



EUROPEAN COMMISSION
EUROSTAT

Directorate E: Social and regional statistics and geographical information
system
Unit E-2: Living conditions



DOC. PAN 105/2003-05

ANONYMISATION CRITERIA APPLIED TO THE USERS' DATABASE

This version of the document includes the anonymisation criteria for the ECHP UDB as from wave 4 onwards (see annex). The 1999 version was limited to the anonymisation of waves 1 and 2.

May 2003

ANONYMISATION CRITERIA APPLIED TO THE USERS' DATABASE

Reminder concerning the need for anonymisation criteria

ECHP micro-data contain information considered “confidential” in terms of the EU Council regulation N° 322/97 of 17/2/97 on Community statistics, the so-called EU “Statistical Law”. Therefore, direct access to these original data has had to be more restrictive than what is desirable for a full exploitation of the data despite the strong interest expressed by many researchers and other users. Also, the structure of the ECHP “production database” is complex and thus very difficult for use outside as well as inside Eurostat, which increases significantly the risk of errors in using the data.

In view of this situation and given the growing demand for ECHP data from both inside and outside the Commission, Eurostat decided to construct back in November 1997 a user-friendly “longitudinal users’ database” (UDB) containing “anonymised” ECHP household’s and personal micro-data. (see PAN document N° 85/97)

Final outcome of the negotiations with ECHP “National Data Collection Units” (NDUs)

The anonymisation of the data has required the definition of “objective anonymisation criteria” to be met by this UDB. By “objective”, it is meant that once these criteria are applied to the various ECHP files, there should be no risk that an individual statistical unit could be identified through “all the means that might reasonably be used by a third party to identify the said statistical unit” (“Statistical Law”).

Provided that all these criteria are met, ECHP data can thus be considered “non-confidential” in the sense of the Statistical Law and therefore made more widely available, although direct access to such anonymised data needs to be restricted by means of **research contracts** stipulating the strict conditions of use and access.

It is in this context that at the 1997 ECHP Working Group (Luxembourg, 13-14/11/97), Eurostat proposed to ECHP NDUs (either national statistical institutes or research centres, depending on the country) a first set of such anonymisation criteria to be applied to the various ECHP files. It also submitted for discussion different rules, contractual requirements that should allow easier access to these “anonymised” data. The objective was therefore to open up access to ECHP micro-data without jeopardising both the necessary conditions of data confidentiality and the value of the data.

Extensive negotiations with all ECHP NDUs concerning the final set of “anonymisation criteria” to be applied to the UDB as well as the rules for accessing this UDB, have been conducted up to November 1998. Apart from some minor differences, all 14 Member States that are taking part in the ECHP have agreed on the same set of criteria and rules, meaning that they have supported both the idea of direct access to anonymised micro-data and the concrete EU harmonised proposal for anonymisation submitted by Eurostat.

Final criteria include: suppression of names and addresses (which is already done in any case in the files Eurostat receives from NDUs), aggregation of ISCO, NACE and NUTS

codes at the appropriate level, any other aggregation of variables or elimination of information to guarantee confidentiality (age, migration, citizenship, etc.). A comprehensive list of these is annexed to the present document; it also forms part of the documentation that is provided on the CD-ROM containing the UDB.

The rules for accessing the UDB, and more specifically the Eurostat dissemination and price policy, are given in PAN document N° 106/99.

As to the UDB itself, a detailed description of its structure together with the main differences between the structures of the “users’ database” and the “production database” can be found in PAN document N° 104/99.

Annex: ECHP UDB - Anonymisation - Impact on variables

1. AGE OF PERSONS; YEAR OF BIRTH, OF LAST CHANGE IN MARITAL STATUS, OF ARRIVAL IN THIS REGION / COUNTRY; MONTH MOVING IN/OUT OF HOUSEHOLD.....	5
2. CITIZENSHIP.....	6
3. MIGRATION TRAJECTORY, FOREIGN COUNTRY OF BIRTH, LAST FOREIGN COUNTRY OF RESIDENCE	6
4. REGIONAL INFORMATION	8
5. NUMBER OF ROOMS	9
6. OCCUPATION CODES (ISCO)	9
7. SECTOR OF ACTIVITY (NACE)	11
8. SAMPLING INFORMATION.....	12
9. FOREIGN LANGUAGES USED IN PROFESSIONAL ACTIVITY	12
10. NUMBER OF NIGHTS SPENT IN HOSPITAL THE YEAR BEFORE.....	12
11. IDENTIFICATION NUMBERS.....	13
12. DATE OF INTERVIEW.....	13
13. INCOME.....	14

1. AGE OF PERSONS; YEAR OF BIRTH, OF LAST CHANGE IN MARITAL STATUS, OF ARRIVAL IN THIS REGION / COUNTRY; MONTH MOVING IN/OUT OF HOUSEHOLD

Information at the upper end of the age distribution has been restricted to avoid identification of individuals in small population groups.

All countries except Germany (original ECHP): Individuals born before 1910 are assigned 1909 as year of birth.

Germany (original ECHP): individuals born before 1925 are assigned 1924 as year of birth. The month of birth has been deleted.

Related information (last change in marital status, year of arrival in this country / region) has been changed accordingly.

Germany (original ECHP): no information is provided on 'year of arrival to current region' and 'year of arrival to current country'.

Germany (converted panel): no information is provided on 'year of arrival to current region'.

Germany (original ECHP): no information is provided on 'month of birth' and 'month moving in / out'.

Germany (converted panel): no information is provided on 'month of birth'.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Change</i>	<i>Applicable for</i>
BIRTHYY	Year of birth	if before 1910 then set to 1909 if before 1925 then set to 1924	all, except DE DE
PD001	Year of birth	= BIRTHYY	all
PD003	Age	based on PD001	all
PD006	Last change in marital status	if before 1910 then set to 1909 if before 1925 then set to 1924	all, except DE DE
PM003	Year of arrival in this region	if before 1910 then set to 1909	all
PM006	Year of arrival in this country	if before 1910 then set to 1909	all
RD001	Year of birth	= BIRTHYY	all
RD003	Age	based on RD001	all

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
BIRTHMM	Month of birth	all, except DE,D1
PD002	Month of birth	all, except DE,D1
PM003	Year of arrival in this region	all, except DE,D1
PM006	Year of arrival in this country	all, except DE
RD002	Month of birth	all, except DE,D1
PIO_MMi	Month moving in / born or out / died (in wave i)	all, except DE

2. CITIZENSHIP

The citizenship has been limited to the 4 modalities "citizenship of the country where the survey was carried out", "another EU citizenship", "other citizenship (extra-EU)" and "Not national, but citizenship unknown".

Germany (original ECHP and converted panel): no information is provided on second citizenship.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
PM010	Existence of a second citizenship	all, except DE,D1
PM011	Second citizenship	all, except DE,D1

3. MIGRATION TRAJECTORY, FOREIGN COUNTRY OF BIRTH, LAST FOREIGN COUNTRY OF RESIDENCE

Migration trajectory

Germany (original ECHP and converted panel): No information is provided.

Foreign country of birth and last foreign country of residence

The restrictions are as follows.

Germany (original ECHP and converted panel), Greece and the Netherlands: No information is provided.

All countries except Germany (original ECHP and converted panel), Greece and the Netherlands: 2 modalities (1 = "another EU Member State" and 2 = "another foreign country") are given.

All countries except Germany (original ECHP and converted panel), Greece, the Netherlands and Finland: 7 modalities (901, 902, 910, 920, 930, 940, 990) are also given. See code below.

All countries except Germany (original ECHP and converted panel), Greece, Italy, the Netherlands, Austria and Finland: 11 modalities (901, 902, 911, 912, 913, 921, 922, 931, 932, 940, 990) are also provided. See code below.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
PM001	Migration trajectory	all, except DE, D1
PM005A	Last foreign country of residence before coming to present country (11 mod.)	all, except DE, D1, EL, I, NL, A and FIN
PM005B	Last foreign country of residence before coming to present country (7 mod.)	all, except DE, D1, EL, NL and FIN
PM005C	Last foreign country of residence before coming to present country (2 mod.)	all, except DE, D1, EL and NL
PM007A	Foreign country of birth (11 mod.)	all, except DE, D1, EL, I, NL, A and FIN
PM007B	Foreign country of birth (7 mod.)	all, except DE, D1, EL, NL and FIN
PM007C	Foreign country of birth (2 mod.)	all, except DE, D1, EL and NL

Country codes used for migration

<i>Value</i>	<i>Description</i>
901	EU
902	Other European Countries
910	Africa
911	North Africa
912	West Africa
913	Central, East and South Africa
920	America
921	North America
922	Central and South America
930	Asia
931	Near and Middle East
932	Other Asian countries
940	Australia, Oceania and Other Countries
990	Any Country, not elsewhere covered

4. REGIONAL INFORMATION

Breakdown

NUTS 2 level is given for Portugal and United Kingdom (converted panel).

For Finland, NUTS 2 level is given except for Ahvenanmaa/Åland (FI2) that is merged with Etelä-Suomi (FI12).

NUTS 1 breakdown is provided for Belgium, Denmark, Greece, Spain, France, Italy, Luxembourg (original ECHP and converted panel), Austria and the United Kingdom (original ECHP).

For Ireland (which forms a single NUTS 1 region), 2 modalities are distinguished: "Dublin" and "Rest of the Country".

For Germany (converted panel), two NUTS 1 regions are grouped together - "Rheinland-Pfalz" and "Saarland".

No regional information is provided for Germany (original ECHP) and the Netherlands.

All countries: As soon as at least one member of a household moves abroad and if they are followed-up abroad, all regional information has been dropped for this household as a whole.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Change</i>	<i>Applicable for</i>
HREGIONi, i = wave number	Region in which the household is situated (wave i)	regroup in 2 categories (‘Dublin’ and ‘Rest of the country’)	IRL
		NUTS 2, but regroup FI2 and FI12	FIN
		NUTS 1, but regroup DEC and DEB into DEX	D1
		NUTS 1	B, DK, EL, E, F, I, LU, L1, A, UK
		NUTS 2	P, U1
HG015	= REGIONi		
<i>Variable</i>	<i>Label</i>	<i>Available for</i>	
HG015	Region in which the household is currently situated	all ¹ , except DE and NL	
HREGIONi	Region in which the household is situated (wave i)	all ¹ , except DE and NL	

¹ Unless moved abroad

5. NUMBER OF ROOMS

Greece, Netherlands and Austria: the number of rooms (excluding kitchen) in the home has been limited to "6 or more", and accordingly, the number of rooms including kitchen has been limited to "7 or more".

For the other Member States the limits are "9 or more" and "10 or more" respectively, though the less detailed versions of variables are also provided for consistency.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
HA006	Number of rooms without kitchen (limited to 9)	all, except EL, NL and A
HA006A	Number of rooms without kitchen (limited to 6)	all
HA007	Number of rooms with kitchen (limited to 10)	all, except EL, NL and A
HA007A	Number of rooms with kitchen (limited to 7)	all

6. OCCUPATION CODES (ISCO)

The most detailed breakdown (28 modalities: 11+12, 13, 21, 22, 23, 24, 31, 32, 33, 34, 41, 42, 51, 52, 53, 61, 71, 72, 73, 74, 81, 82, 83, 84, 91, 92, 93, 94, 01+99) is provided for Belgium, Denmark, France, Ireland, Luxembourg (original ECHP and converted panel), Portugal and the UK (original ECHP and converted panel). For the other countries (Germany - original ECHP and converted panel, Greece, Spain, Italy, the Netherlands, Austria and Finland) some categories have been merged (22 modalities: 11+12, 13, 21+22, 23, 24, 31+32, 33+34, 41+42, 51, 52, 53, 61, 71+74, 72+73, 81+83, 82, 84, 91, 92, 93, 94, 01+99). See code below.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
PE006A	Current job – ISCO (28 modalities)	all, except DE, D1, EL, E, I, NL, A and FIN
PE006B	Current job – ISCO (22 modalities)	all
PE006C	Current job – ISCO (1-digit)	all
PJ007A	Previous job - ISCO (28 modalities)	all, except DE, D1, EL, E, I, NL, A and FIN
PJ007B	Previous job - ISCO (22 modalities)	all
PJ007C	Previous job - ISCO (1-digit)	all

ISCO breakdown used

<i>Value</i>	<i>Description</i>
1	Legislators, senior officials and managers
1112	Legislators, senior officials and corporate managers
13	Managers of small enterprises
2	Professionals
2122	Physical, mathematical, engineering, life science and health professionals
21	Physical, mathematical and engineering science professionals
22	Life science and health professionals
23	Teaching professionals
24	Other professionals
3	Technicians and associate professionals
3132	Physical, engineering, life science and health associate professionals
31	Physical and engineering science associate professionals
32	Life science and health associate professionals
3334	Teaching and other associate professionals
33	Teaching associate professionals
34	Other associate professionals
4	Clerks
4142	Clerks
41	Office clerks
42	Customer services clerks
5	Service workers and shop and market sales workers
51	Personal and protective services workers
52	Models, salespersons and demonstrators
53	5-Miscellaneous (ECHP-specific code)
6	Skilled agricultural and fishery workers
61	Skilled agricultural and fishery workers
7	Craft and related trades workers
7174	Extraction, building trades, other craft and related trades workers
71	Extraction and building trades workers
72	Metal, machinery and related trades workers
7273	Metal, machinery, precision, handicraft, printing and related trades workers
73	Precision, handicraft, printing and related trades workers
74	Other craft and related trades workers
8	Plant and machine operators and assemblers
8183	Stationary-plant and related operators, drivers and mobile-plant operators
81	Stationary-plant and related operators
82	Machine operators and assemblers
83	Drivers and mobile-plant operators
84	8 – Miscellaneous (ECHP-specific code)
9	Elementary occupations
91	Sales and services elementary occupations
92	Agricultural, fishery and related labourers
93	Labourers in mining, construction, manufacturing and transport
94	9 – Miscellaneous (ECHP-specific code)
0199	Armed Forces or Missing

7. SECTOR OF ACTIVITY (NACE)

The most detailed breakdown (A+B, C, DA, DB+DC, DD+DE, DF-DI, DJ+DK, DL-DN, E, F, G, H, I, J, K, L, M, N, O-Q) is provided for Belgium, Denmark, France, Ireland, Luxembourg (original ECHP and converted panel), Austria, Portugal and the UK (original ECHP and converted panel). For the other countries (Germany - original ECHP and converted panel, Greece, Spain, Italy, the Netherlands and Finland), the categories C and E have been merged into C+E. See code below.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
PE007A	Current job - NACE (19 modalities)	all, except DE, D1, EL, E, I, NL and FIN
PE007B	Current job - NACE (18 modalities)	all
PE007C	Current job - NACE (3 major sectors)	all
PJ008A	Previous job - NACE (19 modalities)	all, except DE, D1, EL, E, I, NL and FIN
PJ008B	Previous job - NACE (18 modalities)	all
PJ008C	Previous job - NACE (3 major sectors)	all

NACE breakdown used

<i>Value</i>	<i>Description</i>
1	Agriculture
A+B	Agriculture, hunting, forestry and fishing
2	Industry
C+E	Mining and quarrying; electricity, gas and water supply
C	Mining and quarrying
D	Manufacturing
DA	Manufacture of food products, beverages and tobacco
DB+DC	Manufacture of textiles, clothing and leather products
DD+DE	Manufacture of wood and paper products; publishing and printing
DF-DI	Manufacture of coke, refined petroleum, nuclear fuel, chemicals, chemical products, rubber & plastic and other non-metallic mineral products
DJ+DK	Manufacture of metal products, machinery and equipment n.e.c.
DL-DN	Other manufacturing
E	Electricity, gas and water supply
F	Construction
3	Services
G	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods
H	Hotels and restaurants
I	Transport, storage and communication
J	Financial intermediation
K	Real estate, renting and business activities
L	Public administration
M	Education
N	Health and Social work
O-Q	Other community, social, personal service activities; private households with employed persons; Extra-territorial organisations and bodies

8. SAMPLING INFORMATION

Not given for Germany (original ECHP and converted panel) and the Netherlands.

For the other countries, this information has been recoded.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Change</i>	<i>Applicable for</i>
HG005	Primary strata (anonymised)	Recode	all
HG006	Primary sampling unit (anonymised)	Recode	all
HG007	The order of selecting PSUs (anonymised)	Recode	all

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
HG005	Primary strata (anonymised)	all, except DE, D1 and NL
HG006	Primary sampling unit (anonymised)	all, except DE, D1 and NL
HG007	The order of selecting PSUs (anonymised)	all, except DE, D1 and NL

9. FOREIGN LANGUAGES USED IN PROFESSIONAL ACTIVITY

Not given for Germany (original ECHP and converted panel).

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
PE018	First foreign language used in current job	all, except DE, D1
PE019	Second foreign language used in current job	all, except DE, D1
PE020	Third foreign language used in current job	all, except DE, D1

10. NUMBER OF NIGHTS SPENT IN HOSPITAL THE YEAR BEFORE

Not given for Germany (original ECHP).

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
PH007	Number of nights spent in hospital during the past 12 months	all, except DE

11. IDENTIFICATION NUMBERS

Identification numbers have been anonymised.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Change</i>	<i>Applicable for</i>
PID	Personal identification number	Recode	all
PID1	Personal identification number	Recode	all
PID2	Personal identification number	Recode	all
HID	Household identification number	Recode	all
HIDi	Household identification number	Recode	all
PG001	Basic household identification number	Recode	all
HG001	Personal identification number of the reference person	Recode	all
HG002	Personal identification number of the person responsible for accomodation	Recode	all
HG003	Personal identification number of the person responding to the household interview	Recode	all
RG001	Basic household identification number	Recode	all

12. DATE OF INTERVIEW

Not given for Germany.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
HG012	Day of the interview	all, except DE and D1
HG013	Month of the interview	all, except DE
HG014	Year of the interview	all, except DE
PG005	Day of the interview	all, except DE and D1
PG006	Month of the interview	all, except DE
PG007	Year of the interview	all, except DE

13. INCOME

Some income components (and the corresponding indices) are not provided separately for Germany (original ECHP) as listed in the following table. Of course, these components not provided separately are included in aggregates such as HI100, HI110, HI120, HI130, HI111, HI132.

Impact on variables:

<i>Variable</i>	<i>Label</i>	<i>Available for</i>
HI1111	Household: wage and salary earnings (regular)	all, except DE
HI1112	Household: wage and salary earnings (lump sum)	all, except DE
HI121	Household: capital income	all, except DE
HI121X	Household: capital income (imputation index)	all, except DE
HI123	Household: private transfers received	all, except DE
HI123X	Household: private transfers received (imputation index)	all, except DE
HI133	Household: family related allowances	all, except DE
HI133X	Household: family related allowances (imputation index)	all, except DE
HI134	Household: sickness/invalidity benefits	all, except DE
HI134X	Household: sickness/invalidity benefits (imputation index)	all, except DE
HI135	Household: education related allowances	all, except DE
HI135X	Household: education related allowances (imputation index)	all, except DE
HI136	Household: any other (personal) benefits	all, except DE
HI136X	Household: any other (personal) benefits (imputation index)	all, except DE
PI1111	Persons: wage and salary earnings (regular)	all, except DE
PI1112	Persons: wage and salary earnings (lump sum)	all, except DE
PI121	Persons: capital income	all, except DE
PI123	Persons: private transfers received	all, except DE
PI1321	Persons: old-age related benefits	all, except DE
PI1322	Persons: survivor's benefits	all, except DE
PI133	Persons: family related allowances	all, except DE
PI134	Persons: sickness/invalidity benefits	all, except DE
PI135	Persons: education related allowances	all, except DE
PI136	Persons: any other (personal) benefits	all, except DE