

O-86245

FORPROSJEKTRAPPORT

Vurdering av retningslinjer for etablering av emneord-registre for temaområdet miljøinformasjon

NIVA – RAPPORT

Norsk institutt for vannforskning



NIVA

Hovedkontor
Postboks 33, Blindern
0313 Oslo 3
Telefon (02) 23 52 80
Telefax (02) 39 41 29

Sørlandsavdelingen
Grooseveien 36
4890 Grimstad
Telefon (041) 43 033
Telefax (041) 42 709

Østlandsavdelingen
Rute 866
2312 Ottestad
Telefon (065) 76 752

Vestlandsavdelingen
Breiviken 5
5035 Bergen - Sandviken
Telefon (05) 95 17 00
Telefax (05) 25 78 90

Prosjektnr.:	O-86245
Undernummer:	
Løpenummer:	2065
Begrenset distribusjon:	

Rapportens tittel:	Dato:
Vurdering av retningslinjer for etablering av emneord-registre for tema-område miljø-informasjon	1987-12-18
	Prosjektnummer:
	O-86245
Forfatter (e):	Faggruppe:
Hans Munthe-Kaas	
	Geografisk område:
	Antall sider (inkl. bilag):
	25

Oppdragsgiver:	Oppdragsg. ref. (evt. NTFN-nr.):
Statens Kartverk	

Ekstrakt: I regi av Statens Kartverk planlegges det utvikling av et "nasjonalt referanseregister for miljøinformasjon". I tilknytning til dette er det behov for en integrert samling av emneord-registre på nasjonalt nivå som dekker alle tema innen tema-området "det ytre miljø". Slike lister fins (med få unntak) ikke idag. Det å lage slike integrerte lister er en stor og komplisert oppgave. Retningslinjer for hvordan slike lister bør bygges opp, var ikke kjent. Forprosjektets oppgave var å finne frem til eventuelle eksisterende retningslinjer (uansett tema-område) i Norge og andre land - og vurdere/supplere disse med henblikk på anvendelse for ovennevnte emneord-samling.

Det viste seg at vel gjennomarbeidede, generelle retningslinjer fantes i form av ISO-standarder fra 1985/86. Disse var lite kjent blant miljøinstitusjonene i Norge og Norden, men brukt ganske mye innenfor EF-landene og USA. Flere eksempler på emneord-liste for tema-område miljø basert på ISO-standardene er innhentet fra disse landene.

4 emneord, norske:	Emneordregister	4 emneord, engelske:
1.	Tesaurus	1.
2.	Miljøinformasjon	2.
3.	Retningslinjer	3.
4.	Etablering	4.

Prosjektleder:

For administrasjonen:

Hans Munthe-Kaas

ISBN - 82-577-1326-0

Munthe-Kaas

FORORD

Den økonomiske ramme for dette forprosjektet var på kr 100.000, hvorav Statens Kartverk og NIVA bidro med en halvpart hver. En overskridelse på ca kr 20.000 er dekket av NIVA.

Underveis i forprosjektet ble det avtalt at også prosjektlederens deltagelse i seminarer, møter etc som behandler det bredere og overordnede tema "nasjonale tiltak for samordning av informasjonshåndteringen på miljø-siden", skulle inngå i prosjektet. Omkostningene i forbindelse med denne utvidelsen av prosjektet (arbeidstid og reiseutgifter, tilsammen ca kr 20.000) er dekket av Statens Kartverk.

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

Seksjon	Side
1. BAKGRUNN , FORMÅL OG RESULTATER	3
2. AVGRENSNING AV OPPGAVEN	5
3. HOVED-PRINSIPPER FOR KONSTRUKSJON AV EMNEORD-LISTER	7
3.1 Nødvendige definisjoner	7
3.2 Sentrale og perifere emneord	9
3.3 Kategorier av emneord-lister	9
4. TILLEGGS-PRINSIPPER OG -RETNINGSLINJER FOR NORSKE FORHOLD	10
5. OVERORDNEDE KRAV TIL DEN FREMTIDIGE SAMLING AV EMNEORD-LISTER	12
6. ERFARINGER FRA LETE-RUNDEN	13
6.1 Generelt	13
6.2 Internasjonale standarder	14
6.3 Internasjonal oversikt over eksisterende tesauri	16
6.4 Eksisterende emneord-opplegg for litteratur-referanser	17
6.5 Emneord-opplegg for dataserie-referanser	19
7. SKISSE TIL PROGRAM FOR DET EFTERFØLGENDE HOVEDPROSJEKT	21
LITTERATUR SOM DET ER HENVIST TIL	

Fig 1 Det nasjonale referansesenter for miljø-informasjon

Tabell 1 Fakta om endel utvalgte emneord-opplegg for referanser
av kategori "litteratur"

1. BAKGRUNN , FORMÅL OG RESULTATER

BAKGRUNNEN for dette forprosjekt var behovet for å få etablert en samling av emneord-lister for miljø-området på nasjonalt nivå. Lister av denne type vil være nødvendige elementer i "Det nasjonale referanse-register for miljøinformasjon" som er planlagt utviklet i regi av Statens Kartverk.

FORMÅLET med forprosjektet har vært å fremskaffe retningslinjer for etablering og utnyttelse av en slik samling. Fremskaffelsen skulle bestå i først å finne frem til eksisterende retningslinjer og emneordliste-opplegg (gjennom litteraturen og ved henvendelse til institusjoner som kunne forventes hadde slike). Dernest skulle de funn som ble gjort, vurderes - og om nødvendig bearbeides og viderutvikles - til en skisse til sett av retningslinjer tilpasset norske behov og forhold.

I formålet inngikk også en skisse til program for et etterfølgende hovedprosjekt.

RESULTATETENE fra forprosjektet kan presenteres i fem punkter:

- 1 Leting etter eksterne retningslinjer. Det viste seg at vel gjennomarbeidede, generelle retningslinjer fantes i form av ISO-standarder fra 1985/86. (Sub-kap 6.2). Disse var lite kjent - og derfor lite brukt - blant institusjonene i Norden. Derimot så de ut til å være ganske mye brukt i EF-landene, USA samt av flere FN-organer.
- 2 Vurdering/supplering av disse retningslinjene med henblikk på anvendelse for emneordlister til "det nasjonale referanse-register for miljøinformasjon". Retningslinjene i ISO-standardene synes stort sett å være dekkende for de emneord-listekategoriene som skal brukes for litteratur-referanse-registre. (Det er det formålet de er laget for). Forprosjektet peker på et mulig behov for tilleggs-regler for referanse-registre for dataserier. (Sub-kap 3.3).

Forprosjektet antyder også behov for tilleggs-retningslinjer med sikte på etablering av en hierarkisk overbygning som skal omfatte alle emneord-registrene på miljø-siden. Dette behovet har sammenheng med at man i Norge har tenkt å lage en overordnet hierarkisk struktur for selve referanse-systemet på miljødata-

siden. (Kap 4).

- 3 Innhenting av informasjon om eksisterende emneord-lister på miljø-siden. I tillegg til retningslinjer ble det også innhentet informasjon om eksisterende ferdige emneord-lister. Mange emneord-lister av kategori "litteratur" ble funnet i EF-landene, USA og FN, og informasjon om et representativt utvalg av slike er tatt med i rapporten. (Sub-kap 6.3). Av kategorien "dataserie" ble ingen funnet.
- 4 Krav som bør stilles til den fremtidige samling av emneordliste. Et utkast til slike krav er satt opp. (Kap 5).
- 5 Skisse til program for det etterfølgende hovedprosjekt. En enkel skisse er satt opp som dekker et første trin i hovedprosjektet. (Kap 7). Skisse til de følgende trin kan ikke settes opp før det første er gjennomført.

2. AVGRENSNING AV OPPGAVEN

Med "emneord-lister" menes i denne forprosjekt-rapport EDB-baserte oppslags-registre som omfatter alle "tematiske" begrep som kan være aktuelle søke-begrep ved utsøking av informasjons-referanser fra et referanse-register.

Det foreligger i forprosjekt-beskrivelsen ingen avgrensning mht hvilke "typer" av tematiske begrep som skal tas med under "emneord"-betegnelsen- og heller ikke noen avgrensning av begrepene "miljø" og "informasjon". Selv om det i denne fase av arbeidet ikke er behov for noen presise avgrensninger, ble det ansett nødvendig å ha noen tentative grenser. I regi av prosjektet ble derfor et forslag til foreløpige avgrensninger satt opp:

* Med tematiske begrep er det her i første omgang tenkt på:

- "Artslister" for biologi, geologi, kjemi etc
- Andre begrep fra biologi, geologi, kjemi etc
- Begrep fra relevant teknologi
- Begrep fra arkeologi, etnologi etc
- Begrep fra relevante forvaltnings-aktiviteter

- Prøvetagnings- og prøvebevarings-metoder aktuelle for disse begreps-områder
- Måle-parametre aktuelle for disse begreps-områder
- Måle-metoder aktuelle for disse parametre

* Med informasjon menes her informasjon som er lagret. Informasjon kan være av to hovedtyper:

- "Rene dataserier" (dvs fakiske opplysninger uten tolkninger)
- "Litteratur" (dvs data sammen med tolkninger av disse).

All lagret informasjon vil være knyttet til en eller annen informasjonsbærer. Alle typer av informasjons-bærere som det kan være aktuelt å referere til, skal være med. Eksempler på slike kan være:

- Bøker, tidsskrifter, rapporter etc
- Fotos, film etc
- Tema-kart

- Lyd-bærere (grammofon-plater, kassetter etc)
- EDB-lagrings-medier
- Fysiske prøver fra natur og ikke-naturlige objekter

Som fellesnavn for alle slike informasjons-bærere er det - ihh til ISO-standard (kfr kap ?????) i denne rapport brukt betegnelsen dokument.

- * Med miljø menes her bare det "ytre" miljø - ikke det "indre" miljø.
- * Med miljø-informasjon menes informasjon både om naturen selv og om ikke-naturlige faktorer som påvirker prosesser og tilstander i naturen.

I tillegg til oppslags-registre av "emneord"-typen trengs det også noen andre - hvorav de fleste vil være av "stedfestende" type. Eksempler på slike kan være:

- * Areal-inndeling basert på administrative grenser
- * Areal-inndeling basert på hydrologiske grenser
- * Areal-inndeling basert natur-geografiske forhold
- * Lokalisering av trafikk-årer, ledningsnett, bygninger mv
- * Lokalisering av målepunkter og måleområder i naturen

Disse andre typer av oppslags-registre er ikke behandlet i denne rapport.

3. HOVED-PRINSIPPER FOR KONSTRUKSJON AV EMNEORD-LISTER

3.1 Nødvendige definisjoner

Begrepene "emneord" og "emneord-liste" synes å være et vidt begrep som er brukt for alt fra enkle alfabetiske ordlister til store og kompliserte liste-systemer. Det fins internasjonale ISO-standarder (nærmere omtalt i sub-kapittel 6.2 nedenfor) som definerer og forklarer både de to og beslektede begrep. Disse standardene synes imidlertid å være lite kjent - ihvertfall her i Norge.

For å kunne forstå de senere avsnitt i denne rapport, må man vite noe mer om hva disse to begrepene betyr. En kort orientering (basert på ISO-standardene) er gitt i det følgende:

- 1 Emneord er utvalgte begrep (ord) fra det daglige sprog. Hvert emneord representerer et "tema" og kan brukes ved indeksering (dvs karakterisering) og utsøking av dokumenter i et referanse-arkiv. (Begrepene "tema" og "dokument" slik som brukt her er forklart i kapittel 2).
- 2 Emneord-liste er en liste som inneholder et sett av emneord. Settet må dekke alle aktuelle tema innen et avgrenset "tema-område".
- 3 Alle emneord-lister vil normalt foreligge alfabetisk sortert. Noen vil i tillegg også foreligge i andre former (se nedenfor).
- 4 Emneord-lister innenfor et tema-område kan ha ulike nivåer m h t grad av kompleksitet. Nivåene er her nevnt i rekkefølge etter stigende kompleksitet:
 - a Alle begrep som mist én gang er brukt ved indeksering av dokumentene innenfor et tema-område. Gjerne alfabetisk sortert.
 - b Bare begrep som er godkjent av den som har ansvaret for listen. Slike godkjente begrep kalles deskriptorer. Gjerne alfabetisk sortert.
 - c Alle godkjente begrep (deskriptorer) - og i tillegg alle ikke

godkjente (kalt ikke-deskriptorer) som kan tenkes prøvet ved indeksering eller søking. Alfabetisk sortert sammen. Bak ikke-deskriptorene står det en henvisning (f eks USE) til hvilken deskriptor som må brukes isteden.

- d Som pkt c, men i tillegg står det bak hver deskriptor en henvisning (f eks USED FOR) til hvilke(n) ikke-deskriptor(er) som den har henvisninger fra.
 - e I tillegg til pkt c eller d er det også bygget inn hierarkisk struktur i listen. Det vil si at det bak hver deskriptor står hensinings-koder som forteller hvilke andre deskriptorer denne eventuelt har over og/eller under seg. Kodene kan f eks være BT (= BROADER TERM = overordnet deskriptor) og NT (= NARROWER TERM = underordnet deskriptor). Slike emneordlister kan - i tillegg til den vanlige alfabetiske form - også presenteres i form av tre-strukturer og eventuelt også nettverk-strukturer.
 - f I tillegg til pkt e er det også bygget inn i listen informasjon om andre (dvs ikke-hierarkiske) relasjoner som måtte finnes mellom enkelte deskriptorer. Slike relasjoner er vist ved at det bak hver av slike deskriptorer står en henvisnings-kode (RT = RELATED TERM).
 - x I tillegg til pkt f kan også flere andre konvensjoner være bygget inn i listene som bidrar til ytterligere økning av presiserings-mulighetene. Eksempler på dette kan være bruk av kommentarer (kode SN = SCOPE NOTE) og bruk av sammensatte deskriptorer (deskriptorer som inneholder mer enn ett ord).
- 5 Emneord-lister kan være ett-sproglige eller fler-sproglige. De fler-sproglige har spesielle ordlister som gir oversettelser for de enkelte deskriptorer og ikke-deskriptorer.

Det fins mer presise betegnelser som kan brukes istedenfor betegnelsen "emneord-liste". Det gjelder for alle typene under pkt 4 ovenfor unntatt type "a". Alle typene f o m "b" kan kalles indekserings-sprog - og alle typene f o m "e" kan kalles tesaurus (tesauri i flertall).

3.2 Sentrale og perifere emneord

Praktisk erfaring har vist (kfr dokument fra EF - nærmere omtalt i sub-kapittel 6.3 nedenfor):

- at en emneordliste bør være hierarkisk oppbygget i samsvar med tema-områdetets indre struktur.
- at en emneord-liste ikke bør bli for stor (dvs inneholde for mange emneord). Dette innebærer at en emneord-liste som skal dekke et bredt tema-område, ikke kan være særlig detaljert - og at en emneord-liste som skal være detaljert, bare kan dekke et snevert tema-område.
- at en emneordliste - for å fungere godt - bør dekke ikke bare dens eget definerte tema-område, men i tillegg ha med de overordnede emneord fra alle tilgrensende tema-områder.

3.3 Kategorier av emneord-lister

I sub-kapittel 3.1 står det, under pkt 1, at "emneord" er utvalgte begrep fra det daglige sprog - og under pkt 2 at "emneordlister" er sett av emneord som dekker alle aktuelle tema under avgrensede tema-områder.

I tillegg til at emneord-listene må inndeles etter tema-område, kan det i noen tilfelle være praktisk å inndele dem noe videre. Dette behovet for videre-inndeling avhenger av typen "informasjon" (kfr kap 2 ovenfor).:

- For informasjons-typen "litteratur" vil det antagelig greie seg med den tidligere omtalte tematiske inndeling.
- For informasjons-typen "dataserier" vil det ofte være behov for mer konsise søkeprofiler. Her vil det antagelig være praktisk å videre-dele emneord-listen innenfor det enkelte tema-område basert på "type" begrep. Hvordan reglene for en slik videreinndeling skal utformes, er ikke nærmere vurdert, men et første forsøk gjort for tema-område "vann" kan tjene som et illustrerende eksempel:

o Medium (det "element" som målingene er utført i - f eks grunnvann)

- o Biologiske taxa
- o Fysiske, kjemiske og biologiske parametre
- o Fysiske, kjemiske og biologiske analysemetoder
- o Fysiske, kjemiske og biologiske prøvetagnings-metoder
- o Diverse

I litteraturen er det stort sett bare "litteratur"-typens behov for emneord-lister som er funnet omtalt. I den forbindelse er behovet for videre-inndeling ikke funnet nevnt.

Den eneste litteratur-kilde som har behandlet bruk av emneordlister spesielt for dataserie-referanser, er Nordisk Ministerråd. I et par av publikasjonene herfra er behovet for en videre-inndeling av de tematiske emneord-listene antydnet. (Kfr sub-kap 6.5 nedenfor).

4. TILLEGGS-PRINSIPPER OG -RETNINGSLINJER FOR NORSKE FORHOLD

I følge en eksisterende planskisse for det nasjonale miljødatasamarbeid i Norge, vil det bli opprettet et "nasjonalt referansesenter for miljø-informasjon".

Det antas at dette senteret vil bli strukturert med et overordnet tverrfaglig hovedsenter og flere landsdekkende fag-sektorielle del-sentea underlagt dette (kfr fig 1).

Hvilke sektorielle del-sentre som bør bli etablert, er ennå ikke ferdig diskutert. Noen sektorielle landsdekkende referanse-opplegg fins imidlertid allerede under etablering. Det er rimelig å anta at disse vil gå inn som del-sentre i det kommende felles-opplegget. (Som eksempler på slike kan nevnes Vassdragsregisteret ved NVE (tema-område ferskvann) og referanse-tjenesten for geologisk informasjon ved NGU (tema-område geologi)).

Den nasjonale referanse-tjenesten vil få ansvaret for to drifts-aktiviteter:

- Den skal motta referanser om miljødata-dokumenter. Alle produsenter av miljø-informasjon i Norge skal være forpliktet til å levere referanser for alle sine rapporter, dataserier etc til Referansetjenesten (såfremt produksjonen av dem er skjedd med offentlig støtte eller som følge av offentlig påbud). Den ordning

antas å bli valgt at slike leveranser skal sendes direkte til det faglig sett nærmeste del-senter (altså ikke via hoved-senteret).

- Den skal fungere som et nasjonalt informasjons-senter for miljø-dokumenter. Alle som ønsker det, skal kunne spørre referanse-tjenesten ("on-line" eller på annen måte) om det fins miljø-informasjon "om hva i hva fra hvor og når" - og, hvis slik informasjon fins, om hvem som har den. For slike forespørsler antas at det som hovedregel blir valgt å dirigere dem direkte til hovedsenteret. Om nødvendig, vil forespørslene bli formidlet videre derfra til ett eller flere av delsentrene .

For å få dette til, må referanse-tjenesten få utviklet et integrert EDB-system som kan ta hånd om både hovedsenterets eget og del-sentrenes egne database-behov - og dessuten ta hånd om behovet for rasjonell kommunikasjon mellom disse.

Spørsmålet om hvilke prinsipper som skal brukes for grense-trekking mellom de enkelte del-sentre mht tematisk arbeidsområde er man heller ikke ferdig med å diskutere. For hovedløsning antas det at man vil måtte velge mellom to alternativer:

- En ren tema-basert deling. (Dvs at tverrfaglige institusjoner må fordele sine referanser mellom flere del-sentre).
- En løsning basert på tema-basert gruppering av de institusjoner som skal levere referanser til den nasjonale referansetjenesten. (Dvs at den enkelte institusjon må sende alle sine referanser bare til ett del-senter (selv om disse berører flere tema-områder) - og at det enkelte del-senter må informere de andre del-sentrene om hvilke referanser det har fått innenfor deres respektive tematiske ansvars-områder).

Et slikt hierarkisk opplegg av referanse-systemet for miljø-informasjon som her er foreslått, stiller krav om et tilsvarende hierarki på liste-nivå i den emneord-liste-samling som skal benyttes. Dette kravet kan oppfylles ut i fra de prinsipper som er omtalt i sub-kapitel 3.2 ovenfor. Retningslinjer for hierarki på et så høyt nivå fins ikke i ISO-standardene - og må derfor utarbeides.

5. OVERORDNEDE KRAV TIL DEN FREMTIDIGE SAMLING AV EMNEORD-LISTER

Ut i fra innhentede informasjon, råd og erfaringer fra litteraturgjennomgang og institusjons-besøk - og ut i fra gjennomgåelsen av ISO-standardene - synes det riktig og nødvendig å foreslå følgende krav til den samling av emneordlister som skal utvikles:

- A Samlingen må dekke alle de tema-områder som har tilknytning til det ytre miljø. (Kfr kap 2).
- B Den må kunne brukes for alle aktuelle typer av "dokumenter". (Kfr kap 2).
- C Den må være tilpasset den fler-nivå-struktur som det kommende nasjonale referanse-systemet skal få. (Kfr fig 1 (kapittel 4 ovenfor)).
- D Den må muliggjøre rasjonelt samspill på miljøsidens side med tilsvarende opplegg både internasjonalt og i andre land. Dette må gjelde med hensyn både til samlingen av emneord i seg selv og til det nasjonale referanse-system som denne skal være en del av.
- E Den må være forberedt mht samspill med kommende nasjonale emneord-liste-samlinger og referanse-systemer innenfor tilgrensende tema-områder (som f eks helse og samfunnsvitenskap).
- F Den må være rasjonell både med tanke på dens senere vedlikehold og viderutvikling og på dens funksjon som del av det nasjonale referanse-systemet.
- G Den må kunne bygges ut gradvis - og må allerede f o m dens første versjon kunne være funksjons-dyktig innenfor det nasjonale referanse-systemet.
- H Den skal ha status som nasjonalt "standard emneord-samling" på miljø-siden. Det innebærer at den må holdes ajour - og at den må utvikles videre i takt med endrede behov.
- I Institusjonene innenfor de enkelte tema-områder på miljøsidens side må kunne få kopier fra emneord-samlingen avgrenset til de tema som de selv har bruk for.

6. ERFARINGER FRA LETE-RUNDEN

6.1 Generelt

Målet for leterunden var å finne frem til forskjellige typer av emneord-opplegg som fins eller er under utvikling - og til forskjellige sett av retningslinjer for utvikling av slike.

Fordi det a priori så ut til å finnes bare et begrenset materiale av denne type som var laget spesielt for tema-området miljø, var det aktuelt å lete langs to linjer:

- Etter opplegg og retningslinjer uansett tema-området - for å være sikker på å få med materiale av informasjons-teknisk høy kvalitet.
- Etter opplegg og retningslinjer fra tema-området miljø - for å få med det interessante materiale som måtte finnes der.

Det presiseres at det ikke var noe mål for prosjektet å få kartlagt alt som fins av opplegg og retningslinjer for emneord-lister på miljø-siden. Målet var kun å fremskaffe nok materiale til å kunne velge den strategi som passer best for våre forhold - samt å finne frem til vel gjennom-arbeidede ferdige emneord-lister som kan brukes som grunnlag når våre egne lister skal bygges opp.

Forut for leterunden hadde prosjektet lite informasjon om hvor det burde letes og hva man kunne vente å finne. Den første fasen av leterunden ble derfor noe famlende, og endel tid gikk med til å studere emneord-opplegg som senere ble bedømt som lite interessante. Etterhvert kom det imidlertid frem spor som trinnvis førte frem til en rimelig god oversikt over situasjonen.

Det ble i løpet av lete-runden klart at det ute i verden fins noen ganske gode emneord-opplegg og noen sett av vel gjennom-arbeidede retningslinjer. Samtidig ble det klart at det fins mange emneord-opplegg som ikke er gode nok - sett i relasjon til de krav som er skissert i kapitel 4 foran.

Det vil her i rapporten ikke bli gitt noen beskrivelse av leterundens forløp. Heller ikke vil det bli gitt noen fullstendig oversikt over de

opplegg og sett av retningslinjer som ble studert i løpet av prosjektet. Kun de opplegg og sett av retningslinjer som har hatt betydning for rapportens konklusjoner, vil bli omtalt (avsnittene 6.2 - 6.5). nedenfor).. De viktigste vil bli nevnt først.

6.2 Internasjonale standarder

Det fins tre internasjonale ISO-standarder som angår utvikling av emneord-lister:

ISO 5963 Documentation - Methods for examining documents, determining their subjects and selecting indexing terms.

ISO 2788 Documentation - Guidelines for the establishment and development of monolingual thesauri. (Second edition - 1986-11-15).

ISO 5964 Documentation - Guidelines for the establishment and development of multilingual thesauri.

Disse standarder inneholder ikke, slik standarder ellers ofte gjør, bare et sett med definisjoner og regler. Ut i fra struktur og form er de mer å betrakte som lærebøker, idet de utover regler og definisjoner også omhandler formål samt alternative valgmuligheter og deres konsekvenser.

Som det fremgår av avsnitt 6.3 nedenfor, er disse standarder i ikke liten utstrekning brukt og fulgt rundt om i verden av større nasjonale og internasjonale emneord-opplegg utenom Norden. (Hvorvidt de også er brukt av mindre, lokale opplegg, fremgår ikke av prosjektets kilde). Det er derfor forbausende at ISO-standardene ikke er benyttet av noen av de opplegg som prosjektet har fått kontakt med i Norge eller Norden forøvrig. Og like forbausende at de fleste av våre kontakt-personer ikke engang visste at slike standarder eksisterer.

De retningslinjene som fins i de tre standardene, synes til sammen å dekke alle de problemstillinger som man kan vente å møte i forbindelse med utvikling av et emneord-opplegg. De synes videre å være logiske og meget godt gjennomarbeidet. Og presentasjonen av dem er klar og velformet. Ingen av de andre sett av retningslinjer som prosjektet har gjennomgått, har denne bredde og kvalitet.

Hvorvidt det - i tillegg til disse dokumentasjons-standardene fra ISO

- også fins andre tilsvarende internasjonale sett - er ikke undersøkt spesielt. Siden slike andre sett ikke er omtalt i den gjennomgåtte litteratur, er det imidlertid lite sannsynlig at det i vår del av verden fins noe alternativt sett som har noen vesentlig betydning. (I det oppslagsverk som omtales i neste avsnitt (Thesaurus guide) er riktignok noen tyske DIN-standarder nevnt - men da bare som supplement til ISO-standardene.

Prosjektet anbefaler at retningslinjene i ISO-standardene legges til grunn for det videre arbeid. Det er seks grunner til dette:

- 1 Det er den eneste kilde prosjektet har funnet som dekker hele problem-komplekset.
- 2 De er de beste prosjektet har funnet mht logisk struktur, grad av gjennom-arbeidelse og klar presentasjon.
- 3 Det å bruke standard-regler gir muligheter til samspill med andre informasjons-systemer som er basert på de samme standarder. Dette standard-settet synes å ha en ganske stor utbredelse.
- 4 Det at standard-regler er brukt, gir muligheter til å kopiere ferdige emneord-lister fra andre opplegg som bruker de samme reglene. Dette er et viktig argument, fordi slike lister faktisk fins og fordi egenutvikling av emneord-lister er en meget tung og kostbar prosess.
- 5 Det å bruke standard-regler innebærer også at man selv slipper å lage regler. Det å lage slike regler som er gode nok er en meget vanskelig og tid-krevende prosess.
- 6 Hvis man velger å kopiere et komplett og gjennomarbeidet sett av emneord fra en annen institusjon, vil man til en viss grad kunne bruke en forenklet utgave av dette til å begynne med - og så øke kompleksiteten etterhvert i takt med behov og kapasitet.

På litt lengre sikt er det imidlertid ikke tvil om at det nasjonale referanse-system for miljøinformasjon i Norge vil trenge et emneordliste-opplegg på nivå x i listen i sub-kapitel 3.1

6.3 Internasjonal oversikt over eksisterende tesauri

Innenfor EF-landene foregår det et organisert samarbeid mht registrering og utveksling av bl a vitenskapelig og teknisk informasjon. Ett av de delmål som man har satt seg innenfor dette samarbeid, er å få utviklet et sett av tesauri som samlet dekker alle aktuelle temaområder. Som et ledd i arbeidet mot dette målet ble det i 1985 lagt frem en oversikt over eksisterende tesauri i den vestlige verden (uansett fagområde) som bruker minst ett av de offisielle EF-språkene. Oversiktens tittel er:

Thesaurus Guide. Analytical directory of selected vocabularies for information retrieval. (Komplett referanse fins i litteratur-oversikten - ref 1).

Denne oversikten omfatter primært 462 "levende" tesauri (dvs tesauri som er i bruk). I tillegg nevner den 192 andre emneordlister som av forskjellige grunner ikke er med i primær-listen. (Som eksempel på slike grunner nevnes: Ikke tilgjengelige fra "østblokk"-land, ikke ferdig utviklede, ikke avanserte nok og ikke ajourholdte).

Primær-oversikten er inndelt tematisk i 10 hovedklasser og tilsammen 61 underklasser. Hovedklassene er disse (tallene til høyres viser antall underklasser under hver hovedklasse):

A	General	4
B	Information Area	4
C	Mathematics and Physico-Technical Sciences	8
D	Physico-Chemical Technology	3
E	Astronomi and Geosciences	7
F	Agriculture and Nutrition	6
G	Biomedical Sciences	4
H	Regional and Environmental Sciences	5
J	Social Sciences	10
K	Culture and the Arts	10

For hver av de 462 tesauri er det gitt en "system-beskrivelse" inndelt i 15 punkter. Gjennom disse får man en god oversikt over hva den enkelte thesaurus står for. De 15 punktene har følgende styretekster (endel forenklet):

- 1 Hovedklasse, underklasse, språk, identifikasjons-nummer
- 2 Tittel og publikasjons-detajler
- 3 Ansvarlig organisasjon
- 4 Originator (dvs den som fikk denne thesaurus igang)

- 5 Språk som denne tesaurus fins på - samt, hvis mer enn ett språk, opplegg for samspill mellom disse
- 6 Medium som denne tesaurus er tilgjengelig i (f eks mikrofilm, papirutskrift, floppydisk)
- 7 Faglig deknings-område (totalt, hoved-område og grense-områder)
- 8 Antall termer, deskriptorer, ikke-deskriptorer og modifikatorer. Entall eller flertall. Term-typer. Sikring mot tve-tydighet. Indekserings-hyppighet. Kronologisk historie om innlegging og stryking. Relasjons-former mellom termene.
- 9 Presentasjons-former brukt for denne tesaurus
- 10 Faglig inndelings-prinsipp. Antall klasser og sub-klasser; antall nivåer
- 11 Typer av interne relasjoner (eks: synonymer, ekvivalens-relasjoner, alternativ-relasjoner, hierarki-relasjoner)
- 12 Notasjoner. (Eks: alfanumeriske deskriptorer, numeriske deskriptorer)
- 13 Konstruksjon og vedlikehold. (Bruker-beskrivelse, karakter-sett, maksimal deskriptor-lengde, EDB-program-system brukt, dator-type brukt, oppdaterings-frekvens)
- 14 Implementering. (Navn på "databanken", service-organisasjonen eller typer av sevice som tilbys)
- 15 Diverse. (Eks: Planer om utvidelser, forbedringer etc, henvisninger til andre beslektede tesauri på andre språk)

6.4 Eksisterende emneord-opplegg for litteratur-referanser

Dette sub-kapittel inneholder - som nevnt ovenfor - ikke noen oversikt over alle eksisterende emneord-opplegg for litteratur-referanser etc som ble funnet i leterunden. Kun de opplegg som har hatt betydning for rapportens konklusjoner, vil bli omtalt.

Omtalen har stort sett måttet bli summarisk og lite detaljert. Dette fordi det ikke var mulig å studere oppleggene noe nærmere innenfor forprosjektets ramme. Av denne grunn må det presiseres at hverken utvalget av opplegg eller beskrivelsene av de utvalgte kan danne grunnlag for noen evaluering (absolutt eller relativ) av de enkelte opplegg.

I alt ble det i løpet av lete-perioden sett på 41 emneord-opplegg beregnet for litteratur-referanser etc. Av disse ble 7 funnet ved kontakt med nordiske institusjoner, mens 34 ble plukket ut fra EFs oppslagsverk "Thesaurus Guide" (ref 1).

De 34 oppleggene som ble plukket ut fra Thesaurus Guide, hadde alle "miljøvern" i en eller annen form som et vesentlig aspekt. De ble imidlertid funnet under forskjellige fag-grupper i oppslagsverket:

Gruppe	Antall opplegg
Agrikultur	1
Akvakultur	1
Akvatisk vitenskap og fiskerier	2
By- og region-planlegging	4
Emmisjons-kontroll	1
Fast avfall	1
Hydrografi	1
Hydrologi	1
Luft	2
Marint miljø	1
Miljø-lovgivning	2
Olje og gass	2
Oseanografi	1
Naturmiljø	8
Støy	1
Vann	2
Vann i naturmiljøet	1
Vann-ressurser	1
Vann-teknologi	1

Den Thesaurus Guide som er benyttet, ble utgitt i 1985 og er basert på innsamlet materiale gjennom flere år forut for utgivelsen. Utviklingen vil utvilsomt være kommet betydelig lenger pr idag. Forprosjektet har ikke hatt mulighet til undersøke hvor mye lengre.

Den tilsvarende oppstilling for de 7 opplegg som ble funnet i Norden - og vurdert som noe interessante - er følgende:

Gruppe	Antall opplegg
Geologi	2
Miljø	2
Olje	1
Vann	2

Ved videre leting i Norden kunne antagelig noen fler ha blitt funnet. Letingen her ble imidlertid avbrutt da Thesaurus Guide dukket opp.

Blant de 34 + 7 emneord-oppleggene som er vurdert som noe interessan-

te, fins det flere varianter mht størrelse, kompleksitet, indre struktur mv. Det har ikke vært mulig, innenfor forprosjekt-rammen, å studere noen av dem inngående. Av denne grunn kan det ikke her settes opp noen rangering av oppleggene mht hvor godt de egner seg som studie-materiale og/eller mønster for neste fase i prosjektet. Det eneste som kan gjøres i denne retning, er å velge ut noen av de tilsynelatende best egnede oppleggene og gi en kort presentasjon av disse. Dette er gjort i form av en tabell-opstilling samt noen få kommentarer:

I tabell 1 er 16 utvalgte tesauri ført opp med endel sentrale faktiske informasjoner. Kriteriene for utvelgelse var tildels å få flest mulig av de tema-områder representert som hører inn under begrepet "det ytre miljø" - og tildels å få med opplegg som synes å være gode, sett fra en infoteknisk synsvinkel.

Blant de 16 oppleggene er 12 hentet fra Tesauros Guide og de 4 øvrige fra Norden.

Kommentarer:

- 1 Den samling av emneord-lister av kategori "litteratur" som er funnet for tema-område miljø, må antas å ha fått med de fleste av de nasjonale og internasjonale som fantes i den vestlige verden pr 1983/84. Tabell 1 viser de antatt viktigste av disse - sett fra prosjektets synsvinkel. Men også mange av dem som ikke kom med i tabellen, bør studeres nærmere i det etterfølgende hovedprosjekt.
- 2 Tabell 1 viser at ganske mange tema-områder under "det ytre miljø" er representert i denne samlingen.
- 3 Mange av oppleggene er basert helt eller delvis på ISO-standardene. Det betyr at de må ha en ganske høy infoteknisk standard - og at de strukturelt sett må være nokså like. Dette siste innebærer er at det vil være relativt lett å bruke dem som et felles grunnlag når den norske emneord-samlingen skal utvikles.

6.5 Emneord-opplegg for dataserie-referanser

Det ble i leterunden, hverken på nasjonalt eller internasjonalt nivå, funnet noen ferdige referanse-opplegg for dataserier - og heller ikke ble det funnet noen ferdige emneord-opplegg laget spesielt for slike kategorier av registre.

Av slike referanse-opplegg under utvikling ble det funnet ett - NVEs "referanse-register for dataserier". Noe eget emneord-opplegg for dette registeret vil ikke bli laget foreløbig, idet man der først vil avvente den videre utvikling av Statens Kartverks emneord-prosjekt.

Mht emneord-opplegg under utvikling for dataserie-referanse-registre, er det gjort noe og - gjøres fortsatt noe - i regi av Nordisk Ministerråd:

- I rapportene "Karakterisering av miljödata" (ref 2) og "Characterization of Environmental Data - Handbook Part 1" (ref 3), er de prinsipielle sidene av dette tema utredet. I den sistnevnte fins det dessuten utkast til praktiske retningslinjer.
- En "Nordisk Kodcentral" er opprettet som har som oppgave å utarbeide biologiske taxon-lister ihht disse prinsippene. Endel lister er ferdige.
- En egen arbeidsgruppe med finsk og svensk bemanning (opprettet tidlig i 1984) har som oppgave å utarbeide forslag til "emneord-liste" for fysisk-kjemiske parametere og metoder på miljø-siden. Ingen rapport foreligger ennå fra denne gruppen.

7. SKISSE TIL PROGRAM FOR DET EFTERFØLGENDE HOVEDPROSJEKT

Opprinnelig var det tenkt at forprosjektet skulle fremlegge en skisse til mål og arbeidsprogram for et etterfølgende hovedprosjekt.

Resultatene fra forprosjektet viser imidlertid at det foreløpig ikke fins et godt nok grunnlag for å lage program for hele hovedprosjektet. Før det kan skje, må det materiale som nå foreligger, studeres nærmere. Derneft må delmål og prinsipper vurderes videre både for referansetjenesten som helhet og for emneord-mekanismen spesielt. Samtidig må man diskutere både med dem som er kommet lengre enn oss (ISO, EF mv), med dem som referansetjenesten skal "spille sammen med" i fremtiden (tilsvarende referansetjenester for tilgrensende tema-områder) og med de fremtidige brukere av referansetjenesten for tema-område miljø.

Forprosjektet vil istedet skissere et program bare for den første fase av hoved-prosjektet:

- 1 Det tas kontakt med de institusjoner som står bak de viktigste funnene i denne rapport. Man bør starte med ISO og EF - og der søke råd om hvilke andre institusjoner/ personer som man eventuelt bør ta kontakt med. Hos disse kontaktene bør man be om orienteringer mht status pr idag, planer videre og erfaringer så langt. Videre bør man skissere egne planer for dem og diskutere prinsipper, mål og muligheter i den forbindelse. Og endelig bør man undersøke mulighetene for samarbeid (som bl annet kan innebære at man får anledning til å bruke deres lister som basis for egen utvikling).
- 2 Det må treffes en endelig beslutning om hvilken hovedstruktur "det nasjonale referanseregister for miljø-informasjon" i Norge skal ha.
- 3 På grunnlag av punktene 1 og 2 vil et program for resten av hovedprosjektet kunne utarbeides.

LITTERATUR SOM DET ER HENVIST TIL

-
- 1 Gesellschaft für Information und Dokumentation: Thesaurus Guide, Analytical directory of selected vocabularies for information retrieval. Office for Official Publications of the European Communities, Luxembourg (1985)
 - 2 Nordisk Ministerråd: Karakterisering av miljödata. B 1978:10 Oslo (mars 1978)
 - 3 Nordisk Ministerråd: Characterization of Environmental data, Handbook Part 1 - Introduction and Element Standardization. Oslo (september 1981)

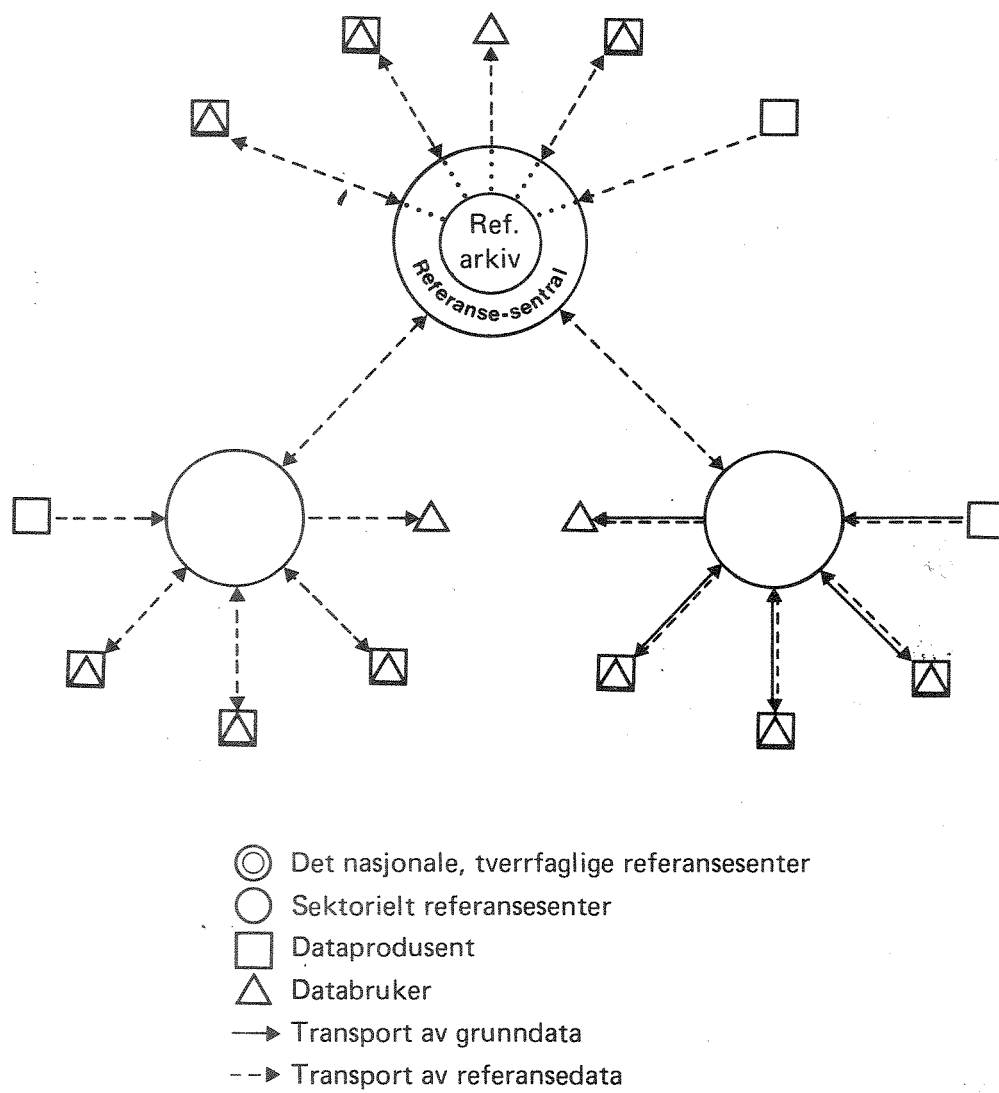


Fig 1 Det nasjonale referansesenter for miljø-informasjon

Tabell 1 FAKTA OM ENDEL UTVALGTE EMNEORD-OPPLEGG FOR REFERANSER AV KATEGORI "LITTERATUR"

Side 1/2

Navn	Land/inter- nasj. organ.	Deknings-område	Språk	Ca antall termer pr språk Totalt/ Deskript.	Hierark. struktur	Følger ISO-st.	Ref nr i T.G.
Water Resources Thesaurus	USA (WRSIC)	Water Resources	Eng	5.500 4.300	ja	ja	E307
API Thesaurus	USA (API)	Petroleum Refining Technology	Eng	11.000 5.700	ja	delvis	E353
A Multilingual Thesaurus of Agricultural Terminology	EF	Agriculture	Eng, Fr, Ger, Ital, Span	17.000 8.700	Ja	ja	F101
Thesaurus of Terms for Aquatic Sciences and Fisheries	FAO	Aquatic Sciences and Fisheries	Eng, Span, (Fr planned)	7.200 4.700	ja	ja	F207
Umweltthesaurus	Tyskland	Umwelt	Ger, (Eng planlagt)	10.000 10.000	ja	?	H101
Infoterra	FN (UNEP)	Miljø	Eng, Fr, Span, Rus, Gresk, Arab, Kin	1.500 1.500	ja	?	H103
Standard Air Pollution Classifi- cation Network: A Thesaurus of Terms	USA (EPA)	Air pollution	Eng	? 2.300	ja	ja	H109
Thesaurus Immissionsschutz	Tyskland	Immissionsschutz	Ger	1.700 1.400	ja	?	H110
Thesaurus Nuisances Acoustiques	Frankrike	Nuisances Acoustiques	Fr	1.400 1.100	ja	delvis	H111

Navn	Land/inter- nasj. organ.	Deknings-område	Språk	Ca antall termer pr språk Totalt/ Deskript.	Hierark. struktur	Følger ISO-st.	Ref nr i T.G.
Solid Wastes Thesaurus	Sveits	Solid Waste Management	Eng	700 500	nei	?	H113
"IUCN-Thesaurus"	IUCN	Environmental Law	Eng, Fr, Ger, Span	2.400 2.300	nei	nei	H114
Thesaurus	Frankrike	Regional and Urban Plan- ning and housing in the developing countries	Eng, Fr	2.500 400	ja	delvis	H206
Petroleum-tesaurus	Norge	Petroleum-fagene	Norsk	? 8.000	ja	delvis?	-
Swedish Environmental Research Index (SERIX)	Sverige	Ytre miljø og arbeids- miljø	Svensk (Eng?)	? ?	ja	delvis?	-
Emneordliste for NGUs referanse- arkiv	Norge	Geologi	Norsk	500 400	ja	delvis	-
GEOREX	Sverige	Geologi	Svensk	? ?	ja	?	-