

# Tilastotuotannon standardointi

## - EG Add-In?

Heikki Rouhuvirta, Tilastollinen t&k yksikkö  
[heikki.rouhuvirta@stat.fi](mailto:heikki.rouhuvirta@stat.fi)

SAS Forum Finland 2009

---

# Esityksen kulku

- kolme näkökulmaa SAS komponentteihin  
järjestelmäkehityksen työkaluina

1. Yleinen toimintaympäristö - mihin menemme?
2. Tilastotuotanto - sen logiikka
3. Välineet - SAS

# EU

Mitä EU:ssa tilastotuotannolla tavoitellaan

■ Harmonisoitu kuva

- keinot: a) sisältöjen harmonisointi  
b) tulkinta ja tulkinnan edellytysten parantaminen

=> Ratkaisu tietojärjestelmätasolla on laaja-alaisen sisällöllisen kuvaustiedon tuottaminen, käsittely ja jakelu

2. Tiedon aitous

- keinot: a) harmonisoitu tuotanto  
b) laatu  
c) tuotantoprosessin läpinäkyvyys

=> Ratkaisu tietojärjestelmätasolla on laaja-alaisen laatu- ja tuotantoinformaation tuottaminen, käsittely ja analysoiminen

*- näkemykset ovat omiani eivätkä ne välttämättä heijasta  
Tilastokeskuksen kuin myöskään Eurostatin kantoja*

# Tilastotiedon käyttäjät

## Tilastotiedon käyttäjien odotukset

- Sisällöltään laajempaa ja syvällisempää tilastotietoa  
keinot: a) tuotannon tehostaminen  
b) menetelmät ja teknologia

=> Ratkaisu tietojärjestelmätasolla on tuotantoprosessin uudistaminen sekä menetelmällisesti, teknologisesti että tuotannollisena prosessina

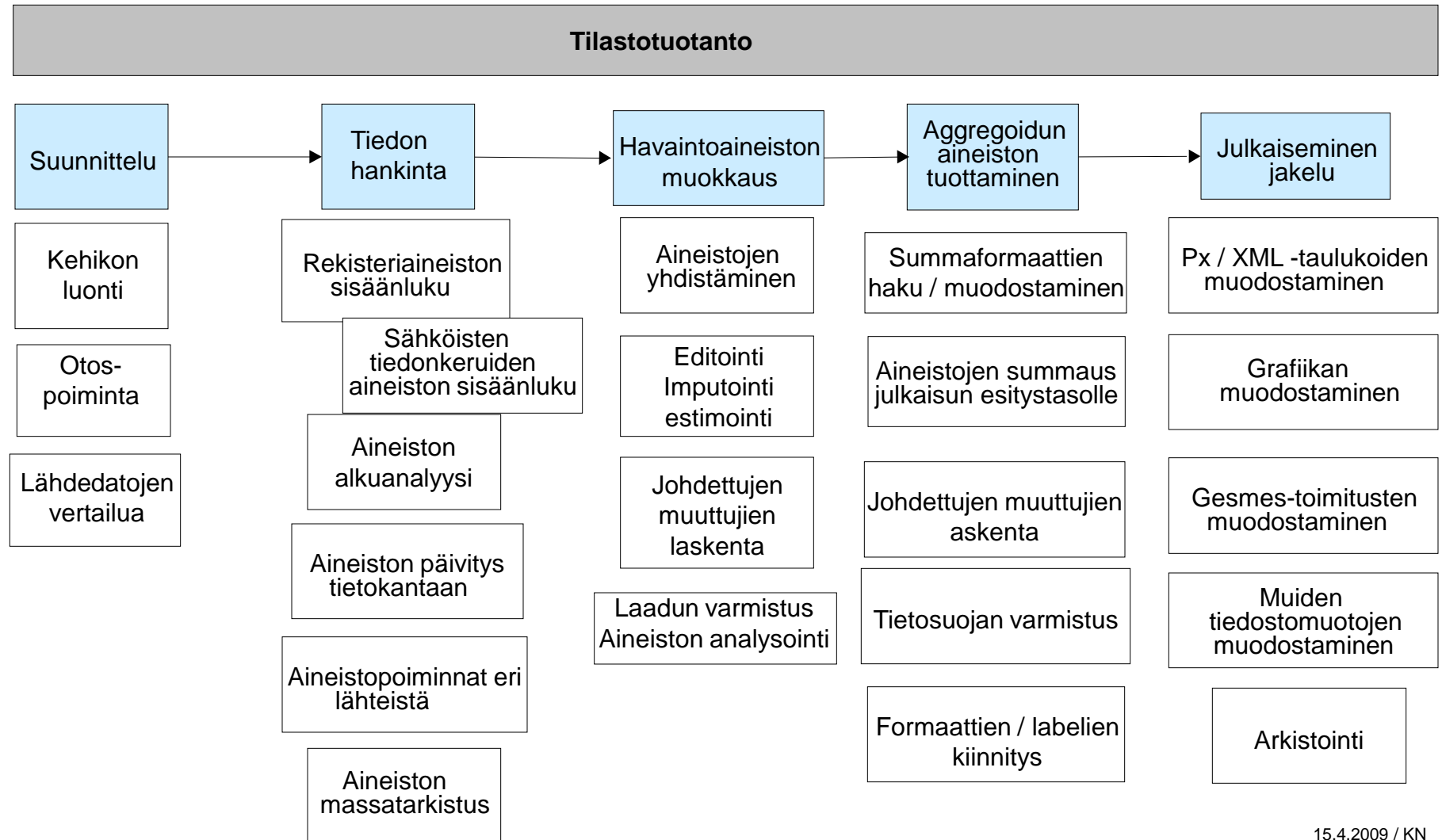
2. Tilastotiedon helppoa saavutettavuutta  
keinot: a) tuotannon tehostaminen  
b) menetelmät ja teknologia

=> Ratkaisu tietojärjestelmätasolla on tuotantoprosessin uudistaminen sekä teknologisesti että tuotannollisena prosessina

*- näkemykset ovat omiani eivätkä ne välttämättä heijasta kaikkien tilastotiedon käyttäjien käsityksiä*

# Tilastotuotanto prosessina

# Tilastotuotanto (I)



15.4.2009 / KN

# Tilastotuotannon prosessipiirteet

1. Toisteisuus

2. Vaiheisuus

3. Niputus

⇒ Kaikki tämä tapahtuu uudestaan ja uudestaan, mutta vain likimuodossa – mikään tuotantokierros ei ole täsmälleen sama kuin edeltävä tuotantokierros

⇒ Joka puolestaan on johtanut tuotannon manuaalisuuteen

# Manuaalisuus tilastotuotannossa

1. ilmenee siinä, että kaikki empiirisen datan tyypilliset käsittelyvaiheet toteutetaan



## Opas SAS Enterprise käyttäjäll

EG versio 4.1

Timo Hurme

avaaminen Ent

oismuuttujien sisään lukemi

0.7 SAS-datan avaaminen Enterprise

6.8 Datojen tallentaminen Enterprise G

6.9 Projektin tallentaminen

7 Datan muokkaus

7.1 Datan muuttaminen muokkaustilaan

7.2 Muuttujien perusominaisuuksien määrittel

7.3 Muuttujan arvojen tulostusmuotojen eli forr

7.4 Datan muokkauksen työkalut

7.5 Muuttujien muunnokset ja uusien muuttujien t

7.6 Muuttujien luokittelu ja luokkien yhdistäminen..

7.7 Muuttujan luominen aritmeettisella operaatiolla ...

7.8 Muuttujien poistaminen

7.9 Muuttujien järjestyksen vaihtaminen datassa

7.10 Datan havaintorivien järjestäminen

7.11 Osa-aineiston eli tiettyjen havaintorivien valitseminen

7.12 Kahden tai useamman datan yhdistäminen

7.13 Datan kääntäminen

7.14 Muuta datojen käsittelyyn liittyvää

8.2 Pylväs

8.3 Viivakuva (Line

8.4 Laatikko-janakuvi (Boxplot)

8.5 Piirakkakuvi (Pie Chart)

8.6 Wizard-kuvatyökalut

8.7 Kuvien tulostusmuoto ja yleisasetukset

9 Datan tarkastelu Describe- ja Analyze-valikoiden työkaluilla

9.1 Datan listaukset

9.2 Datan luonnehdinta -työkalu (Characterize Data)

9.3 Jatkuvan muuttujan jakauman tarkast

9.3.1 Distribution Analysis -työkal

9.3.2 Summary Statistics -työkal

9.3.3 Summary Tables -työkal

9.3.4 Tunnuslukujen taulukoir

9.4 Kategorisen muuttujan jakaur

9.5 Kahden muuttujan yhteyden

9.5.1 Kahden numeerisen

9.5.2 Kahden kategorise

11.1 Tulostusasetukset

11.1.1 Tulostuskohteen valinta

11.1.2 Tulostustyylin valinta

11.1.3 Muut tulostusasetukset

11.2 Yksittäisen tulosteen hallinta ja jatkokäyttö

11.2.1 Kohteen ja tyylin vaihtaminen yksittäiselle tulosteelle

11.2.2 Yksittäisen tulosteen tallentaminen tiedostoon

11.2.3 Tulosteen tai datan siirtäminen toiseen ohjelmaan

11.2.4 Valmiin kuvan muokkaaminen

11.2.5 Selitetekstien ja formaattien lisääminen yksittäiselle tulosteelle

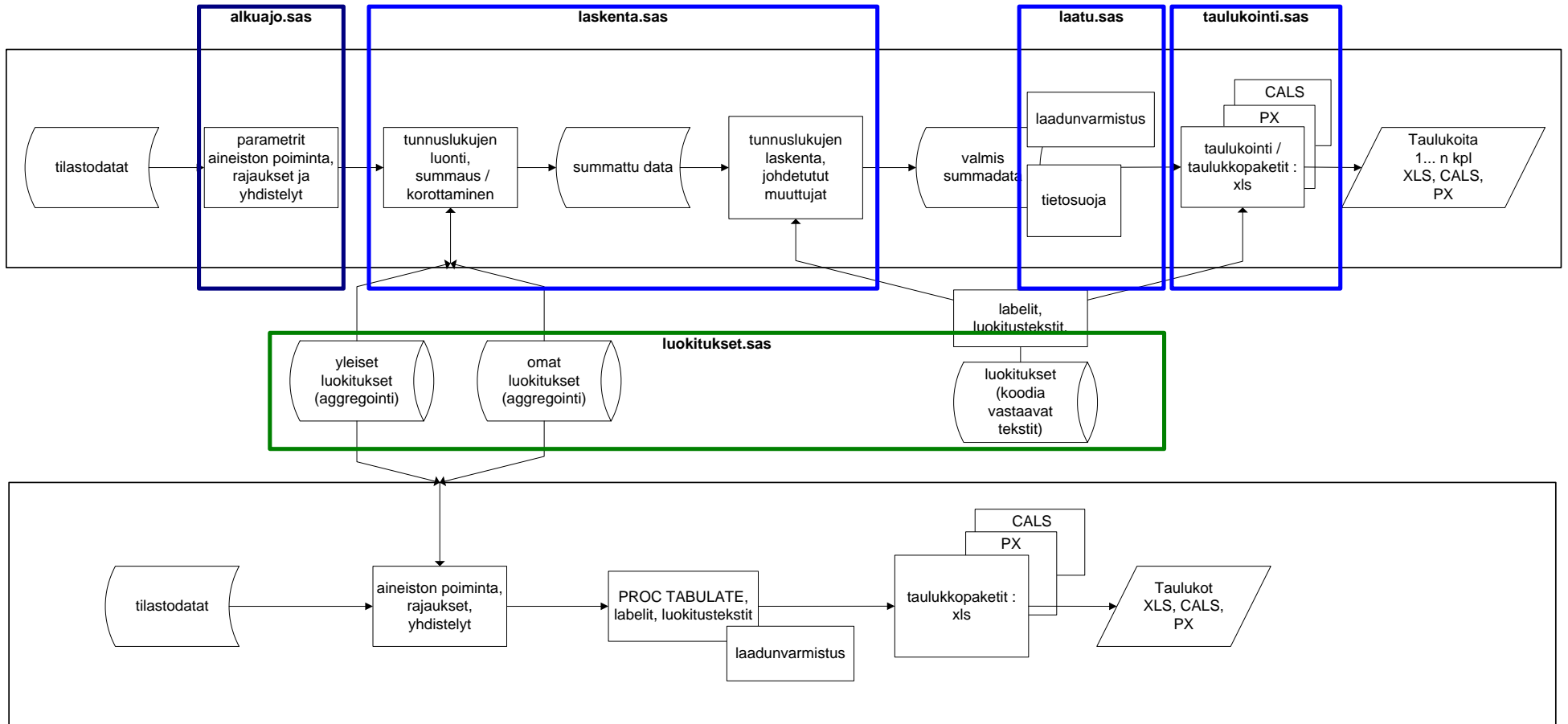
11.2.6 Otsikoiden ja alaviitteiden muokkaaminen

11.3 Yksittäisten tulosten kokoaminen yhteen

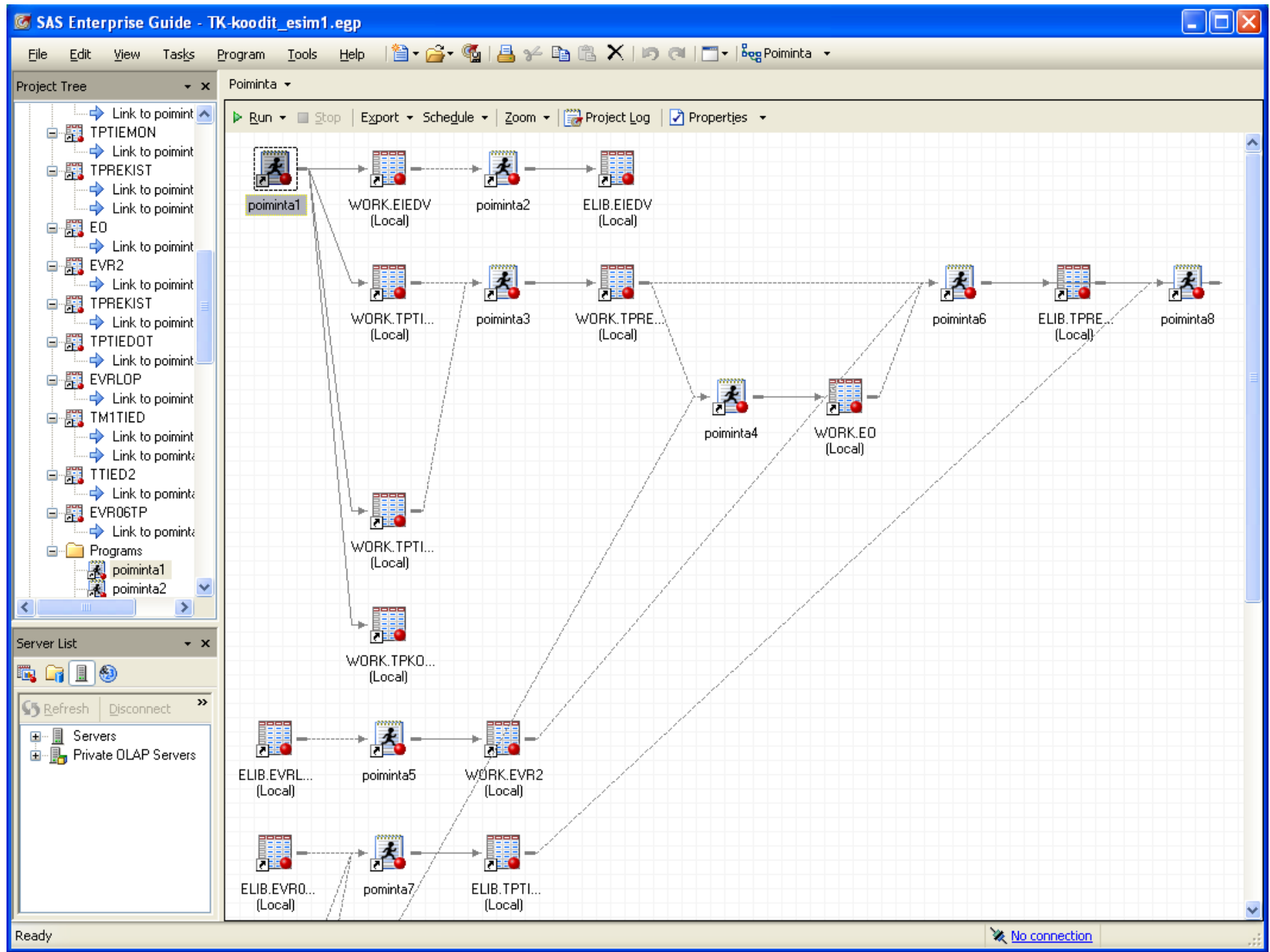
11.3.1 HTML dokumentti

11.3.2 SAS Raportti

# Tilastotuotanto (II)



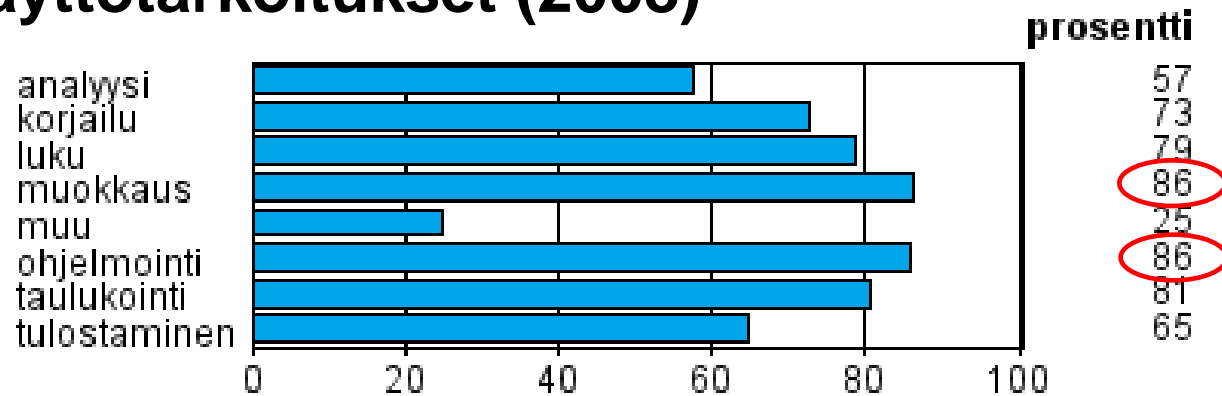
# Tilastotuotanto (III)



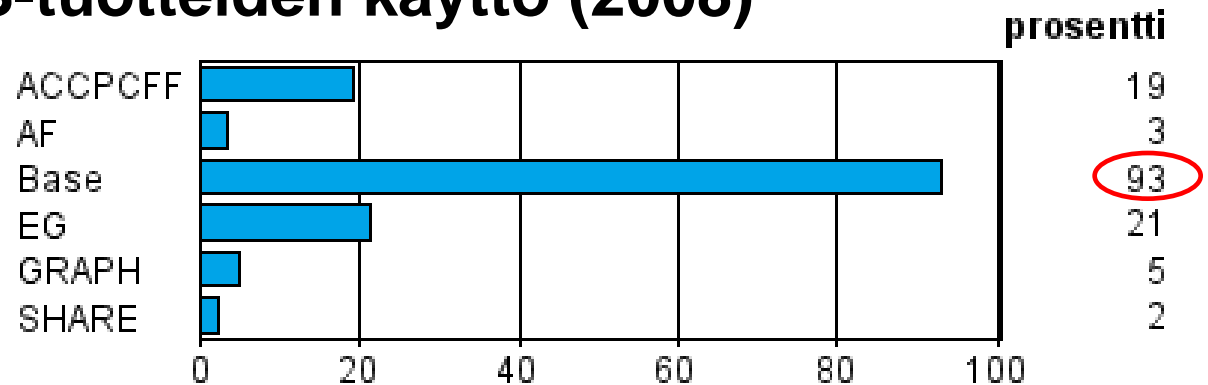
# Manuaalisuus tilastotuotannossa

2. ilmenee siinä, että empiirisen datan käsittelyvaiheet toteutetaan ohjelmakoodia tuottamalla ja sitä editoimalla

## SAS-käyttötarkoitukset (2008)



## Eri SAS-tuotteiden käyttö (2008)



# Mitä pitäisi tehdä ja miten?

Tilastotietojärjestelmän tulisi

- Sisältää laajan sisällöllisen metatiedon
- 2. Sisältää laajan tuotantoinformaation kokoamisen ja analysoimisen
- 3. Tuottaa automaattisesti tuotannon ja tilastotiedon laatukuvaukset
- 4. Luoda edellytykset tuotantoprosessin ohjaamiselle tuotanto- ja laatuinformaation perusteella

# Mitä pitäisi tehdä ja miten?

## Tuotantoparadigmojen muuttaminen

- Tilastotyö on yksilöllistä työtä yksilöllisin työvälinein
- Tilastollinen työ on kertaluonteisilla aineistoilla tehtävää käsityötä
- 2. Tilastotuotanto on automaattinen vaiheittainen prosessi, jossa prosessitietona on vaiheen suoritetuksi tuleminen

## Uusi paradigma

- => tuotanto, jossa numeerisen datan ymmärrettävyys on tuotantoprosessin eri vaiheissa turvattu ja jossa tilastoaineiston käsittelyratkaisut perustuvat menetelmällisiin ratkaisuihin ja jossa aineiston käsittelyn vaikutuksista (automaattisesti) koottu tieto ohjaa tuotantoa laatukriteereiden mukaisella tavalla

# SAS tilastotuotannossa

# SAS ja tilastotuotanto

## Vaihtoehdot

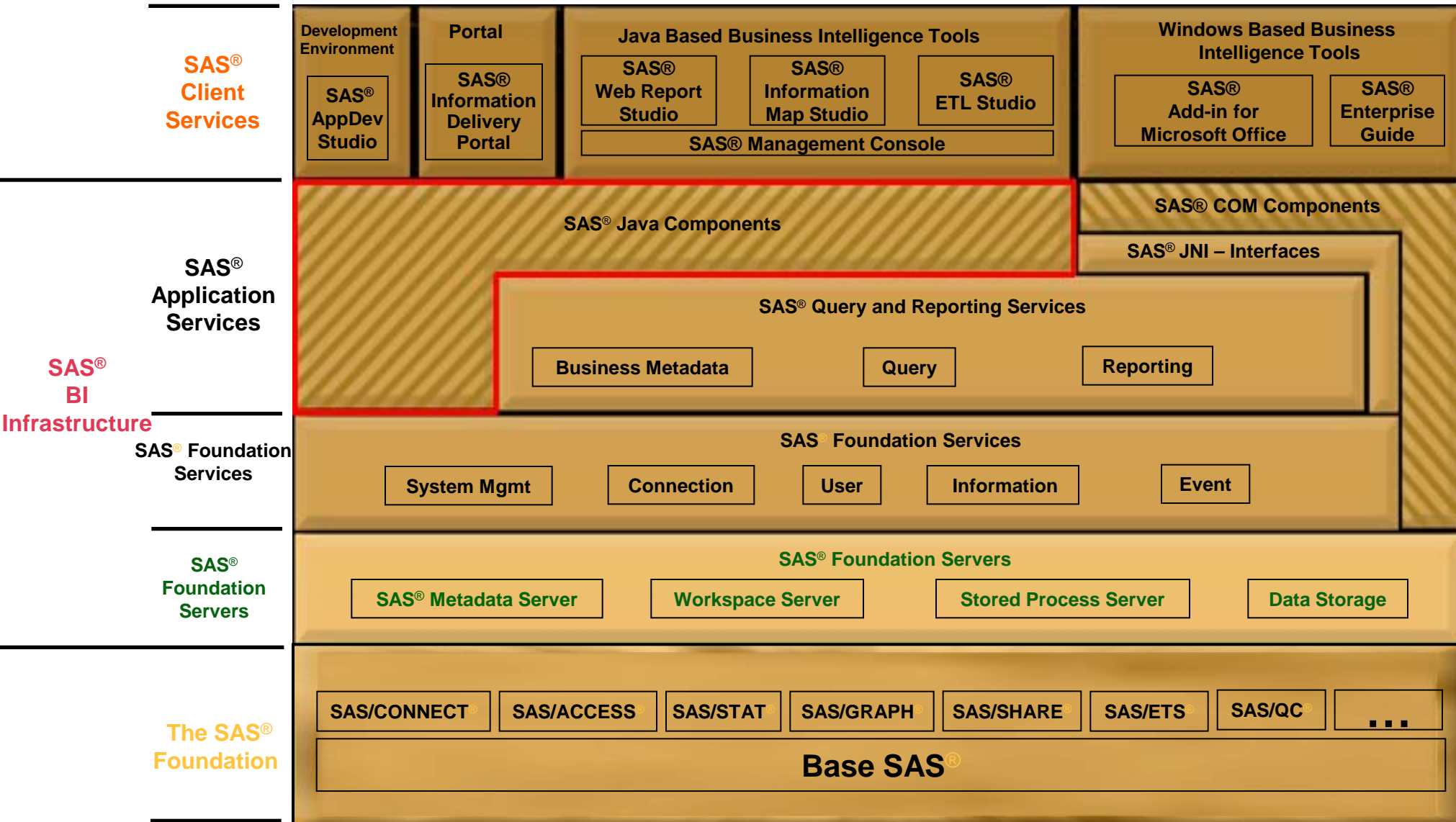
1. Tilastotuotannon sopeuttamien SASiin
    - näin on jo tehty eikä se toimi (Base SAS merkkipohjaisine käyttöliittymineen ja suoritusten hallintoineen)
  2. SASin sopeuttaminen tilastotuotannon tarpeisiin
    - helpompi muuttaa SASin kehityssuuntaa kuin vaikuttaa 800 tilastontekijän tottumuksiin ja mieltymyksiin
- ⇒ SASin keinot prosessien hallintaan
  - ⇒ SASin keinot tilastotuotannon “serialisointiin”
  - ⇒ SASin keinot tuotantoinformaation käsittelyyn
  - ⇒ SASin keinot numeerisen tiedon merkityksen/sisällön kuvaamiseen

# SAS

## Vaihtoehdot

- Base SAS + koodi
- SASin visuaaliset käyttöliittymät
- 2. Muut mahdollisuudet

# What? – SAS® 9.1 Architecture



# SASin visuaaliset käyttöliittymät (I)

## Vaihtoehdot

### 1. SAS IML Studio

- ei prosessin hallintaa
- ei integroitavissa
- mallintajan työväline

### 2. SAS JMP

- stand alone –ratkaisu
- lähes reaaliaikaisten tarkastelujen ja parhaat aineiston ominaisuus- ja niiden graafisten kuvausten tuottamisen mahdollisuudet
- ei prosessin hallintaa
- ei tietosisällön kuvaustiedon käsittelyn tukea
- integroitavissa SAS EG:hen

# SASin visuaaliset käyttöliittymät (II)

## Vaihtoehdot

### 1. SAS Enterprise Guide

- prosessin hallintaa, mutta ei riittävä
- ei tietosisällön kuvaustiedon käsittelyn tukea

### 2. SAS Data Integration Studio

- prosessin hallinnan tuki
- ei integroitavissa
- ei tietosisällön kuvaustiedon käsittelyn tukea

⇒ Sellaisenaan mikään väline ei ole ratkaisu tilastotuotannon kehittämistarpeisiin

⇒ Onko jokin yhdistelmä mahdollinen sovitettavaksi tilastotuotannon tarpeisiin?

# SAS ratkaisun komponentit (työhypoteesi)

## 1. SAS Enterprise Guide

- prosessin hallinnan perusteet
  - parametrit
  - tuotantokaavio/ajokaaviot
  - tehtäväryhmien kokoaminen/ajoryhmät
  - SAS Stored Processes

## 2. SAS JMP

- integroidaan SAS EG:hen
- tuotetaan aineiston ominaisuuskuvaukset ja niitä vastaavat graafiset kuvaukset keskeisten tuotantovaiheiden osalta (aineiston alkuanalyysi, lopullinen taulukoitava aineisto)
- tuotetaan tuotantoinformaation ominaisuuskuvaukset sekä tuotantoinformaation graafiset kuvaukset eri muokausvaiheiden vaikutuksista

# SAS ratkaisun komponentit (työhypoteesi)

## 3. SAS Enterprise Guide Add In

- täydentää prosessin hallinnan ominaisuuksia
  - toimenpideryhmittäiset osasuoritusmäärittelyt
  - toisteisten teknisten muokkausten automatisoinnit
- luo edellytykset tietosisällön kuvaustiedon käsittelyyn
  - eXist -XML-kantayhtyes
  - luokitusten ja uusien muuttujien määrittely ja (tekstuaalinen) kuvaaminen niiden muodostamisen yhteydessä
  - arkistointi (säilytettävien datatiedostojen tallennus ja kuvaus)

# Tilastontekijän työpöytä I (utopia)

1. Mitä hallitaan (i)
  1. Toisteisuus
  2. Vaiheisuus
  3. Niputus
  
2. Mitä hallitaan (ii)
  1. Mitä on tehty, millä perusteella ja millaisin vaikutuksin
  2. Tilastoinformaation jalostaminen – kaikki uusi tieto on kuvattu ja dokumentoitu
  
3. Millä välineillä
  1. SAS EG (Process Flow/project tree)
  2. EG Aadd In (tehtäväkokonaisuudet)
  3. SAS JMP (tilasto ja tuotantoinformaation analyysit)

# Tilastontekijän työpöytä II

## Prosessin hallinta

1. Tilastoaineiston avaus (Add In)
  1. Avaa tilastoajankohdan perusaineiston
  2. Tuottaa peruskuvauksen aineistosta
    - Aineiston koko (havainnot x muuttujat)
    - Puuttuvien tietojen määrä ja kohdentuminen
  3. Suorittaa tekniset tarkistukset ja muunnokset
  4. Yhdistää tilastolliset metatiedot aineistoon (aineistokuvaus muuttujittain ja muuttujien perusluokitukset)
2. Tilastoaineiston alkuanalyysi (Add In + JMP)
  1. Tuottaa perusaineiston jakaumantiedot
  2. Tuottaa vertailun edellisen tilastoajankohdan aineiston ominaisuuksiin
  3. Tuottaa peruseditointimenetelmien valinnan kannalta keskeiset tunnusluvut

# Tilastontekijän työpöytä II .....jatkuu

3. Tilastoaineiston editointi (Add In)
  1. Aineistolle soveltuvien editointimenetelmien ja -toimenpiteiden valinta (parametriohjattu valikko-käyttöliittymä)
  2. Editointitoimenpiteiden suoritus (parametriohjattu valikko-käyttöliittymä)
  3. Tuottaa perusdatan editointitoimenpiteiden suorituksesta (määrä ja laatu sekä kohdentuminen)
  4. Tuottaa perusdatan editointitoimenpiteiden vaikutusten analysoimiseksi
  
4. Tilastoaineiston paikkaaminen (imputointi) ja estimointi; aineiston hyväksyntä (Add In + spesifit menetelmäratkaisut + JMP)
  
5. Taulukointi (Add In)
  1. Tilastotaulukoiden määrittely (visuaalinen käyttöliittymä)
  2. Tuotannon suoritus (parametriohjattu valikko-käyttöliittymä)
  3. Tulostus (XML –taulukointi ODS) – sisältää tilastollisen informaation kuvausten liittämisen taulukkoon XML-kannasta

**Kiitokset kiinnostuksestanne!**