



**Economic and Social  
Council**

Distr.  
GENERAL

CES/AC.71/2001/15  
12 January 2001

ENGLISH ONLY

STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE

COMMISSION OF THE EUROPEAN  
COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE/Eurostat Meeting on the Management of Statistical Information Technology  
(Geneva, Switzerland, 14-16 February 2001)

Topic (ii): Metadata modelling and terminology issues

**CONSIDERATIONS FOR THE IMPROVEMENT OF THE NETWORK ENVIRONMENT IN THE FIELD OF  
DATA COLLECTION  
e-CORD AT THE HUNGARIAN CENTRAL STATISTICAL OFFICE**

Submitted by Hungarian Central Statistical Office <sup>1</sup>

**CONTRIBUTED PAPER**

1. Internet is regarded as a new weapon in the arsenal of the Hungarian Central Statistical Office. It can be used to attack publicity with attractively presented, up-to-date and detailed data, inexpensive to publish and less expensive – compared to traditional publications – to consume.
2. But the Web also serves to defend respondents' interests in the process of data collection. There is a good chance that many respondents will feel a much lighter burden when statistical questionnaires can be filled in without paper, typewriter, pen or floppy disc.
3. The application of network tools in data dissemination at the Hungarian Central Statistical Office

---

<sup>1</sup> Prepared by Tamas Koltai.

(HCSO) dates back to the late eighties. The HCSO was able to keep pace with the rapid development in technology, and today more than 20.000 pages (screens or items to be downloaded) await the Internet user; most of them are in Hungarian, many in English, some free for all, others for subscribers only. The service is developed continuously, improving design, contents and functionality.

4. However, in the area of electronic data collection the first steps were made only at the beginning of 1998 with the setting up of an EDI service between the four largest Hungarian companies and the HCSO. Today, detailed data at the unit level of the companies for a growing number of different monthly and quarterly surveys are collected through this system. The service is available via X400, Internet E-mail and FTP. The data is secured by proprietary encryption procedures. Messages in RDRMES format are converted to in-house format to be accepted by programs originally designed for manual data entry only.

5. The EDI service is based on mutual agreements, tailor-made software on the client's side and an important piece of an EDIFACT message clearing software on the EDI server at the HCSO.

6. The new wave of e-banking, e-commerce and other members of the e-family overrode this robust technology. Both from the marketing viewpoint and to keep pace with the comfort and functionality of the new software tools in the area of e-business, new approaches are necessary for electronic data collection as well.

7. This fact was recognised in the decisions of the HCSO to turn toward Internet solutions in its EDI. It also includes a strong determination to bring electronic data collection up to the level of electronic dissemination and to merge them into a homogenous Internet solution.

8. This year the development of an Internet service began. It will correspond with the emerging requirements of the coming legislation on electronic documents and electronic signature. The pilot project is a common effort by HCSO and the Hungarian Telecom. The latter holds the role of Certification authority and WEB-EDI provider. The service will go into operation early next year. It will be available for all companies and private entrepreneurs with a choice of three different security levels.

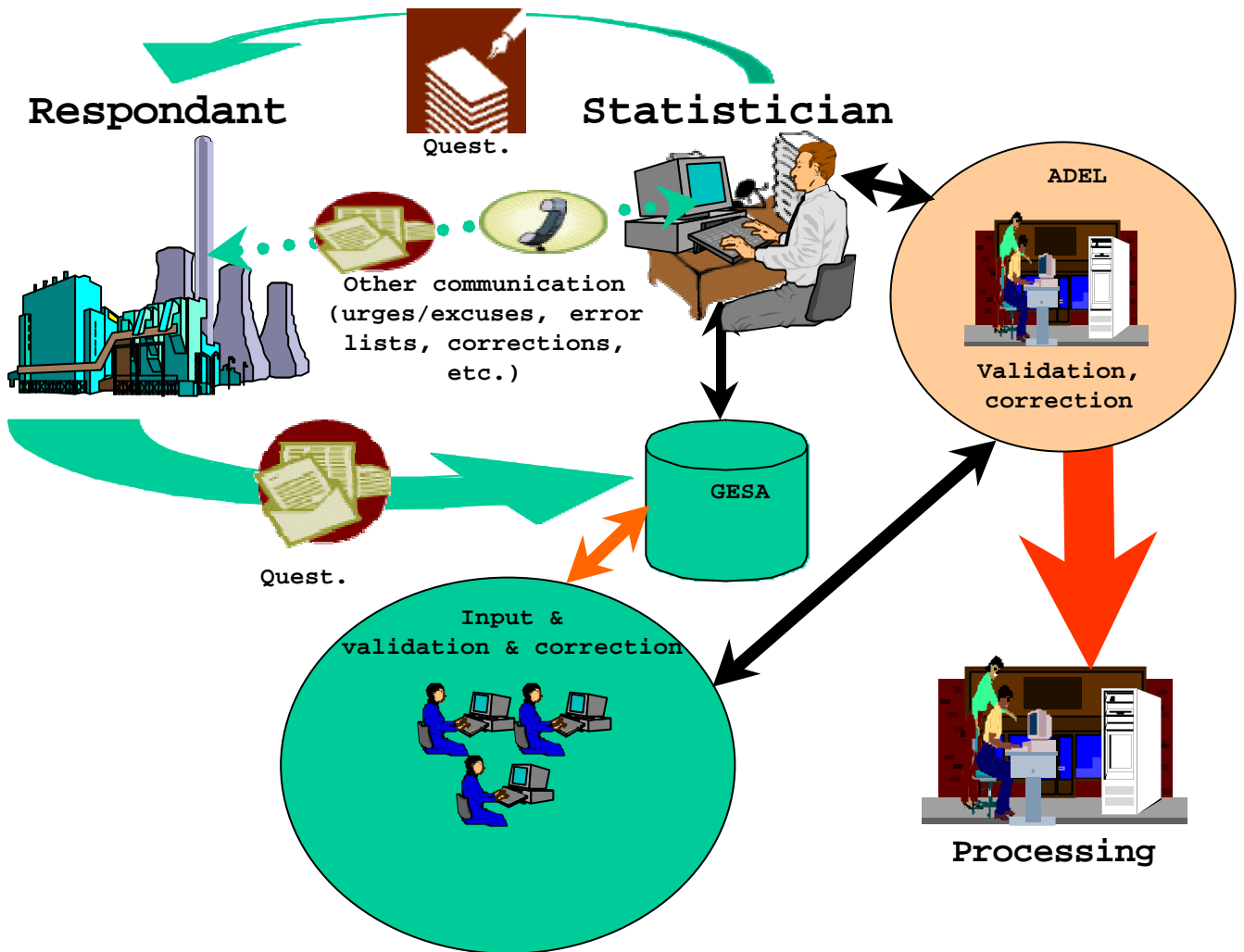
9. To match the claims from a growing number of respondents, HCSO will provide a less developed but still useful method of data collection to avoid the need for typewriters or "hand-writers" to fill in a statistical questionnaire. For the first time in reality, this method will be available for the annual infrastructure survey of 2000. The same facility can also be used from February 2001 for the monthly integrated economical survey. The EXCEL/WORD questionnaires will be available on the net to be downloaded, filled-in on a PC then printed, signed and sent in by post as usual.

10. Below are given presentation slides providing an overview of the projects and development plans in the HCSO in this area.

**Sign posts for the reader:**

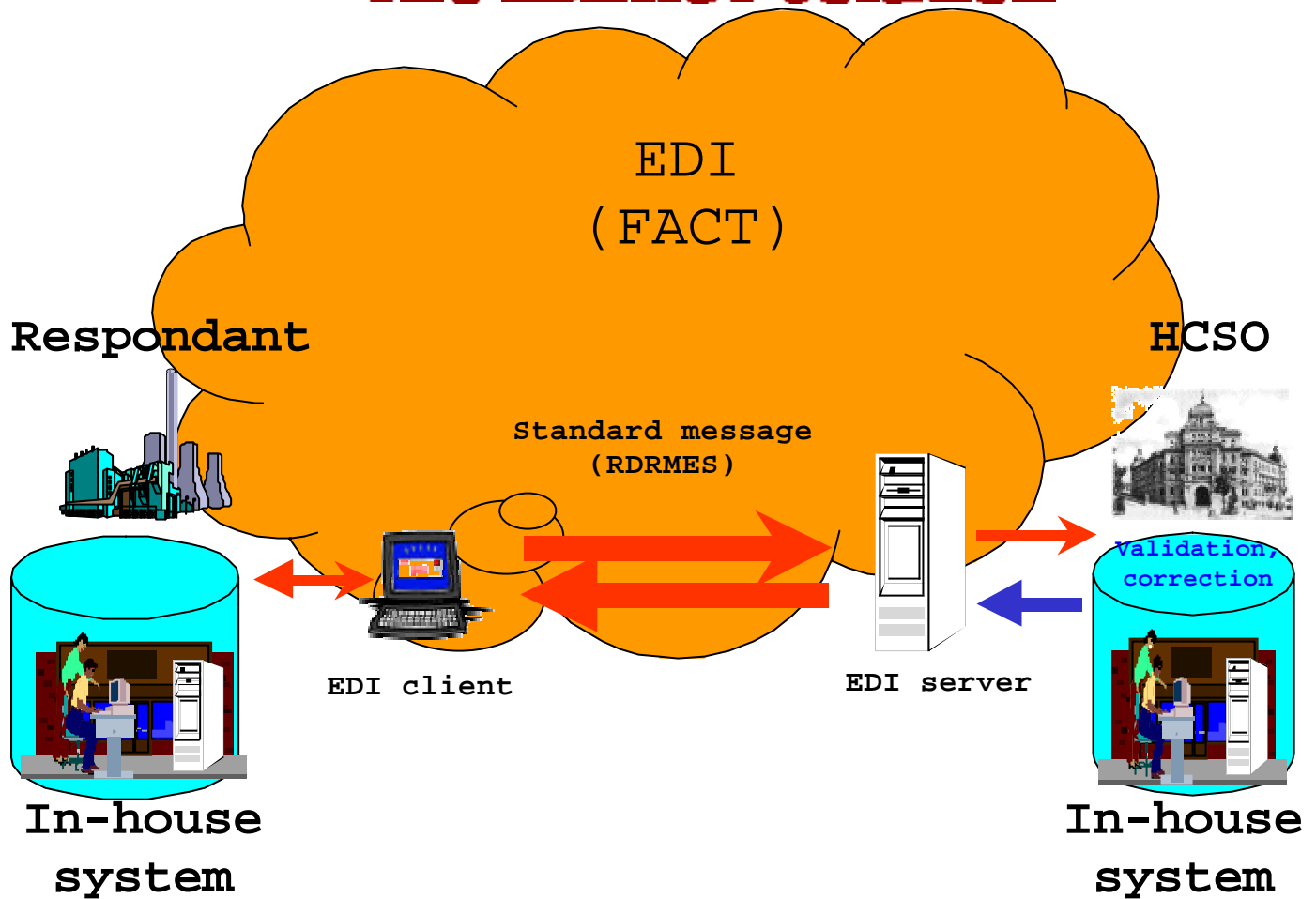
- ◆ The tradition: how we collect data at present
- ◆ The new wave: e-CORD
- ◆ Shadows in the light: certification, data security, privacy
- ◆ A pilot (pro) JECT: The Joint Effort of the CSO and the Telecom
- ◆ Half way through: web + paper
- ◆ Data collection in the present system is based on paper questionnaires
- ◆ The sample and the managing information is in a data base handled by the so-called GESA subsystem

### The traditional way of data collection



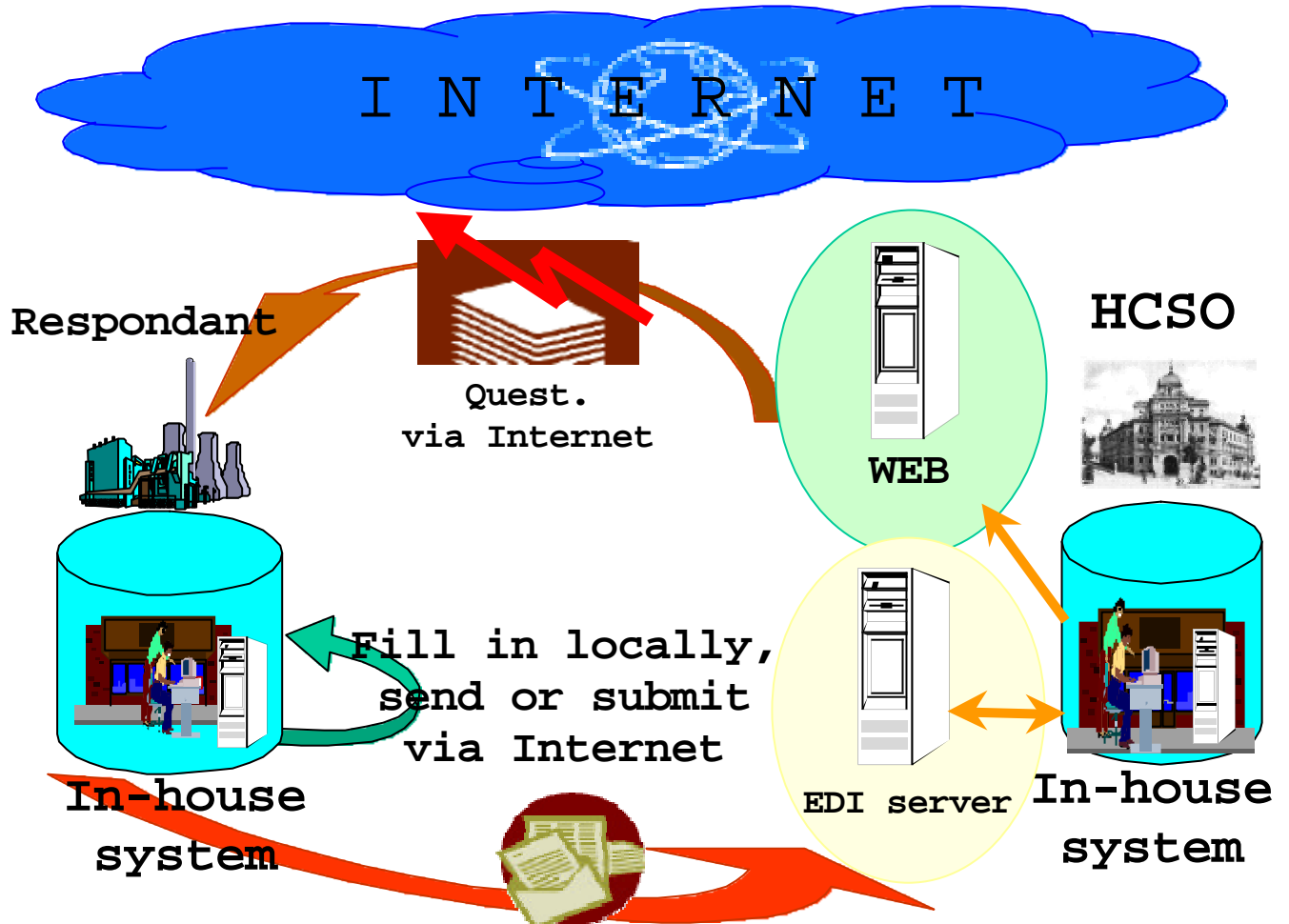
- ◆ Data input is controlled and supported by the so-called ADEL subsystem entering the data into an ORACLE database
- ◆ Electronic raw data collection (e-cord) makes the “input&validation” component unnecessary.

# The EDIFACT solution



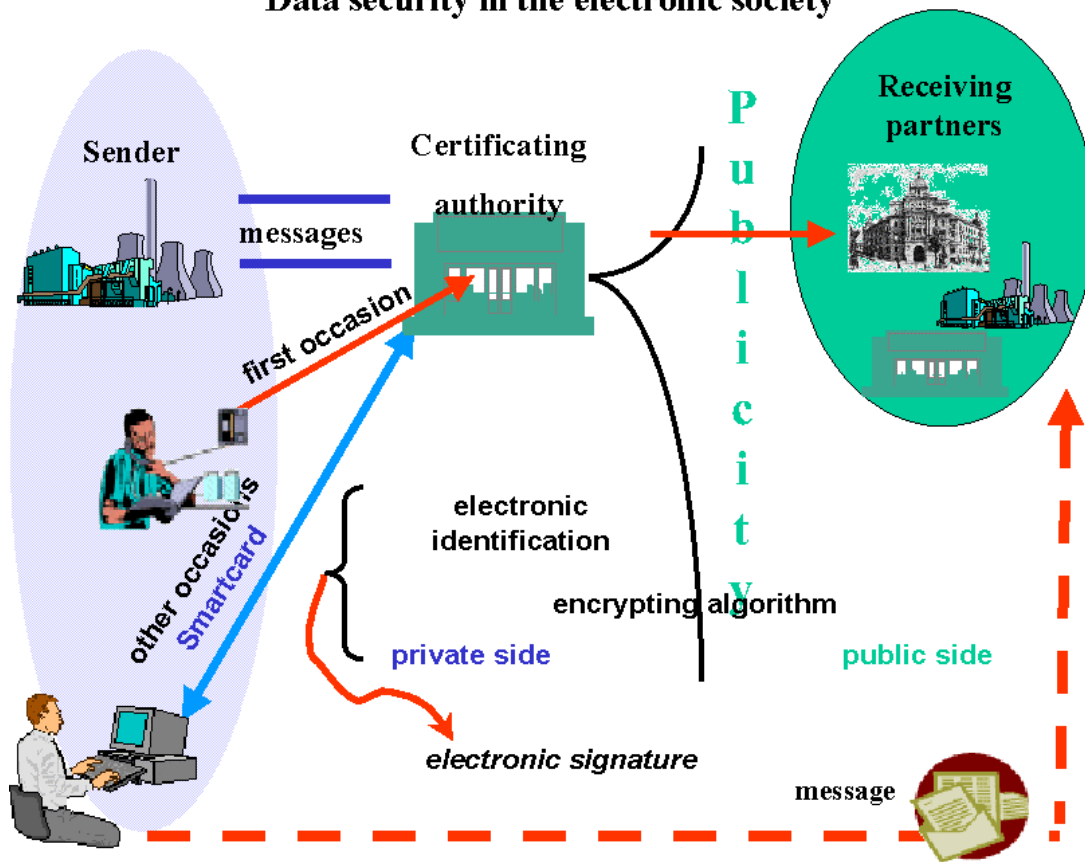
- ◆ The sad reality: two pieces of software (EDI client, EDI Server), and an EDIFACT message to be mapped to and from in both the respondent's and the CSO's in-house systems.
- ◆ The good news: no sign of clerical work.
- ◆ The bad news: to build up and maintain the communication, highly qualified human resources and much effort on their behalf are required (to modify or create in-house programs).
- ◆ The question: can we drop humans (statisticians) from this process? If not, how, when and at which point(s) should they be able to intervene?
- ◆ The EDIFACT communication has been built up for monthly and quarterly questionnaires of economic data for all units of the four largest Hungarian companies.
- ◆ More questionnaires and more companies could be involved.
- ◆ Before expanding to more questionnaires and more respondents, other methods should also be tested on a small scale.

# The Internet solution

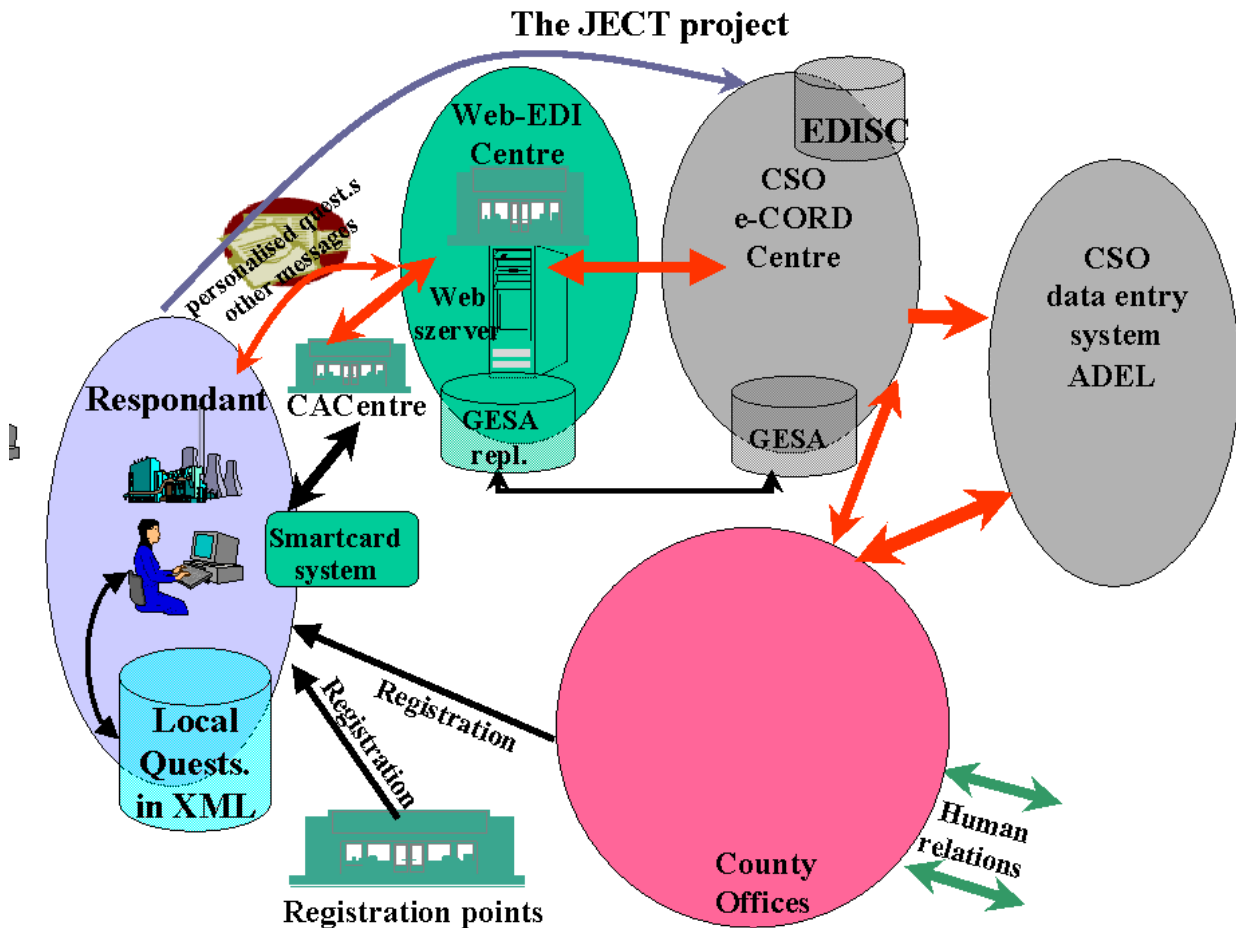


- ◆ The simple change from EDIFACT client-server solution to browser-WEB&EDI servers solution.
- ◆ The respondents - through the browser - have access to the electronic questionnaire. This supports manual input or import by scripts like JAVA, XML/XSL or other active elements interpreted by the browser.
- ◆ The end result is a message to the EDI server of the CSO. The message can be in any format, e.g. RDRMES.
- ◆ The web server beyond the personalised questionnaires may provide additional information such as historical data, relevant statistical data etc.
- ◆ In order to abolish the EDI server, the WEB server should take over all the work.
- ◆ This is WEB-EDI. The server not only reads the back-end data base but updates it as well.

### Data security in the electronic society



- ◆ We can not live without certificates, data security, privacy in an open environment.
- ◆ Premises:
  - Authentication + providing integrity + encryption
- ◆ Conclusion: "I believe the message" between the two partners within a whole community

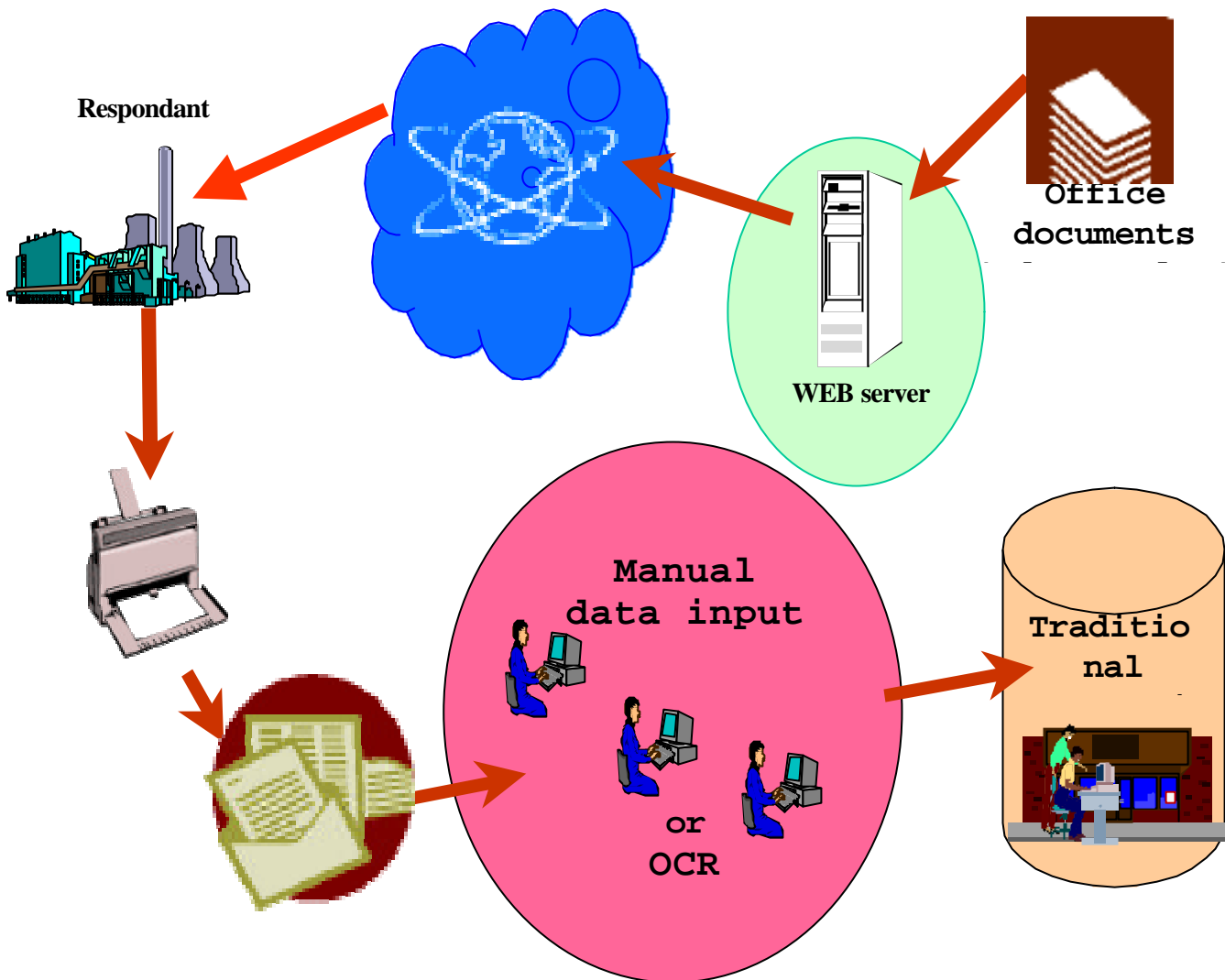


### The JECT project

#### A pilot project of HCSO and the Hungarian Telecom

- ◆ Electronic & paper input are integrated at data entry level (ADEL).
- ◆ Data security services can be high: CA with smart-card; medium: CA without smart-card (keys on floppy); low: CSO registration only with id&password.
- ◆ Semiautomatic error reports, correction either by repeating new version of data or manually as with paper.

## A simple but useful exploitation of the Internet technology



- ◆ Questionnaires can be downloaded as WORD or EXCEL files to be filled in using one of the Microsoft Office applications.
- ◆ Easy for in-house solutions to import files instead of typing in data.
- ◆ The printed questionnaire can be signed and posted in the traditional way avoiding data security problems.
- ◆ No changes required to the present production systems.
- ◆ In the case of OCR input, consider the advantage of completely printed documents vs. partially hand-written documents.
- ◆ 9 different questionnaires in the field of infrastructure statistics and 7 questionnaires of monthly economical surveys will be available at the web-site of the CSO to provide annual data for 2000 and monthly and quarterly data from January 2001.
- ◆ At a later stage E-mail will be used to send in data. Certain add-ins are also planned to help importing data from the in-house system of the respondent to an EXCEL/WORD interface. The final solution (to generate an RDRMES message in the EXCEL/WORD environment) is a better substitute for the present EDI system and an alternative to the WEB-EDI design.