EUROPEAN COMMISSION
EUROSTAT

Directorate D: Single Market, Employment and Social statistics
Unit D-2: Living conditions and social protection

Luxembourg, 15 June 2006
EU-SILC/BB D(2005)

EU-SILC USER DATABASE DESCRIPTION (draft)

1. AIM OF EU-SILC

EU-SILC will gather comparative statistics on income distribution and social exclusion from the 25 EU members states, Norway and Iceland.

It will provide two types of annual data:

Cross-sectional data pertaining to a given time or a certain time period with variables on income, poverty, social exclusion and other living conditions, and

Longitudinal data pertaining to individual-level changes over time, observed periodically over a four years period.

2. SAMPLE DESIGN

EU-SILC allows member states a large flexibility in terms of sampling design.

Depending on the country, micro-data could come from:

- two or more national sources (surveys and/or registers);
- one or more existing national sources combined or not with a new survey;
- A new harmonised survey to meet all EU-SILC requirements called the integrated design.

In this way, the cross-sectional and longitudinal data may come from one or different sources. Consequently, a linkage between longitudinal and cross-sectional data,

Commission européenne, L-2920 Luxembourg. Telephone: (352) 43 01-1.
Office: BECH D2/704. Telephone: direct line (352) 43 01-37330.

http://europa.eu.int/comm/eurostat/
E-mail: bruno.bernard@cec.eu.int

either between two consecutive cross-sectional data is not precluded, and would normally be possible when the two types of records come from the same source.

The only constraint is that for both, the cross-sectional and longitudinal components, all household and personal data will be linkable through unique keys.

3. SAMPLE SIZE

On the basis of various statistical and practical considerations and precision requirements for the most critical variables, the minimum effective sample sizes to be achieved will be as set out in the table I.

Sample size for the longitudinal component refers, for any pair of consecutive years, to the number of households successfully interviewed in the first year in which all or at least a majority of the household members aged 16 or over are successfully interviewed in both the years.

For the cross-sectional component, the plans are to achieve the minimum effective sample size of around 121.000 households in the EU as a whole (127.000 including Iceland and Norway). The allocation of the EU sample among countries represents a compromise between two objectives: the production of results at the level of individual countries, and production for the EU as a whole.

Table I. Minimum effective sample size for countries using a sample of households/address

	Households		**Persons aged 16 or over to be interviewed**	
	Cross-sectional	**Longitudinal**	**Cross-sectional**	**Longitudinal**
	1	2	3	4
EU-Member States				
Belgium	4750	3500	8750	6500
Czech Republic	4750	3500	10 000	7500
Denmark	4250	3250	7250	5500
Germany	8250	6000	14 500	10 500
Estonia	3500	2750	7750	5750
Greece	4750	3500	10 000	7250
Spain	6500	5000	16 000	12 250

France	7250	5500	13 500	10 250
Ireland	3750	2750	8000	6000
Italy	7250	5500	15 500	11 750
Cyprus	3250	2500	7500	5500
Latvia	3750	2750	8500	6500
Lithuania	4000	3000	9000	6750
Luxembourg	3250	2500	6500	5000
Hungary	4750	3500	10 250	7750
Malta	3000	2250	7000	5250
Netherlands	5000	3750	8750	6500
Austria	4500	3250	8750	6250
Poland	6000	4500	15 000	11 250
Portugal	4500	3250	10 500	7500
Slovenia	3750	2750	9000	6750
Slovakia	4250	3250	11 000	8250
Finland	4000	3000	6750	5000
Sweden	4500	3500	7500	5750
United Kingdom	7500	5750	13 750	10 500
Total of EU Member States	**121000**	**90750**	**251 000**	**187 750**
Iceland	2250	1700	3750	2800
Norway	3750	2750	6250	4650
Total including Iceland and Norway	**127000**	**95200**	**261 000**	**195 200**

Table II. Minimum effective sample size for countries using a sample of persons

	Households and selected respondent		Persons aged 16 or over	
	Cross-sectional	**Longitudinal**	**Cross-sectional**	**Longitudinal**
	1	2	3	4
EU-Member States				
Belgium	6500	5000	12000	9250
Czech Republic	7500	5750	15750	12250
Denmark	5500	4250	9500	7250
Germany	11000	8000	19250	14000
Estonia	5750	4500	12750	9500
Greece	7500	5500	15750	11500
Spain	12000	9250	29500	22750
France	10250	7750	19000	14500
Ireland	6000	4500	12750	9750
Italy	11750	8750	25000	18750
Cyprus	5750	4250	13250	9250
Latvia	6500	5000	14750	11750
Lithuania	6750	5000	15250	11250
Luxembourg	5000	3750	10000	7500
Hungary	7750	5750	16750	12750
Malta	5250	4000	12250	9250
Netherlands	6500	5000	11500	8750
Austria	6500	4750	12750	9250
Poland	11250	8500	28250	21250
Portugal	8000	5750	18750	13250
Slovenia	6750	5000	16250	12250
Slovakia	8250	6250	21250	15750

Finland	5000	3750	8500	6250
Sweden	5750	4500	9500	7500
United Kingdom	10250	8000	18750	14500
Total of EU Member States	189000	142500	399000	300000
Iceland	3000	2000	5000	3250
Norway	4750	3500	8000	6000
Total including Iceland and Norway	196750	344750	412000	309250

Countries using a sample of persons (the selected respondent) must select an extra sample to cover population 14 and 15 years old.

4. DATABASE STRUCTURE

The domains and areas covered by the survey are listed below and are collected at two different levels:

Household level:

BASIC DATA (B)	Basic household data including degree of urbanisation
INCOME (Y)	Total household income (gross and disposable)
	Gross income components at household level
SOCIAL EXCLUSION (S)	Housing and non-housing related arrears
	Non-monetary household deprivation indicators, including problems in making ends meet, extent of debt and enforced lack of basic necessities
	Physical and social environment
LABOUR INFORMATION (L)	Child care
HOUSING (H)	Dwelling type, tenure status and housing conditions
	Amenities in dwelling
	Housing costs

Personal level:

BASIC DATA (B)	Basic personal data
	Demographic data
EDUCATION (E)	Education, including highest ISCED level attained
LABOUR INFORMATION (L)	Basic labour information on current activity status and on current main job, including information on last main job for unemployed
	Basic information on activity status during income reference period
	Total number of hours worked on current second/third … jobs
	Detailed labour information
	Activity history
	Calendar of activities
HEALTH (H)	Health, including health status and chronic illness or condition
	Access to health care
INCOME (Y)	Gross personal income, total and components at personal level

Following the structure of the main database, the different variables are distributed in four different files:

Household Register (D)

Personal Register (R)

Household Data (H)

Personal Data (P)

The household register file (D) must contain every household (selected + substituted + split off (longitudinal only)), also those where the address could not be contacted or which could not be interviewed.

In the other files records related to a household will only exist if the household has been contacted (DB120 = 11 (or DB110 = 1)) AND has a completed household interview in the household data file (H) (DB130 = 11) AND at least one member has complete data in the personal data file (P) (RB250 = 11, 12 or 13 => DB135 = 1). This member must be the selected respondent (RB245 = 2) if this mode of selection is used.

The personal register file (R) must contain a record for every person currently living in the household or temporarily absent. In the longitudinal component (initial household) this file must contain also a record for every person moved out or died since previous wave and for every person who lived in the household at least three months during the income reference period and was not recorded otherwise in the register of this household.

The personal data file (P) must contain a record for every eligible person (RB245 = 1, 2 or 3) for whom the information could be completed from interview and/or registers (RB250 = 11, 12 or 13).

5. KEY VARIABLES

The four D, H, R and P files have to be adequately linked:

All observation from P file must have a univocal link to the three other files.

All observations from R file must have a univocal link to a D file observation.

All observations from H file must have a univocal link to a D file observation.

For that purpose, the variables:

- "COUNTRY": DB020, RB020, HB020 and PB020
- "Household ID": DB030 and HB030 and RB040* (* only longitudinal)
- "Personal ID": RB030 and PB030.

are used as key variables.

Note that Personal ID are constructed with Household ID and two more digits.

For longitudinal files, the link between R and D files is done with the variables RB040 and DB030. In case of split-off household, the people who leave the initial household will have two observations in the R file. The first one linked to the initial household and the second one linked to the new (split-off) household. As his personal ID cannot change and is still constructed with the original Household ID, we need this second variable to link the split-off H-observation to the split-off R-observation.

6. LINK BETWEEN YEARS AND WAVES

6.1. Cross-sectional data

In the case of “integrated” or pure panel sample design, all Household ID and Personal ID from people belonging to several cross-sectional data are in principle modified across the different years. But, for variance of estimators calculation, a link between data from Year n-1 and Year n has been created. It has been done with the variable RB041

This variable will be coded with 7 digits. The first 6 digits correspond to the Personal id (variable RB030) of the previous wave or last interview, if not contacted in previous wave.

The digit 7 corresponds to the last wave (number) that person was contacted. If the person has never been contacted the variable will be coded as 0.

6.2. Longitudinal data

It has been agreed by gentlemen agreement that all member states will provide each year the D, H, R and P files from all the panels (subsample) who have at least 2 years of data.

Let see in the example below the data each country has to provide in 2010. We’ll have 4 years data from subsample 1’, 3 from subsample2’ and 2 from subsample3’.

Subsamples to transmit to Eurostat in year 2010

Year of survey	Subsample 4	Subsample 1’	Subsample 2’	Subsample 3’	Subsample 4’	Subsample 1’’	Subsample 2’’
2004							
2005							
2006							
2007							
2008							
2009							

We intend to provide internally and to researchers (and possibly public) each year and for each country:

1 four years panel (subsample1’ 2005-2008)

2 three years panels (subsample1’ and subsample 2’ 2006-2008)

3 two years panels (subsample 1’, subsample 2’ and subsample3’ 2007-2008)

Across the years, in the same subsample, all personal ID and Household ID remain the same. There is no link between different subsamples.

The revision policy of the files has yet to be dertermined..

7. LIST OF VARIABLES

7.1. Primary target variables

Listed below all variables contained in D, R H and P file. On the head of each variable is indicated if the variables are part of the longitudinal files (L), cross-sectional files (X) or both (X-L)

HOUSEHOLD REGISTER (D-FILE)

X-L DB010: YEAR OF THE SURVEY
X-L DB020: COUNTRY
X-L DB030: HOUSEHOLD ID
X-L DB040: REGION
X-L DB050: PRIMARY STRATA
X-L DB060: PSU-1 (FIRST STAGE)
X-L DB062: PSU-2 (SECOND STAGE)
X-L DB070: ORDER OF SELECTION OF PSU
X-L DB075: ROTATIONAL GROUP
X-L DB080: HOUSEHOLD DESIGN WEIGHT
X-L DB090: HOUSEHOLD CROSS-SECTIONAL WEIGHT
X-L DB100: DEGREE OF URBANISATION
L DB110: HOUSEHOLD STATUS
X-L DB120: CONTACT AT ADDRESS
X-L DB130: HOUSEHOLD QUESTIONNAIRE RESULT
X-L DB135: HOUSEHOLD INTERVIEW ACCEPTANCE

PERSONAL REGISTER (R-FILE)

X-L RB010: YEAR OF THE SURVEY
X-L RB020: COUNTRY
X-L RB030: PERSONAL ID
L RB040: CURRENT HOUSEHOLD ID
X RB041: PERSONAL ID
X RB050: PERSONAL CROSS-SECTIONAL WEIGHT
L RB060: PERSONAL BASE WEIGHT
X-L RB070: MONTH OF BIRTH
X-L RB080: YEAR OF BIRTH
X-L RB090: SEX
L RB100: SAMPLE PERSON OR CO-RESIDENT
L RB110: MEMBERSHIP STATUS
L RB120: MOVED TO
L RB140: MONTH MOVED OUT OR DIED
L RB150: YEAR MOVED OUT OR DIED
L RB160: NUMBER OF MONTHS IN HOUSEHOLD DURING THE INCOME REFERENCE PERIOD
L RB170: MAIN ACTIVITY STATUS DURING THE INCOME REFERENCE PERIOD

L RB180: MONTH MOVED IN
L RB190: YEAR MOVED IN
X-L RB200: RESIDENTIAL STATUS
X-L RB210: BASIC ACTIVITY STATUS
X-L RB220: FATHER ID
X-L RB230: MOTHER ID
X-L RB240: SPOUSE/PARTNER ID
X-L RB245: RESPONDENT STATUS
X-L RB250: DATA STATUS
X-L RB260: TYPE OF INTERVIEW
X-L RB270: PERSONAL ID OF PROXY
X RL010: EDUCATION AT PRE-SCHOOL
X RL020: EDUCATION AT COMPULSORY SCHOOL
X RL030: CHILD CARE AT CENTRE-BASED SERVICES
X RL040: CHILD CARE AT DAY-CARE CENTRE
X RL050: CHILD CARE BY A PROFESSIONAL CHILD-MINDER AT CHILD'S HOME OR AT CHILDMINDER'S HOME
X RL060: CHILD CARE BY GRAND-PARENTS, OTHERS HOUSEHOLD MEMBERS (OUTSIDE PARENTS),OTHER RELATIVES, FRIENDS OR NEIGHBOURS
X RL070: CHILDREN CROSS-SECTIONAL WEIGHT FOR CHILD CARE

HOUSEHOLD DATA (H-FILE)

X-L HB010: YEAR OF THE SURVEY
X-L HB020: COUNTRY
X-L HB030: HOUSEHOLD ID
X-L HB040: DAY OF HOUSEHOLD INTERVIEW
X-L HB050: MONTH OF HOUSEHOLD INTERVIEW
X-L HB060: YEAR OF HOUSEHOLD INTERVIEW
X-L HB070: PERSON RESPONDING THE HOUSEHOLD QUESTIONNAIRE
X-L HB080: PERSON 1 RESPONSIBLE FOR THE ACCOMMODATION
X-L HB090: PERSON 2 RESPONSIBLE FOR THE ACCOMMODATION
X-L HB100: NUMBER OF MINUTES TO COMPLETE THE HOUSEHOLD QUESTIONNAIRE
X-L HH010: DWELLING TYPE
X-L HH020: TENURE STATUS
X-L HH030: NUMBER OF ROOMS AVAILABLE TO THE HOUSEHOLD
X-L HH031: YEAR OF CONTRACT OR PURCHASING OR INSTALLATION
X-L HH040: LEAKING ROOF, DAMP WALLS/FLOORS/FOUNDATION, OR ROT IN WINDOW FRAMES OR FLOOR
X-L HH050: ABILITY TO KEEP HOME ADEQUATELY WARM
X-L HH060: CURRENT RENT RELATED TO OCCUPIED DWELLING
X-L HH061: SUBJECTIVE RENT
X HH070: TOTAL HOUSING COST
X-L HH080: BATH OR SHOWER IN DWELLING
X-L HH090: INDOOR FLUSHING TOILET FOR SOLE USE OF HOUSEHOLD
X-L HS010: ARREARS ON MORTGAGE OR RENT PAYMENTS
X-L HS020: ARREARS ON UTILITY BILLS
X-L HS030: ARREARS ON HIRE PURCHASE INSTALMENTS OR OTHER LOAN PAYMENTS 176
X-L HS040: CAPACITY TO AFFORD PAYING FOR ONE WEEK ANNUAL HOLIDAY AWAY FROM HOME
X-L HS050: CAPACITY TO AFFORD A MEAL WITH MEAT, CHICKEN, FISH (OR VEGETARIAN EQUIVALENT) EVERY SECOND DAY
X-L HS060: CAPACITY TO FACE UNEXPECTED FINANCIAL EXPENSES
X-L HS070: DO YOU HAVE A TELEPHONE (INCLUDING MOBILE PHONE)?

X-L HS080: DO YOU HAVE A COLOUR TV?
X-L HS090: DO YOU HAVE A COMPUTER?
X-L HS100: DO YOU HAVE A WASHING MACHINE?
X-L HS110: DO YOU HAVE A CAR?
X-L HS120: ABILITY TO MAKE ENDS MEET
X-L HS130: LOWEST MONTHLY INCOME TO MAKE ENDS MEET
X-L HS140: FINANCIAL BURDEN OF THE TOTAL HOUSING COST
X-L HS150: FINANCIAL BURDEN OF THE REPAYMENT OF DEBTS FROM HIRE PURCHASES OR LOANS
X HS160: PROBLEMS WITH THE DWELLING: TOO DARK, NOT ENOUGH LIGHT
X HS170: NOISE FROM NEIGHBOURS OR FROM THE STREET
X HS180: POLLUTION, GRIME OR OTHER ENVIRONMENTAL PROBLEMS
X HS190: CRIME VIOLENCE OR VANDALISM IN THE AREA
X-L HY010: TOTAL HOUSEHOLD GROSS INCOME
X-L HY020: TOTAL DISPOSABLE HOUSEHOLD INCOME
X-L HY022: TOTAL DISPOSABLE HOUSEHOLD INCOME BEFORE SOCIAL TRANSFERS OTHER THAN OLDAGE AND SURVIVOR'S BENEFITS
X-L HY023: TOTAL DISPOSABLE HOUSEHOLD INCOME BEFORE SOCIAL TRANSFERS INCLUDING OLDAGE AND SURVIVOR'S BENEFITS
X-L HY025: WITHIN-HOUSEHOLD NON-RESPONSE INFLATION FACTOR
X-L HY030G/HY030N: IMPUTED RENT
X-L HY040G/HY040N: INCOME FROM RENTAL OF A PROPERTY OR LAND
X-L HY090G/HY090N: INTEREST, DIVIDENDS, PROFIT FROM CAPITAL INVESTMENTS IN UNINCORPORATED BUSINESS
X-L HY050G/HY050N: FAMILY/CHILDREN RELATED ALLOWANCES
X-L HY060G/HY060N: SOCIAL EXCLUSION NOT ELSEWHERE CLASSIFIED
X-L HY070G/HY070N: HOUSING ALLOWANCES
X-L HY080G/HY080N: REGULAR INTER-HOUSEHOLD CASH TRANSFER RECEIVED
X-L HY100G/HY100N: INTEREST REPAYMENTS ON MORTGAGE
X-L HY110G/HY110N: INCOME RECEIVED BY PEOPLE AGED UNDER 16
X-L HY120G/HY120N: REGULAR TAXES ON WEALTH
X-L HY130G/HY130N: REGULAR INTER-HOUSEHOLD CASH TRANSFER PAID
X-L HY140G/HY140N: TAX ON INCOME AND SOCIAL CONTRIBUTIONS
X-L HY145N: REPAYMENTS/RECEIPTS FOR TAX ADJUSTMENT

PERSONAL DATA (P-FILE)

X-L PB010: YEAR OF THE SURVEY
X-L PB020: COUNTRY
X-L PB030: PERSONAL ID
X PB040: PERSONAL CROSS-SECTIONAL WEIGHT
L PB050: PERSONAL BASE WEIGHT
X PB060: PERSONAL CROSS-SECTIONAL WEIGHT FOR SELECTED RESPONDENT
X-L PB070: PERSONAL DESIGN WEIGHT FOR SELECTED RESPONDENT
L PB080: PERSONAL BASE WEIGHT FOR SELECTED RESPONDENT
X-L PB090: DAY OF THE PERSONAL INTERVIEW
X-L PB100: MONTH OF THE PERSONAL INTERVIEW
X-L PB110: YEAR OF THE PERSONAL INTERVIEW
X-L PB120: MINUTES TO COMPLETE THE PERSONAL QUESTIONNAIRE
X-L PB130: MONTH OF BIRTH
X-L PB140: YEAR OF BIRTH
X-L PB150: SEX
X-L PB160: FATHER ID
X-L PB170: MOTHER ID

X-L PB180: Spouse/partner ID
X-L PB190: Marital status
X-L PB200: Consensual Union
X PB210: Country of birth
X PB220A: Citizenship 1
X PB220B: Citizenship 2
X PE010: Current education activity
X PE020: ISCED level currently attended
X PE030: Year when highest level of education was attained
X-L PE040: Highest ISCED level attained
X-L PH010: General health
X-L PH020: Suffer from any a chronic (long-standing) illness or condition
X-L PH030: Limitation in activities because of health problems
X PH040: Unmet need for medical examination or treatment
X PH050: Main reason for unmet need for medical examination or treatment
X PH060: Unmet need for dental examination or treatment
X PH070: Main reason for unmet need for dental examination or treatment
X PL015: Person has ever worked
X-L PL020: Actively looking for a job
X-L PL025: Available for work
X-L PL030: Self-defined current economic status
X PL035: Worked at least 1 hour during the previous week
X-L PL040: Status in employment
X-L PL050: Occupation (ISCO-88 (COM))
X-L PL060: Number of hours usually worked per week in main job
X PL070: Number of months spent at full-time work
X PL072: Number of months spent at part-time work
X PL080: Number of months spent in unemployment
X PL085: Number of months spent in retirement
X PL087: Number of months spent studying
X PL090: Number of months spent in inactivity
X PL100: Total number of hours usually worked in second, third… jobs
X PL110: NACE
X PL120: Reason for working less than 30 hours
X PL130: Number of persons working at the local unit
X-L PL140: Type of contract
X PL150: Managerial position
X-L PL160: Change of job since last year
X-L PL170: Reason for change
X-L PL180: Most recent change in the individual’s activity status
X-L PL190: When began first regular job
X-L PL200: Number of years spent in paid work
X-L PL210A: Main activity on January
X-L PL210B: Main activity on February
X-L PL210C: Main activity on March
X-L PL210D: Main activity on April
X-L PL210E: Main activity on May
X-L PL210F: Main activity on June
X-L PL210G: Main activity on July
X-L PL210H: Main activity on August
X-L PL210I: Main activity on September
X-L PL210J: Main activity on October
X-L PL210K: Main activity on November
X-L PL210L: Main activity on December
X-L PY010G/PY010N: Employee cash or near cash income

X-L PY020G/PY020N: NON-CASH EMPLOYEE INCOME
X-L PY030G: EMPLOYER'S SOCIAL INSURANCE CONTRIBUTION
X-L PY035G/PY035N: CONTRIBUTIONS TO INDIVIDUAL PRIVATE PENSION PLANS
X-L PY050G/PY050N: CASH BENEFITS OR LOSSES FROM SELF-EMPLOYMENT
X-L PY070G/PY070N: VALUE OF GOODS PRODUCED BY OWN-CONSUMPTION
X-L PY080G/PY080N: PENSION FROM INDIVIDUAL PRIVATE PLANS
X-L PY090G/PY090N: UNEMPLOYMENT BENEFITS
X-L PY100G/PY100N: OLD-AGE BENEFITS
X-L PY110G/PY110N: SURVIVOR' BENEFITS
X-L PY120G/PY120N: SICKNESS BENEFITS
X-L PY130G/PY130N: DISABILITY BENEFITS
X-L PY140G/PY140N: EDUCATION-RELATED ALLOWANCES
X PY200G: GROSS MONTHLY EARNINGS FOR EMPLOYEES

7.2. Derived variables

In the UDB, it is intended to add a few derived variables that would ease the statistical exploitation of the data base. The list below is non exhaustive.

P -AGE (in year)
P -ADULT/CHILD indicator
P -MOST FREQUENT ACTIVITY STATUS (Employed, Unemployed, Retired)

H -HOUSEHOLD TYPE (1 person HH no dependent child, 2 p HH no DCH, 1 p HH with DCH, 2 p HH with 1DCH, 2 p HH with 2 DCH, 2 p HH with >=3 DCH, other HH with DCH)
H -WORK INTENSITY OF THE HOUSEHOLD
H -EQUIVALISED HH SIZE
H -EQUIVALISED DISPOSABLE INCOME
H -TOTAL HOUSEHOLD INCOME by component (employee, self income, social benefit, …)

Contact:

BERNARD, Telephone:(352) 43 01-37330, bruno.bernard@cec.eu.int