



**STATISTICS DIRECTORATE
STATISTICAL INFORMATION MANAGEMENT AND SUPPORT DIVISION**

STD/SIMS(2004)8
Unclassified

**Part 1: Use of Software Consortium Programs GESMES/TS Deliveries from Output Databases at
Statistics Sweden**

OECD EXPERT GROUP ON STATISTICAL DATA AND METADATA EXCHANGE

**1-2 April
Château de la Muette, Paris**

JT00160611

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

English - Or. English

TABLE OF CONTENTS

PART 1: USE OF SOFTWARE CONSORTIUM PROGRAMS GESMES/TS DELIVERIES FROM OUTPUT DATABASES AT STATISTICS SWEDEN 3

- Background 3
- Sweden’s Statistical Databases and the data model 3
- The production database and its use..... 4
- Routines for making Gesmes/TS files from the database environment 4
 - Preparation phase..... 4
 - Execution phase..... 8
- Comments to the present routines and future development 10
- The routines for making Gesmes/TS files where data is not in the database 10
- Alternative production of Gesmes/TS files not in the database 11
- Still things to improve for the data sending 12
- Final remarks..... 12
- Function of the PC-Axis consortium..... 12
 - Step-by-step development for the dissemination of statistics on the web 14
 - Possible transfer of the software to other organisations for statistics dissemination..... 14
 - Further information and questions..... 14

PART 1: USE OF SOFTWARE CONSORTIUM PROGRAMS GESMES/TS DELIVERIES FROM OUTPUT DATABASES AT STATISTICS SWEDEN

Background

Statistics Sweden has a long tradition of using standardised file formats to carry statistical data and metadata. This concept took the form of a data model, based on multi-dimensional matrices in the Statistics Sweden's central database, which was introduced in the mid-1970s. At the end of the 1980s, a file format was introduced for the transmission of statistics from the mainframe database to personal computers. This file format was then used as the input format for the PC-Axis software, developed for users of statistics in the municipalities in Sweden for the population census in 1990. When the development of an EDIFACT (Electronic Data Interchange For Administration Commerce and Trade) for electronic transmission of statistical data and metadata was initiated at the beginning of the 1990s, it was natural for Statistics Sweden to participate in the project. Statistics Sweden also linked the PC-Axis file format to the EDIFACT/Gesmes file format.

Over the years, several versions of such a converter were developed, related to the progress in the development of the Gesmes standard. EDIFACT/Gesmes is a framework within which two organisations can agree on how to transmit data between them. This implies not only converting the data into EDIFACT/Gesmes, but also adding certain agreed parameters to the conversion, according to users' needs. The present version of the Gesmes/TS conversion option in PC-Axis is adapted in accordance with the status of the Gesmes standard. This paper will present an overview of the current routines at Statistics Sweden for the production of Gesmes/TS files, using Sweden's Statistical Databases as the origin for delivering statistics to international organisations. However, it should be noted that certain statistics are not yet in the database and are therefore produced in a different way, which is also described in this paper.

Sweden's Statistical Databases and the data model

Statistics Sweden manages a public database service on the Internet. This database covers all statistical products for which Statistics Sweden is responsible. Some of the authorities, outside Statistics Sweden, that are responsible for official statistics, also have data in Sweden's Statistical Databases. Use of this database is free-of-charge. The database resides in a Sybase relational database, using a data model that covers users' needs concerning information on data at different levels. This generally relates to information from the table stub and headings, to the footnotes on all the different levels, to the table cells. Information is also available on contact persons with telephone and e-mail contacts, etc.

The same data model is used in other countries, although Denmark and Norway, together with Sweden, are the core partners concerning the data model development. The present version 2.0 is under implementation, not only in these three countries, but also at the UN/ECE Statistical Division and the statistical offices in Croatia, Ireland and Taiwan. Statistics Sweden and Ireland use Sybase, Denmark and Norway use Oracle and Croatia, the UN/ECE and Taiwan use MS-SQL Server.

The production database and its use

The external database takes data from an internal production database, using the same data model. The production database does not only consist of the data that is transferred to the public database on the release date, but also contains additional data. This may be data that is not official or data for commissioned services. The production database is used for different purposes in the production process, such as for control or for the production of publications. One of the tasks is to send data to international organisations. The production database can either be accessed using the same interface as that for external users on Internet, or using the PC-Axis SQL macro software. In PC-Axis SQL, it is possible to save queries to the relational database. Saved queries can be used to produce a table in PC-Axis file format or many other formats, such as Excel and dBase. One of these file formats is the Gesmes/TS. Saved queries can be executed in PC-Axis at a later stage or in a batch program named PX-Batch.

Routines for making Gesmes/TS files from the database environment

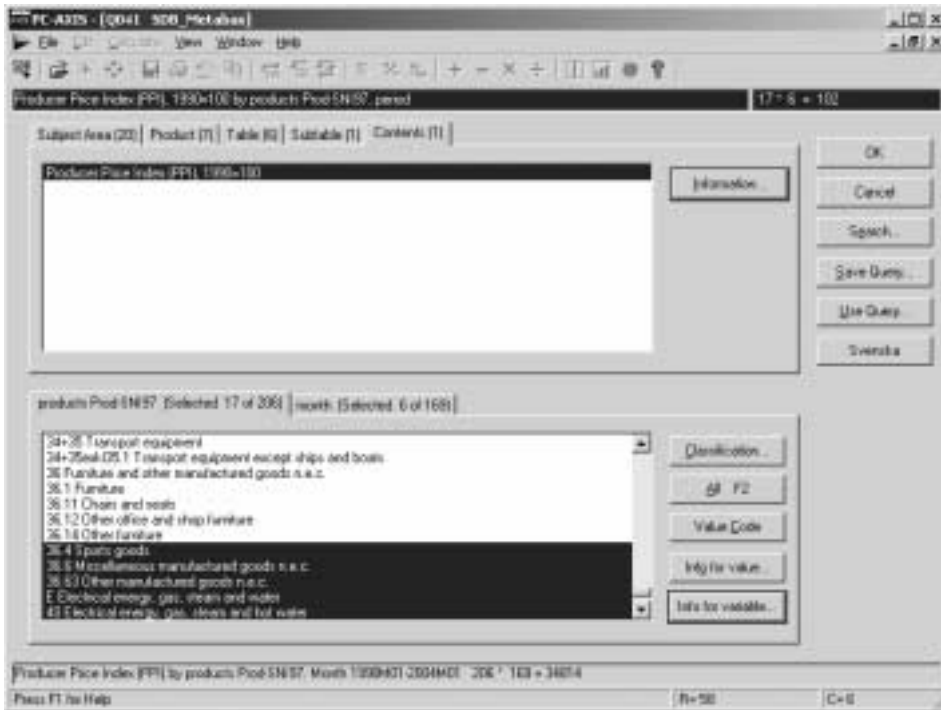
The majority of data that is converted to Gesmes/TS is retrieved from the public database, which is available on Internet, but some data is retrieved from the production database. However, the routines are the same regardless of where the data comes from. The routines are divided into a preparation phase, normally performed only once, and an execution phase, which is repeated each month, each quarter or another cycle, depending on the requirements of the receiving organisation.

Preparation phase

The preparation phase consists of the following steps:

1. Produce saved queries, using PC-Axis SQL;
2. If needed, parameters for a selection of arrays from a rectangular retrieval must be prepared;
3. If needed, concatenation of several Gesmes/TS files within one dataflow must be prepared;
4. If needed, a list of queries for PX-Batch should be developed.

Step 1: start with the selection of data from the database in PC-Axis SQL user interface



When the selection is made, the Save Query button is pressed and the following window will appear:



Select a time option. "Moving time interval" is the normal option for regular reporting. Decide variable order and where the saved query shall be placed. A Gesmes/TS form then follows, accessed by pressing the Gesmes/TS button.

The screenshot shows the 'Gesmes/TS' window with the following fields and values:

- Sender: SE1
- Contact name: Lars Nordback
- E-mail: lars.nordback@scb.se
- Receiver: 400
- Code list maintainer: EUROSTAT
- Interchange ref: IREF000001
- Message ref: MREF000001
- Dataset name: STS_IND_PROD_M
- Data structure id (key family): EUROSTAT_STS
- Observation status: A=Normal value
- Confidentiality: F=Free
- Filename (with path) for code mapping when needed: [L:\T\MIMA\GesmesCB\GesmesCB\ani2naceK.0RR1.bt]

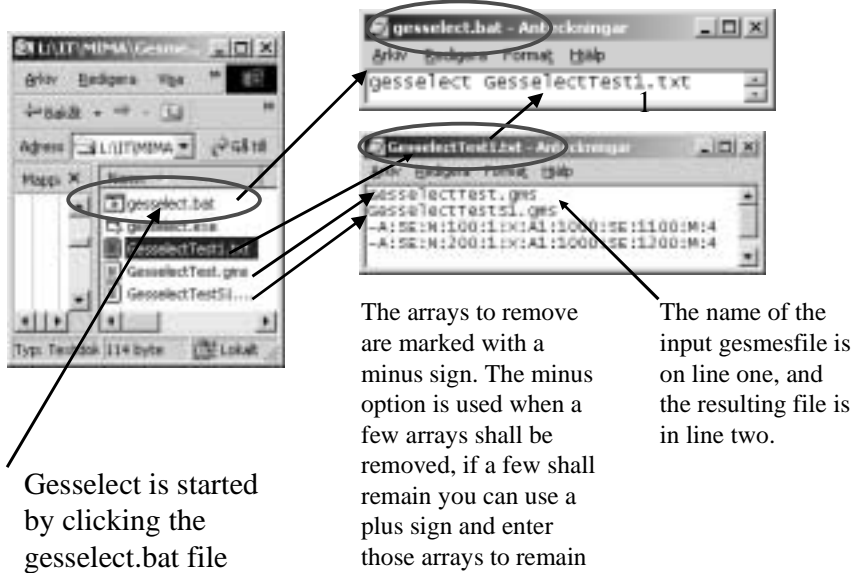
Key dimensions

Fixed in key			Varying in key	
Concept	Code value	Position in key	Concept	Position in key
			First	Last
Frequency	M	1	Table variables	5
RefCountry	SE	2	products Prod-SN197	5
Adjustment	w	3		
Indicator	PROD	4		
Institution	1	6		
Baseline	1995	7		

The different parameters should be filled in, according to the Gesmes/TS manual that describes the data flows requested by the receiving organisation. The names on the texts in the form are taken from the manuals received from Eurostat, concerning short-term statistics. As can be seen, much information to be entered here is not available in the local statistical database but, once this has been entered for a particular retrieval, it will be included in the saved query. One important parameter is the name of the code-mapping file. This is necessary when the codes used in the local database are not the same as those in the receiving organisation.

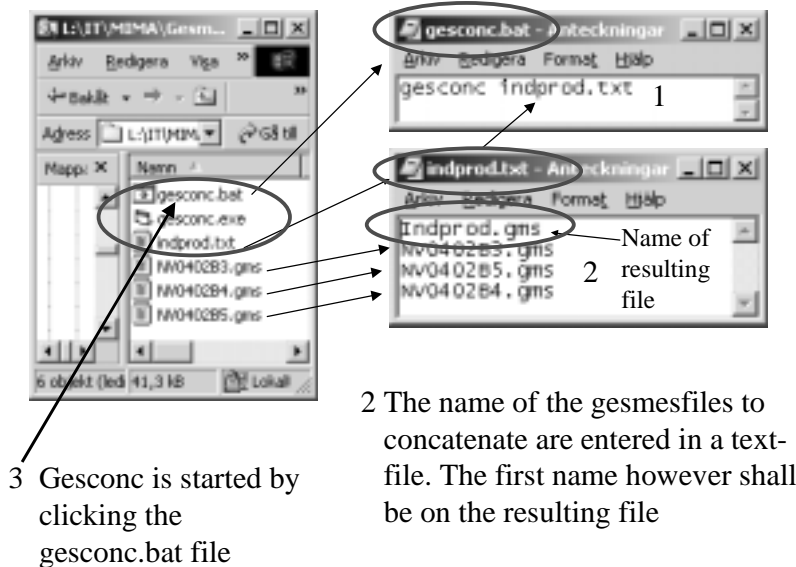
Step 2 in the preparation phase is needed if the multi-dimensional data structure in the database does not fit the receiving organisation's database. If time and two or more dimensions are used, the retrievals become rectangular and sometimes contain combinations that are not required by the receiving organisation. For this purpose, a piece of software named GesSelect has been developed by Statistics Denmark. How it is used can be seen in the picture below.

The Selection program is monitored using a bat- and a txt-file



Step 3 above is needed since the receiving organisation does not want to receive many small files in a data flow but to have one consolidated Gesmes/TS file for the data flow. If this were a true EDIFACT application, the Gesmes/TS messages would automatically enter in the right place in the receiving database. The GesConc program is developed in Statistics Denmark and is easy to use since it is similar to the GesSelect program showed above. It looks like this:

The Concatenation program is monitored using a bat- and a txt-file



Step 4: preparation of PXT-list for PX-Batch

If several queries are to be executed in one batch, it is possible to put the queries together in a list. This list must have the extension PXT. The queries have the extension PXS. The PC-Axis file has the extension PX, which explains the name of some of the programs.



Execution phase

The execution phase consists of the following steps:

1. PX-Batch – run queries that produce Gesmes files
2. GesSelect – selecting series for delivery
3. GesConc – concatenating files into one file
4. Stadium Web Client – sending the Gesmes/TS files and keeping track of deliveries

The first step can involve several retrievals, in which several queries are put together in a long list that can be run in one batch in the PX-Batch program. This was described in the preparation phase above.

The PX-Batch program needs information on which database to access, see 1 in the image below.

The list of queries is picked up in the PX-Batch program, using the function Select list, see 2 in the image below.

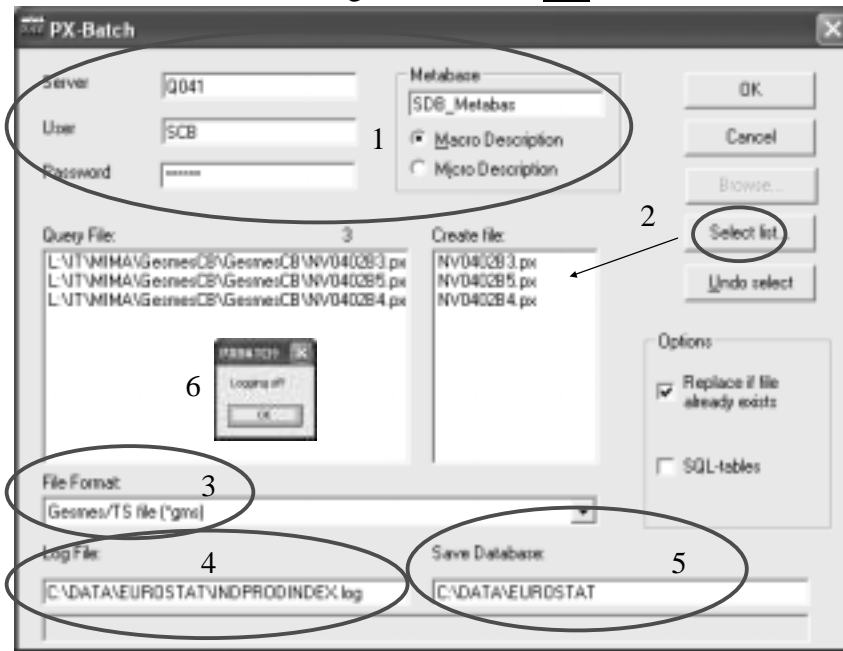
Select the Gesmes/TS file format, see 3 below.

Name a log file, see 4.

Indicate where the Gesmes/TS files should be written, see 5

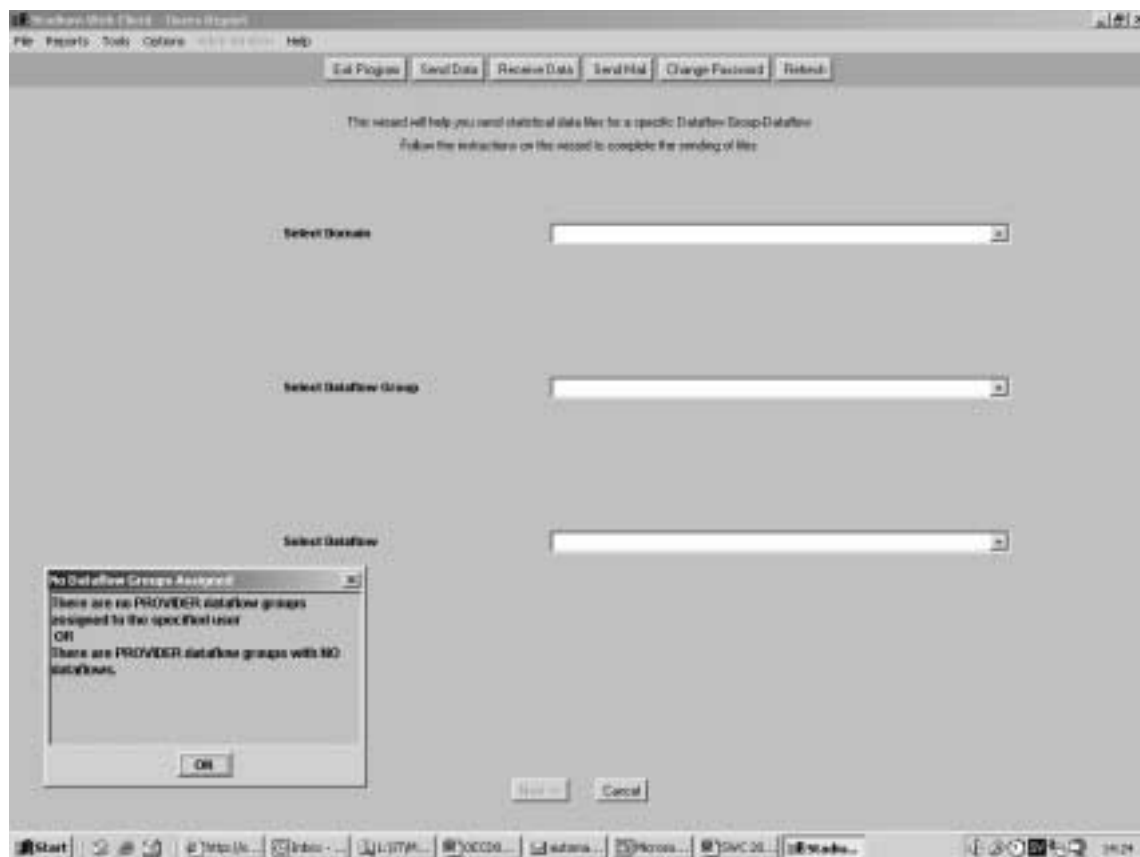
Log out once the batch is completed.

PX-Batch to make a set of gesmesfiles for one dataflow



If necessary, the programs GesSelect and GesconC can be run in one batch when the Gesmes/TS files are produced.

When the final Gesmes/TS file is prepared, it can be sent to Eurostat using the Stadium Web Client, see below.



Comments to the present routines and future development

As can be seen above, there are several steps in the routines, both for the preparation phase and for the execution phase. Statistics Sweden is still fine-tuning the most rational way of handling all the different parameters described in the manuals that have so far been created by Eurostat and the ECB. Progress in the field in the subject area units in Eurostat is very slow. We are still expecting more manuals to be produced, with the hope that nothing in these manuals will make it necessary to change our systems.

The next goal is to make the whole process more automatic. In neighbouring countries, Denmark and Norway, which are using the same data model as Statistics Sweden, projects are ongoing to develop automatic systems for sending Gesmes/TS files to Eurostat. Statistics Denmark is using the same tools as described above. To some extent, the preparation phase is similar but the execution will be automatically triggered when updates are released in the public database. Statistics Norway has another approach and will not use the same tools as we are using. Instead, they will produce the Gesmes/TS files in a specially developed program, directly from the database.

The routines for making Gesmes/TS files where data is not in the database

A consultancy company has developed an "Excel Add-in" to convert data that is not in the database into Gesmes/TS-files.

Microsoft Excel - test.xls

Arkiv Redigera Visa Infoga Format Verktyg Data Fönster Hjälp Årsigt

Formel 10

	A	B	C	D	E	F	G	H	I	J	K	L
1	Industriproduktionsindex (F&Pv), kalenderkorrigerat, 1995=100 efter											
2	tid och näringsgren SM82											
3												
4												
5		C-E	ERVE	C+D	INS&K&ER&V	INV	IKON	VKON	C	10-12	10	11
6	2001M10	124,8	108	126,4	118,2	109	137,6	122,1	136,7	85,9	156,2	156,2
7	2001M11	130,7	111,6	132,6	118,8	104	153,5	122	144	91,1	162,3	162,3
8	2001M12	136,4	114	140,9	121,1	106,8	170,3	127,4	161,7	91	153,3	153,3
9	2002M01	111,9	116,5	111,4	111,4	105,6	110,7	112,7	113,2	79,9	154,4	154,4
10	2002M02	124	112,6	125,1	117,2	106,8	137,9	119	122,4	85,8	161,4	161,4
11	2002M03	134,9	112,7	137,2	119,8	109,7	162,6	127,5	138,8	86,2	163,7	163,7
12												
13												
											Summa=42029,3	

Klar

The Excel Workbook contains three sheets. Sheet 1 contains the table with the data to convert. Sheet 2, named ini, contains the required Gesmes/TS parameters concerning sender, receiver etc.

Microsoft Excel - test.xls

Arkiv Redigera Visa Infoga Format Verktyg Data Fönster Hjälp Årsigt

Formel 10

	A	B	C	D	E	F	G	H	I	J
1	[Sender]	SCB								
2	[Receiver]	Eurostat								
3	[InterchangeReference]	IREF000001								
4	[MessageReference]	MREF000001								
5	[DataSetName]	STS_IND_PROD_M								
6	[DataSetStructureId]	EUROSTAT_STS								
7										
8	[ArrayData]	M	SE	W	PROD	[Activity]	1	1995	[Period]	[Observation] A
9	[CodeConversion]				ans2name					
10										

Klar

Sheet number three contains the code mapping for the actual material.

Microsoft Excel - test.xls

Arkiv Redigera Visa Infoga Format Verktyg Data Fönster Hjälp Årsigt

Formel 10

	A	B	C	D	E	F	G	H	I	J	K	L
1	10	NI1000										
2	10-12	NI00CA										
3	11	NI1100										
4	12	NI1200										
5	13	NI1300										
6	13.1	NI1310										
7	13.2	NI1320										

Klar

Alternative production of Gesmes/TS files not in the database

Having two different systems for the conversion to Gesmes/TS is not very satisfactory. PC-Axis is currently used when the data originates from the databases and the Excel Ad-in is used for data that is not

in the database. The plan is now to use a PC-Axis conversion also for data not available in the database. This will be possible with PC-Axis 2004, where it is possible to save queries out of PC-Axis files. With this technique, the Gesmes/TS parameters will be saved in such a query. The name of such a query file is PXQ. The 2004 version of PC-Axis will be released in June 2004.

To be able to use this routine, the data must be transferred into PC-Axis files. This is simply done, using PX-Make (developed at Statistics Denmark) or even simpler with PX-Edit (developed at Statistics Finland).

Still things to improve for the data sending

In the present database, there is no comprehensive method for handling different type of flags and special values. An extensive project is ongoing, with the cooperation of Denmark, Norway and Sweden, to enhance the capability of the databases in these three countries in the field of special signs in the table cells. This will probably be introduced at the end of 2004 in the Norwegian database and later in Denmark and Sweden.

Final remarks

An ideal situation would be an entirely automatic process for sending statistics to international organisations, as originally proposed at the start of the EDIFACT project. It has so far been a question of trial and error, according to the different needs in the different Eurostat data flows. When using the same tools for sending statistics to other organisations, it has been necessary to develop processes even further. Adjustments have been introduced in the Gesmes/TS conversion option in PC-Axis over the years. We hope to develop a solution similar to the automotive industry, where the monitoring the process simply involves checking the logs.

For an automatic standardised transfer of data to be achieved, it is essential that the organisations involved adopt the standards in practice. Among other things, this involves producing the necessary manuals and initiating the processes, both in international organisations and in the NSOs. In the SPC meeting in May 2003, Statistics Sweden's Director General proposed that all international statistical transfers from January 2005 should be in Gesmes/TS format. This should be implemented in the Eurostat plan of action for the introduction of Gesmes/TS in the organisation.

For those interested in joining the PC-Axis family and benefiting from the development in the field of electronic dissemination of statistics, this document ends with a description of the PC-Axis consortium.

Function of the PC-Axis consortium

The structure of the "consortium" can be described as follows:

- Statistics Sweden is the leading country, managing the cooperation;
- The PC-Axis International Reference Group exists for all parties outside Sweden, that are using PC-Axis as a dissemination tool;
- There is a Nordic core for cooperation in the area of output databases, where Denmark, Norway and Sweden are using the same data model;
- A Swedish user group exists for the use of PX-Web applications in other organisations in Sweden;

- An internal user group is also active at Statistics Sweden;
- There is a board, in which some of the main end users are consulted, concerning the development of Sweden's Statistical Databases;
- A steering committee for the Database Project at Statistics Sweden is involved, where PC-Axis constitutes one of the important ingredients.

The first three instances above are described more in detail below:

1. Statistics Sweden is the architect of the PC-Axis file format and the PC-Axis software. After presentations at different international conferences, other NSIs became interested in the product. Since Statistics Sweden had no funds for supporting other NSIs that are interested in or are using PC-Axis, it was decided to charge other organisations for using PC-Axis for the dissemination of statistics.

After negotiations with the first interested countries, a price list for an annual license fee took form. Since Statistics Sweden is a non-profit organisation and is also using the software, the price is minimal.

The contract for a licence for another NSI also covers the right to disseminate the software on the Internet or on a CD-ROM, together with statistics. It also gives the right for the NSI to sub-license other organisations, in the geographical area covered, to use the product for dissemination of statistics. The NSI may also charge for its redistribution of the software or not. This is the background for the existence of a differentiated price list.

The licence fee covers the cost for the administration of the customer contacts, partly for the development of PC-Axis at Statistics Sweden, arranging the annual PC-Axis Reference Group Meetings and the other meetings and contacts with the different groups mentioned above.

2. The *PC-Axis International Reference Group* was established in 1992. It usually meets once a year.

- The morning session on the first day is focused on the exchange of experiences in the field of dissemination statistics in electronic form and the use of the PC-Axis family software. The hosting organisation gives a deeper presentation of its work and plans in the field.
- The afternoon of the first day is devoted to demonstrations and presentations of new developments in the field. There are developments at all the different organisations, while the main development takes place at Statistics Sweden.
- The second day is completely devoted to discussions on developments that will take place in the coming year in the PC-Axis file dependent software.
- The third day is focused on the SQL extensions related to the PC-Axis family software. Experiences are exchanged and desired developments discussed.

The PC-Axis international reference group meeting took place at the following sites

Meeting in	Year	Vitoria-Gasteiz	1996	Kinsale	2001
Stockholm	1992	Stockholm	1997	Tallinn	2002
Stockholm	1993	Copenhagen	1998	Ljubljana	2003
Copenhagen	1994	Oslo	1999	Not yet decided	2004
Helsinki	1995	Madrid	2000		

3. The Nordic core of Cooperation in the field of Output databases and the PC-Axis family software

As mentioned above, development work is currently distributed among some of the users of the PC-Axis family software. These are Statistics Denmark, Statistics Norway and Statistics Sweden, which usually meet once a year for planning purposes and to discuss the developments. The concept is divided into the common **Metadata models**, the common **PX-Main programs**, the common **PX-components, groups of cooperation** and other **groups for cooperation linked to the database activities**. For the common Metadata models and the PC-Axis file format, certain common rules exist for how these must be used, as well as for the creation of new versions. This is to ensure that users of the data model should not come to a standstill. Similarly, there are rules for the common PX-Main programs and PX-components.

The use of PC-Axis in the different NSIs is based on the common macro Metadata model that is reflected in the PC-Axis file format. Some of the NSIs are only using the PC-Axis file format, while others are also using the macro metadata model in a commercial Database Management System, DBMS, such as Sybase, Oracle or MS-SQL server.

Step-by-step development for the dissemination of statistics on the web

- A simple way of establishing an Internet-based statistical database is to create PC-Axis files using PX-Make and put them together with PX-Web on the website;
- A further step then can be taken establishing a SQL-based output database using the macro metadata model;
- PC-Axis SQL additions can be used internally to make PC-Axis files for the PX-Web;
- Publications can be produced in paper copy or on the web, using PX-Publ;
- An SQL user interface can be established, using components from PC-Axis SQL saved queries, PX-Batch and PX-Web.

Possible transfer of the software to other organisations for statistics dissemination

It is possible to transfer the following programs if a licence is agreed with Statistics Sweden: PC-Axis Main module, PX-Web, PX-Make, PX-Edit, PX-Map, PX-iMap. For organisations using SQL, the SQL-related programs PC-Axis SQL, PX-Batch and PX-Publ also can be transferred.

Further information and questions

For more information and references visit the website <http://www.pc-axis.scb.se/>

Questions on the PC-Axis software family and the license conditions for the use of the software for dissemination of statistics can be put to lars.nordback@scb.se