

Palvelutuotteet/Tietokannat/Jantunen  
veli-matti.jantunen@stat.fi  
09-1734 2326

8.3.2006

<b>PX-EDIT: STRUCTURED FILES.....</b>	<b>2</b>
General .....	2
Basic format .....	3
Basic format with codes .....	6
Table format .....	7
Table format with codes.....	10
Processing of sequential files .....	11
Further information .....	12

## PX-Edit: structured files

### General

In addition to PC-Axis files, tables can be transferred to PX-Edit from other systems: either from text files, direct from Excel or from Excel files or by means of clipboard (e.g. from OpenOffice tables).

All tables to be transferred should be **structured** so that PX-Edit is able to convert the table into PC-Axis format. A structured table contains only the minimum necessary information, and only these:

- table title
- variable names,
- variable values (headings), and
- numerical data.

PX-Edit does not require that all the data values of the table should be in the input data. The missing data values are replaced by a separately specified fill item. The table variables or variable values need not be in any specific order and even duplicated data cell references are allowed (the last found reference is taken as the value of the data cell in question). In addition, the classification codes related to the variable values can be included, as shown later in these instructions.

*NB PX-Edit warns users of duplicated data cell references.*

In text files cells are separated either by a tabulator or semi-colon. A comma and space are also possible, but to avoid confusions, their use is **not** recommended. In data input either a comma or full stop is accepted as a decimal separator, but thousand separators should not be used. Data cells may also have **dot codes** (or replacing representations possibly defined for them), but if the value of an individual data cell cannot be converted into a figure or a dot code, it is replaced by a fill item.

*NB PX-Edit will not warn if it cannot convert a figure.*

Excel tables are mainly used in the following examples, from which the table structure becomes apparent. The same principles also apply to the processing of text files.

Palvelutuotteet/Tietokannat/Jantunen  
veli-matti.jantunen@stat.fi  
09-1734 2326

8.3.2006

## Basic format

At its simplest a transfer file (or table) has in the **first row** (anchor cell) the title of the table, in the next row the names of the variables and below them by row the values of each variable and in the end the numerical data value corresponding to them. The part below the variable names can also be perceived as being consisting of columns, which first have different variable values and in the last column the data value corresponding to the variable value combination of each row.

A file of this type may be the easiest one for transferring tables to be produced from relation databases.

	A	B	C	D	E	F	G	H	I
1	Population 1 January after time, citizenship, age, reporting country and sex.								
2	<b>time</b>	<b>citizenshi</b>	<b>age</b>	<b>reporting</b>	<b>sex</b>				
3	1990	Total	Total	Finland	Total	4974383			
4	1990	Total	Total	Finland	Men	2412760			
5	1990	Total	Total	Finland	Women	2561623			
6	1990	Total	0-4 years	Finland	Total	309419			
7	1990	Total	0-4 years	Finland	Men	158072			
8	1990	Total	0-4 years	Finland	Women	151347			
9	1990	Total	5-9 years	Finland	Total	327075			
10	1990	Total	5-9 years	Finland	Men	167346			
11	1990	Total	5-9 years	Finland	Women	159729			
12	1990	Total	10-14 year	Finland	Total	325775			
13	1990	Total	10-14 year	Finland	Men	166516			
14	1990	Total	10-14 year	Finland	Women	159259			
15	1990	Total	15-19 year	Finland	Total	301284			
16	1990	Total	15-19 year	Finland	Men	153928			
17	1990	Total	15-19 year	Finland	Women	147356			
18	1990	Total	20-24 year	Finland	Total	353391			
19	1990	Total	20-24 year	Finland	Men	180017			
20	1990	Total	20-24 year	Finland	Women	173374			
21	1990	Total	25-29 year	Finland	Total	377940			
22	1990	Total	25-29 year	Finland	Men	193158			
23	1990	Total	25-29 year	Finland	Women	184782			
24	1990	Total	30-34 year	Finland	Total	386877			
25	1990	Total	30-34 year	Finland	Men	197918			
26	1990	Total	30-34 year	Finland	Women	188959			
27	1990	Total	35-39 year	Finland	Total	410165			
28	1990	Total	35-39 year	Finland	Men	209916			
29	1990	Total	35-39 year	Finland	Women	200249			

This table has five variables and each variable value is given in every row.

Palvelutuotteet/Tietokannat/Jantunen  
veli-matti.jantunen@stat.fi  
09-1734 2326

8.3.2006

The variable values can also be given hierarchically. Then the content used for an empty field is the content of the previous non-empty variable value field.

	A	B	C	D	E	F	G	H	I
1	Population 1 January after time, citizenship, age, reporting country and sex.								
2	<b>time</b>	<b>citizenshi</b>	<b>age</b>	<b>reporting</b>	<b>sex</b>				
3	1990	Total	Total	Finland	Total	4974383			
4					Men	2412760			
5					Women	2561623			
6			0-4 years	Finland	Total	309419			
7					Men	158072			
8					Women	151347			
9			5-9 years	Finland	Total	327075			
10					Men	167346			
11					Women	159729			
12			10-14 year	Finland	Total	325775			
13					Men	166516			
14					Women	159259			
15			15-19 year	Finland	Total	301284			
16					Men	153928			
17					Women	147356			
18			20-24 year	Finland	Total	353391			
19					Men	180017			
20					Women	173374			
21			25-29 year	Finland	Total	377940			
22					Men	193158			
23					Women	184782			

Please note that the **first** value has to be given to every variable.

The next example also has a basic format table but now not all the data values are included:

	A	B	C	D	E	F	G	H	I
1	Population 1 January after time, citizenship, age, reporting country and sex.								
2	<b>time</b>	<b>citizenshi</b>	<b>age</b>	<b>reporting</b>	<b>sex</b>				
3	1990	Total	Total	Finland	Total	4974383			
4					Men	2412760			
5					Women	2561623			
6			0-4 years	Finland	Total	309419			
7			5-9 years	Finland	Total	327075			
8			70+ years	Finland	Total	440894			
9		Denmark	Total	Finland	Total	463			
10					Men	280			
11					Women	183			
12			0-4 years	Finland	Total	17			
13			5-9 years	Finland	Total	37			
14			70+ years	Finland	Total	32			
15		Finland	Total	Finland	Total	4953209			
16					Men	2400848			
17					Women	2552361			
18			0-4 years	Finland	Total	308675			
19			5-9 years	Finland	Total	325600			
20			70+ years	Finland	Total	439596			
21		Iceland	Total	Finland	Total	69			

Palvelutuotteet/Tietokannat/Jantunen  
veli-matti.jantunen@stat.fi  
09-1734 2326

8.3.2006

Transferred to PX-Edit, this table looks like this (now two dots are used as a fill item):

[2] Population 1 January after time, citizenship, age, reporting country and sex.					
Tiedosto Muokkaa Ikkuna					
1990	Total	Total	Finland	Total	4974383
				Men	2412760
				Women	2561623
		0-4 years	Finland	Total	309419
				Men	..
				Women	..
		5-9 years	Finland	Total	327075
				Men	..
				Women	..
		70+ years	Finland	Total	440894
				Men	..
				Women	..
	Denmark	Total	Finland	Total	463
				Men	280
				Women	183
		0-4 years	Finland	Total	17
				Men	..
				Women	..

60 riviä x 1 saraketta = 60 lukua

The table is shown in PX-Edit by default with hierarchical titles.

Palvelutuotteet/Tietokannat/Jantunen  
veli-matti.jantunen@stat.fi  
09-1734 2326

8.3.2006

## Basic format with codes

In addition to variable values, codes corresponding to them can be entered in connection with the table to be transferred. This is made by giving the **same** variable name twice, whereby the first column has codes and the next texts.

	A	B	C	D	E	F	G
1	Taulukko 01: Käyttötalous: kunnat, tehtävä ja meno/tulolaji						
2	<b>Kunnat</b>	<b>Kunnat</b>	<b>Tehtävä</b>	<b>Tehtävä</b>	<b>Meno/tulolaji</b>	<b>Meno/tulolaji</b>	
3	004	Alahärmä	110	Yleishallinto	100	Palkat ja palkkiot	2025
4	004	Alahärmä	110	Yleishallinto	370	Eläkevakuutusmaksut	343
5	004	Alahärmä	110	Yleishallinto	380	Muut sosiaalivakuutusmaksut	152
6	004	Alahärmä	110	Yleishallinto	510	Asiakaspalvelujen ostot valtiolta	0
7	004	Alahärmä	110	Yleishallinto	520	Asiakaspalvelujen ostot kunnilta	0
8	004	Alahärmä	110	Yleishallinto	530	Asiakaspalvelujen ostot kuntayhtymiltä	0
9	004	Alahärmä	110	Yleishallinto	540	Asiakaspalvelujen ostot muilta	0
10	004	Alahärmä	110	Yleishallinto	600	Muiden palvelujen ostot	2618
11	004	Alahärmä	110	Yleishallinto	1400	Aineet, tarvikkeet ja tavarat (ml. varastoje	167
12	004	Alahärmä	110	Yleishallinto	1800	Avustukset	999
13	004	Alahärmä	110	Yleishallinto	1920	Vuokramenot, ulkoiset	21
14	004	Alahärmä	110	Yleishallinto	1970	Vuokramenot, sisäiset	189
15	004	Alahärmä	110	Yleishallinto	2700	Muut menot	401
16	004	Alahärmä	110	Yleishallinto	2900	Toimintamenot yhteensä	6915
17	004	Alahärmä	110	Yleishallinto	2910	Siitä: Sisäiset menot yhteensä (ml. men	409
18	004	Alahärmä	110	Yleishallinto	2940	Poistot ja arvonalentumiset	293
19	004	Alahärmä	110	Yleishallinto	2960	Vyörytyserät	14
20	004	Alahärmä	110	Yleishallinto	2965	Palautusjärjestelmän arvonalisävero	247
21	004	Alahärmä	110	Yleishallinto	5210	Myyntitulot valtiolta	1
22	004	Alahärmä	110	Yleishallinto	5220	Myyntitulot kunnilta	38
23	004	Alahärmä	110	Yleishallinto	5500	Maksut	25
24	004	Alahärmä	110	Yleishallinto	6700	Muut tulot	143
25	004	Alahärmä	110	Yleishallinto	6900	Toimintatulot yhteensä	379
26	004	Alahärmä	110	Yleishallinto	6910	Siitä: Sisäiset tulot yhteensä (ml. tulot o	15
27	004	Alahärmä	110	Yleishallinto	6960	Vyörytyserät	418
28	004	Alahärmä	201	Sosiaali- ja terv	100	Palkat ja palkkiot	596

Codes need not be given to all variables.

PX-Edit's import routine treats every possible value-code pair separately, in which case there may be several *Vyörytyserät* or *Yhteensä* value texts in the same variable if they have **different** codes. According to the PC-Axis standard, the variable values should be unique, but because in practice tables contrary to the px standard are used fairly generally, PX-Edit allows importing of such tables.

*Users are always warned against duplicated codes or values.*

Palvelutuotteet/Tietokannat/Jantunen  
veli-matti.jantunen@stat.fi  
09-1734 2326

8.3.2006

## Table format

In the table format presentation both row and column variables are used. The data part is then as a multi-column table. Compared with the basic format, the table will also contain a description of column variables, where the names of column variables are given after the title row by row in the first column and the variable values in their own data columns.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Population 1 January after time, citizenship, age, reporting country and sex.													
2	reporting country	Finland												
3	sex	Total												
4	time	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	
5	age	citizenship												
6	Total	Total	4974383	4998478	5029002	5054982	5077912	5098754	5116826	5132320	5147349	5159646	5171302	5181115
7		Denmark	463	484	514	474	473	476	482	482	517	528	574	580
8		Finland	4953209	4972223	4991423	5008732	5022325	5036742	5048260	5058566	5066749	5074586	5083622	5090041
9		Iceland	69	78	99	100	108	123	118	107	102	107	111	126
10		Norway	465	530	511	527	516	520	539	513	504	507	536	585
11		Sweden	5660	6051	6277	6456	6528	6685	7014	7291	7507	7756	7809	7887
12		other EEA	4495	4837	5151	5309	5500	5740	6229	6356	6903	7447	7946	8190
13		other Euro	3584	5800	12568	18043	24163	27584	30662	33230	36441	39210	41430	45643
14		Africa	942	1247	3230	3947	5244	6215	6970	7608	8547	8740	7791	7600
15		Canada	338	365	395	433	452	447	477	498	549	579	593	600
16		USA	1354	1475	1621	1740	1754	1775	1844	1833	1905	2001	2063	2010
17		other Ame	339	406	481	538	607	678	751	806	881	944	993	1028
18		Asia	1882	3002	4566	6720	8331	9884	11302	12512	14154	14782	15578	14597
19		Oceania	253	285	340	359	362	387	397	402	414	449	495	551
20		Stateless	371	369	388	411	469	613	1064	1405	1518	1368	1256	1172
21		Unknown	959	1326	1438	1193	1080	885	717	711	658	642	505	505
22	0-4 years	Total	309419	312485	317756	324759	326440	327702	324870	320172	312733	304936	297522	291275
23		Denmark	17	18	25	18	13	12	16	14	27	24	29	21
24		Finland	308675	311461	315910	322184	322949	323659	320284	314935	306890	298933	291812	285495
25		Iceland	4	3	10	10	13	11	10	6	6	7	8	6
26		Norway	20	27	25	35	36	33	38	36	38	33	32	33
27		Sweden	201	215	244	237	225	241	261	274	289	274	232	241
28		other EEA	89	100	108	111	114	136	161	157	177	182	227	211
29		other Euro	112	267	766	1159	1735	1870	1961	2148	2271	2367	2486	2711
30		Africa	32	31	201	316	504	700	938	1205	1478	1514	1076	985
31		Canada	9	6	11	11	8	7	7	12	17	20	20	19
32		USA	17	22	28	30	32	26	23	27	32	25	29	26
33		other Ame	20	21	23	19	24	21	23	19	30	33	32	42
34		Asia	76	143	240	498	698	915	1075	1272	1420	1390	1411	1339
35		Oceania	7	7	7	6	4	6	9	9	8	10	12	21
36		Stateless	21	12	13	11	12	14	17	17	9	6	8	11
37		Unknown	119	152	145	114	73	51	47	41	41	118	108	114
38	5-9 years	Total	327075	326827	324511	318487	315622	313943	316643	320975	327353	328311	329375	326424
39		Denmark	37	34	36	26	20	14	11	11	10	15	16	20
40		Finland	325600	325187	322135	315453	311538	309366	311718	315905	322079	323066	324142	321101
41		Iceland	5	7	9	8	9	6	8	8	6	5	5	5
42		Norway	40	37	25	28	23	21	27	22	22	25	32	33
43		Sweden	499	459	404	359	300	266	280	289	273	290	281	261
44		other EEA	302	260	209	165	136	106	128	132	155	163	166	176
45		other Euro	227	363	984	1523	2266	2504	2625	2649	2708	2657	2721	2852
46		Africa	35	45	146	211	471	636	694	702	732	706	589	603

Any logical table presentation can be quite easily converted into structured table format accepted by PX-Edit. Conversion of the presentation into a multi-dimensional PC-Axis table is based on the existence of **empty** cells, so care should be taken in forming the title part of the table.

Palvelutuotteet/Tietokannat/Jantunen  
veli-matti.jantunen@stat.fi  
09-1734 2326

8.3.2006

Here is a practical example of a table to be delivered to the *Toimiala OnLine* service:

	A	B	C	D	E	F	G	H	I	J	K	
1	Henkilöstömäärän kehitys Koillis-Suomessa											
2			ICT-klusteri									Puu-
3	Vuosi	Kuu	Muutos-%	Alkuperäin	Kausitasoi	Trendi	Yritykset	Toimipaika	Vuosi	Kuu	Muuto	
4	1997	01		60,98	59,43	58,03	33	36	1997	01		
5	1997	02		59,91	59,10	57,71	32	35	1997	02		
6	1997	03		56,43	53,44	57,34	31	34	1997	03		
7	1997	04		58,44	57,84	58,05	31	34	1997	04		
8	1997	05		62,89	60,53	59,25	33	36	1997	05		
9	1997	06		60,41	59,01	60,27	30	33	1997	06		
10	1997	07		62,92	64,04	60,98	29	32	1997	07		
11	1997	08		58,15	58,45	61,40	29	32	1997	08		
12	1997	09		57,92	61,62	62,38	29	32	1997	09		
13	1997	10		60,27	65,30	63,86	31	34	1997	10		
14	1997	11		61,45	65,39	64,93	31	34	1997	11		
15	1997	12		67,91	64,90	65,72	33	36	1997	12		
16	1998	01	14,9 %	70,08	68,45	66,35	34	37	1998	01	5,	
17	1998	02	12,8 %	67,56	66,71	66,34	33	36	1998	02	4,	
18	1998	03	19,5 %	67,46	63,55	66,30	35	38	1998	03	19,	

The title row is already in place. The names of the variables need to be added and moved to their right places.

First an empty row is added for row variables:

1	Henkilöstömäärän kehitys Koillis-Suomessa										
2	ICT-klusteri										Puu
3	Vuosi	Kuu	Muutos-%	Alkuperäin	Kausitasoi	Trendi	Yritykset	Toimipaika	Vuosi	Kuu	Muutos-%
4											
5	1997	01		60,98	59,43	58,03	33	36	1997	01	
6	1997	02		59,91	59,10	57,71	32	35	1997	02	
7	1997	03		56,43	53,44	57,34	31	34	1997	03	
8	1997	04		58,44	57,84	58,05	31	34	1997	04	
9	1997	05		62,89	60,53	59,25	33	36	1997	05	

Palvelutuotteet/Tietokannat/Jantunen  
veli-matti.jantunen@stat.fi  
09-1734 2326

8.3.2006

Next the names of column variables are entered in the first column, the names of the row variables in their own rows next to each other and the title area is then tidied up.

	A	B	C	D	E	F	G	H	I	J	K
1	<b>Henkilöstömäärän kehitys Koillis-Suomessa</b>										
2	Toimialat	<b>ICT-klusteri</b>									<b>Puu-</b>
3	Indeksit		Muutos-%	Alkuperäin	Kausitasoi	Trendi	Yritykset	Toimipaika	Vuosi	Kuu	Muuto
4	Vuosi	Kuukausi									
5	1997	01		60,98	59,43	58,03	33	36	1997	01	
6	1997	02		59,91	59,10	57,71	32	35	1997	02	

In the end unnecessarily duplicated year and month columns are deleted from the table.

	A	B	C	D	E	F	G	H	I	J
1	<b>Henkilöstömäärän kehitys Koillis-Suomessa</b>									
2	Toimialat	<b>ICT-klusteri</b>							<b>Puu-klusteri</b>	
3	Indeksit		Muutos-%	Alkuperäinen	Kausitasoitettu	Trendi	Yritykset	Toimipaikat	Muutos-%	Alkuperäinen k
4	Vuosi	Kuukausi								
5	1997	01		60,98	59,43	58,03	33	36		67,25
6	1997	02		59,91	59,10	57,71	32	35		71,42
7	1997	03		56,43	53,44	57,34	31	34		71,18
8	1997	04		58,44	57,84	58,05	31	34		70,42
9	1997	05		62,89	60,53	59,25	33	36		87,58
10	1997	06		60,41	59,01	60,27	30	33		95,74
11	1997	07		62,92	64,04	60,98	29	32		95,06
12	1997	08		58,15	58,45	61,40	29	32		70,72
13	1997	09		57,92	61,62	62,38	29	32		67,28
14	1997	10		60,27	65,30	63,86	31	34		87,96
15	1997	11		61,45	65,39	64,93	31	34		74,41
16	1997	12		67,91	64,90	65,72	33	36		96,19
17	1998	01	14.9 %	70.08	68.45	66.35	34	37	5.2 %	70.75

In Excel tables the right margin of the table area should always be checked for possible legends, etc. (they should be deleted or the table will have unnecessary phantom values of variables.)

Excel tables may in some cases be otherwise unpredictable, for example, the internal *month/year* presentation does not get converted properly, columns or rows cannot be deleted because of pivoting or other table editing, different versions installed onto the same computer may prevent transfer of the table, and so on.

Please also note that part of Excel editing has no effect on the transfer (e.g. colours, fonts and sizes and percentage and currency fields).

## Table format with codes

In the table format it is also possible to give variable codes in the same way as in the basic format. Codes are now given to column variables in rows of their own above the variable values.

Let this example shed light on the matter:

	A	B	C	D	E	F	G
1	Establishment and personnel by Region, Year, Industry and Number and personnel.						
2	Industry			C		D	
3	Industry			C Mining and quarrying		D Manufacturing	
4	Number and personnel			Establishment	Personnel	Establishment	Personnel
5	Year	Region	Region				
6	1995	0	Whole country	1271	3528	26009	380791
7		4	Alahärmä	2	0	42	458
8		5	Alajärvi	6	6	84	548
9	1996	0	Whole country	1409	3961	27442	386001
10		4	Alahärmä	2	0	41	406
11		5	Alajärvi	6	11	88	655

The general structured table format can be described as follows:

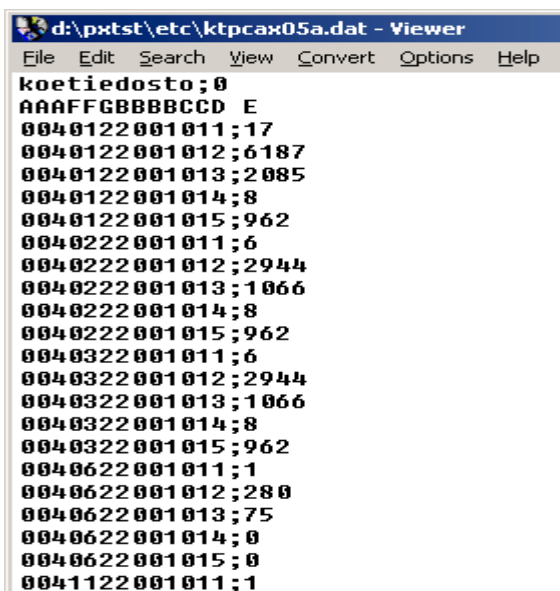
- the table title is in the first cell of the first row
- the names of the column variables are in the first cell of the following rows
- the names of row variables are in the following row next to one another
- the values of column variables are in their own rows
- the values of row variables are in their own columns
- the data part is presented as a table, each data value is at the intersection of its variable value titles
- possible variable codes are given with the same variable name before the value column or row
- the table may be comprised only of row variables (a table consisting of only column variables may in practice be valueless).

## Processing of sequential files

The data mass derived from some systems, especially from old mainframe applications, consists of sequential file records. Breaking down the record description of such a sequential file could be made with the help of **record mask**. The title of the table is entered in the first row of the file and after that a tabulator (or a semi-colon) and **zero** in order that PX-Edit could identify the file correctly.

The column mask is written in the next row, where each record field is masked with a sign that is repeated the required number of times. If the record has fields, punctuation marks, etc. which are not to be included, they are masked by space characters. The data field is always the last one, and pointing to the location of the first sign of the data part is sufficient for marking it (this is because the length of variable-length data parts cannot always be determined on the basis of the initial part of the file). When reading sequential files the variable values found from records are transferred into both values and codes of variables, because codes are usually processed in sequential files.

In the next example the file has six variables, and a semi-colon is not included:



```
d:\pxtst\etc\ktpcax05a.dat - Viewer
File Edit Search View Convert Options Help
koetiedosto;0
AAAFFGBBBBCCD E
0040122001011;17
0040122001012;6187
0040122001013;2085
0040122001014;8
0040122001015;962
0040222001011;6
0040222001012;2944
0040222001013;1066
0040222001014;8
0040222001015;962
0040322001011;6
0040322001012;2944
0040322001013;1066
0040322001014;8
0040322001015;962
0040622001011;1
0040622001012;280
0040622001013;75
0040622001014;0
0040622001015;0
0041122001011;1
```

PX-Edit will now have as “variable names” *AAA*, *FF*, *G*, *BBBB*, *CC* and *D*. These should naturally be replaced by the right variable names (e.g. variable *AAA* in the example would appear to be municipality code and *BBBB* year).

## Further information

Where necessary, spaces are deleted from the ends and starts of variable names, from variable values (such as classification headings) and only from the ends of possible codes; this is because in some systems, leading blanks are used to show the internal hierarchy of variables.

In a structured table the numbers of table row and column variables can be given direct in the cell after the title row (in the control cell), whereby PX-Edit bypasses table edit checking – then the table must not have variable codes.

The documentation of older PX-Edit versions also describes another control mode, where a special format table could be processed with the negative value of the control cell (a basic table where each variable has both codes and values). This format should not be used anymore, although it should still be possible to import it.

The size of the table to be read has no practical limits for PX-Edit. Excel has its own limitations, reading of especially large tables is always slow, and the size of text files can nowadays be over 4GB, which should be enough even for demanding use.

In editing of the table format, especially the basic format, the following additional information may help:

- empty fields at the *ends* of rows need not be expressed with the right number of successive field separators
- the value order of variables is determined only according to the match order of value texts; the order can always be corrected with the PX-Edit sort function, but where necessary, the variable values can be given in the right order right at the *start* of the result table without data cells.