

Distr.
GENERAL

CES/SEM.47/21
29 January 2002

ENGLISH ONLY

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE
EUROPEAN COMMUNITIES**

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

**Joint UNECE/Eurostat Seminar on Integrated Statistical
Information Systems and Related Matters (ISIS 2002)**
(17-19 April 2002, Geneva, Switzerland)

Topic III: Object-oriented technologies, component architecture

IT-CONCEPT OF THE NEW STATISTICAL INFORMATION SYSTEM GENESIS

Contributed paper

Submitted by the Federal Statistical Office of Germany¹

I. INTRODUCTION

1. Statistical Information systems have been developed and operated in the Federal Statistical Office for many years. In our view a Statistical Information system may, in brief be said to consist of a database with statistical data and metadata, some tools for data-retrieval and access, which are able to analyse the data and generate new information out of the basic data and printed and electronic publications. Besides that in many cases there is the possibility to access the database online from outside the office. The Federal Statistical Office reported about different aspects of its "STATIS-BUND" information system at ISIS-Seminars and the ECE Working Party on EDP.

2. Although the system has worked well since its first implementation in 1976 and a lot of additional facilities and applications have been developed, e.g. access via the Internet, some new challenges have arisen, which we have only partly met with the old system. Some points in respect to subject matter or technical fields to mention are:

- after German unification the Statistical Offices of the Federal *Länder* and the Federation decided to have a common solution in the field of information systems to have a uniform appearance for the users of official statistics
- better support for quick tabulation of basic data
- better metadata-support (in the old system data were mainly timeseries which were identified and accessed by a unique number. Even though the user was led to that number by a metadata based retrieval system, the number sometimes changed, e.g. when a new item was added to a classification

¹ Prepared by Ernst Schrey (ernst.schrey@destatis.de).

- the software should be portable and scalable so as to be independent of specific hard - and software – distributors and to be appropriate for offices of both large and small *Länder*
- direct access via the Internet (in STATIS-BUND the timeseries can be ordered and downloaded via the Internet, but it takes about 15 minutes to do so)
- graphical user interface (GUI)

3. Taking into account the above-mentioned demand, we decided to build a totally new system. The following contribution describes the development of that part of the new system, which is called GENESIS and the benefit we derived from the use of client/server architecture and object-oriented technologies in that work.

II. DEVELOPMENT OF THE STATISTICAL INFORMATION SYSTEM GENESIS

A. Short description of GENESIS

4. GENESIS covers the database of statistical figures and a system of metadata including information on surveys, variables and their items, rules for deriving variables, units of measurement and the existing frames for standard tables. The thesaurus includes all relevant catchwords (terms) required to access the information designed for the general public. In addition, information of the evaluation system like table definitions, result tables and results of retrievals are stored in GENESIS.

5. As it is very important to use unambiguous metadata across different statistics and different installed systems in the *Länder*, metadata are defined and maintained centrally in a special unit in the Federal Statistical Office. This unit also disseminates the metadata to the other offices. Although it reduces the speed we can get data into the database, it is our view that it is better than a fully decentralized system.

6. As the logical datamodel we decided to take a data cube, where each dimension (axis) is represented by a statistical variable (fig. 1 shows the different types of variables); each point on an axis represents an item of the appropriate variable and each cell in the cube contains one or more value variables like local units, persons employed, etc.

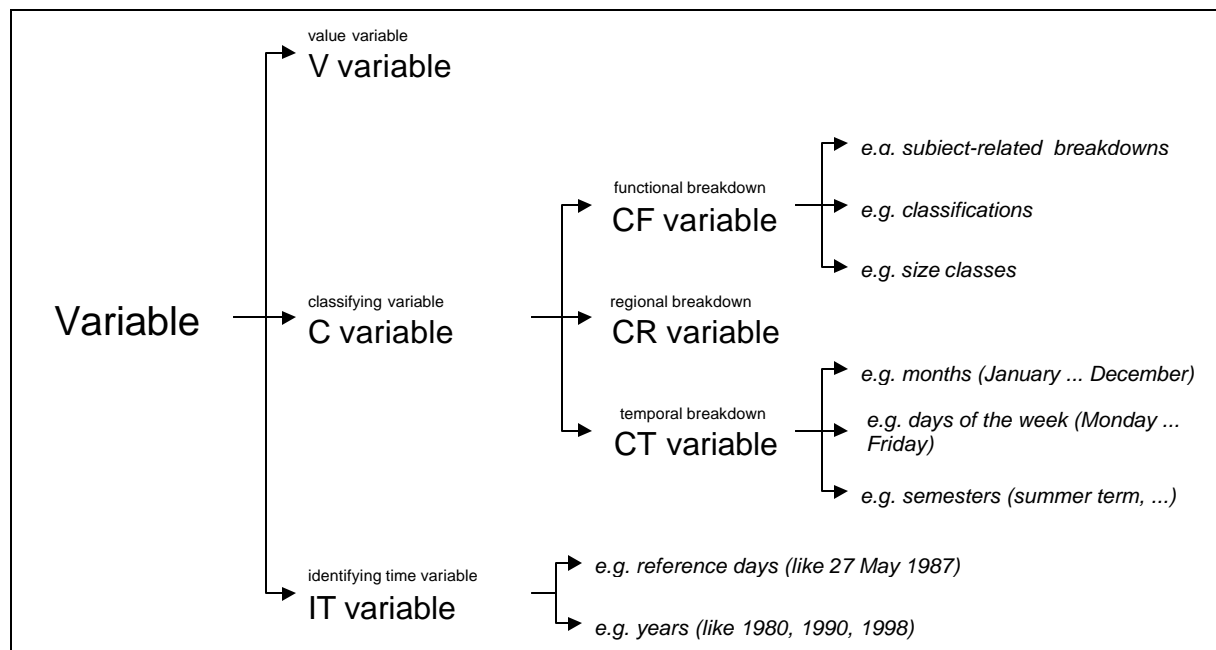


Fig. 1 Different types of Variables

7. Based on these data cubes the GENESIS table preparation component either prepares tables fully in line with user requests or uses ready-made ones, i.e. standard tables offered by the database provider. A table assistant is available to the user to propose a tabulation for a selected data cube.

B. Conditions for the common development

8. Taking as a starting point the aim of having a common system for all sixteen *Länder* and the Federation as well as the general aims of Statistical Information systems, there arose some additional conditions:

- the software developed commonly, had to be implemented on the specific hardware and operating systems in the offices (IBM, Siemens, SUN).
- requirements were very extensive as they had to cover the needs of very different units, e.g. the Federation, the *Länder* of Hamburg or Bavaria, and so in some respect they became the sum of all requirements
- the migration from the different existing systems had to be taken into account
- the amount of data to be stored in the systems differed widely
- after the development of the basic version of the system by a commercial software provider, the system was developed further by five small teams in different statistical offices
- the software development is based on the DBMS ADABAS and the developer tool Natural, as these were the commonly agreed software tools, used until that time in other programming projects in the so called Verbund of the Statistical Offices

C. Development of GENESIS

9. Within these conditions the first version of GENESIS was shipped with a menu based dialog interface (Fig.2). But although most of the facilities needed were available, there was a lack of user

acceptance as the rapid progress in information technology, e.g. the use of personal computers with graphical user interfaces and the Internet had changed user expectations since the start of the project.

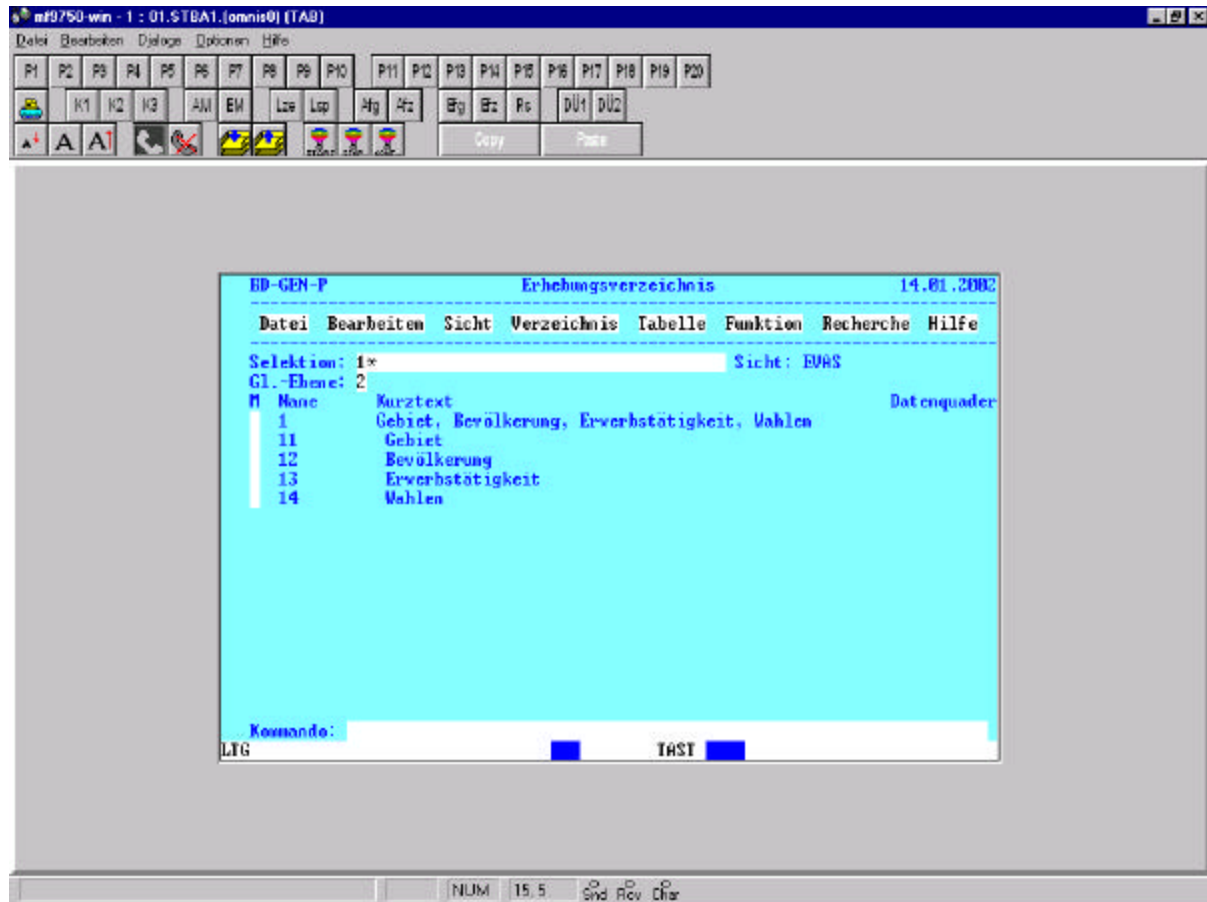


Figure 2: catalogue of surveys – menu based interface

10. Therefore the decision was taken to develop a graphical user interface in addition to and in parallel with the further development of the basic system. This development was done by the Federal Statistical Office on its own, but the results were also disseminated to the *Länder*. At that stage we benefited greatly from the decision of the developers, to opt for a Client/Server architecture at the beginning of the project, even though at that times it was planned to run the software on a mainframe and one operating system.

11. There are many possibilities concerning the distribution of the functionality of a database-, application- and presentation-layer between client and server. After different tests and development of prototypes we decided to develop the GUI with a new presentation-layer realised as a Java-client, the middleware EntireX and remaining the ADABAS/Natural components on the server. It is evident that the server-software had to be changed also in some respect. Figure 3 shows the same user dialog for the GUI as figure 2 menu based interface.

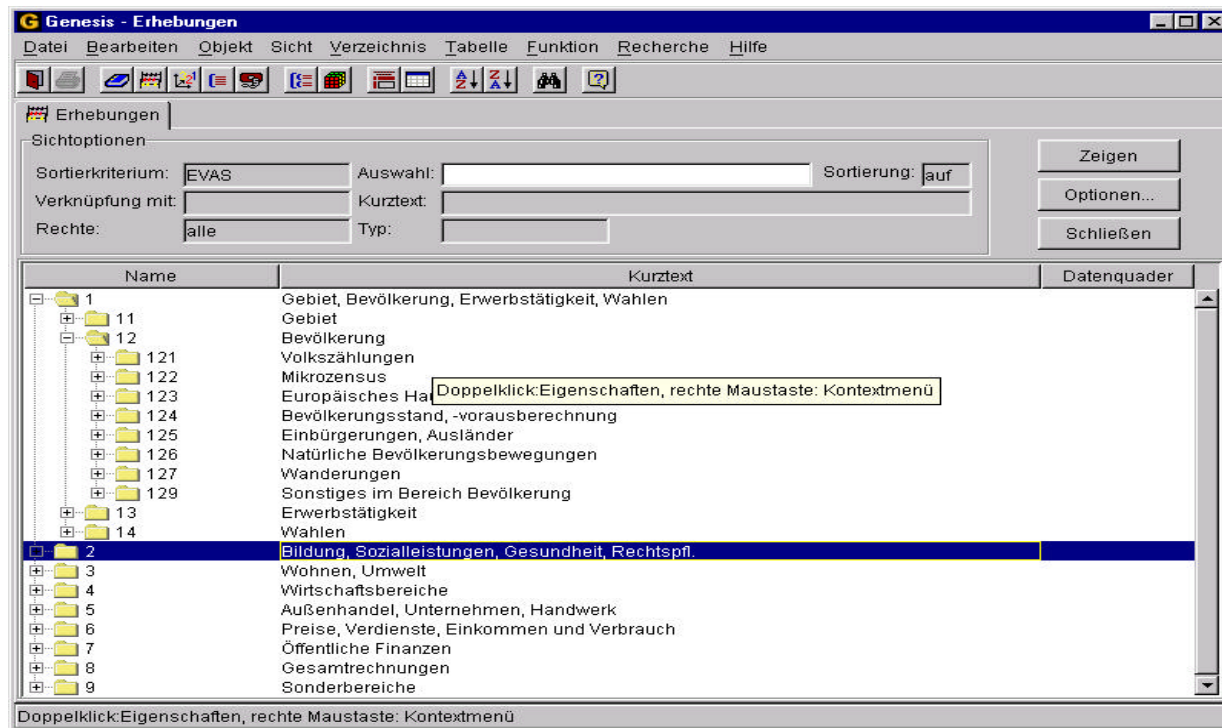


Figure 3: catalogue of surveys – GUI

12. Obviously the graphical user interface with its facilities to navigate by mouse click is more user-friendly and has additional features and information compared with a menu-driven one.

13. Even though in the first phase we planned to give access to GENESIS to all external users via the GUI, after some discussions there arose a requirement for an additional interface. This would be tailored for less experienced external users having access via the Internet. At the same time tools for data access and retrieval would be restricted to the standard tables, which can be changed only slightly by the user. In this development, the user interface is HTML-based and uses a general XML-interface, which was designed during the development of the GUI. For reasons of security this client will get access to a special UNIX-based installation of GENESIS, which holds only part of the data and metadata. It is called GENESIS-Online. In figure 4 the same application as figure 3 is shown for GENESIS-Online.

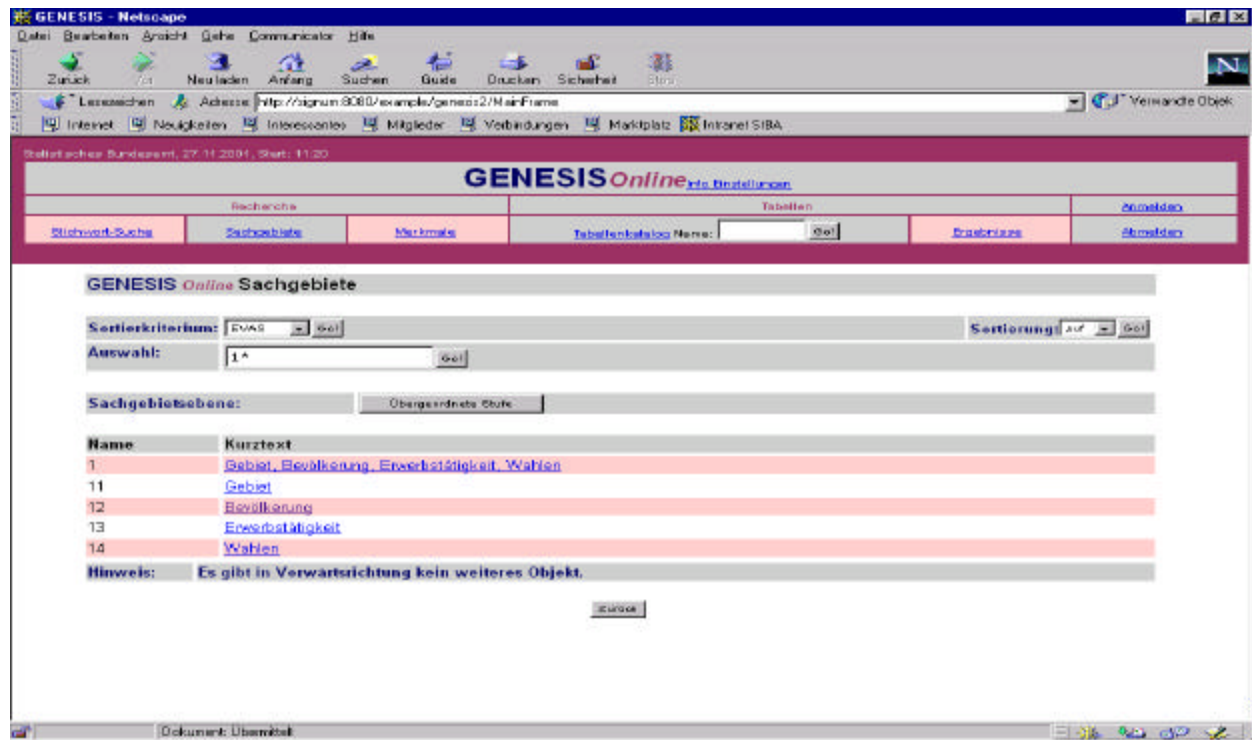


Figure 4: catalogue of surveys – GENESIS-Online

14. Figure 5 shows the different types of users and user interfaces depending on the demands and knowledge of the user and the different communication network used, which are now available.

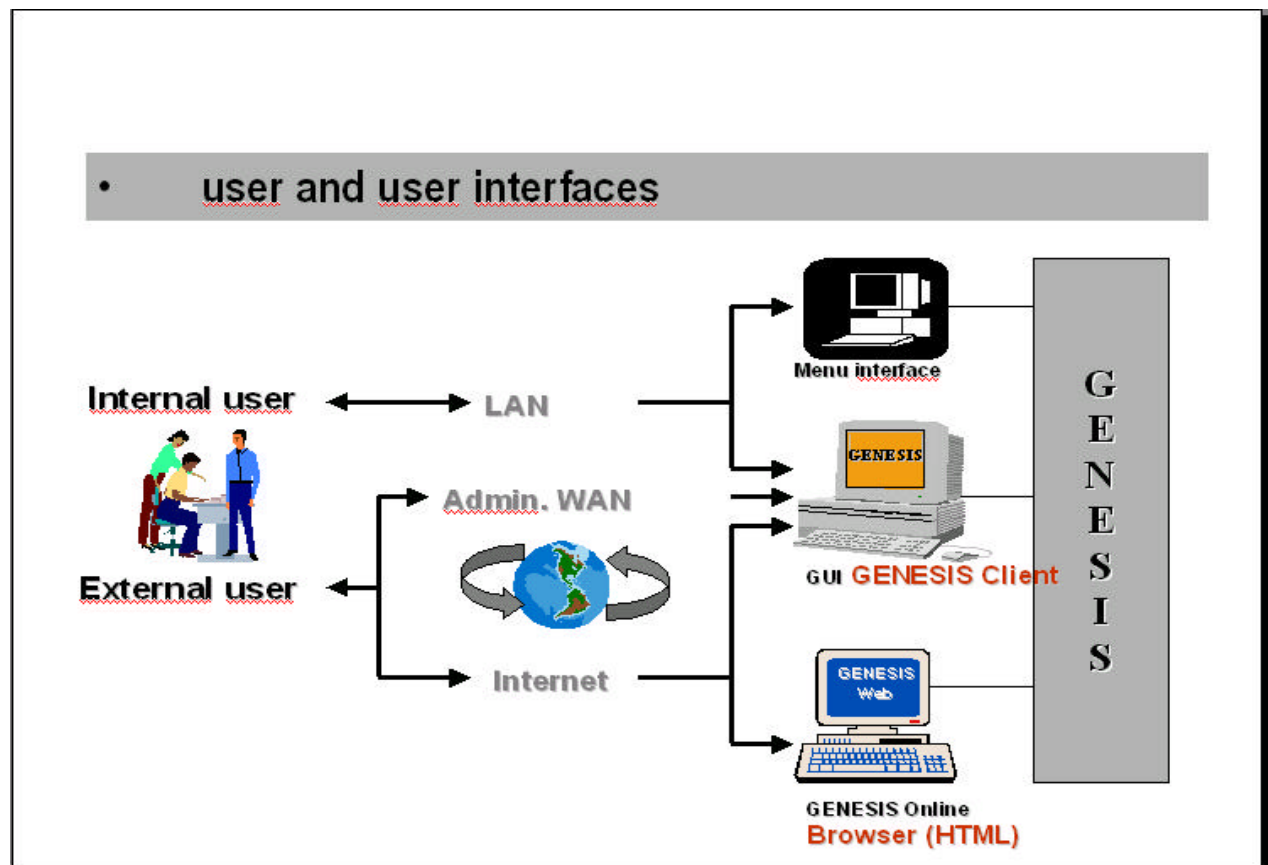


Figure 5

III. CONCLUSIONS

15. What were the lessons learnt during the planning and development of GENESIS? The development of a Statistical Information system is an ambitious project in itself. Additional requirements arose from the multitude of stakeholders involved and the decreasing budget and human resources. Nevertheless the use of new technologies like object-oriented tools, web-technology and client/server architecture enabled us to cope with these problems, as they facilitated a very efficient and flexible development process. The software of the new information system with its different user interfaces is both portable and scalable and therefore appropriate for the differing demands of all the Statistical Offices and applicable to experienced users and people “who need only one figure” as well. It should be mentioned however, that the qualifications of the personnel involved in the development have to be of a high order.

16. Besides speed and flexibility of development the new technologies, especially component architecture, enable new models to be developed organise the work in the decentralised statistical system in Germany. So, for example, using the detailed and advanced access and protection system of GENESIS it would be possible to install different “virtual” data warehouses for several *Länder* within GENESIS on the one server. Although the clients would remain decentralised in different locations, the costs of installing and maintaining the basic hardware and software would have to be done only once.

17. Steps have also been taken to integrate the commercial software package for evaluation, analysis and reporting implemented in the Federal Statistical Office via the client interface to enable access from the evaluation system to GENESIS data and metadata. So the data stored in GENESIS can be used to

generate additional and new information, for example by seasonal adjustment of a timeseries or regression with categorical variables based on a table from the database. The first version is running successfully and we will continue these developments to integrate different applications in the Office.