

# XML- based Data Collection Case: Tourism Statistics

Voorburg Group Meeting, Statistics Finland 28 September 2005

Ville Vertanen  
Toni Räikkönen



## Statistics Finland's strategy

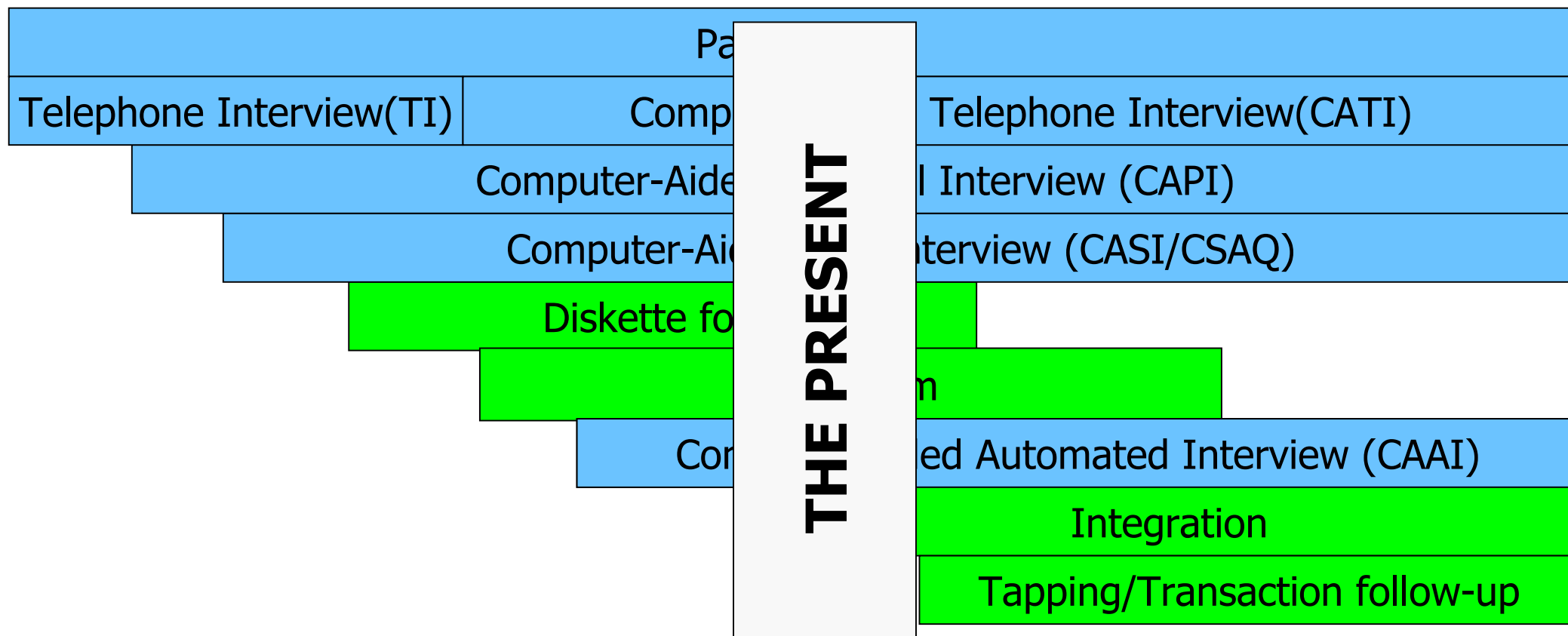
- To offer an electronic option in all data collections by 2007
  - Does not extend to collection of personal data
- It's the respondent's choice whether to use this option or not
  - Difficult to use? Resistance to change?
  - Lack of Internet connection
  - Security

## Business data collections (excl. financial statistics)

- About 45 surveys  
(excluding collections with less than 30 respondents)
- Web (Internet form) collections in use on 1 May 2005: 15
  - By the end of year 2006: 45
- Summer- autumn 2005 piloting of direct data collection from enterprise management systems into Statistics Finland (SF) database.
- Paper option will remain for the foreseeable future  
--> decreases benefits

# Development of Tools and Methods

time



## Case: Tourism Statistics



## Tourism Statistics

- Monthly questionnaire
  - Includes all establishments with 10 or more rooms, cottages or caravan pitches
  - Covers 90% of room capacity in Finland
  
- Enquiry covered 1,500 accommodation establishments in 2003, of which
  - 779 hotels
  - 213 similar establishments
  - 312 camping sites
  - 113 holiday villages
  - 47 youth hostels

## Data collection for Tourism Statistics

- A number of establishments are open only in summer (camping sites, holiday villages)
  - Number of establishments:
    - In summer about 1,500 establishments (max.)
    - In winter about 1,000
- Obligation to supply data is based on the Statistics Act
- Non-response approximately 15%.
- Non-response weighted with capacity approximately 4%

## Data collection for Tourism Statistics

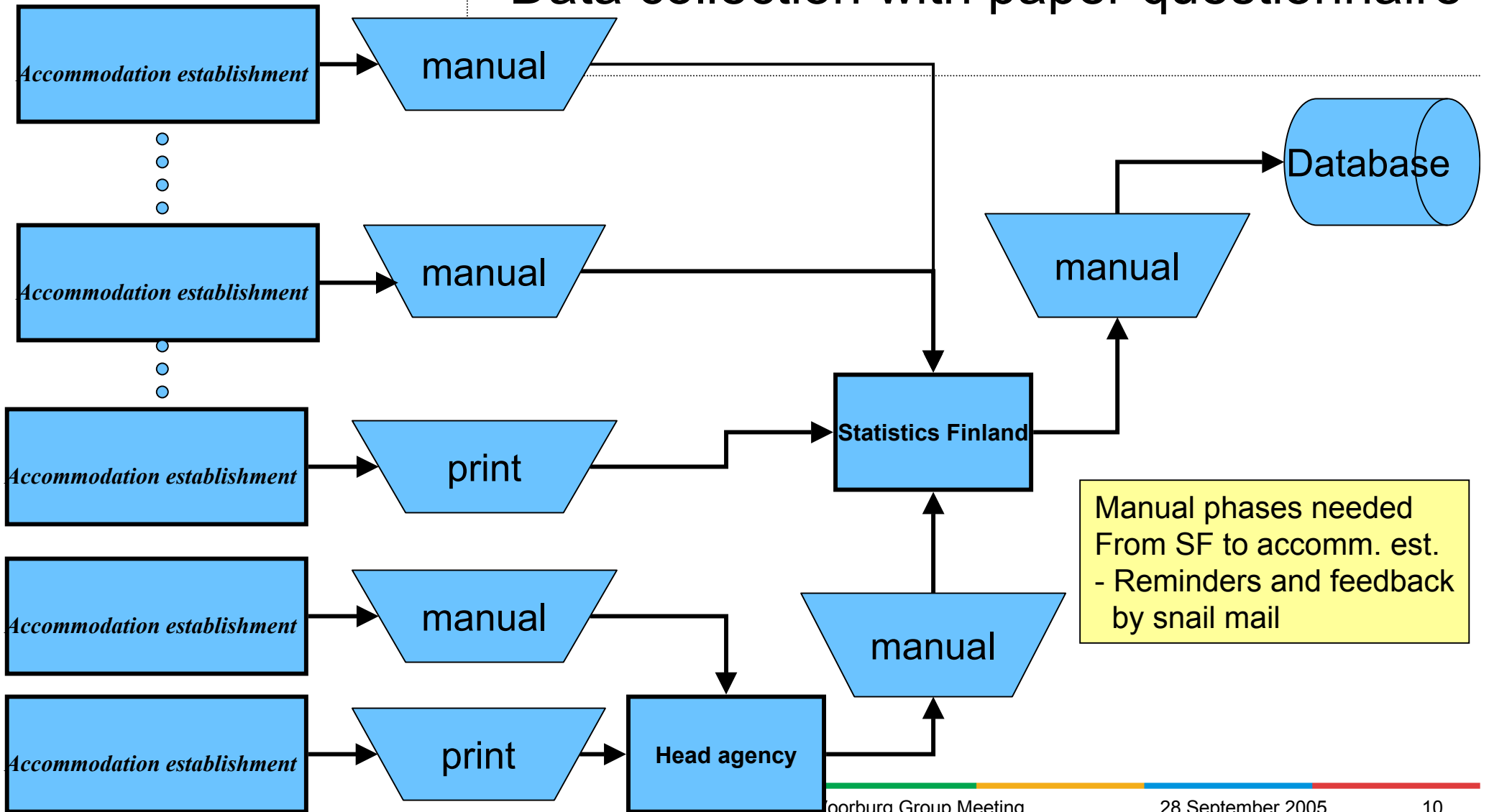
- Information that is collected every month
  - Sales income from accommodation
  - Accommodation capacity and room occupancy rate in the reference month
  - Nights spent by country of residence, purpose of visit and mode of lodging
  - Arrivals by country of residence in reference month



## Data collection for Tourism Statistics

- Up to this year collected information was keyed manually from paper forms into our database
- From the beginning of 2005 it has been possible to respond by an Internet questionnaire (now 170-200 users)
  - Information is entered manually into the questionnaire
  - In July 2005, there were 185 answers via the Internet
- In autumn 2005 it will be possible to send information direct from hotel management systems into our collection database
  - Typically, no manual input needed at Statistics Finland or at hotels

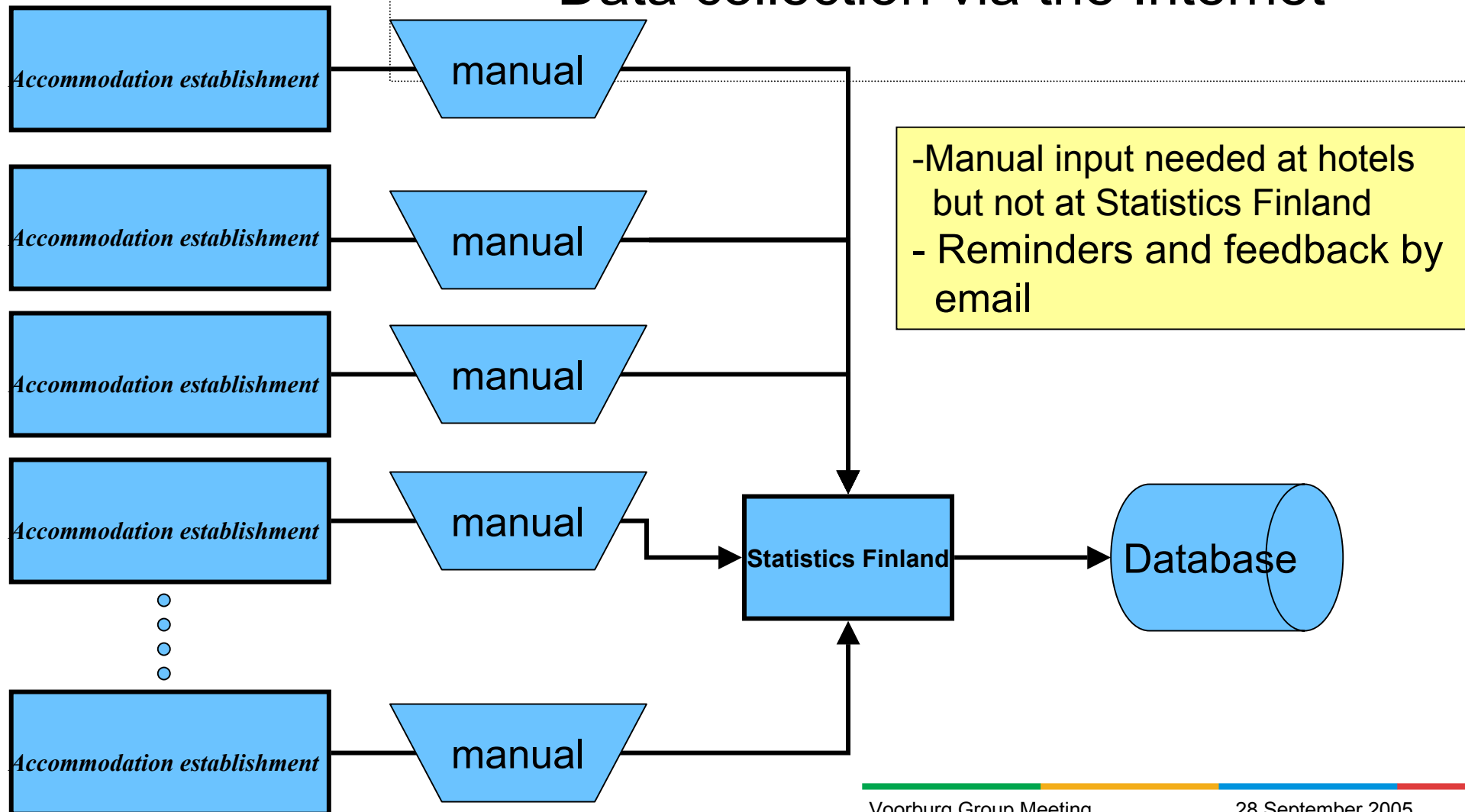
## Data collection with paper questionnaire



## Data collection with paper questionnaire

- Paper questionnaire
  - By mail
  - By fax
- The questionnaire is also available in Excel format
  - By email

## Data collection via the Internet

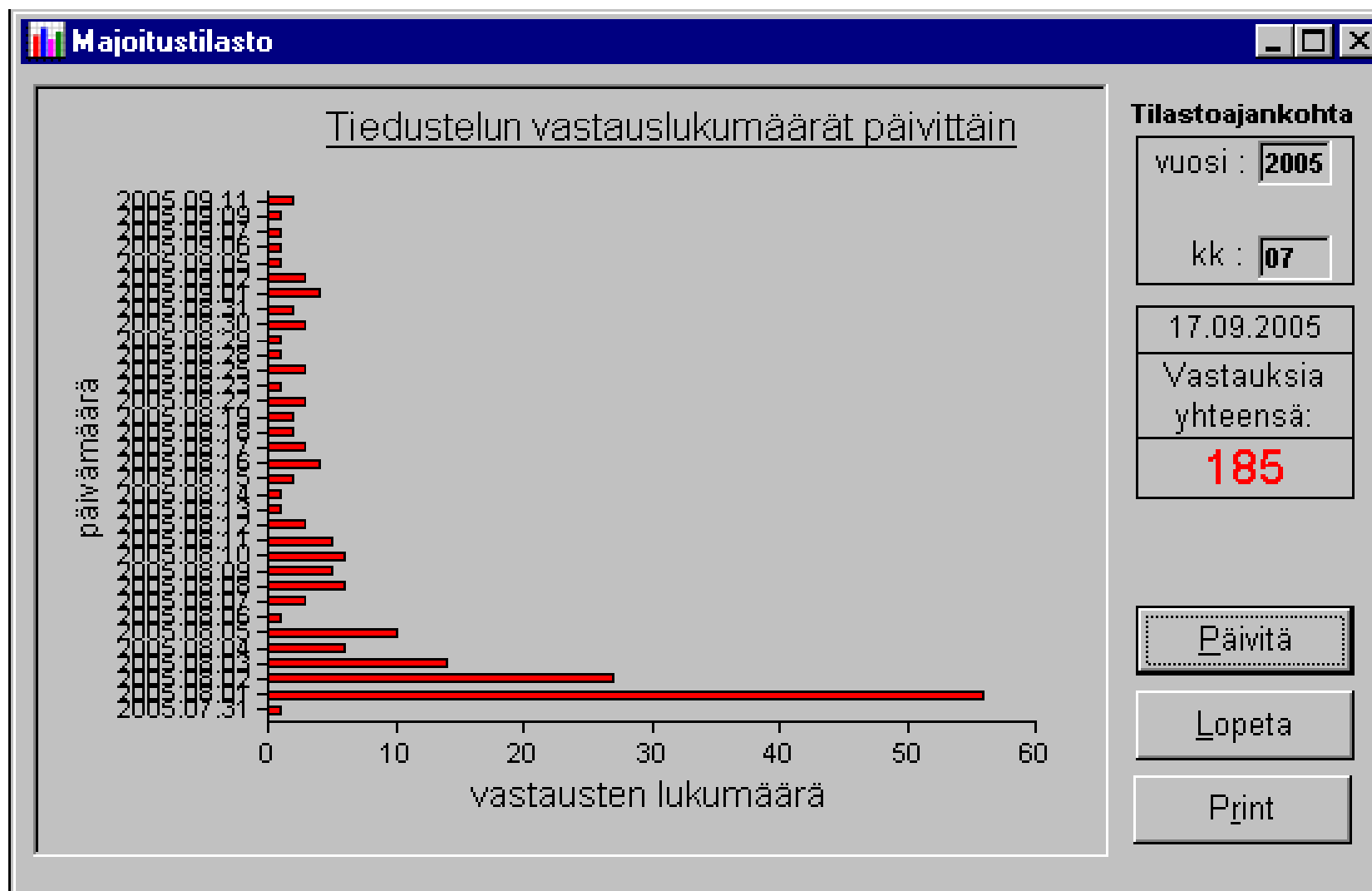


## Data collection via the Internet

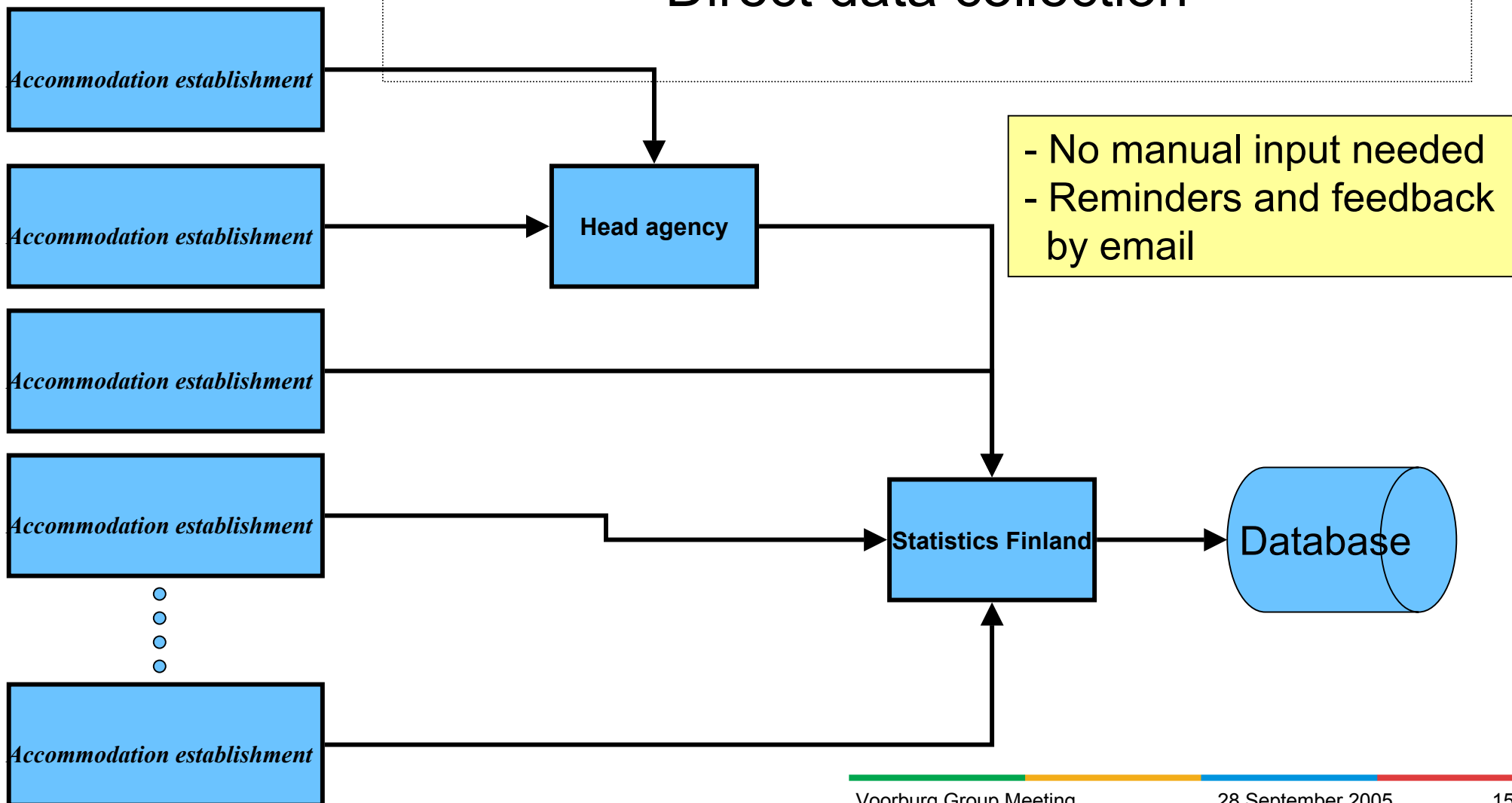
- Internet-based data collection is implemented using Statistics Finland's generic collection application
- Provides the respondent with an Internet questionnaire that can be filled with a browser
- The questionnaire contains basic validation routines for typical errors
- Each respondent has an account that can be used for Internet-based collection
  - Internet questionnaire

## Data collection for Tourism Statistics

### Response activity via the Internet



## Direct data collection



## Direct data collection

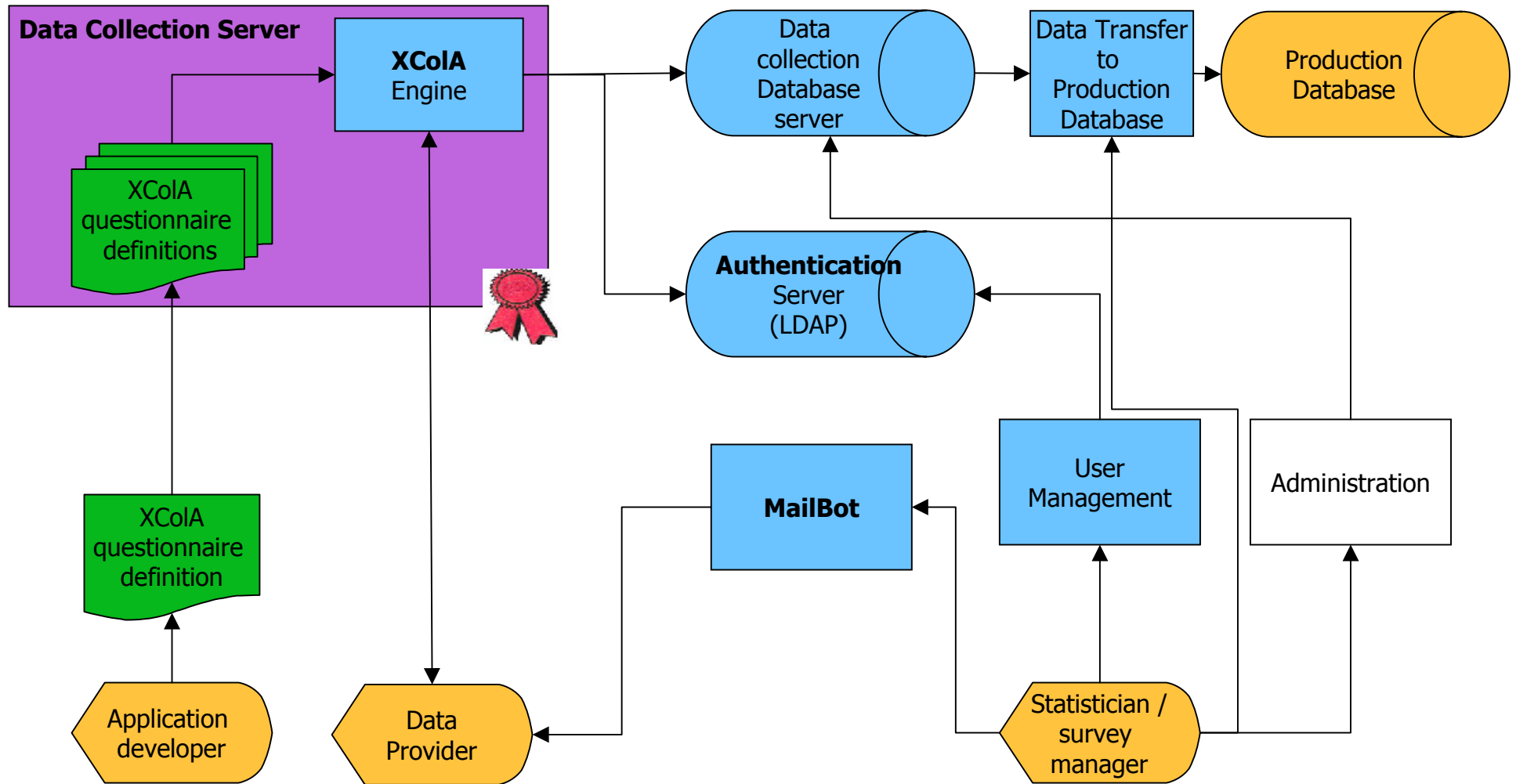
- Data are delivered direct from hotel management systems into our database
  - No manual work needed (except to initiate the transfer)
- After their reception data are submitted to the standard validation process
- Software vendors implement a module for the hotel's management software using Statistics Finland's definitions for data and service interface
- Implemented using typical B2B integration technique: XML Web Services



# Technical solution



# Infrastructure



## What is XCola?

- A generic application for Web surveys
  - Typically for surveys aimed at enterprises
- Meaning that the surveys
  - Conform with each other
  - Are developed using modern techniques and technologies
  - Can be developed “easily”

...and technically?

- An engine that reads XML questionnaires as input files and outputs them as Web forms
- Typically does a lot of processing for the questionnaires between the input and output phases
  - Enabling the dynamics for the surveys

## The power of XML

- The questionnaire described in XML can be used in different dissemination formats
- XML can be used to define the structure and the basic functionality of the questionnaire
- The functionality can be
  - Different question types (dropdown and selection lists, etc.)
  - Standard validations (comparisons, compulsory fields, data types)
  - Customised forms (pre-fill, dynamic tailoring)

## XCola in a nutshell

- Represents an example of a Web survey
- Processes the XML questionnaires and transforms them into Web applications
- Supports client and server side validations
- Executed on the server side, does not require any installation on the respondent side
- Works on every modern browser
- Extremely simple
- Easy to implement questionnaires in just hours

## References:

- Using the Web for collecting data for business statistics in Finland (Prepared by Ilkka Hyppönen), UNECE / CES Plenary Session (Geneva, 13-15 June 2005).
  - <http://www.unece.org/stats/documents/ces/2005/8.e.pdf>