

Interne notater

STATISTISK SENTRALBYRÅ

35/18

21. mai 1985

S Y S T E M F O R D A T A D O K U M E N T A S J O N I S T A T I S T I S K S E N T R A L B Y R Å

Innstilling fra styringsgruppen

I n n h o l d

	side
1. Innledning.....	2
2. Hovedretningslinjer for dokumentasjon av Byråets data.....	3
3. Datadokumentasjonssystem for Byrådet. Rapport fra prosjektgruppen "DATADOK".....	5
4. Publikasjonsdokumentasjonssystem for Byrådet. Rapport fra prosjektgruppen "PUBLDOK".....	70
Vedlegg : Notat fra adm. direktør av 3/11-82. System for datadokumentasjon	83

DEL 1 INNLEDNING.

Statistisk Sentralbyrås perspektivprogram for 1980-årene sier at det skal utarbeides og iverksettes en langsiktig plan for å lette gjenbruken av data. I en første fase tas det sikte på å fastsette generelle retningslinjer for dokumentasjon av dataene.

Administrerende direktør ga i november 1982 to prosjektgrupper og en styringsgruppe i oppdrag å arbeide med dokumentasjonsspørsmål. Den ene prosjektgruppen (DATADOK-gruppen) fikk som mandat å spesifisere de krav som stilles til et dokumentasjonssystem for data på maskinlesbare media, vurdere ulike tekniske løsninger og utarbeide forslag til opplegg av et system for data-dokumentasjon. Den andre prosjektgruppen (PUBLDOK-gruppen) skulle utrede behovet for og opplegget av dokumentasjonen av data lagret i publikasjoner o.l.. Styringsgruppens hovedoppgave skulle være å samordne arbeidet til prosjektgruppene og utarbeide hovedretningslinjer for dokumentasjon av data. Gruppens sammensetning og en nærmere beskrivelse av deres mandat er gitt i vedlegg (notat av 3/11-82 fra adm. direktør).

Prosjektgruppens arbeid med utformingen av et system for dokumentasjon av data er søkt samordnet dels ved drøftinger i fellesmøter mellom prosjektgruppene og styringsgruppen, dels i møter mellom prosjektlederne og styringsgruppens formann. Det er i alt holdt 8 fellesmøter.

Styringsgruppen sluttet seg i møte 6.mars 1985 til hovedpunktene i sluttrapportene fra prosjektgruppene og vedtok at rapportene skulle legges til grunn for utformingen av forslag til retningslinjer for dokumentasjon av Byråets data.

Styringsgruppens forslag til hovedretningslinjer for dokumentasjon av Byråets data er tatt inn som del 2 av dette notatet, mens rapportene fra DATADOK-gruppen og PUBLDOK-gruppen er tatt inn som henholdsvis del 3 og del 4.

DEL 2 HOVEDRETNINGSLINJER FOR DOKUMENTASJON AV BYRÅETS DATA.

Rapportene fra prosjektgruppene inneholder detaljerte beskrivelser av et forslag til system for datadokumentasjon. Det vil imidlertid være sider ved dokumentasjonsspørsmålene som må utredes og drøftes nærmere. Den endelige tekniske løsningen og særlig tilknytningen mellom metadatasystemet og de ulike databehandlingsverktøy i statistikkproduksjonen forutsettes således behandlet i mer detalj i selve utviklingsprosjektet for datadokumentasjon. Også ressurs- og organisasjonsmessige sider ved driften og vedlikeholdet av et system for datadokumentasjonen må utredes nærmere. Det er dessuten planlagt å utarbeide en brukerveiledning.

Selv om utredningen av dokumentasjonsspørsmålene ennå ikke er ferdig, finner styringsgruppen at arbeidet nå er kommet så langt at berørte parter bør informeres og gis høve til å uttale seg om det foreliggende forslag. Etter eventuelle justeringer på bakgrunn av høringsuttalelser, vil styringsgruppen tilrå overfor administrerende direktør at det foreslåtte oversikts-system blir gjeldende dokumentasjonsstandard for Byråets data.

Hovedpunktene i forslaget går fram av følgende utkast til retningslinjer for datadokumentasjon :

1. Formål. Dokumentasjonssystemet er et maskinstøttet system for registrering, lagring, oppdatering og framhenting av metadata om statistiske undersøkelser. Systemet er ment å være et verktøy for spesifikasjon, utvikling og produksjon av statistikk og et hjelpemiddel ved informasjon om og bruk av statistikk.
2. Opplegg. Dokumentasjonssystemet består av et felles og standardisert oversiktssystem og et sett systemer for grunndokumentasjon. Opplegg og retningslinjer for oversiktssystemet er beskrevet i kap. II i DATADOK-rapporten. Grunndokumentasjonen vil i tillegg inneholde opplysninger som er spesielle for det enkelte prosjekt.
3. Omfang. Alle prosjekter (undersøkelser) hvor det innsamles/genereres data som det kan være aktuelt å lagre i Byrået med tanke på senere gjenbruk, skal dokumenteres. Dette gjelder uavhengig av hva slags lagringsmedium som benyttes.
4. Innhold. Dokumentasjonen skal foruten opplysninger om prosjektet generelt, inneholde definisjoner m.v. av enheter, kjennemerker og datasett (filer). For disse metadatatypene dokumenteres de opplysninger som er listet opp i avsnitt II.1 i DATADOK-gruppens rapport.
5. Registrering. Registrering og oppdatering skjer interaktivt fra skjermterminal, men kan også gjøres ved satsvis innlast-

ing. Ved registrering/oppdatering nyttes de skjermbilder / blanketter som er vist i avsnitt II.4 i DATADOK-rapporten. Systemet innebærer at det for hvert prosjekt fylles ut disse skjermbildene :

- ett med generelle opplysninger om prosjektet
- et bilde for hver av de enheter prosjektet gir data om
- et bilde for hvert kjennemerke som karakteriserer den enkelte enhet
- et bilde for hvert datasett (logisk fil) med tilhørende bilder for dennes redigering og fysiske fil. Dersom et datasett er lagret på flere fysiske filer og eventuelt med forskjellige redigeringer, dokumenteres hver redigering/fysisk fil på egne bilder.
- et bilde for hver publikasjon i henhold til retningslinjene i rapporten fra PUBLDOK-gruppen.

Registrering/oppdatering av metadataene foretas i takt med framdriften av prosjektet, slik at dokumentasjonen er ajour til enhver tid.

6. Lagring. Metadataene lagres i en egen database (her kalt metadatabasen).
7. Uttak. Foruten funksjoner for interaktiv og satsvis registrering og oppdatering av metadata, skal metadatabasen ha funksjoner for rapportuttak, for søking etter opplysninger som oppfyller visse kriterier og for generering av data-definisjoner til de viktigste edb-verktøy for produksjon og analyse av statistikk.
8. Ansvar. Prosjektleder er ansvarlig for at dokumentasjonen av et prosjekt gjennomføres etter fastsatte retningslinjer. Ansvar for teknisk vedlikehold og drift tilligger System- og Driftskontoret. Det faglige hovedansvaret for veiledning og håndheving av dokumentasjonsstandarden tilligger en sentral dataadministrator.
9. Sikkerhet. Brukerne av metadatabasen tildeles autorisasjonskoder avhengig av hva slags opplysninger de skal ha tilgang til og hvilke funksjoner de kan utføre.
10. Iverksettelse. Disse retningslinjene trer i kraft fra det tidspunkt administrerende direktør bestemmer.

HST, TSO 18.02.85.

DEL 3 : Datadokumentasjonssystem for Byrået. Rapport fra prosjektgruppen "DATADOK".

I n n h o l d

	side
0. Sammendrag	7
I. Innledning.	
I.1 Logiske og fysiske metadata	8
I.2 Databeskrivelse	8
I.3 Behandlingsregler	9
I.4 Bruk av metadata	11
I.5 Vurdering av eksisterende system	12
I.6 Krav til dokumentasjonssystemet	13
II. Forslag til metadatabase, retningslinjer.	
II.1 Definisjon av metadatatyper	14
II.2 Datamodell	20
II.3 Metadatafunksjoner	21
II.4 Utkast til skjermbilder for metadatatdefinering	22
III. Utvikling og drift av dokumentasjonssystemet.	
III.1 Systemarkitektur	29
III.2 Valg av tekniske løsninger	30
III.2.1 Tekstbehandlingssystem	30
III.2.2 Datakatalogsystem	31
III.2.3 Skreddersydd metadatasystem	32
III.2.4 Konklusjon, utviklingsplan	33
III.3 Drift og vedlikehold av systemet	34
IV. Bruk og utvikling av dokumentasjonssystemer i statistikk-byråer.	
IV.1 Datadokumentasjonssystemer i stat.byråer ...	35
IV.2 Statistiske database- og analyse-systemer ..	41
V. Oversikt/vurdering av datakatalogsystemer.	
V.1.1 Datakatalog-systemer (DD/DS)	49
V.1.2 Oversikt over aktuelle DD/DS	50
V.2 IR-systemer	62

Referanser	64
Vedlegg :	
A. Filbehandlingsoperasjoner - syntaks	65
B. Avledning av variable - syntaks	66
C. Kontroller - syntaks	67

0. Sammendrag.

Det er tidligere lagt frem en rekke notater/rapporter om data-dokumentasjon i Byrået (se bl.a. referansene på side 60). Denne rapporten gir derfor en mer summarisk oversikt over metadata i Byrået som bakgrunn for å sette opp en kravspesifikasjon til et maskinstøttet dokumentasjonssystem. Et forslag til organisering av datadokumentasjon i en metadatabase er gitt i Kapittel II.

Edb-teknologien stiller idag til rådighet mange hjelpemidler for å lagre, bearbeide, gjenfinne og presentere dokumentasjon. I SSB som i andre statistikk-byråer er disse mulighetene tatt i bruk i enkeltprosjekter og for å dekke mer generelle behov. SCB (Sverige), Statistics Canada, ABS (Australia) og flere har opprettet ganske omfattende dokumentasjonssystemer. Også i Finland og Danmark blir denne type arbeid høyt prioritert.

Selv med maskinelle hjelpemidler har det ofte vist seg problematisk å holde dokumentasjonen ajour og å utnytte den i selve statistikkproduksjonen. De metadatasystemene som er nært knyttet til sentrale databehandlingsverktøy, virker best i så måte, men er mer teknisk orientert. Utviklingen mot mer brukervennlige språk og edb-verktøy med integrerte metadata-systemer fører til at dokumentasjon automatisk følger av system-løsningen.

Alternative tekniske løsninger er vurdert i Kap. III. Prosjekt-gruppa foreslår at en bygger videre på ISDS-systemet som er i bruk på finans-, skatte- og befolkningsstatistikk. Dette metadatasystemet ligger nær opp til denne foreslåtte standarden. Ved innføring av et generelt databasehåndteringssystem må en tilpasse metadatabasen med det tilhørende datakatalogsystemet.

I Innledning.

Kapitlet gir en summarisk oversikt over metadata i Byrået. Dette brukes som bakgrunn for en vurdering av dokumentasjonssystemet. For mer utførlig beskrivelse av datadokumentasjonen i Byrået, vises til refansene. Et forslag til organisering av metadata og utkast til retningslinjer er lagt fram i kapittel II.

I.1. Logiske og fysiske metadata.

Metadata er dokumentasjon av datainnholdet og den fysiske lagringen av dataene og skal dekke brukernes og maskinens behov for data om data.

I.1.1. Logiske metadata.

Logiske metadata er dokumentasjon av datainnholdet så som logisk databeskrivelse med behandlingsregler. Logiske metadata beskriver "hva" som er gjort eller skal gjøres (krav-spesifikasjon / problem-beskrivelse).

I.1.2. Fysiske metadata.

Fysiske metadata er dokumentasjon av lagring og bearbeiding så som lagringsmedium, filident, filtype, rekordbeskrivelse og behandlingsregler i maskinlesbarform (programmer). Fysiske metadata beskriver "hvordan" det er gjort eller skal gjøres. (teknisk løsning).

I.2. Databeskrivelse.

I.2.1. Type undersøkelse (prosjekt).

- engangsundersøkelse
- gjentatt undersøkelse
- utvalgsundersøkelse
- telling
- intervjuundersøkelse
- postundersøkelse

- register
(store registre: personregister, bedriftsregister o.l.,
små registre : kommunekatalog, posstedskatalog o.l.).

- data fra administrative registre.

I.2.2. Populasjon (undersøkelsesbestand).

Opplysninger om populasjon blir dokumentert i publikasjonen.

I.2.2.1. Enhet.

Person, husholdning, bedrift, kommune osv. er angitt på filbeskrivelsen og i publikasjonen.

I.2.2.2. Omfang.

Telling eller utvalg. Utvalg er dokumentert ved trekkemetode og datagrunnlag.

I.2.3. Kjennemerke (variabel).

Dokumentert i filbeskrivelsen ved navn forøvrig definert og dokumentert ved behandlingsreglene.

I.3. Behandlingsregler (datainnsamling og databearbeiding).

I.3.1. Generelle behandlingsregler.

1. Standarder (yrkesstandard, næringsstandard o.l.).

Standardene fins som håndbøker. Fagkontorene er ansvarlig for vedlikeholdet av disse. Ingen av standardene er tilgjengelig interaktivt på maskinen.

2. Regler for gruppering av variable (alder, bosted o.l.).

Like regler for gruppering av de mest vanelige forklaringsvariable. Dokumentasjonen er lite systematisert.

3. Retningslinjer for publisering.

Eksempler:

- Retningslinjer for redigering og utstyr i publikasjoner fins i håndbok nr. 40.
- Minste antall observasjoner ved utvalgsundersøkelser.

I.3.2. Spesielle behandlingsregler.

I.3.2.1. Regler for datainnsamling.

1. Data innsamlet av byrået.

Datainnsamlingen er dokumentert ved skjema med regler for utfylling.

2. Data innsamlet av andre, administrative data o.l.

Tilgjengelige metadata fra kilde.

I.3.2.2. Regler for bearbeiding av data på laveste nivå (mikrodata).

1. Overføring data til maskinlesbart medium.

Instruks for dataregistrering som i de fleste tilfellene er angitt ved posisjoner på skjema (filbeskrivelse for grunn-dataene).

2. Avsløring av feil i dataene.

Kontrollbetingelser er dokumentert i notater og parametersett/program.

3. Retting av feil i dataene.

Ved automatisk retting er retteinstruksene dokumentert i parametersett/program. Forøvrig er retting dokumentert i interne notater og i tekstdelen i publikasjonen. I noen tilfeller hvor rettingen er gjenstand for skjønn, er den ikke dokumentert.

4. Gruppering av svar på åpne spørsmål.

Dokumentert ved referanse til standarder som er brukt ved kodingen, eller ved egne standarder.

5. Avledning av variable.

Omkodinger er dokumentert i parametersett/program. Behandlingsregler for omkoding er noen ganger tatt med som vedlegg til filbeskrivelsen.

6. Avledning av filer (ekstrahering av variable og selektering av dataposter).

Dokumentert ved parametersett/program, filbeskrivelser og systemkart.

7. Kopling av filer (overføring av variable og dataposter).

Dokumentert ved parametersett/program, filbeskrivelser og systemkart.

8. Statistisk bearbeiding (veiting, varianser ol.).

Dokumentert ved parametersett/proogram og i publikasjonen.

I.3.2.3. Regler for aggregering til makronivå (makrodata). opptelling/summering.

Dokumentert ved parametersett/program. og ved fil-
beskrivelse når dataene fins på maskinlesbart medium.

I.3.2.4. Regler for bearbeiding av makrodata (gjennomsnittsberegninger, prosentberegninger og beregning av andre statistiske mål).

Dokumentert ved parametersett/program.

I.3.2.5. Regler for presentasjon. Tabellspesifikasjoner (tabellparametere, tabelltekster og redigering).

Dokumentert ved parametersett/program.

I.3.3. Rekkefølge av aktiviteter.

Dokumentert ved systemkart (kjørekart).

I.4. Bruk av metadata.

Metadata brukes og skapes i alle faser av et statistikkprosjekt. De viktigste logiske metadata produseres under prosjektplanleggingen ved definering av undersøkelsens formål, populasjon etc. og ved utforming av spørreskjema (definering av variable). Disse metadata brukes så under datainnsamling, registrering/kontroll og under den videre bearbeiding av data. Under denne prosessen genereres flere metadata ved at avledede variable og datafiler defineres. Ved presentasjon av resultatet lages endelige tabellspesifikasjoner og publikasjonen redigeres. I dette arbeidet brukes mesteparten av metadata tilknyttet undersøkelsen.

Metadataenes viktigste rolle i Byrået er således å fungere som styredata for såvel manuelle som automatiske aktiviteter i produksjonsprosessen. Styredata tolkes og oversettes gjerne fra en form til en annen før de utføres. Ved programmering overføres behandlings-

regler til et maskinlesbart språk som tolkes i datamaskinen og utføres etter ytterligere oversetting.

I tillegg representerer selvfølgelig metadataene informasjon om Byråets dataressurser, og er som sådan viktige i den interne beslutningsprosess og for informasjonsvirksomhet og markedsføring utad.

Bruker-grupper :

fagkontorene
informasjonsavdelingen
forskningsavdelingen
eksterne brukere
produksjonsavdelingen

Funksjonsområder :

Standard for gjennomføring av stat.prosjekter (jfr. Håndbok nr. 23).

Verktøy for statistikk-produksjon.

Arkivsystem for dokumentasjon.

Informasjonssøkesystem.

I.5 Vurdering av det eksisterende dokumentasjonssystemet.

I det nåværende systemet er de fysiske dataene (data- og programarkiv godt dokumentert, men er ikke lett tilgjengelig og egner seg dårlig for moderne edb. Datainnholdet, hva dataene står for, er ofte dårlig dokumentert og vanskelig tilgjengelig.

I mange tilfelle er det nødvendig å gå tilbake til program og kjørekart for å finne fram til kjennemerkedefinisjonen.

Arsakene kan være:

- Metadataene er ikke sterkt nok knyttet til den fysiske delen av statistikkproduksjonen. Det medfører at behandlingsreglene ikke blir oppdatert når det skjer endringer.
- Behovet for metadata er forskjellig for fagkontor og produksjonsavdeling. Det medfører at arbeids- og ansvarsdeling blir flytende.

- Vi har ikke klart å få til et enkelt og oversiktlig system for arkivering og gjenfinning av logiske metadata.
- Det er mange parter med i den manuelle delen av statistikkproduksjonen. Det gjør at metadataene lett kan bli feil.

I.6. Krav til dokumentasjonssystemet.

Metadata av de typer som er relevante for statistiske undersøkelser, skal kunne beskrives og lagres v.h.a. dokumentasjonssystemet. Det skal med andre ord realiseres en dokumentasjonsstandard som er egnet for statistikkproduksjon.

Det må omfatte både logiske og fysiske metadata samt sammenhengen mellom disse.

Metadata må være lett tilgjengelige (databaseløsning med sluttbrukerorienterte uttaksfunksjoner).

Systemet må ha innebygd kontrollapparat til sikring mot ikke- autorisert bruk og for å beskytte metadataene mot ødeleggelse.

Dokumentasjonssystemet må kunne brukes som et verktøy i statistikk- uksjonen / være sterkt integrert med produksjonssystemet. Innmating og ajourhold av metadata bør skje der de dannes (forandres) og mest mulig som et ledd i den naturlige arbeidsprosessen (uten dobbelt- arbeid). Generering og overføring (konvertering) av metadata til maskinelle hjelpemidler (databasesystemer, standardprogram, spesial- programmer) bør skje mest mulig automatisk.

Systemet bør være lett å bygge ut og modifisere slik at nye behov for metadata og metadatagenerering raskt kan dekket.

II Forslag til metadatabase, retningslinjer for dokumentering av data.

Dokumentasjon utarbeides/ajourføres for hver enkel datainnhenting (undersøkelse, uttak fra registre eller administrative data) for statistikkformål, unntatt rene prøveundersøkelser.

Ansvar for dokumentasjon av data på fagstatistisk nivå (definering av enheter og kjennemerker) ligger hos prosjektleder/det fagkontor som har ansvar for det aktuelle prosjektet. Den datatekniske dokumentasjonen (filbeskrivelser etc.) gjøres av system- og driftskontor, eventuelt av datagruppe ved fagkontor.

Dokumentasjonen skal i størst mulig utstrekning foretas i takt med framdriften av prosjektet. Den skal være ajour til enhver tid, og benyttes aktivt ved alle faser i statistikkproduksjonen.

II.1. Definisjon av metadatatyper.

Definisjoner av enheter, kjennemerker og datafiler er det viktigste innholdet i et datadokumentasjonssystem. For å systematisere dokumentasjonen er det nødvendig også å ha med metadatatyper som prosjekt (undersøkelse) og emne. Følgende egenskaper foreslås dokumentert for de ulike metadatatyper :

Prosjekt :

I systemet for prosjektplanlegging inngår :

Nummer	identifikasjonsnummer (4-sifret).
Navn	beskrivende prosjektnavn.
Kontor	prosjektansvarlig kontor.
Formål	tekstlig beskrivelse.
Emne	emnekode (4-sifret).
Hyppighet	periodisitet/engangsprosjekt.

Tillegg for datadok. :

Undernummer	2-sifret årskode + eventuelt 1-sifret løpenr. Brukes for å identifisere ny versjon/utgave av datadokumentasjonen for et prosjekt. Ny versjon/utgave av dokumentasjonen utarbeides ved enhver endring i metadata. Ajourføring av metadata betinger ikke utarbeiding av ny versjon av dokumentasjonen.
-------------	---

Undersøkellesperiode (tellings-referanse, innsamlingsperiode)

periode - fra dato til dato - undersøkelsesopplysningene er samlet inn.

Oppgavegivere/Kilde

Her oppgis type person eller institusjon som har gitt opplysninger. Skriv f.eks. : den uttrukne person, bedriften selv, Rikstrygdeverket o.l.

Hjemmel

Her oppgis hjemmel i samsvar med "Hjemmel for innlevering av oppgaver" (SSB's Håndbok nr. 10). Dersom hjemmel mangler, skrives "På frivillig grunnlag".

Innsamlingsmåte

Registre, administrative data, oppgaveskjema (post), intervju, annen innsamlingsmåte; kombinasjoner av disse.

Revisjon og kontroller Kort generell omtale.

Annet

Emne (inngår i systemet for prosjekt-planlegging) :

Emnekode

Hierarkisk kode for identifisering av emner.
4 sifre for detajerte emnekoder.

Emneområde i 1. siffer. Emnegruppe i pos. 1,2.

Emnebetegnelse**Definisjon**

Enhet (brukt i prosjekt) :

 beskriver de elementer/individer som det er innhentet/bearbeidet data om i et prosjekt.

Prosjekt-nr	referanse til prosjekt.
Enhetsnavn	Kort navn (inntil 12 tegn).
Beskrivende navn	(inntil 30 tegn).
Spesifikasjon	definering, avgrensning av populasjon med f.eks. kjønn, alder for personer, næring, størrelsesgruppe for bedrifter. Referanse til standarddefinisjon.
Antall enheter i populasjon.	
Antall enheter i utvalg (bruttoutvalg).	
Svarprosent	nettutvalg i prosent av bruttoutvalg, 1 desimal.
Utvalgsmetode	Bør inneholde opplysninger om utvalgsenhet, -populasjon, registre, stratifisering, trekkemåte og utvalgsstørrelse.
Identifikasjon	Kjennemerket eller kombinasjonen av kjennemerker som identifiserer enheten.

Kjennemerke (variabel) :

 Kjennemerke uttrykker en egenskap ved en statistisk enhet. Primære opplysninger om kjennemerke er definisjon m.v. for registerbaserte og administrative data, instruks i skjema o.l.. Sekundære opplysninger er verdiliste, avledningsregel osv.

Prosjekt	referanse til prosjekt.
Enhet	referanse til enheten kjennemerket tilhører.
Kjennemerkenavn	Kort entydig navn innen prosjekt (12 tegn).
Beskrivende navn	(inntil 30 tegn).
Alias-navn	alternative navn for bruk i ulike programmeringsspråk.
Definisjon	tekstlig beskrivelse. Referanse til standarddefinisjon.
Verditype	(numerisk, alfanumerisk, antall sifre etc.)
Verdimengde	gyldige verdier, opplisting av kjennemerkeverdier (med betegnelser), ref. til standard.
Kontroller	spesielle (logiske) kontroller.
Avledningsregel	formalisert uttrykk / referanse til program.
Annet	

Forslag til syntaks for kontroller, verditype, verdimengde og avledningsregler er gitt i vedlegg.

Datafil (datasett) :

 Dokumenteres etter gjeldende filbeskrivelsesstandard.
 Til en logisk fil kan det svare mange fysiske filer med forskjellig redigering, omfang, sortering etc. Databasefiler kan tilpasses samme opplegg, men disse vil normalt ikke ha flere utgaver og redigeringer av samme logiske fil. Som hovedregel defineres filene med en normalisert rekordstruktur (redigering).
 I datadok.-systemet registreres bare endelige datasett og filer beregnet på langtidslagring.

Logisk fil :

 Prosjekt referanse til prosjekt.
 Datasett-nr Logisk filnr., bokstav og 3 siffer.
 Kort navn entydig navn på fil innen prosjekt (12 tegn).
 Beskrivende navn (inntil 30 tegn).
 Register-nr for melding til datatilsynet.
 Definisjon tekstlig beskrivelse av filens innhold, omfang.
 Enhet referanse til enheten.
 Sikkerhetsgradering.
 Datakilde

Fysisk fil :

Fysisk filident :

Prosjekt referanse til prosjekt.
 Datasett-nr ref. til logisk fil.
 Redigering-nr ref. til redigering (posttype).
 Utgave-nr bokstav, siffer (f.eks. A1).
 Generasjon-nr =Gnnn (n=siffer)
 Versjons-nr =Vnn (n=siffer)

Statuskode = S (SAS-prod.datasett), P (andre prod.-datasett), T (testdatasett)
 Oppbevaringskode = X (30 dager), 1 (100 dager), 2 (200 dager), 4 (400 dager), 6 (600 dager), L (langtidslagring), B (sikkerhetskopi av L)
 = K eller S (SAS) hvis testdatasett.
 Tidsreferanse tidspunkt for registrering/framstilling.
 Antall poster min, maks og gj.sn.-tall
 Omfang logisk uttrykk (spesifisering) av omfanget (f.eks. fylke = 01 og år = 83).
 Avledningsregel formalisert uttrykk / referanse til program.
 Lagringsmedium (plate, bånd, ..)
 Lagringstruktur (QSAM, VSAM, SAS,)

Sortering Kjennemerke (-kombinasjon) filen er sortert
etter.

Annet

Forslag til syntaks for filoperasjoner (avledninger) er
gitt i vedlegg.

Redigering (rekord/post-type) :

Prosjekt referanse til prosjekt.
Datasett-nr referanse til logisk fil.
Redigering-nr bokstav, siffer (f.eks. A1)
Felt-beskrivelse-liste :
 Kjennemerke (kort navn)
 startposisjon i rekord
 lengde
 lagringsformat (tegnformat, binært, flytende tall..)
 spesielle opplysninger avhengig av filtype

Publikasjon.

Dokumentasjon av publikasjoner og tabeller utgjør et del-system (PUBL-DOK) som primært benyttes til informasjonssøking. Det er også aktuelt å knytte noen produksjonstekniske opplysninger til disse metedatatypene. PUBL-DOK er beskrevet i eget notat.

De viktigste kjennetegn er :

Prosjekt referanse til prosjekt.
Hovedtittel
Undertittel
Serie, år, nummer
Underserie
ISSN og ISBN-nr
Forlag/utgiver
Sidetall, pris
Periodisitet
Forfatter(e)
Henvisning til tidligere tittel.
Innholdsoversikt.

Tabell.

Publikasjon referanse til publikasjon.
Emne
Tabell-nummer
Beskrivelse (tabelloverskrift)
Formell definisjon (parametre til tabellprogram eller referanse)

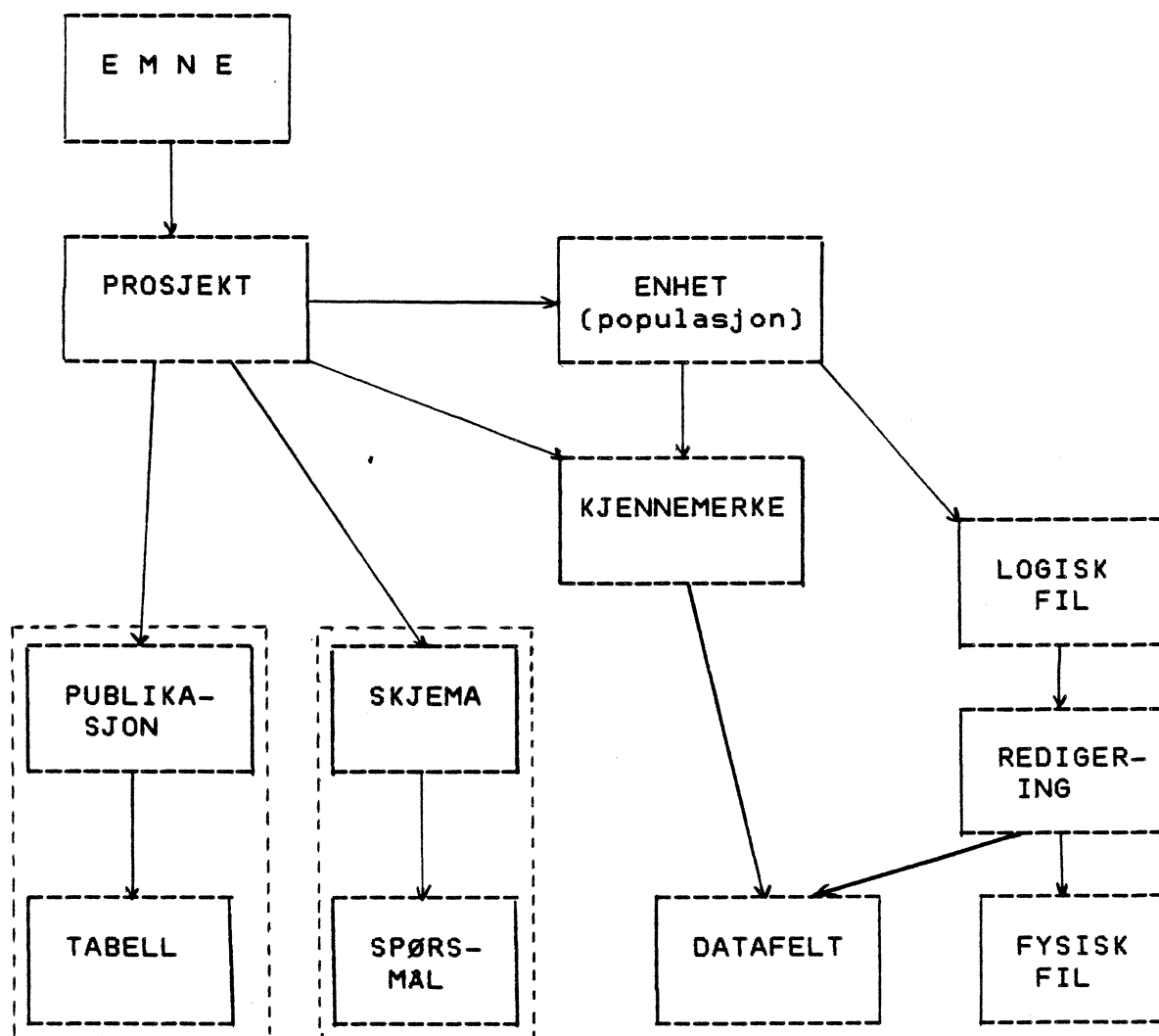
Spørreskjema.

Et sentralt skjemaregistret kan utgjøre et annet delsystem, hvor spørreskjemaer defineres og lagres.

Prosjekt referanse til prosjekt.
Skjema-nr
Spørsmåls-liste

II.2 Datamodell.

De enkelte dokumentasjonsenheter og sammenhengen mellom dem kan fremstilles grafisk i en såkalt datamodell. Metadatabasen konstrueres etter en slik modell. Hver metadatatype er representert med en boks, og relasjoner mellom dem er tegnet inn med piler. (Alle relasjoner er ikke tatt med i denne figuren). Enkelpil indikerer en relasjon av typen en-til-mange. Pilen mellom emne og prosjekt viser at mange prosjekter vil falle inn under ett emne.



II.3 Metadatafunksjoner.

Metadatasystemet må ha funksjoner for registrering og oppdatering av metadata, spørring og rapportuttak og generering av metadata til programmeringsverktøy.

Registrering/oppdatering.

For hver metadata type vil det svare et skjermbilde/blankett for interaktiv registrering og oppdatering av dokumentasjon. Skjermbildene vil inneholde de opplysninger som er angitt over. Menyer, veiledning og kontroller må inngå i systemet. Satsvis innlasting bør også være mulig. Utkast til skjermbilder for metadata definering i II.4. Forslag til syntaksregler for avledninger av filer og variable og for kontroller i vedlegg.

Uttaks-funksjoner.

Det vil være mange og varierte uttaksbehov fra en metadata-base. De enkleste uttak vil være en ren utlistering på skjerm og papir av de blankettene som benyttes ved definering. I tillegg vil en ta ut mer kompliserte sammenstillinger av metadata samt søke frem til opplysninger som oppfyller visse kriterier. Rapporter på flere detaljeringsnivå skal kunne lages og dessuten forskjellige typer kryssreferanserapporter. Brukere av metadatabasen må gis ulike autorisasjonskode avhengig av hva slags opplysninger de skal ha tilgang til og hvilke funksjoner de kan utføre. Fysisk filident må f.eks. undertrykkes for alle unntatt ansvarlig dataadministrator. Utkast til noen metadatarapporter i vedlegg.

Metadatagenerering.

Fra metadatabasen skal det genereres datadefinisjoner (filbeskrivelser ..) til de viktigste edb-verktøy for produksjon og analyse av statistikk. Dette vil være funksjoner som leser aktuelle metadata fra sentrale basen og omformer/tilpasser disse til de syntakskrav mottakersystemet har. I første omgang bør det lages slike grensesnitt til EASYTRIEVE-PLUS, TAB68, PRINTERTAB, SAS, ADDERS, COBOL, PL/I og 4.gen.språket.

II.4 Utkast til skjermbilder for
metadatadefinering.

```
*****
*
*           P R O S J E K T
*          -----
*
* Pr.nr. :           Undernr. :
* Navn   :
*
* Ansvarlig kontor :
*
* Formål :
*
*
* Emne :
*
* Hyppighet :           Undersøkelsesperiode :
*
* Oppgave-       Hjemmel :
* givere :
*
* Innsamlingsmåte :
*
* Kontroller (generelt) :
*
* Annet :
*
*****
```

```
*****
*
*                               E N H E T
*                               -----
*
*   Prosjekt-nr :
*   Kort navn   :
*
*   Beskrivende navn :
*
*   Spesifikasjon :
*
*
*   Tallet på enheter
*   i populasjon :           i utvalg :           Svarprosent :
*
*   Utvalgsmetode :
*
*   Identifikasjonsbegrep :
*
*****
```

```
*****
*
*
*           K J E N N E M E R K E
*           -----
*
*   Prosjekt   :
*   Enhet      :
*   Kort navn  :
*   Beskrivende navn :
*   Aliasnavn :
*
*   Definisjon :
*
*
*
*
*   Verditype :
*   Verdimengde :
*
*
*   Kontroller :
*
*
*   Avledningsregel :
*
*
*   Annet :
*
*****
```



```
*****
*
*                               Datasets (logisk fil)
*                               -----
*   Prosjekt      :
*   Datasets-nr  :
*   Kort navn    :                               Original/Kopi :
*   Beskrivende navn :
*
*   Registernr.  :
*
*   Definisjon  :
*
*
*   Enhet       :
*
*   Sikkerhetsgradering :
*   Datakilde   :
*
*****
```

```
*****
*
*                               Fysisk fil
*                               -----
*   Prosjekt      :
*   Status-kode   :                               Oppbev.-kode  :
*   Datasets-nr  :                               Redigering-nr :
*   Utgave-nr    :                               Generasjon-nr :
*   Versjon-nr   :
*
*   Tidsreferanse :
*
*   Omfang       :
*
*   Antall enheter min. :           maks. :           gj.sn. :
*
*   Avledningsregel :
*
*
*   Lagringsmedium :
*   Lagringsstruktur :
*   Sortering      :
*
*   Annet         :
*
*****
```

* *

Redigering (rekordtype)

* -----

* Prosjekt :

* Datasett-nr :

* Redigering-nr :

*

Kjenne- merke	! Start- pos.	! Lengde	! Lagr. form.	! Spesielle opplysninger
------------------	------------------	----------	------------------	-----------------------------

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

* *

For fullstendig spesifikasjon av publikasjon og tabell henvises til PUBLDOK-rapporten. Her er grovutkast til bilder med de viktigste opplysningene.

```

*****
*
*
*          P U B L I K A S J O N
*          -----
*
*  Prosjekt-nr :
*  Emne-kode   :
*
*  Serie      :           Underserie   :           Nr./år.. :
*
*  Hovedtittel norsk :
*
*
*  Undertittel norsk :
*
*
*  Parallelltittel (engelsk) :
*
*
*  Parallell undertittel (engelsk) :
*
*
*  ISSN-nr :           ISBN-nr :
*
*  Forlag/utgiver :
*
*  Sidetall :           Pris :
*
*  Periodisitet :
*
*  Forfattere :
*
*****
*

```


III Utvikling og drift av dokumentasjonssystemet.

Et edb-system for datadokumentasjon skal være et hjelpemiddel for å lage, arkivere og oppdatere dokumentasjon. Det skal tjene mange brukerkategoriens behov for metadata. Systemet må derfor ha fleksible og brukervennlige uttaksfunksjoner. Systemet skal også i størst mulig utstrekning utnyttes av (være koblet til) andre edb-systemer i statistikkproduksjonen.

III.1 Systemarkitektur.

Det kan være naturlig å tenke seg dokumentasjonssystemet oppdelt i et felles og standardisert oversiktssystem hvor alle de viktigste metadata samles i et databasesystem. I tillegg kan det være manuelle og edb-baserte dokumentasjonssystemer på de ulike statistikkområdene for å dekke mer detaljerte og spesielle dokumentasjonsbehov. Det er "oversiktssystemet" som er tema for denne rapporten.

1. Oversikts-dokumentasjonssystem.

Oversiktssystemet opprettes etter den ramme som er foreslått i II. Ved siden av å være et verktøy for spesifisering, utvikling og produksjon av statistikk, vil systemet benyttes til informasjonssøking i forbindelse med forespørsler, oppdrag og planlegging av nye undersøkelser.

Metadatabasen for publikasjoner (publ-dok) vil være en del-base i et slikt oversikts-system.

De statistiske standarder kan dokumenteres etter samme retningslinjer som statistikkprosjekter og benytte det samme metadatasystemet. Dette er spesielt viktig siden de enkelte statistikk-prosjekter vil benytte og referere til generelle metadata (standarder).

Systemet vil være koblet til planleggings- og oppfølgings-systemet for prosjekter gjennom prosjekt-nr og vil også benytte emne-katalogen herfra.

Systemet for program-dokumentasjon vil være knyttet til metadatasystemet via felles prosjektnummer.

2. Systemer for grunndokumentasjon.

Grunddokumentasjon vil omfatte mer detaljerte opplysninger om undersøkelsene i form av notater, produksjonstekniske beskrivelser etc. Denne dokumentasjonen vil derfor måtte tilpasses spesielle behov for de ulike statistikker. På dette nivået vil det således eksistere flere (manuelle og edb-støttede) systemer.

Oversiktsystemet (den primære metadatabasen) skal være tilknyttet de viktigste edb-verktøy for statistikkproduksjon s.a. ingen metadata må registreres på mer enn ett sted.

III.2 Valg av tekniske løsninger.

Et edb-basert datadokumentasjonssystemet kan opprettes v.h.a. flere typer verktøy og teknikker. Disse er her inn delt i 3 hovedalternativer. De ulike verktøy og systemer på markedet er beskrevet nærmere i kapitlene IV og V.

III.2.1 Tekstbehandlingssystem med lagrings- og søkefunksjoner.

Den edb-teknisk enkleste måten er å benytte vanlig teksteditor og filsystem for manipulering og lagring av metadata. Metadata lagres derved som vanlige tekstfiler som administreres v.h.a. operativsystemets filsystem eller et selvstendig programbiblioteksystem. På IBM-maskinen vil dette svare til å bruke partisjonerte datasett (PDS) og standard ISPF-funksjoner eller tilsvarende verktøy (ROSCOE, LIBRARIAN..). Flere filer (PDS) bør benyttes avhengig av prosjekt, bruker, metadata type e.l. for at filene ikke skal bli for store, og for at det ikke skal bli for tungvint og sårbart å oppdatere metadataene.

For søking/gjenfinning av dokumentasjon vil denne løsningen være noe begrenset. Selve filsystemet gir ikke mulighet for søking på innholdet i dokumentasjonen, bare via navn på tekst-filene (dokumentene). En systematisk navngiving og filkatalogisering vil kunne lette arbeidet med å finne fram i dokumentasjonen. Med hjelp av "editor-funksjoner" kan en søke / "bla" i innholdet av framskaffet materiale. Det vil bli nødvendig å utvikle ekstra funksjoner (i f.eks. ISPF/ DMS) for brukerdiallog, søking og administrering av metadata.

Søkefunksjonene vil være tilstrekkelig for mange formål, men lite egnet for utstrakt referansesøking. Det vil derfor antagelig bli nødvendig å anskaffe et IR-system (se Kap.V.2) for søking og gjenfinning. Viktige utdrag av dokumentasjonen kan lagres i et slikt system slik at en oppnår en effektiv og godt utbygd fritekstsøking.

Med en så ustrukturert teknikk for lagring av metadata vil det være vanskeligere å lage funksjoner som genererer metadata til ulike edb-verktøy, enn hvis en vanlig database-løsning benyttes.

Denne løsningen som her er skissert, vil altså kun benytte programvare som brukes (vil bli brukt) til andre funksjoner i Byrået. Noe systemteknisk arbeid vil gå med til tilpasninger/overbygninger (menyer, søking etc.).

III.2.2 Datakatalogsystem (DD/DS).

En alternativ løsning vil være å bruke et datakatalogsystem (Kap. V.1). Disse er bedre egnet for å opprette en kompleks metadatabasestruktur og har flere innebygde funksjoner for kontroll og analyse av metadata. De fleste av disse systemene inneholder dessuten hjelpemidler for systemutvikling/programmering. Særlig i de integrerte systemene utgjør metadatabasen en sentral og aktiv komponent som de ulike sluttbrukerspråk, database- og skjermhåndtering, systemet for applikasjonsbygging er koblet til.

Datakatalogsystemene har fast modell for datadokumentasjon som en må holde seg innenfor, selv om det er mulig å gjøre visse tilpasninger til egne behov. De fleste systemene er forholdsvis lite brukervennlige.

Ingen eksisterende (kjente) datakatalogsystemer har tilsvarende funksjoner for søking i fritekst som IR-systemene.

Datakatalogsystemene er ikke direkte koblet til vanlige kompilatorer, men de fleste har funksjoner som genererer såkalte "copy"-elementer. En mer dynamisk bruk av et metadatasystem under programutvikling krever tilpasning av (inngrep i) kompilator eller preprossesering. De nyere språkene (såkalte 4. generasjons språk) har integrerte metadatasystem. Datakatalogsystemene har dårlig støtte for de viktigste standardprogrammene i Byrået (DOKS, EASYTRIEVE, SAS, TAB68, PRINTERTAB).

Å anskaffe et datakatalogsystem vil være en ganske stor investering. Et frittstående system som f.eks. DATA-MANAGER vil koste over 200.000 kr. (med nødvendige tilleggsfunksjoner). Et utviklingsverktøy med integrert datakatalogsystem og databasehåndterer koster ca. 2 millioner kr., men er ikke egnet for alle databehandlingsoppgaver i Byrået. NATURAL blir trolig valgt som utviklingsverktøy for online-systemer, men i første omgang antagelig bare med VSAM som databasekomponent og ikke ADABAS med tilhørende datakatalogsystem (PREDICT).

III.2.3 Skreddersydd metadatasystem.

Et tredje alternativ er å utvikle et eget system, som skreddersys Byråets behov. Et metadatasystem kan bygges opp som et vanlig databasesystem ved å benytte standard metoder for utvikling av databaseorienterte informasjonssystemer (datamodellering, funksjonsanalyse, fysisk databasekonstruksjon, dialogutforming etc). Det viktigste grunnlaget for dette arbeidet vil foreligge i forprosjektet i form av kravspesifikasjonen. Utviklingskostnaden vil avhenge av hvor mange funksjoner (for spørring, rapport-uttak, etc) som implementeres, hva slags utviklingsverktøy som benyttes, og tilgjengelig kompetanse. Det er forholdsvis enkelt å programmere funksjoner for registrering, oppdatering, kontrollere og de fleste former for uttak.

En slik løsning vil bety å videreutvikle ISDS-systemet (det eksisterende metadatasystemet, som ble utviklet for Kommunaløkonomisk statistikk). Metadatamodellen i dette systemet ligger tett opp til den foreslåtte dokumentasjonsstandard. Noen mindre justeringer vil være nødvendig, satsvis innlesing av metadata fra ulike kilder og dessuten flere funksjoner for uttak og analyse av metadata.

Søking i fritekst er eller ikke fullt utbygd i generell databaseteknikk. Det er drimot enkelt å generere stikkordregister fra tekstlige beskrivelser og bruke dette i interaktiv søking med standard aksessmekanismer.

Løsningen kan baseres på de verktøy som er benyttet til nå, d.v.s. VSAM, PL/I og CICS (ISPF/DMS), mens NATURAL brukes i den grad det er hensiktsmessig for nye funksjoner og ved omskriving av gamle.

III.2.4 Konklusjon og utviklingsplan.

Prosjektgruppa går inn for at det tredje forslaget, altså skreddersydd metadatasystem, velges. Dette kan enklest tilpasses Byråets spesielle behov både til utforming og bruk av metadata.

Alternativ 1 er ikke et fullgodt verktøy for å lage et databasesystem av denne type. Tekstbehandlingsfunksjoner vil derimot være nødvendig for å håndtere metadata; preparering av tekster som skal inn i metadatabasen, og redigering av metadatauttak. Enkle tekstfunksjoner er også innebygd i systemet.

Datakatalogsystemene er ikke så enkle å tilpasse egne behov, gir liten støtte for de viktigste edb-verktøy i statistikkproduksjonen og er de fleste lite sluttbrukerorientert. Ved en eventuell innføring av ADABAS som databasehåndteringssystem, vil det være mest naturlig å basere metadatasystemet på PREDICT. PREDICT må i såfall suppleres med metadata typer for prosjekt, emne, enhet, kjennemerke og nødvendige relasjoner.

De viktigste aktiviteter for videreutvikling av metadatasystemet med anslag over timeverk :

Aktivitet	Ressursanslag
- Funksjoner for registrering og vedlikehold av metadata (interaktivt og satsvis).	100 timeverk
- Interaktive søkefunksjoner.	180 "
- Satsvise metadatarapporter.	60 "
- Satsvis konvertering og innlasting av metadata fra andre kilder.	120 "
- Metadatagenerering for	
- DOKS (ADDERS)	
- EASYTRIEVE	
- TAB68	
- PRINTERTAB	
- SAS	
- COBOL, PL/I, NATURAL	
- TROLL	200 "
- Diverse hjelpefunksjoner.	100 "
- Lage brukerveiledning	120 "

I alt ... 880 timeverk

De grunnleggende funksjoner for vedlikehold og uttak av metadata gjøres klare først, deretter funksjonene for metadata-generering, konvertering av eksisterende metadata og utvikling av flere uttak/søke-muligheter. Punktet for metadata-generering er i første omgang ment å dekke filbeskrivelser for nevnte programmeringsspråk, men bør etter hvert også omfatte behandlingsregler, kontroller etc.. Dette arbeidet koordineres med systemet for program-dokumentasjon/utvikling. En løsning basert på PREDICT vil kreve et tilsvarende utviklingsarbeid, men noe mindre på uttakssiden.

III.3 Drift og vedlikehold av systemet.

Metadatasystemet vil kreve et apparat for teknisk vedlikehold og drift. Dette vil være sentral (system/drift) funksjon på rundt 1 årsverk som er ansvarlig for vedlikehold og videreutvikling av programsystemet, for at metadatabasen til enhver tid er intakt og effektiv m.h.p. plassutnyttelse, aksesstider etc.

Selve dokumentasjonsarbeidet må nødvendigvis utføres i de enkelte prosjekter, men hovedansvaret for veiledning og "håndheving" av dokumentasjonsstandarden kan ligge under en sentral dataadministratorfunksjon. Hvert fagkontor bør også ha en tilsvarende funksjon som er hovedansvarlig for kontorets metadata. En sentral faglig dataadministratorfunksjon vil utgjøre 1-2 årsverk.

IV Utvikling og bruk av datadok.-systemer i statistikk-byråer.

Dette kapitlet gir en grov oversikt og omtale av noen datadokumentasjonssystemer som er i bruk ved en del statistikk-byråer. Metadatahåndtering i de del database- og analyse-systemer for statistikk blir også omtalt.

IV 1 Datadokumentasjonssystemer i statistikk-byråer.

IV 1.1 VARKAT (SCB, Sverige).

SCB startet utviklingen av et dokumentasjonssystem i 1971, og siden 1974 har et maskinelt system (VARKAT) vært i drift. Det ble opprinnelig spesialprogrammert som et satsvis system, men er senere konvertert til Datamanager (et generelt data-katalog-system, se Kap. V 1.1) som kjøres pseudo-interaktivt under TSO.

VARKAT inneholder dokumentasjon om filer og variable. Det opereres med 6 ulike metadata-typer :

- variabel
- fil
- elementærfil (datafelt)
- statistikk-gren (innsamlende / anvendende st.gr.)
- emne-område
- objekt-type (enhet)

Variable beskrives ved bl. a. følgende egenskaper :

identifikasjon
navn
definisjon
kommentar

For variabelens forekomster i filer (dvs elementærfiler) angis bl.a.

felt-nummer (løpe-nr i posten)
benevning
objekttype
observasjons-tidspunkt
kvalitet

sikkerhetsgradering
Kommentar

De viktigste kjennetegn for fil-beskrivelser :

identifikasjon
benevning
framstillingstidspunkt
statistikk-grener
populasjon
datakilde
kvalitet
div. tekn. opplysninger

Emne-område brukes for søking (klassifisering) av variable, statistikk-gren for fil, objekttype for elementærfil.

Hovedformålene for VARKAT :

1. en samlet og enhetlig dokumentasjon som er lett tilgjengelig.
2. hjelpemiddel i forbindelse med oppdrags- og informasjonsvirksomhet.
3. verktøy for system/program-utvikling.
4. administrasjon av dataarkivet (langtidsarkivet).
5. forenkle kjøring av standardprogrammer ved at metadata (parametre) til disse hentes automatisk fra VARKAT.
6. utgjøre metadatabasen i RAM-modellen (integreert system for stat.prod.) med kobling til database-systemer (RAPID, AXIS) og alle funksjoner for stat.-produksjon.

I praksis har punktene 1 og 2 blitt oppfylt på de områdene dokumentasjon har blitt lagt inn. Forøvrig har VARKAT i liten grad blitt integrert med produksjonssystemet slik at den totale nytteeffekten har vært lav i forhold til kostnadene med å holde dokumentasjonen ajour. Systemet blir derfor idag lite brukt.

På systemutviklingssiden benyttes et annet dokumentasjons-system , DOK . Dette er en katalog (på edb) over arkiverte filer.

Referanser :

PTP-utredningen (1979) bilag 3.
Bo Sundgren: Meta-Information in Stat. Agencies,
ISIS-77.

IV 1.2 SDDS (Statistics Canada).

Statistics Canada har de siste årene gjennomført et dokumentasjonsprosjekt. En kravspesifikasjon til et statistisk data-dokumentasjonssystem (SDDS) foreligger fra dette arbeidet. Dessuten er det laget en prototype av systemet på edb.

SDDS omfatter først og fremst logiske metadata i form av dokumentasjon av statistiske undersøkelser, spørreskjema, publikasjoner, tabeller, enheter og variable. Det er 4 hovedtyper av variable :

- Kvantitativ variabel
- tekst-variabel
- identifikator (for enhet)
- kategorivariabel (grupperinger, stat. standarder)

Kategorivariabeler inndeles videre i 2 grupper :

- klassifikasjonsvariabel (strukturert)
- kvalitativ variabel (ustrukturert, disjunkt)

Makrodata (tabeller) beskrives ved tabell-definisjoner som følger syntaksen til TPL (Table Processing Language) og ved CANSIM-matrise-definisjoner (CANSIM databasesystem for makrodata).

Dokumentasjonsarbeidet foregår ved at de enkelte fag-kontor beskriver sine undersøkelser etter dokumentasjons-standarder i SDDS. Dette materialet blir sjekket, gitt identifiserende koder etc av en egen avdeling (Standards Division). Deretter blir det registrert v.h.a. tekstbehandlingsutstyr og siden overført til en metadatabase (tekst/dokument-base) i hovedmaskinen. Pr. mai 82 var ca. 70 undersøkelser dokumentert, men bare 8 av disse m.h.p. datainnhold (dvs variable, tabeller).

Referanse : R. Graves : More on information modelling
at Statistics Canada. ISIS-82.

IV 1.3 ABS DD/DS (Australian Bureau of Statistics).

ABS DD/DS (Data Dictionary/Directory System) er basert på et kommersielt data-katalogsystem (ADABAS DD) som er tilpasset og utvidet med funksjoner for ABS's behov. ADABAS database-håndteringssystem brukes i stor utstrekning i stat.prod. , og ADABAS DD er et hjelpemiddel for å utvikle og dokumentere applikasjonssystemer i ADABAS (se Kap. V om data-katalog-systemer).

ABS DD/DS inneholder først og fremst beskrivelser av data-filer som er lagret under ADABAS, men også andre fil-beskrivelser (vanlig sekvensielle filer og SAS-filer). Systemet har en rekke koblinger (interfaces) til edb-systemer som brukes i stat.prod. I tillegg til de funksjoner for metadata-generering som ligger i ADABAS DD , dvs. generering av database-beskrivelse (SCHEMA) datadefinisjons-blokker til COBOL, PL/I og NATURAL (et system for applikasjonsutvikling (generering)) , har ABS selv laget tilsvarende funksjoner for TPL (Table Processing Language) og PLEAT (Procedural Language for Edit Amend and Tabulation). De har dessuten preprosessorer for COBOL og PL/I som benytter data-dictionary-systemet mer dynamisk ved at disse systemene automatisk "slår opp i" metadatabasen og henter nødvendige metadata til programmene som kompileres, og dessuten oppdaterer metadatabasen med nye metadata.

ABS har på denne måten klart å opprette et meget aktivt metadatasystem m.h.p. systemutvikling. Det spiller derfor en sentral rolle i statistikk-produksjonen. De eneste systemene som ikke er koblet til ABS DD/DS, er SAS analyse-pakke og et grafisk system (IGL).

Referanse : E.W.W. Miller : The Integrated Statistical computing environment at ABS. Progress problems and prognosis. CES/WP 1983.

IV 1.4 Datadokumentasjon i Finland.

Statistikk-sentralen i Finland er i ferd med å bygge opp et datadokumentasjonssystem. Dette skal omfatte såvel logiske som fysiske metadata. Systemet omfatter ca. 20 metadatatyper, hvorav de viktigste er organisatorisk enhet (i stat. byrået),

statistisk enhet, dataelement (variabel), gruppe (av elementer), rekord, fil, output-beskrivelse (publikasjon/tabell).

DATAMANAGER (se Kap. V.1) brukes for dette systemet og i hovedsak følger metadatastrukturen en av standard-malene i dette systemet. Systemet kjøres på en IBM 4341-maskin (DOS/VSE) under CICS.

I tillegg til nevnte system er det i Finland et fil-katalog-system (RERE) for administrative registre. Dette inneholder administrative opplysninger, opplysninger om enhet, populasjon, innhold osv.. RERE kjøres på Statens datamaskinsentral under MINTTU-systemet (IR-system) og benyttes av flere sentrale myndigheter.

IV 1.5 Datadokumentasjon i Danmark.

Danmarks Statistikk har ennå ikke noe overordnet datadokumentasjonssystem, men har bygd opp endel metadata-systemer i tilknytning til enkelte "SAS-databaser" (om SAS se Kap.II.2.2). Disse består av metadata-funksjonene i SAS, men er utvidet ved hjelp av de generelle datahåndterings- og programmeringsverktøy i SAS.

IV 1.6 Metadata systemet for 3.Kontors Kom.øk.-database.

I forbindelse med prosjektet "Database for Kommunal-økonomisk statistikk" er det utviklet et enkelt interaktivt metadata-system. Det inneholder følgende metadata-typer :

undersøkelse
emne
enhet
variabel
datamengde (beskrivelse av uttak)
bruker
autorisasjon

fil, post, felt

Metadatabasen benyttes for å dokumentere dataene i Kom.øk.-db., men er uavhengig av statistikkområde. Den benyttes dessuten for

å besørge en høy grad av datauavhengighet for uttaks-funksjonene. Disse arbeider på logiske datamengder (samling av variable, uavhengig av filer, post-struktur o.l.).

Systemet kjøres under CICS (skjermbilde-interface) med VSAM som aksessmekanisme. Rapportgenerering foregår satsvis (initieres interaktivt fra skjerm).

IV 2 Statistiske database- og analysesystemer.

Dette kapitlet gir en kort innføring i metadata-bruken i en del stat. DBHS og analysesystemer som er eller kan bli tatt i bruk i Byrået. Formålet er dels å avdekke behovet for metadata i slike systemer og dessuten finne ut om noen av systemene i seg selv kan fungere som et komplett (integrert) dokumentasjonssystem.

Oppsummering, konklusjon :

Ingen av systemene kan uten videre fungere som et overordnet dokumentasjonssystem for Byrået.

RAPID er særlig egnet for lagring og dokumentasjon av primærdata. Det har interfacer til mange standardprogram (EASYTRIEVE, TPL, SAS, TAB68 (via SARA) m.m.). Dokumentasjonssystemet i RAPID er utbyggbart og kan i prinsippet dekke de viktigste metadata i statistikkproduksjonen. Det har derimot begrensede uttaksmuligheter og er lite sluttbrukerorientert.

SIR vil kunne brukes som et enhetlig system for gjennomføring av statistiske undersøkelser (datakontroller, lagring og dok. av mikrodata, tilknyttet tabell-generator (TPL), interface til SAS, SPSS, BMDP). Metadata-systemet i SIR er ikke utbyggbart og har begrensede søkemuligheter.

SAS er primært et analysesystem, men inneholder såvidt mange funksjoner for datahåndtering og presentasjon at det kan fungere som et integrert statistikk-system. Dokumentasjonssystemet i SAS er ganske begrenset.

AXIS og LASD inneholder de mest brukerorienterte metadata-systemene av de omtalte produkter. De dekker derimot primært makrodata lagret i tilhørende databaser.

IV 2.1 SIR (Scientific Information Retrieval).

SIR er laget med tanke på brukere av statistiske analysepakker. I tillegg til programmer for datahåndtering og dokumentasjon har systemet tabell- og rapportgenerator.

Metadatadefinering:

CASE (opplysninger om undersøkelsesenheten):

case id (identifikasjon)
n of cases (antall observasjoner)
recs per case (dataposter pr. undersøkelsesenhhet)
max rec types (max. antall posttyper i databasen)
common vars (lista må alltid inneholde case id)
common security (les-/skrivsikring på common vars)
document (kommentarer)

RECORD (datapost i databasen):

record shcema (posttype)
sort ides (postident)
rec security (les-/skrivsikring)
variable list
input format
compute (for transformering av variable)
if (for transformering av variable)
valid value (gyldige kodeverdier)
var range (gyldige verdier for variabelen)
var security (les-/skrivsikring)
var labels
value labels
accept rec if
reject rec if
document (kommentarer)

Funksjoner for registrering og oppdatering av metadata:

Kommandoene blir skrevet i et SPSS-lignende språk.
Noen egenskaper ved "case" kan modifiserse. "RECORD" kan modifiseres/redefineres.

Funksjoner for spørring, rapportering:

Hele eller deler av databasebeskrivelsen kan listes (kommando-språk).

Datasikring, -integritet:

Data kan sikres på post og variabelnivå mot ulovelig lesing og skrivning med sikkerhetsnivåer fra 0 til 30.

Metadatagenerering.

Systemet genererer metadata statistisk til SPSS, BMDP og SAS.

Krav til utstyr.

"SIR" kjøres på IBM under OS, TSO og VW/CMS.

IV 2.2 SAS (Statistical Analysis System).

SAS inneholder programmer for datahåndtering (avledning, kopling, selektering, redigering), tabell-/rapportgenerering og statistisk analyse.

Metadatadefinering:

DIRECTORY (filkatalog) inneholder opplysninger om :

dsname til library (fysisk lagring av metadata)
filnavn på alle datasettene
antall obs. for hvert datasett
" tracks"
antall tracks totalt

DATASET (datafil-beskrivelse) :

dsname og andre metadata om fysisk lagring av dataene
variabelnr (rekkefølgenr i dataposten)
variabelnavn
type
lengde
posisjon (i dataposten)
format
label
sas-parametrene som genererer datafilen
history (datafilens opprinnelse)

Funksjoner for registrering og oppdatering av metadata:

Metadataene blir registrert samtidig med at datafilen blir dannet. Oppdatering kan skje i begrenset grad ved proc datasets (options) og proc format som er sas utilities.

Funksjoner for spørring, rapportering:

Hele eller deler av databasebeskrivelsen kan listes ved kommandoen proc contents (options).

Datasikring, -integritet.

Data kan sikres på filnivå mot ulovelig lesing og skriving ved bruk av passord.

Metadatagenerering.

Systemet genererer metadata til BMDP dynamisk.

IV 2.3 RAPID MDM (META DATA MEMORY).

RAPID er laget i Statistics Canada og brukes av mange statistikkbyråer. RAPID er et statistisk DBMS (relasjonsmodell) og er særlig beregnet for mikrodatahåndtering.

Metadatadefinering:

RELATION (databasefil) :

navn
variabelliste
variabelspesifikasjoner
kommentarer

FILE (ekstern fil) :

navn
variabelliste
input format
kommentarer

RECORD (brukertilpasset databasefil) :

navn
variabelliste
variabelspesifikasjoner
kommentarer

CODESET (brukes for eksterne filer) :

variabelnavn
kodenavn
kodespesifikasjoner
kommentarer

VALUESET (brukes for databasefiler) :

variabelnavn
kodenavn
kommentarer

CLASS (grupper av logiske variabelverdier) :

overordnet kodenavn definert under valueset
kodenavndefinert under valueset
kommentarer

Brukerdefinerte metadatatyper.

Funksjoner for registrering og oppdatering av metadata:

Satsvis og interaktiv registrering og oppdatering av metadata ved kommandospråk.

Funksjoner for spørring, rapportering:

Hele eller deler av databasebeskrivelsen kan listes ved kommandoen DISPLAY (options) interaktivt og satsvis.

Datasikring, -integritet.

Få mekanismer for sikring/autorisasjon. Vanlige dba-funksjoner for backup/recovery (unload/load).

Metadatagenerering.

Systemet genererer metadata til en rekke programpakker og til programmer skrevet av brukere.

Krav til utstyr.

RAPID kjøres på IBM under OS og TSO.

IV 2.4 AXIS (Auxiliary system for interactive statistics).

AXIS-database består av tabeller på detaljert nivå. Ved dialog med axis kan brukere bearbeide tabelldata (aggregere, beregne) og få resultatet presentert på terminalen med tekst.

AXIS har en operasjonell metadatabase hvor tabellenes/data-matrisenes struktur er beskrevet. Navn, tekster, kommentarer osv lagres i et tilknyttet "lexikon". Metadatabasen er meget omfattende og består av 5 deler : lagrings-, innholds- og koblingsbeskrivelse, konversasjons- og bruker-katalog.

IV 2.5 LASD (Large scale statistical data system).

LASD-database består av mikrodata og makrodata, ved dialog med LASD kan brukere lage tabeller og bearbeide tabelldata (aggregere, beregne) og få tekstede tabeller skrevet ut.

Metadatadefinering:

SEGMENT (gruppe av tabellmatriser)

navn
seleksjonskriterier (omfang skrevet i klartekst)
kategorivariabel
statistikkvariabel (summeringsvariabel)
variabelnavn

CRIT (kategorivariabel)

navn
Kodeverdi
Kodenavn

BASE FILE (sekvensielle filer med grunndata)

filnavn
statistikkvariable
kategorivariabel
kommentarer

Funksjoner for registrering og oppdatering av metadata:

Programmer for registrering og oppdatering av metadata både satsvis og interaktivt.

Funksjoner for spørring, rapportering:

Metadata om segment, crit og base-fil kan listes.

IV 2.6 TROLL.

Data-baser i TROLL organiseres v.h.a. et hierarkisk arkivsystem.
TROLL-variable (tidsserier) defineres ved

navn
periodisitet
starttidspunkt
slutt-tidspunkt
kommentarer.

IV 2.7 EASYTRIEVE .

EASYTRIEVE er et programsystem for rapportgenerering og enklere filbehandling.

Metadatadefinering:

MACRO (fil- programkatalog)

navn
programtekst (parametre)

FILE (datafil)

navn
fysike egenskaper
variabelnavn-liste
start
lengde
type
Kommentarer

Funksjoner for registrering og oppdatering av metadata:

Metadataene blir registrert ved kommandoen ADD og oppdatert ved kommandoene CHANGE eller DELETE.

Funksjoner for spørring, rapportering:

Hele eller deler av filkatalogen kan listes ved DISPLAY-kommandoen.

V Oversikt og vurdering av datakatalogsystemer.

I dette kapitlet behandles såkalte "Data Dictionary / Directory System"-er (DD/DS), som på norsk gjerne benevnes som data-katalog-systemer. Det er også tatt med en kort oversikt over dokumentsøkesystemer (IR-systemer).

V 1.1 Data-katalog-systemer (DD/DS).

Disse systemene er spesielt konstruert for lagring og håndtering av metadata. De består i hovedtrekk av følgende komponenter :

- en database som inneholder beskrivelser av data (navn, definisjon og ulike egenskaper), dvs såkalt "dictionary"-opplysninger, og dessuten opplysninger om hvor og hvordan data er lagret (directory).
- funksjoner for å registrere, oppdatere og hente ut metadata , rutiner for sikring av metadata osv. .
- rapport-generator for å produsere ulike rapporttyper fra metadatabasen.
- verktøy for metadata-generering, som regel av data-beskrivelser til program (COBOL, PL/I), et eller flere databasehåndteringssystemer.

De fleste datakatalogsystemer er avhengig av et bestemt databasehåndteringssystem (DBHS), og er primært utformet som et dokumentasjons- og systemutviklingsverktøy for datasystemer som implementeres v.h.a. det aktuelle DBHS.

De uavhengige datakatalogsystemene er ikke laget v.h.a. et generelt DBHS, men inneholder egne skreddersydde funksjoner for lagring og handtering av metadata. Disse systemene "støtter" som regel et større utvalg av databasehåndteringssystemer enn de DBHS-avhengige systemer, men er selvfølgelig ikke så sterkt integrert med et DBHS. De kan derfor ikke brukes like aktivt i utvikling og drift av database-systemer.

Nytteverdien av et datakatalogsystem beror for en stor del på om det er koblet til andre programsystemer. Det vanligste er at metadatabasen kan brukes av kompilatorer, DBHS og ulike sluttbrukerspråk (rapportgeneratorer, spørrespråk).

Koblingen er som regel statistisk, dvs. databeskrivelser hentes ut av metadatabasen og formateres etter mottakerprogrammets syntaks-krav av en separat funksjon som er knyttet til metadatasystemet. Det dannes derved en fil som er inndata (styredata) til det aktuelle programsystem. Ved dynamisk kobling foregår metadatagenereringen automatisk på det tidspunkt metadatabasene skal brukes av en funksjon. På den måten er en sikret at de riktige metadatabasene brukes til enhver tid. Dynamisk koblede programsystemer må ha innebygde rutiner for å aksessere metadatabasen.

Det er en klar trend i utviklingen av DBHS og DD/DS at metadatasystemet blir mer aktivt og integrert med funksjonene i informasjonssystemene. De dekker mer logiske (problemorienterte) metadatabasene og muliggjør derved også mer høynivå databehandlingsspråk. Også på nye felter som distribuerte databaser og utviklingen av mer intelligente DBHS, er metadatabehandlingen meget viktig.

Alle de nedenfor omtalte systemene er i hovedsak beregnet på administrativ systemutvikling. Spesielle egenskaper ved statistiske data (tidsserier, fordelte variable, komplekse aggregeringer, særskilte integritets- og sikkerhetsregler o.s.v.) blir i liten grad dekket. I de fleste systemene er det derimot mulig å lage sine egne metadatabasetyper.

De generelle datakatalogene er i første rekke edb-tekniske hjelpemidler og er trolig lite egnet for fagkontorenes behov for rask og enkel tilgang til dokumentasjon. Brukergrensesnittet til de fleste systemene består av relativt komplekse kommandospråk selv om de fleste leverandører etterhvert har laget interaktive skjermbilde-språk. De mest brukervennlige er systemer som PREDICT og TIS mens mer omfattende systemer som DATAMANAGER, DATACOM/DD og IDD virker tyngre i bruk.

Som produksjonsteknisk hjelpemiddel gir de fleste systemene god støtte for spesialprogrammering, men derimot ingen for bruk av Byråets standardprogrammer. "Interfacen" til disse må derfor lages spesielt, og i de fleste systemene er det tatt hensyn til slike koblinger. Forøvrig må valg av datakatalogsystem og databehandlingssystem sees i sammenheng. Til et moderne DBHS hører også et integrert datakatalogsystem.

V 1.2 Oversikt over noen aktuelle DD/DS.

Kommersielt tilgjengelige datakatalog-systemer er mest utbredt på IBM og IBM-kompatible maskiner. Denne oversikten begrenser seg til disse samt systemer som kan kjøres på NORD-

maskiner. På IBM-siden er de mest kjente DBHS-uavhengige data-katalogsystemene følgende:

DATA CATALOG 2
 DATAMANAGER
 PRIDE-asdm

Av disse er DATAMANAGER mest brukt. Data Catalog 2 har en viss utbredelse i USA, men liten i Europa.

DBHS-avhengige datakatalog-systemer :

På IBM :

PREDICT	(Software AG, ADABAS)
DATA COM/DD	(ADR, DATA COM/DB)
DB/DC Data Dictionary	(IBM, DL/I)
DCS	(Cincom, TOTAL)
IDD	(Cullinet, IDMS)
TIS Directory	(Cincom, TIS)
Catalog for rel.dbhs.	(SQL, ORACLE etc.)

På NORD :

ND DDS	(Norsk Data, SIBAS)
SYSDOC	(Sysdeco, SIBAS)

I de etterfølgende avsnittene gis en kort (og ufullstendig) presentasjon av de mest aktuelle systemene. Følgende egenskaper behandles :

- metadata-definering.
 (hvilke typer metadata kan beskrives, hvilke egenskaper kan de ha, relasjoner mellom metadata)
- funksjoner for registrering og oppdatering av metadata.
 (kommando-språk, skjermbilder, satsvis, interaktiv)
- funksjoner for spørring, rapportering.
 (søkemuligheter, spørrespråk)
- data-sikring, -integritet.
- metadata-generering.
 (statisk, dynamisk)

- Krav til utstyr (operativsystem o.l).
- pris ved kjøp / leie .

V 1.2.1 DATAMANAGER.

DATAMANAGER fra MSP Inc. ble lansert i 1974 og er inntil nå solgt til over 900 installasjoner (det mest utbredte frittstående datakatalogsystemet på markedet) .

Metadata-definering :

Basis-strukturen består av dokumentasjons-enhetene ("members") fil, gruppe, dataelement og system, program, modul med tilhørende relasjoner. Det er mulig å opprette brukerdefinerte metadata-typer.

Kjennetegn ("for DATAMANAGER-member") :

- navn (inntil 32 tegn)
- status
- beskrivelse (ubegrenset lengde)
- inntil 16 alternative navn (alias) som kan defineres for ulike brukstyper (f.eks. navn i COBOL-progr.).
- nøkkel-ord (emneord) (ubegrenset antall)
- + spesielle kjennetegn avhengig av metadata-type
- + brukerdefinerte kjennetegn

En kan opprette mange metadatabaser (dictionaries) under DATAMANAGER.

Oppdaterings-funksjoner :

Kommandospråk for registrering og oppdatering av metadata (både ved satsvis og interaktiv bruk).
Defaultverdi-mekanisme.
Generering av "dummy-members".
"Setup-facilty" for innlegging av metadata fra eksisterende program.

Uttaks-funksjoner :

Et stort utvalg av rapporter kan lages v.h.a. kommandoene REPORT, LIST, GLOSSARY, PRINT (oversikt-, detalj-, kryss-referanse-, nøkkelord-rapporter m.m.).

Språk for spørring/søking .

Sikkerhets-funksjoner :

Grunnversjonen inneholder integritetsfunksjoner (backup, recovery).

Sikkerhetssystem fåes som tilleggs-modul , hvor eiere av metadata, brukere av metadata, autorisasjoner kan defineres.

Metadatagenerering :

Som tilleggsfunksjoner leveres

- Source-language-generatorer for COBOL, PL/I, BAL, MARK IV.
- DDL-generatorer for ADABAS, IDMS, IMS, SYSTEM 2000, TOTAL.
- User Interface (aksess-rutiner, rutine-kall i inndata-program og rapport-program i DATAMANAGER)
- Beslektede produkter (fra MSP), SOURCEMANAGER og DESIGN-MANAGER, for programutvikling og databasekonstruksjon

Krav til utstyr :

Kan kjøres satsvis under OS, interaktivt v.h.a. CICS, TSO (tilleggsfunksjon). Krever ca 200 Kb primærhukommelse.

Pris :

Grunnversjonen koster ca. kr. 100.000
Hver tilleggsfunksjon -"- 30.000

V 1.2.2 PREDICT.

PREDICT er datakatalogsystemet til databasehåndteringssystemet ADABAS fra Software AG, som er et av de mest utbredte DBHS på markedet. Det brukes på mange områder, også til statistikk. Kan håndtere store datamengder, har fleksible søkemekanismer (inverterte lister).
Til ADABAS hører

ADAMINT makro-DML (for programmering mot ADABAS)
ADASCRIP+, ADACOM spørresystem, rapportgenerator
PREDICT data-katalog-system.

NATURAL et uavhengig system for utvikling av applikasjoner (4.gen.språk).

PREDICT er laget fullstendig v.h.a. NATURAL og ADABAS.
Det er interaktiv og følgelig ganske brukervennlig.

Metadata-definering :

I PREDICT beskrives databaser, filer og de enkelte data-elementer. Dessuten såkalte "user views", dvs ulike brukerorienterte syn på data (fil-beskrivelser).
Brukere og funksjoner beskrives ved entitetene personell, system, program, modul, rapport.

De viktigste metadata-attributter :

- navn 3-32 tegn
- Kommentarer ubegrenset antall tekstlinjer (a 30 tegn)
- inntil 99 synonymer på dataelement-navn
- inntil 200 emneord per filbeskrivelse
- filtype s.a. ADABAS-filer, sekv. filer, VSAM-filer og andre kan beskrives.

Oppdaterings-funksjoner :

- fast format oppdateringstransaksjoner
- skjermbildesystem med editor.
- eksisterende ADABAS-DDL (database-definisjoner) og databeskrivelser (DDM) i NATURAL kan leses for å oppdatere datakatalogen.

Uttaks-funksjoner :

- 20 standard rapporter (spørrespråk, satsvis rapp.-gen.)
- ADASCRIP+, ADACOM, NATURAL kan brukes for å lage rapporter, utføre spørringer.

Sikkerhets-funksjoner :

Funksjonene i ADABAS benyttes siden datakatalogen er implementert som en ADABAS-database. Datasikring på database-, fil-, felt-, verdi-nivå.
(Ikke skille eier/brukere av data)

Metadatagenerering :

Generering av ADABAS databasedefinisjoner.
----"---- COBOL, PL/I datadefinisjoner.
Support for ADASCRIP+, ADACOM, ADAMINT, NATURAL.

Krav til utstyr :

Kan kjøre under DOS, OS ,MVS og CMS.
ADABAS benytter sin egen TP-monitor COMPLETE, men
kan også bruke CICS og TSO.

Pris :

Datakatalogen selges som en tilleggsmodul i ADABAS (ikke
separat). ADABAS koster ca. 1 million kroner og PREDICT
ca. 140.000 kroner.

V 1.2.3 DATACOM/DD.

DATACOM/DD er knyttet til databasehåndteringssystemet DATACOM/DB
fra ADR (Applied Data Research), som er et liknende type DBHS
som ADABAS. ADR's produktserie på databaseområdet omfatter
DATACOM/DB, DATACOM/DC (kommunikasjon), DATACOM/DD, DATAREPORTER
(rapport-gen.), DATAQUERY (spørrespråk), IDEAL (utviklingssystem)
DATADESIGNER (databasekonstruksjon), D-NET (distribuert nett).

Metadata-definering :

I tillegg til standard-entiteter (metadatatyper) som
database, fil, felt, system, program etc kan brukere
definere sine egne metadata-typer og et ubegrenset antall
relasjoner mellom disse (1:n-rel., rekursive rel. med til-
knyttet informasjon).

Metadata-attributter :

- navn (15 tegn)
- versjon
- status (3 typer)
- passord
- dato når inngitt / sist endret
- en beskrivelse (tekst, ubegr.)
- emneord (ubegr.)
- alias (15 stk)
- + spesielle for ulike metadata-typer
- + brukerdefinerte

Oppdaterings-funksjoner :

- Interaktiv oppdatering ved DDOL (DataDictionary On Line). Meny-styrt, men også kommandospråk. Skjerm splitting.
- Ekstrahering av datadef. fra COBOL-, PL/I-program.

Uttaks-funksjoner :

- Stort utvalg av rapporter.
- Visse muligheter for å lage egne rapport-typer.

Sikkerhets-funksjoner :

- funksjonene i DATACOM/DB.
- egne kontrollrutiner ved oppdatering.
- egen sikkerhetsmekanisme for metadata (Person, Authorisation)

Metadatagenerering :

- Generering av datadef. for COBOL, PL/I, ASSEMBLER.
- AKsess-rutiner for metadatabase (DSF Datadict. Serv. Facility)
- Dynamisk kobling (via DSF) for DATAREPORTER, DATAQUERY og IDEAL.
- DATA DESIGNER ,verktøy for databasedesign.

Krav til utstyr :

Krever DATACOM/DB og CICS (el. DATACOM/DC).
Kjøres under OS el. DOS operativsystem.

Pris :

DATACOM/DD koster ca. 26.000 dollar som "Stand-alone"-system.
(rabatt hvis det kjøpes sammen med DATACOM/DB).

V 1.2.4 Data Control System.

DCS fra Cincom er en nyutviklet versjon av TOTAL DD. DCS er et generelt datakatalogsystem som kan brukes interaktivt. Det er laget vha TOTAL (DBHS) .

Metadata-definering :

- TOTAL-databaser, andre filtyper
- funksjoner (system, program, modul)
- opplysninger om brukere m.m.

Oppdaterings-funksjoner :

- Interaktiv (menystyrt, skjermbilder)
- Ekstrahering av datadef. fra eksisterende program og DDL.

Uttaks-funksjoner :

- Interaktivt spørrespråk
- 25 standard spørreprosedyrer
- satsvis rapportgenerering

Sikkerhets-funksjoner :

- automatisk recovery (TLR Transac. Level Rec.).
- passord, autorisasjon av DCS-bruk.
- Sikkerhetssystem i DCS for applikasjonsdatabaser.

Metadatagenerering :

- System Design Facility
systemutviklingsmetode med datamaskinstøtte
(logisk systemkonstruksjon, logisk datamodellering,
data-normalisering, fysisk datastrukturering)
- Programmer Work Station
editering, kodegenerering m.m.

Krav til utstyr :

Kan fåes for de fleste operativsystem på IBM.

V 1.2.5 Integrated Data Dictionary.

IDD er datakatalogsystemet til IDMS (integrated Data Management System) fra Cullinett, det mest utbredte CODASYL-DBMS på stor-maskiner. Til IDMS hører CULPRINT rapportgenerator, Online Query (spørrespråk), Online English (naturlig språk) m.m. . Dessuten ADS applikasjonsutviklingssystem.

Metadata-definering :

Systemet inneholder standard metadatatyper for data og funksjoner (database, fil, felt, program, etc) og også brukerdefinerte metadatatyper.

Metadata beskrives ved bl.a. :

- navn (1-32 tegn)
- beskrivende navn (40 tegn)
- definisjon (tekst ubegr.)
- emneord (40 tegn, ubegr. antall)
- synonymer for fil-navn, rekord-navn, element-navn

Oppdaterings-funksjoner :

Metadata beskrives ved et eget språk (DDDL Dict. Data Def. Language) og det er kommandoer for oppdatering/vedlikehold av metadatabasen (ADD, MODIFY, DELETE, EDIT).

Ekstrahering av datadef. i program (COBOL, PL/I).
DDL-kompilator oppdaterer metadatabasen.

Uttaks-funksjoner :

Rapportgenerator (60-70 rapport-typer).

Sikkerhets-funksjoner :

Sikkerhetsregler for aksessering/oppdatering av metadata.
(brukere/eiere av data).

Metadatagenerering :

Dynamisk kobling for preprocessor (COBOL, PL/I, RPG, BAL)
og sluttbruker-språk.

Krav til utstyr :

IDMS tilgjengelig på de fleste IBM-op.-systemer.

Pris :
Inngår i IDMS .

V 1.2.6 TIS Directory.

TIS (Total Information System) er en relativ ny software-produkt-serie fra Cincom hvor de enkelte komponentene er sterkt integrert v.h.a. en såkalt "In line Directory".

TIS består av :

- Logical User Views
- In line Directory
- DBMS (databasehåndtering)
- Intelligent Query
- Comprehensive Retrieval
- MANTIS (applikasjonsutviklings-system)
- DML (Data Manipulation Language)

Metadatabasen (TIS Directory) inneholder de brukerorienterte data-beskrivelsene (Logical User Views, Queries) og fysiske metadata. Alle funksjoner i TIS kan brukes mot metadatabasen. Det er ikke mulig å opprette brukerdefinerte metadatatyper i TIS Directory. De kan eventuelt implementeres som en vanlig applikasjonsdatabase.

V 1.2.7 Metadata i relasjons-DBHS.

Det har kommet en del DBHS på markedet som bygger på relasjonsmodellen. De mest kjente er DB2 (fra IBM), ORACLE og INGRES. Også ACCESS (fra ND) er et relasjonsorientert spørrespråk. I tillegg er endel av de eldre kommersielle DBHS (ADABAS, DATACOM, IDMS) utbygd i denne retning.

Metadata kan i relasjons-DBHS håndteres mer dynamisk enn i tradisjonelle DBHS. Alle metadata i disse systemene er lagret i en tilhørende relasjonsdatabase som kan behandles med de samme funksjonene (språk) som bruker-databasene.

De viktigste metadatatyper i disse systemene er :

- relation/table (rel.-tabell, fil)
- attribute/column (kollonne, felt)
- view (bruker-modell)
- user (bruker)

- index (aksess-nøkler)

Metadatabasen har ikke brukerdefinerte metadatatyper, men slike kan implementeres ved å opprette egne filer som kan håndteres med de ordinære funksjonene i DBHS.

V 1.2.8 ND Data-Dictionary-System.

ND DDS er et dokumentasjons- og utviklingshjelpemiddel for anvendelser som lages v.h.a. SIBAS (dbhs) og NSHS (skjermhåndtering).

Metadata-definering :

Omfatter SIBAS-database-definisjon, skjermbilde-definisjoner, angivelse av funksjoner og subfunksjoner.

Oppdaterings-funksjoner :

Skjermbilder.

Uttaks-funksjoner :

Rapportgenerator under utvikling.

Sikkerhets-funksjoner :

Funksjonene i SIBAS.

Metadatagenerering :

SIBAS-DRL-schema.
Fortran/COBOL-datadef.
Integrert med Nord Screen Def. System.

Krav til utstyr :

SIBAS, NSHS.

V 1.2.9 SYSDOC.

SYSDOC er en systemutviklingsmetode og et dokumentasjons-system. Det inngår i et applikasjonsutviklingssystem (SYSTEMATOR) utviklet ved SI og markedsføres av SYSDECO a/s.

Metadata-definering :

Konseptuell datamodell (entitet/relasjon-modell) og funksjoner (rapport/skjermbilde, høynivå prosedyrespråk) beskrives og lagres i SYSDOC-databasen (metadatabase). Fysiske metadata genereres i en delvis automatisk design-prosess.

Oppdaterings-funksjoner :

- Skjermbilder (menystyrt) på NORD-versjonen.
- + Kommandospråk

Uttaks-funksjoner :

Stort utvalg rapporter (spesifikasjonsrapporter, kryss-referanserapporter, nøkkelordsrapporter, analyserapporter)

Sikkerhets-funksjoner :

Metadatabasen implementert som en SIBAS-database.
Ved siden av funksjonene i SIBAS finnes skreddersydde recovery-program.

Metadatagenerering (ved andre SYSTEMATOR-funksjoner) :

Generering av database-definisjon (SIBAS-DRL).
Kodegenerering (Fortran).

Krav til utstyr :

- SIBAS.

V 2. IR-systemer.

IR-systemer (Information Retrieval) er beregnet for søking i store registre av "dokumenter" (tekstlig informasjon, kan også inneholde mer strukturerte data).

En database under et IR-system består gjerne av 2 deler :

- dokument-fil (inneholder selve informasjonsmengden).
- invertert fil som benyttes for søking.

I de fleste systemene kan en base kun bestå av en dokument-type. De fleste kan derimot operere på flere enn en database. Den inverterte filen inneholder alle søkbare ord i basen sammen med en referanse-liste til dokumenter hvor de forekommer. Ved definering av dokument-typen (databasen) bestemmes hvilke deler av dokumentet som skal være søkbart. I noen systemer kan alle ord gjøres søkbare eventuelt med unntak av en nærmere angitt liste av "stopp"-ord.

IR-systemer benyttes mest for referansesøking, og spørrespråkene i disse systemene er godt utbygd for slike formål og dessuten som regel svært enkle å bruke.

Alle systemer har satsvise registrerings- og oppdateringsfunksjoner, mange også interaktive. Endel systemer er godt integrert med tekstbehandlingssystemer fra samme leverandør.

Systemer på IBM :

STAIRS (VS,DL/1)
CDS/ISIS
IMDOC
MINTTU
IMDOC
POLYDOC

Systemer på NORD :

NOTIS-IR (NOVA*STATUS, SIFT)
POLYDOC
VERBAL

Disse systemene (og flere) beskrives og evalueres m.h.p. lagring og søking av publikasjoner/tabeller i DOK-PUBL-prosjektet.

Styrken til IR-systemene ligger i de enkle men slagkraftige brukerspråk, fritekstsøking og ganske stor fleksibilitet i definering av attributter (kjennetegn) til metadata (dokument). Å representere en kompleks metadatabase-struktur (mange metadata-typer og relasjoner mellom disse) v.h.a. et slik system kan muligens by på problemer. Ingen av systemene har en eksplisitt mekanisme for å definere relasjoner (unntatt i tesaurus), men ved å opprette innbyrdes referanser i dokumentene kan en ved søking knytte relaterte dokumenter sammen (ingen automatisk sjek ved oppdatering). Systemene av denne type har inegen tilsvarende funksjoner for metadatagenerering som datakatalogene, og har heller ikke interfacer til andre programmer med unntak av tekstbehandlingssystemer.

Referanser :

1. Dokumentasjon av data i Statistisk Sentralbyrå.
RHA/EAU/EH, 9/12-74.
2. Forarbeider til et variabelkatalogprosjekt i SSB. IO 73/8.
3. Notat om spørsmålsarkiv. TH/WJ, 22/3-74.
4. Spørsmålarkivet. Notat om bakgrunnskjenntegn. TH/GH, 3/9-74.
5. Håndbok for systemkontoret. Avsnitt 3: Dokumentasjon.
6. Håndbok nr. 40. Især avsnitt VII.2.c til VII.2.g.
7. Håndbok nr. 10. Hjemmel for innkreving av oppgaver.
8. Håndbok nr. 23. Utsnitt om prinsipper og definisjoner i offisiell statistikk.

VEDLEGG A Filbehandlingsoperasjoner.

- selektering av rekorder i fil.
- ekstrahering av felter fra rekord.

- Kobling av filer (join,match).
- sammenslåing av filer (union,merge).
- sortering av fil.

- aggregering av fil (opptelling, summering).

- omstrukturering av fil.

Syntaks for filoperasjonene :

- Selektering angis ved en betingelse for aktuelle felt (variable i rekorden (bruk av relasjonsoperatører =,<,>,... og logiske operatører og, eller, ikke).

- Ekstrahering av felter angis ved navn på felt (variabel) eller posisjon i rekord.

- Kobling :

```
      KOBLE  fil1(nøkkel-1),fil2(nøkkel-2),.. TIL utfil  
      -----
```

- Sammenslåing (union) :

```
      UNION AV fil1(nøkkel-1),fil2(nøkkel-2) ... TIL utfil  
      -----
```

- Sortering :

```
      SORTER  fil  ETTER var-liste  TIL utfil  
      -----
```

- Aggregering :

```
      AGGREGER  fil  ETTER var-liste  TIL utfil  
      -----
```

- Omstrukturering :

```
      normalisering av ikke-normalisert fil.  
      denormalisering av normalisert fil.
```

VEDLEGG B Avledning av variable (Forslag til syntaks).

Avledning av nye variable beskrives formelt ved aritmetiske uttrykk. Disse kan være betingede s.a. en variabel gis verdi utfra hvilke betingelser som er oppfylt (gruppering, omkoding).

SYNTAKS :

variabel = <aritmetisk uttrykk>;

eller for betinget avledning :

variabel = <aritm.-uttrykk-1> hvis <betingelse-1> ,

<aritm.-uttrykk-2> hvis <betingelse-2> ,

.....

<aritm.-uttrykk-n> ellers ;

Aritmetiske uttrykk består av variable og/eller konstanter (som operander) satt sammen av operatorer (+,-,*,/) og eventuelt paranteser (rundt deluttrykk).
eks. : $x + 5*(y+z)$.

Et uttrykk kan bestå av kun en variabel el. konstant (aktuelt ved gruppering).

En betingelse består av variable, konstanter, relasjonsoperatore (=,<,>,...), logiske operatorer (og,eller,ikke) samt paranteser.
eks. : $Kj\ddot{o}nn = 1$ og $alder > 20$

VEDLEGG C Kontroller (Forslag til syntaks).

- Validitetskontroller

Innebærer å sjekke at en variabel-verdi er av den type (numerisk,...) som variabelen er definert som, og at den hører til den verdi-mengden som variabelen kan anta verdier fra. Validitetskontroller spesifiseres derfor indirekte ved definering av verditype (datatype) og verdi-mengde for variabel.

Verdityper :

numerisk $\{ k \mid k.d \}$ (k: antall sifre, d: antall desimaler)

alfanummerisk k (k: antall tegn)

tekst

logisk (sann/gal)

Kodet $k_1 [/k_2 /k_3 \dots]$ (k: antall sifre i kode, hierarkisk koding angis ved spesialtegn (/))

+ spesielle verdityper for f.eks. dato(er), pers.nr. etc.

Verdi-mengde :

angis ved

- intervaller $v_1 : v_2$

- diskrete verdier v_1, v_2, v_3

- katalog, register, kodeliste (navn på fil/publ.)

eks. : 0,3,5:7,9

Ved koding kan en forklarende tekst angis for hver verdi (kode)
eks. :

treslag :=1 (furu), 2 (gran), 3 (bjørk), 0 (annet)

Lange verdi-lister lagres i kataloger som refereres ved variabel-definering. Sjekking mot kataloger benevnes gjerne som eksistens-kontroll .

- Krysskontroller (logiske kontroller)

Betyr å sjekke sammenhengen mellom ulike variable, at visse betingelser mellom variable er oppfylt. Mange slike kontroller består egentlig i å sjekke at variable som egentlig er avledet av andre variable, er blitt utregnet riktig. I såfall benyttes den definerte avledningsregel i kontroll-prosedyren. Når det gjelder fordelte variable er en del sum-kontroller gitt ved (den hierarkiske) oppbyggingen av fordelingsvariablene (klassifikasjonsvariable).

Forøvrig kan logiske kontroller defineres v.h.a. følgende syntaks :

```

{variabel
{<aritmetisk uttrykk>}}

<rel.-op.> { <verdi-mengde>
             { <aritm.-uttrykk> } hvis <betingelse> ,
-----"-----
.....
-----"-----
                                ellers ;

```

Alternativt kan betingelsene (premissene) settes først s.a. en får konstruksjonen

```

hvis <betingelse> så (skal) .....

ellers-hvis -----"-----
.....

ellers (skal) .... ;

```

I de enkleste tilfellene vil kontrollen kun bestå av en betingelse.

Spesielle kontroller :

Toleranser (tillatte avvik) i tidsserier / tversnittsdata
(fra forrige verdi , gj.sn.-verdi).

LiA/MeS, 4/3-85

Korr. 28/3-85

DEL 4 PUBLIKASJONSDOKUMENTASJONSSYSTEM FOR STATISTISK SENTRALBYRÅ
RAPPORT FRA PROSJEKTGRUPPE PUBL.DOK.

	side
1. Innledning og programvalg	2
2. Kravspesifikasjon	3
2.1 Innledning	3
2.2 Generelle krav og definisjoner	3
3. Omfang	5
4. Dataelementene	8
4.1 Oversikt over dataelementer for beskrivelse av den enkelte publikasjon og/eller publikasjonstittel	8
5. Oppdatering	9
6. Utprodukter og lagring	10
6.1 On-line	10
6.2 Mikrofiche	10
6.3 Teledata	10
6.4 Trykte oversikter	10
6.5 Lagring	11
6. Ressurs- og framdriftsplan	12
7. Registrering av eldre materiale	12
8. Avsluttende kommentarer	13

1. Innledning og programvalg

I henhold til notat om System for datadokumentasjon av 3/11-82 fra Adm. direktør skal gruppen for dokumentasjon av publikasjoner o.l. utarbeide et detaljert forslag til dokumentasjonssystem for data lagret i publikasjoner eller annen lesbar form og beregnet for publisering.

Tre medlemmer ble oppnevnt

/ Dag Album
Liv Argel (prosjektleder)
Karen Thorshaug

Det ble forutsatt at sekretæren for styringsgruppen, Sindre Børke, skulle delta også i denne gruppens arbeide.

I perioden fra oppnevning og fram til oktober 1983 hadde gruppen en rekke møter, og disse notater er lagt fram for styringsgruppen:

- "Utkast til kravspesifikasjon for dokumentasjon av publikasjoner (DOK-PUBL)" (LiA/MoS, 15/3-83)
- "Notat nr. 2 om kravspesifikasjon for dokumentasjon av publikasjoner" (LiA/HaR, 10/5-83)
- "Notat om program for dokumentasjon av publikasjoner" (LiA/HaR, 25/8-83)

Ifølge notat til Styringsgruppen av 7/10-83 SBø/LHå om Publ-dok, Oppstartning/Innmating var utredningen da kommet så langt at det var avklart hva systemet burde inneholde og at et fri-tekstsøkeprogram ville være ønskelig.

I møte den 13. oktober 1983 ga Styringsgruppen sin tilslutning til tilrådingen om å anskaffe fritekstsøkeprogrammet STAIRS-IBM og tekstbehandlingsprogrammet DCF.

Framdriften av Publ-dok-prosjektet ble drøftet i møte med Styringsgruppen 12. januar 1984 på grunnlag av ressurs- og framdriftsplanen fra prosjektgruppen (datert 12/1-84).

Det ble besluttet at prosjektet skulle behandles uavhengig av Data-dok-prosjektet, og at videre arbeid skulle utstå til programvalget var endelig avklart med Systemkontoret.

På Informasjonsavdelingens rulleringsmøte 26/11 1984 ble det besluttet at IR-programmet STAIRS skulle anskaffes fordi det fremdeles er uklart når IBM-versjonen av fritekstsøkeprogrammet SIFT vil foreligge.

Som testgrunnlag foreligger oversikt over Interne notater 1979-1984 og Statistisk årbok 1984.

De følgende avsnitt er utdrag fra de notater som er nevnt foran.

2. Kravspesifikasjon

2.1 Innledning

Ønsker og krav er spesifisert primært ut fra eksterne brukerbehov slik de er kommet til uttrykk i Statistikkbrukerundersøkelsen fra 1980-81 (Rapport 82/13) og IN 1983/3 Markedsføringsnotatet til langtidsprogrammet.

Forutsetningen er at det bygges opp en base som blir søkbar on-line både innen og utenfor Byrået, og at det ut fra denne metadatabasen skal produseres oversikter og lister både i trykt form og på mikrofiche over den statistikk og de analyser og rapporter som er offentlig tilgjengelig, uavhengig av publiseringsform.

Det er videre en forutsetning at den skal gå inn i et overordnet dokumentasjonssystem for Byrået parallelt med, eller som del av den planlagte metadatabase for statistikkproduksjonen. Publ-Dok bør videre koordineres med Bibliotekets eventuelle deltakelse i det nordiske List-stat prosjektet og med det automatiserte biblioteksystemet som velges. Dette gjelder primært innmatingen, slik at dobbeltarbeide unngås.

2.2 Generelle krav og definisjoner

Basen skal i første omgang bygges opp som en bibliografisk base med maksimalt antall søkemuligheter på interne terminaler og on-line for eksterne storbrukere. I tillegg kommer utproduktene trykte lister, oversikter på mikrofiche, og det vil være en fordel om systemene også gir mulighet for overføring til fotosats og eventuelt også til teledata.

Når det gjelder søkemuligheter vil vi i det følgende betegne den bibliografiske beskrivelse av dokumentene (publikasjonene) som dokumentfil. I tillegg kunne det være ønskelig med en parallell søkeordsfil (tesaurus), men dette er et spørsmål som det kan være aktuelt å ta opp når Biblioteket har fått sitt system. Statistiska

Centralbyråen har utarbeidet en tesaurus for statistikk, men den er foreløpig ikke tilgjengelig for andre. Innmating av innholds- eller tabellregistrene fra de enkelte publikasjoner kan erstatte en tesaurus. Brukerne vil ved hjelp av fritekstsøkeprogram kunne oppnå like gode resultater som ved bruk av emneordlister. Forskjellen ligger i at basen kan bli noe mer omfattende fordi tabellregistrene sannsynligvis krever noe mere plass.

De generelle krav kan oppsummeres slik:

1. Basen må være tilgjengelig on-line innen Byrået og senere også for eksterne brukere. Det må være mulighet for utmating i ulike former: lister, mikrofiche, fotosats. Listene skal tilsvare de månedlige og årlige oversikter som nå utarbeides foruten listene i publikasjoner.
2. Innholdet skal gi oversikt over alle Byråets publikasjonstitler med innholdsfortegnelser og opplysninger om regularitet, frekvens, serie, registreringsperiode samt vanlige bibliografiske data. (Basen bør også inneholde dokumentasjon av de av SSB's data som vil bli tilgjengelig for eksterne brukere fra baser, disk, bånd eller mikrofiche).
3. Alle felt må være søkbare, og fritekstsøking må være mulig i alle felt og alle felt bør i prinsippet kunne brukes som sorteringsfelt.
4. Søkekriterier: Alle faste felt eller et utvalg av de faste felt. Mulighet for fritekstsøking i hele basen, i deler av basen og/eller enkelte felt etter logisk og/eller/ikke
Søkeordsfil: Tesaurus kan bygges opp. Høyretrunkering mulig.
5. Mulighet for koordinert innmating til Publ-dok-basen fra Data-dok-basen ansees viktig. Likeledes bør innmating til List-Stat (nordisk dokumentasjonsbase), bibliotekskatalog, database med oversikt over alle offentlige publikasjoner (FAD-prosjekt), Teledata og Teletex kunne samordnes ved at utvalg av data fra Publ-dok overføres til disse via maskinlesbart medium for redigering/supplering i de enkelte baser.
6. Oppdatering bør kunne foretas interaktivt eller i batch hver uke.

3. Omfang

Omfanget av den informasjon som skal lagres om den enkelte publikasjon vil avhenge av faktorer som bl.a. type (serie), regularitet og utgivelsesfrekvens. Faktorer av denne type vil også ha betydning for hvor lenge informasjonen skal lagres, et spørsmål som må tas opp når systemet er i drift.

Våre beregninger av omfanget er basert på disse forutsetninger:

1. Registrering av publikasjontittel/nr. og underserier i serien NOS av monografiseriene (SA/ART/SØS/RAPP) med nødvendige bibliografiske data, supplert en kort innholdsbeskrivelse tilsvarende omtalen i Månedens publikasjoner, og tabellregister og/eller innholdsfortegnelse, samt søkeord.
2. For uke-, måneds- og kvartals- og oversiktspublikasjoner registreres tabellregistrene for de enkelte nr. (hvis uendret tabellregister ett pr. årgang) etter emne slik at den enkelte tabell eller kapittel i innholdsregisteret er registreringsenhet, supplert med henvisning til publikasjonsnummer for periodiske publikasjoner eller oversiktspublikasjoner. Den enkelte publikasjon registreres med de nødvendige bibliografiske opplysninger.

Vi har foretatt en grov beregning av omfanget målt som gjennomsnitt tilgang til basen i løpet et et år, basert på nåværende produksjonsnivå.

Oversikt over antall titler og tabeller i henhold til nåværende produksjonssystem pr. årgang

Publikasjon/ serie	Ca. antall titler pr. år	Frekvens pr. år	Ca. antall tabeller/ titler	Kapitler	Beregnet publ. titler	Antall tabeller/ kapitler
I alt	-	-	-	-	184	6 069
<u>Periodiske</u>						
SU	1	52	10		1	520
SM	1	12	102		1	102
ND	18	12	5		18	60
Bank og kreditt	1	40	10		1	400
KT	1	10	-	11	1	110
MU	1	12	15		1	15
<u>NOS</u>						
SÅ	1	1	550	26	1	576
Øk. utsyn .	1	1	1	26	1	26
Årspann. ..	75	1	29		75	2 175
Andre	15	Sjelden	29		15	435
<u>Monografier</u>						
SA	3	-	63	8	3	220
ART	11	-	-	6	11	70
SØS	5		36	24	5	300
RAPP	30		20	6	30	780
IN	20		8	6	20	280

Med utgangspunkt i denne oversikten blir omfanget:

1. Bibliografiske opplysninger om ca. 185 publikasjoner à anslått feltlengde 400 pos. (ekskl. innholds- oversikt/emneord)	74 000 pos.
2. Innhold periodiske	230 000 "
3. Ca. 3 200 tabell- eller kapittel titler fra serien NOS og ca. 1 600 fra monografiene. Anslått felt- lengde (inkl. henvisninger) 170 pos.	816 000 "
	<hr/>
	1 120 000 pos.
	<hr/>

For engelsk tekst beregnes tillegget som for punkt 1 idet tabell/kapittel titler erstattes med innholdsoversikt, abstract og/eller emneord

80 000 pos.

Innhold i periodiske publikasjoner forutsettes lagret maksimum 2 eller 3 år. "Utgående" års innføring brukes såvidt mulig for ajourføring av ny årgang.

4. Dataelementene4.1 Oversikt over dataelementer for beskrivelse av den enkelte publikasjon og/eller publikasjonstittel

Feltbeskrivelse	Brukes for		Anslått feltlengde (gj.snitt antall pos.)
	<u>dokumentasjon i:</u>		
	Innholds- oversikt SU/NO o.l.	Dok- publ.	
(Dato for oppdatering	(x)	x	6)
[Prosjektnr.	x	x	6]
[Emnekode	x	x	6]
[Emnebetegnelse*	x	x	50]
Seriekode	x	x	5
Seriens fulle navn*	x	x	35
Underserie	x	x	50
Nr./år eller rekkenummer	x	x	10
Hovedtittel norsk*	x	x	100
Undertittel norsk*		x	150
Parallelltittel (engelsk)*		x?	100
Parallell undertittel (engelsk)*		x?	150
Forfatter(e) (monografier)*		x	50
ISSN-nr.		x	9
ISBN-nr.	x	x	13
Registreringsperiode (innholdsår)	x	x	30
Utgitt (mnd./år el. dato/år)	x	x	10
Periodisitet (kode)		x	2
Geografisk kode		x	2
Sidetall		x	4
Pris		x	12
Utgiver*	x	x	25
Forlag/kommisjonær*		x	70
Forrige tittel - norsk*		x	100
Forrige tittel - engelsk*		(x)	100
Første år med samme tittel		(x)	6

TABELL (KAPITTEL I OVERSIKTSPUBLIKASJONER)

Emnekode (max 2)	x	30
Søkeord	x	80
Tabell/kapittelnr. og -tittel norsk*	x	150
Tabell/kapittelnr. og -tittel engelsk*	x	50
Innholdsoversikt til Månedens publikasjoner* ..	x	100

* Felt med varierende lengde.

[] Felt som testes i startfasen, vil senere tas fra Data-dok.

Grovutkast til skjermbilde for publikasjon og tabell er gjengitt på side 27 og 28 i Datadokrapporten foran.

Emne er ment som overordnet enhet.

Serien er her brukt etter Byråets terminologi, dvs. som fellesbetegnelse for seriene NOS, SA osv. Bibliografisk og bl.a. i List-Stat brukes serie eller underserie om tittel på publikasjon som kommer regelmessig.

Supplement. Pressemeldingene AS er supplement til Statistisk ukehefte.

Tittelendringer er det meget viktig å få registrert. Selv om også innholdet delvis endres samtidig med tittelen bør henvisning til tidligere (delvis) tilsvarende publikasjon tas med.

Innholdsoversikt vil for seriepublikasjonenes vedkommende tilsvare ingressen i Månedens publikasjoner.

Katalog brukes for emnebetegnelse, serienavn og ellers hvor det er praktisk.

5. Oppdatering

Oppdateringen bør skje mest mulig kontinuerlig hver uke, gjerne i forbindelse med utgivelsen av Statistisk ukehefte. Det bør være et mål at innholdsregisteret for SU kan tas ut fra basen.

Metadata om statistikkprodukter som skal inn i en(eller flere) publikasjoner, må kunne overføres fra Data-dok og bare suppleres med henvisninger for Publ-dok-basen. Hvis en slik overføring blir mulig vil innmatingsarbeidet reduseres vesentlig. Hvordan Data-dok og Publ-dok kan integreres utredes nærmere i notat om Data-dok.

Generelt må oppdateringen foretas slik at vi til enhver tid har oversikt over hva Byrådet kan tilby brukerne, og slik at månedsoversikter o.l. som skal distribueres i trykt form kan tas ut og sendes regelmessig tidlig i måneden.

6. Utprodukter og lagring

6.1 On-line

Basen må være tilgjengelig on-line for internt bruk for dem som skal gi informasjon til eksterne brukere. Den bør også være tilgjengelig for profesjonelle brukere eksternt - større offentlige etater, kommunedatasentralene, bibliotek, forskningsinstitutter, næringslivsorganisasjoner og bedrifter.

6.2 Mikrofiche

Mikrofiche-utgaver bør lages hver måned for Forlagssentralen for videre distribusjon til bokhandlerne (utalt ønske), for mellomstore bibliotek og andre vitkige brukere som har tilgang til leses, men ikke til basen.

Mikroficheutgaven kan begrenses til bibliografiske opplysninger om publikasjoner og innhold i periodiske publikasjoner.

6.3 Teledata

Hovedopplysninger om publikasjoner, de nyeste, legges over på Teledata. Oppdateres hyppig, minimum hver måned.

6.4 Trykte oversikter

Det forutsettes at grunnlaget helt eller delvis for følgende lister skal produseres fra Publ-dok.

- Månedens publikasjoner. Liste med innholdsoversikt
- Publikasjoner fra SSB. En gang pr. år (eller hyppigere hvis behov). Lister med noen få sentrale opplysninger om den enkelte publikasjon
- Veiviser i norsk statistikk. Data fra basen delvis grunnlag. Utgis hvert 3. år
- Oversikter over innhold i periodiske publikasjoner og årbok
- Månedsoversikter over det som er i produksjon (utarbeides nå av Trykningskontoret)
- Diverse andre oversikter bl.a. til Melding om virksomheten
- Publikasjonsoversiktene bak i publikasjonene
- Gjerne også statistikk (antall tabeller, trykksider o.l.)

- Publikasjonsliste for forsendelsesregister (delvis grunnlag)
- Ajourføring av Fortegnelse NOS hvert 5. år

6.5 Lagring

Som nevnt bør det være mulighet for søking også etter eldre titler på seriepublikasjoner. Etter hvert som ny utgave av Fortegnelse NOS etc. lages (hvert 5. år), kan fullstendige opplysninger om den enkelte publikasjon lagres på mikrofiche og i trykt form, mens bare titler/årgang ligger igjen i basen. Tilsvarende vil innholdsfortegnelser for periodiske publikasjoner bli uaktuelle etter et par år. (Data-dok forutsettes å ta vare på opplysningene om statistikkproduktene.)

6. Ressurs- og framdriftsplan

Denne planen er satt opp med sitke på at Publ-dok-basen skal kunne brukes internt on-line ved årsskiftet 1985/86.

Periode 1985	Arbeidsoppgave	Anslått tidsbruk timer	Utførende (ktr.initialer)
April/juni	Innmating av bibliografiske opplysninger. 1984-årgang.		
	Forarbeid (koding)	25	Info.
	Innmating	40	Man.db.
	Tilrettelegging program	20	System
April/oktober	Testing program og -system	50	System
	Utkjøring og test	20	Drift
	Oppløring/testing	40	Info.
Sept./des.	Innmating av 1985-årgang, publikasjoner med tittel og tabelltekst norsk/engelsk		
	Forarbeid/koding	200	Info.
	Innmating	90	Man.db.
	System	40	System
	Utkjøring	30	Drift

Innføringen forutsettes etter hvert tastet inn hver uke, slik at de utprodukter som er skissert i foregående avsnitt etter hvert skal tas ut. Vi foreslår denne rekkefølgen for 1986

Innhold SU kvartal

Innhold ND halvår

Månedens publikasjoner siste kvartal

7. Registrering av eldre materiale

For at Publ-dok-basen skal være nyttig for dem som søker tidsserier og tilbakegående tall, vil det være nødvendig å registrere tidligere utkommet materiale.

Fortegnelsen over Norges offisielle statistikk dekker perioden fram til 1975. For årene 1976 t.o.m. 1980 foreslås at publikasjonene registreres med rene bibliografiske opplysninger uten innholdsoversikt eller stikkord, og etter den nye emneklassifiseringen - eventuelt med henvisning til den som nå brukes.

Denne innmatingen bør skje som batch bestående av en halv årgang hver, og bør påbegynnes senest i 1986.

8. Avsluttende kommentarer

I det foregående er det bare tatt med resyme av kravspesifikasjoner og programbeskrivelser. For mer utførlige opplysninger vises til de notater som er nevnt i avsnitt 1 foran, og til notater fra Forprosejktgruppa om Datadokumentasjonsprogram for Byrådet av 25. februar 1985.

N O T A T

Fra Adm. direktør

SYSTEM FOR DATADOKUMENTASJON

1. Bakgrunn

Byråets perspektivprogram for 1980-årene sier at det skal utarbeides og iverksettes en langsiktig plan for å lette gjenbruken av data. I en første fase som regnes avsluttet i 1983, skal det fastsettes generelle retningslinjer for dokumentasjon av data.

Som et ledd i gjennomføringen av dette program er det vedtatt å sette i gang et forprosjekt ved Systemkontoret (jfr. prosjektskriv av 3/8-82). Forprosjektet skal gi en kartlegging av de krav som stilles til datadokumentasjonen i Byrådet og en vurdering av ulike tekniske løsninger. Det er videre vedtatt at Informasjonsavdelingen skal arbeide med et prosjekt om dokumentasjon av publikasjoner, innholdsfortegnelser m.v. (jfr. prosjektskriv av 7/9-82).

Disse to prosjektene har en ulik bakgrunn, noe som allerede er et uttrykk for at dokumentasjonsbehovet i Byrådet er mangesidig. De forundersøkelser som skal utføres vil med sikkerhet bekrefte av dokumentasjonsbehovene er forskjellige innen ulike brukergrupper, og de vil vise at behovene dessuten avhenger av forhold som dataenes agrergeringsnivå og det medium dataene ligger lagret på.

Et system for datadokumentasjon i Byrådet bør av forannevnte grunner bygges opp av en rekke elementer. Utformingen i første hand av enkelt-elementene antas mest hensiktsmessig å kunne skje i egne prosjektgrupper, mens samordningen til ett system bør foretas av en styringsgruppe.

2. Styringsgruppe

Mandat

Ut fra mål og retningslinjer gitt i perspektivprogram, EDB-plan og andre plandokumenter for Byråets virksomhet i 1980-årene, skal Styringsgruppen trekke opp hovedretningslinjer for, styre framdriften av og samordne arbeidet med utformingen av retningslinjer for dokumentasjonen av data. Styringsgruppen skal ellers

- (i) påse at datadokumentasjonssystemet tilpasses best mulig til de dokumentasjonsbehov ulike brukergrupper har,

- (ii) arbeide for at dokumentasjonen kan framkomme som et ledd i arbeidet med planleggingen og gjennomføringen av et prosjekt,
- (iii) legge forholdene til rette for at dokumentasjonen kan bli lett tilgjengelig for dem som trenger den,
- (iv) sørge for at berørte funksjonærer og deres organisasjoner gis slik informasjon og slik medinnflytelse som særavtale om medbestemmelse bestemmer,
- (v) sørge for utarbeiding av kostnads- og tidsplaner for framdriften av arbeidet med utformingen av dokumentasjonssystemet og foreslå nedsatt/oppløst de prosjekt- eller arbeidsgrupper som framdriften av arbeidet tilsier,
- (vi) rapportere til administrerende direktør og
- (vii) utarbeide en plan for igangsetting og videreføring av dokumentasjonsarbeidet.

Medlemmer

Styringsgruppen sammensettes av representanter for Informasjonsavdelingen, Forskningsavdelingen, Fagavdelingen, Produksjonsavdelingen og funksjonærorganisasjonene. Disse oppnevnes til medlemmer av styringsgruppen:

Erik Botheim
 Sindre Børke (sekretær)
 Erik Hernæs
 Sverre Hove
 Idar Møglestue (formann)
 Halvard Skiri
 Arne Faye

Tilføyelse: Som følge av Botheims død og Hernæs' permisjon, ble Svein Brenna og Ketil Sørli oppnevnt til nye medlemmer av styringsgruppen i desember 1983. Børke ble i 1984 fritatt for medlemskap i og sekretær for styringsgruppen av arbeidsmessige grunner.

3. Prosjektgrupper

Det nedsettes to prosjektgrupper; en forprosjektgruppe for å spesifisere de krav som stilles til et dokumentasjonssystem for data på maksinlesbare media og vurdere ulike tekniske løsninger for et slikt delsystem, og en annen gruppe for å utrede behovet for og opplegget av dokumentasjonen av data lagret i publikasjoner og på annet papirmedium.

Forprosjektgruppens mandat er som beskrevet i prosjektskriv av 3/8-82. Til medlemmer oppnevnes:

Halvard Skiri
 Tormod Solvin (prosjektleder)
 Halvor Strømme

Sekretæren for styringsgruppen pålegges å delta i gruppens arbeid.

Gruppen for dokumentasjon av publikasjoner o.l. utarbeider et detaljert forslag til dokumentasjonssystem for data lagret i publikasjoner eller annen manuell lesbar form. Arbeidet utføres innenfor de tids- og ressursrammer som er trukket opp i prosjektskriv av 7/9-82 og med utgangspunkt i den drøfting som er gjort i notat av 31/8-82 om markedsføring. Til medlemmer oppnevnes disse:

Dag Album
Liv Argel (prosjektleder)
Karen Thorshaug

Sekretæren for styringsgruppen forutsettes å skulle delta også i denne prosjektgruppens arbeid.