EUROPEAN COMMISSION
EUROSTAT

Directorate F: Social Statistics and Information Society
Unit F-3: Living conditions and social protection statistics

Luxembourg,
EU-SILC/BB D(2005)

BERNARD
Data management

EU-SILC USER DATABASE DESCRIPTION

Version 2007-2 from 01-08-09

Table of content

1. INTRODUCTION

EU-SILC (Community Statistics on Income and Living Conditions) is an instrument aiming at collecting timely and comparable cross sectional and longitudinal multidimensional micro data on income poverty and social exclusion. This instrument is anchored in the European Statistical System (ESS).

EU-SILC was launched in 2004 in 13 MS (all except NL, DE, UK and the 10 new MS except EE) + NO and IS. This first release of the cross sectional data refers mainly to income reference year 2003 and fieldwork operation in 2004 operation. EU-SILC will reach its full scale extension with the 25 MS + NO, IS in 2005. Later it will be completed by TR, RO, BG and CH.

The instrument aims to provide two types of 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, typically, a four years period.

The launching of EU-SILC foresees a transition period till 2007 during which NSI can adapt their tool to common standard, for instance, imputed rent, employer social contribution, income component at gross level. The full implementation of EU-SILC will thus be completed in 2007. The first 4 years individual trajectories will be available by July 2009.

The current information was designed to be general and to be reusable for subsequent releases. It defines the framework that has defined the birth of this instrument. This framework allowed for flexibility and different implementations. Information on current status of the implementation in the different MS is provided as addendum to the different sections and complements the general presentation. The transitional measures valid till 2007 are also underlined whenever relevant.

2. LEGAL BASIS

The introduction of a legal act for EU-SILC was decided by the Directors of social statistics in June 2000. A draft Framework Regulation, prepared by a Task Force together with Eurostat, approved by the Commission in December 2001, was adopted by the European Parliament (EP) in first reading with some minor amendments in May 2002. The common position was approved by the Council at unanimity in March 2003 and the EP approved it in second reading in May 2003. The Framework Regulation was signed by the Council and EP on 16 June 2003 and published in the Official Journal on 3 July 2003.

In parallel, Eurostat and the MS developed the technical aspects of the instrument. More concretely, 5 Commission Regulations ('Sampling and tracing rules', 'Definitions', 'List of primary target variables', 'Fieldwork aspects and imputation

procedures', and 'Quality reports') implementing the Framework Regulation were elaborated. The first four Commission Regulations were approved by the Statistical Programme Committee (SPC) in written procedure in August 2003 and published in the OJ on 17 November 2003. The CR on quality reports was published in OJ on 9 January 2004.

The starting date for the EU-SILC instrument under the Framework Regulation of the EP and of the Council is 2004 for the 12 MS, Estonia, Norway and Iceland, with a derogation for Germany, Netherlands, the UK and 10 New countries with the exception of Estonia to start in 2005 under the condition that they supply comparable data for the year 2004 for the cross-sectional common EU indicators that have been adopted by the Council before 1 January 2003, in the context of the open method of co-ordination.

3.   REFERENCE POPULATION

The reference population of EU-SILC is all private households and their current members residing in the territory of the MS at the time of data collection. Persons living in collective households and in institutions are generally excluded from the target population.

Small parts of the national territory amounting to no more than 2% of the national population and the national territories listed below may be excluded from EU-SILC.

National territories that may be excluded from EU-SILC

Country	Territories
France	French Overseas Departments and territories
Netherlands	The West Frisian Islands with the exception of Texel
Ireland	All offshore islands with the exception of Achill, Bull, Cruit, Gorumna, Inishnee, Lettermore, Lettermullan and Valentia
United kingdom	Scotland north of the Caledonian Canal, the Scilly Islands

4.   SAMPLING

4.1.   LEGAL ASPECTS

According to the Commission Regulation on sampling and tracing rules (N°1982/2003 of 21 October 2003), the sample selection has to fulfil the following requirements:

- *For all components of EU-SILC (whether survey or register based), the cross-sectional and longitudinal (initial sample) data shall be based on a nationally representative probability sample of the population residing in private households within the country, irrespective of language, nationality or legal residence status. All private households and all persons aged 16 and over within the household are eligible for the operation.*

- *Representative probability samples shall be achieved both for households, which form the basic units of sampling, data collection and data analysis, and for individual persons in the target population.*

- *The sampling frame and methods of sample selection shall ensure that every individual and household in the target population is assigned a known and non-zero probability of selection.*

Besides, the EU-SILC Framework Regulation (N°1177/2003 of 16 June 2003) sets out minimum effective sample sizes which shall be achieved by the countries both for the cross-sectional and the longitudinal components.

The cross-sectional sample sizes were calculated in order to achieve an effective size of 121.000 households at the European level (127.000 including Iceland and Norway). Then, the allocation among the countries aims to ensure a minimum precision for each of them.

The longitudinal sample sizes refer, 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.

4.2. IMPLEMENTATION

The EU-SILC Regulations let much flexibility to the countries regarding the sample design.

At the first wave, two ways of sample selection are generally used:

- A sample of households is drawn and all the current members are eligible to be surveyed.

- A sample of persons ("selected respondents") is first drawn and their corresponding households are surveyed. Income information will be collected from registers whereas non-income data will be collected only on the "selected respondents".

From second wave on, some countries have chosen to collect longitudinal information on a pure panel. As for the cross-sectional component, a new sample will be drawn and surveyed every year. Other countries have chosen an "integrated" approach (see next) where both cross-sectional and longitudinal dimension are dealt with through a single design.

4.3. THE "INTEGRATED" DESIGN

This design is recommended by Eurostat and consists in selecting a fixed number of panels at the first wave (Eurostat recommends four panels). Each subsequent year, a panel is dropped and replaced by a new replication.

Figure: Illustration of the "integrated" design (case of four panels)

SUCCESSIVE PANELS OF LIMITED DURATION
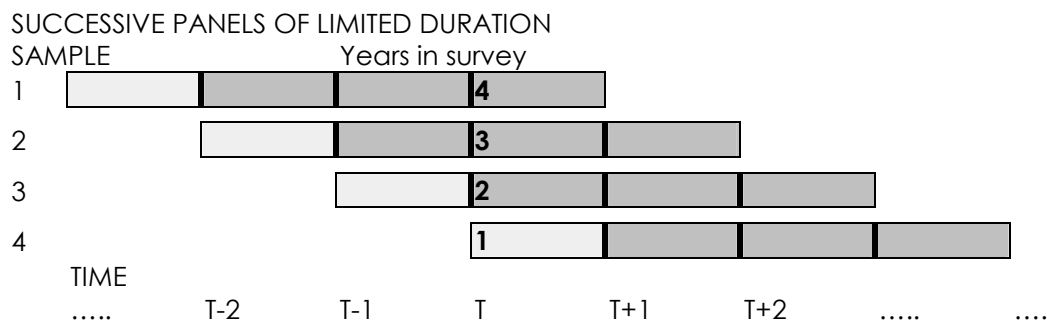
Figure 1 illustrates a simple rotational design (once the system is fully established). The sample for any one year consists of 4 replications, which have been in the survey for 1-4 years (as shown for 'Time = T' in the figure). Any particular replication remains in the survey for 4 years; each year one of the 4 replications from the previous year is dropped and a new one added. Between year T and T+1 the sample overlap is 75%; the overlap between year T and year T+2 is 50%; and it is reduced to 25% from year T to year T+3, and to zero for longer intervals.

This structure is suitable for meeting both the cross-sectional and longitudinal dimensions.

Indeed, it enables to follow up persons over 2, 3 or 4 consecutive years. As for the cross-sectional component, all the households which have at least one current member in a panel are selected and eligible for cross-sectional surveying. It is important to point out the cross-sectional samples designed so are clearly representative of the whole target population because of the renewal of one panel every year.

4.4. THE SAMPLING DESIGNS IMPLEMENTED BY THE COUNTRIES

(See annex 3)

Most of them rely on an "integrated" model with four rotational groups, as Eurostat recommends. But, some other ones chose alternative approaches.

Table: Sampling design implemented by each country

Country	Type of units drawn	Sampling design
Austria (AT)	Households	Integrated design with 4 groups
Belgium (BE)	Households	Integrated design with 4 groups
Cyprus (CY)	Households	Integrated design with 4 groups
Czech Republic (CZ)	Households	Integrated design with 4 groups
Germany (DE)	Households	Integrated design with 4 groups
Denmark (DK)	Persons	Integrated design with 4 groups
Estonia (EE)	Households	Integrated design with 4 groups
Greece (EL)	Households	Integrated design with 4 groups
Spain (ES)	Households	Integrated design with 4 groups
Finland (FI)	Persons	Integrated design with 4 groups (*)
France (FR)	Households	Integrated design with 9 groups
Hungary (HU)	Households	Integrated design with 4 groups
Ireland (IE)	Households	Integrated design with 4 groups
Italy (IT)	Households	Integrated design with 4 groups
Lithuania (LT)	Households	Integrated design with 4 groups
Luxembourg (LU)	"Social Security Households"1 + Households	Longitudinal dimension: pure panel Cross-sectional dimension: pure panel + additional sample selected every year
Latvia (LV)	Households	Integrated design with 4 groups
Netherland (NL)	Households	Integrated design with 4 groups
Poland (PL)	Households	Integrated design with 4 groups
Portugal (PT)	Households	Integrated design with 4 groups
Sweden (SE)	Persons	Longitudinal dimension: rotating four-year panel Cross-sectional dimension: a new survey every year
Slovenia (SI)	Persons	Integrated design with 4 groups
Slovakia (SK)	Households	Integrated design with 4 groups

---

[1] It refers to a sub-group of members of a household who depends on the same Social Security system.

United Kingdom (UK)	Households	Integrated design with 4 groups
Iceland (IS)	Persons	Integrated design with 4 groups
Norway (NO)	Persons	Integrated design with 8 groups

(*) Finland has included SILC in their own "Income distribution survey (IDS)" which is a 2 years rotational panel. To achieve this fusion, a part of the sample from IDS is followed 4 years instead of 2. The difference with the integrated design is that person from wave 3 and 4 are not longer part of the cross-sectional sample (see figure below).

The following figure presents the relations between the longitudinal Income Distribution Survey (IDS) (areas with bold lines) equal to the cross-sectional sample and the wave structure of SILC (shaded). The assumptions are 76 % response for the first wave and 92 % response for other waves.

	2004	2005	2006	2007	2008	2009	2010
	1. year	2. year	3. year	4. year	5. year	6. year	7. year

Gross sample							
3 200							
2 500							
2 500	1 900						
2 500	1 900	1 748					
2 500	1 900	1 748	1 608				
	5 000	3 800					
	2 500	1 900	1 748	1 608			
		5 000	3 800				
		2 500	1 900	1 748	1 608		
			5 000	3 800			
			2 500	1 900	1 748	1 608	
				5 000	3 800		
				2 500	1 900	1 748	
					5 000	3 800	
					2 500	1 900	
						5 000	
						2 500	

5. THE SURVEYS

5.1. SURVEY UNITS

In terms of the units involved, four types of data are gathered in EU-SILC:

(a)   variables measured at the household level;

(b)   information on household size and composition and basic characteristics of household members;

(c)   income and other more complex variables termed 'basic variables' (education, basic labour information and second job) measured at the personal level, but normally aggregated to construct household-level variables; and

(d)   variables collected and analysed at the person-level 'the detailed variables' (health, access to health care, detailed labour information, activity history and calendar of activities).

For set (a)-(b) variables, a sample of households including all household members is required.

Among these, sets (a) and (b) will normally be collected from a single, appropriately designated respondent in each sample household – using a household questionnaire for set (a) and a household member roster for set (b). Alternatively, some or all of these may be compiled from registers or other administrative sources.

Set (c) concerning mainly but not exclusively the detailed collection of household and personal income – must be collected directly at the person level, covering all persons in each sample household. In most countries, i.e. in the so-called 'survey countries', these income variables will be collected through personal interviews with all adults aged 16+ in each sample household. This collection will be normally combined with that for set (d) detailed variables, since the latter also must also be collected directly at the person level.

By contrast, in 'register countries', set (c) variables will be compiled from registers and other administrative sources, thus avoiding the need to interview all members (adults aged 16+) in each sample household.

Set (d) variables will normally be collected through direct personal interview in all countries. These are too complex or personal in nature to be collected by proxy; nor are they available from registers or other administrative sources. For the 'survey countries', this collection will normally be combined with that for set (c) variables as noted above – consequently both normally based on a sample of complete households, i.e. covering all persons aged 16+ in each sample household.

However, from the substantive requirements of EU-SILC, it is *not essential* that – in contrast to set (c) variables – set (d) variables be collected for all persons in each sample household. It is possible to do this collection on a representative sample of persons (adult members aged 16+), such as by selecting one such person per sample household. It is expected that this option will normally be followed in 'register countries', since for these countries interviewing all household members for set (c) is not involved. In countries which choose to do so, the sampling process involved will be the selection of *persons* (usually one adult member aged 16+ per household) directly or through a sample of households. The selected individuals may be termed **'selected respondents'**. Randomised selection procedures must be used to

ensure that a representative sample of persons is obtained from the representative sample of households.

Table 1 summarises the type of survey units for sampling, analysis and data collection involved in EU-SILC. The ultimate units used in the sample selection may be addresses, households or persons, each unit selected with a known probability. From these, it is always necessary to construct a sample of households, the probability of each household in the sample being determined through its association (or identity, as the case may be) with units in the sample selected. The analysis units can be households, all members, adult members, or possibly a sub sample of adult members; these are the units to which the information collected pertains. Their probabilities of selection (or the corresponding sample weights) are determined through their association with the sample household. The collection unit refers to the person or source providing the information.

Table 1 Survey units for sampling, analysis, and data collection

Sampling unit		Analysis units	Collection unit/source	
Selected	constructed		'survey country'	'register country'
Address Or Household Or *Person* (aged 16+)	*Household*	Set (a): household	Household respondent (HR)	Registers +HR
		Set (b): all household members	Household respondent*	Registers +HR
		Set (c): household and personal income and basic variables	Personal interview (all members 16+)	Registers (all members 16+)
		Set (d): detailed variables		
		All members 16+	Personal interview**	
		Selected respondent		Personal interview

* combined with set (a) household interview

** combined with set (c) personal interview

In each country, EU-SILC involves the provision of cross-sectional and longitudinal data, both for 'income and basic variables' (I) and 'detailed variables'(S). Combining these dimensions gives four basic data components in EU-SILC:

(**CI**)   Cross-sectional income component (included basic variables).

(**CS**)   Cross-sectional detailed component.

(**LI**)   Longitudinal income component (included basic variables).

(**LS**)   Longitudinal detailed component.

Substantive requirements of EU-SILC impose certain conditions on the samples for these components. The basic (essential, minimum) condition which must be satisfied by any data structure in EU-SILC can be expressed as:

$$\boxed{\begin{aligned}(a) \quad & CS \subseteq CI \\ (b) \quad & LS \subseteq LI\end{aligned}}$$ *… the basic condition of EU-SILC data structure.*

The basic condition means that the detailed data must be collected on the same sample as the income data, or on a sub sample of the latter. The condition applies separately to both the cross-sectional and longitudinal components.

5.2. MODES OF COLLECTION

EU-SILC data are collected from different sources and by different modes.

They may have been:

- Constructed

- Deducted from sample frame

- Deducted from sample design

- Settled by interviewers

- Collected from household respondent

- Collected from household members

- Collected from a proxy

Some recommendations have to be applied on the following data and mode of collection:

5.2.1.   *Household respondent and household level information*

The <u>household respondent</u> is the person from whom household level information is obtained. Given that the household-level response is going to be attributed to all household members, it is essential that

the information be collected from someone who can, in some sense, 'speak for' the household.

For instance, if the 'selected respondent' is the 16-year old son or daughter, this person is highly unlikely to be able to provide good quality information on such issues as the mortgage or rent payments, housing costs, income from family and other benefits.

The household respondent will be chosen according to the following priorities:

Priority (1): the person responsible for the accommodation.

Priority (2): a household member aged 16 and over who is the best placed to give the information.

For the second and following waves of the longitudinal component of EU-SILC, the household respondent will be chosen according to the following list of priority:

Priority (1): the household respondent in the last wave.

Priority (2): a 'sample person' aged 16 and over giving priority to the person responsible for the accommodation or the best placed to give the information.

Priority (3): a 'non-sample person' aged 16 and over.

5.2.2. Household members and personal level information

For the information on all of household members 16 and over the following modes of data collection are stipulated:

- For basic data, education and labour information: personal interview, proxy on a normal procedure or registers

- For income variables: personal interview (proxy as an exception for persons temporarily away or in incapacity) or extraction from registers

For the information on at least a household member 16 and over (the selected respondent), i.e., health and detailed labour information, only personal interview, proxy as an exception (for person temporarily away or incapacity), or extraction from register will be permitted.

When by special circumstances (absence, illness, incapacity, ...) the individual may not directly provide the information that is requested, through personal interview, it will be chosen:

- to make a personal interview with another member of the household trained to facilitate the data (proxy)

- to make a telephone interview with the individual (CATI or telephone)

- to leave the questionnaire in the household to be self-administered by respondent.

If the information is carried out through personal interview with another member of the household or is self-administered by respondent, the interviewer, if possible, should try to arrange a later interview with that person or, if it is not possible, to contact him/her by phone in order to check the information provided in the questionnaire.

In the case that a proxy interview is carried out, the identification number of the person who has provided the information has to be recorded.

5.3. SURVEY DURATION AND TIME

The following rules about survey duration and time are laid down in the CR on fieldwork aspects and imputation procedures.

(1)   The interval between the end of the income reference period and the time of the interview for the respondent concerned shall be limited to 8 months as far as possible. This applies both to the household and personal samples, and irrespective of whether the reference period used is fixed in terms of calendar dates for the whole sample or is a moving reference period determined according to the timing of the interview for the household or person concerned.

(2)   By way of exception to paragraph 1, if the income variables are collected from registers the interval between the end of the income reference period and the time of interview for current variables shall be limited to 12 months.

(3)   Where all the data are collected through field interviewing and a fixed income reference period is used, the total duration of the data collection of the sample shall be limited to 4 months as far as possible.

(4)   Where the data are collected through field interviewing using a moving income reference period and the fieldwork duration exceeds 3 months, the total annual sample shall be shared approximately equally between the fieldwork months. In this case, the total fieldwork duration for the cross-sectional component and each wave of the longitudinal component shall not exceed 12 months.

(5)   For the longitudinal component, the collection or compilation of data, for a given unit (household or person), between successive waves shall be kept as close as possible to 12 months.

5.4.   SURVEY CHARACTERISTICS BY COUNTRY

Notes:

- Fieldwork duration: it is expressed in quarters. More than 97% of the data collection is done in the specified quarters (see variable HB050 for more details).

- Income reference period: For Ireland, as the "income reference period" is "12 month prior the date of interview", the end of income reference period is the date of the interview.

- Personal level data collection: "Both interview and register" means that for the same respondent, information have been collected partly by interview and partly from register. For Austria, in some few cases, data is only taken from registers.

Country	Income reference period	Personal level data collection
Austria	01/01/ to 31/12/ previous to fieldwork	Interview (mostly)
Belgium	01/01/ to 31/12/ previous to fieldwork	Interview
Cyprus	01/01/ to 31/12/ previous to fieldwork	Interview
Czech Republic	01/01/ to 31/12/ previous to fieldwork	Interview
Germany	01/01/ to 31/12/ previous to fieldwork	Interview
Denmark	01/01/ to 31/12/ previous to fieldwork	Both interview and register
Estonia	01/01/ to 31/12/ previous to fieldwork	Interview
Greece	01/01/ to 31/12/ previous to fieldwork /	Interview
Spain	01/01/ to 31/12/ previous to fieldwork	Interview
Finland	01/01/ to 31/12/ previous to fieldwork	Both interview and register
France	01/01/ to 31/12/ previous to fieldwork	Interview
Hungary	01/01/ to 31/12/ previous to fieldwork	Interview
Ireland	12 month prior the date of interview	Both interview and register

Italy	01/01/ to 31/12/ previous to fieldwork	Both interview and register
Lithuania	01/01/ to 31/12/ previous to fieldwork	Interview
Luxembourg	01/01/ to 31/12/ previous to fieldwork	Interview
Latvia	01/01/ to 31/12/ previous to fieldwork	Interview
Netherlands	01/01/ to 31/12/ previous to fieldwork	Both interview and register
Poland	01/01/ to 31/12/ previous to fieldwork	Interview
Portugal	01/01/ to 31/12/ previous to fieldwork	Interview
Sweden	01/01/ to 31/12/ previous to fieldwork	Both interview and register
Slovenia	01/01/ to 31/12/ previous to fieldwork	Both interview and register
United Kingdom	01/01/ to 31/12 from year of fieldwork	Interview
Iceland	01/01/ to 31/12/ previous to fieldwork	Both interview and register
Norway	01/01/ to 31/12/ previous to fieldwork	Both interview and register

5.5. TRACING RULES

The main objective of the longitudinal component of EU-SILC is to study changes over time at individual level, such as transitions from school to work and from work to retirement, flows into and out of economic activity and work and, above all, changes in the level of income and poverty of individuals and households.

As a consequence, it is necessary in the longitudinal component of EU-SILC to trace individuals on a minimum of a four-year period.

Longitudinal surveys require a set of procedures that indicate who is traced and interviewed through time.

5.5.1. Target population

In each wave, the longitudinal component of EU-SILC should ideally represent the current target population, i.e. the population of all persons living in private households within the national territory of the country concerned. Excluded from the target population are persons living in collective households and in institutions. For practical reasons, small parts of the national territory (the excluded areas) may also not be covered in the survey.

In practice the target population which can be covered will differ in certain respects from the above as a result of the manner in which the longitudinal sample is constructed. The longitudinal component of EU-SILC will comprise one or more panels. Each panel will begin with the selection of an initial sample representing the target population at the time of its selection, in the same way as the cross-sectional survey. This initial sample is then followed-up over time (for a minimum of duration of 4 years; or the duration may be longer or indefinite depending upon the design adopted in the country), according to specified tracing rules defined below. The objective of the tracing rules is to reflect in the initial sample any changes in the target population and to follow-up individuals over time. The sample for the EU-SILC longitudinal component at any given time (year) will in general consist of (i) follow-up of the initial sample(s) selected at earlier times, plus (ii) any new 'initial sample' selected at the time concerned. The latter covers 'rotational designs', as well as any supplements which may be added to the sample from time to time to compensate for panel attrition.

Thus, depending on the tracing rules, the longitudinal sample at any given time may not exactly represent the current 'cross-sectional' target population. The type of demographic changes which need to be reflected include births to individuals in the original population, movements of persons from outside the original population (from collective households, institutions or abroad) into private households containing individuals from that population, and into new private households not containing such individuals. With the possible exception of sample supplements added specially for the purpose, the last mentioned category of in-migrants is generally not covered by the panel tracing procedures. Deducted from the population are individuals who have died, moved out of scope (abroad or outside the private household sector), or become ineligible for other reasons.

5.5.2. Initial sample and sample persons

As already mentioned in 'Survey units', the information collected in EU-SILC pertains to the following types of units. This applies to both the cross-sectional and longitudinal components.

(1)   Households, for the collection of household level variables

(2)   All household members, for the collection of demographic and other basic information on household members, including on household size and composition

(3)   All household members aged 16+, for the collection of income and basic information

(4)   Selected respondents, which may include all members aged 16+ or a random sub sample thereof (usually one such person per household), for the collection of detailed information, and

(5)   Former household members (for the longitudinal component only), on whom some elementary information on activity status and time spent in the household during the income reference period may be collected.

The information for Set 4 concerns 'detailed' variables which must be collected through a personal interview in all countries, irrespective of whether or not registers are used for other purposes. EU-SILC permits two types of samples for this purpose:

(1)   An initial sample of 'complete' households, i.e. covering all persons in each household. Among these only persons aged 16+ at the time are eligible for the detailed personal interview.

(2)   A random sample of persons. Again, only persons aged 16+ at the time are eligible for the detailed personal interview.

Both these schemes are meant to represent the entire target population of persons (and hence also all private households) at the time of sample selection. They differ only in the type of sample selected from that population.

Set 4 defines samples for the other sets. These consist of all households containing at least one Set 4 person (Set 1), all current members of these households (Set 2). Among current members, only persons aged 16+ at the end of the income reference period are eligible for the collection of income and related information under Set 3.

Individuals selected for the purpose of Set 4 are termed sample persons. These are all or a subset of persons in the initial sample which are followed up over the duration of the panel to obtain the longitudinal sample of observations. Thus, in principle, all members of households in the initial sample of 'complete' households are sample persons. For an initial sample of persons, the term applies only to the individuals selected (normally one per sample household). Other individuals in sample households are termed co-residents. A sample household is defined as a household containing at least one sample person.

For those countries where a sample of complete households is selected, exactly the same information (all Sets 1-4) is required from sample persons and from co-residents. For countries using a random sample of persons (normally one person per household), Sets 1-3 apply to sample persons as well as co-residents in the households,

while the personal interview (Set 4) applies only to the sample persons.

5.5.3. *Follow-up of sample persons*

To study changes over time at individual level, it is necessary that all sample persons are followed-up over time, despite the fact that they may move to a new location during the life of the panel. However, in the implementation of EU-SILC some restrictions will be applied for practical reasons, as explained below.

5.5.3.1. Movement

Ideally, all sample persons once selected should be followed up to whatever new place they move to. However, for cost and other practical reasons, it has been decided that in EU-SILC persons moving only within the confines of the target population as defined above will be followed-up: in other words, *person remaining or moving within private households in the national territory covered in the survey*. Sample persons moving to a collective household or to an institution, moving to national territories not covered in the survey, or moving abroad (to a private household, collective household or institution, within or outside the EU), would normally not be traced. The only exception would be the continued tracing of those moving temporarily (for actual or intended duration of less than 6 months) to a collective household or institution within the national territory covered, who are still considered a member of the household.

5.5.3.2. Age range

The longitudinal sample must also remain representative of all age groups in the population. This means than in principle, persons of all ages should be followed up. However, in view of cost and other practical considerations, separate follow-up may be restricted to persons above a certain age. The appropriate choice of the age cut-off will depend on the type of EU-SILC design adopted by the country.

The minimum EU-SILC requirements are for a follow-up of individuals in the longitudinal sample for a period of four years. For panels of such short duration, it is acceptable (in view of cost and other practical reasons) *to separately follow-up only persons aged 14 or over at the time of selection of the initial sample for a panel*.

The practical effect of this limitation is that children aged under 14 in the initial sample will not be covered in the longitudinal sample – but only if they move 'independently' to a new household containing no member aged 14+ from their original household. Also, since households in the longitudinal sample include all private households

containing at least one sample person, when the follow-up is confined to sample persons above a certain age (such as 14+), the resulting sample will fall short of the ideal by excluding households which contain only sample person(s) below that age limit (and no older sample persons).

In addition, to reflect demographic changes in the population accurately, it is also necessary that provision is made to include new-born children into the sample. This can be achieved by including children born to sample women also as sample persons and following them up using the normal procedures. For short panels of 4-year duration, it has been decided not to follow-up new born children. This results in under-coverage of babies who move to households containing no person aged 14+ from their original household – a circumstance which should be rare in EU countries.

The implication of these restrictions on the follow-up of children is that longitudinal (persistent) poverty among them cannot be estimated exactly. However, as noted, the approximation will be confined to children moving into new households not containing any person aged 14+ from their original household.

Hence it is not sufficient to confine the selection to persons aged 16+ within each household for the purpose of follow-up. *For a 4-year panel, the selection should at least cover persons aged 14+.*

Longer the duration of the panel, more necessary it would be to lower the age limit above which all sample persons will need to be followed. It is recommended that if the panel duration exceeds say 8 years, the follow-up covers persons of all ages, including children born to sample women during the course of the panel survey.

As noted earlier, two types of sample designs are possible under EU-SILC for the detailed personal interview survey: a sample of 'complete' households, in which all persons aged 16+ are eligible for the detailed personal interview; or a sample of persons, in which normally one person aged 16+ is selected per sample household for the purpose.

It is important to emphasise that in the design employing a sample of persons, the inclusion of persons aged under 16 is a more critical requirement than that in a sample of complete households. This is because in a sample of persons, those aged under 16 can enter the interview sample on achievement of age 16 only if they were already selected into the sample for this purpose

(1)   Persons aged 14-15 at the time of selection will not be interviewed in detail till they reach the age of 16 - but must be followed-up (traced) even though no detailed personal interviews at all is involved in the household. Household and income information should nevertheless be collected for such households, normally using registers.

(2)   If panels of duration longer than four years are employed, the age limit for the selection of individuals would need to lowered further.

(3)   The size of the sample selected would need to be appropriately increased to achieve the required number of interviews with persons aged 16+.

5.5.3.3.   Non-respondents

A household which refuses interview may be dropped from the sample. Any sample persons are automatically dropped from further follow-up.

For a short panel of 4 years duration, a household which has not been enumerated for two consecutive years or non-contacted the first year of the panel (due to the impossibility of accessing address, because the whole household is temporarily away or is unable to respond due to incapacity or illness) may be dropped, along with any sample person in it. Non-enumerated a single year due to the impossibility of locating the address, the address being non-residential or unoccupied, lost (no information on what happened to the household) may be dropped.

In countries using panels of longer duration, more thorough follow-up procedures are recommended because of the greater danger of panel attrition. As a general recommendation only household after two *consecutive* non-interviews may be dropped.

5.5.4.   *Precise tracing rules*

Based on the above, the EU-SILC tracing rules are summarised below:

(1)   Children aged under 14 will not be traced if they move to a new household containing no sample person aged 14+. In this sense, they are not considered 'sample persons'. Sample persons aged 14+ will be normally traced.

(2)   Detailed interview will be conducted with persons aged 16+, and in the case a sample of complete households is used, with all persons 16+ (whether sample persons or co residents) in the household.

(3)   Sample persons aged 14+ who have moved to another private household within the country are traced to the new location. Those aged 16+ are interviewed.

(4)   Strictly, the reference here is to territory of the country included in the target population. Those moving to certain small and specified excluded areas are dropped from the survey, as are persons moving out of the country.

(5)   Sample persons aged 14+ temporarily in a collective household or institution but still considered as members of a private household are traced and, if aged 16+, are to be interviewed by proxy.

(6)   However, sample persons aged 14+ who have moved to a collective household, are institutionalised or moved abroad on a permanent or indefinite basis, for actual or intended duration of 6 months or more, or for a short stay but who cannot be considered a member of any private household, are dropped from the survey. (In that case, the following information should be recorded from someone who was a member of the person's household at the previous wave: to where did the person move, date of movement, number of months spent in the household during the income reference period and main activity status during the income reference period).

(7)   For sample persons who died, no information other than date of death, number of months spent in the household during the income reference period and main activity status during the income reference period will be collected.

(8)   Sample persons aged 14 + who have not been contacted in the previous wave because of the impossibility to access the address (for atmospheric reasons) or because the whole household was temporary absent or unable to respond (illness, incapacitated,…) or for other reasons, a new contact will be attempted in the present wave. (If the sample has not been contacted the first year or two consecutives years due to the reasons mentioned above, the sample person may be dropped from the survey).

(9)   Also will be dropped from the survey sample persons aged 14 + who have not been contacted because of the impossibility to locate the address or because the address was non-residential or unoccupied, lost (no information on what happened to the survey) and who or whose household refused to co-operate

(10) Co-residents are included in EU-SILC as long as they continue to live with a sample person. Personal information is required, using normal procedures, if aged 16+. However, co-residents are not traced if they move to a household not containing a sample person (aged 14+).

(11) For former residents (''the former household members''), who spent at least 3 months in the household during the income reference period, the following information will be

required(only initial households): number of months spent in the household during the income reference period and main activity status during the income reference period.

(12) The age cut-off of 14 years will be lowered if panels of duration longer than four years are used. Persons of all ages in the initial sample (including children born to sample women) should be treated as sample persons to be followed-up in panels of duration exceeding 8 years.

The age refers to the age that person is in the first wave of each panel.

The following table summarize the follow-up of sample persons, sample households and co-residents:

Table 1.Rules for the follow-up of sample persons, sample households and co-residents

Sample persons	To be
Moving to a private household within the national territory covered in the survey	Followed to the new location of the household
Other persons temporarily away but who are still considered as members of the household	Covered in the household they belong to
Persons no longer members of a private household, or those who have moved outside the national territory covered in the survey	Dropped from the survey
Sample households	**To be**
Non enumerated a single year due to the impossibility to locate the address, the address being non-residential or unoccupied, lost(no information on what happened to the household), or the household refusing to co-operate	Dropped (can be kept for more than 4 years longitudinal design)
Non contacted the first year of the panel or non contacted two consecutive years due to the impossibility to access the address, because the whole household is temporarily away or is unable to respond due to incapacity or illness	Dropped (can be kept for more than 4 years longitudinal design)
Co-residents	**To be**

Living in a household containing at least one sample person	Followed
Living in a household not containing any sample person	Dropped

5.5.5. Organisation of the tracing

For countries where a sample of households/address was selected, the tracing will be done from the address that exist in the previous wave.

As the main risk of attrition in a panel survey is linked to the movers, measures to avoid this risk have to be taken by the NDUs to collect the maximum possible information when a sample person is moving. The NDUs have to establish special procedures to trace all moving/split-off households.

Most importantly, every effort is to be made to trace moving people before the interviewers visit. Several measures can be taken, e.g. (a) asking about intention or expectation of move at the previous interview; (b) contact by mail or phone in the intervening period between the waves; (c) requesting the household to inform of a move (with appropriate financial incentives) etc.

In order to be able to trace moving/split households, the first task of the interviewer , when coming to the address of the household in previous wave, is to get all the information for the identification of the household and on the changes in the household composition. It is important to obtain the date, reason of and the new address of the movers.

If the interviewer is not able to get the new address, then an attempt has to be made by the supervisor and or by the central team. It is recommended that within each NDU, at least one person is concerned only with finding the new addresses of these households in the population, using the postal system/other sources.

Another proposal which may be considered is to use specialised interviewers for follow the movers: they could be paid more, and have a closer relationship with the supervisor.

5.5.6. Information to be collected

In the initial household, the whole information required for current household members, basic information for former household members and also basic information on households members in

previous wave that are no longer household members will be collected.

In the split-off household, only the whole information required for current household members will be collected.

The whole information required for current household members, the basic information for former household members and the basic information on household members in previous waves that are no longer household members are laid down in the Commission Regulation on the list of target primary variables.

Where a sample person is in the survey for more than one year, information will be obtained on whether the person remained at the same address or moved to a different address from one year to the next.

6. WEIGHTS

6.1. LEGAL ASPECTS

According to the Commission Regulation on sampling and tracing rules (N°1982/2003 of 21 October 2003, §7.4):

Weighting factors shall be calculated as required to take into account the units' probability of selection, non-response and, as appropriate, to adjust the sample to external data relating to the distribution of households and persons in the target population, such as by sex, age (five-year age groups), household size and composition and region (NUTS II level), or relating to income data from other national sources where the Member States concerned consider such external data to be sufficiently reliable.

6.2. THEORETICAL ASPECTS

Sample weights $\{\omega_i, i \in s\}$ aim to draw inference from a sample s to the target population. A common simplified view of the estimation problem is to make up a population model based on the sample units replicated according to their weight. Basically, a unit i "represents" ω_i units of the target population.

Figure 2: Population model based on the sample weights

ω_i units i $\quad$ ω_j units j $\quad$ ω_k units k
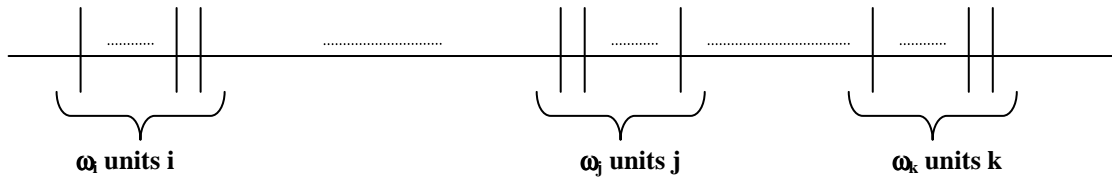
6.3. EU-SILC weights

Cross-sectional weights

- *The household cross-sectional weights (target variable DB090)* will be used to draw inference from the effective sample to the target population of private households. Those weights had to be corrected for household non-response and possibly calibrated to external data source(s).

- *The personal cross-sectional weights for all household members, of all ages (target variable RB050)* will be used to draw inference on individual basic demographic variables for the population of all individuals living in private households.

  Because all the current members of any selected household are surveyed, the personal weights RB050 have to be equal to the corresponding household cross-sectional weight DB090.

- *The personal cross-sectional weights for all household members aged 16 and over (target variable PB040)* will be used to draw inference on the variables included in the personal questionnaire. These weights had to be corrected for individual non-response.

- *The children cross-sectional weights for childcare (target variable RL070)* enable inference on the population of children aged 0 to 12. The individual weights RB050 can be used as children weights. However, it is recommended to adjust them to external sources pertaining to the child population.

- *The personal cross-sectional weights for selected respondents (target variable PB060)* apply to situations where a sample of persons ("selected respondents") is used to collect information about more complex non-income variables, as possibly in countries where income and some other information are obtained from registers (see 1.2.). These weights will be used to draw inference on all the variables defined at the selected respondent level (pertaining to detailed labour information, health status, access to health care…).

Longitudinal weights

- *The personal base weights (target variable RB060)* will apply to all the panel persons regardless the age. They had to be corrected for non-response and calibrated to external data source(s).

- *The personal base weights (target variable PB050)* will apply to all the persons above a certain age limit (normally 16) which replied to the individual questionnaire.

7. IMPUTATION

In EU-SILC missing data problems can arise from diverse sources in a number of forms. The discussion is concerned with the problem of imputation for item non-response, particularly with the problem encountered in constructing total household income in the presence of missing information on some income components. Similar problems arise when the information is available on some but not all the members of a household.

The discussion is confined to cross-sectional context only. Editing and imputation of longitudinal data involves taking into account auxiliary values from current wave, previous and future waves (countries using a rotational or long-term panel will apply a common imputation method for the cross-sectional and longitudinal component).

There are two types of reasons for impute missing data, one may be called statistical and other practical. The statistical reason of imputation is to minimise the mean squared error of survey estimates, in particular the non-response bias component that arises when the pattern of missing data is not random. The practical reasons concern consistency between the results from different analyses (which may handle - and be affected - differently by the problem of missing data), and the convenience of not having to deal with the missing data problem at the analysis stage.

In certain situations, such as when the incidence of item non-response is low and/or when the non-response happens not to be selective, it may be a reasonable option to ignore the problem and confine the analysis only to cases with complete information.

This, however, is not a general option in the case of EU-SILC. This is because total household income is made up of a large number of components. In a large proportion of the cases, information on some but not all components may be available. It is not acceptable to reject a case if the information is incomplete, as that would result in the loss of much valuable information. Hence it was required to impute missing values in the income variables where that can be reasonably done. Furthermore, since the total income of a household is made-up of incomes of its individual members, it is also necessary to take into account the problem of missing individual interviews within otherwise completed households.

7.1. Missing data in EU-SILC

7.1.1. Coverage and sample selection errors

These arise for instance when units in the target population are not represented (whether implicitly or explicitly) in the sampling frame, or when the unit selection probabilities are distorted, or other sample selection errors occur. Generally, such distortions are extremely difficult to correct. Some correction may be possible on the basis of information external to the sampling frame. Attempts to control for such bias are called bench-marking, post-stratification, calibration etc.

7.1.2. Unit non-response

This refers to absence of information on whole units (households and/or persons) selected into the sample. In EU-SILC, the unit non-response has been normally addressed by reweighing the responding cases. Some of the information for weighting comes from within the survey, such as information on units' selection probabilities, and unit non-response rates for different subgroups in the sample. In addition, weighting normally also makes use of external control distributions of population characteristics (e.g. by household size, location, age and sex, activity status) though the use of calibration techniques.

7.1.3. Partial unit non-response.

EU-SILC involves two levels of units of analysis: household and persons. In analysis involving the distribution units at either of these levels, non-response can be dealt with through weighting. However, a special feature of EU-SILC is that a number of variables at the household level are not collected directly with the household as the unit, but are constructed by aggregating information on individual members of the household. An example is the variable 'number of economically active members in the household according with self-defined current activity status', which requires information on the activity status of all household members.

The most important of this class of variables concern components of household income. It can be constructed only if information on income is available for all members of the household. The term 'partial unit non-response' is introduced to describe the situation where some but not all individual members of a household selected for the survey have been successfully enumerated. Two possible approaches of dealing with this problem are:

(1)   Adjustment of weights of enumerated individuals in the household with the objective of compensating for members not enumerated.

(2)   Imputation of the required variables for each non-enumerated person in the household through imputation.

7.1.4.   Item non-response

This refers to the situation when a sample unit has been successfully enumerated, but not all the required information has been obtained.

Various approach of imputation can be envisaged:

 (a) Deductive methods

(b) Deterministic methods

(c) Stochastic methods

Deductive methods refer to imputation procedures in which the true value of a missing item is logically deduced. This means that the value is either deduced from other variables of the survey or is derived from legal regulations. An example for the first mode of deductions is the net-gross-net conversion, when either the gross value or the net value is given and the corresponding missing value is calculated by applying general rules. An example for the latter mode is when the value of the child care benefit is missing and the effectual value can be inserted.

The difference between deterministic and stochastic methods is whether the calculation procedure to calculate the missing item produces a random output as, e.g. simulating the error term using a regression approach.

7.2.   EU-SILC target variables for imputation

According with EU-SILC Framework Regulation :'MS shall transmit to the Commission (Eurostat) in the form of micro data files weighted cross-sectional and longitudinal data which has been checked , edited and **imputed in relation to the income**'.

The Commission Regulation on sampling and tracing rules mentions, in relation to the imputation:

1. Where non-response to income variables at component level results missing data, appropriate methods of statistical imputation shall be applied,

2. Where any gross income variable at component level is collected directly, appropriate methods of statistical imputation and/or modelling shall be applied to obtain the required target variables.

3. When non-response to an individual questionnaire occurs within a sample household, appropriate statistical procedures for weighting, or imputation shall be used to estimate the total income of the household.

Also the Commission Regulation in fieldwork aspects and imputation procedures refers to the imputation as follows:

1. The procedure applied to the data should preserve variation of and correlation between variables. Methods that incorporate 'error components ' into the imputed values shall be preferable to those that simply impute a predicted value.

2. Methods which take into account the correlation structure (or other characteristics of the joint distribution of the variables) shall be preferable to the marginal or univariate approach.

7.2.1.   Desirable characteristics of an imputation procedure

A set of rules is needed as a guide to generate acceptable imputation results. The quality of the results always requires considerable amounts of good judgement during the imputation process, in the identification of patterns, in the selection of the appropriate techniques, choice of auxiliary variables, etc.

Various approaches to the imputation of missing values are possible. It is neither necessary nor possible to insist on any particular methodology in the case of EU-SILC. However, there are clearly some desirable properties which the procedure should have, and some procedures are better than other in terms of those properties.

The procedure should preserve variation of and correlations between variables. Methods that incorporate into the imputed values some 'error component' are preferable to those which simply impute a predicted value. Similarly, methods which take into account the correlation structure (or other characteristics of the joint distribution of the variables) are preferable to the marginal or univariate approach which deals with the imputation of each variable separately. On the other hand, it is also desirable to limit the complexity or the computational work involved in the construction of the imputations. Special techniques such as multiple

The choice of the strategy for imputation is thus dependent of the national context. The EU-SILC UDB contains only information about the presence of imputation for income component (partial or total) which is materialised by the imputation factor in the data base. The imputation factor is detailed later on. At the moment there is a

project to refine the description of imputation procedure by first distinguishing between the different modes of imputation and second collecting systematic meta information on imputation procedure

7.2.2. *Partial unit non-response*

It is necessary to correct for the effect of non-responding individuals within a household in aggregating personal level income variables to construct the corresponding variables at the household level. Otherwise, income of individuals not interviewed is not added up into the total household income.

The same applies to other variables constructed at the household level through aggregation of person-level variables.

In the context of EU-SILC and in case of partial unit non response, the variable HY025, the household income inflation factor, is designed for collecting the multiplicative factor to be applied to the collected household disposable income, HY020, to compensate for partial unit non response at the household level. Different approaches have been recommended by Eurostat to compute HY025. Implementation may vary form country to country. One possible approach is full imputation of missing personal income components to adjust the value of HY025. Adjustment of sample weights can provide an alternative. The ECHP procedure using previous wave could be envisaged for the second wave only if a small proportion (around 3% overall) of the households were affected by the problem.

From 2006 operation onwards, EUROSTAT recommends as much as possible to impute income components of partial unit non response or directly impute household income components instead of using HY025 inflation factor variable

8. THE DATABASE

8.1. DATA AVAILABILITY

Data are not available for all countries depending on the year they started SILC

C: means cross-sectional data

L: means longitudinal data

N/A: no data

COUNTRY	2004	2005	2006	2007
BE	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
CZ	N/A	C	C + L (2 years)	C + L (3 years)
DK	C	C	C + L (2 years)	C + L (3 years)
DE	N/A	C	C + L (2 years)	C + L (3 years)
EE	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
EL / GR	C	C + L (3 years)	C + L (4 years)	C + L (4 years)
ES	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
FR	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
IE	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
IT	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
CY	N/A	C	C + L (2 years)	C + L (3 years)
LV	N/A	C	C + L (2 years)	C + L (3 years)
LT	N/A	C	C + L (2 years)	C + L (3 years)
LU	C	C + L (3 years)	C + L (4 years)	C + L (4 years)
HU	N/A	C	C + L (2 years)	C + L (3 years)
MT	N/A	N/A	N/A	N/A
NL	N/A	C	C + L (2 years)	C + L (3 years)
AT	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
PL	N/A	C	C + L (2 years)	C + L (3 years)
PT	C	C	C + L (2 years)	C + L (3 years)
SI	N/A	C	C + L (2 years)	C + L (3 years)
SK	N/A	C	C + L (2 years)	C + L (3 years)
FI	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
SE	C	C + L (2 years)	C + L (3 years)	C + L (4 years)
UK	N/A	C	C + L (2 years)	C + L (3 years)
IS	N/A	C + L (2 years)	C + L (3 years)	C + L (4 years)
NO	C	C + L (3 years)	C + L (4 years)	C + L (4 years)

8.2. DOMAINS & AREAS

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
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

8.3. THE FILES

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)

Their name have the following structure: UDB_*XYYT*.CSV

With: X = C(cross-sectional) or L(longitudinal)

YY = Year of the survey

T = Type of file (D, R, H or 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.

The household data file (H) must contain a record for every household who:
• Have been contacted AND

- Have completed a household interview AND
- At least one household member has completed a personal interview

The personal register file (R) must contain a record for every person currently living in the household or temporarily absent. In the <u>longitudinal</u> component (initial household) this file must contain also a record for every person moved out or died since previous wave (or since last interview) 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).

8.4. FORMAT

The files are in CSV-format (comma separated values). Most, following rules apply:

- header row (first record with the variable names)

- delimiter of variables is comma (,)

- decimal separator is point (.)

- character values are NOT enclosed by quotes

- blank variables are represented by nothing between the commas (…,,…)

- the first three variables should be Year, Country and ID (for the rest of the variables no fixed order is required)

e.g.

DB010,DB020,DB030,DB040,DB040_F,DB050,DB060,DB050_F,DB060_F, DB090,DB090_F,…

2003,BE,1,BE01,1,,,-2,-2,1.25,1,…

2003,BE,2,BE05,1,536,,1,-2,1.12345,1,…

2003,BE,3,BE01,1,,,-2,-2,1,1,…

8.5.   THE DATASETS

8.5.1.   *Cross-sectional dataset*

Member states annually transmit in year N+1 the sample data or all sub sample data in case of integrated or panel design from surveyed units in year N

8.5.2.   *Longitudinal dataset*

8.5.2.1.   Rotational panel scheme with 4 sub samples

Under a rotational panel scheme with 4 sub samples, Member States shall annually transmit in year N+2 not only the rotational sub sample up to year N with four year duration to the Commission (Eurostat), but also the current rotational sub samples with shorter duration (3 years and 2 years).
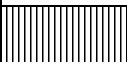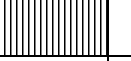
In this way, the Commission (Eurostat) will yearly get 3 over 4 sub samples that will cover at least the most recent 2 years. From the third year of data transmission, the Commission (Eurostat) will get 3 over 4 sub samples that will respectively cover the four, three and two most recent years. Each year all the sub samples will be transmitted together.

For each sub sample, data of the previous years will be updated according to the longitudinal controls.

The Commission (Eurostat) will annually make available for scientific purposes micro-data files at Community level of these sub samples. In this way, the Commission (Eurostat) will yearly make 3 over 4 sub samples that will cover at least the most recent 2 years available for scientific purposes. From the third year of data dissemination, 3 over 4 sub samples that will respectively cover the four, three and two most recent years will be disseminated.

Tables 1 to 5 illustrate the case of a Member State which starts the longitudinal survey in 2004:

Table 1: Sub samples to transmit to Eurostat in year 2007

Year of survey	Subsample 1[2]	Subsample 2	Subsample le 3	Subsample 4	Subsample 1'	Subsample 2'	Subsample 3'	Subsample 4'
2004	▓	▓	▓	▓				
2005		▓	▓	▓	▓			
2006								

---

[2] Needs to be collected for the cross-sectional component in the case of an integrated survey

2007								
2008								
2009								

Table 2: Sub samples to transmit to Eurostat in year 2008

Year of survey	Sub sample 1	Sub sample 2	Sub sample 3	Sub sample 4	Sub sample 1'	Sub sample 2'	Sub sample 3'	Sub sample 4'
2004			▓	▓				
2005			▓	▓	▓			
2006			▓	▓	▓	‖		
2007								
2008								
2009								

▓	Subsamples to transmit
‖	Subsample not to transmit, because it does not cover at least 2 years

Table 3: Sub samples to transmit to Eurostat in year 2009

Year of survey	Sub sample 1	Sub sample 2	Sub sample 3	Sub sample 4	Sub sample 1'	Sub sample 2'	Sub sample 3'	Sub sample 4'
2004				▓				
2005				▓	▓			
2006				▓	▓	▓		
2007				▓	▓	▓	‖	
2008								
2009								

Table 4: Sub samples to transmit to Eurostat in year 2010

Year of survey		Sub sample 4	Sub sample 1'	Sub sample 2'	Sub sample 3'	Sub sample 4'	Sub sample 1''	Sub sample 2''
2004								
2005			▓					
2006			▓	▓				

2007							
2008							
2009							

Table 5: Sub samples to transmit to Eurostat in year 2011

Year of survey		Sub sample 4	Sub sample 1'	Sub sample 2'	Sub sample 3'	Sub sample 4'	Sub sample 1''	Sub sample 2''
2004								
2005								
2006								
2007								
2008								
2009								

	Sub samples to transmit
	Sub sample not to transmit, because it does not cover at least 2 years

8.5.2.2.   Panel scheme (pure panel)

Under a panel scheme such as ECHP, Member States will yearly transmit updated longitudinal data covering the preceding four survey years to the Commission (Eurostat), the two first data transmissions will cover respectively the two and three most recent years.

If during the panel new households (not split off households) are added to panel (substitute the loss of other panel households, …), the new households should be considered the same way than a new sub sample in a rotational survey. They should only be transmitted when they cover at least 2 years.

8.6.   VARIABLES

8.6.1.   Variables names

The variable names are composed by two letters and one number

- First letter corresponding to the file name (D, R, H or P)

- Second letter corresponding to the domain and areas (B, E, H, L, S or Y) or X in the case of added derived variables

- The number is sequential and has no special meaning.

It may have an additional character at the end, the same variable has been split for some reason.

8.6.2. Flags variables

All variables will be completed by a flag variable (the flag-variable name is the variable name with the suffix "_F").

Exemption: the key variables (year of survey [xB010], country [xB020], IDs [xB030, RB040], and additionally constructed variables will NOT have a flag-variable.

The flag-variable gives supplementary information to the value of the main variable (codes may be different by variable or group of variables) and must be filled in a coherent way. For that purpose, the following set of rules has to be applied:

(1)   All the flag-variables are filled with a value.

(2)   A negative flag variable specifies the reason why the main variable is blank (codes are the same for all variables)

(3)   A positive flag (including zero) indicates that the variable is filled and may give some other information on the variable

8.6.3. Imputation factor variables

Income variables have an imputation factor. For gross variables, it is designed to tell whether the data was *collected* net or gross, and the imputation factor to give the amount of imputation because of non-response or net/gross conversion. For net variables, the flags are additionally designed to tell whether the *recorded* net value is net of taxes, social contributions or both.

The imputation factor-variable name is the variable name with the suffix "_I").

The imputation factor is a positive number, result of the division between collected value (during the interview) and the recorded value (in the variable).

Fully imputed value has an imputation factor of "0" (collected value is null)

No imputed value has an imputation factor of "1" (collected and recorded value are the same)

8.6.4. Link 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, HB030, RX030, PX030, and RB040* (* only longitudinal)

"Personal ID": RB030 and PB030.

are used as link variables.

Note that Personal ID is 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.

8.6.5. Income variables

The EU-SILC Commission Regulation specified the requirements about income as well as the Derogations given to some of Members States:

(1) *A key objective of EU-SILC is to deliver robust and comparable data on total disposable household income, total disposable household income before transfers (except old age and survivor's benefits; including old age and survivor's*

benefits), total gross income and gross income at component level

(2)   This objective will be reached in two steps, in that Member States will be allowed to postpone the delivery of some of the above data until after the first year of their operations. The only data for which delivery will not be compulsory as from the first year of the operation are as follows:

(3)   non-monetary components of employee (with the exception of company cars that is to be calculated as from the first year of the operation) and self-employed income, imputed rent and interest payments that shall be optional from the first year of the operation and compulsory from 2007;

- *gross employers' social insurance contributions shall only be included from 2007 if results of feasibility studies are positive.*

- *By way of exception to paragraph 2, Greece, Spain, France, Italy, Portugal, Latvia and Poland may be allowed not to deliver any gross income data as from the first year of their operation. These countries will, however, do their utmost best to deliver this data as soon as possible and definitively no later than 2007.*

(4)   In the case that Greece, Spain, France, Italy and Portugal can not deliver a gross income data component as from the first year of their operation, the corresponding net income component shall be required.

In this way, an income component shall always be recorded in the same form (gross, net of tax on income at source and social contributions, net of tax on income at source, net of tax on social contributions) according to the usual specification for this income component in the country.

In case a given income component is collected as the sum of subcomponents, some of them gross and the others net, the total amount will have to be recorded (and imputed where necessary) either in a gross or net form according with the usual specification for this income component in the country and preferably gross.

If an income component is collected, for all households within a country, in both forms gross and net, both should be provided to Eurostat.

As a consequence, at component level, same income variables exist for both net and gross amount. The only exception is for HY140 "tax on income and social contribution" which hasn't the same meaning if it has been declared "net" or "gross" (see variables description). Both variables may not be filled at the same time.

8.6.6. *Household and personal identification variables*

(see annex 2)

In both, the cross-sectional and the longitudinal survey, every household will receive a household number. This number is the base to construct the Household ID and the Personal ID. It should be a sequential number and not contain other information. It must NOT contain any information that conflicts with confidentiality rules. In the longitudinal survey this number must be unique for all the years of the survey.

Construction

- Household number: 1 - 999999 (maximum 6 digits)

- Household ID (cross-sectional): = Household number

- Household ID (longitudinal): = Household number + split number (2 digits)

    The split number for the first wave will always take value '00'.

    In the case of the household remaining entire, it keeps the Household number and Split number from one wave to the next.

    In the case of a split-off, the initial household will keep the Household number and Split number from one wave to the next. The other households, i.e. the split-off households will keep the same Household number, but will be assigned the next available unique Split number in sequence.

    In the case of a fusion of two sample households, if the new household is still at a previous address, it shall retain the Household number and Split number of the household that was at that address in the previous wave. If the new household is at a new address, the Household number and Split number of the household of the sample person who now has the lowest person number in 'the household register' will be retained.

- Personal ID : = Household ID + personal number (2 digits)

    Personal number: for every new person in the household add 1 to highest used persons number (for all the years of the survey and the Household ID)

    It refers to the number assigned in 'the household register' to the person the first time he/she is recorded as a household member. In the cross-sectional component, and in new households in the longitudinal component, it should correspond to the person's line position in 'the household register'.

    In the longitudinal survey Household ID and Personal ID never change, even not when the person moves to another household.

8.7. LIST OF VARIABLES

8.7.1. 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 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 DB090: HOUSEHOLD CROSS-SECTIONAL WEIGHT
X-L DB100: DEGREE OF URBANISATION
  L DB110: HOUSEHOLD STATUS

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: QUARTER 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: QUARTER 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: QUARTER 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 HB050: Quarter 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

  L PB080: Personal base weight for selected respondent

X-L PB100: Quarter of the personal interview

X-L PB110: Year of the personal interview

X-L PB120: Minutes to complete the personal questionnaire

X-L PB130: Quarter 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   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

8.7.2.   Derived variables

In the UDB, some derived variables have been added in order to ease the statistical exploitation of the data base. 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)

R file

X-L	RX010:	Age at the date of interview
X-L	RX020:	Age at the end of income reference period
X	RX030:	Household identification number (= HB030)

H file

X-L	HX010:	conversion factor: euro*rate = national currency
X	HX020:	Work intensity status of the household
X-L	HX040:	Household size
X-L	HX050:	Equivalised household size
X	HX060:	Household type
X	HX070:	Tenure status
X	HX080:	Poverty indicator
X-L	HX090:	Equivalised disposable income
L	HX100:	Equivalised disposable income quintile

P file

X-L	PX010:	conversion factor: euro*rate = national currency
X-L	PX020:	Age at the end of income reference period
X-L	PX030:	Household identification number (= HB030)

X-L PX040: selected respondent status (= RB245)
X PX050: Activity status

ANNEX 1: GENERAL DEFINITIONS

For the cross-sectional and longitudinal components of EU-SILC, the following definitions will be applied:

Year of survey

Means the year in which the survey-data collection, or most of the collection, is carried out.

Fieldwork period

Means the period of time in which the survey component is collected.

Reference period

Means the period of time to which a particular item of information relates.

Cross-sectional data

Means the data pertaining to a given time or a certain time period. The cross-sectional data may be extracted either from a cross-sectional sample survey with or without a rotational sample or from a pure panel sample survey (on condition that cross-sectional representativeness is guaranteed); such data may be combined with register data (data on persons, households or dwellings compiled from a unit-level administrative or statistical register).

Target primary areas

Means the subject areas to be collected on an annual basis.

Target secondary areas

Means the subject areas to be collected every four years or less.

Gross income

Means the total monetary and non-monetary income received by the household over a specified 'income reference period', before deduction of income tax, regular taxes on wealth, employees', self-employed and unemployed (if applicable) compulsory social insurance contributions and employers' social insurance contributions, but after including inter-household transfers received.

Disposable income

Means gross income less income tax, regular taxes on wealth, employees', self-employed and unemployed (if applicable) compulsory social insurance contributions, employers' social insurance contributions and inter- household transfers paid.

Collective household

Refers to a non-institutional collective dwelling such as a boarding house, dormitory in an educational establishment or other living quarters shared by more than five persons without sharing household expenses. Also included are persons living as lodgers in households with more than five lodgers.

Institution

Refers to old persons' home, health care institutions, religious institutions (convents, monasteries), correctional and penal institutions. Basically, institutions are distinguished from collective households, in that in the former, the resident persons have no individual responsibility for their housekeeping. In some cases, old persons' home can be considered as collective households on the basis of this last rule.

Age

Refers to the age at the end of the income reference period.

The following definitions will be applied for the longitudinal component of EU-SILC:

Longitudinal data

Means the data pertaining to individual-level changes over time, observed periodically over certain duration. The longitudinal data may come either from a cross-sectional survey with a rotational sample where individuals once selected are followed-up or from a pure panel survey; it may be combined with register data.

Initial sample

Refers to the sample of households or persons at the time it is selected for inclusion in EU-SILC.

Sample persons

Means all or a subset of the members of the households in the initial sample who are over a certain age.

Age limit used to define sample persons

In case of a four-year panel, this age limit shall not be higher than 14 years. In countries with a four-year panel using a sample of addresses or of households, all household members aged 14 and over in the initial sample shall be sample persons. In countries with a four-year panel using a sample of persons, this shall involve the selection of at least one such person per household.

The above mentioned minimum age limit shall be lower in case of a longer panel duration. For a panel duration exceeding eight years, members of all ages in the initial sample shall be sample persons, and children born to sample women during the time the mother is in the panel shall be included as sample persons.

Panel duration

Means the number of years over which sample persons, once selected into the sample, belong to the panel to obtain or compile longitudinal information.

Rotational design, integrated design

Refers to the sample selection based on a number of sub-samples or replications, each of them similar in size and design and representative of the whole population. From one year to the next, some replications are retained, while others are dropped and replaced by new replications.

In the case of a rotational design based on 4 replications with a rotation of one replication per year, one of the replications shall be dropped immediately after the first year, the second shall be retained for two years, the third for 3 years, and the fourth shall be retained for 4 years. From the second year onwards, one new replication shall be introduced each year and retained for 4 years.

Sample household

Means a household containing at least one sample person. A sample household shall be included in EU-SILC for the collection or compilation of detailed information if it contains at least one sample person aged 16 or more.

Co-residents or non-sample persons

Co-residents are all current residents of a sample household other than those defined above as sample persons.

Entire household

A sample household is said to be entire (whole) if it remains as one household, without forming an additional household and without the household disappearing, even though there might have been changes in its composition from the previous wave due to deaths, members moving out of scope or co-resident leaving the household, people joining the household, or births.

Initial/Split-off household

Sample household from wave x is said to have been 'split' if its sample persons from wave x reside at the time of wave x+1 in more than one private household within the national territories included in the target population.

When a split has occurred, one (and only one) of the resulting households shall be defined as the "initial" household, while one or more of the others are termed "split-off" households.

The following approach shall be followed in order to distinguish between "initial" and "split-off" households:

(1)   If any sample person of the wave x still lives at the same address as the last wave, then his/her household shall be defined as the "initial" household. All

sample persons who have moved shall form one or more "split-off" households;

(2)   If no sample person lives at the address of the last wave, then the household of the sample person who had the lowest person number in the register for the last wave shall be the initial household. In the case in which this person is no longer alive or in a private household within the national territory of the target population, the initial household shall be the household of the sample person with the lowest person number.

Fusion

Sample persons from different sample households from the previous wave join together to form a new household.

ANNEX 2: EXAMPLES OF HOUSEHOLD AND PERSONAL ID NUMBERS

1. CROSS-SECTIONAL

Household number = 123

Household ID = 123

Personal ID (person 1) = 12301

Personal ID (person 2) = 12302

Personal ID (person 3) = 12303

2. LONGITUDINAL

	WAVE 1	WAVE 2	WAVE 3	WAVE 4
Household	Jean Mary Elaine Lucas Peter	Jean Mary	Jean	Jean
			Mary Marcus	Mary Marcus
		Elaine Lucas Anne	Elaine Lucas Anne	Elaine
				Lucas
		Peter Sara	Peter Sara	Peter Sara

2.1. Wave 1

Household number = 123, Split = 00, Household ID (original)= 12300

Lives in Paris, is composed of five members all of them sample persons :

Line	Person number	First name	Personal id
A	01	Jean	123-00-01
B	02	Mary	123-00-02

Line	Person number	First name	Personal id
C	03	Elaine	123-00-03
D	04	Lucas	123-00-04
E	05	Peter	123-00-05

2.2. Wave 2

Jean and Mary stay in the same accommodation.

Elaine and Lucas moved out to Metz. They live with their aunt Anne (a co-resident).

Peter moved out to her sister house, Sara (a co-resident). They live in Bordeaux.

Household number = 123, Split = 00, Household ID (original and initial)= 12300

Line	Person number	First name	Personal id
A	01	Jean	123-00-01
B	02	Mary	123-00-02

Household number =123, Split = 01, Household ID (split)= 12301

Line	Person number	First name	Personal id
A	03	Elaine	123-00-03
B	04	Lucas	123-00-04
C	01	Anne	123-01-01

Household number =123, Split = 02, Household ID (split)= 12302

Line	Person number	First name	Personal id
A	05	Peter	123-00-05
B	01	Sara	123-02-01

Observation: There are two splits. The household composed of Jean and Mary is the initial household (they live in wave 1 at the same address) and keeps the household number and the split number.

The household composed of Elaine, Lucas and Anne is one split-off household, keeps the same household number but has a different split number '01'. The personal ID assigned to Anne is the Household ID =12301 plus add one to the highest used person number (for all the years of the survey and the Household ID), as the household did not exist in previous wave the person number is 01.

The household composed of Peter and Sara is one split-off household, keeps the same household number but has a different split number '02'(add 1 to the highest split). The personal ID assigned to Sara is the Household ID =12302 plus add one to the highest used person number (for all the years of the survey and the Household ID), as the household did not exist in previous wave the person number is 01.

2.3. Wave 3

Jean and Mary divorced. Mary stays at home with a new partner Marcus.

Jean moved out to a new household in the same city.

Rests of households from wave 2 stay at the same accommodation and with the same composition.

Household number = 123, Split = 00, Household ID (original and initial)= 12300

Line	Person number	First name	Personal id
A	02	Mary	123-00-02
B	06	Marcus	123-00-06

Household number = 123, Split = 03, Household ID (split)= 12303

Line	Person number	First name	Personal id
A	01	Jean	123-00-01

Household number =123, Split = 01, Household ID = 12301

Line	Person number	First name	Personal id
A	03	Elaine	123-00-03
B	04	Lucas	123-00-04
C	01	Anne	123-01-01

Household number =123, Split = 02, Household ID = 12302

Line	Person number	First name	Personal id
A	05	Peter	123-00-05
B	01	Sara	123-02-01

Observation: There is a split. The household composed of Mary and Marcus is the initial household (she lives at the same address) and keeps the household number and the split number. The personal ID assigned to Marcus is the Household ID =12300 plus add one to the highest used person number (for all the years of the survey and the Household ID). As the household exited from previous wave and for this household numbers 01-05 person number was already assigned, the person number for Marcus is 06.

The household composed by Jean is the split-off household, keeps the same household number but has a different split number. The split number is formed adding to 1 to the highest used split number (for all the years of the survey), as we had already 2 split the, new one will be 3.

2.4. Wave 4

Elaine moved out to Nice.

Lucas and Anne moved out to Nancy.

Rests of households from wave 2 stay at the same accommodation and with the same composition.

Household number = 123, Split = 00, Household ID (original)= 12300

Line	Person number	First name	Personal id
A	02	Mary	123-00-02
B	06	Marcus	123-00-06

Household number = 123, Split = 03, Household ID = 12303

Line	Person number	First name	Personal id
A	01	Jean	123-00-01

Household number =123, Split = 01, Household ID (initial)= 12301

Line	Person number	First name	Personal id
A	03	Elaine	123-00-03

Household number =123, Split = 04, Household ID (split)= 12304

Line	Person number	First name	Personal id
A	04	Lucas	123-00-04
B	01	Anne	123-01-01

Household number =123, Split = 02, Household ID = 12302

Line	Person number	First name	Personal id
A	05	Peter	123-00-05
B	01	Sara	123-02-01

Observation: There is a split. The household composed of Elaine is the initial household (she had the smallest line in the register in previous wave) and keeps the household number and the split number. The household composed by Lucas and Anne is the split-off household, keeps the same household number but has a different split number. The split number is formed adding to 1 to the highest used split number (for all the years of the survey), as we had already 3 split the new one will be 4.

2.5. Record of Persons

For determining the household identification numbers, it is necessary in case of household split to distinguish between the initial household (household that split remains identical to the parent household) and split-off households.

The initial household will be also very useful for the data collection purpose.

When a household split from wave x to wave x+1, it will be in the initial household where all information about the moving of the persons from previous wave will be collected (i.e. in the initial household, the full information required for current household members, the basic information for former household members and the basic information on household members in the previous wave that are no longer household members will be collected).

In the split-off household, only the full information required for current household members will be collected.

In the previous example, in wave 2, the initial household is the household composed of Jean and Mary, because they live at the same address than in wave 1. In this household information about moving of the persons from previous wave will be collected.

The household composed of Elaine, Lucas and Anne is a split household. In this household only information on current households members will be collected.

The household composed of Peter and Sara is also a split household. But only information about these members will be collected.

If nobody was living in the household during the income reference period as former household member.

Household ID = 12300

#		RB030	RB070	RB080	RB090	RB100	RB110
Line	Person number	Personal Id	Month of Birth	Year of Birth	Sex	Sample person or co-resident	Member-ship Status
A	01	123-00-01	05	1959	1	1	1
B	02	123-00-02	02	1962	2	1	1
C	03	123-00-03	01	1980	2	1	5
D	04	123-00-04	09	1982	1	1	5
E	05	123-00-05	12	1930	1	1	5

Household ID = 12301

#		RB030	RB070	RB080	RB090	RB100	RB110
Line	Person number	Personal Id	Month of Birth	Year of Birth	Sex	Sample person or co-resident	Member-ship Status
A	03	123-00-03	01	1980	2	1	2
B	04	123-00-04	09	1982	1	1	2
D	01	123-01-01	05	1963	2	2	3

Household ID = 12302

#		RB030	RB070	RB080	RB090	RB100	RB110
Line	Person number	Personal Id	Month of Birth	Year of Birth	Sex	Sample person or co-resident	Member-ship Status
A	05	123-00-05	12	1930	1	1	2
B	01	123-02-01	03	1926	2	2	3

ANNEX 3: EU-SILC SAMPLING DESIGNS

Belgium

The Belgian EU-SILC survey is a stratified two-stage sampling. There is no clustering of sampling units. The stratification is done by NUTS2 region (10 provinces plus the Brussels Capital region).

- Primary units: the municipalities (or part thereof in the larger ones) with probability proportional to size.

- Secondary units: private households by systematic sampling.

Czech Republic

A sample of dwellings is selected using a stratified two-stage design. The stratification of the Census Enumerations Units (CEUs-small geographical units) is done by region (NUTS4) and by number of residents.

- At the first stage, CEUs are sampled as primary sampling units (PSU) with probability proportional to their size.

- In the second stage, 10 dwellings are sampled in each sampled CEU by simple random sampling without replacement.

All the households and the individuals living in the selected dwellings are then eligible for interview.

Denmark

The sampling design is simple random sampling. The sample is a one stage sampling being the sampling unit the individual person. The sampling frame is all individuals aged 14 or more but only households where the selected person is 16 or more at the beginning of the survey year are included in the indicators computation of that year.

Germany

In 2005 the survey started with three quota samples and one random sample. Each year one quota sample is replaced by a further random sample. The sampling frame for the random subsamples is the permanent sample (DSP), a sampling frame recruited among former participants of the German Microcensus.

All the individuals living in the selected addresses are eligible for interview.

Estonia

The design used is one-stage stratified unequal probability sampling of household, with a household selected with probability proportional to the number of persons aged 14 and more in it. The EU-SILC sample is selected according to the following sampling procedure:

- Stratification by county level into three strata by the population size: "big" counties, "small" counties and the Hiiu County, which forms a separate stratum as the smallest county in terms of population size.

- A sample of persons aged 14 and more is selected with equal probabilities within strata.

All the households of the selected persons are identified and all eligible persons in the household are interviewed.

Ireland

In 2004, the Irish EU-SILC sample is selected according to a stratified two-stage selection. The stratification is done by County and degree of urbanisation.

- At the first stage, simple random selection of dwelling blocks.

- At the second stage, simple random selection of households.

Greece

In 2003, a sample of addresses is drawn according to a stratified two-stage selection. The stratification is done by NUTS2 region and degree of urbanisation.

- At the first stage, a sample of blocks is selected with probability proportional to the number of dwellings.

- At the second stage, households are systematically selected within each block.

All the persons living in the selected addresses are then interviewed in order to obtain information at personal level.

Spain

A sample of dwellings is drawn according to a stratified two-stage selection. The stratification of the Census sections is done by administrative region and number of dwellings.

- At the first stage, selection of Census sections with probability proportional to the number of dwellings.

- At the second stage, systematic selection of dwellings within each section.

All the persons living in the selected dwellings are eligible for interview.

France

The type of sampling design is a stratified three-stage sampling. In 2004, a sample of dwellings is drawn from the 1999 Master Sample updated for the "new" dwellings (i.e. the units that came out after the 1999 Census). The selection is done so as to make the sample self-weighted.

- At the first stage, selection by groups of municipalities proportional to size (stratified according geographical criteria as NUTS2 and degree of urbanisation).

- At the second stage, the systematic selection is of dwellings for the urban areas and ad-hoc groups of municipalities for the rural areas.

- The third stage only exists for the rural areas and the dwellings are selected by systematic sampling.

All the households and the individuals living in the selected dwellings are interviewed.

Italy

In 2004, a sample of households is drawn according to a stratified two-stage selection. The stratification of the municipalities is done by administrative region and number of residents.

- At the first stage, selection of four municipalities with probability proportional to the number of residents.

- At the second stage, systematic selection of households within each municipality.

All the persons living in the selected households are then eligible for interview.

Cyprus

The sample design is one-stage stratification. The sampling units are private household which are selected by simple random sampling within each stratum (9 strata based on District).

All the individuals that are current members of the selected households are eligible for interview.

Latvia

The Latvian EU-SILC sample is according to a stratified two-stage design. The stratification is based on the degree of urbanisation.

- At the first stage, the primary sampling units (PSU, Population Census counting areas) are selected in each stratum with probability proportional to the number of households.

- At the second stage, a simple random sample of units (addresses) is selected within each area.

In Latvia several households can be registered in one address. All households and individuals living in the selected address are included in the survey.

Lithuania

The new subsample of households is selected by stratified sample design. The stratification is based on degree of urbanisation into seven strata.

- A simple random sample of non-institutional persons aged 16 and over is selected in each stratum from the Population Register.

Households that live in the selected persons addresses are surveyed.

Luxembourg

The type of sampling design is stratified simple random sampling. In 2003, first year of the survey, two samples are drawn independently:

- A sample of "tax" households, which are in fact a group of persons who depends on the same Social Security system.

- A sample of dwellings wherein none of the members depends on Luxembourgish Social Security system.

A "tax household" is basically a group of persons living in the same dwelling and who depend on the same Luxembourgish Social Security system.

The samples are selected by stratified simple random sampling.

Hungary

EU-SILC sample is selected by a stratified two-stage sampling in one part of the population and by a stratified one-stage sampling in the other part. Localities are stratified by General Election Districts and size (in terms of number of dwellings).

- In the first part of the population, one locality is selected with probability proportional to the number of dwellings. Within each selected locality, a systematic selection of dwellings is done.

- In the other part of the population, a systematic selection of dwellings is done in each stratum.

The final sampling units are the dwellings and, in each of them, every household is observed.

The Netherlands

The EU-SILC sample is composed of the addresses that took part in the Labour Force Survey (LFS) and are willing to cooperate to EU-SILC. The LFS sample is selected according to a stratified three-stage sampling design. The stratification of the municipalities is done by geographical criteria (COROP and interviewer region).

- At the first stage, municipalities are selected with a probability proportional to the number of addresses and according to the above mentioned stratification. At the second stage, there is a simple random selection of addresses within each municipality.

- At a third stage, persons of 16 and older are selected by simple random sampling.

The LFS has a panel structure with five rotational groups. When the first wave (face-to-face interviews) has been completed, addresses with all residents aged over 64 are removed from the sample. In order to get full covering of the target population, an additional sample of addresses with all residents aged 65 and over is drawn for the EU-SILC sample.

All the households and the individuals living at the selected addresses are then eligible for interview.

Austria

The sampling design is simple random sampling. The sample is stratified by geographical units. These units are used in Austrian microcensus to distribute addresses among the pool of interviewers. Implicitly this procedure achieves both a regionally stratified sample and a control of the number of addresses allocated to each interviewer. Sampling units are dwelling units registered in the Central Residence Register.

All the households and the individuals living in the eligible addresses are interviewed.

Poland

The Polish EU-SILC sample is selected according to a stratified two-stage design. The stratification is based on NUTS2 region and degree of urbanisation.

- At the first stage, Census areas are selected with probability proportional to the number of dwellings.

- At the second stage, a simple random sample of dwellings is selected.

All the households and the individuals living in the selected dwellings are eligible for contact.

Portugal

The EU-SILC sample follows a stratified two-sage cluster sampling design.

- At the first stage, Census sections are systematically selected. Primary Sampling Units are the areas of the Master Sample (made of census enumeration areas) and they are stratified by a regional criterion.

- At the second stage, a simple random sample of households is selected in each Census section.

All the persons living in the same dwelling are interviewed.

Slovenia

The sample for the Slovenian EU-SILC is selected according to a stratified two-stage design. The strata are defined according to the size of the settlement and its proportion of agricultural households.

- In each stratum, Primary Sampling Units (PSU) are firstly systematically selected.

- In the second phase, seven persons aged 16 and over are selected in each PSU.

Finally, all the households the selected persons belong to are eligible for contact.

Slovakia

One-stage stratified sampling is used in EU-SILC. Stratification is based on geographical criteria (NUTS3 region and degree of urbanisation).

The proportional number of households is selected by simple random sampling in individual strata.

All the households and the individuals living in the selected dwellings are contacted.

Finland

The sampling design of the Finnish EU-SILC survey is a two-phase sampling design. A systematic sampling of persons aged 16 years and more is carried out in the Population Register in order to get the basis for a Master Sample. Then, all the dwellings with at least one selected person are included in that Master Sample. The Master Sample is stratified according to socio-economic criteria.

- A simple random sample without replacement of dwellings is selected in each stratum.

Finally, all the households and the individuals living in the selected dwellings are eligible for interview.

Sweden

A systematic sample of persons aged 16 and over is drawn from the Population Register (RTB). The final EU-SILC sample also includes a panel of persons that was drawn in 1980 and are re-interviewed every 8 year. In order to cover the whole target population, this panel has been supplemented every 8 year with a systematic sample of immigrants and a systematic sample of individuals aged 16-23.

Finally, all the households the selected persons belong to are then interviewed.

United Kingdom

Data is collected from two sources. First, data is collected by the Office of National Statistics (ONS), using the General Household Survey. Second, a sample of 300 households is collected by NISRA (Northern Ireland Statistics and Research Agency) using the Continuous Household Survey (CHS).

EU-SILC uses a probability, stratified two-stage sample design. Households are sampled from the small users Postcode Address File (PAF).

- The postcode sectors are the Primary Sampling Units. The Postcode address file is ordered by postcode sector, which are similar in size to a UK electoral ward area.

- The Secondary Sampling Units are addresses within those sectors.

All adults aged 16 or over from every household at the sampled address are interviewed.

Iceland

The sampling design is one-stage simple random sample without stratification. The sampling units are persons aged 16 years and more living in private households selected from the Population Register.

All the households the selected persons belong to are then interviewed.

Norway

The EU-SILC in Norway comprises two parts. First, an "old" panel was drawn in 1997 according to a stratified two-stage design. Municipalities are stratified by socio-economic criteria and municipalities are drawn with probability proportional to the population size.

- A systematic sample of registered persons aged 16 and over is selected in each municipality so as to make the final sample self-weighted.

For the "new" part, the sample units are the persons aged 16 years and over that are registered in the Central Population Register. The sample is systematically drawn within one-year groups so as to maintain self-weighting.

All the households the selected persons belong to are then interviewed.

Table: Sampling designs by country

	Type of sampling design	*Number of sampling stages*	*First-stage*			*Final stage*		
			Type of unit	*Selected by*	*Stratification*	*Type of unit*	*Selected by*	*Stratification*
Belgium	Stratified two-stage sampling	2	Municipalities	Pps sampling	NUTS2 Region	Households	Systematic sampling	N
Czech Republic	Stratified two-stage sampling	2	Census sections	Pps sampling	NUTS4 and size of municipality	Dwellings	Simple random sampling	N
Denmark	Simple random sampling	1				Persons 14+	Simple random sampling	N
Germany	Quota sampling + random part							
Estonia	Stratified systematic sampling	1				Persons 14+	Systematic sampling	County level ("big" counties, "small" counties and Hiiu)
Ireland	Stratified two-stage sampling	2	Dwelling blocks	Simple random sampling	NUTS2 and degree of urbanisation	Households	Simple random sampling	N
Greece	Stratified two-stage sampling	2	Dwelling blocks	Pps sampling	NUTS2 and degree of urbanisation	Households	Systematic sampling	N
Spain	Stratified two-stage sampling	2	Census sections	Pps sampling	Administrative region and size of the municipality	Dwellings	Systematic sampling	N
France	Stratified three-stage sampling	3	Groups of municipalities	Pps sampling	NUTS2, degree of urbanisation and rural/urban	Dwellings	Systematic sampling	N

	Type of sampling design	*Number of sampling stages*	First-stage			Final stage		
			Type of unit	*Selected by*	*Stratification*	*Type of unit*	*Selected by*	*Stratification*
Italy	Stratified two-stage sampling	2	Municipality	Pps sampling	Administrative region and number of residents	Dwellings	Systematic sampling	N
Cyprus	Stratified simple random sampling	1				Households	Simple random sampling	Geographical criteria
Latvia	Stratified two-stage sampling	2	Census sections	Pps sampling	Degree of urbanisation	Dwellings	Simple random sampling	N
Lithuania	Stratified simple random sampling	1				Persons 16+	Simple random sampling	Degree of urbanisation
Luxembourg	Stratified simple random sampling	1				"Tax" households	Simple random sampling	Social Security data
Hungary	Stratified two-stage sampling	2	Localities	Pps sampling	Election district and number of dwellings	Dwellings	Systematic sampling	N
The Netherlands	Stratified three-stage sampling	3	Municipalities	Pps sampling	COROP and interviewer region	Persons 16+	Simple random sampling	N
Austria	Simple random sampling	1				Dwellings	Simple random sampling	N
Poland	Stratified two-stage sampling	2	Census sections	Pps sampling	NUTS2 and degree of urbanisation	Dwellings	Simple random sampling	N
Portugal	Stratified two-stage sampling	2	Census sections	Pps sampling	NUTS3	Dwellings	Simple random sampling	N

	Type of sampling design	Number of sampling stages	First-stage			Final stage		
			Type of unit	Selected by	Stratification	Type of unit	Selected by	Stratification
Slovenia	Stratified two-stage sampling	2	Census sections	Pps sampling	Size of the settlement and proportion of agricultural households	Persons 16+	Systematic sampling	N
Slovakia	Stratified simple random sampling	1				Households	Simple random sampling	NUTS3 and degree of urbanisation
Finland	Post-stratified unequal probability sampling	1				Dwellings	Pps sampling	Socio-economic criteria [1]
Sweden	Systematic sampling	1				Persons 16+	Systematic sampling	N
United Kingdom	Stratified two-stage sampling	2	Postcode sectors	Pps sampling	2001 Census data	Dwellings	Systematic sampling	N
Iceland	Simple random sampling	1				Persons 16+	Simple random sampling	N
Norway	Systematic sampling	1				Persons 16+	Systematic sampling	One-year age group

Source: Quality Reports.

Pps sampling = proportional-to-size sampling.

(1) Stratification *a posteriori*, according to socio-economic criteria.

Contact:

BERNARD, Telephone:(352) 4301-37330, bruno.bernard@ec.europa.eu