

# Tietokanta täydentää ETLAn tietotarjontaa\*

JUKKA LASSILA

Elinkeinoelämän Tutkimuslaitos (ETLA) on koonnut kotimaisista ja kansainvälisistä taloustilastoista aikasarjatietokannan. Tietoja käytetään ETLAn tutkimus- ja ennustetoiminnassa, ja niitä päivitetään jatkuvasti. Tietokanta on myös ulkopuolisten käytössä maksua vastaan.

Useimmista muista aikasarjatietokannoista ETLAn kanta poikkeaa siinä, että taloudellisia ilmiöitä koskevista tilastosarjoista säilytetään myös vanhat ennakkotiedot. Täten käyttäjä voi tutkia mm. ennakkotietojen tarkentumista ajan kuluessa. Tietokannassa on myös tärkeimpien kotimaisten ennustajien sekä OECD:n suhdanne-ennusteet.

Tietokantaan liittyy haku- ja muokkausohjelmisto, jonka avulla käyttäjä voi muuntaa ja yhdistellä eri lähteistä peräisin olevia tietoja. Aineisto voidaan sen jälkeen siirtää käyttäjän omalle mikrolle jatkoanalyysia varten. Tilasto- ja ennustenumeroiden lisäksi taloustietoja voidaan siirtää myös valmiiden kuvioesitysten muodossa.

## 1. Tietokannan sisältö

Tietokannan keskeisenä lähtökohtana oli ennustetoiminta, jota ETLAssa on tehty 1970-luvun alkuvuosista alkaen. Tietopankin rakennustyön alkuvaiheessa vuonna 1985 arvioitiin, että Suhdanteessa julkaistavien ennusteiden laadinnan yhteydessä seurattiin säännöllisesti 2000 aikasarjaa. Näiden ajateltiin muodostavan tietokannan perustan. Kun ryhmäkohtaisesta ja osittain tutkijakohtaisesta tiedonhallinnasta siirryttiin koko laitoksen laajuiseen keskitettyyn tiedon ylläpitoon, datan määrä kuitenkin kasvoi. Mukaan otettiin

sarjoja, joiden seurantaan ei aiemmin ollut riittänyt resursseja. Hyppäys datan määrässä tapahtui, kun ETLA solmi sopimuksen OECD:n tietojen edelleen myynnistä tietokannan kautta. Aikasarjojen määrä onkin kasvanut alun perin arvioidusta yli kymmenkertaiseksi; tammikuussa 1991 lukumäärä on suunnilleen 25 000. Kasvu ei ole ollut mikään itsetarkoitus, syynä on ollut datan lisäämisen aiheuttamien rajakustannusten huomattava pieneneminen. Elektronisesta tiedonsiirrosta johtuen rajakustannus on usein negatiivinen: on helpompaa ottaa tietty datamäärä kokonaan kuin karsia siitä osia pois.

Talouden ennustamisessa ja mallintamisessa kansantalouden tilinpito on tärkein tilasto. Kansantalouden vuositilinpito ja neljännesvuositilinpito ovatkin koko julkaistussa laajuudessaan ETLAn tietokannassa. Muut kotimaiset tilastot ovat kannassa vain valikoituilta osin. Ne päivitetään julkaistuihin tilastoista manuaalisesti. Poikkeuksena mainittakoon Teollisuuden Keskusliiton suhdannebarometri, joka siirretään disketeillä kantaan ja on käytettävissä samaan aikaan kuin barometrin tulokset julkistetaan.

OECD:n toimittamaa aineistoa ETLAn tietokannassa ovat mm. Annual National Accounts, Quarterly National Accounts ja Main Economic Indicators, joka sisältää keskeistä kuukausittaista tietoa OECD:n jäsenmaiden talouksista. Economic Outlook sisältää OECD:n ennusteet eri maista tasomuotoisiksi sarjoiksi muutettuina. Näiden lisäksi tietokannassa ovat Financial Statistics Monthly sekä Main Science and Technology Indicators. Kaikki nämä tilastot ovat siinä laajuudessa kuin OECD niitä myy disketeillä.

Helmikuussa 1991 ETLA aloitti yhteistyön Telerate-järjestelmän kanssa, josta saadaan tietokantaan päivittäistä dataa rahoitusmark-

\* Järjestelmäkuvaus

kinoilta. ETLAlla on myös ollut yhteistyötä Ruotsin työnantajaliiton SAF:n kanssa kansainvälisten palkkatietojen osalta.

## 2. Ennakkotiedot ja ennusteet tutkimuksen kohteena

Muutokset ja korjaukset ovat monien taloustilastojen ominaispiirre. Silti tutkimustietoa näiden korjausten suuruudesta ja merkittävyydestä on hyvin vähän. Luultavasti datan keräämiskustannukset verrattuna siihen, että asiaa ei ole pidetty kovin tärkeänä, ovat estäneet useampien tutkimusten teon. Tilanne muuttuu, jos data on saatavilla olennaisesti aiempaa pienemmin kustannuksin.

Ennakkotietoja koskevaa kysymyksenasettelua voidaan havainnollistaa kuvion 1 avulla. Oletetaan, että tuotantoa koskeva tilasto ilmestyy kerran vuodessa, ja kussakin tilastossa esitetään ennakkotieto edelliseltä vuodelta, tarkistettu ennakkotieto sitä edelliseltä vuodelta ja lopulliset luvut aiemmasta tuotannosta. Kuviossa vaaka-akseli kertoo tilaston julkaisuvuoden ja pystyakseli havainnon vuosiluvun.

Kuvioon on merkitty neljä nuolta, jotka kuvaavat neljää aineistolle asetettavaa kysymystä. Ensimmäinen nuoli viittaa tavallisimpaan aikasarjatietokannoille esitettävään kysymykseen: mitä asiasta, tässä siis tuotannosta, tie-

detään nyt? Toinen nuoli kuvaa tutkimuskysymystä: miten yhtä vuotta koskevat tuotantotiedot ovat tarkentuneet ajan kuluessa? Kolmannen nuolen kuvaama kysymys on hyvin lähellä kysymystä kaksi, mutta nyt aikasarjamielessä: mikä on ollut tarkistettujen ennakkotietojen suhde lopullisiin lukuihin? Neljäs kysymys voi tulla eteen, kun yritetään rekonstruoida jokin aiempi päätöksentekotilanne, esimerkiksi talouspolitiikkaa koskevan tutkimuksen yhteydessä: mitä tiedettiin tuotannosta vuonna 1989?

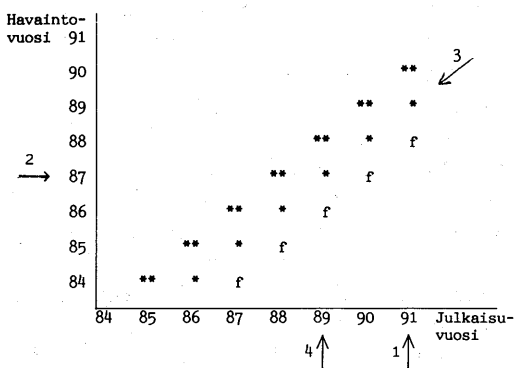
Tiedot voidaan suhteellisen yksinkertaisesti tallentaa siten, että edellä esitettyihin kysymyksiin voidaan myöhemmin aineiston puolesta vastata. Havaintojen lisäksi tallennetaan se, milloin ja mistä lähteestä tieto on tullut. Nämä milloin ja mistä -tiedot ovat ainoat asian vaatimat lisätallennukset. Ennakkotiedoitan tietokantoihin yleensä heti talletetaan; ne vain poistetaan sieltä tarkistettujen tietojen saapuessa.<sup>1</sup>

Päätös vanhojen ennakkotietojen säilyttämisestä on sikäli tärkeä, että se suurelta osin määrää tietokannan rakenteen. ETLAn tietokannassa tiedot on jaettu tauluihin alkuperäisten tilastolähteitten mukaisesti. Yhdestä tilastosta peräisin oleva data voi olla useammas- sa taulussa, mutta yhden tietokantataulun päivityslähteenä on vain yksi lähde-tilasto (lähäs aina, muutamia poikkeuksia on). Täten tieto siitä, milloin ja mistä kukin havainto on tullut, koskee aina koko havaintoriviä taulussa.

Ennakkotietojen lisäksi on muutakin samantyyppistä dataa, jonka tallettaminen sujuu edellä kuvatuilla periaatteilla. Tällaista aineistoa ovat esimerkiksi ennusteet, joita moni laitos tekee useampia vuodessa kutakin vuotta kohden. Myös ketjutettu data ja ajassa aggregoidut havainnot, esimerkiksi kuu- kausihavainnoista lasketut neljännesvuositiedot, kuuluvat tähän ryhmään.

<sup>1</sup> Ennakkotietojen ja ennusteiden tallentamista on tarkemmin käsitelty artikkeleissa *Lassila (1990)* ja *Harni et al (1991)*. Ennakkotietojen käyttöä ennustamisessa ja mallien rakentamisessa on pohdittu mm. julkaisuissa *Boucelham ja Teräsvirta (1989)* sekä *Trivellato ja Rettore (1986)*.

Kuvio 1.



\*\* = ennakkotieto  
\* = tarkistettu ennakkotieto  
f = lopullinen tieto

### 3. Tilastojulkaisut ja ETLAn Suhdanne hakujärjestelminä

Käyttäjien ei ole aina helppoa löytää halua- maansa aineistoa tietopankeista. ETLA on yrittänyt varmentaa löytämistä rakentamalla kaksi erilaista valikkoja käytävää hakutapaa. Ensimmäinen perustuu lähdetilastoihin, toisessa tiedot on jaoteltu ETLAn Suhdanteen sisällysluettelon mukaisesti.

Kuten edellä ennakkotietojen käsittelyn yhteydessä todettiin, ETLAn tietokannalla ja tilastojulkaisuilla on tietty rakenteellinen yhtäläisyys: kukin tietokannan taulu päivittyy vain yhdestä lähdetilastosta. Täten on ollut luontevaa ja suhteellisen vaivatonta rakentaa valikkojärjestelmä, jossa tilastojulkaisuja tunteva käyttäjä löytää tiedot myös tietokannasta. Tietojen dokumentointi on myös helppoa, koska yleensä voidaan esittää tietyn taulun tilastolähde yhdellä kertaa, ovathan kaikki muuttujat peräisin samasta tilastosta. Tämän lisäksi tietokantaan on vaihtelevassa määrin koottu tietoja havaintojen laskutavoista, muuttujien määrittelystä, mahdollisista tilastovirheistä ja muusta vastaavasta. Nämäkin tiedot on jaoteltu taulukohtaisesti. Vielä tarkempia tietoja kaipaavan täytyy kääntyä alkuperäistilaston tuottajan puoleen.

Datan nopean löytämisen ja väärinymmärrysten välttämisen kannalta saattaisi olla hyvä mennä vielä pitemmälle kuin ETLAn tietokannassa on menty ja pyrkiä sellaiseen tietokantaan, joka olisi kultakin osaltaan vastaavan alkuperäistilaston mahdollisimman täydellinen näköispainos. Muuttujilla olisi samat selitykset, sama järjestys, mahdollisuuksien mukaan sama nimilyhenne kuin vastaavassa tilastossa. Alkuperäiset tilastot olisivat samalla tietokannan käsikirjoja. Oppimiskustannukset jäisivät pieniksi.

Asiaan liittyy kuitenkin pari vaikeutta. Ensinnäkin kaikissa tilastoissa ei enää ole kirjallista alkuperäislähdettä, vaan data on saatavissa vain konekielisessä muodossa. Toiseksi, eräiden tilastotuottajien on rahoitettava kasvava osa toimintaansa myyntituloilla. Osa

näistä tuloista voi tulla tuottajan omasta tietopankkitoiminnasta. Tilastokeskus onkin nostanut esille tekijänoikeuskysymykset, erityisesti tekijänoikeuslain ns. luettelosuojapykälän, turvatakseen asemaansa (ks. esimerkiksi Janhunen 1990). Julkaistut tilastotiedot ovat vapaasti käytettävissä, mutta tietopankin rakentajan on syytä varmistaa, ettei liian läheinen yhteys tilastoihin johda raastupaan. Voidaan tietenkin kysyä, toteuttaisiko julkisilla varoilla tietoja tuottava tilastoviranomainen tällä tavoin perimmäistä toiminta-ajatus- taan oikein vai käyttäisikö se monopoliase- maansa väärin.

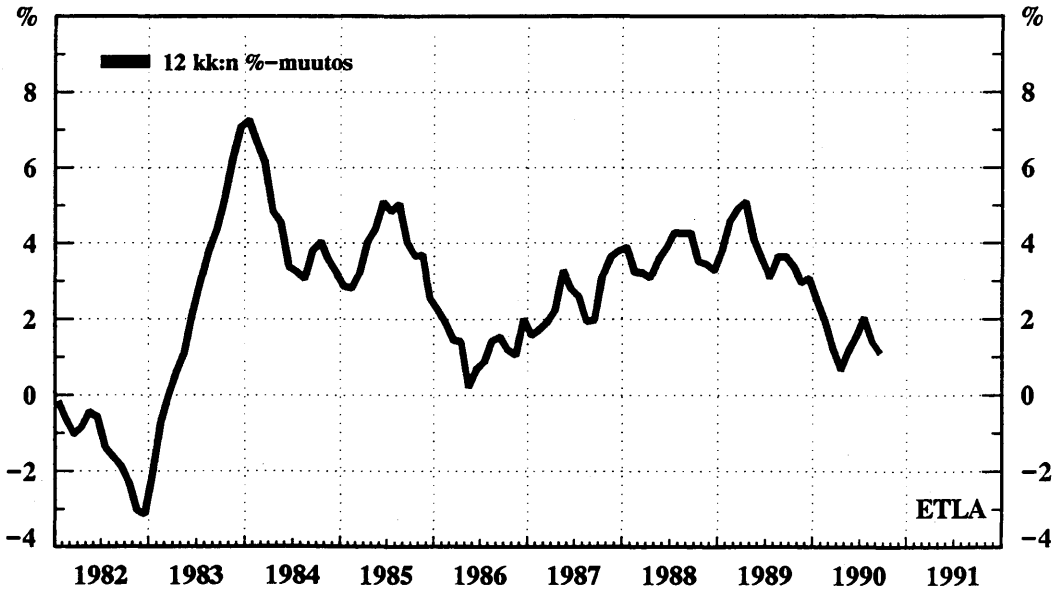
Toinenkin ETLAn tietokannan hakuvalikko palauttaa tilastotiedot muusta yhteydestä tuttuun rakenteeseen eli Suhdannejulkaisun sisällysluettelon. Muitakin hakutapoja on, esimerkiksi SQL-tyyppinen kyselykieli.

Enenevässä määrin käyttäjille laaditaan yksilöllisiä hakuvalikoita, jonne toistuvat haut muunnoksineen on koottu valmiiksi kohdiksi. Valikot voivat sisältää myös kuvion 2 kaltaisia graafisia esityksiä. Kuviotiedostot ovat kooltaan yllättävän pieniä ja niiden siirto käy helposti päinsä valintalinjojakin pitkin. Kuvion 2 esimerkit kuvaavat myös aineistoja, joiden hallinnassa tietokannasta on käyttäjille erityisen suurta hyötyä: yksinkertaisuudestaan huolimatta kuvioiden takana on dataa monesta eri tietolähteestä. Työlästä on tällaisen aineiston pito ajan tasalla, mutta itse kuvion teko on helppoa.

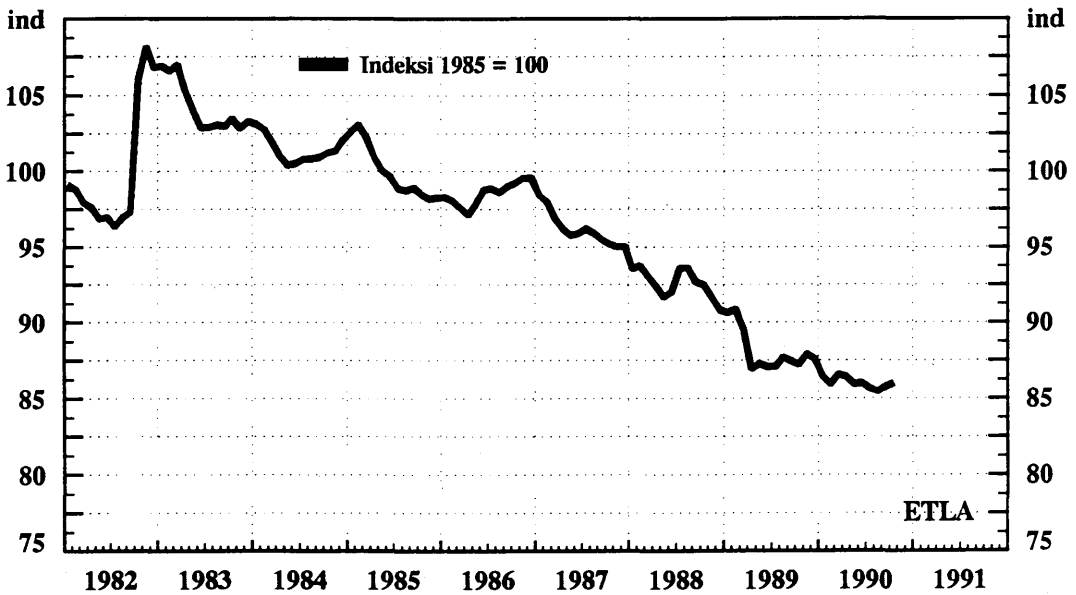
### 4. Tietopankin käyttömaksuista

Numeeristen tietokantojen käyttömaksut muodostuvat yleensä monesta osasta. Useimmiten veloitetaan vuosi- tai kuukausimaksujen lisäksi poimituista havainnoista, yhteys- ajasta, tietokoneressurssien käytöstä ja/tai taulukoista (ks. Wolski et al. 1990). Monimutkainen hinnoittelu tekee kustannusten ennakoinnin ja budjetoinnin hankalaksi. Tämä vähentää tietopankkien käyttäjämääriä, kuten mm. Järvelin (1986) toteaa.

Kuvio 2. Esimerkkejä ETLAn tietokannan yhteydessä olevasta kuviopankista.



11 OECD-maan teollisuustuotanto (kausitasoitettu 3 kk:n liukuva keskiarvo Suomen edellisen vuoden vientiosuuk-silla painotettuna, ei keskitetty)



Suhteelliset kuluttajahinnat, OECD/Suomi, valuuttakurssikorjattu (vuoden 1980 kursssein). OECD-maat painotettu Euroopan markkinaosuuksilla

ETLAn tietopankin hinnoittelussa on pyritty kustannusten ennakoitavuuteen. Tietokonesurssien käytöstä ei veloiteta erikseen, ei myöskään poimituista havainnoista. Käyttäjä maksaa vuosittaiset käyttöoikeusmaksut haluamiinsa osiin tietokannasta, ja sen lisäksi pienen maksun yhteysajan mukaan. Täten kokonaiskustannukset ovat hyvin tarkasti ennakoitavissa ja marginaalikäyttö on halpaa. Käyttöoikeusmaksut eivät ole korkeita: esimerkiksi kansainvälisen aineiston saa ETLAn kautta halvemmalla kuin tilaamalla suoraan OECD:stä. Yliopistoille ja korkeakouluille käyttömahdollisuus tarjotaan vielä tästäkin alennettuun hintaan. Tietokantayhteyden perustamisesta peritään kertamaksu, joka kattaa tietoliikenneohjelman, käsikirjat ja perusopetuksen.

Tietopankit eivät yleensä ole tarjoajilleen kultakaivoksia. Käyttäjämäärät jäävät usein pieniksi. *Thygesen (1990)* mainitsee EY:n CRONOS ja COMEXT tietokannat esimerkkeinä kannoista, joihin on investoitu paljon rahaa ja kovaa työtä, joiden hinnat on painettu keinotekoisena matalalle ja joiden käyttäjämäärät siitä huolimatta ovat jääneet niin pieniksi että toiminnan lopettaminen on edessä.

ETLAssa tietokantaa tarvitaan laitoksen omaan työhön. Kannattavuutta onkin arvioitava siltä pohjalta, tuottaako kannan tarjoaminen laitoksen ulkopuolisille niin paljon, että toiminnasta aiheutuvat lisäkustannukset tulevat katetuiksi. Tarkan vastauksen antaminen ei ole helppoa, mutta kannattavuus tässä mielessä lienee jo nyt positiivinen, ja asiakasmäärän kasvu näyttää jatkuvan.

ETLAn henkilökunnan asiantuntemus tilastoista ja niihin läheisesti liittyvistä asioista sekä valmius keskustella ovat ETLAn valtteja kilpailussa markkinaosuuksista. Erittäin tärkeitä ovat myös mahdollisuudet oheispalveluihin kuten esitelmiin, alustuksiin, kuvioesityksiin ja asiakaskohtaisesti räätälöityihin tuotteisiin. Keskeisiä ovat myös hyvät yhteistyökumppanit kuten OECD, SAF, TKL ja Telerate.

## 5. Teknisiä tietoja

ETLAn tietokanta on Hewlett-Packard 9000-sarjan minitietokoneella, jonka käyttöjärjestelmänä on unix. Kanta on rakennettu ruotsalaiselle relaatio-ohjelmistolle MIMER. Tietokantaan liittyvät hakujärjestelmät ja laskentaohjelmat on tehty ETLAssa.

Relaatio-ohjelmistolle rakennettu aikasarjatietokanta on harvinaisuus. Relaatiomallin käyttöön yleisemminkin tieteellisissä tietokannoissa liittyy hyviä ja huonoja puolia, mutta hyvät puolet ehkä dominoivat muihin kaupallisiin vaihtoehtoihin verrattuna (ks. NSF, 1990). ETLAssa relaatiokannan käyttö ei ole aiheuttanut ongelmia, ja esimerkiksi ennakkotietojen tallentaminen ja käyttö on ollut helppoa.

Aikasarjojen pituudet vaihtelevat huomattavasti. Joitakin tietoja on 1700-luvun alkupuolelta saakka. Useimmat sarjat toki alkavat 1960- ja 1970-luvuilta. Useista sarjoista on samassa taulussa eriperiodisia havaintoja, esimerkiksi neljännesvuosi- ja vuositietoja. Heikki *Vajanteen* kehittämässä *Table*-ohjelmassa onkin pantu erityistä huomiota siihen, että tällaisten sarjojen yhdistely käy käyttäjältä mahdollisimman vaivattomasti. Sarjoja valittaessa ei esimerkiksi tarvitse välittää aikaväleistä tai yksikköperiodeista. *Table* osaa rinnastaa sarjat oikein eikä johda käyttäjää harhaan. Keskeistä ohjelmassa on puuttuvien havaintojen käsittely.<sup>2</sup>

ETLAn tietokantaan ollaan yhteydessä *Reflection*-ohjelmiston avulla. Kantaan on tehty valmiit talletusmuodot esimerkiksi *Lotuksen wr1*-tiedostoiksi ja *Survotiedostoiksi*. Kantaan liittyvä grafiikka on siirrettävissä myös mikrolle ja edelleen *Canonin* lasertulos-timelle. Grafiikka on myös VideoShow-yhteensopiva ja esitettävissä televisioissa ja videotykin avulla korkeatasoisina värikuvina ja edelleen tulostettavissa VideoShow'n tukemilla tulostuslaitteilla. Tietokannan hakuohjelmista ja grafiikasta ks. *Harni et al (1991)* ja *Vajanne (1989)*.

<sup>2</sup> Tablesta enemmän ks. *Harni et al (1991)* ja *Vajanne (1990)*.

## 6. Tutkimuslaitos tietopankin isäntänä

Vaikka tilastotietojen tehokas välittäminen käyttäjille on tärkeää toimintaa, ei pelkästään sitä ole pidetty riittävänä perusteluna sille, että ETLA tarjoaa tietopankkiaan ulkopuolisten käyttöön. Myöskään toiminnan mahdollinen kannattavuus ei ole ollut ratkaiseva motiivi. ETLA on aiemmin tarjonnut eri tavoin jalostettua tietoa tutkimusten ja ennusteiden muodossa. Yritysten ja muiden tiedon käyttäjien omien tiedonhallintajärjestelmien kehittyessä ETLAlta on yhä useammin toivottu myös keskeisten perustietojen liittämistä täydentämään muuta tiedon tarjontaa. Tietopankin ylläpito lisää mahdollisuuksia palvella yksittäisiä asiakkaita räätälöidyin tuottein; erityisesti tätä toimintaa varten on pari vuotta sitten perustettu ETLATIETO Oy.

Suorakäyttöinen tietokanta on myös uusi jakelukanava jalostetulle tiedolle. Keväästä 1991 alkaen kannasta on saatavilla ajan tasalla olevia talouden seurantakuviota sekä kommentteja ajankohtaisesta tilustilanteesta. Tutkimustiedon osalta tietokantatyypin toiminnan toivotaan lisäksi auttavan taloudellisen tiedon osalta tunnetusti vaikeaa tiedon kumuloimista.

### Kirjallisuus

Boucelham, J. ja T. Teräsvirta (1989): »How to use preliminary values in forecasting the monthly in-

dex of industrial production?», *ETLA Discussion Paper* No 284.

Harni, M., J. Lassila ja H. Vajanne (1991): »Transformation and Graphics in ETLA's Economic Database System», *ETLA Discussion Paper* No. 351, Helsinki.

Janhunen, O. (1990): »Kommentar till Lars Thygesens inlägg» teoksessa *Det 18:e nordiska statistikermötet i Esbo 1989*. Nordisk statistik skriftserie 53, Helsinki.

Järvelin, K. (1986): »User charge estimation in numeric on-line databases», *Acta Universitatis Tamperensis, ser A vol 212*, Tampere.

Lassila, J. (1990): »Preliminary Data in Economic Databases. Proceedings of the 5th International Conference on Statistical and Scientific Database Management», *Lecture Notes in Computer Science 420*, Springer-Verlag, Berlin.

NSF (1990): »National Science Foundation Workshop on Scientific Databases», Workshop Report, Washington.

Thygesen, L. (1990): »EDP og spredningen af statistikken til brygerne» teoksessa *Det 18:e nordiska statistikermötet i Esbo 1989*. Nordisk statistik skriftserie 53, Helsinki.

Trivellato, U. ja E. Rettore (1986): »Preliminary data errors and their impact on the forecast error of simultaneous-equation models», *Journal of Business & Economic Statistics* 4, 445—453.

Vajanne, Heikki (1991): *Tablen käyttöohje*. ETLA.

Wolski, K., T. Koivula, M. Lukkari ja M. Mäkinen (1990): »Numeroita näytöllä. Ulkomaisia numeerisia tietokantoja», *Tilastokeskus, Käsi-kirjoja 25*, Helsinki.