



Development of Electronic Data Collection in Statistics Finland



Topics

- General background of EDC in Statistics Finland
- W3 pilot project
- .NET -based collection systems in Business Trends – statistical unit

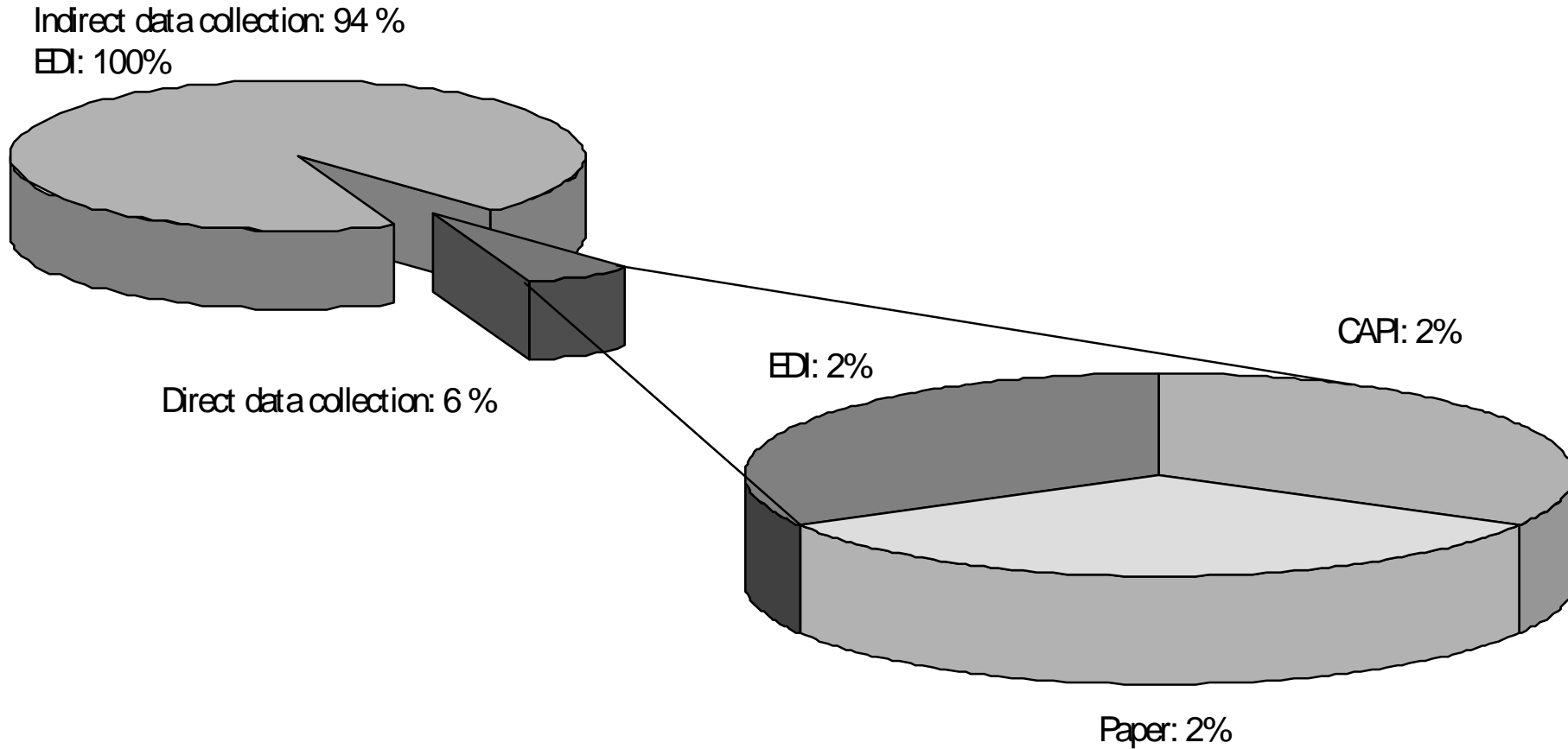


Background

- About 94% of data collection is made from administrative registers
- About 6% of data collection is carried out using direct data collection
 - paper forms
 - machine readable data / Primary EDI
 - interviews by CATI/CAPI systems, mainly using Blaise software
- Guideline from the Ministry of Finance
 - All parties presently still supplying their data on paper forms should also have the possibility to transmit them electronically by the end of 2006.



Data collection in Statistics Finland by type and media used





Electronic data collection

- Primary objectives
 - simplifies data collection process
 - reduces need for human resources
 - reduces other data collection costs
 - improves the quality of collected data
 - reduces response burden
 - speeds up the data accumulation
 - decreases non-response
 - enables direct individual feedback for respondents
 - enables previously submitted data browsing
 - removes overlapping collection and promotes joint use of the collected data between different authorities



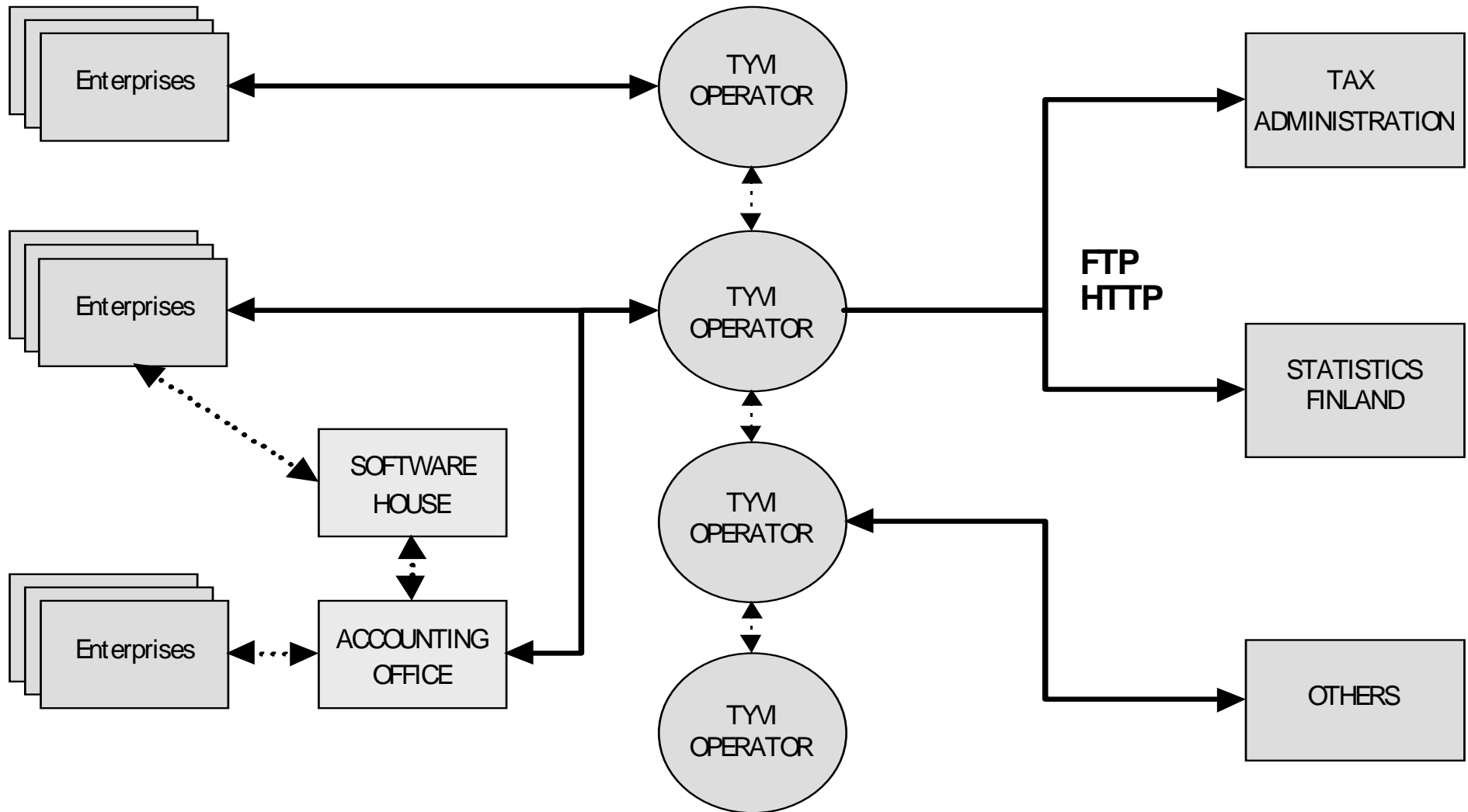
Data flows

- Different types of data flows
 - data are needed only by Statistics Finland
 - the same data are needed by several administrative organizations
 - interviews made by CATI/CAPI system
- Different solutions
 - using external teleoperator for distributing data to different data collectors (TYVI model)
 - self-made web-based system
 - Blaise solution for carrying out interviews



The TYVI model

- Data Flows from Enterprises to Authorities
 - interfaces and transmission
 - data capture
 - data refining
 - management of user accounts
- Participants
 - The enterprises
 - The TYVI-operators
 - The authorities
- The authority needs not to be in relationship of many to many with the respondents



The TYVI-model (Vallaskangas 1998)



Internet -based collection of data

Case: Building Cost Index



General background

- Fall 2000
 - All existing electronic data collections were handled by 3rd party operators (TYVI model)
 - The production system of Building Cost Index was under re-construction and lacked web-based data collection
- About Building Cost Index (Business Trends)
 - ~300 respondents (hardware stores, wholesale stores, plumbing stores etc.)
 - Price information of 1-15 products collected from each respondent every month
 - Paper forms are usually sent on the 15th day and expected back around the 25th day



The design goals of the web system

- Provide means of web based collection of statistical data
- No extra burden (no installations, no javascript based solutions etc.)
- “Live” feedback to the respondents (upon validations etc.)



Hardware architecture

- Running on Windows NT server
 - Web server: Microsoft Internet Information Server 4 (IIS4)
 - Component Server: Microsoft Transaction Server 2.0
 - Anonymous access (No NT-authentication)
- Database server
 - Windows 2000 server
 - Running Microsoft SQL Server 2000
 - Deployed on DMZ, accessible only through firewall



Application architecture

- Built using Microsoft Windows DNA (Distributed iNternet Application Architecture)
- Standard 3-tier architecture that consists of
 - Presentation layer: HTML, ASP
 - Business layer: COM components
 - Database layer: Relational database
- System consists of two separate modules (both self-made)
 - User authentication
 - Data collection



Experiences

- Beta phase from 5/2001 - 9/2001, 30 respondents
- 9/2001 - 2/2002, 70 users
- In 3/2002 the systems was opened to all respondents
 - 147 users at the moment (nearly 50%)



Internet -based collection of data

CASE: Business Trends' collection systems



Sale inquiry: Background

- Monthly inquiry using paper forms (~2000 enterprises)
- Data collection process (7-8 persons: 3,0 working years)
 - Printing and mailing the questionnaires
 - Receiving the questionnaires (mail, fax, e-mail, TYVI)
 - Validating and entering the data
 - Printing and mailing the reminders
 - Phone inquiry
- Quarterly non-individual feedback for respondents
- Previously submitted data preprinted to questionnaires for the last 2 months



Volume index of industrial output: Background

- Monthly inquiry using paper forms (~1000 establishments)
- Data collection process (4 persons: 2,5 working years)
 - Printing (once in a year) and mailing the questionnaires
 - Receiving the questionnaires (mail, fax, e-mail)
 - Validating and entering the data
 - Printing and mailing the reminders
 - Phone inquiry
- Monthly non-individual feedback for respondents
- Previously submitted data in questionnaires for the current year



Inventory inquiry: Background

- Quarterly inquiry using paper forms (~900 enterprises)
- Data collection process (2 persons: 0,7 working years)
 - Printing and mailing the questionnaires
 - Receiving the questionnaires (mail, fax, e-mail)
 - Validating and entering the data
 - Printing and mailing the reminders
 - Phone inquiry
- No feedback for respondents
- No previously submitted data preprinted to questionnaires



The inquiry on incentive stock options: Background

- Monthly inquiry using paper forms (~200 enterprises)
- Data collection process (1 person: 0,3 working years)
 - Printing and mailing the questionnaires
 - Receiving the questionnaires (mail, fax, e-mail)
 - Validating and entering the data
 - Printing and mailing the reminders
 - Phone inquiry
- No feedback for respondents
- Previously submitted data printed in questionnaires for last 5 months



Design goals

- Create framework for similar systems
- Multi-language support
- LDAP -based user authentication w/ centralized administration
- Create generic method for transferring data between collection and production databases
- Create "mass emailer" for all kinds of collection systems



Software & hardware architecture

- Built using Microsoft.NET and ASP.NET
- Generic 3-tier architecture w/ presentation, business and database logic
- Collection database separated from the production database
- 128 bit encryption used for communication between respondents and Statistics Finland



Framework of the collection system

- The modular structure of the framework allows to
 - Change menus, headers, footers and other styles
 - Add custom functionality (using ASP.NET user controls) on the pages
 - Add and load different languages for the pages
- The base use cases are more or less same in different collection systems (login, questionnaire, feedback, instructions and contact information)



Multi-language support

- Most of the textual information on the web pages is stored in the database
 - Texts are loaded on the server's memory on the system startup
 - Only long descriptions are kept as files
- Page language can be changed "on the fly"
- Every element has a tag on the page template and the relevant text is attached to the element upon the page load



User authentication

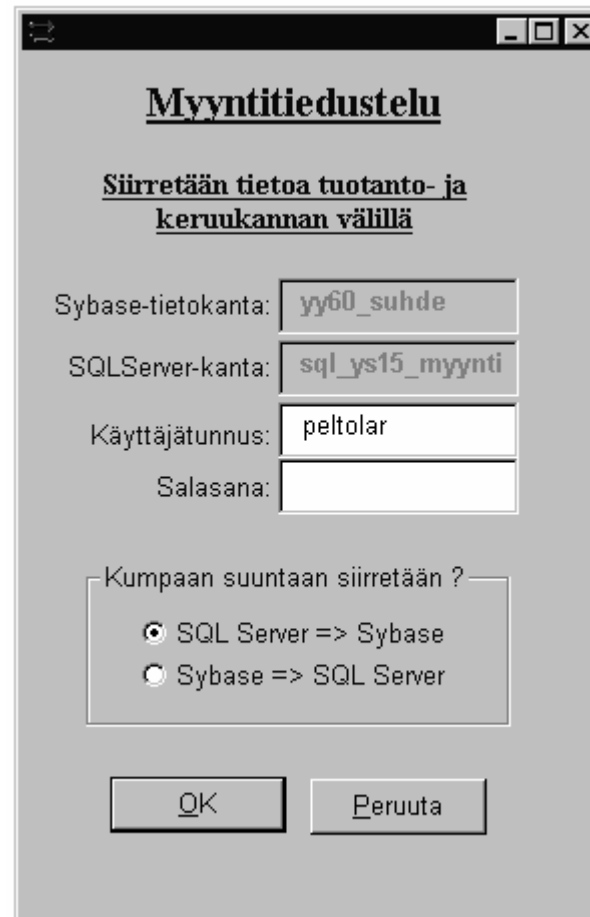
- The objective was to use LDAP (lightweight directory access protocol) for the user authentication
- The development for this didn't proceed in the schedule, so it was temporarily replaced with database-based user authentication and administration
- Authentication thru LDAP has been tested and it seems to be an ideal solution
- At the moment we're building a simple web administration application to finish the LDAP part



Data transfers

- Data transfers between collection and production databases are handled with an external win32 -application
- Built with PowerBuilder using pipeline feature (data flow)
- Data from collection database is transferred to the temporary tables in the production database and then synchronized with the actual tables
- Solution is quite customizable, allowing new functionality by adding new pipelines

Data transfers



Myyntitiedustelu

Siirretään tietoa tuotanto- ja keruukannan välillä

Sybase-tietokanta:

SQLServer-kanta:

Käyttäjätunnus:

Salasana:

Kumpaan suuntaan siirretään ?

SQL Server => Sybase

Sybase => SQL Server



Mass emailer

- An external application was built with Visual Basic 6 to send emails to the respondents
- Modular approach
 - New systems can be added using textual configuration files
 - Reply requests can be added by writing sql statements to the configuration files
- Supports attachments
- Replaces traditional letters

Mass emailer

Meilirobo v1.1 [Käännös 2003-12-17]

Valitse järjestelmä
 Editoi... Päivitä

Valitse lähetystyyppi
Lähetystyyppi

Lähtäjän sähköpostiosoite:

Lähtäjän nimi

Aihe:

Tekstiosan polku
 ...

HTML-muoto
 Kyllä Ei

Palautusosoite

Vastaanottajat

Valmista lähetys
Poista osoite
Lähetä

Sulje

Liitetiedostot

Lisää Poista



Development experiences

- Microsoft.NET was just released when the development began
- Development environment wasn't always stable and the developers experienced quite a lot of unexpected behavior
- Despite this, ASP.NET is quite an improvement when comparing to other web application methods (asp, php, perl etc.)
 - Although inter-browser compatibility is still quite poor



Development ideas

- Although the framework is quite good, some ideas have arisen
- Use of XML to
 - Define the concepts of the questionnaires
 - Define the presentation (XSLT)
 - Define the validations
- Provide means of graphical feedback to the respondents (charts etc.)
- Replace the user authentication with LDAP



Benefits

- Enables
 - Complex validations of the data
 - Dynamic creation of presentation layer logic
 - Displaying of pre-fetched data to individual respondents
 - Live feedback to the respondents (validation errors etc.)



Drawbacks

- Requires user/customer administration for
 - Maintaining user profiles
 - Helpdesk/Support services



Effects of the electronic data supply system on data collection process

- Printing the questionnaires ⇒ **Transferring data to collection database**
- Mailing ⇒ **E-mail informing (mass emailer)**
- Receiving the questionnaires (mail, fax, e-mail, TYVI) ⇒ **(Electronic data supply)**
- Validating and entering the data ⇒ **Mass validation**
- Printing and mailing the reminders ⇒ **E-mail reminder (mass emailer)**
- Phone inquiry ⇒ **Phone inquiry**
- Non-individual delayed feedback ⇒ **Individual direct feedback**
- Limited access to previous own data ⇒ **Previous own data available**

- *Manual exclusive treatment* ⇒ ***Electronic mass treatment***



Current situation of implementation

- Sale inquiry
 - Available for all respondents (~2000)
- Volume index of industrial output
 - Available for ~200 test users
- Inventory inquiry
 - Available for all respondents (~900)
- The inquiry on incentive stock options
 - Available for all respondents (~200)



Results: Sale inquiry

- Electronic data supply system users of all respondents:
 - 1. month: 48%
 - 2. month: 59%
 - 3. month: 61%
 - 4. month: 70%
- Reminders sent:
 - before electronic data supply system: ~1000
 - 1. month: ~800
 - 2. month: ~700
 - 3. month: ~650
 - 4. month: ~550



Experiences

- Feedback from respondents has been very positive: Response burden has reduced remarkably
- Enthusiasm of persons involved in data collection
- Manual data treatment has reduced (at least by 50%)
- Quality of data has improved: Validation, additional information if data is not comparable etc.

- Number of enquires made by respondents concerning electronic data supply system:
 - first two months: ~100 / month (mainly questions concerning base settings)
 - since third month: ~30 / month (mainly forgotten passwords)