

Data-rectie

THE CONTINUOUS SURVEY PROJECT +

Henk Schrik
Swidoc/Steinmetzarchief
Kleine-Cartmanplantsoen 10
1017 RR Amsterdam
tel: 020-225061

Abstract.

This paper presents the Continuous Survey Project of the Social Science Faculty of the University of Amsterdam. This project, originally set up as a quarterly survey project, started in January 1972 and is still going on.

Processing and archiving is done by the Steinmetz Archives. After an introduction, this paper reports documentation instruments that have been developed to have easy access to and use of the datasets. It reports also the cleaning of a large part of all datasets and it goes shortly into weighting procedures developed, in order to get representative national samples.

The paper is completed with a (provisional) bibliography to the project.

+

Dit artikel is in mei te Ottawa gepresenteerd op een conferentie van de International Association for Social Science Information Service and Technology (IASSIST).

1. Introduction.

1.1. history.

The first survey with the title 'kwartaalonderzoek' (quarterly research project), as a part of a longitudinal research project, dates back to January 1972. It was the result of an initiative of the Baschwitz Institute for Masspsychology and Public Opinion and the Department of Methodology of the Institute for Political Science, both of the University of Amsterdam.

Both institutes felt the need to collect data from the adult population throughout the country, on a continuous basis ('t Hart, 1976).

The aims were:

1. to obtain data on the development of opinions within the Dutch population on relevant matters;
2. to consider wants and needs of teaching and research in the Faculty of Social Sciences, that is, being able to put questions concerning specific ends (ad hoc);
3. the possibility of being able to experiment in view of other research;
4. the possibility to perform replications of earlier research.

The participants to the researchgroup decided upon a questionnaire, of which one part was meant to be a fixed part, to appear in every following survey of the project, the other part variable, depending upon actuality, teaching and research needs and wants of the participating departments of the Subfaculty. Other departments of the subfaculty than the initial ones therefore were invited to send in questions for the variable part of the questionnaire and/or to participate in the researchgroup.

Sofar, 18 surveys have been carried out. The fieldwork for the first 14 surveys of the project was done by the

Netherlands Foundation of Statistics unlimited (NSS)⁺ at The Hague. The fieldwork for the latter five surveys of the project was done by the Netherlands Institute for Public Opinion (NIPO)⁺⁺ at Amsterdam. So, in fact the total number of surveys carried out so far is nineteen. Two surveys were carried out at the same time by each of the two bureaus. The main reason to pass on from the one bureau to the other was a financial one.

From the start until spring 1977 the processing of the data from this project was done by the local Social-Science Data Archive of the University of Amsterdam (ASDA)⁺⁺⁺.

From spring 1977 until now the data are being processed by the Steinmetz Archives. The ASDA ceased to exist after early 1977, as it became obsolete, due to the development of the Steinmetz Archives as the one national Social-Science Data Archive, having its residence in the same city and working on the same computers.

Two representatives of the Steinmetz Archives participate (as 'outside University' representatives) in the research-group.

1.2. recent developments.

With the change from the one research bureau to the other, the researchgroup also critically evaluated the questionnaire (that is, for its fixed part). This evaluation resulted in a reduction of this part of the questionnaire, and a simplification, leaving more space for the variable part of the questionnaire.

The subfaculty financed the 'cleaning' of the first fourteen 'waves' of the research project. The cleaning has been carried out by the Steinmetz Archives.

+ NV. v/h. Nederlandse Stichting voor Statistiek

++ Nederlands Instituut voor Publieke Opinie

+++ Amsterdams Sociaal-Wetenschappelijk Data Archief

One part of the cleaning project was the cleaning of the data from the several surveys themselves. This part of the project has been finished. Weighting procedures (that have been developed) form another part of and an addition to the cleaningproject. This part of the whole project will be finished before long.

Documentation, a third part of the cleaning project, has been finished for about 70%, and awaits supplementary financing to be completed.

Documentation, the cleaning itself, and weighting procedures will be described and illustrated in the following sections.

1.3. future plans and developments.

As a result of the cleaning, the weighting procedures developed and the documentation schemes developed, the data have reached such a level that they are now more inviting for research and teaching. Further plans have been made for a profound evaluation of all questions in the data-sets, an evaluation that is started already, and is also being carried out by the Steinmetz Archives, and financed by the Subfaculty of Political Science of the University of Amsterdam.

In future, cleaning and documentation will be carried out shorter after data-collection than was done until now. It seems, that the value of the dataset is also increasing to fit longitudinal analyses, such as time-series analyses and cohort analyses. This, for instance, will be one of the subjects for future publications of the researchgroup. However, the actual possibility of more publications, and especially more surveys to be held is rather uncertain. The necessary plans for economy measures of the University of Amsterdam seem to be, to economize on research in the Social Science Faculties. A discontinuation of the project can, beyond any doubt, be seen as a waste of money already spent on cleaning and documentation.

1.4. the data.

Ever since the start of the Continuous Survey Project, the budget for the project has never been sufficient to honor the initial name of the project : "kwartaalonderzoek" (quarterly survey project).

In table I we show the dates/periods that data-collection of the several surveys took place. In addition we show for each of the surveys the number of respondents and the number of variables included. The surveys are called 'waves' and numbered from one to eighteen.

As we mentioned before, all surveys until march 1977 (including nr.14a) have been carried out by the NSS. Ever since (to start with nr.14b) the surveys are carried out by NIPO.

Table I.

wave	dates/period	number of respondents	number of variables
1	january 1972	600	112
2	september 1972	631	148
3	december 1972	606	101
4	may 1973	567	115
5	september 1973	585	116
6	december 1973	588	127
7	may/june 1974	600	192
8	november 1974	593	189
9	june 1975	594	140
10	october 1975	620	164
11	january 1976	597	128
12	june 1976	608	149
13	november 1976	668	163
14a	march 1977	692	159
14b	march 1977	630	159
15	november 1977	583	191
16	january/february '78	594	252
17	june/july 1978	606	133
18	december 1978	599	200
19	total	11561	2938

The total number of respondents, including wave18, amounts to 11561, with an average of 608 respondents per sample. The total number of variables is 2938. All the surveys contain about 60 variables that belong to the fixed part of the questionnaire and therefore appear in every survey of the project. So, the total number of distinct variables in these 18 (19) surveys amounts to almost 1800. However, a deduction has to be made of this number because in some surveys replication for a number of subjects has taken place, although not belonging to the fixed part of the questionnaire. The average number of 'variable' variables per survey is about 90.

All datasets of this project form a part of study-number P0111 in the Steinmetz Archives. All the datasets from the CSP, i.e. every wave, except for those where data-collection has taken place less than two years ago, are free for public use. The datasets from the CSP that are less than two years old can only be obtained after consultation and consent of the original researchers, although until now there have never been serious problems concerning use of the data. The way the study is announced in the Catalogue and Guide of the Archives follows below(not yet completely updated).

STUDY NUMBER PD111
WAVE 1 UP TO WAVE13, WAVE14A, WAVE14B, WAVE15 UP TO WAVE18

TITLE KONTINU-ONDERZOEK

ACCESSIBILITY A: NO RESTRICTIONS (EXCEPT LESS THAN 2 YRS. OLD)

DEPOSITORS INSTITUUT VOOR WETENSCHAP DER POLITIEK - IWP, UNIV. V. AMSTERDAM * AMSTERDAM * BACHUITZ INSTITUUT, UNIV. V. AMSTERDAM * AMSTERDAM

DATA-COLLECTOR(S) HSS - NV. VH. HED. STICHTING VOOR STATISTIEK * DEN HAAG * UP TO WAVE14A * NIPD - NEDERLANDS INSTITUUT VOOR PUBLIEKE OPINIE * AMSTERDAM * FROM WAVE14B ON

YEAR 1972 * 1972 * 1972 * 1973 * 1973 * 1973 * 1974 * 1974 * 1975 * 1975 * 1976 * 1976 * 1976 * 1977 * 1977 * 1977 * 1978 * 1978 * 1978

SAMPLED UNIVERSE NAME OF COUNTRY: NETHERLANDS, AGE LIMITS: POPULATION ABOVE 17 YEARS, LOCATION OF UNITS OF OBSERVATION: NATIONAL

STATUS SPSS-SYSTEM FILE

NO. OF CASES 600 * 631 * 606 * 567 * 585 * 588 * 600 * 593 * 594 * 620 * 597 * 603 * 668 * 692 * 630 * 583 * 594 * 606 * 599

NO. OF VARIABLES 112 * 148 * 101 * 115 * 116 * 127 * 192 * 189 * 140 164 * 128 * 149 * 163 * 159 * 159 * 191 * 252 * 133 * 200

ABSTRACT *
POLITICAL INTEREST EXPOSURE TO RADIO-TV PROGRAMS AND INFORMATION IN PAPERS OPINION LEADERSHIP SENSE OF POLITICAL EFFICACY POLITICAL INFORMATION PARTISANSHIP ATTITUDE TO SOCIETAL CHANGE LOCAL-COSMOPOLITAN ATTITUDE TO COUNTRIES THAT THREATEN PEACE WAYS TO INFLUENCE PARLIAMENT ENGAGED IN ORGANIZATIONAL MEMBERSHIP ATTITUDE TO HOMOSEXUALITY KNOWLEDGE OF HOMOSEXUALITY. WAVE2 : POLITICAL INTEREST EXPOSURE TO TV-PROGRAMS AND INFORMATION IN PAPERS
OPINION LEADERSHIP SENSE OF POLITICAL EFFICACY POLITICAL INFORMATION PARTISANSHIP RANK-ORDERING OF GENERAL VALUES IMPORTANCE OF LOCAL-COSMOPOLITAN ATTITUDE COUNTRIES THAT THREATEN PEACE MOST ATTITUDE TO ATOMIC WAR PEACE INTERNATIONALISATION EUROPEAN GOVERNMENT WAYS TO INFLUENCE PARLIAMENT ENGAGED IN ORGANIZATIONAL MEMBERSHIP SATISFACTION WITH INCOME LEVEL CONSUMPTION ATTITUDE TO MARRIED WOMEN HAVING A JOB.
WAVE3 : POLITICAL INTEREST EXPOSURE TO TV-PROGRAMS INFORMATION IN PAPERS OPINION LEADERSHIP SENSE OF POLITICAL EFFICACY PARTISANSHIP POLITICAL EFFICACY OF LOCAL ORGANIZATIONS AID TO OTHER COUNTRIES IN CASE OF MILITARY ATTACK ECONOMIC CRISIS COUP D'ETAT LEADING TO DICTATORSHIP TREATY ON AID TO DEVELOPING COUNTRIES POLITICAL INFORMATION WAYS TO INFLUENCE PARLIAMENT ENGAGED IN ORGANIZATIONAL MEMBERSHIP SATISFACTION WITH INCOME LEVEL CONSUMPTION.
WAVE4 : POLITICAL INTEREST EXPOSURE TO TV-PROGRAMS AND INFORMATION IN PAPERS OPINION LEADERSHIP SENSE OF POLITICAL EFFICACY POLITICAL INFORMATION PARTISANSHIP POLITICAL EFFICACY OF LOCAL ORGANIZATIONS TREATY ON AID TO OTHER COUNTRIES IN CASE OF MILITARY ATTACK COUNTRIES THAT THREATEN PEACE MOST WAYS TO INFLUENCE PARLIAMENT ENGAGED IN ATTITUDE TO SOLDIERS' PRESSURE-GROUP TRADE-UNION ACTIVITIES WOMEN
ATTITUDE TO S-S PSYCHIC INSTABILITY AS A REASON NOT TO BE ENLISTED.

2. documentation.

2.1. masterfile.

A first access to the information of the data from the CSP is the 'masterfile'. Because of the structure of the information in all the files of the project, a fixed part and a variable part, a masterfile (spss-system file) was constructed containing only one case. However, the masterfile contains by now, after 18 waves, a total number of 1335 fully labelled variables (var labels and value labels). So, every distinct question that was put to the respondents since January 1972 appears in the masterfile. All different questions in a new survey of the project will be added to the masterfile.

The information from the -list fileinfo- in spss, is stored on tape, after superfluous information was stripped off. The information is available for everyone interested in these datasets. An example of a page of this information, by now well-known to a lot of people, is shown in figure 1.

2.2. wave.

Until now, as we mentioned before, 19 data files from this project have been processed. The raw data files were checked on total number of cards, and number of cards per respondent. Recoding was done and an spss-file was made of each of the datasets. Labelling was and is done out of the masterfile (write fileinfo for selected variables), that was updated already before processing the new data and creating a new spss-file.

All the spss-files are stored on tape. In addition to this, the frequencies output including statistics for all the waves is stored on tape.

Within due course a weighting variable with weightingfactors will also be included in every wave of the project (see page 25).

DOCUMENTATION ON THE 1335 VARIABLES IN FILE MASTER18

(CONTINUED)

ACTIVE VARIABLES ARE MARKED WITH AN *
 LABELS ARE REPORTED FOR ACTIVE VARIABLES ONLY

REL	VARIABLE	VARIABLE LABEL	MISSING	PRT
POS	NAME		VALUES	FMT
1235 *	VAR1218	CONCERN WITH PUBLIC HEALTH		9. 0
		0 NO PROBLEM AT ALL		
		1. NO VERY GREAT PROBLEM		
		2. GREAT PROBLEM		
		3. VERY GREAT PROBLEM		
		9. NA		
	MISSING			
1236 *	VAR1219	CONCERN WITH EDUCATION		9. 0
		0 NO PROBLEM AT ALL		
		1. NO VERY GREAT PROBLEM		
		2. GREAT PROBLEM		
		3. VERY GREAT PROBLEM		
		9. NA		
	MISSING			
1237 *	VAR1220	CONCERN WITH PUBLICITY OF ADMINISTRATION		9. 0
		0 NO PROBLEM AT ALL		
		1. NO VERY GREAT PROBLEM		
		2. GREAT PROBLEM		
		3. VERY GREAT PROBLEM		
		9. NA		
	MISSING			
1238 *	VAR1221	CONCERN WITH PRICE_RISES		9. 0
		0 NO PROBLEM AT ALL		
		1. NO VERY GREAT PROBLEM		
		2. GREAT PROBLEM		
		3. VERY GREAT PROBLEM		
		9. NA		
	MISSING			
1239 *	VAR1222	GOVERNMENT SHLD HANDLE:HOUSING SHORTAGE		9. 0
		0 SHLD DO NOTHING		
		1. A LITTLE		
		2. MUCH		
		3. VERY MUCH		
		9. NA		
	MISSING			
1240 *	VAR1223	GOVERNMENT SHLD HANDLE:UNEMPLOYMENT		9. 0
		0 SHLD DO NOTHING		
		1. A LITTLE		
		2. MUCH		
		3. VERY MUCH		
		9. NA		
	MISSING			
1241 *	VAR1224	GOVERNMENT SHLD HANDLE:POLLUTION		9. 0
		0 SHLD DO NOTHING		
		1. A LITTLE		
		2. MUCH		
		3. VERY MUCH		
		9. NA		
	MISSING			

2.3. datalib.

In addition to the archives tapes, all the waves are stored on another set of tapes, modlib tapes, for public use. These modlib tapes are used in a computerprogram called: 'datalib'.

Datalib is a public (computer) library, in which all studies of the Steinmetz Archives that have reached the status of a spss-file are stored. It was specially developed for this purpose by the Technical Centre of the Faculty of Social Sciences of the University of Amsterdam. Those who have access to the CD6-computers of the Academic Calculation Centre Amsterdam (SARA)⁺, therefore also have easy access to all data (spss-files) of the Steinmetz Archives, by simply copying the files out of the library, provided there are no restrictions on usage of the data of course. For use of files that do have restrictions one has to contact the Archives first. Evidently, studies that have restrictions can't be copied by users. Of course usage of all 'public' files has to be reported to the Archives.

A file that is copied out of the library can be used directly with spss or stap (Amsterdam appendix to spss). In figure 2 we give two examples using datalib.

Figure 2a.

```

HENK,CM22000,T17,IO177,NT1,NP.
ACCOUNT,$$$$$$$$$$.
ATTACH,TCLIB,ID=TCLIB.
LIBRARY,TCLIB.
DATALIB,WAVE7.
CATALOG,GTFILF,PFNM,ID=OWNER.

```

Figure 2b.

```

HENK,CM22000,T17,IO277,NT1,NP.
ACCOUNT,$$$$$$$$$$.
ATTACH,TCLIB,ID=TCLIB.
LIBRARY,TCLIB.
DATALIB,WAVE7=WAVE7,P0015=P0015.
CATALOG,WAVE7,PFNM,ID=OWNER.
CATALOG,P0015,PFNM,ID=OWNER.

```

+ Stichting Academisch Rekencentrum Amsterdam

2.4. kwicindex.

A second access to the contents of the data concerns the kwicindex. This index is based on all variable labels (in English) residing in the masterfile. First all the labels from the masterfile are written on disk. For further processing of this information, the Technical Center of the Social Sciences supplied another useful computerprogram, called : 'varkwic'. All the words in the labels are used as index terms, except for user-specified stop words. The variable number or short name is used as a reference on every line. This index is used as a resource during the processing of data from another survey of the project. The index is also a handy resource for users in finding particular variables. This index is also stored on tape and easy accessible for users. In figure 3 we show a small section of the index.

2.5. quick-reference-list.

A third entrance to the contents of the datasets of the CSP consists of the quick-reference-list (QRL). Actually, it is a matrix in which the variable numbers have been brought together with the numbers of the questions in a particular questionnaire, i.e. for all eighteen waves (processed sofar).

Using spss, these data have been formatted in such a way, in a matrix, that for each variable the matrix shows in what particular wave(s) the variable appears and the corresponding questionnumber in that/those particular wave(s). The QRL is one of the strongest entrances to the information of the datasets. Part of this matrix is shown in figure 4. Like the other documentation this QRL is also stored on tape and made easily accessible.

2.6. riqs documentation.

Documentation of the CSP is completed with a so-called Riqs-file. Riqs is used according to one of its two functions, namely as an information retrieval system.

IN FORMR TIMES,R VISITED	FOOTBALL_MATCHES	VAP1088
R SWITCHES	FOOTBALL_PROGRAMMS ON TV OFF	VAR1062
USING CARS SHLD BE	FORBIDDEN AT ALL	VAR714
LAST TIME R RD APT	FOREIGN NEWS	VAR312
R RDS APT:	FOREIGN NEWS,INT. RELATIONS	VAR008
NETH SHOULD HAVE INDPONT	FOREIGN POLICY	VAR135
MENTIONED PROBLEM:	FOREIGNERS	VAR972
ABORTION RULES SHLD BE DIFF. TO	FORGHNRS	VAR1319
IN A WHILE THIS CRISIS WILL BE	FORGOTTEN	VAR331
IN A WHILE ALL THIS WILL BE	FORGOTTEN	VAR336
INFLUENCE PARLIAMENT:	FORM AN ACTIONGROUP	VAR057
	IN FORMR TIMES,R VISITED FOOTBALL_MATCHES	VAR1088
	IN FORMR TIMES,R VISITED OTHER SPORTS	VAR1089
SPLIT AGR.	FORMULATION_SEE CODE_800K	VAR1015
	FORMULATION_SEE CODE_800K	VAR1006
RS LEFT-RIGHT RATING OF B.D.GAAY-FORTMAN		VAR475
	FRANCE HAS NUCLEAR WEAPONS	VAR129
R'S PLACE HAS	FREE NATURE	VAR694
CITY CENTERS SHLD BE	FREE OF CARS	VAR713
R RANKS:	FREEDOM	VAR118
PEOPLE'S CONCERN:	FREEDOM OF SPEECH	VAR912
R'S CONCERN WITH:	FREEDOM OF SPEECH	VAR001
IN ARMY	FREEDOM O.SPEECH AS IN SOCIETY	VAR263
	FREQUENCY R KEEPS INFORMED APT. POLLUTION	VAR691
R'S PLACE HAS	FRESH CLEAN AIR	VAR495
R LIKES TO BE INTERVIEWED ON	FRIDAY	VAR305
	FRIENDS ASK R ABOUT GOVT PROBLEMS	VAR012
	FRIENDS ASK R APT PLANS IN RESIDENCE ?	VAR010
SHLD HOMOSEXUAL TELL FAMILY OR	FRIENDS ?	VAR097
PEOPLE'S CONCERN:	FRIENDSHIP	VAR930
R'S CONCERN WITH:	FRIENDSHIP	VAR919
R RANKS:	REAL FRIENDSHIP	VAR120
RLIKES _HATES_ TO COME TO THE	FRONT	VAR1033
R LIKES TO COME TO THE	FRONT	VAR1017
	R HAS FUCTION IN SPORTS AS COACH	VAR359
CENTRAL HEATING NEEDS .. FOR	FUEL	VAR275
<IF NO CENTRAL HEATING>R USES .. F.FUEL		VAR276
SUNNY WEATHER IS NO	FUN ANYMORE	VAR495
	R HAS FUNCTION IN SPORTS AS ESCORT	VAR358
	R HAS FUNCTION IN SPORTS AS MANAGER	VAR357
	R HAS FUNCTION IN SPORTS AS REFEREE	VAR360
R THROWS DOORS WHEN HE IS	FURIOUS	VAR1026
R SAYS ANYTHING WHEN	FURIOUS	VAR629
R UNCONTROLLED WHEN	FURIOUS	VAR1028
R THROWS DOOPS WHEN	FURIOUS _DOES NOT_	VAR1043
R UNCONTROLLED WHEN	FURIOUS_ALWAYS CONTR.	VAR1047
R SAYS ANYTHING WHEN	FURIOUS_IS CONTROLD	VAR1049
PRESENT DRNG INTERVIEW: CHLD4 _F_		VAR1107
RS' LEFT-RIGHT RATING OF B.D.GAAY-FORTMAN		VAR475
	R IS NOT GAMBLE_MINDED	VAR1029
R DOESN'T LIKE _LIKES_ GAMBLING		VAR1048
NOISY SUPPORTERS ARE PART OF THE	GAME	VAR1073
DISADV.RESOURCES:	GAS. 2ND. ANSW.	VAR1201
DISADV.RESOURCES:	GAS. 1ST. ANSW.	VAR1200
DISADV. RESOURCES:	GAS. 3RD. ANSW.	VAR1202
ADVANTAGES OF RESOURCES:	GAS. 2ND. ANSW	VAR1189
ADVANTAGES OF RESOURCES:	GAS. 3RD. ANSW	VAR1190
ADVANTAGES OF RESOURCES:	GAS. 1ST. ANSW	VAR1183
	SURVEYS IN GENERAL IMPORTANT FOR RESP	VAR289
	SURVEYS IN GENERAL INTERSTING	VAR287
	SURVEYS IN GENERAL UNDERSTANDABLE	VAR283
	SURVEYS IN GENERAL USEFULL	VAR286
	R OFTEN GETS ANGRY WITH ARBITER	VAR633
R OFTEN NEVER_	GETS ANGRY WITH OPPONENTS	VAR1068
	R OFTEN GETS ANGRY WT OPPOSITE PLAYERS	VAR637

By using Riqs (Remote Information Query System) a documentation file was created.

The structure of the documentation file 'Kontinu file' is according to the Steinmetz Archives version of the international study description scheme(SDS)⁺, last discussed and agreed upon at the Copenhagen conference of 1977. Therefore this documentation file contains amongst others : information on sampling, number of variables, presence of backgroundvariables, English abstract of the contents etc.,etc. In addition to the SDS the file contains for all waves the complete questionnaire, together with the numbers of the questions, corresponding variable numbers and the English labels of the variables. Finally, if necessary, additional information per variable is included.

A Riqs-file consists of records. Records contain items and subitems. Every wave is considered as a record, further information as items and subitems. Using Riqs, one may search⁺⁺ for specific words in questionnaires/texts or SDS thereby using the retrieval facility. A small section of standard output from Riqs (search procedure with Riqs-online) is shown in figure 5. It is also possible to index⁺⁺⁺ words in the file, for instance words in the text of the questionnaire or variable labels using stop or key words, and using the particular wave(s) as a reference. Exact position of words can be traced then, |

⁺The new Steinmetz Archives version 3.0 of the international SDS is only 2 months old. All the information in the own former SDS from the Archives' holdings has been rearranged and adapted as much as possible to the agreed structure. The Steinmetz version contains additional items and leaves others out. The Riqs system is used for storage and retrieval.

⁺⁺One disadvantage of Riqs is that, unlike STAIRS, the program/system can't handle truncation (version 2.0).

⁺⁺⁺A second disadvantage of Riqs is that when indexing over a number of items for one subitem, the same words appearing in other subitems are also found in the index (version 2.0).

FIGURE 5.

1.	TITLE OF STUDY	KONTINU-ONDERZOEK
100.	_SUB_FILE_S_	WAVE1
101.	VRAAGNR	003
	VRAAGTEKST	SAMENSTELLING VAN HET GEZIN NAAR GESLACHT EN LEEFTIJD
	VARNR	VAR001 VAR002 VAR003
	VARLABEL	NR. OF PERSONS IN HOUSEHOLD O F 1M YRS. AND OLDER/ AGE OF RESPONDENT IN YEARS/ FAMILY SIZE
102.	VRAAGNR	008
	VRAAGTEKST	KUNT U MIJ ZEGGEN HOE VAAK U ACTUALITEITEN PROGRAMMA KIJKT OP DE TELEVISIE
	VARNR	VAR004
	VARLABEL	RESP. SEES NEWSREEL ON TELEVISION
103.	VRAAGNR	009
	VRAAGTEKST	HOE VAAK LUISTERT U NAAR ACTUALITEITEN PROGRAMMA'S OP DE RADIO
	VARNR	VAR005
	VARLABEL	RESP. LISTENS TO NEWS ON RADIO
104.	VRAAGNR	010
	VRAAGTEKST	HOE VAAK LEEST U IN DE KRANT NIEUWS OVER UW WOONPLAATS
	VARNR	VAR006
	VARLABEL	RESP. READS ABOUT : NEWS OF RESIDENCE
105.	VRAAGNR	011
	VRAAGTEKST	ALS ER IN DE KRANT NEDERLANDS NIEUWS STAAT ZOALS LOON - EN PRIJSBELEID OF REGERINGSPROBLEMEN IN DE HAAG HOE VAAK LEEST U ZELF OVER DAT SOORT ZAKEN
	VARNR	VAR007
	VARLABEL	RESP. READS ABOUT : GOVERNMENT PROBLEMS
106.	VRAAGNR	012
	VRAAGTEKST	EN ALS ER IN DE KRANT BUITENLANDS NIEUWS STAAT ZOALS SPANNINGEN OF BESPREKINGEN TUSSEN VERSCHILLENDE LANDEN KUNT U WEER VOLGENS KAARTJE AANGEVEN HOE VAAK U ZELF OVER DAT SOORT ZAKEN LEEST
	VARNR	VAR008
	VARLABEL	RESP. READS ABOUT : NEWS, INTERNATIONAL
107.	VRAAGNR	013
	VRAAGTEKST	ALS ER IN GEZELSCHAP OVER PLANNEN OF MAATREGELEN IN UW WOONPLAATS WORDT GESPROKEN DOET U DAN MEESTAL MEE AAN HET GESPREK LUISTERT U MET BELANGSTELLING OF LUISTERT U NIET HERT U GEEN BELANGSTELLING
	VARNR	VAR009
	VARLABEL	RESP. TALKS ABOUT PLANS OR DECISIONS IN RESIDENCE
108.	VRAAGNR	014
	VRAAGTEKST	STEL DAT KENNISSEN IETS OVER PLANNEN OF MAATREGELEN IN UW WOONPLAATS WILLEN WETEN IS ER DAN EEN GOEDE KANS DAT ZIJ DAT AAN U VRAGEN OF EERDER AAN IEMAND ANDERS
	VARNR	VAR010
	VARLABEL	FRIENDS ASK RESP. ABOUT PLANS IN RESIDENCE
109.	VRAAGNR	015

Figures 6 and 7 show two examples of indices on words in questionnaire respectively in variable labels.

This file is one of the strongest resources of information to the CSP. However the status of the file is still optional because finishing this part of the cleaning project awaits additional financing. The file, when completed and updated will be stored on tape and made available for experienced users of Riqs. Others may ask the Archives to perform a search procedure or construct indices using specific key words. The file will replace the one available and provisional codebook with several appendices.

2.7. original_forms.

One remaining access to the data of the CSP has been skipped so far. This entrance consists of the original forms of all the surveys, filled in partly by the respondents and partly by the interviewer. These forms of all the waves are still available.

These forms were of vital importance to the cleaning project of course. Further, the forms contain still information that has not been coded yet. This concerns for instance answers to open-ended questions. Every now and then students take care of this in small research projects, of which some are to be found in the bibliography at the end of this paper.

3. Cleaning.⁺

3.1. to start off with.

October 1977 the Steinmetz Archives agreed with the research-group of the CSP to carry out the cleaning project. Upon starting the control work all then present information was : fourteen rawdata files and fourteen spss-system files. In addition there was the so-called masterfile, the spss-system file containing one dummy case, and in principle all variables ever used in the CSP so far. Finally the former ASDA had produced a written codebook with some background information.

3.2. control.

The cleaning started with a number of control operations to be performed:

- a. control on the presence of all questions from the several waves in the masterfile;
- b. checking the correspondence of text and categories of questions appearing in more than one wave, the inventory of differences, and if necessary, creation of new variables;
- c. checking the correspondence of Dutch text and categories with English variable and value labels in the masterfile and the waves;
- d. control over and standardisation of 'missing values' of all variables in the masterfile and existing waves.
- e. an inventory was made of all wild codes and inconsistencies in the existing spss-system files.

In fact, this stage of the cleaning project meant answering about 16000 questions. For every variable in every wave (up to 14b) questions had to be answered like: is this variable appearing in this wave, if so, what is the questionnumber; is the text in correspondence with the text in other waves.

+ for more detailed information see the report of De Vries and De Vries.

And many questions more had to be answered. The same goes for the categories. This all resulted in a manuscript of about 150 pages with instructions for carrying out the actual cleaning work. It resulted in a preliminary form of the quick-reference-list, and a drawer full of notes and memos. This stage of the cleaning project took about two-third of the time/budget.

3.3. the cleaning work

The second stage of the cleaning project started with revising and improving the masterfile on the basis of the results found in the first stage. Wild codes and inconsistencies in the different files were listed for the respective respondents. Improvements were carried out on the basis of the information in the original questionnaires. Every wave was revised and improved on the basis of the results gained in the first two stages (improved masterfile and checked inconsistencies). Some variables were split in more, i.e. new variables were created. Some variables were added anew from the existing raw data files, because there were serious doubts on proper coding. For a number of variables the non-response category could be reduced substantially. Nevertheless, at this point the 'cleaners' started to doubt the gain in reliability, because of the impossibility to perform this kind of operation systematically for all variables, in view of time and budget. The quality of the original spss-system files differed substantially, because they were made at different times by different people. This part of the project, the cleaning itself, took about one-third of the total time/budget.

3.4. documentation

The third stage of the cleaning project can be defined as the documentation phase. Broadly, this meant that all the waves of the CSP had to be made accessible and usable in most optimal form.

In an early stage of the project it was already decided

that as much documentation as possible should be in machine readable form. Therefore the already discussed instruments were constructed. These instruments are available now for users.

Some of the instruments were already present, like the 14 existing (but to be revised) spss-sytem files. And datalib was and is not only meant for the CSP-waves.

Some of the instruments -though present- were not yet available for users, like the kwic-index and the masterfile. Some of the instruments were not present at all, and had to be build up on the basis of the results gained in the first two stages. This concerns for instance the quick-reference-list.

The Riqs-file was, in a narrow sense, already existing as part of the larger documentation file of all the Steinmetz holdings, including the questionnaire. Now it consists as a seperate Riqs-file, called 'Kontinu file'. The file is not yet completed and updated, because as we mentioned already, this part of the work awaits additional financing. Description of differences between questions that are brought together in one variable, categories to open-ended questions or categories not covered by value labels could not completely be fitted in the already discussed instruments. An additional document (in the masterfile or Riqs-file or as seperate file) will be created.

3.5. summary

At the end of their report the 'cleaners' made some general remarks concerning the cleaning.

They say that their initial enthusiasm faded away in view of the complexity of the problems, problems that are so specific because of the continuous character of the research project at hand.

At the same time they state that because of this continuous character of the project it makes the cleaning all the more necessary.

Another remark concerns the checking of inconsistencies. The reason why this made up such a large part of the cleaning work was the existence of 'blocks of questions', filter-questions, extensive interviewer instructions etc. Mistakes

caused by those structures are sometimes irretrievable. Consistency can be gained in such instances at the cost of growing non-response categories. Consistency in the wording of the questions was lacking more than was necessary, because often the wording was not intended to be different.

4. Weighting procedures.⁺

4.1. why weighting.

Sampling procedures for the CSP have always been carried out, using families as sampling units, so meant to be representative for family size. In fact this is not the case, because the difficulty of a sample of persons, drawn as a sample of families, is that single persons are mostly underrepresented, as there is more chance of finding single persons not at home.

In order to get to a representative sample, we may make adjustments where sample data deviate from population data, which means weighting the sample.

As an addition to and follow-up of the cleaning project, it was decided that all waves should be weighted, each of them forming a representative sample of persons of the Dutch population.

4.2. defining conditions for weighting.

If we decide to weight a sample to a population for which the sample should be representative, amongst others we have to consider the following:

- a. what variables will be used to perform the weighting on ;
- b. what point(s) in time to choose for comparing sample data with population data ;
- c. collection of population data ;
- d. drafting a strategy, that is, the way the weighting will be carried out and the order of the variables in the strategy that is developed ;
- e. check where the sample(s) deviate from the population data ;
- f. the weighting itself ;
- g. investigating the effectivity of the weighting.

+ for more detailed information see the report of Bakker.

4.3. population_data.

The selected variables for weighting the CSP were the following : family-size, age, sex, civil status, province, and degree of urbanisation. But problems of non-correspondence between categories in the sample data and in the population data from the Central Bureau of Statistics concerning degree of urbanisation, caused this variable to be excluded from the weighting procedures.

4.4. point_in_time.

The several waves (up to 14b) of the CSP cover a time-period from January 1972 to March 1978. For most variables population data were available for each year via the 'Statistisch Zakboek'. However for the variable family-size only data from the Census of 1971 were and are available. Population data (reduced to the size of 600) were compared for several years. None of the four variables changed significantly during the years 1972-1978. On the basis of these results it was decided to control and weight the samples for population data stemming from January 1976, except for the variable family-size (February 1971).

4.5. weighting_strategy.

After the decision on what variables the weighting procedure will be carried out, and the collection of the population data, a weighting strategy was developed. One strategy was developed to apply for all the waves. Weighting will be carried out (is actually being done now, mid-april) according to the following order : first; family size or in fact now the number of persons of 17 yrs. and older in the family, second; age and civil status, third ; province and sex. One problem, that of the transition of a sample of families to a sample of persons was satisfactorily solved. A second problem in this weighting strategy was the number of categories on which the weighting procedure to be carried out.

The three step procedure will be carried out on the basis of adjustment of $5 \times 16 \times 22 = 1760$ cells (by way of assigning weighting factors). This implies, with an average of 600 respondents per wave, many empty cells. Therefore the number of categories has to be reduced drastically. A number of criteria has been decided upon for reduction.

4.6. the weighting.

For every wave of the CSP will be calculated what variables deviate significantly from the population data. A specially developed computerprogram for performing weighting procedures (Koens, 1975 ; De Boon, 1978), will adjust frequencies for blocks in the samples to frequencies for blocks in the population by assigning weightingfactors to respondents inside such defined blocks.

The program calculates also the effective sampl size, of which the variance is equal to the variance of a simple random sample of that size (Kish, 1965, p.88), and adjusts the weightingfactors in such a way that this size will be the result.

A variable containing weightingfactors will be added to every wave (up to 14b). It is up to the users of the datasets to work with unweighted or weighted files.

4.7. evaluation of weighting.

After weighting procedures have been completed, it will be investigated in what way the weighting has been effective, by controlling for every sample, deviation from the population data before and after weighting.

Another point of evaluation is to investigate if the adjustments concern systematic or accidental deviations.

5. Actual status of the project.

By now, mid-april, number 18 in the series of the CSP is completed. Fieldwork for number 19 is scheduled for mid-may next. Financing is also insured for another survey mid-september.

The cleaning project as agreed on in october 1977 has been completed, i.e. up to and including wavel4b. In case the waves from number 15 onwards should be cleaned too in the next future, new arrangements between the researchgroup and the Steinmetz Archives have to be made.

As far as documentation is concerned, the masterfile and kwic-index have been updated to wavel8, the quick-reference-list has been updated to wavel7.

Weighting the waves up to wavel4b is in its final stage. This goes also for the first investigations into longitudinal analyses of the files that have been cleaned sofar.

(Provisional)

6. Bibliography to the Continuous Survey Project.

- 1.-Bakker, T.P.V., De weging van het kontinu-onderzoek, Amsterdam, november 1978.(b)
- 2.-Baschwitz Instituut, leeronderzoek studenten, superv. Tellegen, S. and Koopman, M., Onderzoek naar arbeidssatisfactie werkende vrouw, Amsterdam, march 1975
- 3.-Baschwitz Instituut, leeronderzoek studenten, superv. Tellegen, S. and Meyer zu Schlochteren, B. Meningen over afkeuring op S-5, Amsterdam, april 1974.
- 4.-Baschwitz Instituut, leeronderzoek studenten and medewerkers,
Vrouwen en mannen in werk en politiek, Amsterdam, september 1974-march 1976.
- 5.-Boon, A. den, and Hart, H.'t, de ombudsman van de VARA, Amsterdam, 1976.(a)
- 6.-Boon, M.E., and Niemöller, K., Volgorde van dimensionaliteit in:
Acta Politica, vol XI. nr.2., april 1976.
- 7.-Brug, H.M. van den, and Koopman, M.J., Sport en politiek, Amsterdam, 1976.(a)
- 8.-Brug, H.M. van den, and Koopman, M.J., Sport en politiek II.
(replikaties en aanvullende bevindingen), Amsterdam, 1976. (a)
- 9.-Brug, H.v.d., Over de betekenis van de vraagstelling bij de konstruktie
van een vragenlijst, Amsterdam, november 1978.(b)
- 10.-Eijk, C. van der, Opvattingen over defensie in het Nederlandse massapubliek,
Amsterdam, 1976.(a)
- 11.-Eijk, C. van der, Media als bron van politieke informatie, Amsterdam 1978.(b)
- 12.-Hart, H. 't, Milieubeheer en Sociale Wetenschappen, in: "Handboek voor het
milieubeheer" (hfst. III), Vermande, IJmuiden, 1976.
- 13.-Hart, H. 't, Inleiding, in: Kontinu-onderzoek, enquetes 1972-1976,
Amsterdam, december 1976.(a)
- 14.-Hart, H. 't, Non-response in het kontinu-onderzoek FSW-A, Amsterdam, 1976.(a)
- 15.-Hart, H. 't, De energiecrisis, Amsterdam, 1976.(a)
- 16.-Hart, H. 't, Kontinu-onderzoek, enquetes 1976-1978, Inleiding, Amsterdam, december
1978.(b)
- 17.-Hart, H. 't, Andere aanwezigen bij het vraagggesprek, Amsterdam, november 1978.(b)
- 18.-Hart, H. 't, Political efficacy, nog maar eens geexploreed, Amsterdam, november
1978.(b)
- 19.-Kok, W.J.P., Meningen over de VVDM, in Twintig, vol.7 nr. 11, 1977.
- 20.-Koopman, M., Verschiller en overeenkomsten tussen twee vergelijkbare te zelfder
tijd afgenomen enquetes; het bureau-effekt op survey data, Amsterdam, november 1978
- 21.-Meilof-Oonk, S., and Valkman, O., Meningen over homosexualiteit, publikatie
Stichting tot Bevordering Onderzoek Minderheden, Amsterdam 1973.
- 22.-Middendorp, J., Non-respondenten in het Kontinu-onderzoek, Amsterdam, 1979.

- 23.-Schmidt, O., Het kwartaalonderzoek 1971-1975, (verslag t.b.v. subfaculteitsraad FSW-A), Amsterdam, june 1975.
- 24.-Schmidt, O., Het kwartaalonderzoek 1971-1975 Acta Politica, vol. XI, nr.1, january 1976.
- 25.-Schmidt, O., De energiecrisis in 1973, enkele politieke aspecten, in: Acta Politica vol XI, october 1976.
- 26.-Schrik, H., and Vries, G. de, De toegankelijkheid van en toegang tot het kontinuonderzoek, Amsterdam, november 1978.(b)
- 27.-Tubbergen -Ligthart, H., Kontinuonderzoek, IWP-Baschwitz Instituut, controle en weging van de steekproef, Amsterdam, 1974.
- 28.-Tubbergen-Ligthart, H., De buitenshuis werkende gehuwde vrouw in de publieke opinie. Doktoraal werkstuk psychologie, superv. Hart. H. 't, and Rijswijk, M.J.van, Amsterdam, december 1975.
- 29.-Veenhuizen, J.P., Evenementen lijst (jan '72-dec'78) t.b.h. het kontinuonderzoek, i.v.m. analyses in de tijd, Amsterdam, march 1979 .
- 30.-Vries, G. de, and Vries, R. de, De cleaning van het kwartaal-onderzoek, Amsterdam, november 1978 .(b)
- 31.-Wetering, K. van de , Stage verslag naar aanleiding van openvraagstelling in wave 14 A, betreffende energie, Zevenbergen, march 1979.

(a) Publication by the research workgroup "kwartaalonderzoek" of the subfaculty of Political Science of the University of Amsterdam, printed by University Press, Amsterdam, december 1976.

(b) Publication by the researchgroup "kwartaalonderzoek" of the subfaculty of Political Science of the University of Amsterdam, printed by University Press, Amsterdam , may 1979.

References.

- Bakker, T.P.V., De weging van het Kontinu-onderzoek, Amsterdam, november 1978.
- Boon, A. den , Kortweg, computerprogramma voor het wegen van steekproeven, Baschwitz Instituut, Amsterdam, 1978.
- Bormann, L., Remote Information Query System, Vogelback Computing Center Northwestern University, Evanston, september 1976.
- Hart, H. 't , Inleiding, in: Kontinu-onderzoek, enquetes 1972-1976, Amsterdam, december 1976.
- Hart, H. 't , Inleiding, in: Kontinu-onderzoek, enquetes 1976-1978, Amsterdam, april 1979.
- Kish, L., Survey sampling, New York, 1965.
- Koens, M., Weging, computerprogramma voor het wegen van steekproeven, Baschwitz Instituut, Amsterdam, 1974.
- Nie, N.H., c.s., Statistical Package for the Social Sciences, New York, 1975.
- Schrik, H., and Vries, G. de, De toegankelijkheid van en de toegang tot het Kontinu-onderzoek, Amsterdam, november 1978.
- Vries, G. de and Vries, R. de, De cleaning van het kwartaalonderzoek, Amsterdam, november 1978.