EUROPEAN COMMISSION
EUROSTAT

Directorate E: Social statistics
Unit E-1: Labour market

Structure of Earnings Survey 2002

Eurostat's arrangements for implementing the Council Regulation 530/1999 and the Commission Regulation 1916/2000

DATE: 19-DEC-08

CONTENTS

1. GENERAL REMARKS 4

1.1. Objective of the SES 2002 4

1.1.1. Goal of the survey 4

1.1.2. Data sources and reference period 4

1.2. Sampling design: a two-stage sample 5

1.3. Scope of the survey 5

1.3.1. Economic activities covered 5

1.3.2. Population and sample of employees to be covered 5

1.3.3. Categories of workers which are to be **included** in the SES 6

1.3.4. Categories of workers which are **excluded** from the SES 6

2. INFORMATION REQUESTED IN REGULATION 1916/2000 7

2.1. Information about the local unit to which the sampled employees are attached 7

2.2. Information relating to each employee in the sample 10

2.3. Information concerning earnings, hours paid and days not worked, etc 13

2.4. Grossing-up factors 19

3. PROCESSING OF MICRO-DATA 20

3.1. Flowchart 20

3.2. Data Coding 21

3.3. Data Validation and Eurostat Calculations 21

3.3.1. Introduction 21

3.3.2. Global checks 22

3.3.3. Plausibility checks 22

3.3.4. Calculations that Eurostat will undertake 27

3.4. Technical format and transmission of the SES micro-data 29

3.4.1. The requirements 29

3.4.2. Contents of records A and B 29

3.4.3. Variables 29

3.4.4. Content of record A 31

3.4.5. Content of record B 32

3.4.6. Transmission 33

3.5. Treatment of confidentiality 34

3.5.1. Eurostat's strategy: a systematic approach to protect data confidentiality 34
3.5.2. Anonymisation of the 2002 SES micro-data 35
3.5.3. Protecting SES tabular data : another layer of protection 36
3.6. Tables 37
3.6.1. Dissemination of the 2002 SES results 37
3.6.2. New Cronos tables: introductory notes about the attached tables 37
3.6.3. Information about the variables in the attached New Cronos tables 38
3.6.4. Planned 2002 SES tabular analyses 42
4. ANNEXES 44
4.1. Annex 1: Commision Regulation 1916/2000 44
4.2. Annex 2: Codes for Alphanumeric Variables 45

1. GENERAL REMARKS

1.1. Objective of the SES 2002

1.1.1. Goal of the survey

The Structure of Earnings Survey (SES) for 2002 is the first of a series of four-yearly surveys to be conducted under the Council Regulation 530/1999 and the Commission Regulation 1916/2000. The latter Regulation is attached at Annex 1, as there will be frequent references to it.

The objective of this legislation is to provide accurate and harmonised data on earnings in EU Member States and Candidate Countries for policy-making and research purposes. The 2002 SES will give detailed and comparable information on relationships between the level of remuneration, individual characteristics of employees (sex, age, occupation, length of service, highest educational level attained, etc) and their employer (economic activity, size and location of the enterprise).

The SES obtains the earnings actually received by an employee of a business in the reference month and year. The information collected relates to the earnings paid to each "job holder". It does not cover earnings by the same employee elsewhere in a second or third job.

1.1.2. Data sources and reference period

The collection of data for the 2002 SES can be obtained from “tailor-made” questionnaires, existing surveys, administrative data or a combination of such sources, which provide the equivalent information. While accepting a degree of flexibility in the means employed for collecting the survey data, the information obtained must be of acceptable quality and be comparable between European countries.

The reference year is 2002. For most countries, the financial year corresponds to the calendar year. In some countries, however, the accounting year does not necessarily coincide with the calendar year and therefore for these countries the financial year which gives the best match with the calendar year 2002 should be used. Exceptionally, Germany was given a derogation to carry out their SES for the reference year 2001.

The reference month is October for the majority of the countries, this being the month which is least affected by absences owing to annual leave or public holidays. The choice of another month is acceptable if the month can be justified as being representative.

1.2. Sampling design: a two-stage sample

The 2002 SES is based on samples of employees drawn from a stratified sample of local units. For each individual employee, their characteristics (e.g. sex, age, occupation, etc) and remuneration received (hourly, monthly & annual earnings, etc) is collected, together with information about the local unit (e.g. region, economic activity, etc). The collection of micro-data at the local unit level instead of enterprise level is necessary in order to provide results at the NUTS 1 level. For enterprises without decentralised activities, the distinction between local unit and enterprise is superfluous.

The reporting unit (the local unit or the enterprise) fills out a questionnaire and provides the information collected. When the enterprise reports, it gives information concerning the local unit where the individual employee works.

The sampling procedure used for the SES contains two stages. In the first stage, a stratified random sample of local units is drawn. Stratification criteria used by the countries include economic activity (at the 2-digit level of NACE Rev.1), the number of employees in the enterprise to which the local unit belongs and region (at the NUTS 1 level). For many countries, the latter coincides with the national level. For the second stage, a simple random sample of employees is taken within each of the selected local units[1]. Where stratification is employed, criteria used could include sex, full-time or part-time and occupation.

1.3. Scope of the survey

1.3.1. Economic activities covered

The statistics of the 2002 SES refers to enterprises with at least 10 employees in the areas of economic activity defined by sections C-K of NACE Rev.1[2]. The inclusion of sections L–O is optional for 2002, as is the inclusion of enterprises with fewer than 10 employees.

1.3.2. Population and sample of employees to be covered

The SES collects data both for the reference year 2002 and the reference month (October in most countries). The population of employees to be targeted for the SES are those employed in the observation unit in the reference month, which have an employment contract. **Specifically, the employees to be covered in the 2002 SES are those who actually received remuneration during the reference month**. For sampled employees who have period(s) of unpaid absence during the reference month, their earnings should be adjusted on to a full month's basis. Where it is not feasible to adjust their monthly earnings, then such employees should be excluded from the sample.

---

[1] Some countries use only a one-stage sample and cover all employees within the selected reporting units.

[2] While the population of enterprises to be covered concerns enterprises with 10 or more employees, the population of local units to be covered are all units belonging to enterprises with 10 or more employees. Where an enterprise has just one local unit, that unit will clearly have 10 (or more) employees. Where an enterprise has more than one local unit, some of the local units may have fewer than 10 employees.

*1.3.3. Categories of workers which are to be **included** in the SES*

- employees having a direct employment contract with the enterprise or local unit and which receive remuneration, irrespective of the type of work performed, the number of hours worked (full or part-time) and the duration of the contract (fixed or indefinite).

- apprentices and trainees with an employment contract with the reporting unit.

- seasonal or occasional workers who are working pre-defined hours on a contractual basis with the local unit or the enterprise.

- interim or temporary workers employed by/through agencies - providing the reporting unit is the agency actually employing them.

- outworkers, but only if there is an explicit agreement that they are remunerated on the basis of the amount of hours worked.

- employees on maternity leave as long as they receive remuneration from the employer.

*1.3.4. Categories of workers which are **excluded** from the SES*

- employees without an employment contract with the enterprise/local unit.

- apprentices and trainees without an employment contract with the reporting unit (and who are considered to be government employees).

- seasonal or occasional workers which are employed without pre-defined working hours.

- interim or temporary workers employed by/through agencies (unless, of course, the agency employing them is the reporting unit itself).

- persons wholly remunerated by way of fees, or commission.

- members of the board of directors of the enterprise.

- unpaid owners or directors or managers (who are paid by way of profit share or fee).

- employees of the observation unit who have been permanently working abroad (i.e. for more than one year) in an affiliated company.

- the self-employed.

- family workers.

- voluntary workers.

2. INFORMATION REQUESTED IN REGULATION 1916/2000

Annexes 1 and II of Regulation 1916/2000 list and define the variables for which data is required. The countries have asked many questions about these variables. In response, Eurostat has produced a booklet "**Answers to Questions about the 2002 SES**". This is not attached here, but is separately available.

The information presented in this chapter is broken into four sections (2.1 to 2.4), corresponding to the following four blocks of variables specified in Regulation 1916/2000:

- Information relating to the local unit to which the employees are attached
- Information relating to each employee in the sample
- Information concerning earnings, hours paid and days not worked
- Grossing-up factors

The variable numbers below correspond to those listed in Regulation 1916/2000. Variables in ***italics*** highlight the items which are ***optional***.

Mandatory variables

It is essential that Eurostat receives complete information for each employee for all mandatory variables. Eurostat cannot accept micro-data records from a country if there is missing data for any of the mandatory variables, as this affects the grossing-up factors.

Optional variables

Clearly, not all optional variables will be supplied by all countries. Eurostat will make use of the data supplied, but obviously very much prefers that data should be provided (as for mandatory variables) for all observation units or employees.

2.1. Information about the local unit to which the sampled employees are attached

Annex 2 gives the codes for the alpha-numeric variables in this section.

Geographical location of the observation unit (variable 1.1)
Each region should be classified according to the nomenclature of territorial statistical units (NUTS level 1).

Size of enterprise to which the local unit belongs (variable 1.2)
The size of the enterprise (in terms of number of employees) should be classified to one of the following bands: 1-9*, 10-49, 50-249, 250-499, 500-999, 1000 and more employees. *This first band is optional for the 2002 SES.

Principal economic activity of the observation unit (variable 1.3)
The main economic activity should be coded at the 2-digit level of NACE Rev.1 for sections C-K. NACE sections L, M, N and O are optional for the 2002 SES.

Form of economic and financial control (variable 1.4)
The financial control of the enterprise should be coded as **one** of the following:
Public Control (public ownership is more than 50%)

Private Control (private ownership is more than 50%);

Shared Control (public and private ownership are both 50%)

'Shared control' should be used only when the public and private ownership are both exactly 50%; this will happen only rarely. For further background, see the booklet "Answers to Questions about the 2002 SES".

Existence of collective pay agreements covering the majority of employees in the observation unit (variable 1.5)
This is to identify the type of pay agreement covering at least 50% of the employees in the local unit. **One** of the following should be chosen:
– National level or interconfederal agreement;

– Industry agreement;

– Agreement for individual industries in individual regions

– Enterprise agreement;

– Single observation unit (local unit) agreement.

– Other type of agreement, not covered above;

– No collective agreement exists (see the coding instructions in Annex 2)

See the booklet "Answers to Questions about the 2002 SES" for examples.

Optional: Total number of employees in the local unit (variable 1.6)
A simple head count of the total number of employees in the reference month is required, covering all employees, including apprentices/trainees.

Optional: Principal market for the enterprise's products (variable 1.7)
One of the following should be selected:
– Local or regional market;

– National market;

– European Union market* (i.e., the geographical area of the Member States)

– World market.

*In contrast, the Candidate Countries (CCs) should take the "EU market" to include the CCs as well as the Member States (see the booklet "Answers to Questions about the 2002 SES").

Optional: Size of the group of enterprises (variable 1.8)

This relates to the group of enterprises (at the world level) to which the local unit belongs. Size here means the total number of employees.

Optional: Country of residence of the entity controlling the group of enterprises (variable 1.9)

This refers to the country of residence of the legal unit that publishes the accounts for the full group of enterprises.

2.2. Information relating to each employee in the sample

As indicated in section 1.3.2, the employees to be included in the SES sample are those who actually received remuneration during the reference month. Employees who did not receive any remuneration in the reference month should be excluded. The grossing-up factors (variable 4.2) should relate to those who received remuneration in the reference month.

Annex 2 gives the codes for alphanumeric variables in this section.

Sex (variable 2.1)
This is to be coded **F** for females, and **M** for males.

Employee's age (variable 2.2)
The employee's age in the reference month should be given in complete years, eg a person aged 30 years 9 months is still **30** (ie, under 31).

Occupation (variable 2.3)
This should be coded according to the ISCO-88 (COM) classification at the two digit level - or three digits, if possible.

Apprentices and trainees with an employment contract will be classified to the occupation in which they carry out their apprenticeship or their training period. Foremen are also classified to the occupation in which they supervise.

Optional: Management position or supervisory position (variable 2.4)
A simple 'Yes' or 'No' is required here, to indicate whether or not the employee has management and supervisory activities.

Highest completed level of education and training (variable 2.5)
The information required concerns the level of general, professional or higher education, which the employee has completed. "Completed" implies the successful completion of the training and is normally (but not invariably) accompanied by an appropriate paper qualification. It is only necessary to code the highest level reached. Annex II of Regulation 1916/2000 provides definitions and codes to be used for each of specified ISCED 97 groupings.

Code 03 is the **general** code to be used for ISCED 3 (Upper Secondary Education). On an *optional* basis, for countries that are able to provide a breakdown of ISCED 3, four codes (04 to 07) are provided for this purpose.

Length of service in the enterprise (variable 2.6)
The total length of service in the reference month should be based on the number of completed years of service. Eg, total service of 5 years 10 months represents ***under*** 6 years and should be given as **5 years**.
Short periods away from work of less than 12 months (e.g. on maternity or sick leave) should be included as part of the total service. However, career breaks of 12 months or more (e.g. for study leave) should be excluded.

Is the employee full-time or part-time? (variable 2.7)
Variable 2.7 is now slightly different from Regulation 1916/2000 and will be used purely to distinguish full time (FT) from part-time (PT) employees. Apprentices can be either FT or PT. The "share of part-time" is now broader (covering FT and PT employees) and is collected by a new variable (2.7.1).

Share of a full-timer's normal hours (variable 2.7.1) This is a new variable. For a FT employee, the share is 100%. For a PT employee, the hours worked should be expressed as a % of the number of normal hours worked by a FT employee in the local unit (in a similar job to the PT employee). In most cases, this part-time % will be less than 90%**.** Please give var.2.7.1 to two dec. places.

While the actual monthly and annual earnings of PT employees are of interest, these take no account of the hours worked by part-timers. The percentages given for variable 2.7.1 will therefore be used by Eurostat to gross up monthly and annual earnings (i.e., variables 3.1 & 3.2) of PT employees on to a full-time basis – for comparison with corresponding earnings of FT employees.

Type of contract of employment (variable 2.8)
One of the following types of employment contracts should be selected:
- indefinite duration (i.e. contract duration has not been agreed in advance);
- fixed-term (i.e. the contract duration is temporary or of a fixed nature);
- apprentice/trainee* (i.e. contract of fixed duration drawn up between the employer and apprentice/trainee);
- Other (i.e. any other type of contract, not specified above).

*Variable 2.8 will be used to 'flag' apprentices/trainees.

Optional: Citizenship (variable 2.9)
Citizenship is defined as the legal nationality of each person, and a citizen is a person who is a legal national by birth or naturalisation, whether by declaration, option, marriage or other means. One of the following should be selected:
Resident with citizenship;

Resident with foreign citizenship;

A commuter from another country.

Optional: Coverage by a government scheme designed to promote employment (variable 2.10)
A simple 'Yes' or 'No' is required here, to indicate whether or not the employee is covered by a government scheme. There is no need to distinguish between the different types of government scheme featured in Annex II of Reg.1916/2000.

Optional: Total period of career breaks – in months (variable 2.11)

The total period of career breaks should be given **in months**, to the nearest month. Only career breaks of 12 months or more should be counted. If several career breaks have occurred, they should be cumulated. Short term breaks (each lasting less than 12 months) should be excluded.
Enter '0' (zero) if the employee has had no long-term careers breaks. Otherwise, enter the cumulated period of breaks in months**, eg 28 mths** (not 2 yrs 4 mths).

2.3. Information concerning earnings, hours paid and days not worked, etc

As previously indicated, the population of employees covered in the SES are those who received remuneration during the reference month. Employees without any remuneration in the reference month should be excluded. The grossing-up factors (see var.4.2 below) should relate to the number of sampled employees who received a full month's remuneration in the reference month.

Average gross hourly earnings in the representative month (variable 3.0)
As gross hourly earnings does not feature in Regulation 1916/2000, this is an **extra variable** that Eurostat is asking the countries to supply. The figure required is the average gross earnings per hour paid to the employee in the reference month. **For accuracy, please give var. 3.0 to two decimal places**.

Total gross earnings for a representative month (variable 3.1)
All employees will have received remuneration during the reference month. In most cases, the employee's total gross earnings will represent a full month's earnings, that is, the employee's gross earnings will not have been reduced by periods of unpaid absence (due to sick, maternity, study leave, etc) or because the employee joined or left the firm during the month.

Where the employee's earnings are affected by unpaid absence, then the monthly earnings should be suitably adjusted in order to provide an estimate of the employee's earnings for a full month**.**

Where it is not feasible to adjust the employee's monthly earnings so that the estimated figure corresponds to a full month's earnings, then the employee should be excluded from the sample. Where necessary, the grossing-up factors (variable 4.2) should be re-calculated so that it reflects the exclusion of such employees from the sample.

Variable 3.1 should be consistent with variable 3.4 (the number of hours paid during the reference month).

Earnings related to overtime (variable 3.1.1)
The amount of overtime earnings paid for overtime hours is required. The full overtime rates should be taken into account and not just the premium element added to the basic/normal hourly rate.

If the employee's earnings are affected by unpaid absence, then the overtime earnings should be adjusted to obtain overtime earnings for a full month**.**

Where necessary, provide a rough estimate of overtime earnings using: Adjusted 3.1.1 = Unadjusted 3.1.1 * (Adjusted 3.1/ Unadjusted 3.1). Where it is not feasible to adjust variable 3.1, then such employees should be excluded from the sample and the grossing factors (variable 4.2) should be re-calculated.

Variable 3.1.1 should be consistent with variable 3.4.1 (the number of overtime hours paid during the representative month.).

Special payments for shift work (variable 3.1.2)
These are premium payments during the reference month for shift work, night work or weekend work where these are not treated as overtime. The amount to include is the premium element or supplementary payment, not the total payment for such shift work.

Where these special payments are affected by unpaid absence, they should be adjusted to provide an estimate of the shift payments for a full month**.**

Where necessary, provide a rough estimate of payments for shift work using: Adjusted 3.1.2 = Unadjusted 3.1.2 * (Adjusted 3.1/ Unadjusted 3.1). Where it is not feasible to adjust variable 3.1, then, as indicated above, this employee should be excluded from the sample and the grossing factors re-calculated.

Total gross annual earnings in the reference year (variable 3.2*)*
The actual gross earnings for the calendar year should be supplied, not the gross annual salary featured in the employee's contract.

Data for variable 3.2 should be provided for all employees for which gross monthly earnings (variable 3.1) has been supplied. Variable 3.2 should not be supplied if variable 3.1 is not completed. This is because the reference population and grossing factors (variable 4.2) both relate to employees that have actually received remuneration in the representative month.

It does not matter if the employee's earnings do not always relate to a full year. Some employees will have periods of unpaid absence, or will have joined or left the enterprise during the year. Give the actual gross earnings in the year 2002.

Do not adjust the actual gross annual earnings. When var.3.2.1 (number of weeks to which the annual earnings relate) is less than 52 weeks, Eurostat will use var.3.2.1 to "gross up" var.3.2 (plus variables 3.2.2 and 3.2.2.1 to 3.2.2.3), providing 3.2.1 is '**x**' or more weeks. Eurostat previously indicated that '**x**' should be 40 weeks. However, evidence from the countries suggests that a threshold of 40 weeks is too high. Eurostat will review the threshold to be used when the 2002 SES data has been received.

Number of weeks to which the gross annual earnings relate (var. 3.2.1)
Variable 3.2.1 refers to the paid working time of the employee during the year and should correspond to the gross earnings supplied for variable 3.2.

Please give the **number of weeks to 2 dec. places** (eg, 365/7 = 52.14 weeks).

Because Eurostat will use var.3.2.1 to "gross up" variables 3.2, 3.2.2 and 3.2.2.1 to 3.2.2.3, it is therefore essential to complete var.3.2.1 for all employees – and to insert '52' when the employee's gross annual earnings relate to a full year.

PT employees should be treated like FT employees, irrespective of the hours worked. If a part-timer has been paid for a full year, insert '52' weeks. If another part-timer has been paid for 6 months, insert '26' weeks.

Total annual bonuses (variable 3.2.2)
The word "bonuses" is a bit too narrow. Variable 3.2.2 includes any periodic, irregular, ad-hoc and exceptional bonuses and other payments that do not feature every pay period. Typical examples are Christmas & holiday bonuses, 13th or 14th month payments, allowances for leave not taken, occasional commissions, productivity bonuses and profit-sharing premiums.

The main difference between annual earnings and monthly earnings is the inclusion of payments that do not regularly occur in each (monthly) pay period. Although variable 3.2.2 is already subsumed within total annual gross earnings, Eurostat asks for "total annual bonuses" to be separately distinguished because they account for a significant proportion of total annual earnings in a number of countries.

Total annual bonuses (variable 3.2.2) = sum of variables 3.2.2.1 to 3.2.2.3

Providing variable 3.2.1 is 'x' or more weeks[1], Eurostat will use variable 3.2.1 to "gross up" total annual bonuses.

Optional: Regular bonuses not paid at every pay period (variable 3.2.2.1)
The words "regular" and "bonuses" are not entirely satisfactory here. Variable 3.2.2.1 covers (periodic) bonuses which do not occur each pay period, such as holiday bonuses, 13th and 14th month payments, but 3.2.2.1 also covers allowances for leave not taken and occasional commissions.

Providing variable 3.2.1 is 'x' or more weeks[1], Eurostat will use variable 3.2.1 to "gross up" variable 3.2.2.1.

Optional: Annual bonuses based on productivity (variable 3.2.2.2)
This component refers to irregular bonuses paid to the employee during the year, which are linked to individual performance or piecework.

Providing variable 3.2.1 is 'x' or more weeks[1], Eurostat will use variable 3.2.1 to "gross up" variable 3.2.2.2.

Optional: Annual premium related to profit sharing (variable 3.2.2.3)
The component refers to ad hoc bonuses or other exceptional payments linked to the overall performance of the enterprise, which are made under incentive schemes.

Providing variable 3.2.1 is 'x' or more weeks[1], Eurostat will use variable 3.2.1 to "gross up" variable 3.2.2.3.

[1] Eurostat will decide which threshold of **x** weeks to be used when 2002 SES data has been received.

Optional: Employees' social security contributions and taxes paid by the employer on behalf of the employee to government authorities during the representative month (variable 3.3)

Variable 3.3 (split between variables 3.3.1 and 3.3.2) are the deductions made by the employer during the reference month. This information is requested in order to obtain net monthly earnings for each employee.

If the employee's earnings are affected by unpaid absence, then variable 3.3 should be adjusted to obtain the deductions for a full month**.**
Where necessary, provide an approximate estimate of variable 3.3, using: Adjusted 3.3 = Unadjusted 3.3 * (Adjusted 3.1/ Unadjusted 3.1). Where it is not feasible to adjust variable 3.1, then this employee should be excluded from the sample and the grossing factors (variable 4.2) should be re-calculated.

Optional: Compulsory Social Security Contributions (variable 3.3.1)

The amount relates to the employee's social security contributions for the reference month.

If the employee's earnings are affected by unpaid absence, then variable 3.3.1 should be adjusted to obtain the social security contributions for a full month**.**
Where necessary, provide an approximate estimate of variable 3.3.1 using: Adjusted 3.3.1 = Unadjusted 3.3.1 * (Adjusted 3.1/ Unadjusted 3.1). Where it is not feasible to adjust variable 3.1, then this employee should be excluded from the sample and the grossing factors (variable 4.2) re-calculated.

Optional: Taxes (variable 3.3.2)

This relates to all taxes on the employee's earnings withheld by the employer for the representative month paid by the employer on behalf of the employee to the government authorities.

If the employee's earnings are affected by unpaid absence, then variable 3.3.2 should be adjusted to obtain the taxes for a full month**.**
Where necessary, provide an approximate estimate of variable 3.3.2 using: Adjusted 3.3.2 = Unadjusted 3.3.2 * (Adjusted 3.1/ Unadjusted 3.1). Where it is not feasible to adjust variable 3.1, then this employee should be excluded from the sample and the grossing factors (variable 4.2) should be re-calculated.

Number of paid hours during the representative month (variable 3.4)

Paid hours include all normal and overtime hours worked and remunerated by the employer during the month. Hours not worked but nevertheless paid are counted as 'paid hours' (e.g. for annual leave, sick leave, public holidays, etc).

Variable 3.4 should be consistent with monthly earnings (variable 3.1). If the employee's paid hours are affected by unpaid absence, then they should be adjusted to obtain paid hours for a full month**.**
Where necessary, provide an approximate estimate of paid hours using: Adjusted 3.4 = Unadjusted 3.4 * (Adjusted 3.1/ Unadjusted 3.1). Where it is not feasible to adjust variable 3.1, then this employee should be excluded from the sample and the grossing factors (variable 4.2) should be re-calculated.

Number of overtime hours paid in the reference month (variable 3.4.1)
Overtime hours include those worked in addition to those of the normal or conventional working month.

Variable 3.4.1 should be consistent with overtime earnings (variable 3.1.1). If the employee's overtime hours are affected by unpaid absence, then they should be adjusted to obtain the paid overtime hours for a full month**.**
Where necessary, provide a rough estimate of paid overtime hours using: Adjusted 3.4.1 = Unadjusted 3.4.1 * (Adjusted 3.1.1/ Unadjusted 3.1.1). Where it is not feasible to adjust variable 3.1 or 3.1.1, then this employee should be excluded from the sample and the grossing factors (variable 4.2) re-calculated.

Annual days of absence (variable 3.5)

Var.3.5 is not defined in Reg.1916/2000, but is the sum of 3.5.1 to 3.5.3.

Eurostat previously indicated that variables 3.5.1, 3.5.2, 3.5.2.1, 3.5.2.2 & 3.5.3 would be used only when the value for var. 3.2.1 is 52 weeks. However, with the introduction of a ***new variable*** (**var.3.5.1.1**), this restriction is no longer required and these variables will be grossed up when **y** ≤ var.3.2.1< 52 weeks. Eurostat will decide which threshold of '**y**' weeks is to be used when the SES data has been received.

Annual days of holiday leave excluding days of sick leave (variable 3.5.1)
This relates to the "total number of paid annual holidays actually taken by the employee, excluding sick leave and public holidays", expressed in days.

However, Eurostat recognises that many employers are better able to provide the "annual holiday entitlement" rather than the number of holidays actually taken by the employee during the year. Therefore, Eurostat will accept figures for variable 3.5.1 based on either of the following:

Normal annual holiday entitlement had the employee worked for 52 weeks;

Total number of paid holidays actually taken by the employee during 2002.

If an employee has had period(s) of unpaid absence or joined/left the firm during the year, this will affect the figure provided for (2), even if it is unlikely to affect the figure for (1). So, in order the "gross up" the number of holidays under basis (2) when var. 3.2.1 is less than 52 weeks, it is necessary to distinguish between bases (1) and (2). This distinction will be provided by a ***new variable (var. 3.5.1.1)***, for which information is given below.

For both bases (1) and (2), leave should be expressed in '**full**' days for employees. Eg, 5 weeks' holiday entitlement for a **full-timer** represents 25 days of leave. **In contrast**, 5 weeks' holiday entitlement for a **part-timer** (who works 60% of the normal full-timer's hours) represents only 15 'full' days of leave. If the available information for a PT employee is say 5 weeks (or 25 days), then the countries are asked to calculate the correct number of 'full' days, using variable 2.7.1 (the percentage share of a full-timer's normal hours).

Holiday entitlement *or* Number of holidays actually taken (variable 3.5.1.1)
This is a ***new*** variable requested by countries, in order to distinguish between:
(1) Normal annual holiday entitlement (**E**); and
(2) The number of paid holidays actually taken (**T**) by the employee in 2002.

The following table gives illustrative figures on bases (1) and (2), for a FT and for a PT employee. To avoid complications connected with the grossing-up of figures on basis (2) when var.3.2.1 is less than 52 weeks, the figures in the table assume that the FT and PT employees both have earnings for 52 weeks:

Illustrative values for variable 3.5.1 and the coding for variable 3.5.1.1

Basis 1) or 2) ?	FT or *PT employee ?*	Number of days of holiday leave (var. 3.5.1)	Coding for var. 3.5.1.1
1 = normal annual entitlement	**FT**	25 full days	**E**
2 = no. of days actually taken	**FT**	27 full days	**T**
1 = normal annual entitlement	***PT***	15 full days	**E**
2 = no. of days actually taken	***PT***	14 full days	**T**

Eurostat much prefers that each country uses the same code (either E or T) for all microdata records. However, where it is clear some reporting units have supplied the employee's entitlement (E) while other units have provided the days actually taken (T), Eurostat will accept either code.

Optional: Annual days of sick leave (variable 3.5.2)
This refers to the actual number of days of sick leave taken by the employee during 2002. Variable 3.5.2 is the sum of variables 3.5.2.1 and 3.5.2.2

Optional: Annual days of sick leave paid by the employer (variable 3.5.2.1)
This refers to the actual number of days of sick leave taken by the employee and paid totally or partially by the employer.

Optional: Annual days of sick leave not paid by the employer (var. 3.5.2.2)
This refers to the actual number of days of sick leave taken by the employee, but not paid by the employer.

Optional: Annual days of vocational training (variable 3.5.3)
This is the actual number of days of vocational training spent by the employee on all vocational training activities, expressed in full day equivalents. Vocational training of apprentices/trainees can be ignored because training is an integral part of their working time.

Optional: Annual estimation for payments in kind (variable 3.6)
Estimation of the value of all payments in kind given to the employee during the reference year.

2.4. Grossing-up factors

Grossing-up factor for the local unit (variable 4.1)

Within each sampling stratum, the grossing factors for each local unit are calculated as follows:

(Number of local units in the population) / (Number of local units in the sample)

For accuracy, please give var. 4.1 to two decimal places.

Grossing-up factor for employees (variable 4.2)

The grossing-up factor for employees is calculated as follows:

(Number of employees in the population) / (Number of employees in the sample)

For each local unit, the grossing-up factor for employees is based on:

*(Variable 4.1) * (Total no. of employees in the local unit / No. of employees in the sample)*

For accuracy, please give var. 4.2 to two decimal places.

As previously indicated, it is essential that the denominator ("*No. of employees in the sample*") corresponds to those employees that have received a full month's remuneration in the representative month.

Where it is necessary to exclude some of the sampled employees that have periods of unpaid absence(s) in the reference month, then the grossing factors supplied should be re-calculated to take account of the employees that have been excluded.

In general, whenever it is decided that the micro-data for an individual business or individual employee should be withdrawn (for whatever reason), then the grossing factors should be re-calculated by the country concerned.

Mandatory variables
Complete information must be supplied for all mandatory variables on all micro-data records. Otherwise, the grossing factors supplied will not be suitable for all variables.

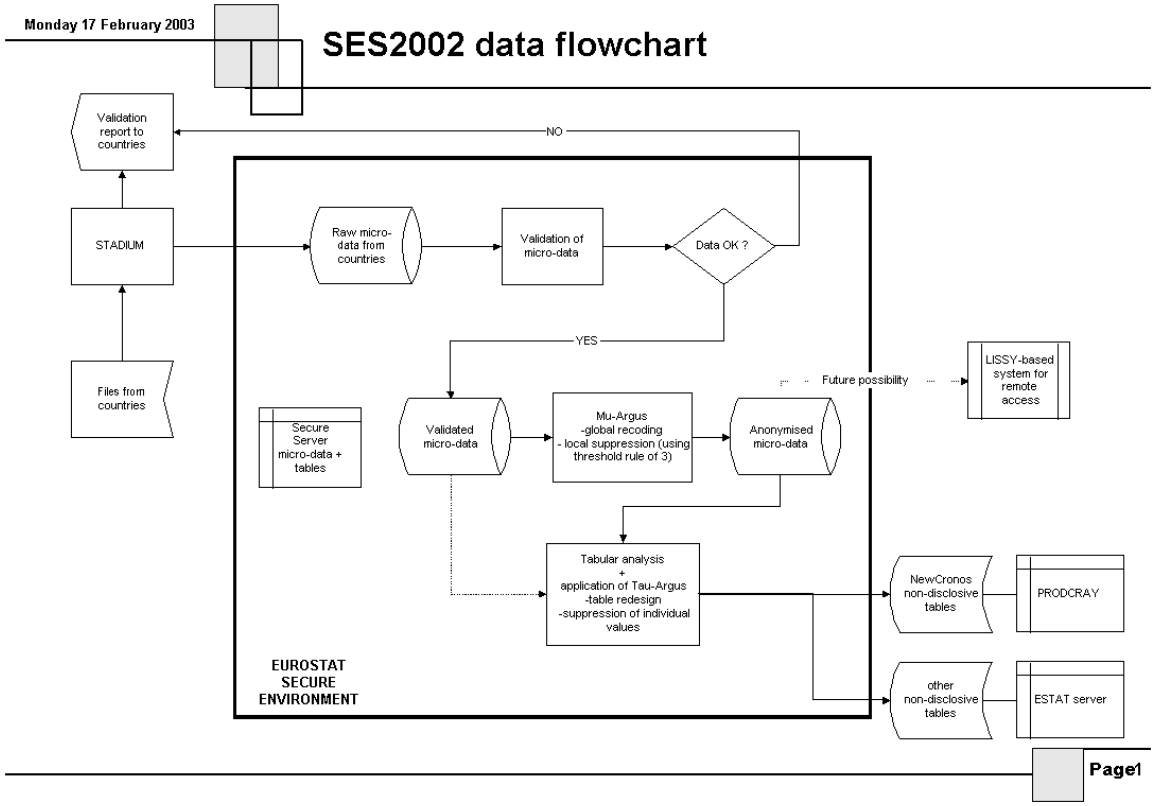
Optional variables
For those optional variables that a country chooses to supply, Eurostat will make use of all the data supplied, but obviously much prefers that data should be provided (as for mandatory variables) for all observation units or employees.

3. PROCESSING OF MICRO-DATA

3.1. Flowchart

This flowchart shows Eurostat's plans for handling the 2002 SES data from the initial receipt of the microdata, right through to the final dissemination of the tabular results. Data protection is a key integral part of the systematic approach that is being programmed.

In order to guarantee confidentiality the raw SES micro-data will be transmitted through Stadium only and the micro-data will then remain in a secure Eurostat database. Only one person has access to this data – and the micro-data will never leave Eurostat. Eurostat will only disseminate tabular SES data via New Cronos or publications.



3.2. Data Coding

Section 3.4 presents the technical format to be used for the transmission of the 2002 SES variables. The **codes** to be used for alphanumeric variables are given in **Annex 2**. "Alphanumeric" variables include: (i) variables which are coded only with letters (eg M, F, FT, PT etc) and (ii) variables where the codes contain a combination of letters and numbers (eg DE7, E50-249, R34).

Two general points about coding should be made:

- For each variable, there should be an entry on every micro-data record. '**Blank**' information for any variable will **not** be accepted.

- For optional variables for which data are not available, insert '**OPT**' for alphanumeric variables and '**99999999**' for numeric variables, the number of nines corresponding to the length of the field.

3.3. Data Validation and Eurostat Calculations

3.3.1. Introduction

Section 3.3 deals with the following issues:

- The **data validation** that will be undertaken by Eurostat on receipt of the 2002 SES micro-data from each country.

- The **adjustments, grossing-up and other calculations** that Eurostat will undertake on the SES micro-data.

Data validation

This consists of Global checks and Plausibility checks.

'Global checks' are necessary to ensure that complete data is received on micro-data records from each country. Missing entries/values for individual variables cannot be accepted because this affects the grossing up factors and the population estimates. The global checks are presented in section 3.3.2.

'Plausibility checks' on each variable are needed to ensure that the data are reasonable and consistent with other SES variables (see section 3.3.3)

These global and plausibility checks by Eurostat will automatically generate an ***output report*** on the micro-data received from each country. These output reports will be transmitted to the country, indicating where there is a data problem and the action required by the country.

Before transmitting the micro-data to Eurostat, each country should therefore carry out the same global and plausibility checks as listed here. This is essential in order to avoid returning the micro-data to a country for correction and revisions to the grossing factors (variables 4.1 and 4.2). These checks will also help to ensure that the treatment and quality of the data is harmonised across the countries.

Data adjustments, grossing up and other calculations

This relates to calculations that Eurostat will undertake, for example, to convert data for part-time employees to full-time units, or to convert data for the reference year on to an annual basis in those cases where the employee has worked for less than 52 weeks. These calculations are specified in section 3.3.4.

3.3.2. Global checks

It is necessary to distinguish between mandatory and optional variables:

Completeness of information for each mandatory variable:

For each country, all micro-data records should contain data for mandatory variables. Missing data or codes will not be accepted. The mandatory variables are: 1.1, 1.2, 1.3, 1.4, 1.5, 2.1, 2.2, 2.3, 2.5, 2.6, 2.7, 2.7.1, 2.8, 3.0, 3.1, 3.1.1, 3.1.2, 3.2, 3.2.1, 3.2.2, 3.4, 3.4.1, 3.5, 3.5.1, 3.5.1.1, 4.1 and 4.2. Please note that variable 2.7 has changed slightly and 2.7.1, 3.0 and 3.5.1.1 are new variables that do not feature in Regulation 1916/2000.

Completeness of information for each optional variable:

Each country will decide which optional variables it is able to supply. Eurostat will make use of the data supplied, but clearly very much prefers that all micro-data records should contain data for that optional variable. The optional variables are: 1.6, 1.7, 1.8, 1.9, 2.4, 2.9, 2.10, 2.11, 3.2.2.1, 3.2.2.2, 3.2.2.3, 3.3, 3.3.1, 3.3.2, 3.5.2, 3.5.2.1, 3.5.2.2, 3.5.3 and 3.6

3.3.3. Plausibility checks

Apart from checking that the correct codes have been used for the variables, the following plausibility checks are planned for the variables listed below: Each country will undoubtedly have many more checks and more demanding checks than listed below. Each country is encouraged to use its own checks, providing that the resulting SES microdata satisfies the checks below.

Optional: Total number of employees in the local unit (variable 1.6)

- If data is available, the following plausibility checks will be carried out:

Var.1.6 ≥ 1
If var.1.2 = E1_9, then var.1.6 ≤ 9
If var.1.2 = E10_49, then var.1.6 ≤ 49
If var.1.2 = E50_249, then var.1.6 ≤ 249
If var.1.2 = E250_499, then var.1.6 ≤ 499
If var.1.2 = E500_999, then var.1.6 ≤ 999
If var.1.2 = E1000, then var.1.6 ≤ 1000, or var.1.6 ≥ 1000

Optional: Size of the group of enterprises (variable 1.8)

- If data is available, the following plausibility check will be carried out:

Var.1.8 > 1 If var.1.2 = E1_9, then var.1.8 > 1 If var.1.2 = E10_49, then var.1.8 > 10 If var.1.2 = E50_249, then var.1.8 > 50 If var.1.2 = E250_499, then var.1.8 > 250 If var.1.2 = E500_999, then var.1.8 > 500 If var.1.2 = E1000, then var.1.8 > 1000

Age (variable 2.2)

14 **years** ≤ var.2.2 ≤ 80 **years**

Length of service in the enterprise in the reference month (var. 2.6)

0 **years** ≤ var.2.6 ≤ 60 **years**

% Share of a full-timer's normal hours (variable 2.7.1)

if var 2.7 = FT then var 2.7.1 = 100 if var 2.7 = PT then var 2.7.1 < 100

Type of employment contract (variable 2.8)

if var.2.8 =code C (employee is an apprenctice/trainee), then 14 ≤ var.2.2 ≤ 60

Optional: Total period of career breaks – in <u>months</u> (var. 2.11)

var.2.11= 0 ***or*** var.2.11 ≥ 12 months

Gross hourly earnings (variable 3.0)

Var. 3.0 ≥ 0 If var. 3.0 > 0 then 0.90 * (var.3.1/ var.3.4) ≤ var 3.0 ≤ 1.10 * (var.3.1/ var.3.4)

Total gross earnings for a representative month (variable 3.1)

Var. 3.1 > 0 Var.3.1 > var. 3.1.1 + var.3.1.2 Var. 3.1 > var. 3.3 If var.3.0 > 0 then (0.90 * var. 3.0 * var. 3.4) ≤ var 3.1 ≤ (1.10 * var. 3.0 * var. 3.4)

Earnings related to overtime (variable 3.1.1)

Var. 3.1.1 ≥ 0
Var.3.1.1 < var. 3.1
If (var 3.4.1>0 and var.3.1.1>0), then (var.3.1.1 / var.3.4.1) ≥
(var.3.1 – var.3.1.1) / (var.3.4- var. 3.4.1) * 0.60
(Workers can receive a lower hourly rate for overtime than for normal hours.)

Special payments for shift work (variable 3.1.2)

Var. 3.1.2 ≥ 0
Var.3.1.2 < var.3.1

Total gross annual earnings in the reference year (variable 3.2)

Var. 3.2 > 0
Var.3.2 > var.3.2.2
Var.3.2 > (var.3.1 – var.3.1.1) * 0.70 * var. 3.2.1 / (4.345238)
[This last check is to ensure that annual earings are at least 'z' times the value of monthly earnings, where the term "var.3.2.1 / (4.345238)" represents the number of months, z, that the employee worked in 2002. However, several countries reported some problems with this check, so it has been modified by excluding overtime pay in the reference month and by applying of coefficient of 0.70, in order to cope with those employees with unusually high earnings in the reference month.]

Number of weeks to which the yearly earnings relate (variable 3.2.1)

0.1 < var.3.2.1 < 53

Total Annual Bonuses (variable 3.2.2)

Var.3.2.2 < var.3.2

Optional: Regular Bonuses not paid at every pay period (var. 3.2.2.1)

- If data is available, the following plausibility check will be carried out:

Var.3.2.2.1 ≤ var.3.2.2

Optional: Annual bonuses based on productivity (variable 3.2.2.2)

- If data is available, the following plausibility check will be carried out:

Var.3.2.2.2 ≤ var.3.2.2

Optional: Annual Premiums related to profit sharing (variable 3.2.2.3)

- If data is available, the following plausibility check will be carried out:

Var.3.2.2.3 ≤ var.3.2.2

Optional: Compulsory social security contributions and taxes paid by the employer on behalf of the employee (variable 3.3)

- If data for variable 3.3 is available, the following plausibility checks will be carried out:

If var. 3.3.1 and var. 3.3.2 are both available, Var.3.3 = var.3.3.1 + var.3.3.2; Otherwise, Var.3.3 ≥ 0. Var. 3.3 < Var.3.1

Optional: Compulsory social security contributions (variable 3.3.1)

- If data is available, the following plausibility check will be carried out:

Var.3.3.1 < var.3.1 – (var.3.1.1 + var.3.1.2)

Optional: Taxes (variable 3.3.2)

- If data is available, the following plausibility check will be carried out:

Var.3.3.2 < var.3.1 – (var.3.1.1 + var.3.1.2)

Number of hours paid during the representative month (variable 3.4)

Var. 3.4 > 0 0.90 * var. 3.1/ var. 3.0 ≤ var. 3.4 ≤ 1.10 * var. 3.1/ var. 3.0 If var. 2.7 = FT, then 130 hours < (var. 3.4 – var.3.4.1) < 215 hours If var.2.7 = PT, then 130 * (var.2.7.1/100)< (var. 3.4 – var.3.4.1) <215 * (var.2.7.1/100)

Number of overtime hours paid in the representative month (var. 3.4.1)

Var. 3.4.1 < Var.3.4. If Var.3.1.1 > 0 then Var.3.4.1 > 0. Var. 3.4.1 < 0.65 * (var. 3.4 – var.3.4.1). [This last check recognises that in some countries the number of overtime hours represents a very high proportion of the total hours paid (var. 3.4).]

Annual days of absence (variable 3.5)

If var. 3.5.2 & 3.5.3 are available, Var.3.5 = var.3.5.1 + var.3.5.2 + var.3.5.3 Otherwise, Var. 3.5 ≥ var. 3.5.1

Annual days of 'full' holiday leave excluding sick leave (var. 3.5.1)

If Var.2.7 = FT, then Var. 3.5.1 < 80 for NACE sections C-K, L, N & O. If Var.2.7 = PT, then Var. 3.5.1 < 80 * (var. 2.7.1/100) for sections C-K, L, N & O. Var. 3.5.1 < 120 days for NACE section M.

The employee's normal annual holiday entitlement (E) *or* the number of holidays actually taken (T) by the employee (variable 3.5.1.1)

Var. 3.5.1.1 = E *or* T

Optional: Annual days of sick leave (variable 3.5.2)

- If data is available, the following checks will be carried out:

If var. 3.5.2.1 & 3.5.2.2 are available, Var.3.5.2 = var.3.5.2.1 + var.3.5.2.2 Otherwise, Var. 3.5.2 ≥ 0

Optional: Annual days of sick leave paid by the employer (var. 3.5.2.1)

If var.3.5.2 > 0, then var. 3.5.2.1 ≥ 0

Optional: Annual days of sick leave not paid by employer (var. 3.5.2.2)

If var.3.5.2 > 0 then var.3.5.2.2 ≥ 0

Optional: Annual days of vocational training (variable 3.5.3)

Var.3.5.3 ≥ 0

Optional: Annual estimation of payments in kind (variable 3.6)

Var.3.6 < 0.20 * var.3.2

Grossing-up factor for the local unit variable (4.1)

Var. 4.1 ≥ 1

Grossing-up factor for employees (variable 4.2)

Var.4.2 ≥ 1

3.3.4. Calculations that Eurostat will undertake

This section deals with the adjustments, grossing-up and other calculations that Eurostat will undertake on the SES micro-data. These relate to calculations, for example, to convert data for part-time employees into full-time units, or to adjust data for the reference year on to an annual basis in those cases where the employee has worked for less than 52 weeks.

PT employees: conversion to full-time units

Variable 2.7 provides a simple head count of PT employees. Variable 2.7.1 will be used to convert PT employees into full time units (FTUs).

PT employees: adjusting gross monthly and annual earnings (variables 3.1 and 3.2) on to a full-time basis.

The actual monthly and annual earnings of PT employees provided by the countries are of interest and will be disseminated. Additionally, because the actual earnings take no account of the hours worked by part-timers, Eurostat will use the percentages for part-timers (given by variable 2.7.1) to gross up the gross monthly earnings (var.3.1) and gross annual earnings (var.3.2) of PT employees on to a full-time basis**.** This will afford an approximate comparison with corresponding earnings of FT employees. This grossing up procedure for PT employees will not be undertaken for other monthly or annual variables.

Annual earnings and bonuses (variables 3.2, 3.2.2 & 3.2.2.1 to 3.2.2.3): adjustments of these variables on to an annual basis where the employee has worked for less than 52 weeks

- Variables 3.2, 3.2.2, 3.2.2.1, 3.2.2.2 and 3.2.2.3 will not be used if variable 3.2.1 < **x** weeks[1]

- If **x** [1] ≤ var.3.2.1 < 53 weeks, then the above variables will be adjusted on to an annual basis. For example, for variable 3.2:

  Adjusted var. 3.2 = unadjusted var.3.2 * (52.143 / var.3.2.1)

Likewise, for variables 3.2.2, 3.2.2.1, 3.2.2.2 and 3.2.2.3.

[1] Eurostat previously indicated that '**x**' should be 40 weeks. However, evidence from the countries suggests that a threshold of 40 weeks is too high. Eurostat will review the threshold to be used when 2002 SES data has been received.

Annual days of absence (variables 3.5 and 3.5.1 to 3.5.3)

- Variables 3.5, 3.5.1, 3.5.2, 3.5.2.1, 3.5.2.2, 3.5.3 will be used by Eurostat when 52 ≤ var. 3.2.1 < 53 weeks**.

- When **y**[1] ≤ var.3.2.1 < 52 weeks**, variables 3.5, 3.5.1, 3.5.2, 3.5.2.1, 3.5.2.2, 3.5.3 will also be used (after being adjusted on to annual basis).

** Variables 3.5.1, 3.5.2.1 & 3.5.3 all relate to periods of paid absence. However, variable 3.5.2.2 relates to annual days of sick leave **not paid by the employer**. Hence, if var.3.5.2.2 > 0, Eurostat will convert the number of days into weeks (assuming 5 days = 1 week) and these weeks will be added to var. 3.2.1 before applying the above rules.

Annual estimation of payments in kind (variable 3.6)

- Variable 3.6 will be used when 52 ≤ var. 3.2.1 < 53 weeks.

- When **y**[1] ≤ var.3.2.1 < 52 weeks, variable 3.6 will also be used (after being adjusted on to annual basis).

Use of the grossing factors (variable 4.1):

The grossing factors (variable 4.1) will be applied by Eurostat to the variables on each micro-data record for the local units to obtain population estimates of the total number of local units, including breakdowns by region, NACE activity, etc.

Use of the grossing factors (variable 4.2):

Eurostat will apply variable 4.2 to the variables on microdata records for the employees to obtain population estimates of the total number of employees and their aggregate earnings (broken down by sex, age, FT/PT, etc). The grossed up number of employees will be used for weighting purposes, including the calculation of European averages. Likewise, the grossed up number of employees will be used as the denominator for the calculation of employees' average earnings (hourly, monthly, annual), average paid hours, holidays, etc.

[1] Eurostat previously indicated that use would be made of variables 3.5.1, 3.5.2, 3.5.2.1, 3.5.2.2, 3.5.3 & 3.6 only when the value for var. 3.2.1 is 52 weeks. This restriction has now been dropped. Eurostat will decide which threshold of '**y**' weeks is to be used when the SES data has been received.

3.4. Technical format and transmission of the SES micro-data

3.4.1. The requirements

The individual data concerning each local unit and each employee should be provided in the form of two types of micro-data record:

***A*: records for the local units;**

***B*: records for the employee.**

The first record of the file with local units has to contain the column names as in section 3.4.4. The first record of the file with employees has to contain the column names as in section 3.4.5.

Regulation 1916/2000 requires that the employee records have to be linked to the local unit records by a key. You are free to use this field as you like, providing, of course, that the key itself does not disclose the identity the business. This could be an artificial number, or an existing key, as long as the same key is used in both the local unit and employee records.

Additionally, in order to be able to quickly identify an employee, Eurostat is requesting a key for each employee. Again, the key itself should not identify the person.

3.4.2. Contents of records A and B

The content and sequence of the SES variables in records A and B are given below in sections 3.4.4 and 3.4.5.

The records should contain a field per variable. All records have to be either fixed-length or comma-separated and should strictly follow the numerical order of the variables. To separate the variables, **tabs** should be used. In a **fixed-length record**, numeric variables should be "right-justified" and alphanumeric variables should be "left-justified". "Alphanumeric" variables include: (i) variables which are coded only with letters (eg M, F, FT, PT, etc) and (ii) variables where the codes contain a combination of letters and numbers (eg BE1, E10-49, R12).

3.4.3. Variables

Except for the new variables 2.7.1, 3.0 and 3.5.1.1, all variables are defined in the Commission Regulation 1916/2000. Variable 2.7 is slightly changed from Reg. 1916/2000.

All items for records A and B should be completed in full. There should be entries for all individual items, including optional variables (see below).

Mandatory variables

Data should be provided for all non-optional variables in Regulation 1916/2000. Please ensure that there are no missing values.

Optional variables

These should be **coded** strictly according to the following rules:

When information **is** available for an optional variable, Eurostat clearly very much prefers that data should be provided (as for mandatory variables) for all observation units or employees.

When information is **not** available for an optional variable, **please insert for every observation unit or employee 'OPT'** for alphanumeric variables and **'99999999'** for numeric variables, the number of 9s corresponding to the length of the field.

Zero values

'0': this should strictly be used **only** for those variables where a zero value may sometimes genuinely occur (e.g. when an employee has no overtime or shift premium payments in the reference month).

Units to be used for alphanumeric and numeric variables

The contents of records A and B are given in sections 3.4.4 and 3.4.5, respectively. An '**A**' identifies all alphanumeric variables and a '**N**' identifies all numeric variables.

'N' variables ***should be expressed in absolute terms, that is, giving the numbers in full*** (and not in decimals, or in tens, hundreds, thousands, millions, etc). ***However, because of the need for precision, the values for variables 3.0, 3.2.1, 4.1 and 4.2 should be given to two decimal places.***

There is one numeric variable that is different. This is the "**share of a full-timer's normal hours**" (variable 2.7.1) which ***should be expressed as a percentage, and also given to two decimal places, e.g. 43.27***

Where **'N'** variables relate to money values (e.g. hourly, monthly annual earnings & bonuses, taxes, social security contributions, payments in kind) these ***should be expressed in units of the national currency*** of the country concerned.

3.4.4. Content of record A

Regulation Code	Column Name	Data Type	Data length for fixed length	Data Label	Mandatory (M) or Optional (O) variable
	TABLE	A	1	Table identification (A)	M
	YEAR	N	8	Identification of the reference period (eg 2002)	M
1.1	A11	A	3	Region	M
1.2	A12	A	8	Size of the enterprise to which the local unit belongs	M
1.3	A13	A	3	Economic activity	M
1.4	A14	A	1	Economic and Financial control	M
1.5	A15	A	1	Collective Pay Agreement	M
1.6	A16	N	8	Total number of employees in the local unit	O
1.7	A17	A	3	Principal market for the enterprise's products	O
1.8	A18	N	8	Size of the group of enterprises	O
1.9	A19	A	3	Country of residence controlling the group of enterprises	O
4.1	A41	N	8	Grossing-up factor for local units (to 2 decimal places)	M
	KEY_L	A	6	Key identifying the local unit	M

3.4.5. Content of record B

Regulation Code	Column Name	Data Type	Data length for fixed length	Data Label	Mandatory (M) or Optional (O) variable
	TABLE	A	1	Table identification (B)	M
	YEAR	N	8	The reference period (eg 2002)	M
	KEY_E	A	6	Key identifying the employee	M
2.1	B21	A	1	Sex	M
2.2	B22	N	8	Age (in years)	M
2.3	B23	A	3	Occupation	M
2.4	B24	A	3	Management position / supervisory position	O
2.5	B25	A	2	Highest completed level of education and training	M
2.6	B26	N	8	Length of service in enterprise (in years)	M
2.7	B27	A	2	Full-time or part-time employee	M
2.7.1 (new)	B271 (new)	N	8	% share of a full-timer's normal hours (to 2 decimal places)	M
2.8	B28	A	1	Type of employment contract	M
2.9	B29	A	3	Citizenship	O
2.10	B210	A	3	Coverage by a government scheme designed to promote employment	O
2.11	B211	N	8	Total of career break (given in months)	O
3.0 (new)	B30 (new)	N	8	Average gross hourly earnings (to 2 decimal places)	M
3.1	B31	N	8	Total gross earnings in reference month	M
3.1.1	B311	N	8	Earnings related to overtime	M
3.1.2	B312	N	8	Special payments for shift work	M
3.2	B32	N	16	Total gross annual earnings in the reference year	M
3.2.1	B321	N	8	Number of weeks to which the gross annual earnings relate (to 2 dec. places)	M
3.2.2	B322	N	16	Total Annual Bonuses	M
3.2.2.1	B3221	N	16	Regular bonuses paid not at every pay period	O
3.2.2.2	B3222	N	16	Annual bonuses based on productivity	O
3.2.2.3	B3223	N	16	Annual premiums related to profit sharing	O

Regulation Code	Column Name	Data Type	Data length for fixed length	Data Label	Mandatory (M) or Optional (O) variable
3.3	B33	N	8	Compulsory social contributions and taxes paid by the employer	O
3.3.1	B331	N	8	Compulsory social security contributions	O
3.3.2	B332	N	8	Taxes	O
3.4	B34	N	8	Number of hours paid during the representative month	M
3.4.1	B341	N	8	Number of overtime hours paid in the representative month	M
3.5	B35	N	8	Annual days of absence	M
3.5.1	B351	N	8	Annual days of holiday leave (in full days)	M
3.5.1.1 (new)	B3511 (new)	A	1	Normal annual holiday entitlement (E) *or* Days of leave actually taken (T)	M
3.5.2	B352	N	8	Annual days of sick leave	O
3.5.2.1	B3521	N	8	Annual days of sick leave paid by the employer	O
3.5.2.2	B3522	N	8	Annual days of sick leave not paid by the employer	O
3.5.3	B353	N	8	Annual days of vocational training	O
3.6	B36	N	8	Annual estimation for payment in kind	O
4.2	B42	N	8	Grossing-up factor for employees (to 2 decimal places)	M
	KEY_L	A	6	Identification key of the local unit the employee belongs to	M

3.4.6. Transmission

The 2 data files with micro-data records will be validated and transmitted using standard Eurostat transmission tools and services. These tools provide: validation of the data, generation of a standard statistical message to send the data, a guaranteed delivery of the data to Eurostat, encryption if needed, acknowledgements and notifications.

Tools are provided and supported free of charge by Eurostat.

A detailed presentation of the tools will be given in the next Working Group meeting on 7-8 October 2003.

3.5. Treatment of confidentiality

This section gives a summary of the procedures that Eurostat will follow to ensure that the 2002 SES microdata is anonymised and that the disseminated data is not disclosive.

3.5.1. Eurostat's strategy: a systematic approach to protect data confidentiality

All 15 Member States (MS) and some 12 Candidate Countries (CCs) will send Eurostat SES micro-data. Eurostat has adopted the following strategy to protect the confidentiality of the SES data

First, the raw SES micro-data that each country transmits to Eurostat should not contain any personal identifiers. The micro-data will then remain in a secure Eurostat database. Only one person has access to this data – and the micro-data will never leave Eurostat.

Secondly, Eurostat will only disseminate tabular SES data via New Cronos or publications. There will be no risk at all to the countries because the tabular data on New Cronos or in Eurostat publications will be too aggregated to permit any disclosure. However, to ensure all such tabular information is non-disclosive, appropriate software (e.g. Tau-Argus, CiF) will be applied and action taken to remove any residual risks to individual businesses or individual employees. The tabular information to be disseminated by Eurostat (via New Cronos) is presented in Section 3.6.

Thirdly, Eurostat will want to analyse the microdata for its own (i.e. Commission) purposes. The tests undertaken to date indicate that close to 100% anonymisation of the microdata can be obtained without much sacrifice. Section 3.5.2 describes the anonymisation procedures and shows, in particular, that the loss of the regional breakdown (NUTS1) was sufficient to remove any risk of identification to all but a small number of individual businesses or employees. On receipt of the 2002 SES data, tests will be undertaken to ensure that a similarly high degree of anonymisation is achieved for each country.

Fourthly, while close to 100% anonymisation of the microdata is achievable, there will inevitably be a limited number of individual businesses or employees whose identities remain at risk. To deal with these remaining cases, another 'layer' of protection will be applied to any tabular analyses produced by Eurostat. Section 3.5.3 explains how suitable software can be applied to tabular output – in order to remove these remaining risks of disclosure. As indicated above, this software will also be applied to the tabular data that Eurostat disseminates via New Cronos and publications.

To put the treatment of confidential data into context, Section 3.1 contains aflowchart showing Eurostat's plans for handling the 2002 SES data from the initial receipt of the microdata, right through to the final dissemination of the tabular results. Data protection is an integral part of the processing of the SES data.

3.5.2. Anonymisation of the 2002 SES micro-data

This section gives a summary of the proposed anonymisation procedures:

3.5.2.1. Protection of businesses: the EG's proposals

- Sample threshold rule of 3 units (cells with 1 or 2 units are considered disclosive).
- Region: restricted to the national level;
- Sacrifice: Form of economic and financial control; Total number of employees in local unit; Size of the group of enterprises; Country of residence of the entity controlling the group.

The three most important identifying variables are region (NUTS 1), size (in bands of employee numbers) and economic activity (NACE, 2 digits). The NUTS 1 variable is the most disclosive (in combination with the other two variables). A very high degree of anonymisation is achievable by simply dropping the regional variable. As NUTS 1 corresponds to the national level for the Candidate Countries and some EEA countries, anonymisation problems for these countries may be limited. But this will be tested.

It should be stressed that the loss of the regional variable (or other variables) in order to anonymise SES microdata does not mean that the NUTS 1 data is permanently sacrificed. SES tabular analyses can be produced which give a regional breakdown, but the simultaneous presence of other identifying variables in the tables (like size and economic activity) will need to be collapsed or removed if any of the cells are disclosive.

3.5.2.2. Protection of individual employees: the EG's proposals

The anonymisation of the business units simultaneously goes a very long way towards anonymising the employees sampled in those units (because the identifying variables for individuals are less likely to be disclosive in the absence of the firm's details).

To further reduce the residual risks to employees, the following anonymisation will occur:

- Sample threshold rule of 3 units (cells with 1 or 2 units are considered disclosive)
- Citizenship: sacrificed;
- Age: restricted to 6 bands (14-19, 20-29, 30-39, 40-49, 50-59, 60+);
- Occupation: restricted to 9 categories of the ISCO 88 (1-digit) classification;
- Earnings: restricted to bands of income, including top coding.

Together, these further measures are expected to reduce risks of disclosure to negligible proportions. Section 3.5.3 illustrates how any residual risks can be 'mopped-up' by applying suitable software to the tabular data produced from the anonymised microdata.

3.5.3. Protecting SES tabular data : another layer of protection

While the global recoding procedures proposed above for businesses and employees achieve close to 100% anonymisation, there will inevitably be some businesses/employees which can still be identified. Another layer of protection will be applied whenever analytical tables are produced from the anonymised microdata files for each country. Any outstanding risks of disclosure will be identified and eliminated using suitable software.

Software will be applied to tabular analyses produced from anonymised microdata files for each country. A significant part of these tabular analyses will be those disseminated via New Cronos and publications (e.g. Statistics in Focus articles). Because it is not feasible to present the same detail as is available in the anonymised microdata files, the tables disseminated via New Cronos will inevitably show **restricted** breakdowns for individual variables. Most variables (e.g. age, occupation, length of service, earnings, etc) will be displayed in bands. The proposed New Cronos tables are presented in Section 3.6.

3.6. Tables

3.6.1. Dissemination of the 2002 SES results

The 2002 SES results will be disseminated in the following ways:

- **New Cronos**: Detailed results will be available via the Eurostat New Cronos database in the form of multidimensional tables. There will be links to information about the methodology.

- **Statistics in Focus articles**: SiF articles will present summaries of the main results for the EEA and Candidate Countries.

- **Further publications**: Other dissemination of the 2002 SES results is envisaged via publications like the Eurostat Yearbook.

3.6.2. New Cronos tables: introductory notes about the attached tables

Section 3.6.4 gives a summary of the planned 2002 SES tabular analyses that will be accessible via New Cronos. It is stressed that these analyses are provisional. Eurostat is keen to get feedback from the countries about these proposed tables (see section 3.6.3).

The tables may also be revised later, following the tests that Eurostat will undertake to protect the confidentiality of the 2002 SES data. Section 3.5 sets out Eurostat's proposals to avoid any risk of disclosure to individual businesses or employees. There is clearly also a physical limit on the number of variables that can be simultaneously extracted from New Cronos and reasonably viewed in tabular form.

Each row represents a table (see the table numbers in the final column). The crosses ("x") in each row identify the variables available for analysis. This is the maximum number of variables that it is feasible to present in one New Cronos table; other analyses can be produced using any combination of the variables specified in each row. The user is free to decide which variables he/she wants.

The two variables "Country/region" and "Sex" feature in every table but of course, individual regions or countries can be aggregated. Likewise, males and females can be combined, if desired.

In section 3.6.4, the first block of tables - "**1. Numbers of Employees**" – provides analyses of the population of employees broken down by characteristics of the observation unit and employee. This first block of tables also includes breakdowns of total employees for specified bands of hours paid, annual holidays and of hourly/monthly/annual earnings.

The second block of tables - "**2. Gross Earnings, Paid Hours and Annual Days of Leave**" - shows the proposed tables on hourly, monthly and annual earnings, monthly paid hours and annual days of leave. These analyses will provide measures of location (mean and median values), together with measures of dispersion (lower & upper quartiles, 1st & 9th deciles and the coefficient of variation).

3.6.3. Information about the variables in the attached New Cronos tables

The variable numbers in the attached tables correspond to those in Regulation 1916/2000.

The 'dependent' variables - hourly, monthly & annual earnings, monthly hours paid and annual days of leave - feature in the first column of the two attached tables. The 'independent' variables are variables 1.1 to 1.5 and 2.1 to 2.8; these explain differences in the levels of the dependent variables.

Geographical location of the observation unit (variable 1.1)
As a minimum, this variable will always provide country data plus aggregates for the EU Member States (MS) and Candidate Countries (CCs). Subject to confidentiality considerations, variable 1.1 will additionally provide a regional breakdown (at the NUTS 1 level) for 11 MS. (For the other 4 MS and all the CCs, the NUTS 1 level corresponds to the national level).

Size of the enterprise - in terms of the number of employees (var. 1.2)
Variable 1.2 has six bands: 1-9, 10-49, 50-249, 250-499, 500-999, 1000 +

Economic activity for NACE Rev.1 sections (variable 1.3)
This is available at the 2-digit level. But for the New Cronos tables we propose to limit the breakdown to the 1-digit level, ie the individual NACE sections C, D, …, K, plus the optional sections L, M, N, O. The NACE aggregates C-F, G-K, L-O, C-K and C-O will also be available.

Economic control (variable 1.4)
Three categories: public, private and shared control.

Collective pay agreement (variable 1.5)
There are 7 categories of agreements: National level or interconfederal agreement; Industry agreement; Agreement for individual industries in individual regions; Enterprise agreement; Single observation unit (local unit) agreement; Other type of agreement, not covered above; No collective agreement exists.

Sex (variable 2.1)
Male, female plus total employees.

Age in years (variable 2.2)
Six bands: under 20, 20-29, 30-39, 40-49, 50-59, 60 years and over.

Occupation - ISCO 88 (COM) at the 1-digit level (variable 2.3)
Under Regulation 1916/2000, as a minimum, occupation is available at the 2-digit level of ISCO. However, for the New Cronos tables we propose to limit the breakdown to the 1-digit level.

Highest level of education achieved, ISCED 97 (variable 2.5)
We propose to offer a breakdown of seven ISCED levels: 0 & 1 combined, 2, 3, 4, 5B, 5A and 6.

Length of service in years (variable 2.6)
8 bands are proposed: under 1 year, 1-5, 6-9, 10-14, 15-19, 20-29, 30-39 and 40 or more years.

FT or PT (variable 2.7) together with variable 2.7.1 (% share of a normal full-timer's hours)
Variable 2.7 distinguishes "FT" and "PT" employees. Using a simple head count, the total number of employees = FT + PT. Variable 2.7.1 (which is a new variable) enables part-timers to be converted into full-time units (FTUs). Variable 2.7.1 also permits the monthly and annual earnings of PT employees to be adjusted on to a full time basis. These adjustments will be made only for variables 3.1 (gross monthly earnings) and 3.2 (gross annual earnings).

Employment contract (variable 2.8)
There are 4 contract types: indefinite duration, fixed term, apprentice/trainee, 'other' contract types. Variable 2.8 will be used to separately identify apprentices.

Gross hourly earnings (new variable 3.0)
Variable 3.0 is the average gross earnings per hour paid in the representative month. There are major differences between countries, so three sets of hourly earnings (in euro) are proposed, each containing 11 bands (see table 9):

- Under 5.00, 5.00-9.99, 10.00-14.99, …,40.00-44.99, 45.00-49.99, 50.00 +
- Under 2.50, 2.50-4.99, 5.00-7.49, …, 20.00-22.49, 22.50–24.99, 25.00 +
- Under 1.00, 1.00-1.99, 2.00-2.99, … , 8.00-8.99, 9.00-9.99, 10.00 +

Depending on the country, one of the three sets of hourly earnings would be selected. As the three sets are consistent with each other, comparisons are possible between the countries. Eurostat would like to know if the countries are content with these bands - or would prefer a reduced number of bands.

Total gross earnings for the representative month (variable 3.1)
Gross monthly earnings differ between FT and PT employees, the latter working shorter hours. The following three sets of 15 bands of monthly earnings (in euro) are proposed (see table 10):

- Under 500, 500-999, 1000-1499, … , 6000-6499, 6500-6999, 7000 +
- Under 250, 250-499, 500-749, … , 3000–3249, 3250–3499, 3500 +

- Under 100, 100-199, 200-299, … , 1200-1299, 1300–1399, 1400 +

Eurostat would like to know if the countries are content - or would prefer eg a smaller number of bands.

Total gross annual earnings in the reference year (variable 3.2)
All tables for gross annual gross earnings will be programmed to allow for two separate distributions: (a) data for employees with annual earnings relating to 52 weeks and (b) data for employees with 'estimated' annual earnings (ie based on annual earnings of '**x**' or more weeks[1], the grossing up being done by Eurostat, using variable 3.2.1). The decision about the separation or non-separation of these two distributions will be made after an analysis of the 2002 SES annual data.

The following three sets of 15 bands of gross annual earnings (in euro) are proposed (see table 11):

- Under 5000, 5000-9999, 10000-14999, …, 65000-69999, 70000 +
- Under 2500, 2500-4999, 5000-7499, … , 32500-34999, 35000+
- Under 1000, 1000-1999, 2000-2999, … ,13000-13999, 14000+

Eurostat would like to know if the countries are content.

Number of hours paid during the representative month (variable 3.4)
The number of hours paid differs between FT and PT employees. The following 15 bands of monthly number of hours paid represent an attempt to cover the hours worked by PT as well as FT employees (see table 7) :

- Less than 50 hours, 50-59, 60-69, 70–79, ….,140-149, 150-159, 160-169, 170- 179, 180 or more hours.

Annual days of holiday leave (variable 3.5.1)
Again, the annual number of days of leave differs between FT and PT employees. The following bands of annual holiday leave are proposed (see table 8) :

- Less than 5 days, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35 or more.

With the introduction of the new variable (3.5.1.1), annual days of holiday leave (3.5.1) will be programmed to allow for two separate distributions relating to: (a) the normal holiday **entitlement,** and (b) the number of days actually **taken**. The decision about whether to combine these two distributions will be made after an analysis of the 2002 SES annual data.

[1] As indicated earlier (in section 2.2), variable 3.2 will be grossed up when variable 3.2.1 shows that the employee has worked for less than 52 weeks. Eurostat will decide later which threshold should be used.

3.6.4. Planned 2002 SES tabular analyses

This section gives an overview of the proposed NewCronos tables.

2002 SES results: the proposed New Cronos tables 25 July 2003

1. NUMBERS OF EMPLOYEES

Number of employees broken down by variables related to the local unit and the employee (Tables 1-6). Additionally, a further breakdown of total employees is proposed into bands of hours paid, or annual leave, or earnings (see Tables 7-11)	**Variables related to the local unit**					**Variables related to employees**							Table no
	1.1 Country / Region (NUTS 1)	**1.2** Size of the enterprise (6 bands)	**1.3** Economic activity (NACE Sections)	**1.4** Economic control	**1.5** Collective pay agreement (7 categories)	**2.1** Sex	**2.2** Age (6 bands)	**2.3** Occupat. ISCO-88 (COM) (I digit)	**2.5** Level of educ., ISCED 97 (7 levels)	**2.6** Length of service (8 bands)	**2.7** FT/ PT	**2.8** Employmt contract (4 types)	
	x		x		x	x					x	x	1
	x		x			x	x				x		2
	x	x				x		x			x		3
	x		x			x			x		x		4
	x		x			x				x	x		5
	x		x	x		x					x		6
by 15 bands of **monthly hours paid**	x		x			x					x		7
by 8 bands of **annual leave days**	x		x			x					x		8
by 15 bands of **hourly earnings**	x		x			x		x			x		9
by 15 bands of **monthly earnings**	x		x			x		x			x		10
by 15 bands of **annual earnings**	x		x			x		x			x		11

2. GROSS EARNINGS, PAID HOURS and ANNUAL DAYS OF LEAVE

Gross earnings, monthly paid hours and annual days of leave	Variables related to the observation unit					Variables related to employees							Table no
	1.1 Country / Region (NUTS 1)	**1.2** Size of the enterprise (6 bands)	**1.3** Economic activity (NACE Sections)	**1.4** Economic control	**1.5** Collective pay agreement (7 categories)	**2.1** Sex	**2.2** Age (6 bands)	**2.3** Occup. ISCO-88 (COM) (1 digit)	**2.5** Level of educ., ISCED 97 (7 levels)	**2.6** Length of service (8 bands)	**2.7** FT/ PT	**2.8** Employmt contract (4 types)	
***Mean hourly earnings (var.3.0)** and hourly overtime pay (var. 3.1.1 / 3.4.1), the latter being expressed in absolute terms and also as % of mean hourly earnings); mean hourly earnings of women as % of those of men (including and excluding overtime payments).	x		x		x	x					x		12
	x		x			x	x				x		13
	x					x	x	x			x		14
	x		x			x		x			x	x	15
	x		x			x			x		x		16
	x		x			x				x	x		17
	x	x				x		x			x		18
***Mean monthly earnings (var.3.1)** and overtime & shift pay (variables 3.1.1 and 3.1.2), the latter two being expressed in absolute terms and also as % of variable 3.1.	x		x		x	x					x		19
	x		x			x	x				x		20
	x					x	x	x			x		21
	x		x			x		x			x	x	22
	x		x			x			x		x		23
	x		x			x				x	x		24
	x	x				x		x			x		25
***Mean annual earnings (var.3.2)** and total annual bonuses (variable 3.2.2), the latter being expressed in absolute terms and also as % of variable 3.2.	x		x		x	x					x		26
	x		x			x	x				x		27
	x					x	x	x			x		28
	x		x			x		x			x	x	29
	x		x			x			x		x		30
	x		x			x				x	x		31
	x	x				x		x			X		32
*__Mean monthly hours paid (var.3.4)__ plus overtime hours (var.3.4.1), expressed in absolute terms and as % of var. 3.4. and * **Mean annual holidays** (var. 3.5.1)	x		x		x	x					x		33
	x		x			x	x				x		34
	x					x	x	x			x		35
	x		x			x		x			x	x	36
	x		x			x			x		x		37
	x		x			x				x	x		38
	x	x				x		x			X		39

* Together with the corresponding medians, 1st & 9th deciles, lower & upper quartiles and coefficients of variation.

4. ANNEXES

4.1. Annex 1: Commision Regulation 1916/2000

Annex 1 is not attached here, but is provided as a separate file.

4.2. Annex 2: Codes for Alphanumeric Variables

Annex 2 is not attached here but is provided as a separate file.