

Technological Trends in Statistics: the Dutch Perspective

Wouter J. Keller
Statistics Netherlands

Marcel P.R. Van den Broecke
International Statistical Institute

ABSTRACT: Present day statistics are developing rapidly as regards the media used for input of statistical data, the media used for the output of results, and the organisation of the data structures needed for throughput. Present day trends that can be discerned in this respect are expected to accelerate. These developments of electronic data interchange will further reduce the role of paper, allow for more extensive, automatic input, will require a central database construction including metadata and “data cubes”, and will lead to the possibility of rapid dissemination of custom-tailored results.

1. Trends And Expectations in the Input of Household and Establishment Survey Data

Three types of input are used at Statistics Netherlands (SN) for Household Surveys:

- paper in the form of questionnaires,
- Primary Electronic Data Interchange (P-EDI) via floppy, tape, etc., and
- Secondary Electronic Data Interchange (S-EDI) via direct electronic transfer from sources such as Internal Revenue Service, Social Security Agencies, etc.

As can be seen in Table 1, SN’s use of paper questionnaires has declined significantly over the past ten years in favor of P-EDI. Over the next ten years, SN expects the use of P-EDI and paper to be cut in half in favor of S-EDI. For Establishment Surveys, we see a similar, if somewhat less dramatic, development.

Table 1. Percent of Data by Method of Input for Household and Establishment Surveys, Statistics Netherlands

Year	Household Surveys			Establishment Surveys		
	Paper	Primary EDI	Secondary EDI	Paper	Primary EDI	Secondary EDI
1987	75	5	20	95	0	5
1997	10	60	30	65	20	15
2007 (expected)	5	30	65	20	45	35

Different sources yield varying results on the basis of different questionnaires. The various sources provide input for one large database which in turn generates a variety of statistics as demanded by the user. A combination of P-EDI and S-EDI yields virtual censuses.

Thus, P-EDI and S-EDI input are expected to further grow at the expense of paper input, reducing response burden and input error.

2. Dissemination of Statline and Other Data by Statistics Netherlands

Statline is a database containing about 80 percent of SN data which can be accessed by anyone through a (hyper)search on (meta)data. Access occurs via:

- a one-stop database, yielding results on-line, disk, fax or paper,
- CD-ROM, and
- data warehouse containing some 300 data cubes with *region*, *time* and *branch/topics* as dimensions.

General statistical data are disseminated by SN via:

- periodicals (paper) appearing at intervals,
- thematic books (paper), some with data disk,
- disks, floppy, CD-ROM, and
- on-line — fax, e-mail and Internet.

Table 2 shows that over the past ten years, SN's dissemination of data has been slowly shifting to more electronic means. In 1987, many separate publications were based on separate databases. By 1997, many data were combined in one large database and custom-made results could be obtained on a variety of media. By 2007, SN expects the majority of data to be disseminated electronically via disks and on-line.

Table 2. Percent of Data by Method of Dissemination, Statistics Netherlands

Year	Journals	Books	Disk	On-line
1987	70	25	5	0
1997	45	40	10	5
2007 (expected)	10	20	20	50

3. Electronic Data Interchange in 2007

EDI-input requires high volume editing and allows various imputation techniques to be applied through matching of survey and register data. The output requirements are consistency and comparability. The “cube” concept allows various views on some unit database, which can be opened per type of object (person, establishment, etc.) Thus, sufficiently comprehensive input of, for example, business accounts leads with the appropriate throughput to the output of National Accounts, which can be disseminated through the various media available.

Combining surveys (characterised by many variables and few units) and registers (with few variables and many units) permits the imputation of a virtual census.

4. Technological Trends and Expectations for 2007

- Forget about paper and focus on Electronic Data Interchange.
- Leave numerous separate bottleneck structures behind and focus on integration.
- For versatility, construct metadata.
- Catch data elsewhere and focus on registers.
- Combine primary and secondary data aimed at virtual census.