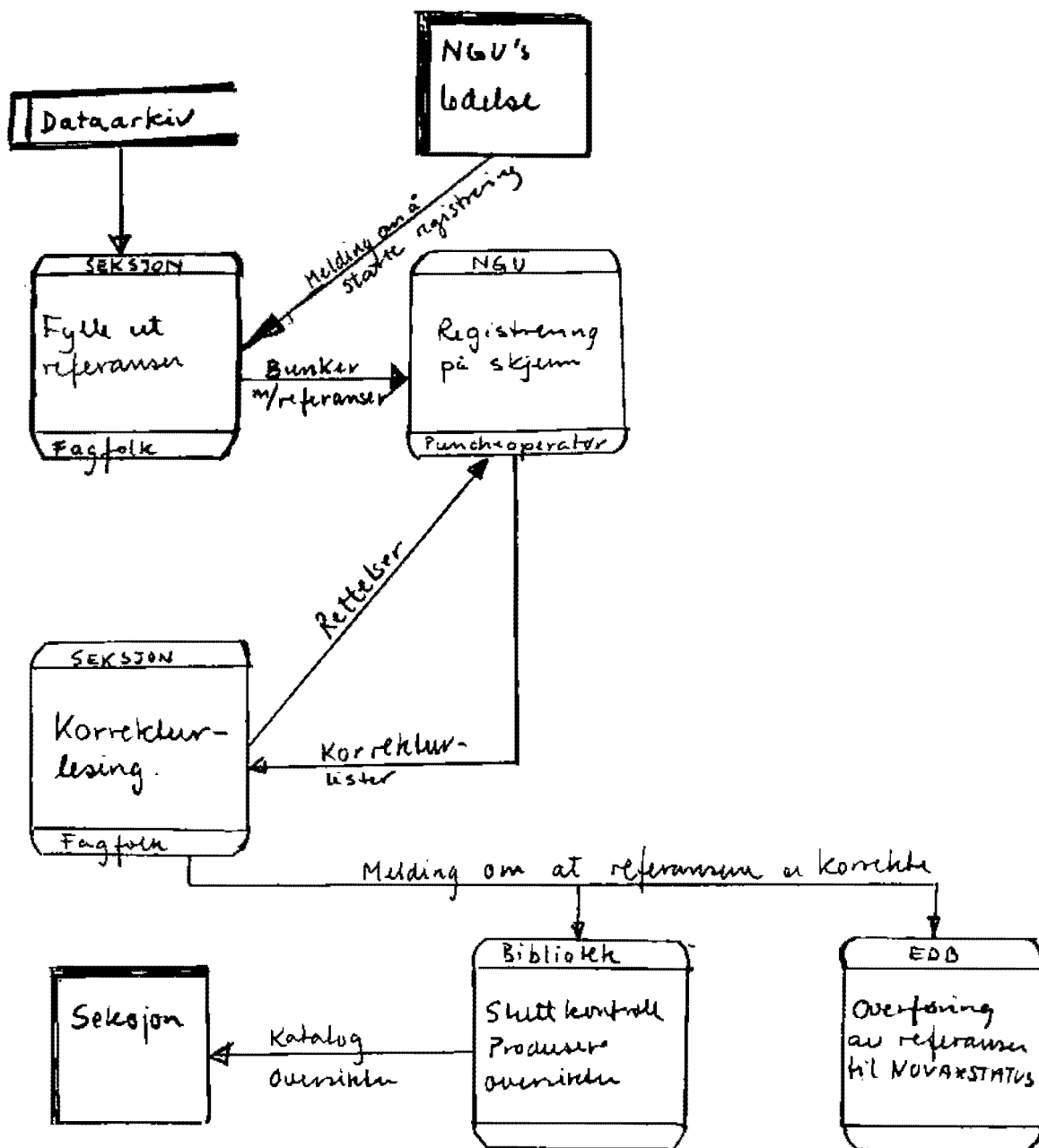
Når det gjelder NGU-rapporter, er lagning av referansen en grei jobb ettersom referansene vil kunne tas direkte fra rapportforsiden. Riktig nok mangler rapportforside på eldre rapporter.

Den videre rutinen i engangsregistreringen vises best i et dataflytdiagram. La oss først forklare hva symbolene i dataflytdiagrammet betyr.

Tegnforklaring:

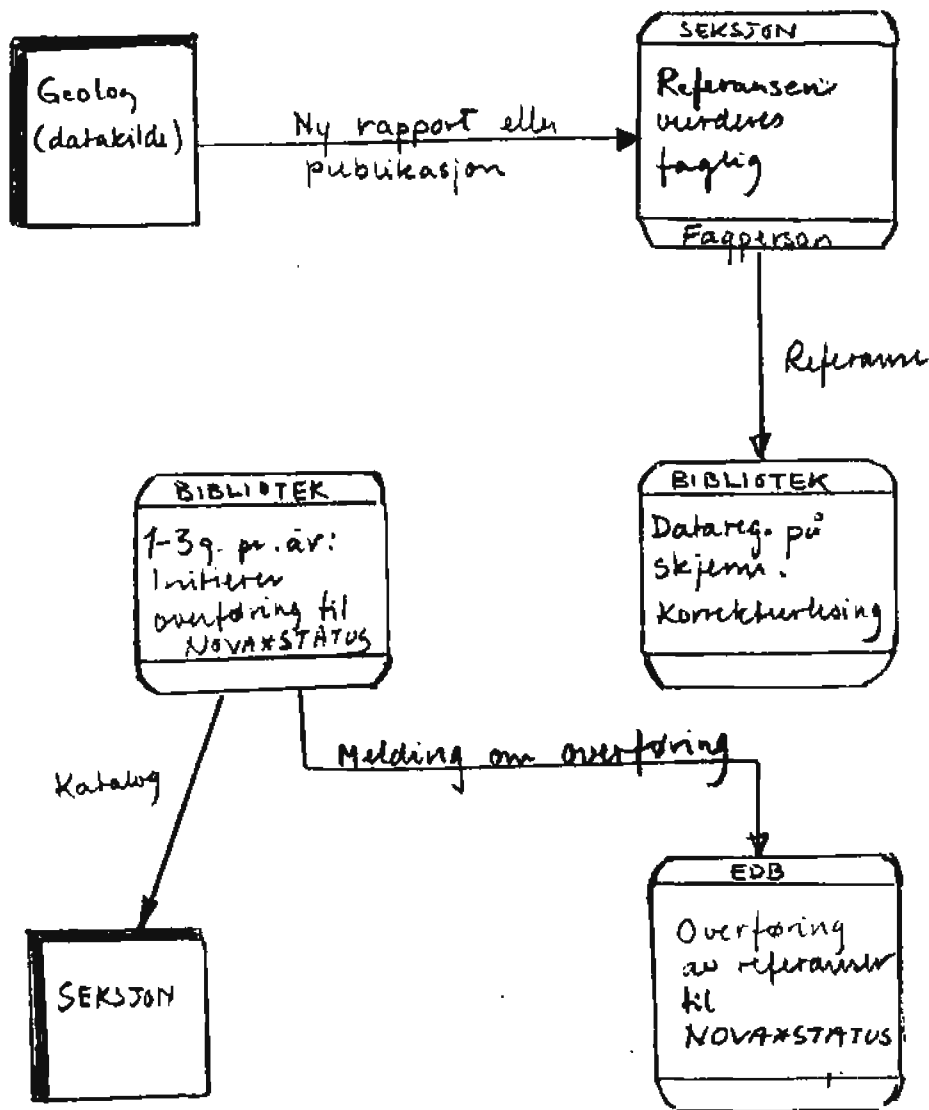


Dataflyttdiagrammet ved engangsregistreringen vil se slik ut:



Etter førstegangsregistreringen vil referansearkivet oppdateres løpende etter hvert som nye rapporter produseres.

En slik rutinemessige innlegging av nye referanser i referansearkivet vil i dataflyttdiagrammsform se slik ut:



I den rutinemessige innleggingen av nye referanser har biblioteket en mer sentral rolle enn ved engangsregistreringen. Men seksjonene vil også i dette arbeidet være ansvarlig for utfyllingen av referansene.

#### 4.5. Ajourhold fra dataskjerm

Med ajourhold menes det å registrere nye referanser inn på lagringssystemet, korrigere feil og legge til nye opplysninger på allerede eksisterende referanser. Ajourhold av referansene utføres fra dataskjerm. Ved førstegangsregistrering av en referanse er informasjonen å finne utfyllt på et skjema. Operatøren vil kunne kalle fram et tilsvarende skjema på dataskjermen. Dataregistreringen består i å fylle ut feltene på skjermen.

Når en referanse skal korrigeres, kaller operatøren fram et skjermbilde med referansen utfyllt. Deretter kan de felter som er feil, korrigeres og referansen blir lagt tilbake i lagringssystemet.

#### 5. Bruken av referansearkivet

##### 5.1. Søking

Søking etter referanser er den viktigste funksjonen i referansearkivet. Etter hvert vil arkivet inneholde anselig store mengder informasjon. Søkemekanismen er derfor utformet slik at den som er ute etter bestemte problemstillinger lett kan søke seg fram til de referanser som er av interesse og få disse utskrevet.

Søkespråket i NOVA\*STATUS er laget slik at brukeren gir noen stikkord (søkeord) som han regner med finnes i de referanser han er ute etter.

Eksempel

```
|
| GI KOMMANDO
| NAMSOS?
| 87 dokumenter funnet
|
```

NOVA\*STATUS angir at den er klar for stikkord ved teksten GI KOMMANDO.

Brukeren har her foreslått NAMSOS som stikkord fordi han er interessert i referanser der NAMSOS er brukt som stedfestingsbegrep.

Systemet kvitterer med å si at "87 dokumenter funnet", altså 87 referanser har ordet NAMSOS ett eller annet sted i teksten.

Eksempel

```
|
| GI KOMMANDO
| LESJA. OG. LØSMASSER?
| 7 dokumenter funnet
| GI KOMMANDO
| LIST
| 1051.03.42
|   Forekomstreg. LESJA 14192 DOMBÅS
|   Nyregistrering nr. 24
|   NGU/Furuhaug, 0. 1981
|   NGU/Grusregisteret; Furuhaug, 0.
|   Løsmasser, sand og grus, kvalitet,
|   volum, arealfordeling
|
```

Her har brukeren kombinert to stikkord, LESJA og LØSMASSER.



Ettersom den logiske operatoren OG er brukt, vil systemet søke fram de referanser som inneholder begge stikkord.

Resultatet her var som vi ser 7 referanser. Dette er et antall som vil være interessant å kikke nærmere på. Derfor gir brukeren kommandoen LIST og første referanse skrives ut på skjermen.

Slik kan man "blå" seg igjennom referansene eller man kan få dem utskrevet på papir, eller fortsette søkingen.

## 5.2. Ekstern bruk

Eksterne henvendelser om opplysninger om norsk geologi vil fortsatt skje ved skriftlig, telefonisk eller direkte henvendelse til NGUs bibliotek. Biblioteket vil så bruke referansearkivet til å finne fram til de aktuelle opplysningene. Etter avtale med kunden vil biblioteket deretter formidle utskrifter fra arkivet, utlån av aktuelle litteratur eller salg av NGUs kart, rapporter eller publikasjoner. I noen tilfelle er det også aktuelt å formidle direkte kontakt med NGUs fagpersonell innen det aktuelle fagfelt/geografiske område, eller ansvarlige for databaser.

Men referansearkivet er også åpent for direkte tilkobling fra eksterne brukere. Det er to muligheter for tilkobling til NGUs HP3000:

- Modem 300 baud
- Datex 1200 baud

For å bruke systemet trenger man bare å kjenne påloggingsprosedyren:

```
<trykk return>  
:HELLO XX,NGU.REF
```

Deretter vil brukeren komme direkte inn på referansearkivet. Ved hjelp av menyer kan brukeren styre seg inn på de funksjoner han ønsker.

I NTNf-regi arbeides det med å etablere en referansesentral for all stedfestede data. NGUs referanser eller deler av dem kan lett overføres til en slik sentral. Det er heller ikke store problemer med å utbygge systemet slik at det kan ta seg av henvendelser fra denne referansesentralen.

For NGU-ansatte vil det være mest vanlig at den enkelte medarbeideren selv foretar søkingen i referansearkivet.

## 6. Sikring mot ødeleggelse og misbruk av referansearkivet

Sikkerhetsaspektet i forbindelse med edb-systemer har to sider:

- sikring mot ødelegges
- sikring mot misbruk

Referansearkivet vil som andre viktige systemer ved NGU inngå i Seksjon for data og systemtjeneste sine "back-up"-rutiner. Vanligvis vil man kunne gå tilbake til gårdagens kopi hvis arkivet skulle bli ødelagt p.g.a. teknisk feil.

Når det gjelder sikkerhet mot misbruk, bør det ikke være noe problem. Informasjonen som er å finne i referansene er i prinsippet åpen for alle. Angående data som er av fortrolig karakter, bør en del av referansen til dem likevel kunne registreres etter tillatelse fra oppdragsgiver. Det gjelder opplysninger om tittel, stedfesting og hvem som har utført undersøkelsen. En bruker som er interessert i disse dataene vil dermed kunne få nok informasjon til å henvende seg til de som disponerer dataene for å få avklart om man kan få tilgang til materialet. Aktuelle data av mer eller mindre fortrolig karakter vil være data i forbindelse med prospektering og private konsulentfirmaers oppdragsrapporter m.v.

Sett under ett vil firmaer og etater som bidrar med referanser til mer eller mindre fortrolige data på lengre sikt ha nytte av

åpenheten, sammenholdt med alternativet om fullstendig hemmeligholdelse av opplysningene.

## 7. Framdriftsplan

Målsetningen med referansearkivet kan synes ambisiøs. Det bør den også være. NGU er en viktig produsent og distributør av geologisk informasjon til samfunnet. Informasjonen og informasjonsformidlingen må være lagt til rette slik at man kan gi denne tjenesten på en tilfredsstillende måte.

Vi må likevel innse at oppbyggingen av referansearkivet vil ta tid. Framdriftsplanen for de 3 ulike typene referanser er:

### 7.1. Dokumentasjon om data (kart, rapporter og publikasjoner)

- 1984 Alle referanser til rapporter ved NGU om norsk geologi legges inn
- 1984-85 Alle referanser til artikler om norsk geologi i publikasjoner ved NGU legges inn
- 1984-85 Alle referanser til foreløpige og ferdige kart ved NGU i M 1:250 000, M 1:100 000, M 1:50 000 og M 1:20 000 legges inn
- 1985-87 Innlegging av referanser til kart, rapporter og publikasjoner om norsk geologi, hvor dataene er lagret ved andre institusjoner enn NGU

## 7.2. Bearbejdede data

For enkelte typer data er innleggingen av referanser allerede påbegynt, men samlet sett vil dette arbeidet ha lavere prioritet enn pkt. 7.1.

1983- Referanser til feltdagbøker og "feltkart", grusforekomster, malforekomster, borebrønner, kjemiske og sedimentologiske analyser og andre dataserier.

## 7.3. Primærdata

For enkelte typer data er innleggingen av referanser allerede påbegynt, men samlet sett vil dette arbeidet ha lavere prioritet enn pkt. 1 og 2.

1983- Referanser til bergartsprøver lagret ved NGU