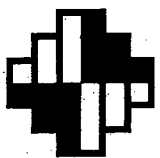


F153



muistio

Tilastokeskus

Tekijä	Päiväys	N:o
Tuula Nurminen	13.6.1977	45

TIETOKANTAMENETELMIEN KÄYTTÖMAHDOLLISUUDET TILASTOTUOTANNOSSA

- 1 Nykyinen tuotantotapa ja sen puutteet
- 2 Eräitä peruskäsitteitä
- 3 Tietokantamenetelmien soveltuvuus tilastotuotantoon ja käytön tavoitteet
- 4 Tilastotuotannossa käytettävän tiedonhallintajärjestelmän ominaisuudet
- 5 Tulevaisuuden näkymät
- 6 Siirtyminen tietokantapohjaiseen tuotantotapaan



1 Nykyinen tuotantotapa ja sen puutteet

Tilastokeskuksen tietojenkäsittelyjärjestelmä muodostuu useista kymmenistä toisistaan melko riippumattomista tilastokohtaisista tietosysteemeistä:

- Tiedot kerätään ja käsitellään yleensä kutakin tilastoa varten erikseen.
- Jokaista tuotettavaa tilastoa varten on omat sovellutusohjelmansa ja tiedostonsa.
- Tiedostot on suunniteltu lähinnä palvelemaan tiedonvälitystä eri käsittelyvaiheiden välillä eikä käsiteltävien tietojen loogiseen rakenteeseen ja fyysiseen ympäristöön olla paneuduttu.

Tällaiseen tuotantotapaan liittyy useita ongelmia ja puutteita.

Redundanssi (päällekkäisyys, ylimäärä)

Tarpeettomasta tietojen päällekkäisestä keruusta, tarkistuksesta ja korjauksesta olisi yritettävä päästä, jotta tilastotuotannolle tyypillistä erittäin työlästä tiedon syöttövaihetta pystyttäisiin helpottamaan. Päällekkäinen keruu aiheuttaa myös tiedonantajien ylimääräistä vaivaa.

Kun tietyt tiedot ovat talletettuna usean eri sovellutuksen tiedostoihin, ei tietojen yhtäpitävyyttä pystytä takaamaan. Tietyn tietosysteemin piirissäkin on tuotantotavasta johtuen tiedostoista useita eri versioita, jolloin tiedostoja koskeva kirjanpito saattaa tuottaa hankaluuksia.

Samoja tietoja eri tietosysteemeissä käsittelevät ohjelmat voivat sisältää samanlaisia osia, mikä aiheuttaa ylimääräisiä ohjelmointi- ja ylläpitokustannuksia.

Kielteisenä ei sen sijaan voida pitää tietojen tallettamista sekä mikro että makrotasolla, koska tämä nopeuttaa käsittelyä.

Heikot edellytykset uusien tietotarpeiden tyydyttämiseen ja huono ylläpidettävyys

Tilastokohtaiset tietosysteemit tuottavat tulosteensa ennalta määrättyssä muodossa tietyin aikaväleihin. Voidaan kuitenkin haluta tulosteita, joissa esim. tilastoyksikön valintakriteeri, luokittelu, tarkasteluajaväli tai tuottamisajankohta poikkeaa aikaisemmasta. Ennakoimaton tietojen käyttö tai tietosisällön muuttaminen aiheuttaa yleensä kohtuuttoman suuren lisätyön. Ylläpitotoimenpiteen aiheuttajana voi olla myös tietosysteemin käytössä havaittu virhe tai halu kehittää tietosysteemiä.

Valtion tilastotoimen kehittämishjelmassa esitetään tavoitteeksi, että eri ajankohtina ja eri yhteyksissä kerättyjä perustietoja tulisi voida käyttää toistuvasti ja toisiinsa yhdistellen tilastotuotannossa. Tiettyä tarkoitusta varten kerättyjä ja talletettuja tietoja ei kuitenkaan helposti voida käyttää toisiin tarkoituksiin erillisiin tilastokohtaisiin tietosysteemeihin perustuvassa järjestelmässä. Tietojen huonoon yhdisteltävyyteen on syynä mm. käsitteiden, määritelmien ja luokitusten erilaisuus eri sovellutuksissa, tietojen erilainen järjestys eri tiedostoissa ja runsas peräkkäistiedostojen käyttö.

Kokonaisvaltainen ohjauksen ja valvonnan vaikeus

Erillisiin tietosysteemeihin perustuva tietojenkäsittelyjärjestelmä ei tarjoa kovin hyviä mahdollisuuksia kokonaisvaltaiseen ohjaukseen ja valvontaan. Tietosysteemien suunnittelusta ja toteuttamisesta vastaavat projektit ovat yleensä suhteellisen itsenäisiä ja niiden toiminnan ohjaaminen ja valvonta on työlästä. Tietosysteemien määrän kasvaessa vaivaa tietojenkäsittelyjärjestelmää helposti yleinen koordinoimattomuus.

Tilastovirastossa käsitellään niin suuria tietomääriä, että tietojen hoito ja valvonta on erityisen vaikeata. Eri ajankohtina ja eri tarkoituksiin kerätyt tiedot sijaitsevat hajallaan sovellutuskohteisissa tiedostoissa eikä tiedonhallintaan ole apuvälineitä. Ensimmäiseksi törmätään vaikeuteen hallita tiedon kuvaaminen. Tietoja koskevien tietojen saaminen on työlästä ja usein jopa mahdotontakin. Tietosysteemien kunnollinen kuvaaminen laiminlyödään usein, jolloin tavallisesti syytetään aika- tai resurssipulaa.

Toisaalta vaikka halua systemaattiseen kuvaamiseen olisikin, niin vaikeuttaa kuvaamisen suorittamista tavallisten kuvausvälineiden (kynä, paperi, kansio) kykenemättömyys tyydyttää kuvaajan ja toisaalta kuvausten hyväksikäyttäjän tarpeet.

Ongelmia ei välttämättä esiinny yksittäisiä tilastoja tuotettaessa, vaan ne ilmenevät, kun yhteiskunnan toimintoja ja tiloja pyritään kuvaamaan kokonaisvaltaisten tilastojärjestelmien avulla ja halutaan rakentaa tilastokohtaisten tietosysteemien muodostamalle perustalle tarvittavat tiedot tuotavia järjestelmiä, tai kun yritetään tyydyttää satunnaisia tietotarpeita.

Puutteellinen ohjaus ja valvonta saattaa aiheuttaa mm. seuraavia tilanteita (/1/ s. 9):

- Sama tietoalkio esiintyy erinimisenä eri tiedostoissa.
- Eri tietoalkiot esiintyvät samannimisinä eri tiedostoissa.
- Tietoalkion arvot on kooditettu eri tavoin eri tiedostoissa.
- Tietoalkion mahdolliset arvot poikkeavat eri tiedostoissa toisistaan.
- Tietoalkion tietyllä koodilla on eri merkitys eri tiedostoissa.
- Samaa merkitsevien tietojen arvojen yhtäpitävyyttä eri tiedostoissa ei pystytä turvaamaan.
- Tietojen päivityksen oikeellisuutta ei pystytä valvomaan.
- Tiedon redundanssia ei pystytä valvomaan eikä vähentämään.
- Tietojen käyttäjiä ei pystytä informoimaan tarvittavassa määrin tietoja koskevissa asioissa.
- Eri käyttäjien tarpeita ei pystytä koordinoimaan.
- Tietojen ominaisuuksiin kohdistuvat muutokset aiheuttavat kohtuutonta ohjelmien ylläpitotyötä.
- Tiedon suojausta ja varmistusta ei pystytä järjestämään riittäväksi.

2 Eräitä peruskäsitteitä

Käsitteen tietokanta merkitystä on erittäin vaikeaa ellei mahdotonta määritellä siten, että voitaisiin yksikäsitteisesti ilmoittaa, muodostaako tietty tietojoukko tietokannan.

Seuraavassa on lueteltu ominaisuuksia, joita tietokannalla yleensä katsotaan olevan (/1/ s. 22):

- (1) Tietokannan tehtävänä on vastaanottaa, säilyttää ja tuottaa reaalimaailman jotain osaa koskevaa tietoa. Tietokanta kuvaa reaalimaailman rakennetta, ts. reaalimaailman yksilöitä ja niiden välisiä yhteyksiä. Tietokantaa päivitetään reaalimaailman tapahtumia koskevilla tiedoilla.
- (2) Tietokannan avulla pyritään tuottamaan tietoja useille toimintajärjestelmän toimintatasoille ja palvelemaan useita eri tietojenkäsittelysovellutuksia, mutta sen rakenne ja käyttö on suunniteltu keskitetysti yhdenmukaisia periaatteita noudattaen ja eri sovellutusten kokonaistietotarve huomioonottaen.
- (3) Tietokannan kanssa kommunikoivat sovellutusohjelmat käyttävät hyväkseen keskitettyä tietokannan kuvausta, joka sijaitsee erillään näistä ohjelmista.
- (4) Tietokannan rakenteen ja käytön valvonta on keskitetty erilliselle organisaatioyksikölle. Ainoastaan tämä organisaatioyksikkö voi muuttaa tietokannan rakennetta.
- (5) Tarpeeton tiedon redundanssi on pyritty eliminoimaan tietokantaa suunniteltaessa.
- (6) Tietokanta sisältää joitain sellaisia tietoja, joihin voidaan päästä käsiksi useita eri saantipolkuja käyttäen.

- (7) Tietokanta koostuu loogisesti yhteenliittyvistä tiedoista, jotka voi olla talletettu useisiin käyttöjärjestelmän tuntemiin tiedostoihin, joiden välillä vallitsee riippuvuussuhteita.
- (8) Suurin osa tietokannasta on talletettu suorasaanti-muistiin.

On huomattava, että mikäli tietojoukolla on joitain tietokannalle ominaisia piirteitä, ei ole mitään estettä sille, että sitä kutsutaan tietokannaksi, vaikka

- siihen ei kohdistuisi minkäänlaisia on-line -kyselyitä ja -päivityksiä,
- se olisikin kokonaan talletettu magneettinauhoille tms. sarjamuistilaitteille,
- se sisältäisikin esim. tehokkuus- tai varmistussyistä redundanttia informaatiota.

Tietokannan käsitteeseen yhdistetään usein sellainen harhaluulo, että tietokanta käsittää kaikki toimintayksikön tietoalkiot. Tästä harhaluulosta seuraa toinen virheellinen käsitys, jonka mukaan toimintayksikössä voi olla korkeintaan yksi tietokanta. Sen sijaan voidaan kyllä puhua koko toimintayksikön kattavasta tietojen organisointiperiaatteesta.

Tietokantajärjestelmäksi tai tietokantapohjaiseksi järjestelmäksi voidaan kutsua kokonaisuutta, jonka muodostavat

- tietokanta,
- tietokannan kuvaukset,
- tiedonhallintajärjestelmä,
- tietohakemistojärjestelmä ja
- käyttäjäliitännät (vrt. /1/ s. 26 ja /2/ s. 1).

Tiedonhallintajärjestelmä (data base management system) on ohjelmisto, joka suorittaa tiedon talletusta, hakua ja ylläpitoa.

Yleistä tiedonhallintajärjestelmää voidaan käyttää useissa eri sovellutuksissa.

Erityistä tarkoitusta varten tehdyt l. räätälintyönä tehdyt tiedonhallintajärjestelmät on suunniteltu jonkun tietyn käyttäjän tai käyttäjäryhmän tiettyjen tarpeiden mukaisiksi.
(kts. /1/ luku 5.5.2)

Tietohakemistojärjestelmä (data dictionary/directory system) on ohjelmisto, jota käytetään apuvälineenä tiedon kuvausten tallettamisessa, ylläpidossa ja analysoinnissa. Tietohakemistojärjestelmän tehtävänä on tuottaa tietoa tiedoista (meta-tietoa).

Käyttäjäliitännöillä tarkoitetaan niitä systeemin osia, joiden välityksellä käyttäjät ovat yhteydessä tietokannan kanssa. Näitä osia ovat tietokantaa käsittelevät sovellutusohjelmat sekä kielet, joiden avulla eritasoiset tietojen hyväksikäyttäjät voivat esittää omia informaatiopyyntöjään. Nämä kielet katsotaan usein kuuluviksi tiedonhallintajärjestelmään, jonka kiinteään yhteyteen myös sovellutusohjelmistoa on voitu rakentaa.

Tiedonhoitaja (data base administrator tai data administrator) on henkilö tai henkilöryhmä, joka on vastuussa tietojen hoidosta ja suorittaa keskitetysti tietokannan tietosisällön, rakenteen ja käytön sekä tiedonhallintaohjelmiston suunnittelua, toteuttamista ja ylläpitoa sekä tietokantaan liittyvää koordinoitua, koulutusta, yhteydenpitoa ja dokumentointia.

3 Tietokantamenetelmien soveltuvuus tilastotuotantoon ja käytön tavoitteet

Tarkastellaan tietokantamenetelmien soveltuvuutta tilastotuotantoon ensin tietojen käsittelyyn liittyvien toimintojen ja sitten tilastoaineistojen kannalta (vrt. /3/ luku 3.1).

Tilastojen tuottaminen on prosessi, jossa tietyt vaiheet toistuvat sovellutuksesta toiseen vain pienin muutoksin. Tiedot kerätään, tarkistetaan, korjataan, taulukoidaan sekä esitetään vaaditussa muodossa. Varsinaisia tiedostojen päivitystarpeita ei esiinny, koska tietyn kauden tiedot sisältävä tiedosto luodaan aina uudestaan tietyin aikaväleihin. Eri ajankohtien tietoja voidaan liittää aikasarjatiedostoihin, kun halutaan tallettaa tietyn ilmiön kehitystä koskevat tiedot. Voisi hyvinkin tuntua siltä, että tietokantamenetelmillä ei olisi käyttöä tilastollista tietojenkäsittelyjärjestelmää rakennettaessa.

Tilastokohtaisia tietosysteemejä suunniteltaessa onkin yleensä kiinnitetty ennen kaikkea huomiota tietojenkäsittelyyn liittyviin prosesseihin, eikä näissä vaiheissa käsiteltävän tiedon loogiseen ja fyysiseen rakenteeseen olla erityisemmin paneuduttu. Tällaisessa tuotantotavassa esiintyy luvussa 1 kuvattuja ongelmia.

Tietojenkäsittelyjärjestelmä voitaisiin kuitenkin suunnitella sellaiseksi, että sovellutukset rakennetaan tietojen ympärille (kts. /4/ luku 3.4). Tällöin muodostaa tiedostojen ja niiden hallinnan suunnittelu koko järjestelmän toteuttamisen perustan. Tilastotuotannon eri toiminnot suorittavat modulit voidaan tällöin suunnitella itsenäisiksi ja toisistaan riippumatta tietokannan tietoja käsitteleviksi.

Tällöin ei enää välttämättä synny jokaisen käsittelyvaiheen seurauksena uusia tiedostoja. Myöskään ei tällöin ole olemassa mitään teknisiä syitä, jotka vaatisivat tuotannon eri vaiheiden suorittamisen tarkasti tietyssä järjestyksessä, kuten nykyisin on usein asianlaita.

Tiedot voitaisiin tallettaa suoraan tietokantaan ja tarkistaa ne vasta tämän jälkeen sekä päivittää tietokantaa korjauksilla. Tiedot olisivat tällöin koko tarkistus ja korjausvaiheen ajan käytettävissä esim. ennakkotulosten tekemiseen. Eri tuotantovaiheita voitaisiin tällöin suorittaa samanaikaisesti.

Tilastollisella informaatiojärjestelmällä on oltava käytettävissään tietoja tilastoyksiköiden olemassaolosta ja ominaisuuksista sekä tietoja antavista yksiköistä. Näiden tietojen tallettamiseksi on rakennettu erilaisia rekisterijärjestelmiä.

Tietoja antavien yksiköiden nimi, osoite ym. tiedot vaativat jatkuvaa päivittämistä. Rekistereiden tietoihin tulisi joustavasti voida liittää yksiköitä koskevia muita tietoja, ja tietojen tulisi olla kaikkien niitä tarvitsevien sovellustusten käytettävissä.

Tietokantamenetelmät tuntuisivat soveltuvan käytettäväksi rekisterijärjestelmien toteutuksessa.

Kun ilmiöitä halutaan tarkastella kokonaisvaltaisesti, joudutaan yleensä käyttämään hyväksi eri tilastojen keräämiä ja tallettamia tietoja. Tallettamiseen tulisi tapahtua yhtenäisten organisointiperiaatteiden mukaisesti, jotta tietojen käyttöön-saattaminen ei tuottaisi kohtuutonta työmäärää. Tietokantajärjestelmien avulla voitaisiin ilmeisesti mahdollistaa tilastoaineistojen joustava ja riittävän nopea saanti esim. analyttisiin tarkoituksiin.

Tilastoaineistot voidaan jakaa kahteen ryhmään:

- mikrotiedot eli alkuperäiset yksikkötason tiedot
- makrotiedot eli aggregoidut tiedot.

Tilastoaineistojen erityispiirteenä on yksikkötason tietojen valtava määrä. Näiden tietojen määrä kasvaa koko ajan, koska eri kausien tiedot on talletettava, jotta voitaisiin tarvittaessa kuvata jonkin ilmiön kehitystä. Historiatiedot muodostavat siis tärkeän osan järjestelmän tiedoista.

Tämä aiheuttaa erityisiä vaatimuksia muistilaitteistoille. Perusaineistot talletetaan tavallisesti magneettinauhoille, jotka eivät ole kovin sopivia välineitä, mikäli halutaan käyttää nykyisin markkinoilla olevia tiedonhallintajärjestelmiä.

Edelleen ei perusaineistojen käsittelyssä esiinny tarpeita monimutkaisten tietorakenteiden käyttöön, vaan luotaessa tiedostoja sekä tarkistettaessa ja korjattaessa niitä on yleensä luonnollisinta käyttää yksinkertaisia lineaarisia rakenteita ja peräkkäiskäsittelyä.

Tietokantajärjestelmien avulla voitaisiin ilmeisesti kuitenkin vähentää tiedon redundanssia ja parantaa tietojen käytettävyyttä muihin kuin alkuperäisiin tarkoituksiin sekä päästä siitä suuresta määrästä tosiasiallisen tehtävän kannalta epäoleellisia apuohjelmia, joita nykyään tarvitaan etsimään ja lajittelemaan tietueita ennen varsinaista käsittelyä.

Makrotason tiedot eli varsinainen tilastollinen tieto koskee yksikköryhmiä ja esitetään taulukkomuodossa, jolloin luontevin tapa kuvata tietojen välisiä loogisia riippuvuussuhteita on käyttää matriisimuotoista tietorakennetta. Nykyisin saatavissa olevat yleiset tiedonhallintajärjestelmät eivät kuitenkaan salli tämän rakenteen käyttöä tietokannan loogisen strukturoinnin välineenä.

Aikasarjatietoja on syytä tarkastella erillään muista makrotason tiedoista, koska niiden tallettamisessa joudutaan ottamaan huomioon erityispiirteitä.

Makrotason aineistojen joukossa on sellaisia, joita melko usein voidaan tarvita nopeasti tiettyihin erityistarkoituksiin, joten tietokantamenetelmien hyväksikäyttö näiden aineistojen tallettamisessa tuntuisi luontevalta.

Ei voida ajatellakaan, että tilastolliset tietoaineistot muodostaisivat yhden tietokannan. Käytännössä voisi olla useita ehkä kymmeniä tietokantoja. Kuitenkin vaikka nämä tietokannat olisivatkin loogisesti ja fyysisesti erillisiä, saattaisivat ne olla toteutettu samaa tiedonhallintajärjestelmää käyttäen. Edelleen tulisi tiedon kuvausten olla laadittu ja talletettu yhtenäisten periaatteiden mukaisesti.

Tietokantamenetelmien käyttöön tilastotuotannossa voidaan liittää seuraavia tavoitteita (vrt. /1/ s. 44, /2/ s. 1 ja /3/ s. 9):

- Järjestelmän yhtenäistäminen

Tavoitteena on eri ajankohtina ja eri yhteyksissä kerättyjen tietojen yhdisteltävyyden lisääminen. Edelleen olisi pyrittävä siihen, että useat sovellukset voivat käyttää samoja tietoja, jolloin on mahdollista pienentää kerättävien tietojen määrää.

- Tietojen tallettaminen siten, että ne ovat tarvittaessa käytettävissä

Jo vuosia on puhuttu yhtenäistettyjen tilastoaineistojen järjestelmästä, joka asettaa tavoitteeksi sen, että yksikkötason tiedot varastoidaan systemaattisesti siten, että ne ovat toistuvasti käytettävissä. Tällaisen järjestelmän toteuttaminen ei ole mahdollista erillisten tilastokohtaisten tietosysteemien pohjalta, vaan se vaatii kokonaisvaltaista lähestymistapaa eli tietokantaajattelun soveltamista tilastotuotantoon.

- Tietojen käytön helpottaminen ja automatisointi tallettamalla tietoja koskevat kuvaukset

- Käyttäjystävällisemmän järjestelmän luominen

Tietokantamenetelmien avulla voidaan pyrkiä tiedontarvitsijoiden parempaan palvelemiseen, mutta niitä voidaan käyttää myös sisäisenä työkaluna, jonka avulla tilastohenkilöstö saa aikaisempaa paremman kosketuksen itse tietoihin tuotannon eri vaiheissa.

- Tietojen hoidon ja valvonnan keskittäminen

Mikäli halutaan pienentää järjestelmässä esiintyvää tarpeetonta redundanssia sekä hoitaa tietojen varmistus ja suojauskysymykset, on luotava edellytykset kokonaisvaltaiseen ohjaukseen ja valvontaan.

- Tietoriippumattomuuden lisääminen

Tietoriippumattomuutta pidetään eräänä tärkeimmistä tietokantamenetelmien käytön tavoitteista. Tietosysteemien joustavuus lisääntyy, kun tiedot ja niitä käyttävät sovellutusohjelmat ovat mahdollisimman riippumattomia toisistaan. Tällöin voidaan tietoja muuttaa ilman sovellutusohjelmien muuttamista.

- Tietosysteemien suunnittelu-, ohjelmointi ja ohjelmien ylläpitokustannusten pienentäminen

Tietokantamenetelmien käytöllä voidaan vaikuttaa systeemityöhön. Tavoitteena on luoda ympäristö, jossa uusien sovellutusten toteuttaminen helpottuu valmiiden apuvälineiden ansiosta. Tietokantatekniikka vapauttaa sovellutusohjelmoijan tiedon fyysisten talletusmenetelmien aiheuttamista ongelmista.

4 Tilastotuotannossa käytettävän tiedonhallintajärjestelmän ominaisuudet

Tilastoaineistoihin sekä niiden hallintaan ja käsittelyyn liittyy erityispiirteitä, joiden johdosta eivät nykyisin tarjolla olevat yleiset tiedonhallintajärjestelmät ole erityisen soveliaita.

Tarkastellaan vaadittavia ominaisuuksia kolmessa ryhmässä (kts. /3/ luku 5):

- tiedonhallinta
- tietohakemistot
- käyttäjäliitännät.

Tiedonhallinta

a) On tarpeen kuvata tiedot eri tasoilla, koska tavoitteena on, että käyttäjän tarvitsee tietää mahdollisimman vähän tietojen talletusrakenteesta sekä se, että käyttäjän tarvitsee tuntea ainoastaan se osa tietokantaa, jota hän käsittelee.

Tilastotuotannon kannalta on oleellista, että kuvaamiseen ei saa vaikuttaa se, talletetaanko tiedot suorasaantilaitteille tai magneettinauhoille. Tiedon kuvaukset tulee voida laatia samaa tiedonkuvauskieltä käyttäen.

Kuvaamisessa on otettava huomioon myös tietohakemistojen käyttöön liittyvät tarpeet. Tiedon kuvauskielen on perustuttava tilastotuotannon käsitteistöön.

b) Erilaisten tilastoaineistojen määrä on niin suuri, että ne käytännössä muodostavat useita ehkä kymmeniä tietokantoja, jotka ovat loogisesti ja fyysisesti toisistaan erillisiä. Kuitenkin olisi pyrittävä koko tilastotuotannon kattavaan yhtenäiseen tietojen organisointiperiaatteeseen.

Nykyisin tarjolla olevien kaupallisten tiedonhallintajärjestelmien käyttöön liittyy vielä monia ongelmia (kts. /4/ s. 29). Niiden eräänä tärkeimpänä etuna pidetään yleensä niiden joustavuutta. Ne yrittävät sallia mahdollisimman monenlaisten tietorakenteiden sekä talletusrakenteiden käytön. Tiedonhallinnan tehokkuus edellyttää tietokannan tietojen käytön kannalta sopivaa organisointia. Onnistuneen tieto- ja talletusrakenteen löytäminen on nykyisin tietokantasovellutusten suunnittelun eräs vaikeimpia tehtäviä. Tiedonhallinnan tehokkuuteen joudutaan tilastotuotannossa kiinnittämään erityistä huomiota, koska käsiteltävät aineistot ovat yleensä suuria.

Tarvitaan menetelmiä, joiden avulla tietokannan tiedostojen suunnittelijat voivat löytää eri tyyppisiin aineistoihin sopivat tietorakenteet ja edelleen talletusrakenteet. Tarvitaan apuvälineitä tietokannan loogiseen ja fyysiseen strukturointiin.

Tietorakenteiden teorian alalla on kehityssuunta ns. relaatiomalli (tietynlainen taulukkoesitys, kehittäjä E.F. Codd), joka on tilastotuotannon kannalta mielenkiintoinen. Voidaan nimittäin osoittaa, että tiedosto, olipa sillä kuinka monimutkainen tietorakenne tahansa, voidaan hajoittaa osiin, joilla kaikilla on samanlainen erittäin yksinkertainen tietorakenne (/4/ s. 30).

Näyttää siis siltä, että on mahdollista standardoida käytettävät tieto- ja talletusrakenteet ja näin suunnitella tietokantajärjestelmiä, jotka ovat tarpeeksi tehokkaista tilastotuotannossa käytettäviksi. Edelleen kokoaisi tietojen yhteinen organisointitapa ja saman tiedonhallintajärjestelmän käyttö toisistaan loogisesti ja fyysisesti erilliset tietokannat yhtenäiseksi kokonaisuudeksi.

c) Tietojen tallettamisessa tietokantaan ja tietojen ylläpidossa sekä tietojen esille saamisessa on otettava huomioon:

- Tilastotuotannossa lisätään olemassa oleviin tiedostoihin säännöllisesti suuria tietomääriä, joten tämä toimenpide täytyy voida suorittaa mahdollisimman yksinkertaisesti ja tehokkaasti.

- Tietojen tarkistusta ja korjausta varten tarvitaan eri tyyppisille aineistoille erityisohjelmistoa.

- Varsinaista tietojen päivitystä tarvitaan, kun kyseessä ovat rekisterijärjestelmät, jotkut makrotason tietokannat sekä tietokannat, jotka sisältävät korjaamatonta tietoa.

- On tyypillistä, että tietojen haku tietokannasta halutaan tapahtuvaksi sisällön perusteella. Avaimen perusteella tapahtuvaa suoraa hakua tarvitaan yleensä vain eri tiedostojen tietoja yhdisteltäessä. Ajankohdista, johon tietyt tiedot liittyvät, on myös usein valintakriteerinä.

- Tilastolliseen tietokantaan kohdistuva tietopyyntö edellyttää yleensä suuren tietojoukon läpikäynnin.

- Koska etenkin mikrotason aineistot ovat hyvin suuria, tarvitaan erilaisia tiedon tiivistämismenetelmiä. On kuitenkin huomattava, että näiden menetelmien käyttö yleensä lisää käsittelyaikaa. Mahdollisuudet talletusrakenteen optimointiin sopivan muistitarpeen ja käsittelyyn kuluvan ajan välisen suhteen löytämiseksi onkin erityisen tärkeitä.

- Tietokannan varmistukselle ja suojaukselle asettaa lainsäädäntö erityisvaatimuksia. Tiedot on talletettava tavalla, joka suojaa ne luvattomalta käytöltä, ja toisaalta on luvallisen käytön tapahduttava siten, että tietoja on mahdotonta yhdistää yksityiseen tilastoyksikköön.

Tietohakemistojärjestelmä

Tiedon kuvauksen hallinta muodostaa perustaavaa laatua olevan ongelman, jonka ratkaisemiseksi tietohakemistojärjestelmät on tarkoitettu. Tietohakemistojärjestelmää tarvitaan tietosysteemien valmistamisessa, käytössä ja ylläpidossa (kts. /5/ luku 3.2).

On voitava saada selville, mitä tietoja kerätään tai tullaan keräämään sekä mitä tietoja tilastoaineistot sisältävät ja kuinka tiedot saadaan käyttöön. On pystyttävä myös kertomaan, mitkä ovat tietojen tarkat merkitykset. Tarvitaan ylläpidettävä sanakirja, jossa määritellään käsitteet ja niiden väliset suhteet sekä annetaan käsitteille yksikäsitteiset nimet. Edelleen on voitava tallettaa tilastotuotannossa käytettävät määritelmät ja luokitukset. On myös kyettävä esittämään loogisten arvojen ja fyysisten koodien väliset vastaavuudet.

Tietohakemistojärjestelmän avulla on voitava tallettaa myös moduleja, ohjelmia, osasysteemejä ja systeemejä koskevat tiedot.

On myös tärkeätä, että käytettävissä on tietojen käyttöä koskevia tietoja, jotta voidaan tarkkailla, onko tiedot talletettu käytön kannalta tehokkaasti.

Siirryttäessä tietokantapohjaisiin tietosysteemeihin huomataan tietosysteemien kuvausten ja tietohakemistojärjestelmän välttämättömyys. Kun tietokantaan yhdistetään useiden sovellustusten tiedot, on tiedoille oltava yhteiset määrittelyt ja ne on talletettava. Tietokannan eritasoisten rakenteiden (tietokannan käsiterakenne, tietorakenne ja talletusrakenne) hallitseminen ja kuvaaminen manuaalisesti on vähänkään monimutkaisemmissa tapauksissa mahdotonta.

Tietoriippumattomuuden toteuttamisessa on tietohakemistojärjestelmä tärkeä apuväline. Tietojen kuvaukset on pystyttävä erottamaan ohjelmista ja kuvaamaan tiedot siten, kuin jokin ohjelma ne haluaa, ja toisaalta siten, kuin ne talletetaan. Tiedon talletusrakennetta on voitava muuttaa ilman, että se vaikuttaa tietoa tar-

vitseviin ohjelmiin ja toisaalta on ohjelman käsitystä tietokannan tiedoista voitava muuttaa ilman, että se vaikuttaa muihin samaa tietokantaa käyttäviin ohjelmiin.

Käyttäjäliitännät

Tietokantamenetelmien avulla voidaan tilastotuotannossa pyrkiä saattamaan tiedot nopeasti ja joustavasti tiedontarvitsijoiden käyttöön, mutta niiden avulla tulisi voida helpottaa myös itse tietojen käsittelyä tuotannon eri vaiheissa.

Tiedonhallintajärjestelmä voi olla suunniteltu lähinnä tietojen tarvitsijoiden ei-ohjelmoivaan käyttöön tai pitäen sovellutusohjelmointia tärkeimpänä käyttömuotona. Jotkut järjestelmät pyrkivät tyydyttämään molempia käyttötarpeita.

Varsinaisten tiedontarvitsijoiden käyttöön suunniteltuihin tiedonhallintajärjestelmiin on yleensä hankalaa liittää uusia toimintoja. Toisaalta on sellaisten sovellutusten toteuttaminen helppoa, joissa tarvitaan tiedonhallintajärjestelmään jo sisältyviä toimintoja.

Edelleen voidaan tiedonhallintaohjelmisto suunnitella siten, että se palvelee lähinnä tiettyä vaihetta tilastotuotannossa. Tällöin liitetään tiedonhallintajärjestelmän kiiteään yhteyteen ohjelmia, joiden avulla tämän tuotantovaiheen tietojenkäsittelytehtäviä voidaan suorittaa.

Tilastotuotannon eri vaiheet asettavat erilaisia vaatimuksia, joita siis eri tiedonhallintajärjestelmät voivat erilaisessa määrin pyrkiä tyydyttämään. Esimerkiksi Ruotsin tilastovirastossa on kehitteillä useita eri lähtökohdat omaavia järjestelmiä, koska ei vielä pystytä kokemuksen perusteella sanomaan, mitkä olisivat tärkeimmät ominaisuudet, jotka tilastotuotantoon soveltuvalla tiedonhallintajärjestelmällä tulisi olla (kts. /2/).

Joka tapauksessa olisi nykyistä useampien henkilöiden voitava itse käyttää tietokoneita hyväkseen työskennellessään. Tarvitaan tehtävänläheisiä kieliä, joiden syntaksi on mahdollisimman yksinkertainen. Kehitettävien kielten olisi sovelluttava keskustelunomaiseen käyttöön.

Tarvittavat välineet voidaan ryhmitellä käyttäjien mukaan seuraavasti (kts. /1/ s. 36 ja /3/ s. 18):

- a) Sovellutusohjelmoijat tarvitsevat välineitä helpottamaan tietojen saantia ja käsittelyä. Tietokannan käsittely voidaan toteuttaa siten, että tavalliseen ohjelmointikieleen (isäntäkieleen) lisätään tietokannan käsittelyn mahdollistavia komentoja. Joissain järjestelmissä on itsenäinen tiedonkäsittelykieli (omavaraiskieli).
- b) Tietojen varsinaisille käyttäjille sekä tilastohenkilöstölle olisi luotava mahdollisuudet tietokantajärjestelmien ei-ohjelmoivaan käyttöön. Tarvitaan kieliä kyselyjen, taulukointien, graafisen tulostuksen, tilastollisten analyysien, aikasarjojen käsittelyn sekä tietojen kuvaamisen ja muuntamisen suorittamiseen.
- c) Tiedonhoitaja tarvitsee sopivia apuvälineitä tietokannan uudelleen organisointia, varmistusta, suojausta, vedostusta ym. varten.

5 Tulevaisuuden näkymät

Tulevaisuudessa olisi pystyttävä nykyistä paremmin huolehtimaan siitä, että tilastoaineistot ovat tarvittaessa todella käytettävissä. On pystyttävä luomaan tuotantojärjestelmä, jossa käyttäjät saavat haluamansa tiedon kyllin nopeasti. Edelleen tarvitaan tietoja, joiden avulla käyttäjä saa selville, mitä tietoja on olemassa, minkälainen tietojen laatu sekä kuinka ne saadaan käyttöön.

On odotettavissa, että tuotantoprosessissa tulee tapahtumaan muutoksia, joiden seurauksena tiedon keruuseen ja tallettamiseen liittyvät toiminnot eivät enää liity niin kiinteästi kuin nykyisin tiedon tulostusvaiheeseen. Tiedon keruu tulee olemaan säännöllistä, kun taas tulostusvaiheessa on odotettavissa tiettyä erityistarkoitusta varten tarpeen mukaan laadittavan tulostuksen lisääntyvän.

Tietojen keruuta suunniteltaessa, joudutaan kiinnittämään erityistä huomiota oletettavaan yhä kasvavaan tiedonantajien haluttomuuteen antaa tietoja sekä siihen, että ulkopuoliset tulevat kiinnittämään yhä suurempaa huomiota tietojen salaisuuskysymyksiin. On odotettavissa alunperin muita tarkoituksia varten laadittujen aineistojen yhä lisääntyvä käyttö tilastollisiin tarkoituksiin.

Tarkastellaan seuraavassa eräiden odotettavissa olevien teknologisten ja menetelmällisten kehityssuuntien mahdollisia vaikutuksia tilastotuotantoon ja erityisesti tiedonhallintaan (kts. /3/ luku 6 ja /4/).

- On mahdollista, että osa tiedonhallintaan liittyvistä toiminnoista toteutetaan itse laitteistojen tai käyttöjärjestelmien avulla.
- On odotettavissa, että tulevaisuudessa on saatavissa muistilaitteita, joilla on nykyisiä laitteita huomattavasti suurempi kapasiteetti. Koska tietojen saanti näistä uusista muisteista tulee lisäksi olemaan nopeata

niin paranevat yksikkötason laajojen aineistojen talletusmahdollisuudet ratkaisevasti. Tilastotuotannon kannalta erittäin käyttökelpoisia olisivat laitteistot, joissa kohtuullisin kustannuksin voitaisiin säilyttää suuria aineistoja ja joista sisällön perusteella tapahtuva tietojen haku olisi mahdollista.

Tietoliikenneyhteyksien kehittyessä on odotettavissa, että tietojenkäsittely hajautuu. Tällöin voivat tilastolliset tietokannat olla ositettuja (distributed), jolloin useat maantieteellisesti erillään sijaitsevat osat muodostavat loogisen kokonaisuuden, jonka osien välinen tietojen siirto tapahtuu tiedonsiirtoverkon välityksellä (kts. /6/ luku 4.5). Tiedot voidaan tällöin sijoittaa mahdollisimman lähelle niiden hyväksikäyttöpaikkaa siten, että talletuksesta, käsittelystä ja siirrosta aiheutuvat kustannukset ovat mahdollisimman pienet. Nykyiset tiedonhallintajärjestelmät eivät vielä pysty ositetun tietokannan hallintaan, mutta näiden periaatteiden soveltaminen käytäntöön tulee varmasti ajankohtaiseksi tietokoneverkkojen yleistyessä.

Tiettyä erityistarkoitusta varten suunniteltuja pientietokoneita tullaan ilmeisesti yhä enemmän käyttämään erilaisten toimintojen suorittamisessa. Pientietokoneita on mahdollista käyttää ositetun tietokannan eri osien tiedonhallinnassa, aineistojen tarkistuksessa, etäiskäsittelyn päätelaitteena ym.

Tilastotuotannolle on tyypillistä erittäin työläs ja aikaa vievä syöttövaihe, joka nykyisin vaatii paljon manuaalilityötä. Tietoliikenneyhteyksien ja päätteiden kehittyessä on odotettavissa, että tietojen saattaminen konekieliseen muotoon sekä niiden tarkistus ja korjaus tulee hajautumaan. Edelleen tullaan tiedot yhä enemmän saamaan täysin automaattisesti osana sitä toimintaa, jonka yhteydessä tiedot syntyvät tai sitten kerätään tietotavalla, jolla tarvittavien manuaalitoimenpiteiden määrä on mahdollisimman pieni (optinen luku). Myös

tarkistus- ja korjausvaihetta on ilmeisesti mahdollista nopeuttaa ja suorittaa se pitkälle koneellisesti erilaisten hakemistojen, koodistojen ja rekisterien avulla. Tallettamalla tiedot tietokantaan jo ennen lopullista tarkistamista luodaan mahdollisuudet erittäin tuoreiden ennakkollisten tietojen saantiin.

Nykyisten tiedonhallintajärjestelmien avulla saavutettava tietoriippumattomuus vaihtelee. Täydellistä tietojen ja niitä käyttävien sovellutusohjelmien riippumattomuutta ei ole mahdollista saavuttaa, mutta yleisesti ottaen on sitä nykyisestä mahdollista huomattavasti lisätä (kts. /1/ luku 5.7).

Tietojen loogisen rakenteen muuttamiseen on tarvetta, kun tietokantaan halutaan liittää täysin uusia tietoja tai luoda olemassaolevien tietojen välille uusia yhteyksiä. Mikäli muutokset eivät poista tietokannasta mitään sellaista tietoa, jota sovellutusohjelmat käyttävät, on tietokannan uudelleenstrukturoida mahdollista toteuttaa tarvitsematta muuttaa sovellutusohjelmia.

Tietojen fyysisen organisointitavan muuttamiseen antaa aiheutta se, kun tietokannan käytöstä kerättyjen tilastotietojen perusteella havaitaan, että esim. tietokannan talletusvälineen tai talletusrakenteen vaihtamisella voitaisiin ohjelmien suorituskykyä lisätä. On odotettavissa, että tiedonhallintajärjestelmät kykenevät tulevaisuudessa automaattisesti (tai kenties tiedonhoitajan avustamana) suorittamaan tietokannan uudelleenorganisoinnin, kun käyttöä koskevat tilastotiedot sitä edellyttävät.

Tietoriippumattomuuden toteuttaminen edellyttää tehokkaan tietohakemistojärjestelmän olemassaoloa. Tietokantojen kuvaukset talletetaan kuvaustietokantaan. Sekä varsinaisten tietokantojen että kuvaustietokannan käsittelyyn tarvitaan tiedonhallintajärjestelmä. Ongelmana on tietokantojen hallinnan tarkoituksenmukainen yhteensovittaminen (kts. /1/ s. 27 ja /5/ s. 19).

Nykyään eivät tiedonhallintajärjestelmät yleensä pysty automaattisesti käyttämään hyväksi tietohakemistojärjestelmän tallettamia kuvauksia, vaan kuvausten välittäminen tiedonhallintajärjestelmälle vaatii manuaalisia välivaiheita. Kehitys tulee kuitenkin johtamaan siihen, että tietokantajärjestelmien toiminta perustuu siihen, että tietohakemistojärjestelmä tallettaa ja ylläpitää tiedon kuvauksia sekä tuottaa tietoja koskevia metatietoja olipa niiden hyväksikäyttäjä ihminen tai ohjelmisto (tiedonhallintajärjestelmä).

Tietohakemistojärjestelmien käyttömahdollisuudet tilastotuotannossa tuntuvat huomattavilta. Niiden avulla voidaan tuottaa tietosysteemien suunnittelua, toteuttamista ja hyväksikäyttöä varten tarvittavaa tietoa tilastoaineistoista ja niiden suhteesta ympäristöönsä. Niitä voidaan käyttää apuna tietoihin liittyvässä standardointityössä sekä erillisten tietokantojen yhtenäistämässä.

Kuvaamisen määrittäminen ei ole kuitenkaan mielekästä, ennen kuin on selvä käsitys siitä, mitä aiotaan kuvata. On muodostettava tilastotuotannon käsitteistö, jossa määritellään, mitä osia tilastollisissa tietosysteemeissä on, mitä ominaisuuksia näillä osilla on ja miten ne suhtautuvat toisiinsa.

Tietokantojen yhteydessä muodostaa niiden hallinta ja toiminnan valvonta suuren ongelman. On pystyttävä turvaamaan talletettujen tietojen oikeellisuus sekä estämään tietojen tuhoutuminen ja luvaton hyväksikäyttö.

Tietokannan suojauksella (privacy, security) tarkoitetaan niitä toimenpiteitä, joilla pyritään estämään tietokannan luvaton käyttö (kts. /1/ luku 4.2.2). Koska tietokannalla saattaa olla suuri joukko erilaisia käyttäjiä, tarvitaan usealla eri tasolla tapahtuvaa suojausta. Suojauksen kohteen lisäksi suojausta määriteltäessä on ilmoitettava ne toimenpiteet, joilta suojauksen kohde halutaan suojata. Eritasoisiin suojauksen kohteisiin liittyy erilaisia toimenpiteitä, jotka suojauksella voidaan estää. Varsinaisen käsittelytiedon lisäksi voi esiintyä tarvetta suojata tietokannan kuvaukset esim. muuttamista vastaan.

Varmistus (integrity) on yhteisnimitys niille toimenpiteille, joilla pyritään turvaamaan tietokannan virheettömyys ja käytettävyys (kts. /1/ luku 4.2.3). Varmistus on suojausta esim. laitteiston, tiedonhallintaohjelmiston, tietoliikenneohjelmiston ja sovellutusohjelmien virheellistä toimintaa sekä ihmisen tahallisia tai tahattomia toimenpiteitä vastaan.

Tietojen suojausta ei nykyisin menetelmin pystytä hoitamaan tilastotuotannon kannalta alkuunkaan tyydyttävästi. Tietojen käyttöä rajoittavat useissa tapauksissa tiukat salassapitomääräykset, joita on pystyttävä noudattamaan. Ennen kuin voidaan ajatella toteutettavaksi järjestelmä, jossa tietojen tarvisijat voivat itsenäisesti käyttää mikrotason aineistoja saadakseen haluamansa tilastotiedot tai suorittaakseen tilastollisia analyyseja, on huomattavaa kehitystä tapahduttava verrattuna niihin mahdollisuuksiin, joita nykyiset tiedon-

hallintajärjestelmät tiedon suojaukseen tarjoavat. Eikä yksin tiedonhallintajärjestelmien kehittäminen riitä, vaan työtä on tehtävä myös täysin teoreettisen menetelmäkehittelyn, laitteistojen, käyttöjärjestelmien sekä sovellutusohjelmistojen piirissä.

6 Siirtyminen tietokantapohjaiseen tuotantotapaan

Sovellettaessa tietokantamenetelmiä aiheuttavat seuraavat tekijät kustannuksia:

- tietohakemiston kerääminen,
- organisatoristen muutosten toteuttaminen,
- tietokannan suunnittelu ja toteuttaminen,
- tietokantaan pohjautuvien tietosysteemi~~en~~ suunnittelu ja toteuttaminen,
- standardointi,
- mahdollisesti tarvittavat uudet laitteistoresurssit,
- lisähenkilökunta,
- tarvittava konsultointiapu,
- tiedonhallintajärjestelmän osto, vuokraus tai toteuttaminen itse,
- koulutus,
- tietojen keruu,
- vanhojen systeemi~~en~~ konvertointi,
- käyttökustannukset.

Tietokantamenetelmien käyttöönoton kannattavuus riippuu erittäin vaikeasti mitattavissa olevista asioista. On lähes mahdotonta kvantitatiivisesti arvioida, mitä hyötyä (kustannussäästöjä, parempaa tietoa käyttäjille ym.) niiden soveltamisesta olisi (kts. /3/ luku 3.3 ja /6/ luku 3).

Hyötyjä arvioitaessa on tärkeätä selvittää, voidaanko samoihin parannuksiin päästä mahdollisesti muilla keinoin. Nykyisiin tilastokohtaisiin tietosysteemeihin liittyviä haittoja on ilmeisesti mahdollista paljonkin pienentää ilman varsinaisten tietokantamenetelmien käyttöä (Kts. /6/ luku 3).

(1) On pyrittävä tehostamaan tietojen käytön ohjausta ja valvontaa keskittämällä tähän liittyvien tehtävien suorittaminen. Tilastovirastossa on tietojen hoitoon liittyvä tehtäväkenttä hyvin laaja. Tarvitaan tilastojärjestelmien, tilastojen ja tilastollisessa kuvauksessa käytettävien menetelmien perusteellinen tuntemus sekä toisaalta systeemin suunnittelun ja ohjelmoinnin hallinta. Edelleen on pystyttävä kommunikoimaan tietojen hyväksikäyttäjien kanssa. Tämän vuoksi ei ole mahdollista, että yhdellä henkilöllä olisi tarvittavat tiedot ja taidot, vaan tietojen hoitoon liittyvien tehtävien suorittamiseen tarvitaan erityinen tiedonhallintaryhmä.

Tilastollisen kuvauksen perustana on käsitejärjestelmä, jonka avulla hahmotetaan tilastoinnin kohteet. Aineistojen hankintaa ja käsittelyä varten tarvitaan lisäksi tarkat määritelmät ja luokitukset. Näiden käsitteiden, määritelmien ja luokitusten yhtenäistämistyössä sekä tilastotoimen tietosisällöllisessä koordinoinnissa tarvitaan tilastollista asiantuntemusta. Tietojen koordinaattorin (data co-ordinator, kts. /3/ s. 7) tehtäviin kuuluu yleinen tilastotietojen käytön ohjaus ja valvonta. On pidettävä yhteyttä tietojen käyttäjiin ja avustettava käyttäjiä tiedottamalla ja selvittämällä käyttäjien tietosisältöön liittyviä ongelmia. Tehtäviin kuuluu edelleen tietokantojen tietojen loogisen rakenteen suunnittelu ja kuvaaminen sekä tietoihin liittyvien standardien luominen ja standardien noudattamisen valvonta.

Tiedonhoitajan (data administrator) tehtävänä on vastata tiedonhallinnan teknisestä puolesta. Tehtäviin kuuluu tietojen loogisten rakenteiden muuttaminen sopiviksi talletusrakenteiksi ja näiden kuvaaminen sekä mahdollisesti tiedonhallintaohjelmiston suunnittelu ja toteuttaminen. On huomattava, että tiedonhoitajan tehtäviin kuuluu myös toimiminen tiedostojen suunnittelun konsultina eri suunnitteluprojekteissa. Tiedonhoitajan tehtäviin kuuluvat tietokannalle suoritettavat toiminnot ja hän vastaa myös tietojen suojauksen ja varmistuksen suunnittelusta ja toteuttamisesta (vrt. /1/ luku 4.5).

Ainakaan aluksi ei ehkä ole tarkoituksenmukaista, että kaikki tilastupuolen henkilöt voisivat itse käyttää hyväkseen tietokoneita. Sen sijaan voitaisiin tilastupuolelle nimetä henkilöitä, jotka koulutetaan hakemaan tietoja tietokannasta.

Keskitetty tietojen hoito on välttämätöntä tietokantaympäristössä, mutta aivan ilmeisesti hyödyllistä myös erillisiin tilastokohtaisiin tietosysteemeihin perustuvassa järjestelmässä. Monien seuraavaksi mainittavien toimenpiteiden onnistuminen edellyttää tiedonhallinnasta vastaavien henkilöiden olemassaoloa.

(2) Jotta tilastokohtaisten tietosysteemien tietoja voitaisiin tarvittaessa yhdistellä, on käsitteiden, määritelmien ja luokitusten yhtenäistäminen välttämätöntä. Tietoaineistojen yhdisteltävyyttä parantaa edelleen standardointi, joka koskee nimiä, esitysmuotoja, kooditusta, suunnittelun ja ohjelmoinnin periaatteita ja menetelmiä, suojausta, varmistusta, dokumentointia jne.

Standardointitoimenpiteisiin kuuluu mm.

- standardien suunnittelu,
- standardeihin liittyvä koulutus ja niistä tiedottaminen,
- standardien noudattamisen valvonta,
- standardien seuranta ja kehittäminen.

Tietokantamenetelmien menestyksellinen soveltaminen edellyttää erilaisten standardien asettamista ja noudattamista. Standardoinnin tulisi edeltää tietokantajärjestelmän suunnittelua ja toteuttamista. Tietokantoihin liittyvät mekaaniset apuvälineet: tiedonhallintajärjestelmä ja tietohakemistojärjestelmä helpottavat jo sinänsä tiedonhallintaan liittyvää standardointia pakottaessaan käyttäjän noudattamaan tiettyjä periaatteita ja rajoittaessaan mahdollisten vaihtoehtojen lukumäärää, mutta pelkkä niiden käyttöönotto ei vielä merkitse sitä, että tarvittavasta standardoinnista olisi riittävässä määrin huolehdittu (/6/ luku 5).

(3) Tavallisten tiedostojen rakenne kuvataan yleensä nykyäänkin niitä käsittelevissä ohjelmissa. Tällöin joudutaan aina, kun tiedoston rakenteeseen tulee muutos, muuttamaan kaikkia kyseistä tiedostoa tarvitsevia sovellutusohjelmia. Käytettäessä tiedonhallintajärjestelmiä on tietojen kuvaukset sen sijaan erotettu ohjelmista.

Tietoriippumattomuutta voidaan kuitenkin lisätä ilman varsinaisten tiedonhallintajärjestelmien käyttöönottoa pyrkimällä mahdollisimman pitkälle erottamaan käsiteltävien tietojen kuvaukset ohjelmista.

(4) Jotta tietoriippumattomuutta voitaisiin lisätä tarvitaan apuvälineitä tiedon kuvausten tallettamisessa. Tilastollisten tietosysteemien tietosisältö on niin laaja ja monipuolinen, ettei tietojen kuvauksen tyydyttävää hallintaa voida ajatella saavutettavan ilman automaattisten välineiden käyttämistä. Tarvitaan tehokas tietohakemistojärjestelmä.

Tietohakemistojärjestelmän merkitys on riippumaton varsinaisten tietokantamenetelmien soveltamisesta. On kuitenkin huomattava, että tietohakemistojärjestelmän luominen vaatii runsaasti henkilö- ja koneresursseja ja sen toteuttaminen vastaa sinänsä laajaa tietokantaprojektia.

(5) Tiedon redundanssia on pyrittävä vähentämään. Tämä edellyttää kuitenkin, että tietoihin kohdistuva ohjaus ja valvonta keskitetään. Edelleen jotta pystyttäisiin selvittämään systeemien ylimääräinen tiedon talletus on näissä käytettyjen käsitteiden, määritelmien, luokitusten, tietoalkion merkintöjen ym. oltava vertailtavia.

(6) Tietosysteemien ylläpidettävyyttä voidaan lisätä supistamalla tietosysteemin rakentamisen aikana tehtyjen ratkaisujen vaikutusta. Yksi tapa on eristää toiminta ja toiminnan ohjaus. Ohjelmiin sovellettuna se tarkoittaa yleisohjelmien luomista. Ylläpitotilanteessa ohjelma säilyy koskemattomana, vain parametreja (toiminnan kuvausta) muutetaan (kts. /5/ luku 3.2.4).

Samoja tietoja eri sovellutuksissa käsittelevät ohjelmat sisältävät usein samanlaisia osia. Sopivien yleisohjelmien ja -modulien avulla on tätä päällekkäisyyttä mahdollista vähentää.

(7) Valmistautumisessa mahdolliseen tietokantamenetelmien käyttöönottoon on ensiarvoisen tärkeitä riittävän pitkä ja monipuolinen henkilökunnan koulutus.

Valmistautumisen merkitystä aliarvioidaan yleensä siksi, koska ajatellaan, että tietokannan käyttöönotto on ainoastaan tekninen muutos. Tiedonhallintajärjestelmä onkin ennen kaikkea tekninen apuväline, mutta tietokantojen rakentaminen ja tiedonhallintajärjestelmän tehokas hyväksikäyttö edellyttää tietosysteemien rakentamista uudelle pohjalle (kts. /6/ luku 1).

Lähdeluettelo

- /1/ Tietokanta ja tiedonhallintajärjestelmä, Valtion tietokonekeskus, koulutus, Markku Nokso-Koivisto, 1976-08-13
- /2/ Databaser och databashanterare vid SCB, SCB, P/DBM, L. Olsson, alc, 1977-05-12
- /3/ Report on the use and future need for data base management in national statistical services, CES/WP.9/150, 1977-01-26
- /4/ Anticipated computer hardware and software developments and their relevance to statistical offices, Statistics Canada, 1976, CES/SEM.8/8/Add., 1977-01-12
- /5/ Tietohakemistosysteemi, Eripainos, DIFO-vuosikirja 5/1976, luku 3, Heikki Kämäräinen
- /6/ Valmistautuminen tietokannan ja tietohallintajärjestelmän käyttöönottoon, Valtion tietokonekeskus, koulutus, Markku Nokso-Koivisto, 1976-09-01