

Norsk Oseanografisk Datasenter (NOD)

SYSTEM

FOR DATAADMINISTRASJON VED

N O D

Januar, 1977



FISKERIDIREKTORATETS HAVFORSKNINGSINSTITUTT
BERGEN

Strand

SYSTEM

FOR DATAADMINISTRASJON VED

N O D

Januar, 1977

INNHold

	Side
0. INNLEDNING	1
1. DEFINISJONER OG FORKORTELSER	2
2. MÅLFORMULERING FOR SYSTEMET	3
3. TEKNISKE FORHOLD	4
4. DELSYSTEMER	5
4.1 KOMMENTAR GRAF-NR 0 - objektsystemets virkemåte	12
4.2 GRAF-NR 7 - lagring	15
4.3 GRAF-NR 8 - framhenting og beregning	17
4.4 GRAF-NR 9 - arkiv-log	19
5. DATABASE LØSNINGEN	22
5.1 KOMMENTARER GRAF AS	24
5.2 TABELL 1/ARKIVINFORMASJON OM DATA	25
5.3 TABELL 2/INFORMASJON-INNHOLD/GRAF AS	27
5.5 TABELL 3/KOMMENTARER TIL RELASJONER/GRAF AS	32
6. KOMPRIMERT SYSTEMOVERSIKT	34
7. VIDERE SYSTEMERING	37
8. KONKLUSJON - KOSTNADSOVERSLAG	39
APPENDIX	42

0. INNLEDNING

I september 1976 ble det ved Norsk Oseanografisk Datasenter i Bergen nedsatt en komite bestående av :

cand.mag	Arvid Hilstad (formann)
	stud./Institutt for Informasjonsvitenskap, Universitetet i Bergen.
Vit.ass.	Øyvin Strand /(NOD)
Førstesekretær	Arne Sandahl /(NOD)
Havforsker ass.	Kjell Seglem /Havforskningsinstituttet

Komiteens mandat var følgende:

- Diskutere opplegg for et system for dataadministrasjon ved NOD.
- Komiteens arbeide skal være avsluttet ved utgangen av 1976
- Legge fram en skisse av det system man har funnet

Det ble holdt ukentlige møter der man diskuterte opplegg som den enkelte hadde forberedt til hvert møte.

Komiteen mener på ingen måte å ha framstilt ferdig beskrivelsen av det framtidige system. Systembeskrivelsen i denne rapporten må kun oppfattes som en ramme for det framtidige system.

Hensikten med rapporten er å kort beskrive de systemdeler man på nåværende tidspunkt har behandlet samt å forsøke å gi et kostnads-overslag for det framtidige systemet. Dette overslaget må nødvendigvis bli usikkert på grunn av den grove detaljeringsgraden av systemet.

1. DEFINISJONER OG FORKORTELSER

Datatype

Dette er et samlingsnavn/-identifikasjon på PARAMETRE som vanligvis lagres sammen på grunn av logisk samhörighet eller på grunn av brukernes forespørselsprofil (brukerprofilen).

Parameter

Brukt om observert størrelse, f.eks. temperatur, dybde, osv., eller en størrelse som er beregnet ut fra observerte størrelser.

NOD

Norsk Oseanografisk Datasenter

UIB

Universitetet i Bergen.

2. MÅLFORMULERING FOR SYSTEMET

Systemet skal :

1. motta korrigerede måledata fra sine omgivelser.
2. automatisk finne fram til en passende lagringsplass for mottatte data
3. omformattere de mottatte data til det lagringsformat som brukes i systemet for den aktuelle datatypen.
4. Automatisk vedlikeholde en oversikt over de data som er lagret i systemet og hvor de er lagret
5. holde oversikt og kontroll med systemets brukere og deres bruk av systemet
6. automatisk fremhente data som etterspørres. Man må imidlertid være forberedt på forespørsler som er så spesielle at en kontrollert ("manuell") framhenting og beregning må foretas
7. holde oversikt over hvor ofte de enkelte datatyper og parametre framhentes
9. periodisk gi rapporter om systemets tilstand og bruk.

3. TEKNISKE FORHOLD

A. De data som NOD administrerer har blant annet følgende egenskaper:

1. Datamengden er stor
2. En gitt (vilkårlig) datamengde etterspørres sjelden.
3. Kravene til ekspedisjonstid er som regel av størrelsesorden dager.

Dette peker tydelig mot at data bør lagres på tape.

En konsekvens av dette er at framhenting av data vanligvis ikke kan foretas direkte fra terminal. Den del av systemet som holder oversikt over lagringssted for data, skal også kunne generere den jobb som er nødvendig for å hente dataene fra tape. Systemet blir implementert på dataanlegget ved Universitetet i Bergen, som har en UNIVAC-1110 maskin.

B. Fra de parametre som observeres er det ofte mulig å beregne andre parametre.

Slike "beregnete parametre" vil også etterspørres med varierende teidsfrekvens av brukerne. En beslutning som periodisk må foretas i systemet er dermed:

- Hvilke beregnede parametre skal lagres eksplisitt, og hvilke skal beregnes for hver gang de etterspørres?

Systemet må vite hvilke parametre som kan beregnes ut fra de lagrede parametre for å kunne gi svar på brukernes spørsmål. Videre må systemet overvåke framhentingsfrekvensen for de enkelte parametre slik at man lettere kan foreta beslutningen om hvilke beregnede data som skal lagres eksplisitt.

4. DELSYSTEMER :

På bakgrunn av målformuleringen, de tekniske forhold og objekt-systemets virkemåte (se graf 0), avtegner det seg 3 delsystemer :

- 7 Lagringssystem som består av de prosesser og den informasjon som er nødvendig for lagring av nye data.

- 9 Arkivsystem som består av de prosesser og den informasjon som er nødvendig for å administrere:
 - plassering av nye data
 - oversikt over systemets brukere og den bruk de gjør av systemet
 - oversikt over de data som finnes i systemet
 - angi prosedyre for framhenting av etterspurte data.

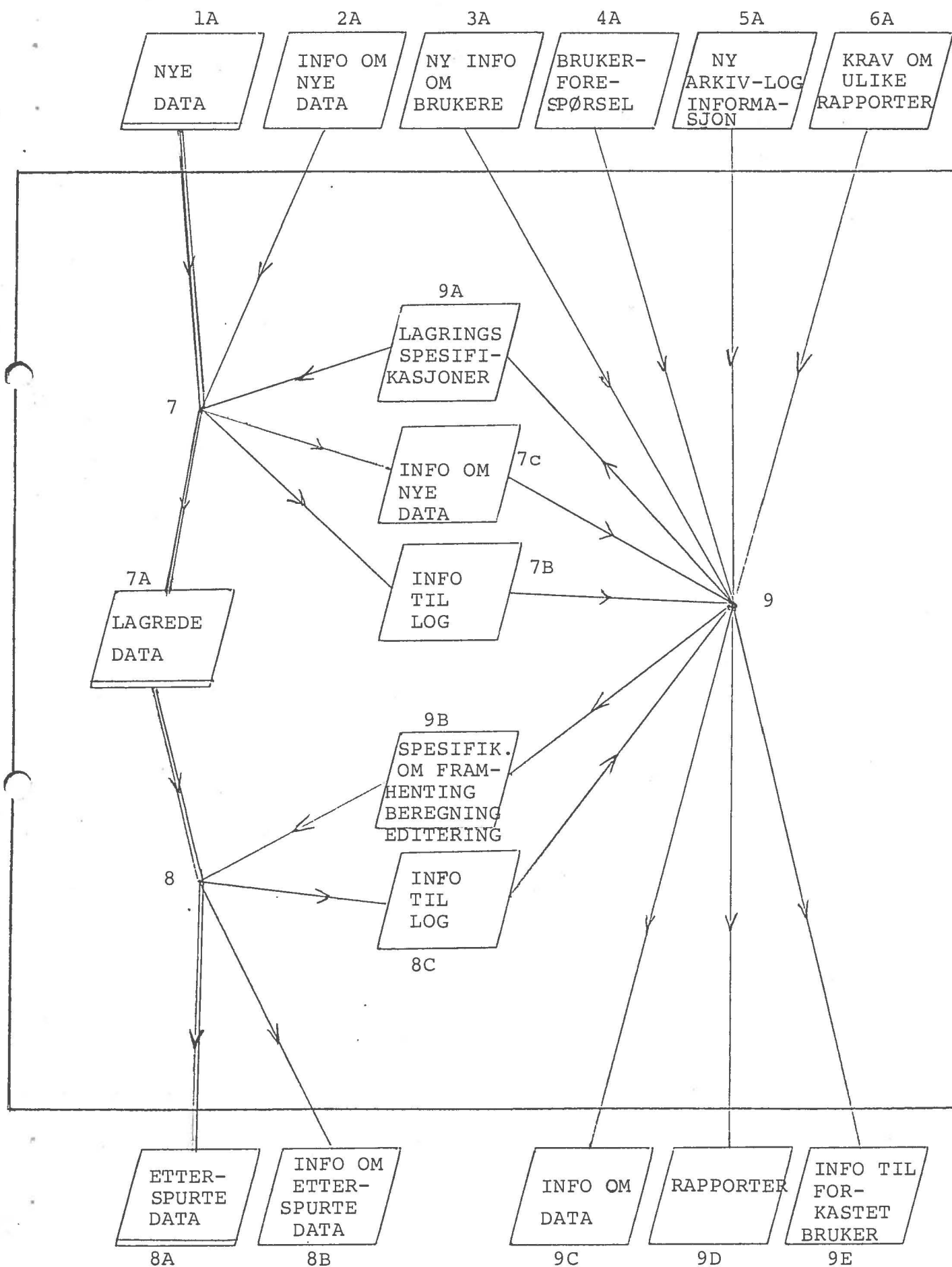
- 8 Framhentingssystem som består av de prosesser og den informasjon som er nødvendig for å
 - hente fram data fra angitt lagringsposisjon
 - utføre aktuelle beregninger på framhentede data (se "tekniske forhold")
 - redigere og levere de etterspurte data til brukeren.

Numrene 7, 8 og 9 er brukt som merking på grafene.

Kommentarer til grafene finnes bak grafene.

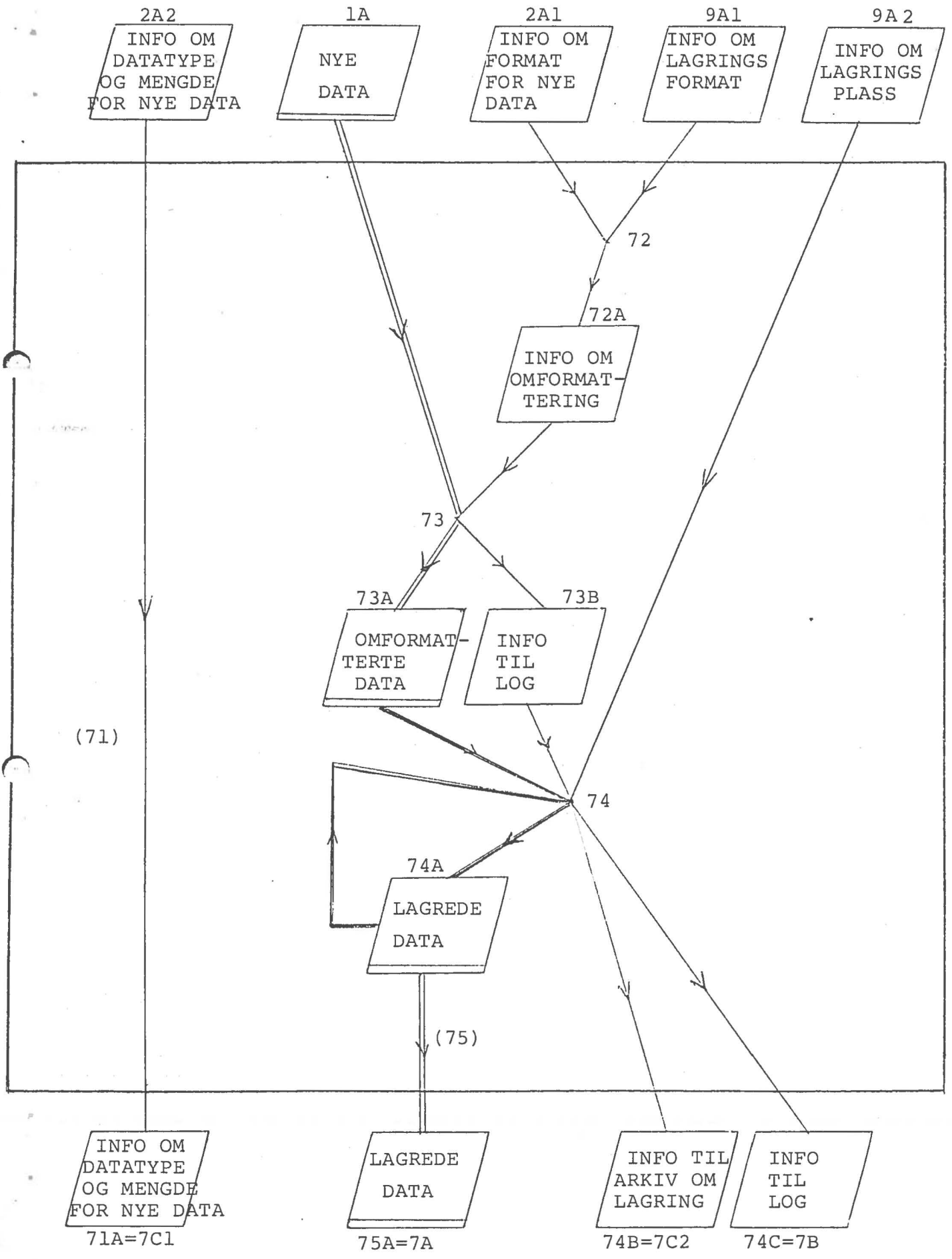
UTFØRT AV	A.H.	DATO: 1/11-76	PRESEDENS	-GRAF
SYSTEM			GRAF-NR	0
LAGRING/FRAMHENTING/ARKIV			VERSJON	1

SIDE: 6



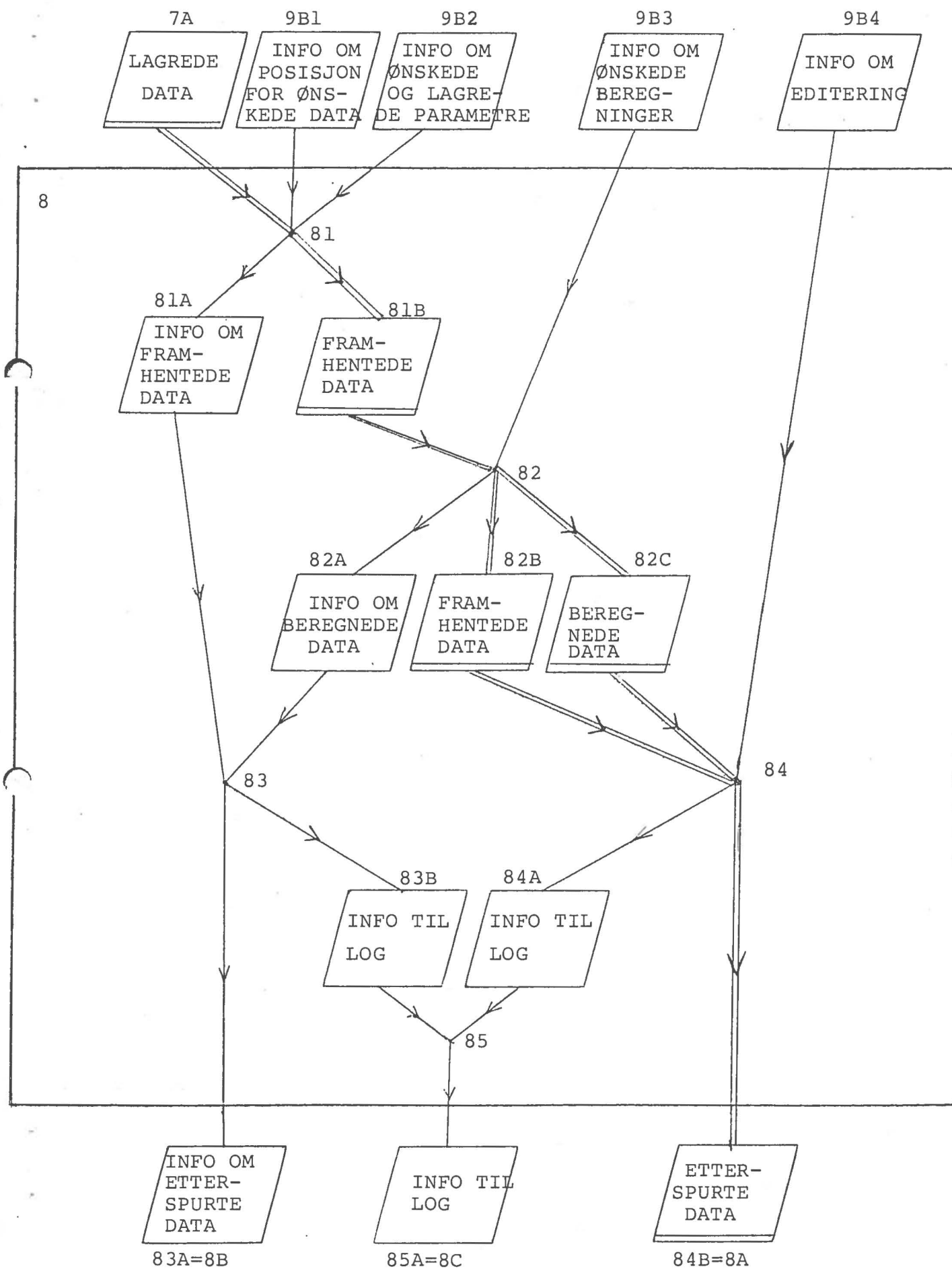
UTFØRT AV	A.H.	DATO: 1/11-76	PRESEDENS	-GRAF
SYSTEM			GRAF-NR	7
LAGRING			VERSJON	4

SIDE: 7

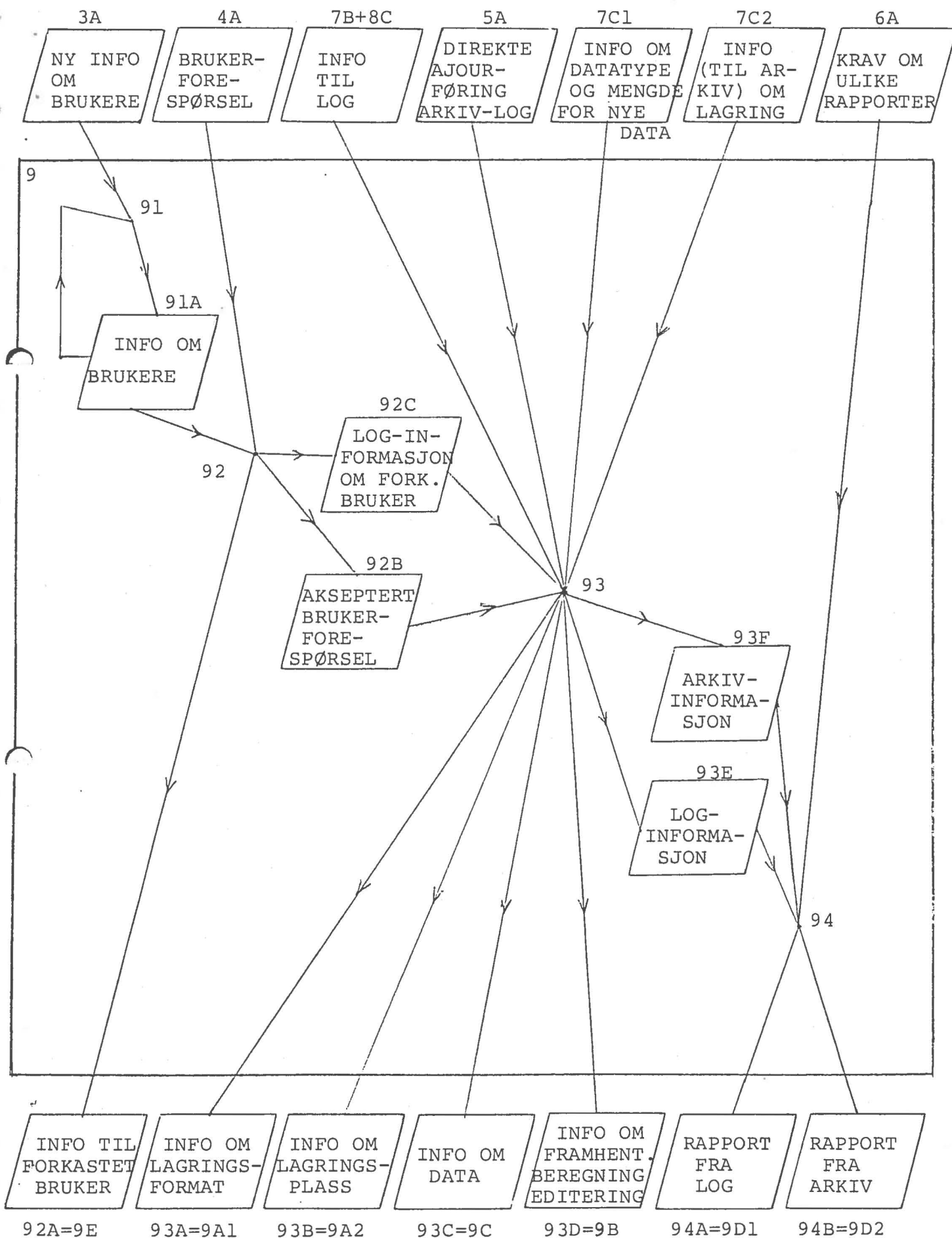


UTFØRT AV	A.H.	DATO: 1/11-76	PRESEDENS	-GRAF
SYSTEM			GRAF-NR	8
FRAMHENTING OG BEREGNING			VERSJON	3

SIDE: 8

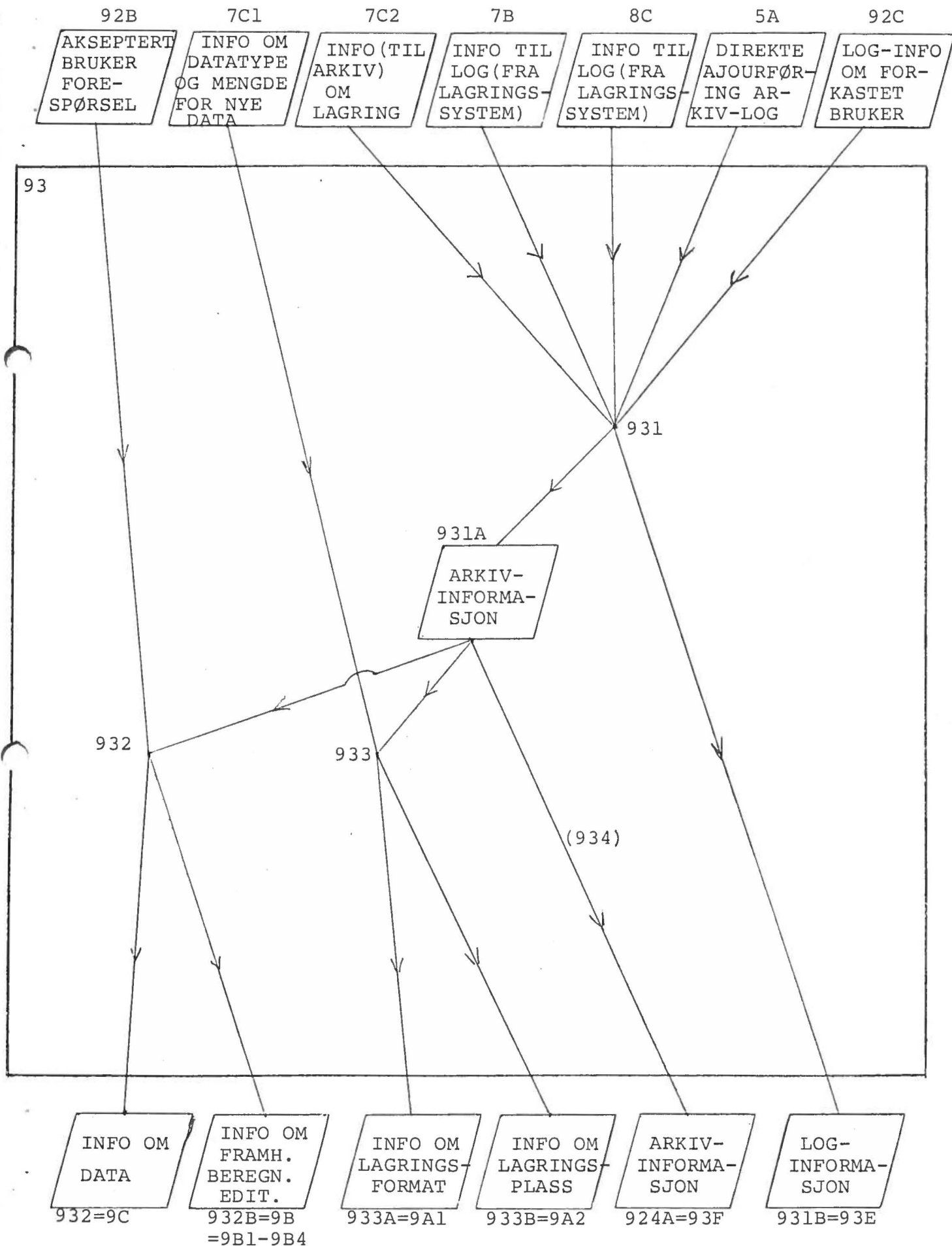


UTFØRT AV	A.H.	DATO: 1/11-76	PRESEDENS	-GRAF
SYSTEM			GRAF-NR	9
ARKIV-LOG			VERSJON	3
				SIDE: 9



UTFØRT AV	A.H.	DATO: 1/11-76	PRESEDENS	GRAF
SYSTEM			GRAF-NR	93
ARKIV-LOG			VERSJON	3

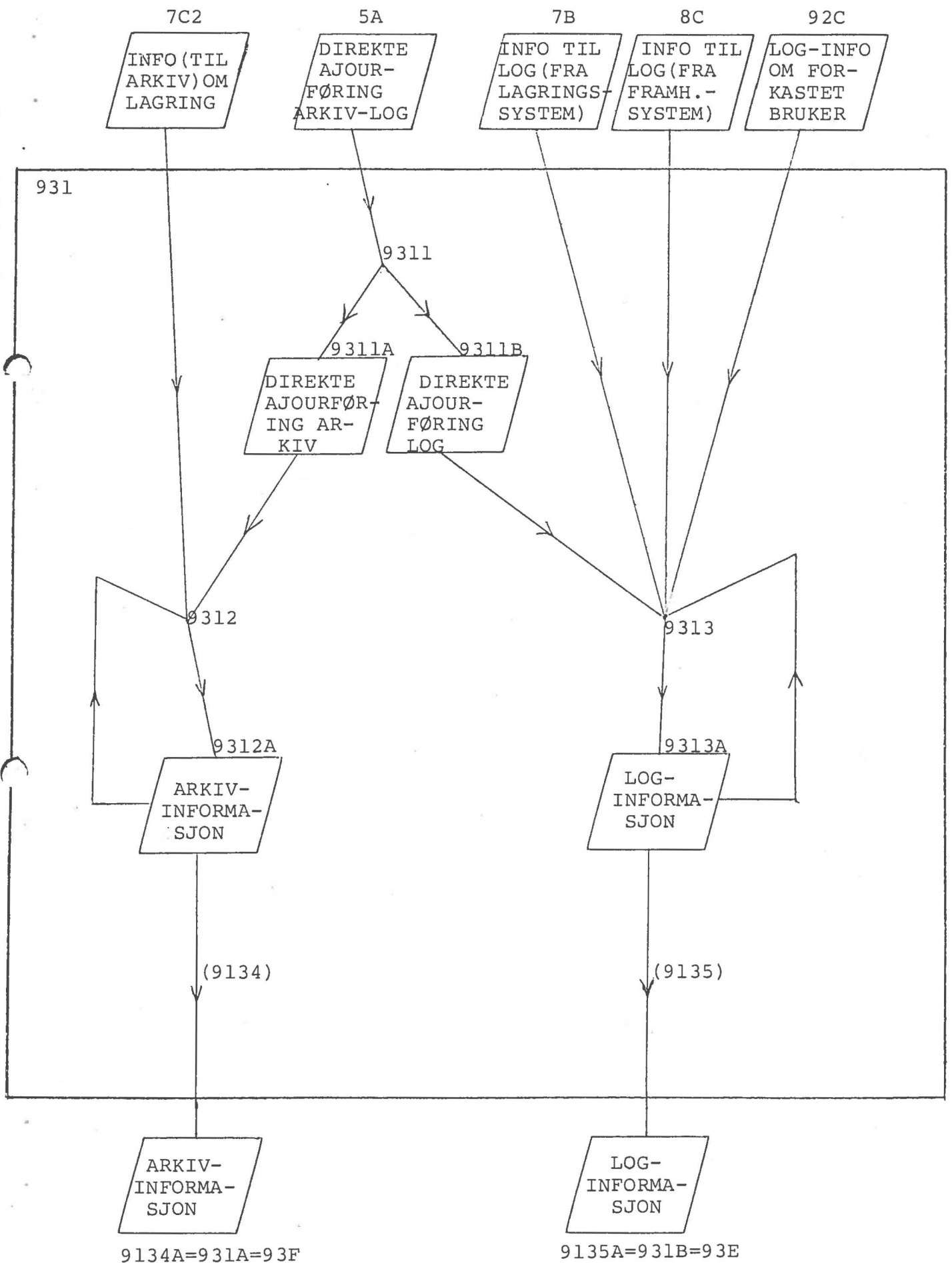
SIDE: 10



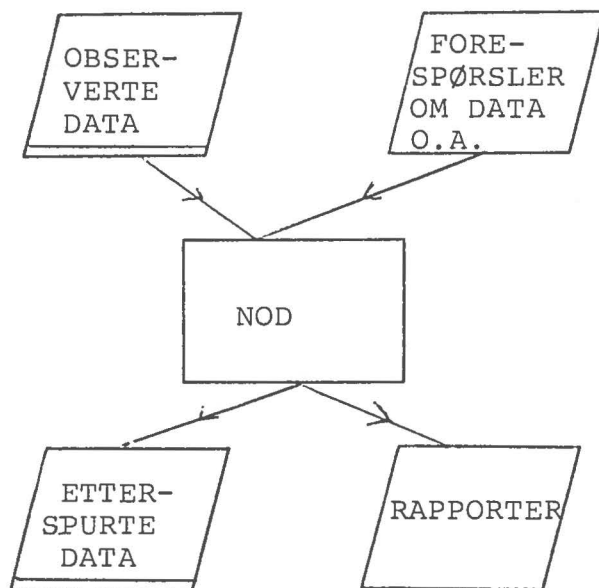
KOMM: Ingen nye mengder eller dekomponeringer

UTFØRT AV	A.H.	DATE: 12/11-76	PRESEDENS	GRAF
SYSTEM			GRAF-NR	931
	ARKIV-LOG		VERSJON	1

SIDE: 11



4.1 KOMMENTAR GRAF-NR 0 -objektsystemts virkemåte



Grafene ovenfor er en enkel skisse av NOD's kontaktflater med omgivelsene.

- Data observeres av en eller flere institusjoner som arrangerer tokt/ekspedisjoner. De observerte data kan gjennomgå bearbeidelse og korreksjoner før NOD kommer i befatning med dem. Data er av ulik kvalitet alt etter hvilket måleinstrument som er brukt, hvilken institusjon som er ansvarlige for datas opprinnelse osv.
- Videre mottar NOD forespørsler om data fra interesserte personer/-institusjoner. Disse forespørslene skal behandles av det framtidige system.
- På bakgrunn av de data NOD rår over og de forespørsler som kommer inn, avgir NOD etterspurte data. Foruten dette vil NOD også periodevis eller på forespørsel utgi rapporter bl.a. om de data som er tilgjengelige.

Graf 0 viser noe mer detaljert de deler av objektsystemet som er av interesse i forbindelse med det framtidige system.

- De NYE DATA (1A) ankommer systemet sammen med en del INFORMASJON OM NYE DATA (2A).
- På bakgrunn av 2A gir lagringssystemet (7) beskjed til arkivsystemet (9) om ankomsten av 1A (7C).
- 7C gir den informasjon som arkivsystemet trenger for å finne ut hvor disse nye data skal lagres og hvordan de skal lagres. Dette resulterer i LAGRINGSSPESIFIKASJONER (9A).
- 1A, 2A og 9A går nå sammen i prosess 7 der lagringen foretas (mengden 7A ajourføres), og rapport om lagring gis til arkivet via 7B.
- Systemet skal holde oversikt over sine brukere og derfor oppstår mengden 3A i INFO OM NYE BRUKERE.
Den består av nødvendig informasjon for registrering av nye brukere og for sletting av gamle brukere.
Dessuten kan visse opplysninger om eksisterende brukere forandres.
- 5A og 6A er opphav til direkte ajourføring av arkivet (5A) og initiering av rapporter (6A).
- Brukerne stiller spørsmål (4A) til arkivet.
Disse spørsmål er av flere typer:
 - Krav om framhenting av gitte datamengder.
 - Spørsmål om eksistens av gitte datamengder.
 - Spørsmål fra systemets eiere (NOD) av administrative årsaker.
- I de tilfeller at 4A initierer en framhenting av data, vil arkivsystemet (9) finne ut hvor disse data ligger, hvilke begrensninger som må foretas osv. Dette gir opphav til 9B.

Beregninger:

Det er ikke sikkert at systemet har eksplisitt lagret de parametre som brukeren spør etter.

I de tilfeller aktuelle parametre ikke er lagret må disse beregnes ut fra de lagrede parametre. Dette skal systemet som regel kunne utføre automatisk.

På bakgrunn av den LOG som tas av systemets bruk vil man via de opplysninger man får fra 9D være i stand til å finne ut hvilke parametre som etterspørres oftest. Dette kan resultere i at man finner der formålstjenlig å eksplisitt lagre parametre som tidligere ble beregnet ut fra andre parametre (fordi brukerne så ofte spør etter disse). Man har flere eksempler på slike parametre i dag (f.eks. SIGMA-T).

- 9B går inn i framhentingssystemet (8) der de nødvendige operasjoner foretas for å produsere 8A og 8D.
Etter avslutning av disse prosessene rapporteres tilbake til arkivet via 8C.
- 9C: INFO OM DATA

Ikke alle forespørsler forårsaker framhenting av data fra 7A. Svaret på slike forespørsler som bare omhandler opplysninger om data gis i 9C.

9D : er ulike rapporter om systemets tilstand og bruk.

9E : gis til bruker som av en eller annen grunn nektes adgang til systemet får beskjed.

4.2 GRAF-NR 7 Lagring

7C : INFO OM NYE DATA

2A1 : INFO OM FORMAT FOR NYE DATA

2A2 : INFO OM DATATYPE OG MENGDE FOR NYE DATA

9A : LAGRINGSSPESIFIKASJONER

9A1 : INFO OM LAGRINGSFORMAT

9A2 : INFO OM LAGRINGSPLASS

1A : NYE DATA

Korrigerte data av ulike datatyper ankommer systemet.

Disse data er av varierende format og kvalitet.

2A1 : INFO OM FORMAT FOR NYE DATA

De nye data kan befinne seg i et annet format enn det format de skal lagres i. (Jmfr. 9A1 og 72).

2A2 : INFO OM DATATYPE OG MENGDE FOR NYE DATA

Systemet må vite hva slags data og hvor mye data som er inneholdt i 1A. Det er på bakgrunn av denne informasjon at 9A2 bestemmes.

71C=7C1 : INFO OM DATATYPE OG MENGDE FOR NYE DATA

Denne mengden er presedent til de prosesser i arkivsystemet som framfinder en passende tape for å lagre de nyankomne data. Denne delprosessen trenger informasjon om datatyper for å plassere de nyankomne data "nært" tilsvarende data i systemet, og informasjon om mengde for å finne en tape med plass nok til de nyankomne data.

9A1 : INFO OM LAGRINGSFORMAT

Data skal lagres i et spesielt format, og hvilket format gis informasjon om fra arkivsystemet.

9A1 er indirekte suksedent til 7C1.

9A2 : INFO OM LAGRINGSPLASS

Spesifiserer hvilken tape og fil de nyankomne data skal lagres på.

9A2 er indirekte suksedent til 7C1.

75A=7A : LAGREDE DATA

En ajourført mengde av eksisterende lagrede data. Disse data er kjent av arkivsystemet, og kan automatisk framhentes.

74B=7C2 : INFO (TIL ARKIV) OM LAGRING

Etter at nyankomne data er lagret i systemet sendes 7C2 som bekreftelse til arkivsystemet. De nye data registreres dermed i arkivet.

74C=7B : INFO TIL LOG

Aktiviteten i forbindelse med omformattering (73) og lagring (74) registreres i LOG for kontroll.

4.3 GRAF-NR 8. FRAMHENTING OG BEREGNING

9B : SPESIFIKASJON OM FRAMHENTING, EDITERING, BEREGNING

9B1 : INFO OM POSISJON FOR ØNSKEDE DATA

9B2 : INFO OM ØNSKEDE OG LAGREDE PARAMETRE

9B3 : INFO OM ØNSKEDE BEREGNINGER

9B4 : INFO OM EDITERING

7A : LAGREDE DATA

Se graf 7.

9B1 : INFO OM POSISJON FOR ØNSKEDE DATA

Grunnlaget for de data som etterspørres befinner seg på en eller flere gitte taper og filer.

9B2 : INFO OM ØNSKEDE OG LAGREDE PARAMETRE

Man vil behøve noen av de lagrede parametre sammen med de parametre som beregnes for å tilfredsstille kravet til de etterspurte data (8A).

9B3 : INFO OM ØNSKEDE BEREGNINGER

Generelt sett vil noen parametre måtte beregnes for å tilfredsstille etterspørselen etter data (8A).

9B4 : INFO OM EDITERING

Brukeren vil ha data (8A) redigert på bestemt måte, eller det finnes en systembestemt editering.

81A : INFO OM FRAMHENTEDE DATA

Består av aktuell informasjon til brukeren (f.eks. kvalitet på data) og systembestemt informasjon til LOG.

81B : FRAMHENTEDE DATA

Består av de parametre som er et tilstrekkelig grunnlag for å tilfredsstille brukerforespørselen.

82A : INFO OM BEREGNEDE DATA

Analogt til 81A.

82B : FRAMHENTEDE DATA

Består av 81B men minus de parametre som i 81B bare var nødvendig for å få utført beregningene som var påkrevet.

82C : BEREGNEDE DATA

Består av de parametre som måtte beregnes for å tilfredsstille brukerforespørselen.

82B og 82C utgjør til sammen 8A som brukeren har etterspurt.

83A : INFO TIL LOG

84A : INFO TIL LOG

85A=8C INFO TIL LOG

Den aktuelle informasjon som skal lagres for administrasjon i LOG. Omhandler bl.a. datatype framhentet, mengde, tidspunkt, osv.

8A : ETTERSPURTE DATA

De data eller nærmeste match til de data som brukeren har etterspurt.

4.4 GRAF-NR 9 ARKIV-LOG

9D : RAPPORTER

9D1 : RAPPORT FRA LOG

9D2 : RAPPORT FRA ARKIV

TEMPORÆR KONSOLIDERING :

93D=9B=9B1-9B4: INFO OM FRAMHENTING- BEREGNING- EDITERING

7B+8C : INFO TIL LOG

3A : NY INFO OM BRUKERE

Inneholder informasjon om nye brukere og forandringer av status for eksisterende brukere.

Disse forandringene går bl.a. ut på:

- temporær utelukkning
- nye tider på dagen der tilgang er tillatt osv.

4A : BRUKERFORESPØRSEL

En bruker kommer inn i systemet med en forespørsel og identifiserer seg samtidig for systemet.

91A : INFO OM BRUKERE

Består av :

- brukeridentifikasjon (passord)
- tidsbegrensning (dato der tilgang ikke lengre aksepteres)
- tidsintervall på dagen der bruker har tilgang
- status (temporært åpent eller lukket)
- registreringsdato
- antall ganger brukeren har brukt systemet
- personlig informasjon (navn, adresse, telefon osv.)

92A=9E : INFO TIL FORKASTET BRUKER

Dersom brukeren er ukjent for systemet eller de øvrige betingelser (se 91A) ikke er oppfylt, forkastes brukeren.

92B : AKSEPTERT BRUKERFORESPØRSEL

Dersom brukeren er kjent av systemet og øvrige betingelser (se 91A) er oppfylt, aksepteres brukeren og hans forespørsel.

92C : LOG-INFORMASJON OM FORKASTET BRUKER

Alle brukere som forkastes merkes av i LOG for system administrasjon.

7D+8C : INFO TIL LOG

Se graf 7 og graf 8.

5A : DIREKTE AJOURFØRING ARKIV LOG

I praksis kan det bli behov for å ajourføre arkivet eller LOG direkte. Dette oppstår bl.a. ved tap av data, ikke-systembestemt innlegging av data osv. Dette gir en viss fleksibilitet i systemet for å kunne takle uventede eller unormale situasjoner.

7C1, 7C2 : se graf 7.

6A : KRAV OM ULIKE RAPPORTER

Systemets administratorer vil ha behov for å få rapporter om driften og situasjonen både periodisk og på kommando.

93A=9A1 : INFO OM LAGRINGSFORMAT

93A=9A2 : INFO OM LAGRINGSPLASS

Se graf 7.

93C=9C : INFO OM DATA

Enkelte brukerforespørsler vil ikke nødvendigvis betinge en framhenting av data, men gå på informasjon om de data som er lagret. (F.eks. om en viss datatype fra et visst geografisk område eksisterer i systemet).

93D=9B=9B1-9B4 : INFO OM FRAMHENTING, BEREGNING, EDITERING

Se graf 8.

93E : LOG INFORMASJON

Interessante hendelser i systemet er registrert her. F.eks. resultatet av en framhenting, innleggelsen av nye data osv.

93F : ARKIV INFORMASJON

Her finnes alle opplysninger som er nødvendig for å bestemme hvilke data en bruker er interessert i. Avgjøre om disse data finnes og bestemme nøyaktig hvor de finnes, samt informasjonen om hvilket format data er lagret i. Dessuten holdes oversikt over hvilke parametre som kan beregnes ut fra lagrede parametre, og en slik beregning kan automatisk foretas i forbindelse med framhenting.

94A=9D1 : RAPPORT FRA LOG

De hendelser systemet har gjennomgått den siste tiden blir her rapportert.

94B=9D2 : RAPPORT FRA ARKIV

Systemets tilstand og innhold rapporteres her.

5. DATABASE LØSNINGEN

Arkivsystemet egner seg bra for en database løsning. Grunnene til det er flere, og noen kan nevnes her:

- Arkivet vil bestå av flere ulike objekttyper (noe man vil samle informasjon om).
- Det vil eksistere mange, og til dels kompliserte, relasjoner mellom disse objekttypene.
- Programmering av spesial rutiner for å vedlikeholde struktur og informasjonsinnhold vil bli tidkrevende og kostbar.

I og med at man ved dataanlegget ved UIB har tilgjengelig et generelt databasesystem (DMS-1100) som er i stand til å vedlikeholde nettverksstrukturer er det rimelig å satse på dette systemet. Man håper dermed å få vesentlig lavere design- og implementeringskostnader enn dersom man satser på en spesialløsning av arkivsystemet.

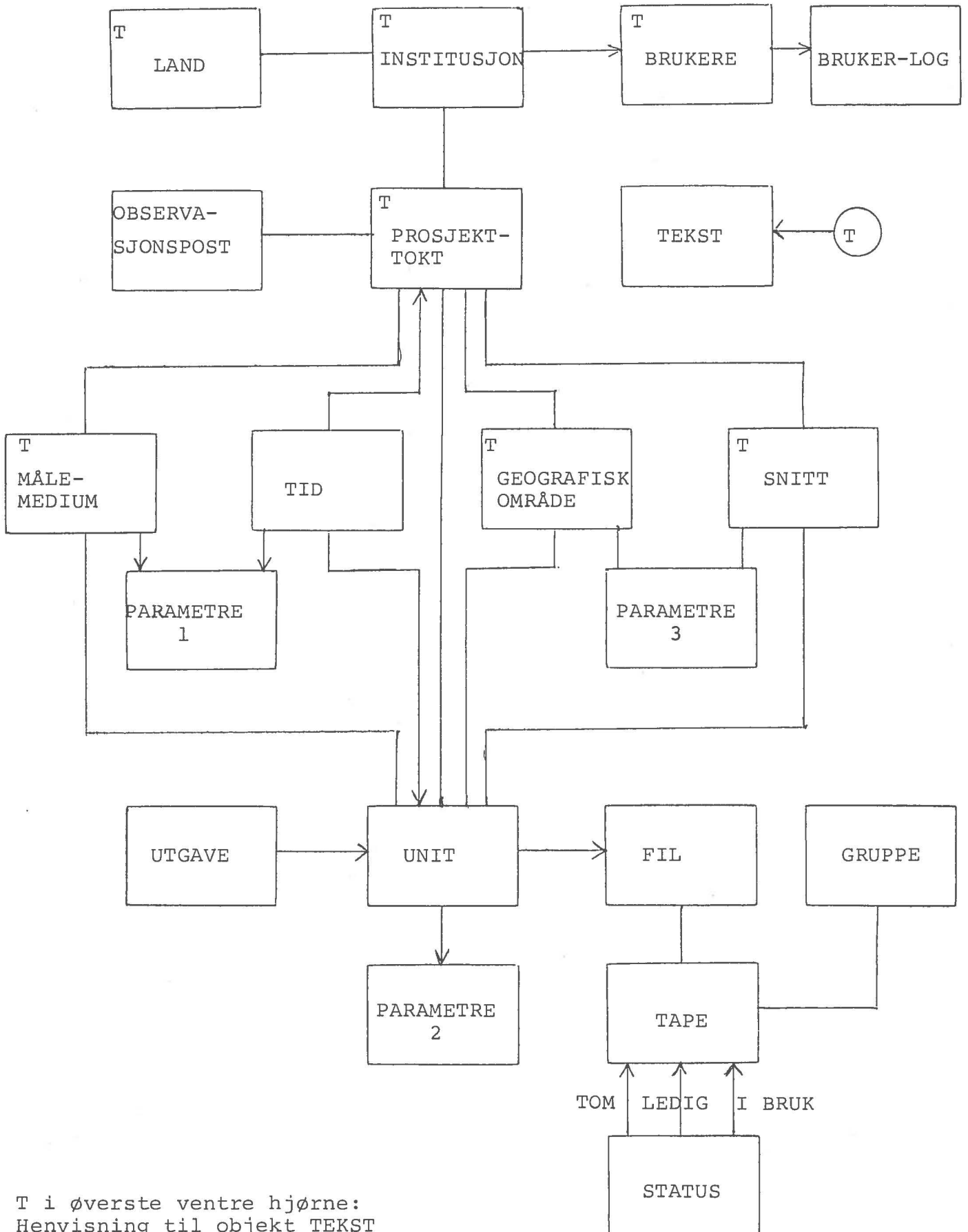
Det som heretter kalles "databasen" vil da inneholde informasjon om (bl.a.) :

- Systemets brukere
- systemets lagrede data
- en vesentlig del av det som tidligere er kalt log-informasjon med unntak av :
 - rapporter om mislykkede kjøring (som skal rapporteres direkte til systemets administrasjon)
 - forkastede brukere
 - øvrig informasjon som ikke har direkte betydning for systemets drift.

Bak følger nå en graf over denne databasen samt kommentarer til denne. Framstillingen må oppfattes som

en mellomting mellom en infologisk modell og databasemodell og et grunnlag for videre systemering.

"INFOLOGISK MODELL" FOR ARKIVSYSTEM + DATABASE/GRAF AS.



T i øverste ventre hjørne:
Henvisning til objekt TEKST

5.1 KOMMENTARER GRAF AS

Relasjonenes retning :

- a) I de tilfeller at relasjonene er gitt retning skal dette oppfattes som at man her har funnet å kunne angi relasjoner ut fra logiske grunner.

- b) De relasjoner som ikke er gitt retning er relasjoner som på nåværende tidspunkt ikke er tilstrekkelig gjennomdiskutert til at man kan uttale seg om retning.
Flere av disse relasjonene kan altså være "mange til mange" relasjoner.

Flere av relasjonene er heller ikke navngitt. Grunnen til det er bl.a. punkt b ovenfor, men også at man med herværende graf har bare en relasjon mellom de fleste objekttyper. Man skal derfor ikke utelukke at arkivsystemets struktur kan framstilles hierarkisk, noe som kanskje kan forenkle programmering og vedlikehold.

I tabell 1 neste side er det satt opp en summerisk oversikt over hvilken informasjon man kunne tenke seg å ha bruk for i arkivet. Hensikten med tabellen er å få en følelse av arkivsystemets utseende.

Med bakgrunn i tabell 1 og graf AS kan følgende skisse av arkivsystemets informasjon settes opp (tabell 2). (se 4.3)

4.2 TABELL 1

01 ARKIVINFORMASJON OM DATA.

- 02 MÅLEMEDIUM/OBJEKTTYPE (sjøvann, ferskvann, fisk...)
- 02 POSISJON I MÅLEMEDIUM (lever, rogn, overflate...)
- 02 FAGOMRÅDE
- 02 DATAS OPPRINNELSE
 - 03 LAND
 - 03 BÅT - OBSERVASJONSPOST
 - 03 FAST ELLER MOBIL OBSERVASJONSPOST
 - 03 INSTITUSJON
 - 03 PROSJEKT
 - 03 TOKTIDENTIFIKASJON
 - 03 SNITTIDENTIFIKASJON
- 02 MÅLEINSTRUMENT
- 02 ANALYSEMETODIKK
- 02 MÅLEFREKVENS (kontinuerlige eller diskont. målinger)

- 02 GEOGRAFISK POSISJON (FOR MÅLEMEDIUM/OBJEKT)
- 02 TIDSINTERVALL (ÅR, DATO1 - DATO2)
- 02 KVALITETSKODE
- 02 FRIE ELLER BEGRENSEDE DATA (m.h.t. utlevering)
- 02 PARAMETRE
 - 03 LAGREDE PARAMETRE
 - 04 TAPE-IDENTIFIKASJON (kan være flere taper)
 - 04 FILNUMMER PÅ TAPE (der data er lagret)
 - 04 BLOKKNUMMER
 - 04 FORMAT
 - 05 HEADING (Makroopplysninger for en måleserie)
 - 05 PARAMETRE
 - 04 FRAMHENTINGSRUTINE
 - 04 SØKEMETODE (ev. henvisning til register e.l.)
 - 04 LAGRET DATA, TID
 - 04 ANTALL GANGER FRAMHENTET
 - 04 SIST FRAMHENTET (dato, tid)
 - 04 MENGDE DATA

TABELL 1 forts.

- 03 PARAMETRE SOM KAN BEREGNES fra de lagrede parametre
- 04 RUTINE FOR BEREGNING
- 04 INFORMASJON SOM BEREGNINGSRUTINEN TRENGER

01 INFORMASJON OM TAPER

- 02 LEDIGE TAPER (antall)
 - 03 TAPE-ID
 - 03 LEDIG Plass (mengde)
- 02 TAPER I BRUK (antall)
 - 03 TAPE-ID
 - 03 LEDIG Plass
 - 03 (BRUKT Plass)
 - 03 SIST BRUKT (dato, tid)
 - 03 ANTALL FILER PÅ TAPEN

5.3 INFORMASJONS-INNHOLD/GRAF AS

RG = "repeterende gruppe", REF = Referanse

UNIT = en identifikasjon på en gruppe filer med identisk innhold

TABELL 2

OBJEKT	EGENSKAP	KOMMENTAR
LAND	KODE-LAND	Kode for land.
INSTITUSJON	KODE-INST	Kode for institutt.
	REF. TIL TEKST	Institutt navn, leder, osv.
PROSJEKT- TOKT	KODE-PROSJEKT	
	REF/TEKST	Prosjektleder, finansiering, kontakt- personer, osv.
	STARTDATO	For prosjekt.
	TIDSRAMME	
	AVSLUTN.DATO	
	KODE-TOKT (RG)	
	REF/TEKST	Toktleder, kontaktpersoner, osv.
	STARTDATA	For tokt.
	AVSLUTN.DATO	
	ANT-UNITS	Antall UNITS med data innen PROSJEKT/ TOKT.
OBSERVA- SJONSPOST	KODE	
	TYPE	Båt, bøye,....
	NAVN	
BRUKERE	FAST/MOBIL	- observasjonspost
	BRUKERID/PASSORD	
	DEADLINE	Dato da brukeren ikke lenger har til- gang.
	TID1-TID2	
	TID1-TID2	Tidsintervall på dagen da brukeren har tilgang (kan bli aktuell dersom NOD vil reservere systemet til egen bruk)

TABELL 2 forts.

OBJEKT	EGENSKAP	KOMMENTAR
(BRUKERE)	STATUS	Temporært åpent eller lukket for bruker.
	REGISTRERINGSDATO	
	ANT-BRUKT	Antall Log-on for denne bruker.
	SIST-BRUKT	Dato og tid for siste log-on
TEKST	REF-INFO	Referanse til personlig informasjon, navn, telefon, ...
		Uformalisert tekst til hjelp for menneskelige brukere.
LOG		Identifikasjon av hvilke data brukerne har hentet fra systemet osv.
MÅLEMEDIUM	KODE-MEDIUM	
	REF-TEKST	
	POSISJON I MEDIUM-(RG)	Eks. lever, rogn,.. for MEDIUM fisk.
PARAMETRE1	KODE-PARAMETRE	
	DATAMENGDE	Innenfor et målemedium.
	ANT-MÅLINGER	" " " "
SNITT	KODE-SNITT	Et snitt er en definert linje mellom to geografiske punkter, eks. FEIE-SHETLAND.
	ANT-UNITS	Antall UNITS.
	ANT-REF	Antall ganger etterspurt.
	SIST	Siste gang etterspurt.
GEOGRAFISK OMRÅDE	REF-TEKST	
	KODE-GEO	Et geografisk område kan være f.eks. MSQ-rutennummer, geografisk område (eks. Haltenbanken) osv.
	REF-TEKST	
	ANT-UNITS	Antall UNITS

TABELL 2 forts.

OBJEKT	EGENSKAP	KOMMENTAR
SIST PARAMETRE3	SIST	Siste gang etterspurt.
	KODE-PARAMETRE	(Innenfor et
	DATAMENGDE	(Innenfor et SNITT eller GEOGRAFISK OMRÅDE)
	ANTALL MÅLINGER	(Innenfor et SNITT eller GEOGRAFISK OMRÅDE).
TID	KODE-TID	F.eks. årstall, årstid,....
	ANT-UNITS	Antall UNITS
	ANT-REF	Antall ganger etterspurt
UNIT	SIST	Siste gang etterspurt
	UNIT-ID	En unit er en samlet identifikasjon på filer med identisk innhold (m.a.o. er grunnfil+backupfiler). Det er likegyldig for en bruker hvilken av disse filer (innen UNIT) han får data fra.
	DATO1-DATO2	Tidsintervall for data i UNIT.
	BREDDE1-BREDDE2	Intervall/breddegrad for data i UNIT,
	LENGDE1-LENGDE2	" /lengdegrad " " " "
	KVALITETSKODE	For data i UNIT.
	STATUS	Frie eller begrensede data m.h.t. utlevering.
	DATAMENGDE	F.eks. i ord.
	ANT-OBS	Antall observasjoner i UNIT
	ANT-PARAMETRE	Antall parametre i UNIT
	ANT-STASJ.	Antall stasjoner/måleserier
	RUTINE	Eventuell/spesiell leserutine for filen
	REGISTER	Eventuell henvisning til register over filene for søking før oppslag i fil.
FORMAT	Format spesifikasjon for UNIT.	

TABELL 2 forts.

OBJEKT	EGENSKAP	KOMMENTAR
PARAMETRE2	REF-TEKST	Referanse til tekst.
	KODE-PARAMETER	Innenfor en UNIT
	INSTRUMENT	Måleinstrument
	KODE-METODIKK	(Eks. titrering) for observasjon.
	TYPE	Lagret eller beregnet parameter
	RUTINE	Eventuell beregnings rutine (dersom ikke lagret).
	INPUT	Data som RUTINE trenger for beregning.
UTGAVE	ANT.REF	Antall ganger etterspurt (av bruker).
	SIST	Siste gang etterspurt (data, tid).
		Det kan eksistere flere utgaver av en datatype. F.eks. observasjoner ved standarddyp, observasjoner pr. 10 meter
FIL	FIL-ID	
TAPE	FIL-NR	Hvilket filnummer filen har på TAPE.
	TAPE-ID	
	LEDIG Plass	Mengde (f.eks i tracks) av ledig plass.
	(BRUKT Plass)	
	ANT-REF	Antall ganger brukt.
	SIST BRUKT	Dato og tid.
	ANT-FILER	Antall filer på tapen.
GRUPPE		En framtidig bestemmelse om gruppering av data forventes. I så fall er det fordelaktig å gruppere taper analogt til data for å få et "balansert" lager (Fagområde kan inngå i denne objektypen)

5.4 KOMMENTARER TIL RELASJONER/GRAF AS

TABELL 3

MELLOM OBJEKTER		VISER/KAN VISE
LAND	INSTITUSJON	Hvilke institusjoner tilknyttet hvilke land.
INSTITUSJON	BRUKERE	Hvilke institusjon de ulike brukere er tilknyttet.
BRUKERE	TEKST	Personalialia for en bruker.
BRUKERE	LOG	Den bruk (hvilke data framhentet osv) som bruker har gjort av systemet
INSTITUSJON	PROSJEKT-TOKT	Hvilke prosjekter og tokt hver institusjon har gjennomført og hvilke institusjoner et prosjekt-tokt er tilknyttet
PROSJEKT-TOKT	OBSERVASJONSPOST	Hvilke båter osv. hvert prosjekt har nyttet og hvilke prosjekter-tokt hver båt osv. har deltatt i.
PROSJEKT-TOKT	MÅLEMEDIUM	Hvilke media som er observert i et prosjekt og hvilke prosjekter har observert et gitt medium.
PROSJEKT-TOKT	TID	Hvilke prosjekter/tokt ble gjennomført i et gitt tidsintervall.
PROSJEKT-TOKT	SNITT	Analogt PROSJEKT-MÅLEMEDIUM.
PROSJEKT-TOKT	GEOGRAFISK OMRÅDE	Analogt PROSJEKT-MÅLEMEDIUM/SNITT.
SNITT	PARAMETRE3	Hvilke parametre er observert innen et snitt og hvilke snitt er en gitt parameter er observert innenfor.
GEOGRAFISK OMRÅDE	PARAMETRE3	Analogt til SNITT-PARAMETRE3
MÅLEMEDIUM	PARAMETRE1	Hvilke parametre er observert innen et medium.

TABELL 3 forts.

MELLOM OBJEKTER		VISER/KAN VISE
TID	PARAMETRE1	Hvilke parametre ble observert i et gitt tidsintervall
MÅLEMEDIUM	UNIT	Hvilke UNITS (se tabell 2) som omhandler et målemedium og hvilke målemedia som inngår i en UNIT
PROSJEKT-TOKT	UNIT	Analogt MÅLEMEDIUM-UNIT
SNITT	UNIT	" " "
GEOGRAFISK OMRÅDE	UNIT	" " "
TID	UNIT	Hvilke UNITS som omhandler en gitt tidsperiode
UTGAVE	UNIT	Hvilke UNITS som er av en gitt utgave
UNIT	PARAMETRE2	Hvilke parametre som inngår i en UNIT
UNIT	FIL	Hvilke filer inngår i en UNIT
FIL	TAPE	Hvilke taper inneholder en gitt fil og hvilke filer inneholder en gitt tape
GRUPPE	TAPE	Hvilke grupper (se tabell 2) inneholder en gitt tape og hvilke taper inneholder en gitt gruppe.
STATUS	TAPE	TOM : Hvilke taper er tomme (uten data) LEDIG : Hvilke taper er ikke i bruk (men inneholder data) I BRUK: Hvilke taper er i bruk (og dermed utilgjengelige i øyeblikke)

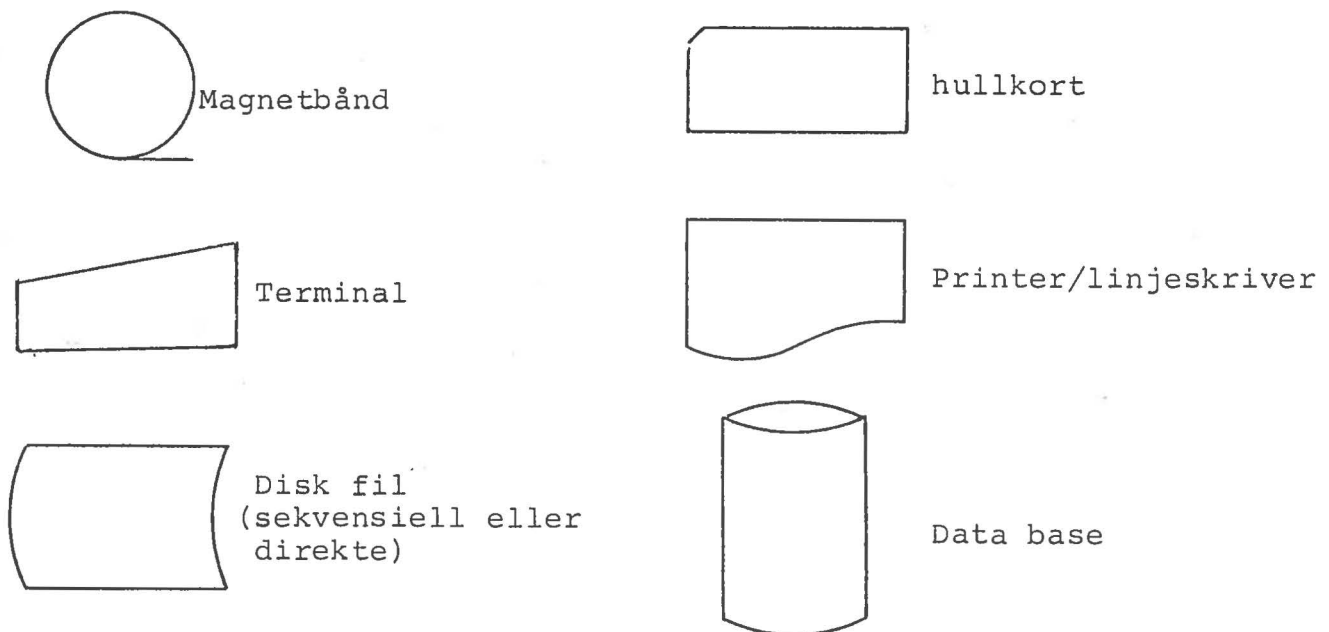
Som tidligere nevnt må det foretas en grundigere analyse for å avgjøre hvilke retninger i relasjonene som er aktuelle, og også om andre relasjoner er aktuelle.

6. KOMPRIMERT SYSTEMOVERSIKT

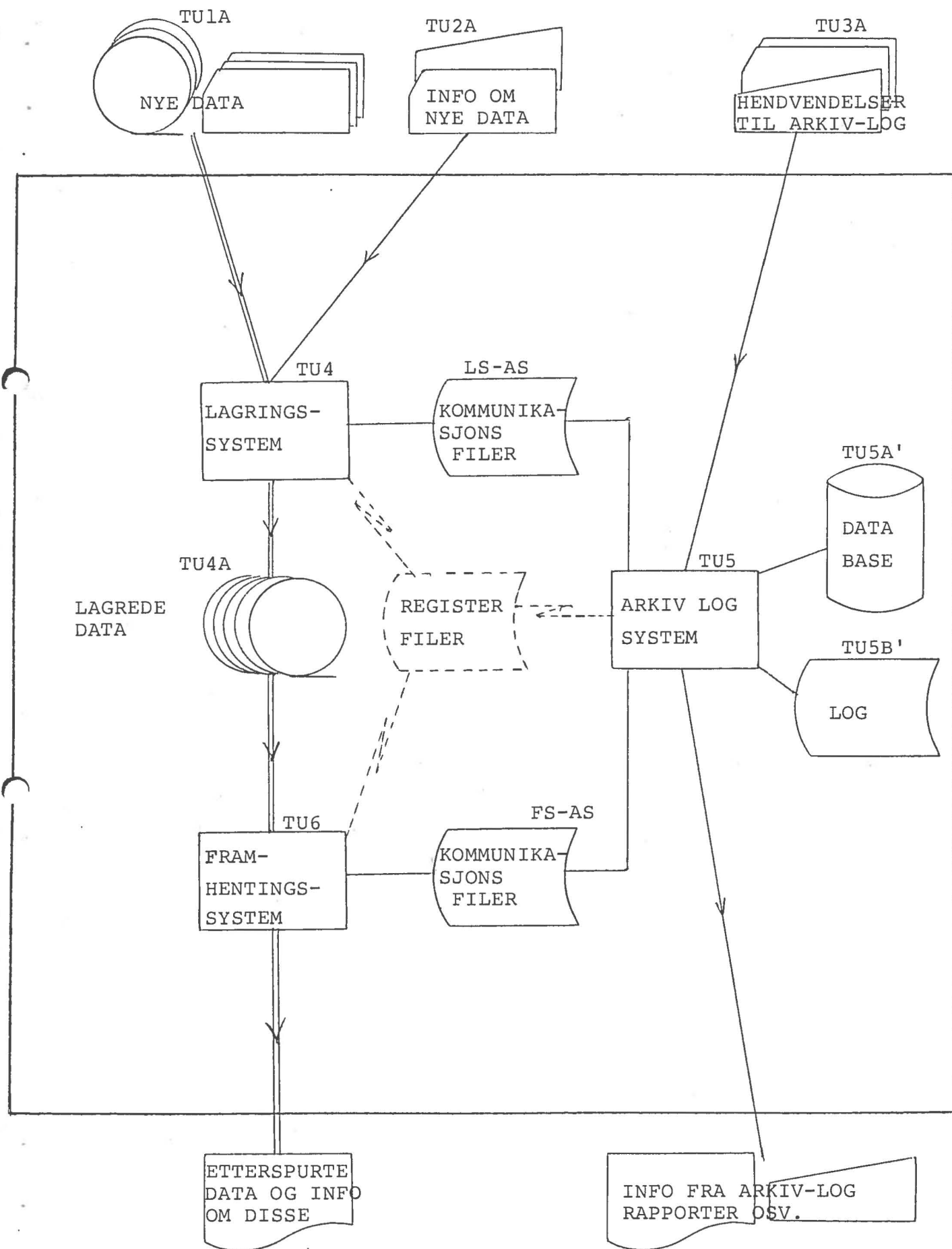
I store trekk befinner systemutviklingen seg på det stadium som er beskrevet foran.

Flere andre spørsmål er diskutert (se kapittel 9).

GRAF TU viser en grov oversikt over systemet på dette stadium. Som vanlig har figurene følgende betydning:



UTFØRT AV	A.H.	DATE: 10/12-76	TEKN.UTRUSTNING	GRAF
SYSTEM		GRAF-NR	TU	
SYSTEMOVERSIKT		VERSJON	2	SIDE:35



KOMM:

GRAF TU er sammenholdt med GRAF 0 selvforklarende.

Informasjonsutveksling mellom systemets deler kan forsinkes med (overlegg) på kommunikasjonsfilene LS-AS og FS-AS. Særlig gjelder dette LOG-informasjon til ARKIV-LOG systemet.

Innholdet i kommunikasjonsfilene LS-AS er i store trekk:

- informasjon til ARKIV-LOG systemet om ankomst, mengde og type for nye data.
- LOG informasjon til ARKIV-LOG systemet fra LAGRINGSSYSTEMET
- Informasjon fra ARKIV-LOG systemet til LAGRINGSSYSTEM om plassering og format for nye data.

Innholdet i kommunikasjonsfilene FS-AS er

- En runstream generert av ARKIV-LOG systemet for å tilfredsstille en bruker forespørsel
- Log-informasjon fra FRAMHENTINGSSYSTEMET til ARKIV-LOG systemet.

Databasen TU5A' (se GRAF AS) inneholder den informasjon som aktuelle rutiner i TU5 trenger for å styre resten av systemet.

REGISTER FILER, som er oppslagsfiler for de egentlige datafilene TU4A, er antydnet i grafen. Det kan ved visse systemutforminger være fordelaktig med slike filer for å minske søkingen i store måledatafiler i TU4A. Dette er imidlertid et spørsmål under diskusjon. Man kan også tenke seg løsninger der tilstrekkelige opplysninger om datas beliggenhet befinner seg i arkivet. Dette vil i så fall øke arkivets størrelse. Andre løsningstyper kan gå ut på å overføre noen av de opplysninger, som i denne rapporten er angitt i arkivet, til disse REGISTERFILENE. Dette vil i så fall minke arkivets størrelse, men kan også føre til en mer komplisert styring av systemet og mer komplisert administrasjonsarbeid for systemets administrator.

Dersom slike registerfiler finnes formålstjenlige, så bør disse også organiseres som en database og kan som sådan ansees som en del av arkivet. Betingelsen er da at den totale databasen organiseres slik at registerfilene ligger adskilt (på andre AREA's) fra det øvige arkiv.

Datamengden i disse filene vil nemlig raskt bli større enn datamengden i resten av arkivet, og vil øke responstiden fra arkivet dersom de ikke skilles.

7. VIDERE SYSTEMERING

Systemeringen har ennå mange uløste spørsmål der noen er av fundamental karakter. Av slike fundamentale avgjørelser kan nevnes (i vilkårlig rekkefølge) :

1. Hvilket format skal man lagre data i (TU4A)?
2. Hvordan skal dette/disse format representeres slik at automatisk formattering, omformattering og initiering av leserutinen kan foretas?
3. Hvilke brukerspørsmål skal arkivsystemet kunne løse?
4. Hvilke forhold skal gjelde mellom Arkivet og de eventuelle registerfilene (se graf TU) - skal registerfiler brukes?
5. Hvilke sorterings- og grupperingskriteria skal gjelde for data? (Disse vil i prinsippet variere med varierende brukerinteresser).
6. Hvilke rapporter over systemets tilstand og dynamikk skal Arkivsystemet kunne generere? (Dette er i prinsippet et underspørsmål av 3).
7. Hvordan klassifiseres/identifiseres (se graf AS)
 - geografiske områder
 - snitt
 - tidsspesifikasjoner
 - utgaver av data (og hvilke standardutgaver skal eksistere)
 - parametre (2)
 - grupper av taper?

I videre sammenheng blir det spørsmål om å konstruere et kodesystem for alle elementer skissert i graf AS.

8. Hvordan skal arkivsystemet behandle brukerforespørsler?
Med andre ord: Hvordan konstrueres effektive utsøkningskriteria og utsøkningsprosesser i arkivet. Dette har særlig betydning ved utsøking etter aktuelle filer for å tilfredsstille en brukerforespørsel.
9. Hvordan skal kommunikasjonsspråket mellom brukerne og arkivet se ut?
Hvordan skal "kommunikasjonsspråket" mellom systemets (automatiske) deler se ut?

Punktene 1 - 9 ovenfor inneholder spørsmål som på dette nivå i systemutviklingen lett faller i øynene, og det er selvfølgelig meget sannsynlig at flere spørsmål vil oppstå etterhvert som disse løses.

Noen av disse spørsmål er allerede diskutert mer eller mindre inngående. Her skal nevnes spesielt spørsmål 1, der tre ulike forslag eksisterer. Disse forslag er gjengitt i appendix 1.1 bak. For ordens skyld bemerkes at framstillingen i app. 1.1 er interne arbeidsrapporter framlagt som forslag under komiteens arbeide, og har ikke til formål å kunne forklare spørsmålet for personer uten kjennskap til den totale problemstillingen. Disse framstillingene forventes likevel å være til hjelp i det videre arbeidet, og er derfor tatt med.

Løsning av spørsmål 1 er sannsynligvis det viktigste i det nærmeste arbeidet fordi det har konsekvenser for løsningen av flere andre spørsmål (f.eks. 2, 4 og 9).

8. KONKLUSJON - KOSTNADSOVERSLAG

Dersom det skisserte (eller tilsvarende) system blir implementert ved NOD/BERGEN venter man å oppnå bl.a. følgende fordeler:

1. Den datamengde NOD rår over kan utnyttes bedre i og med at man får bedre oversikt over data.
2. De ansatte ved NOD avlastes i sitt rutinearbeid i forbindelse med framhenting av data og øvrig ekspedisjon av brukere. Dette oppnås ved at brukere kan registreres i systemet og få tilgang på NOD's premisser.
3. Punkt 2 resulterer i at de ansatte får bedre tid til nytenking og forbedring av eksisterende systemer samt utvikling av nye systemer.
4. Punkt 2 medfører at utnyttelse av data blir billigere (siden administrasjonen forenkles).
5. Punkt 1 medfører at mer konstruktiv planlegging for datautnyttelse kan foretas (når man har oversikt over tilbud, etterspørsel og forbruk).
6. Man kan forvente en større etterspørsel etter data og opplysninger om data. Dette betyr at NOD som institusjon trekker til seg nye arbeidsoppgaver - noe som bør føre til en utvikling innen området.
7. NOD vil i det hele styrke sin stilling som datasenter på det oseanografiske område.

KOSTNADSOVERSLAG

Et kostnadsoverslag på grunnlag av en så grov systemmodell som denne må bli heftet med feil. I den grad man finner å kunne legge signifikans i tidsrammene kan følgende oversikt settes opp:

Det videre arbeide består i :

- a) Avslutning av systemering
- b) Planlegging av implementasjon - modulering av systemet i rutinen
- c) Programmering
- d) Implementering (testkjøringer)
- e) Innkjøringsperiode. (se nedenfor).

Flere av disse tidsperiodene vil overlappe hverandre uten at den totale tiden kan forventes å bli kortere. I tillegg til disse bør eksplisitt nevnes tid for dokumentasjon som lett kan undervurderes.

Følgende ramme foreslås: (am = en person i en måned)

For a : 4 am
for b : 2 am
for c : 10 am
for d : 4 am
for e : 4 am

Med "innkjøringsperiode" menes her den tid det tar å innlemme så mye av NOD's data i systemet at man kan uttale seg om at systemet virker tilfredsstillende. Det vil ta mye lenger tid før alle NOD's data ligger i systemet da man må forvente omformatteringer og omorganiseringer i stor grad.

KOSTNADER

Kostnadene med systemarbeidet kan deles i

- lønninger til ansatte
- maskinkostnader

Fra et økonomisk synspunkt er maskinkostnadene ikke interessante fordi NOD ikke betaler for den maskintid som brukes. For fullstendigheten tas likevel maskinkostnader med som egen kostnad.

En vanlig maskinkostnad ved implementering er ca. 50 000 betalingsenheter (dvs. 50 000 kr) pr. kvartal pr. person, og vi antar en årslønn på kr. 90 000,-. Av dette får vi følgende oversikt (avrundet til 1000 kroner).

OPPGAVE	ARBEIDS- MENNGDE	LØNN	MASKIN- KOSTNADER
Avslutning/systemering	4 am	30 000,-	0,-
Planlegging implementasjon-modulering	2 am	15 000,-	0,-
Programmering og implementering	14 am	105 000,-	233 000,-
Innkjøringsperiode	4 am	30 000,-	67 000,-
SUM	24 am	180 000,-	300 000,-

APPENDIX I

App. I inneholder tre forslag til løsning av formatteringsspørsmålet (se kapittel 9).

Som nevnt i kapittel 9 er disse rapportene interne rapporter framlagt som arbeidsgrunnlag i komiteen. Rapportene har dermed ikke til hensikt å orientere personer uten kjennskap til den totale problemstillingen.

De tre forslag har mange felles likhetspunkter, noe som vil framgå av teksten, men ulikhetene er såvidt mange at et felles forslag ennå ikke eksisterer.

Det felles utgangspunkt for alle rapportene er :

- a) Måledata i systemet skal lagres på filer på tape
- b) De aller fleste data må kunne hentes fram automatisk av systemet
- c) Det må finnes muligheter for å behandle/administrere datatyper som hittil er ukjørte, uten for store forandringer i systemet.

INNHOLD

1.1	ORGANISERING AV DATA PÅ TAPE Arne Sandahl, 3/11-76	side	1
1.2	TANKER OM LAGRING AV DATA VED NOD Øyvin Strand, 11/11-76	"	7
1.3	FORELØPIG LØSNING AV FORMAT-SPØRSMÅLET Arvid Hilstad, 20/11-76	"	11

1.1 Organisering av data på tape

Forfatter : Arne Sandahl

Dato : 3/11 -76

Data i en fil består av flere måledataserier (stasjoner). En stasjon deles i en masterpost og en eller flere dybdeposter. En masterpost består av parametre som er felles for alle data i måleserien. Dybdepostene består av parametre som er felles for dypet. Flere målinger av samme parameter kan lagres for samme dyp.

Alle måledatatyper plasseres i samme type format. Dersom det skulle dukke opp datatyper som ikke passer i formatet, skrives det spesialrutiner.

Dataseriene blir lagret i blokker. De leses og skrives med NTRAN som er en rutine som kan kalles fra FORTRAN. Lengden av blokkene bør være fast, i hvert fall innenfor hver fil. For at lesing/skriving skal kunne gå hurtigst mulig, bør en velge lange blokker.

Hver blokk består av:

1. en tekststreng (kan utelates og tar da ikke plass)
2. en eller flere stasjoner eller en del av en stasjon
3. en hale som gir opplysninger ved administrering av blokken.

Første blokk har dessuten et hode som gir opplysninger om filen. Hodet viser hvilke parametre som er på filen.

Når dataene sorteres, bør en ta hensyn til brukerens spørrefrekvens. Det kan være aktuelt å blande data med delvis forskjellige parametre fra forskjellige tokt. Derfor nyttes det variabel postlengde.

Masterpostene har en fast og en variabel del. ybdepostene har variabel lengde. Den variable delen er avhengig av antall parametre som er tatt. Dette gjør at vi får en effektiv lagring selv om det er tatt forskjellige parametre i ulike dyp og selv om forskjellige stasjoner har ulike parametre.

Format for lagring av data på magnetbånd.

Hode Tekst Første stasjon Andre stasjon . . Siste stasjon Hale	1. blokk
Tekst Stasjoner Hale	2. blokk
Tekst Stasjoner Hale	n. blokk

Hodets oppbygging.

Ord nr.

1	5 ord som viser hvilke parametre som er på filen. Hver av de 180 bitene representerer en parameter.
2	
3	
4	
5	
6	Brukeridentifikasjon
7	Filidentifikasjon (filanvn) 3 ekstra ord til identifikasjon av filen
8	
9	
10	
	Format for tolking av resten av filen
	Annen Klassifikasjon

De 5 parameterordene inneholder 180 biter som viser hvilke parametre som er på filen. Hver bit representerer en parameter og blir satt lik 1 når parameteren er på filen.

Definisjon av hver parameter står i en egen tabell.

Blokknr. settes negativ dersom blokken er den siste i filen.

Stasjonenes oppbygging.

7 ord for faste masterdata: Land, skip, st.nr., posisjon Kvadrant, dato, stasjonstid dybde til bunn og marsden square.	Masterkort
Parameterord for dypet med antall parametre Parametrene i dypet	1.dyp
Parameterord med antall parametre Parametrene i dypet	2.dyp

Parameterord med antall parametre Parametrene i dypet	Siste dyp.

Parameterordet

Bitene viser hvilke parametre av de som er angitt i hodet som er med i dette dypet.	Antall parametre
-------------------------------------------------------------------------------------------	---------------------

Dersom et parameterord er negativt (bit null er satt), er samme parameter tatt flere ganger. Parameterordet viser da hvilken parameter det gjelder og hvor mange ganger den er tatt.

Halens oppbygging

Ord nr.

	Adresse til siste stasjon
	.
	.
	.
4993	Adresse til andre stasjon.
4994	Adresse til første stasjon.
4995	Adresse til siste ledige ord.
4996	Adresse til første ledige ord bak siste stasjon
4997	Nr. på siste stasjon i blokken.
4998	Nr. på første stasjon i blokken.
4999	Blokknr. på filen
5000	Filnr. på bandet.

Nr. på første stasjon i blokken er negativ dersom stasjonen begynner i foregående blokk.

Nr. på siste stasjon i blokken er negativ dersom stasjonen fortsetter i neste blokk.

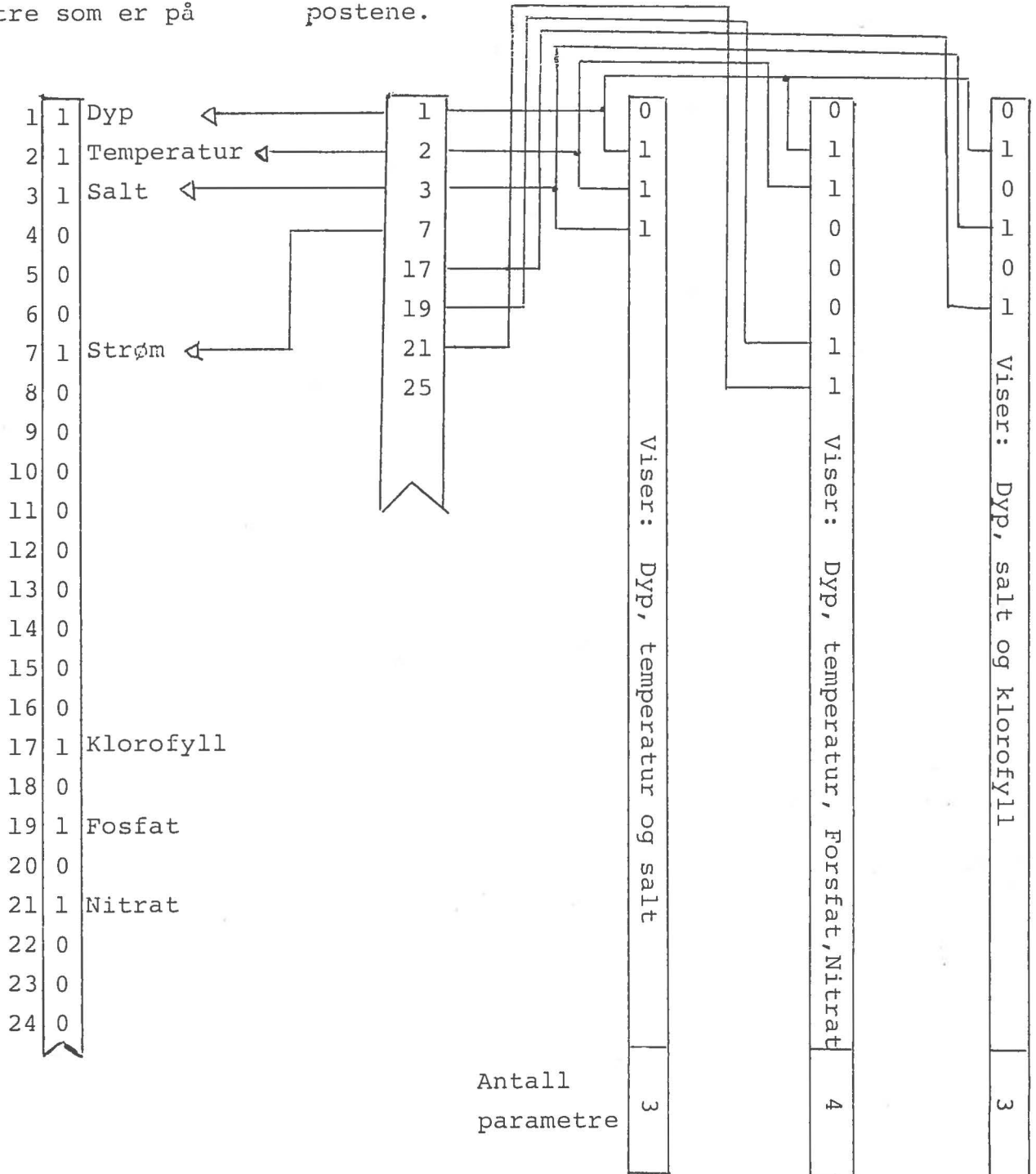
Adresse til første og siste ledige ord settes lik null dersom det ikke er ledig plass i blokken.

Eksempel på sammenhengen mellom parameterordet i hodet og parameterordet i dypene.

Første delen av de 5 ordene som viser hvilke parametre som er på filen.

Første delen av 30 ord i hukommelsen som viser parameterbitene i hver av dybdepostene.

3 eksempler på parameterord for dypene.



1.2 Tanker om lagring av data ved NOD

Forfatter : Øyvind Strand

Dato : 11/11 -76

Klassifisering av data:

- a) Data som representerer en tilstand (fysisk, kjemisk, biologisk) på et bestemt sted (posisjon) i et bestemt øyeblikk (eks. salt, temperatur, vannets øvrige kjemiske parametre, biologiske parametre (mengde, artsfordeling), kjerneprøver, kjemisk sammensetning i levende organismer etc.).
- b) Data som representerer en tidsserie for en tilstand i en bestemt posisjon. (Eks.: strøm, bølger (som NOD ikke skal lagre) samt nye typer bøyer etc.).

Disse data vil vanligvis være etterspurt i grupper. F.eks. vil en hyppig få forespørsel om temp., salt og avledede parametre (sigma-t, delta-alfa, delta-d, lydshastighet). Disse typer bør derfor lagres samlet. Ved forespørsel etter kjemiske data vil brukeren vanligvis være interessert i alle kjemiske parametre som er målt på stedet (også temp. og salt). Følgelig bør også disse lagres samlet. De to typene forespørsler er imidlertid hver for seg så spesielle, at det ikke skulle være noen grunn til ha felles datalager. Jeg foreslår derfor følgende lagringsstruktur for data som kan klassifiseres under pkt. a: Det opprettes separate hovedsystemer) for:

- 1) Fysiske målinger (temp., salt (O_2 (?), forsfat(?)) og avledede størrelser) i et begrenset antall dyp pr. stasjon (): standarddyp).
- 2) Samme målinger som (1) men i større antall dyp (): sonder). Utdrag av disse lagres i (1). (Oppsplitting i 1 og 2 skyldes det betydelig større volum som 2 vil få).

- 3) Kjemiske data i tradisjonell forstand (): data som kan kvantifiseres med ett eller noen få tall pr. dyp (dybdeintervall) (ref. ICES-kort for kjemiske data).
- 4) Biologiske data i den utstrekning disse kan kvantifiseres med noen få tall.
- 5) Bunnprøver
- 6) Kjemisk sammensetning i organismer (Hg etc.)
- 7) Spesielle datatyper (): volumiøse datatyper (vanligvis) må behandles med skreddersydde systemer (type 2 (fullstendige) sondemålinger er forsåvidt en variant av dette).

For å spare tidkrevende søking gjennom volumiøse måledatafiler, bør det for hver måledatafil opprettes en registerfil med en post (henvisning) til hver stasjon. Innholdsregisterfilen bør inneholde posisjon (i flere versjoner?) dato, tidspunkt (klokken), stasjonsnummer (hvis dette er relevant) etc. samt en grov oversikt over målte parametre (bunndyp, max. obs. dyp, hovedgrupper av data etc.). Likeledes bør der være en henvisning til om stasjonen er "representert" også i andre filer.

Filoppdeling/sorteringsrekkefølge.

En bør ikke binde seg til ett fast sorteringskriterium for alle typer data. Det kan ofte være fordelaktig å ha data sortert på flere forskjellige måter. Som hovedregel bør likevel alle data finnes i geografisk/kronologisk sorteringsorden (): geografisk område, år, skip, dato, tid/stasjonsnummer). Geografisk område kan tenkes spesifisert på flere måter, men tigradersruter (eller underinndelinger ($5^{\circ} \times 5^{\circ}$)) gir en logisk og systematisk oppdeling som har vist seg velegnet for formålet (unntak: fjordområder). Ved siden av hovedsorteringssystemet bør en og ha muligheter for en finere geografisk inndeling. F.eks. er sortering etter tigradersruter, en-gradersruter, måned osv. mye brukt.

NOD's nåværende lagrinssystem tillater sortering etter en valgfri rekkefølge av engradersruter innen hver Marsdenrute (): engradersrutene kan samles i grupper av valgfri størrelse) mens den videre sorteringsrekkefølge er den samme som i hovedarkivet (): år, skip, dato, stasjonsnummer). En sortering etter årstid vil sannsynligvis være mer praktisk ved en slik sorteringsmåte.

Videre bør en ikke helt droppe tanken om et fjordarkiv basert på den inndeling av kysten som er gjort ved NOD.

Postformater/blokkformater.

Den foran skisserte oppdeling av datalageret vil gjøre det hensiktsmessig å operere med faste postformater. Disse kan være på karakterformat, binære eller en blanding. For de mest typerike registre (f.eks. kjemi) kan en fordelaktig dele opp i flere typer poster. Faste postformater vil gjøre det betydelig lettere å hente fram data, videre vil det også kreve mindre plass. Innholdsregisterpostene bør ha fast lengde og, i et hvert fall for hovedparametrenes vedkommende, fast format. Blokk lengden for måledata bør gjerne være fast uansett datatype. Den bør imidlertid ikke være for stor, da det kan by på fordeler å bruke parallellprosessering i NTRAN, og en ofte vil operere med flere blokker samtidig i primærhukommelsen. Med et system basert på det foran skisserte, skulle et innviklet blokkformat for måledata være unødvendig, stasjonene kan bare "fylles på" fra blokken begynnelse. En "hale" med adresse til første (eller siste ord i hver stasjon kan dog by på fordeler ved oppdatering (): "fletting" av to filer) ved at en da slipper å søke seg gjennom stasjonen for å finne slutten (basert på at en ikke behøver å ha innholdsregisterfilene (med adresseparametre) tilgjengelig for å "flette" to sorterte måledatafiler).

Konklusjon

1. Datalageret organiseres etter de foran opptrukne linjer med faglig/datateknisk sammenhørende data i felles filer.
2. Blokk lengder/blokkformat er felles for alle typer.
3. Postformat/postlengde defineres for hver type. Oppbyggingen av dette skjer før lagringssystemet overtar (i bearbeidings/korrigeringsfasen).
4. Systemet bygges opp slik at rutinene blir mest mulig uavhengig av stasjonenes interne postformater.

1.3 FORELØPIG LØSNING AV FORMAT-SPØRSMÅLET

Forfatter : Arvid Hilstad

Dato : 22/11 1976

Måledata ansees å falle i to kategorier :

- a) Data som består av en eller flere måledataserier - såkalte stasjoner.
En stasjon kan deles i en "heading" samt måledata.
En heading består av størrelser som er felles for alle data i måleserien. Av slike størrelser kan nevnes dato, klokkeslett, geografisk posisjon osv. (Det er imidlertid ingenting som motsier at f.eks. geografisk posisjon varierer i andre forbindelser og som sådan blir å betrakte som en parameter).

Måledataserien består av størrelser på en eller flere parametre. Antallet av parametre (som kan både være observerte og beregnede) kan variere fritt men skal være fast innenfor hver fil.

- b) Data som ikke faller under kategori a.
Disse typene (som på dette tidspunkt er uoversiktlige) kan leses og skrives med spesialrutiner. Navnet/identifikasjonen på hver slik rutine kan avmerkes i arkivet og brukes når slike data skal leses/skrives. Dette gir også muligheter for å innlemme nye datatyper i systemet ved å programmere nye lese/skriverutiner.

De fleste kjente datatyper (ved NOD) faller i kategori a. Dersom det er nødvendig med forklarende tekst i forbindelse med datafiler kan denne teksten avmerkes i arkivet slik at datafilene alltid inneholder bare måledata (pluss kontroll data).

Oppbygging av filer

Hver fil på magnetbånd bør ikke være større enn at den kan gjennomleses/behandles på maksimum 5-10 minutter I/O-tid. En fil er bygd opp av blokker som kan variere i størrelse men forutsettes fast innenfor hver fil. Blokkene skal bestå av reelle vektorer (dersom lesing/skriving skal foregå med NTRAN). (Ellers må blokkene bygges opp i henhold til det programmeringsspråk som velges - men best slik at alle størrelser i blokken er av samme type). Antallet blokker kan variere fra fil til fil.

Formatspesifikasjon for data i kategori a.

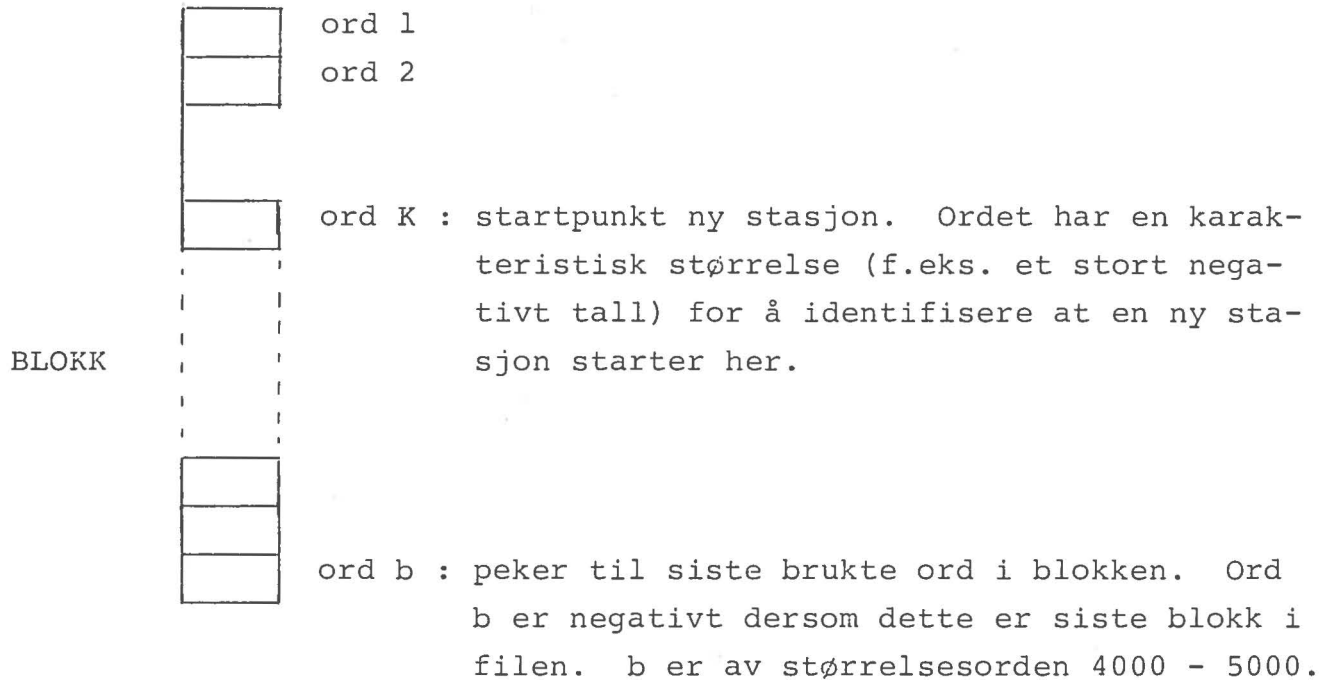
Etter det som er sagt kan formatet spesifiseres med følgende størrelser

- Blokkstørrelse i filen
- antall størrelser (ord) i headingen
- antall parametre i måledataserien (dette antallet viser antall ulike parametre og har ingenting med antallet observasjonr å gjøre.

En vanlig størrelsesorden for dette antallet er 5-6).

Oppbygging av hver blokk og identifikasjon av stasjoner

I hver blokk må finnes en del kontrolldata slik at lese/skrive-rutinen kan orientere seg i blokken. Her kan løsningene dele seg i to typer alt etter om registerfilen (se graf TU) blir implementert eller ikke. Dersom slike filer implementeres vil det være tilstrekkelig med denne oppbygging av blokken: (b = blokkstørrelse).



En typisk gjennomlesning av en slik blokk kan da skisseres slik der:

- h er antall ord i heading
 - p er antall parametre
 - ny stasjon identifiseres med at ord K (se ovenfor) er mindre enn N. (N > 0).
1. Se på de p neste ord i filen (m.a.o. de resterende i denne blokk + de nødvendige antall ord fra den/de neste blokkene).
 2. Er det første av disse p ord mindre enn N?
Ja : Ny stasjonsheading (bestående av de neste h ord) er påtruffet.
Nei: Alle disse p ord er parametre.

Dersom registerfilen ikke implementeres, kan det være fordelaktig med en slags direkte adressering innenfor hver blokk: (blokkstørrelse b).

b-5	
b-4	
b-3	adresse til 2. stasjon i blokk
b-2	adresse til 1. stasjon i blokk
b-1	antall stasjoner som starter i denne blokken
ord b	peker til siste brukte ord i blokken (negativt for siste blokk)

Størrelsen av ord (b-1) vil altså også være lik antall kontrollord minus 2.

Spesifikasjon av parametre

Hver parameter gis en entydig kode (gjerne heltall).

Til hver fil avmerkes i arkivet de koder som tilsvare de parametre som finnes i filen (i samme rekkefølge som de framkommer i måledata-serien). Rekkefølgen av parametre skal være konstant innenfor hver stasjon på hver fil.

Det har vært en del diskusjon om det forhold at enkelte stasjoner kan mangle enkelte parametre innenfor en fil. Dette er mer et unntak enn en regel, og det foreslås to typer løsning for dette:

- parametre grupperes slik at dette unngås i størst mulig grad
- der manglende parametre oppstår fylles tilsvarende ord med "umulige verdier for å angi "parameter ikke observert/beregnet". Disse "nullverdiene" kan f.eks. være meget høye tall som sprengen utskriftsformatet.