

SEMINAIRE

E+; 3=! C

SEMINAR

STATISTICAL COMMISSION AND
ECONOMIC COMMISSION FOR EUROPEDistr.
GENERALCONFERENCE OF EUROPEAN
STATISTICIANSCES/SEM.43/5 (Summary)
14 January 2000

Original: ENGLISH

Seminar on integrated statistical information
systems and related matters (ISIS 2000)

(Riga, Latvia, 29-31 May 2000)

Topic I: Data warehousing and the development and use
of statistical databases in a network environment

APPLYING DATA WAREHOUSING TECHNIQUES IN A STATISTICAL ENVIRONMENT

Invited paper

Submitted by Statistics Netherlands¹

SUMMARY

1. It can be profitable when data is stored in databases according to a dimensional model (as opposed to the traditional relational model). In essence, the dimensional model is the storage of textual data in so-called dimensions tables in a denormalized fashion. The measured facts are stored in a table with foreign keys to the dimension tables. In this way, every measured fact is linked to its textual descriptions on an occurrence basis. One central fact table with its dimensions is called a datamart.

2. The Department of Population Statistics of Statistics Netherlands has concluded a pilot project with positive results using Microsoft software on Intel machines. For production purposes and ad hoc queries, a production data warehouse with two data marts has been built. The first results are a much better understanding by statisticians of the statistical material because of the tremendous power of analysis that these means provide.

3. The presentation will explain in more detail the advantages of storing data in this way and the possibilities for better analysis of statistical data using data warehousing techniques. There will be also a brief discussion about the problem of changing coding and naming schemes as well as the way in which these problems are dealt with in a data warehouse.

1 Prepared by Marton Vučan.