Design of the User Interface for an Object-oriented Statistical Database

Erik Malmborg

SCB

R & D Report
Statistics Sweden
Research - Methods - Development
1988:11

INLEDNING

TILL

R & D report : research, methods, development / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1988-2004. – Nr. 1988:1-2004:2.
Häri ingår Abstracts : sammanfattningar av metodrapporter från SCB med egen numrering.

Föregångare:

Metodinformation : preliminär rapport från Statistiska centralbyrån. – Stockholm : Statistiska centralbyrån. – 1984-1986. – Nr 1984:1-1986:8.

U/ADB / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1986-1987. – Nr E24-E26

R & D report : research, methods, development, U/STM / Statistics Sweden. – Stockholm : Statistiska centralbyrån, 1987. – Nr 29-41.

Efterföljare:

Research and development : methodology reports from Statistics Sweden. – Stockholm : Statistiska centralbyrån. – 2006-. – Nr 2006:1-.

R & D Report 1988:11. Design of the user interface for an object-oriented statistical database / Erik Malmborg.
Digitaliserad av Statistiska centralbyrån (SCB) 2016.

urn:nbn:se:scb-1988-X101OP8811

Design of the User Interface for an Object-oriented Statistical Database

Erik Malmborg

SCB

R & D Report
Statistics Sweden
Research - Methods - Development
1988:11

Published	August, 1988
Printer	STATISTICS SWEDEN, Dept of Research and Development, EDP Systems Unit
Publisher	Staffan Wahlström
Questions answered by	Bo Sundgren

ISSN 0283-8680

Printed in Sweden
Garnisonstryckeriet, Stockholm 1988

DESIGN OF THE USER INTERFACE FOR AN OBJECT-ORIENTED STATISTICAL DATA BASE

Erik Malmborg
Statistics Sweden, U/ADB
S-11581 STOCKHOLM Sweden

Abstract

The concept of an Object-Oriented Statistical Database (OOSD) is rather new. The approach taken in this paper unifies two different traditions leading to OOSD. The first tradition is from Object-Oriented Programming Systems and languages in the Smalltalk tradition. The other tradition is from database systems able to handle more complex objects than records (or segments). Examples of such systems can be found in e g CAD/CAM and Office Information Systems.

The emphasis of this paper is on the user interface. The development towards modern Window-Icon-Mouse interfaces is described. The specific interfaces to be designed are a graphical meta-data browser, a graphic table-design language and the interaction between these.

The paper includes an analysis of different approaches for the interaction between query (or table-design) languages and meta-data handling, leading to the proposed "Macintosh-style" interaction.

0. Introduction and structure of paper

This paper describes parts of the design of a planned software system named TBE-2. TBE stands for Table By Example and alludes to the famous IBM language Query By Example (described in e.g. Date-85). At Statistics Sweden there has been a great deal of discussion on how to extend the ideas from QBE into a Table Design Language. TBE(-1) (Nilsson-84) is a result from these discussions. TBE-1 is an algebraic language working on boxes (matrices). It is theoretically based on the box-theory of Sundgren-73. A modern interactive user interface was planned but never materialized. ABE (Klug-82) has a user interface similar to that planned for TBE-1. QBE, TBE-1 and ABE are based on text terminals with full-screen handling. TBE-2 is based on the technology of bit-mapped graphic screens and will have a "Macintosh-style" interface.

The paper is divided into six sections. The first three give background material on Object-Oriented Statistical Databases, Macintosh style user interfaces and on the interaction between statistical query languages and meta-data.

Sections 4 - 6 describe the design of the TBE-2 user interface. Section 4 describes the graphical meta-data browser. Section 5 describes the table design language. Section 6 contains further material on the interaction between table design, meta-data and data. Some indications of further plans are given at the end of section 6.

1. The meaning and goal of an Object-Oriented Statistical Database

There seem to be two traditions leading to the concept of an Object Oriented Statistical Database (OOSD):

- Traditional databases handle records (segments, tuples . . .), i.e. simple concatenations of terms. In many areas there has been a demand for databases handling more complex "objects". Such areas are e.g. VLSI design, CAD/CAM and picture analysis. The storage/retrieval of statistical tables (matrices, boxes etc) is a similar problem. The emphasis is on data base (on secondary storage), and the demand comes from the users. This is a problem in search of a solution. Scientific papers in this tradition normally have references to papers on "data structures for VLSI" and/or "CAD databases". An example from VLDB-85 is "An Extensible Object-oriented Approach to Databases for VLSI/CAD" by Afsarmanesh et al. (Afsarmanesh-85)

- The concept of object oriented systems (not specifically "databases") has evolved from programming languages (SIMULA-67, Smalltalk . . .) and user interfaces (Smalltalk, MacIntosh . . .). The "object" in such systems are further developments of "procedures" and "classes" from earlier languages. A basic idea is that an object (class) should handle all operations on the instances of that class. This gives highly modular systems. The storage/retrieval on secondary storage of the instances could be examples of such operations, but the emphasis is on "object-oriented". The "object-oriented" approach in this sense is almost a solution in search of problems. Papers in this tradition typically refer to e.g. the Rank Xerox Smalltalk project (Goldberg, Kay, Ingals . . .). An example from VLDB-85 is "An Object-Oriented Environment For OIS Applications" by Niestrasz/Tsichritzis (Niestrasz-85).

To me it seems that these are two traditions, possibly leading to research on "Object Oriented Databases" in general and "Object Oriented Statistical DBMS" in particular.

A simple solution to the problem of the unclear concept is to propose a definition that is pertinent to both traditions above. This is basically what is done in the report from the working group on Object Oriented Statistical DBMS from the third conference in this series (Luxembourg July-86). The report can be found in Statistical Software Newsletter 1,1987.

The system to be designed is somewhere inbetween a Data Base Management System and a tool for System Development. Typically the requirement on a DBMS is to give the user an answer within seconds. In the case of complex questions or reports minutes might be acceptable. System Development, on the other hand, is normally considered a complex process with users and analysts together building an application giving the required functionality. This process takes weeks, months or even years.

In the tool to be designed users and/or analysts should be able to create applications giving the desired information within hours or perhaps days. Technically the system is not to include its own database format, but generate programs able to use existing file formats. The technical aspects of program generation are not the topic of this paper, but were discussed in Malmborg/Chowdhury-84.

2. The development towards Macintosh-style user interfaces

The Macintosh computer from Apple has become the symbol of a new style of man-machine interface. This is somewhat unfair to XEROX, as the major research and development behind this type of "WIMP"-interface was made at the XEROX PARC laboratories. The Smalltalk system from XEROX (Kay-77, BYTE-81, Goldberg-83,-84, Krasner-83) is a major development. As will be elaborated later in the paper Smalltalk is the language of choice for my project. STAR (Smith-82) is another important XEROX-system on the way to modern "WIMP" interfaces. Smith-82 is strongly recommended reading for those interested in the goals and strategies for WIMP-interfaces. WIMP stands for:

- Windows, i.e. the dividing of the CRT-screen into several areas, and the use of flexible methods for moving, opening, closing, enlarging the windows.

- Icons are symbols on the screen, denoting the different "objects" hidden inside the computer (programs, data...).

- Mouse is the device for pointing on the screen.

- Pull-down or Pop-up menues are flexible ways for giving commands to the computer (or to the objects chosen by pointing to an icon and "selecting" it before pulling down or popping up the menu).

There are further aspects of this type of interface than can be seen on the WIMP-surface. The interface should e.g.

- be mode-less, i.e. the user should not be locked into modes as in a hierarchical menu-system.

- use a familiar conceptual model (as e.g. the desktop metaphor of the Macintosh).

- be What You See Is What You Get (WYSIWYG)

The developments in this area are far from complete. IBM's Systems Application Architecture (SAA) will probably in a few years provide this type of interface to the users of IBM mainframes, office computers and personal systems.

There is a rather strong coupling between WIMP-interfaces and Object Oriented Programming (OOP). In OOP, programming is by sending messages to objects rather than by invoking procedures. In a WIMP-environment objects are selected by pointing to the icon representing them. The appropriate command is then chosen as an option in a menu. Cox-86 gives a good discussion on the operator/operand model of traditional languages (programming and command) versus the message/object model.

3. The interaction between statistical query languages and meta data

The project described in this paper is heavily influenced by two rather different papers. Both share however a common understanding of the problems with traditional query or table design languages against statistical data bases. This problem view is also similar to that of e.g. Shoshani-82, Chan-80, McCarthy-82.

The sheer volume and complexity of metadata needed for finding appropriate data and formulating a query against a large statistical database puts extreme demands on the user. If there is no integrated meta-database, the documentation of the contents in a census, environmental, world trade or other large statistical database consists of hundreds of pages.

One solution to the problem is to create an integrated meta-database used by the Data Base Management System to drive a menu-oriented interface for the user. This relieves the user from much search in the written database documentation. Solutions in this category can be found in the SUBJECT-systen (Shoshani-82, Chan-80) and in the AXIS-system (Nordbäck-82,-83).

Another slightly different solution is to have an interactive meta-database where the user can search for relevant data, and then specify his questions or table needs. The query language and the meta-data system can be integrated. The system designed in this paper emphasizes such integration.

One of two main influences mentioned above is Sato-86, which describes an implemented statistical database for the Japanese National Land Agency. When retrieving information from this database the user is helped by the Land Information Dictionary, i.e. the meta-database. This meta-database is structured by using a semantic data model based on the extended relational data model RM/T (Codd-79, Date-85). In the Land Information Dictionary the user can search ("browse") the data on two levels.

The first level is the "conceptual" level where all conceptually obtainable data can be found. This corresponds to all possible questions that can be formulated using the concepts introduced.

The second level clarifies which parts of the conceptual data correspond to data actually stored in the Data Base. One reason for the distinction is that often no complete cross-section table exists, but merely a number of subtables. This is due to the demands of statistical accuracy or the convenience of survey routines (Sato-86).

One important aspect of Sato-86 is the handling of aggregate data on the semantic level. The key idea is not to use any specific constructs such as summary sets (Johnson-79). A similar approach was suggested in Malmborg-86.

In Sato-86 the potential explosion of the number of different objects (or entities) is handled by only modelling the most detailed cross products (of the category attributes) on the conceptual level. This anyhow allows questions and data for the physical database to work on all subsets of the set of category attributes in the conceptual file. The figure below (fig 6 from Sato-86) illustrates this.

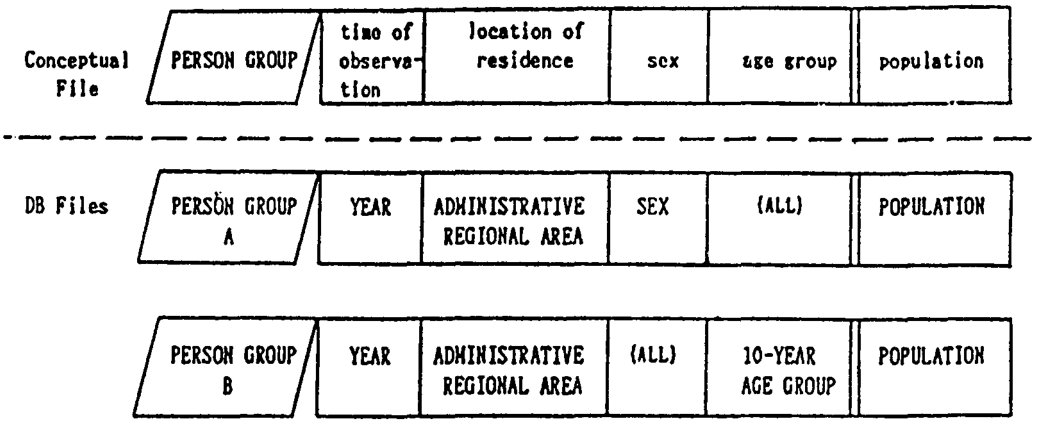


Fig.6 A Partial DB schema of Sample Statistical DB

The second paper that has had a major influence on my project is Wong-82. The system described in that paper is a Graphical User Interface for Database Exploration (GUIDE). The system is an Interactive Metadata Browser, where the user can interactively explore a data model presented on a graphical display. The data model is based on a version of the Entity-Relationship model (ER-model, Chen-76, Teoray-86). The user has options for building his own graphical view of the part of the database relevant for him. He can set schema detail, change focus, change radius, hide, zoom and move schema (see Wong-82 for details). When the view is ready it can be used for queries against the data base.

The GUIDE schema is a "normal" ER-model. There is no provision for summary sets and aggregate functions. The paper indicates plans for introducing summary sets referring to Johnson-79.

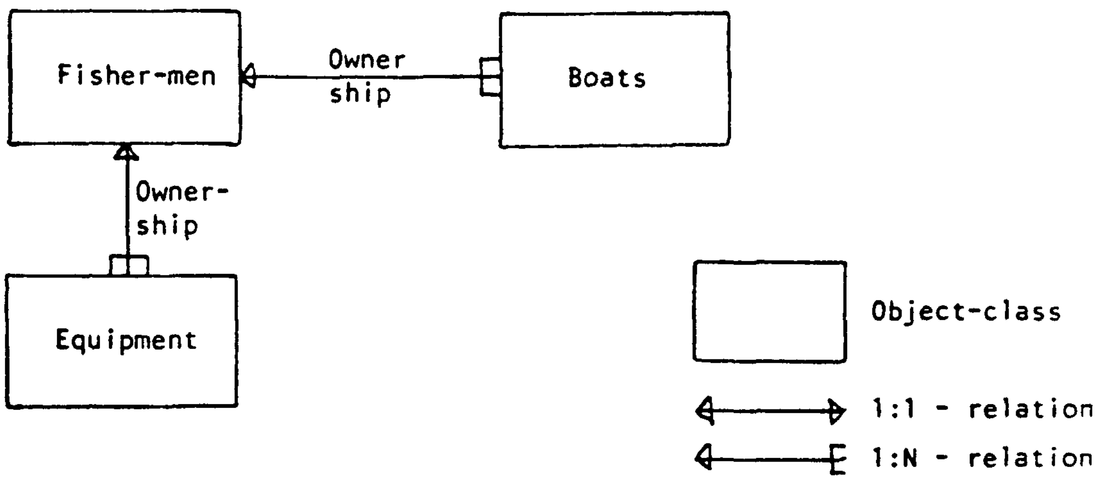
4. The graphical metadata browser

This and the next two sections describe the planned software system. The project is in its early phases and only simple "mock-ups" exist. Where the project will end is a matter of financial funds. The planned software system in this first phase will be used as a demonstration tool to create an interest in later phases. The prototype will be fully executable, and will show some main ideas of the project. As the system is programmed in Smalltalk-80 with the superior facilities of this language for structuring and reusing code there is good hope of beeing able to reuse code from the prototype in the final system. The prototype will work within the Smalltalk environment, with its drawback in the form of rather low speed. As the power of PC's and workstations grow this problem might vanish. For reasons of portability the final system will however generate code in the language C rather than work in the Smalltalk environment. The tool and the metadata will still be in Smalltalk.

The first part of the system to be described is the Interactive Graphical Metadata Browser. At Statistics Sweden we have an established System Development Model (the SCB-model, described in Sundgren-84, Malmborg-82,-83,-84). The SCB-model is similar to the ER-approach (Chen-76), but has its own roots (Sundgren-73). The graphic notation is somewhat different from the ER-approach. The SCB-model has gradually evolved towards a mix of an Extended Entity Relationship model (as in Teorey-86) and Jackson structured Design (as in Jackson-83). The models introduced in the first version of TBE-2 (the name of the tool, cf Nilsson-84 where TBE(-1) is described), will correspond to the simple ER-models.

The SCB model contains a special notation for specifying statistical tables. The user of TBE-2 will not use this "alpha/beta/gamma-notation", but it will be used in the explanation of the tool. The following extract from Malmborg-84 contains a short presentation of a simple example:

> In the most simple situation we have an information base containing static information on a population of entities. The entities are grouped into disjoint classes. The input is obtained by observation (including questionnaires). There are no dynamics to be considered. All input and output are by problem definition "snapshots". This type of problem exists in practice, and can be handled by e.g. a simple object-oriented approach.
>
> The SCB System Development Model is seen to be fully adequate for this class of situations. As an example I will present a (simplified) model used for a survey of fishing and fishing equipment.



Note that this simple "infological" model contains a lot of agreement on semantics:

- It is assumed that each boat has a single owner. This is for the purpose of registration. Of course it might not be true in an economic sense.

- Boats and Equipment are seen to be disjoint classes of entities. Of course this is only an agreement on concepts made for this survey. In another situation they might be considered as different entities from a common class.

The output from the survey will be statistical tables. The content of these can be expressed in e.g. the alpha/beta/gamma-notation (Sundgren-73) which is a part of the Systems Development Model. An example:

α -part (selected population) : fisher-men in fresh-water (lakes)

β -part (selected variables) : number of equipment

γ -part (distributed by) : equipment type and district

	seine	trawl	fyke net	...
Lake Vänern	100	...	...	
Lake Vättern	...	...	...	
Lake Mälaren	...	...	...	
...				

TBE-2 will contain a graphic editor for creating the infological models. When using the browser, objects (entities) are selected by moving the pointer onto the symbol and clicking the mouse-button. The result will be a pop-up menu where different options will be given:

- Present a verbal description of the object class.
- Present a list of files (or relations) from the database representing extensions of the object class. Each file is described in several aspects (population, level of detail ..)
- Present a list of variables (attributes) for the object.

By clicking on a variable in the last list, a list of categorizations for the variable pops up on the screen. In our example geographic location might be categorized by Lake District (as in the table above) or Region code. The use of generalization hierarchies for grouping of category attributes is discussed in Sato-86.

The categorizations are presented as lists of values (the lake names in our example). These lists correspond to the values (texts) in the table's column- and row labels (stub and heading). This visual aspect will be used in the next section.

5. The table design language

The idea for the table design part of the program is based on an observation of the similarity between a "tiled" window in Smalltalk or on the Macintosh and a statistical table. Some common aspects:

- both can be seen as a rectangular area subdivided into a number of rectangular parts. In this paper the outer rectangle is called a window, and the inner rectangles are called panes.
- The contents of the panes within a window are interdependent. On the Macintosh we might typically have two panes, one with a grapical and one with a textual representation of the same model (or document). When we move around in the model both views (panes) are updated. The same is true for a statistical table. If we change the chosen values for the stub, the central pane of the table should be updated.

	heading
stub	central pane (table pane)

Fig 5.1

	large boat		small boat	
	trawl	net	trawl	net
Lake Vänern	...	...	...	..
Lake Vättern	...	...	..	.
Lake Mälaren	...	...	...	...
:				

Fig 5.2

The way for the user to work is to pop up an empty table-window with panes for stub, heading and a center pane (for the figures to be computed). Additional panes may be introduced for table title, footings, etc..

To specify the contents and structure of the table the user browses around in the metadata-window (cf section 4). When the appropriate set of labels for the stub or heading is found, it is selected in the metadata window. Selections are by default indicated by text inversion both in Smalltalk and on the MacIntosh. Moving this selection into the table window is accomplished either by cut & paste (standard MacIntosh technique) or by dragging a symbol between the windows (another standard technique). Both alternatives will be tested to find out which is most natural to the user. The process is repeated for each gamma-variable (category attribute). If more than two gamma-variables are used, the stub and/or the heading is structured into the appropriate number of levels (cf fig 5.2 where the heading is 2-level). The number of levels is chosen from the pane's pop-up menu.

The appropriate function for computing the cells (count, sum, mean ...) is chosen from the pop-up menu of the central pane in the table-window. If the beta-part is based on a variable this is chosen in the meta-data window (e.g. age in a table with mean age of equipment).

If the table is more complex and contains several concatenated subtables and/or totals the approach can be extended by more options on the table-pane's menu.

The goal is to create a very intuitive and natural interface for the user. A key point is that the "visual" finding of the needed labels for the table (stub or heading) in the meta data makes the user feel confident in his design.

To describe this type of interaction in a text is awkward. The idea is that the user should need only minimal documentation, but find the way of specifying the information needs (the table) obvious. The problem of describing user interfaces of this kind is well known (Goldberg-84, Schmucker-82). The specification of the "intuitive"

MacIntosh user interface standard in Schmucker-86 is a hefty 50 page appendix. The process of developing this kind of interface from the early XEROX-research into the present MacIntosh standard represents more than 10 years of intensive research and development.

In other words, it is surprisingly difficult to design a really simple and intuitive user interface. Further experience shows (Smith-82) that the designer and the user often come to different opinions on what is simple and intuitive. Of course the user's opinion (by definition) is correct, but it shows the need for prototyping and experiments.

6. The interaction between table design, meta data and data.

In the preceding two sections I have presented a simple interaction between the meta data (the browser) and the table design. Actually I think this is only a first level of interaction. As was discussec in section 3, Sato-86 and Malmborg-86 makes the proposal that aggregate data (summary sets in Johnson-81) should be represented on the conceptua level as orinary entities, rather than as special modelling constructs (such as summary sets).

Aggregated data can both on logical and physical level be looked upon in two different ways. Logically one alternative is as entities (objects) or relations (Codd-meaning). The other logical alternative is as matrices or boxes (Sundgren-73). In the first alternative category attributes are explicit, in the second implicit (i.e. part of meta-data rather than data). The same two choices exist on the physical level, giving four possible combinations.

What is done in the table window (section 5) can be seen as the specification of an aggregate data set. This specification can be taken back to the meta-data window to constitute an (aggregate) object. This object can either correspond to a physically created summary file or only be a specification without extensions. The inheritance mechanisms in Smalltalk makes it possible to handle such entities as "abstract objects". We can use the specification to generate the program to create the summary file if needed. It is possible to have several such specifications for a single entity type, giving several ways to create a specific summary file.

The idea of "abstract objects" can be extended to cater for the handling of event-oriented databases. We can specifiy abstract objects representing the states of objects at any specific time point as a function of the events. Malmborg-82, -83, -84 present the modelling issues behind this potential extension of TBE-2. Such extensions would make the tool into a general purpose system development tool, rather than a table design tool. The ideas in this last paragraph will be elaborated in another paper.

REFERENCES

Afsarmanesh, H., McLeod, D., Knapp, D., Parker, A.
An Extensible Object-Oriented Approach To Databases For VLSI/CAD
Proc 11th Int Conf on Very Large Data Bases, Stockholm 1985

BYTE (-81)
Special Issue of the journal BYTE presenting Smalltalk.
Several articles by the creators of the language.
BYTE Aug-81

Chan, P., Shoshani, A. (-80)
SUBJECT: A directory driven system for organizing and accessing large statistical databases.
Proc 6th Int Conf on Very Large Data Bases, Montreal 1980

Chen, P.P-S: (-76)
The Entity-Relationship Model. Toward a Unified View of Data.
ACM Transactions on Database Systems, Vol. 1, No 1, March 1976

Codd, E.F. (-79)
Extending the Database Relational Model to Capture More Meaning.
ACM Transactions on Database Systems, Vol. 4, No 4, Dec 1979

Copeland, G., Maier D. (-84)
Making Smalltalk a Database System
ACM Sigmod Record, June 1984

Cox, B.
Object-oriented Programming: An Evolutionary Approach
Addison-Wesley 1986

Date, C. (-85)
An Introduction to Database Systems, 4th ed., Vol 1 and 2.
Addison-Wesley, 1985

Goldberg, A. (-84)
Smalltalk-80 - The Interactive Programming Environment.
Addison-Wesley, 1984

Goldberg, A.,Robson, D. (-83)
Smalltalk-80 - The Language and its Implementation.
Addison-Wesley, 1983

Jackson, M.A. (-83)
System development
Prentice Hall, 1983

Johnson, R.R. (-81)
Modelling Summary Data
Proc ACM SIGMOD Int Conf on Management of Data, 1981

Kay, A.C. (-77)
Microelectronics and the Personal Computer
Scientific American, September 1977

King R.
A Database Management System Based on an Object-Oriented Model.
Proc Conf on Expert Database Systems 1986. Benjamin/Cummins 1986.

Klug, A.
Abe, A Query Language for Constructing Aggregates-by-example.
Proceedings of the First LBL Workshop on Statistical Database Management. Menlo Park, California 1982.

Krasner, G. (-83)
Smalltalk-80 - Bits of History, Words of Advice.
Addison-Wesley, 1983

Malmborg, E. (-82)
The OPREM-Approach - An extension of an OPR-approach to include dynamics and classification.
Statistiska Centralbyr}n S/SYS-E12

Malmborg, E. (-83)
An analysis of system design methodologies using the ISO-framework.
Second Scandinavian research seminar on information modelling and data base management (Tampere 1983), also as SCB S/SYS-E14.

Malmborg E. (-84)
Stepwise formalization of information systems specifications by extending a simple object-oriented approach.
Seventh Scandinavian research Seminar on Systemeering (Helsinki-84), also as SCB P/ADB-E20

Malmborg, E., Chowdhury, S. (-84)
Program generation as a method for the production of portable software - an experiment and a proposal for a methodolgy. Paper in Swedish
SCB, ADB-METOD 1984:04

Malmborg, E. (-86)
On the Semantics of Aggregated Data.
SCB, U/ADB - E25, also in Proc 3rd Int Workshop on Stat Database Man, Luxembourg 1986

McCarthy, J.L. (-82)
Metadata Management for Large Statistical Databases
Proc 8th Int Conf on Very Large Data Bases, Mexico City 1982

Niestrasz, O., Tsichritzis, D.
An Object-Oriented Environment For OIS Applications
Proc 11th Int Conf on Very large Data Bases, Stockholm 1985

Nilsson, G. (-84)
Table by Example - TBE
SCB, Feb 1984

Nordbäck, L., Widlund, A. (-82)
AXIS - The Manager of Very Large Statistical Databases.
Proc COMPSTAT 1982 (5th Symposium at Toulouse)
Physica-Verlag, Wien 1982

Nordbäck, L. (-83)
Problems, Plans and Activities concerning the Economic Databases at Statistics Sweden.
Proc 2nd Int Workshop on Stat Database Man, 1983 (Los Altos)

Olsson, L. (-82)
The role of databases in the dissemination of statistics
SCB, S/SYS - E13, also in Proc ISIS-82 seminar

Sato, H., Nakano, T., Fukazawa, Y.,Hotaka, R. (-86)
Conceptual Schema for a Wide-Scope Statistical Database and Its Applications
Proc 3rd Int Workshop on Stat Database Man, Luxembourg 1986

Schmucker, K.J.
Object-Oriented Programming for the Macintosh.
Hayden Book Company, New Yersey 1986

Shoshani, A.
Statistical Databases: Characteristics, Problems, and some Solutions.
Proc 8th Int Conf on Very Large Data Bases, Mexico City 1982

Smith, D.C., Harslem, E.
Designing the Star User Interface
BYTE, April 1982

Sundgren, B. (-73)
An Infological Approach to Data Bases
SCB, URVAL nr 7

Sundgren, B. (-84a)
Conceptual design of data bases and information systems
SCB P/ADB-E19

Sundgren, B. (-81)
Statistical data processing systems - Architectures and Design Methodologies
SCB S/SYS - E11

Sundgren, B. (-85)
Outline of an Algebra of Base Operators for Production of Statistics.
SCB, P/ADB-E21

Teorey, J.T., Yang, D., Fry, J.P.
A Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model.
ACM Computing Surveys, Vol 18, No 2, June 1986

Teitel, R.
Statistical Databases and Database Systems.
Proc Bureau of the Census First Annual Research Conference
Reston, Virginia 1985

Wong, H.K.T., Kuo, I. (-82)
GUIDE: Graphical User Interface for Database Exploration.
Proc 8th Int Conf on Very Large Data Bases, Mexico City 1982