



Methodological Manual for **statistics** on the **Information Society**

Survey year 2008, v1.1

Methodological manual for statistics on the Information Society

implementation of Art. 5 of Regulation (EC) No 808/2004 of the European Parliament and of the Council of 21 April 2004 concerning Community statistics on the information society (OJ L143, 30.04.2004, p. 49)

Published by Eurostat

Survey year 2008

Version 1.1, October 2008

This publication is an update of the 2007 methodological manual and has been prepared by Unit F6 "Information Society Statistics and Services" of Eurostat. The opinions expressed are those of the authors alone and do not necessarily reflect the position of the European Commission.

Contact

Part I: Hendrikus Storm (Eurostat, Unit F6),
+352 4301 34838
hendrikus.storm@ec.europa.eu

Part II: Heidi Seybert (Eurostat, Unit F6),
+352 4301 37416
Heidi.Seybert@ec.europa.eu

Reproduction is authorised, except for commercial purposes, provided the source is acknowledged.

I.2.4.	Survey type	84
I.2.5.	Questionnaire – Data collection tool	87
I.2.6.	Quality control systems.....	88
I.2.7.	Data processing	90
I.2.7.1	Misclassification treatment	90
I.2.7.2	Non-response treatment.....	91
I.2.7.2.1.	<i>Unit non-response</i>	92
I.2.7.2.2.	<i>Item non-response</i>	94
I.2.7.3	Computing accuracy measures	95
I.2.8.	Survey execution reports	96
I.3.	Annexes.....	97
I.3.1.	Model questionnaire	97
I.3.2.	Transmission format.....	102
I.3.3.	Reporting template.....	103
Part II	Household survey	105
II.1.	Statistical product	107
II.1.1.	Statistical unit	107
II.1.2.	Target population.....	108
II.1.3.	Periodicity.....	109
II.1.4.	Observation variables	109
II.1.5.	Summary measures, aggregated variables, indicators and tabulation 109	
II.1.6.	Explanatory notes	110
II.1.6.1	Module A: Access to Information and Communication Technologies.	111
II.1.6.2	Module B: Use of computers, location and frequency of use	116
II.1.6.3	Module C: Use of the Internet.....	118
II.1.6.4	Module D: Use of Advanced Services	124
II.1.6.5	Module E: Use of e-Commerce.....	134
II.1.6.6	Module F: Socio-demographic background variables	137
II.2.	Production methodology	150
II.2.1.	Timetable – Survey period.....	150
II.2.2.	Frame population	152
II.2.3.	Sampling design	156
II.2.3.1	Stratification	165

II.2.3.2	Sample size	165
II.2.3.3	Weighting – Grossing up methods	167
II.2.4.	Survey type	167
II.2.5.	Questionnaire – Data collection tool	171
II.2.6.	Quality control systems.....	171
II.2.7.	Data processing	173
II.2.7.1	Misclassification treatment	173
II.2.7.2	Non-response treatment.....	174
	II.2.7.2.1 Unit non-response	175
	II.2.7.2.2 Item non-response	177
II.2.7.3	Computing accuracy measures	181
II.2.8.	Survey execution reports	182
II.3.	Annexes.....	184
II.3.1.	Model questionnaire	184
II.3.2.	Transmission format.....	195
II.3.3.	Reporting template.....	196
Part III	EU Legislation	198
III.1.	Regulation (EC) No 808/2004 of the European Parliament and of the Council of 21 April 2004 concerning Community statistics on the information society	200
III.2.	Commission Regulation (Ec) No 847/2007 of 18 July 2007 implementing Regulation (Ec) No 808/2004 of the European Parliament and of the Council concerning Community statistics on the Information Society.....	201

Preface

Information and Communications Technologies (ICT) account for a significant part of the EU productivity and growth and are transforming our societies and economies in a profound and unprecedented way. Official statistics are indispensable for an informed understanding of the implications of the transformations underway. However, the right set of variables is crucial. The measurement of the Information Society by statistics on society, on business processes and productivity is an area for continuous revision and improvement.

From 2002 the European Commission established annual Information Society surveys to benchmark the ICT-driven development in enterprises and by individuals. Eurostat developed two model surveys, one on enterprises, one on individuals, in close collaboration with Member States and the OECD, which have been regularly adapted to the changing needs of users and policy makers. These model surveys are composed of model questionnaires and accompanying methodological guidelines for their implementation. In 2004 the European Parliament and the Council adopted Regulation (EC) No 808/2004 covering the above mentioned surveys. That regulation is a framework regulation: it allows flexibility to adapt the surveys to newly evolving needs by users and decision makers. Annual implementing measures such as Commission Regulation (EC) No 847/2007 for the survey year 2008 are forming the basis for the Eurostat model surveys and will ensure harmonized data for all EU-27 Member States until 2010.

The first manual, which was related to the 2006 survey, was compiled in a joint effort of Eurostat, the NSI's and the OECD. The current version is drawn up as a tool for helping national statistical institutes (NSI) to translate the two Eurostat model questionnaires into national languages and to implement the national surveys following a harmonised methodology. It is a hands-on tool with recommended guidelines and does not replace statistical handbooks. Its main focus is on explanatory notes to the variables and attributes of the model questionnaires for the year 2008. As the original manual, it does however include a set of recommendations to NSI's to be taken into account when carrying out the survey to assure the production of high quality and comparable results as foreseen in the Regulatory framework. Annexes with the model questionnaires, the transmission formats and the reporting templates as well as the legal background documents are included.

The concept of this methodological manual is to use it as a rolling document. Any proposal to the authors to improve the manual is welcome, also from interested readers outside the European Community, in view of reaching for harmonized information society statistics at global level.

Albrecht Wirthmann

Information Society Statistics

Part I

Enterprise survey

I.1. STATISTICAL PRODUCT

The statistical product is the clear and precise definition of the statistical information to be produced. It distinguishes itself from the production methodology. The production methodology is the way or method of doing, while the statistical product is its direct result. Different statistical methodologies can produce the same statistical product, being only different ways of doing the same thing. This means that as long as we guarantee that two figures concern the same statistical product, for instance for two different countries, they are comparable. This way, this distinction between the statistical product and the statistical methodology helps us to focus on those elements more important to assure comparability between the several national statistics and produce new ones at the EU level, i.e. the statistical product, while leaving to the discretion of the national statistical institutes the choice of the better statistical methodology to be applied in its own country taking into consideration its own specificities.

The elements that make up the statistical product, at an input level, are the statistical unit, the target population and the observation variables, and at the output level, the periodicity and the summary measures, aggregate variables and tabulation. Covering all the elements of the statistical product, the statistical concepts and the nomenclatures are the additional needed elements to assure harmonization and comparability of statistics.

I.1.1. Statistical unit

The statistical unit is the base type of the elements of a group (also called population) that we want to observe or analyse. The basic statistical operations of classification, aggregation and ordering are done on the statistical unit.

The choice of the statistical unit is a matter of both the data collection process (namely the operational restrictions associated to collecting data from each type of statistical unit) and the conceptual framework chosen to observe and analyse the phenomenon. The statistical unit is the bearer of statistical characteristics or attributes, which we ultimately want to measure.

There are several types of statistical units, according to its usage. An *observation unit* represents an identifiable entity, about which data can be obtained. During the collection of data, this is the unit for which data is recorded. It should be noted that this may, or may not be, the same as the reporting unit. The *reporting unit* is the unit that reports to the survey authority. It reports information for the observation unit(s). In certain cases it may be different from the observation unit. A reporting unit is a unit that supplies the data for a given survey instance.

The observation statistical unit in the "Community Survey on ICT usage and e-Commerce in Enterprises" is the enterprise, as defined in the Council Regulation (EEC) No 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community defines a list of statistical units:

"The enterprise is the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit."

The enterprise thus defined is an economic entity which can therefore, under certain circumstances, correspond to a grouping of several legal units. Some legal units, in fact, perform activities exclusively for other legal units and their existence can only be explained by administrative factors (e. g. tax reasons), without them being of any economic significance. A

large proportion of the legal units with no persons employed also belong to this category. In many cases, the activities of these legal units should be seen as ancillary activities of the parent legal unit they serve, to which they belong and to which they must be attached to form an enterprise used for economic analysis.

However, the definition of the enterprise as the appropriate statistical unit poses some limitations. Some enterprises, especially the larger ones, are composed of several local units (establishment). Because of this, a geographical breakdown of the results (although still possible using the local of the main headquarters of the enterprise) is of limited use. Nevertheless, ICT usage is not easily attributable to the different establishments of an enterprise, and for this reason the enterprise is the statistical unit adopted, also for statistical products with a geographical breakdown; use of another statistical unit – e.g. the local unit - for regional results would also increase the burden on enterprises and NSI's in an unjustified way.

I.1.2. Target population

A population is a collection of objects of the same class, which in statistical terms means a group of elements of the same statistical unit. There are two types of populations to be considered when producing statistics: the target population and the frame population.

The **target population** is the population of interest. It is defined by delimiting clearly the group of statistical elements for which some information concerning the all group we want to know. That delimitation is based on one or more attributes of the statistical unit. For example, for the enterprise, some commonly used attributes to delimit the target population are the size (e.g. number of persons employed), the economic activity and its location. An example of a target population could be "enterprises with 10 or more persons employed, classified in Division 72 of NACE, localised in the EU".

The **frame population** is an operationalisation of the target population, consisting ideally in a complete list of the elements of the target population. Although a target population can be easily defined, in practise a list of all its elements is needed for its complete or partial (in case a sample is used) observation, and that can be very difficult to obtain. That list should be complete and include only and every element of the target population only once. However, most of the times it will suffer from both under-coverage and over-coverage. Usually files of statistical elements (registers) are maintained and updated, containing lists of statistical elements and also information on some attributes, usually used for delimiting target populations. Frame populations are usually extracted from those registers. The frame population will be further explained in chapter I.2.2.

The target population of the "Community Survey on ICT usage and e-Commerce in Enterprises" is the group of enterprises delimited by the following attributes:

- **Economic activity:**

Enterprises classified in the following categories of NACE Rev. 1.1:

- Section D – "Manufacturing";
- Section F – "Construction";
- Section G – "Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods";
- Groups 55.1 and 55.2 – "Hotels" and "Camping sites and other provision of short-stay accommodation";

- Section I – “Transport, storage and communication”;
- Classes 65.12, 65.22, 66.01 and 66.03 – “Banking, financial leasing and insurances”;
- Section K – “Real estate, renting and business activities”;
- Groups 92.1 and 92.2 – “Motion picture and video activities” and “Radio and television activities”.

Optionally the target population can be extended to the following categories:

- Section E – “Electricity, gas and water supply”;
- Groups from 55.3 to 55.5 inclusive;
- Classes 67.12, 67.13 and group 67.2 – “Activities auxiliary to financial intermediation, except administration of financial markets”;
- Groups from 92.3 to 92.7 inclusive; and
- Division 93 – “Other service activities”.

The enterprises are classified in one of these categories according to their **principle** activity.

- **Enterprise size:**

Enterprises with 10 or more persons employed;

Optionally the target population can be extended to enterprises with number of persons employed between 1 and 9.

Please note that the number of persons **employed** is defined in Commission Regulation (EC) No 2700/98 and should not be confused with the number of **employees**.

- **Geographic scope:**

Enterprises located in any part of the territory of the Country.

I.1.3. Periodicity

The periodicity is annual, meaning the data is collected and compiled once per year. However, some variables can be observed with a lower frequency, e.g. variables which tend to be stable over time.

An annual survey should be a compromise or trade-off between response and collection burden and the need for relevant and recent information on a “fasting moving” study domain like the information society.

This periodicity is laid down in paragraph c) of Annex II of Regulation 808/2004 and in section 5 of Regulation 847/2007 (see also *Part III – EU Legislation*).

I.1.4. Observation variables

The attributes of the statistical unit are what in the end we want to observe and the observation variables hold numerical measures of those attributes. Attributes and observation variables shouldn't be confused. An attribute is some property of the statistical unit and each attribute may have one or more observation variables which hold numerical measures of that attribute.

For example, for the statistical unit "Enterprise" an attribute is the "Percentage of orders that were sent electronically in relation to the total purchases' value". We want to measure this attribute and can just use a variable to hold the percentage of the purchases via computer networks. But collecting the percentage of an enterprises' purchases ordered via computer networks might not be feasible. In this case we can use another variable which is feasible, like a categorical variable with 5 different ranges of percentages of purchases, which information enterprises are willing and able to give. In this case two different observation variables can be used to measure the same attribute.

There are many ways of classifying observation variables, but the most relevant one for the ICT usage surveys is the distinction between qualitative and quantitative variables.

Qualitative variables contain non-numeric information. They serve merely as labels or names for identifying special attributes of the statistical unit. An example is the "Use (Yes or No) of computers" by the enterprise. Sometimes qualitative variables can be rendered numeric by coding the non-numeric values. An important type of qualitative variables is that of binary variables. Binary variables assume only two different values, which usually are rendered numeric by attributing the values "0" and "1" in a meaningful way.

Quantitative variables contain information as to how much or how many. Hence they are always numeric. An example is the number of persons employed that used computers with access to the WWW at least once a week.

The observation variables in the "Community Survey on ICT usage and e-Commerce" are mainly qualitative, namely binary.

For the 2008 enterprise Community Survey the observation variables to be collected are the following.

Question number in model questionnaire	Variable	Optional	Core indicator
A1	Usage of Computers		
A2	Percentage of persons employed using computers at least once a week	x	
A3	Usage of an internal computer network (e.g. LAN)		*
A4	Usage of wireless access within internal computer network (e.g. wireless LAN)		*
A5	Usage of internal home page (Intranet)		
A6a	Usage of internal home page for sharing general policy or strategy of enterprise	x	
A6b	Usage of internal home page for sharing internal company newsletters or daily news	x	
A6c	Usage of internal home page for sharing day-to-day working documents (e.g. for meeting)	x	
A6d	Usage of internal home page for sharing manuals, guides or training material	x	
A6e	Usage of internal home page for sharing product or services catalogues	x	
A7	Usage of dedicated applications for employees to access human resources services		
A8	Usage of an extranet		*
A9	Usage of third party free or open source operating systems, such as Linux		*
B1	Access to Internet		
B2	Percentage of persons employed using computers connected to the WWW at least once a week		*
B3a	Connect to the Internet via modem (dial-up access over normal telephone line) or ISDN		*
B3b	Connect to the Internet via DSL		*
B3c	Connect to the Internet via another fixed connection (e.g. cable etc.)		*
B3d	Connect to the Internet via mobile connection (e.g. mobile phone, GPRS, etc)		*
B4a	Purpose of the Internet (as a consumer): banking and financial services		
B4b	Purpose of the Internet (as a consumer): training and education		
B5	Usage of Internet for interaction with public authorities		*

B6a	Usage of Internet for interaction with public authorities - for obtaining information		*
B6b	Usage of Internet for interaction with public authorities - for obtaining forms		*
B6c	Usage of Internet for interaction with public authorities - for returning filled in forms		*
B6d	Usage of Internet for interaction with public authorities – for treating administrative procedure completely electronically		*
B6e	Usage of Internet for interaction with public authorities – to submit a proposal in a public electronic tender system (e-procurement) (in the system itself and not by email)		*
B7	Have a Web Site or a Home Page		
B8a	Facilities of the Web Site (as a provider): product catalogues or price lists		
B8b	Facilities of the Web Site (as a provider): possibility for visitors to customise or design products		
B8c	Facilities of the Web Site (as a provider): online ordering, reservation or booking		
B8d	Facilities of the Web Site (as a provider): online payment		
B8e	Facilities of the Web Site (as a provider): personalised content in the Web Site for regular/repeated visitors		
B8f	Facilities of the Web Site (as a provider): advertisement of open job positions or online job applications		
B9	Usage of a digital signature in any message sent, i.e. using encryption methods that assure authenticity and integrity of the message		*
C1	Usage of automated data exchange between your enterprise and other ICT systems outside your enterprise		*
C2a	Automated data exchange used for sending orders to suppliers		*
C2b	Automated data exchange used for receiving e-invoices		*
C2c	Automated data exchange used for receiving orders from customers		*
C2d	Automated data exchange used for sending e-invoices		*
C2e	Automated data exchange used for sending or receiving product information		*
C2f	Automated data exchange used for sending or receiving transport documents		*
C2g	Automated data exchange used for sending payment instructions to financial institutions		
C2h	Automated data exchange used for sending or receiving data to / from public authorities		
C3a	Used EDIFACT or similar formats for automated data exchange	x	
C3b	Used XML or similar formats for automated data exchange	x	
C3c	Use agreed proprietary standards for automated data exchange	x	
C4a	Barrier to use automated data exchange: No interest	x	
C4b	Barrier to use automated data exchange: Lack of expertise	x	
C4c	Barrier to use automated data exchange: Return on investment too low or unclear	x	
C4d	Barrier to use automated data exchange: Lack of appropriate software	x	
C4e	Barrier to use automated data exchange: Difficult to agree common standards	x	
C4f	Barrier to use automated data exchange: Uncertainty of the legal status	x	
D1	Regularly share electronically information on the supply chain management with suppliers or customers		
D2a	Regularly share electronically information with suppliers on inventories, production plans or demand forecasts		
D2a1	Regularly share electronically information with suppliers on demand forecasts	x	
D2a2	Regularly share electronically information with suppliers on inventories	x	
D2a3	Regularly share electronically information with suppliers on production plans	x	
D2b	Regularly share electronically information with suppliers on progress of deliveries		
D3a	Regularly share electronically information with customers on inventories, production plans or demand forecasts		
D3a1	Regularly share electronically information with customers on demand forecasts	x	
D3a2	Regularly share electronically information with customers on inventories	x	
D3a3	Regularly share electronically information with customers on production plans	x	
D3b	Regularly share electronically information with customers on progress of deliveries		
D4a	Usage of Web Sites for electronic exchange of information		
D4b	Usage of automated data exchange for electronic exchange of information		
E1a	Share electronically information on sales with software used for management of inventory levels		*
E1b	Share electronically information on sales with software used for accounting		*

E1c	Share electronically information on sales with software used for production or services management		*
E1d	Share electronically information on sales with software used for distribution management		*
E2a	Share electronically information on purchases with software used for management of inventory levels		*
E2b	Share electronically information on purchases with software used for accounting		*
E3	Usage of ERP software package to share information on sales and or purchases with other internal functional areas		*
E4a	Usage of CRM to capture, store and make available to other business functions the information about its clients		*
E4b	Usage of CRM to analyse information about clients for marketing purposes		*
F1	Did receive orders via computer networks		*
F2	Percentage of total turnover resulting from orders received electronically via computer networks		*
F3a	Percentage of total turnover resulting from orders received electronically via a website	x	
F3b	Percentage of total turnover resulting from orders received electronically via automated data exchange over the internet	x	
F3c	Percentage of total turnover resulting from orders received electronically via automated data exchange over other computer networks	x	
F4	Usage of secure protocol, such as SSL or TLS, for reception of orders via Internet		*
F5	Did send orders via computer networks		*
F6	Percentage of total purchases resulting from orders sent electronically via computer networks		*
G1a	Perceived benefits of ICT projects: reorganisation and simplification of work routines	x	
G1b	Perceived benefits of ICT projects: release of resources	x	
G1c	Perceived benefits of ICT projects: higher earnings for the enterprise	x	
G1d	Perceived benefits of ICT projects: development of new products and services	x	
X1	Main economic activity of the enterprise		
X2	Average number of persons employed		
X3	Purchases of goods and services		
X4	Turnover		
X5	Location (Convergence/non-Convergence region)		

I.1.5. Summary measures, aggregated variables, indicators and tabulation

Summary measures are numerical values defined by a statistical measure that is used to summarize the values for a specific variable for all statistical units in a specific group. Such measures can take the form of aggregates (e.g. total number of yes-answers on a specific question) or indicators (e.g. percentage of yes-answers).

Aggregates can be compiled for the total population or for the different subpopulations defined by the background variables (e.g. NACE category or size class) or for subpopulations defined conditionally on the answers of other study variables (e.g. broadband users versus non broadband users).

E.g.:

- number of enterprises with Internet access
- number of enterprises with between 10 and 49 persons employed with Internet access
- number of enterprises having a LAN and either an Intranet or an Extranet

To obtain indicators (proportions, percentages, ratio, etc.), the aggregates need to be divided by the corresponding total population or subpopulation.

The denominator of such ratio can for instance be:

- total population of enterprises in the selected strata (e.g. "proportion of enterprises with between 10 and 49 persons employed with access to the Internet" = the *number of enterprises with between 10 and 49 persons employed with access to the Internet* divided by the *total number of enterprises with between 10 and 49 persons employed*)

The detailed transmission format for sending data to Eurostat can be found in 0 the *Transmission format (annex I.3.2*. In that document, aggregates are specified; on the basis of these aggregates, Eurostat calculates and releases indicators.

I.1.6. Explanatory notes

The explanatory notes in this chapter refer to the questions in the 2008 model questionnaire (see Model questionnaire in section I.3.1). The structure of this chapter follows the model questionnaire, i.e. the explanatory notes are grouped per *module* and per *question*. It is recommended to have the model questionnaire at hand while reading this section.

I.1.6.0 PRELIMINARY, GENERAL REMARK: USE, HAVE OR HAVE ANOTHER ENTERPRISE USE FOR YOU

In many of the following questions and notes on these questions, reference is made to the **use** of computers, networks, systems etc. Use never refers to ownership: computers, systems etc. might belong to the enterprise, be leased or shared with another organisation.

Enterprises sometimes outsource **ICT** functions (see for instance the Methodological Manual for survey year 2007, section I.1.6.5, question E11 for a list of ICT functions); in such cases, where computers etc. are used on behalf of the responding enterprise, these computers etc. should be considered to be used by the responding enterprise; the corresponding question on the use of computers etc. should thus be answered by "yes".

I.1.6.1 MODULE A: USE OF COMPUTERS AND COMPUTER NETWORKS

A1: Did your enterprise use computers, in January 2008?

[Scope: all enterprises]

[Type: one single answer needed, i.e. *Tick only one* ; binary (Yes/No); filter question]

A **computer** is a programmable device or machine capable of receiving, processing and outputting various types of data, performing calculations and other manipulations on it, under the control of a stored set of instructions. A computer typically consists of one or more processing units, memory units, and associated peripheral input and output devices.

Depending on size and power, computers can be generally divided in the following categories:

- **Supercomputer:** An extremely fast computer that can perform hundreds of millions of instructions per second.

- **Mainframe:** A powerful multi-user computer capable of supporting many hundreds or thousands of users simultaneously.
- **Minicomputer:** A multi-user computer capable of supporting from 10 to hundreds of users simultaneously.
- **Workstation:** A powerful, single-user computer....like a personal computer, but it has a more powerful microprocessor and a higher-quality monitor.
- **Personal computer:** A small, single-user computer based on a microprocessorhas a keyboard for entering data, a monitor for displaying information, and a storage device for saving data.

Amongst personal computers one can find:

- **Desktop computers:** A desktop computer can be defined as a computer that usually remains fixed in one place. Normally the user is placed in front of it, behind the keyboard with the monitor placed on top.
- **Portable computers (laptop):** A portable computer (or laptop) can be defined as a computer that is small enough to carry. A portable computer (or laptop) can usually carry out the same tasks as a desktop computer. Includes notebooks but not handheld computers, although portable.
- **Handheld computers (palmtop):** A handheld computer is a computer that can be used while you are holding it and stored in a small bag or for instance in a pocket. Personal Digital Assistants (PDAs) are another term for handheld computers. A PDA uses a pen rather than a keyboard.

All the above types of computers are in the scope of this question. On the other hand, electronic tills which are only used for monetary transactions (EPOS) and computer controlled machinery are not considered for this question.

A2: How many persons employed used computers at least once a week, in January 2008?

or

Please indicate an estimate of the percentage of the number of persons employed who used computers at least once a week, in January 2008.

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: numerical, in real or percentage values]

The concept of **computer** to be used in this question is the same as the one for question A1.

This question refers not to **access** only, but to effective **use** of computers by the personnel in the enterprise activities. It intends also to capture only the relevant use of computers by workers, setting a minimum level of frequency: at least once a week. Although there was in the past reference to the use of computers by persons employed in their normal work routine, this concept shouldn't be used as it would, for instance, include those which routine includes using a computer every three months when this wouldn't be a relevant use of computers by the enterprise personnel.

The concept of person employed to be used in this variable is the one described in the explanatory note of variable X2 ("Average Number of Persons Employed").

This variable can be collected in **real** or in **percentage** values. The variable in real values might be more appropriate for smaller enterprises, while the one in percentage terms indicated for bigger ones. When the number of persons employed is small it's easier for the respondent just to identify which ones work with computers, instead of having to calculate the percentages afterwards. The collection in percentage terms might even introduce measurement errors in those cases, because of the rounding that is necessary to provide an answer in percentages made of integer numbers between 1 and 100. The collection in percentage terms can also introduce other measurement errors as it induces the respondent to make an educated guess of its value introducing the bias usually associated to this type of answer. On the other hand, for bigger enterprises, with a high number of persons employed, the identification of every worker using (and not just having access) a computer can be burdensome or not possible and result in non-response. In this case the collection of this information in percentage terms inducing to an educated guess may be preferable.

Eurostat in its model questionnaire (annex I.3.1) recommends to use a combination of both, giving to the respondent the opportunity to choose to answer in real or percentage values.

Independently on how this variable is collected, in real or percentage values, the background variable "number of persons employed" is needed for grossing up the overall percentage of persons employed using computers. In order to compute the overall percentage both the total number of persons employed and the total number of persons using computers need to be grossed up. If collected in percentage terms, the corresponding real value needs to be calculated afterwards multiplying the answer with that background variable to gross up the number of persons employed using computers.

A3: Was your enterprise using an internal computer network (e.g. LAN - Local Area Network) in January 2008?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. Tick only one; binary (Yes/No); filter question]

This question refers to the use of an internal computer-mediated network within the enterprise at whatever moment during January, regardless property and purpose of its use.

An internal computer network is a group of at least two computers connected together using a telecommunication system for the purpose of communicating and sharing resources within an enterprise. It typically connects personal computers, workstations, printers, servers, and other devices. It is used usually for internal file exchange between connected users, intra business communications (internal e-mail, internal web based interface, etc), shared access to devices (printers, etc) and other applications (databases) or for joint business processes.

A LAN is a network for communication between computers confined to a single building or in closely located group of buildings, permitting users to exchange data, share a common printer or master a common computer, etc.

LAN is a proxy for internal computer network and it's used in this question as a typical example of such network. The relevant concept to be measured is internal connectivity in the enterprise. This connectivity facilitates the sharing of information among different people

in an organisation and it's a pre-condition for the automation and integration of business processes in the enterprise.

However, internal computer network is a concept which might be understood in very broad terms. Computer networking is sometimes understood as including also the linking of computers to other devices, such as telephones and personal digital assistants (PDA's), through for example USB or FireWire, and can include only two computer devices. Such networking is not very informative about the enterprise's ability to integrate its business processes and the concept used should be narrower, restricting it to the interconnection of at least two computers.

On the other hand, LANs are a basic and common way of connecting computers within a company. And larger scale computer networks, at enterprise level, are generally made up of several LAN's connected or at least one LAN to which single computers at different sites connect. For these reasons, LAN is used to proxy the existence of a reasonable level of connectivity in the enterprise.

Compared to Eurostat model questionnaires of the previous years, this question instead of referring directly and exclusively to the use of a LAN, refers to the use of an internal computer network, referring to LAN as a typical example of such network. This change occurred because some statistical institutes have the experience that sometimes LAN is not a term well known by the respondents. However, the reference to LAN is still very important and should be used because it helps to clarify the concept we want to measure for those respondents which are familiar with the term.

The use of the term "internal" in this question should not be here understood as restricting the variable to those cases where there is an access to the computer network restricted to employees. It only means that several computers used by the enterprise are interconnected.

A4: Did your enterprise use wireless access within its internal computer network (e.g. wireless LAN), in January 2008?

[Scope: enterprises which use an internal computer network, i.e. A3 = Yes]

[Type: one single answer needed, i.e. Tick only one; binary (Yes/No)]

Wireless access to the internal computer network refers to the use of wireless technologies such as radio-frequency, infrared, microwave, or other types of electromagnetic or acoustic waves, for the last internal link between users devices (such as computers, printers, etc) and the internal network backbone line(s) within the enterprise's working premises. A computer network which includes wireless access is also called a wireless network.

The backbone of the internal computer network is a line or set of lines which constitute the main wire based segment of the internal network. It usually provides high capacity communications and it's used to interconnect several dispersed sub-networks to span distances efficiently (for example, between buildings).

In a typical wireless network there are two types of wireless stations: access points and clients. Access points are fixed stations which are connected to the network backbone through cables and are prepared to transmit and receive radio frequencies or other signals so they can communicate with wireless enabled devices. Wireless clients are usually mobile devices such as laptops, personal digital assistants or IP phones, with a hardware device

which allows them to communicate via wireless with the access points and in this way to connect to the internal network.

Wireless access also includes the cases where wireless devices directly communicate with each other (also called peer-to-peer or P2P). Wireless devices, such as computers, within a certain range of each other can communicate directly without using a network backbone, and form in this way a network.

Two typical technologies used in wireless networks are Wi-fi and Bluetooth.

Wireless access to the internal computer network provides greater mobility to staff for moving around within a broad area while still being connected to the network.

A5: Did your enterprise have in use an internal homepage (Intranet), in January 2008?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. Tick only one; binary (Yes/No); filter question]

Intranet refers to the general use of internet protocols (TCP/IP and others) to share and exchange company information in the same way as over the Internet but with access restricted to internal users, (organisation's members, employees, or others with authorisation). Internet protocols commonly used are ftp (file transfer protocol), email and http (hyper text transfer protocol), this latter is used to form an internal website.

However, typically the term intranet is understood in a stricter sense as referring only to the internal website, or home page. An intranet's website looks and acts just like any other websites in the public internet. The same concepts and technologies of the World Wide Web such as web browsers and servers running on the internet protocol suite are used to build an internal website.

It was chosen in Eurostat's community survey to use the stricter meaning of intranet, i.e. an internal home page. The reason is that the use of internet protocol in LANs is very common nowadays. Therefore, the indicators on the percentage of enterprises using a LAN and percentage of enterprises using an intranet in the broader sense should give almost the same information. However, experience in the previous years show that only in a couple of countries the indicators are very close. In most countries the adoption level of intranet is much lower than that of LAN, as would be expected for the adoption of internal homepages but not for intranet. The probable reason for this difference is that respondents understand intranet as the internal home page, and not as the use of internet protocol in internal computer networks. The intranet as the internal homepage is also the most relevant concept. The technology used (internet protocol) is not important, but the function that it performs, in this case as a communication and integration tool within the enterprise.

Compared to Eurostat's model questionnaire of the previous years, this question was changed in order to give priority to the reference to internal homepage, providing still the term intranet as a parenthesis. Homepage leads respondents to think of web (html) interfaces.

Although in general the internal homepage looks like a private version of the World Wide Web, there does not necessarily have to be any access from the organisation's internal network to the internet itself.

When part of the internal homepage is made accessible to customers, partners, suppliers, or others outside the company, that part becomes an extranet.

In question A6, a – non-exhaustive - selection of specific uses of the internal homepage are measured.

A6: In January 2008, was your enterprise using the internal homepage for sharing the following information?

[Scope: enterprises which use an internal homepage, i.e. A5 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

For those enterprises using an internal homepage, this question focuses on the measurement of its specific use. This information is of particular importance, because the sole existence of the internal homepage provides only the *potential* for better communication and information sharing in the enterprise. The *impact* of the use of such tool depends on the *specific information shared* in the internal homepage.

The purpose of this question is to complement the measurement of e-business and, more concretely, the integration of business processes. In fact, the main variables for the measurement of the integration of business processes are those focusing on the automatic share of information within the enterprise in module E. However, such automation is more appropriate for those cases where there is a sequential chain of processes where the precedents trigger the subsequent ones. There are other processes which are not chained sequentially and which are integrated by making information (e.g. its results) readily available to be used whenever it is needed. This happens in several auxiliary activities inside the enterprise, but also at operational level in certain services activities where automatic sharing of information is not applicable. Therefore, this question includes several items which intend to capture the integration of business processes by making information readily available in the internal homepage.

a) The general policy or strategy of the enterprise

The purpose of this item is to measure the use of the internal homepage to share internally with the staff general purpose information about the enterprise, such as the mission, strategy, policies and overall objectives.

b) Internal company newsletters or daily news

This item refers to the use of the internal homepage as a mean to broadcast information of interest to all the staff of the enterprise. As specific references the item refers to company newsletters and daily news, either internal or external, but of particular interest for the enterprise, and which can be made available in the entry page of the intranet or on a specific page.

c) Day-to-day / working documents (e.g. for meetings)

This item includes information or documents that are made available in the internal homepage and which are needed during the normal work of the staff. These might include minutes of internal or external meetings, conclusions of specific working groups, reference material, tabulated data, etc. This item does not include material with a didactic purpose directed to support learning.

d) Manuals, guides or training material

The purpose of this item is to measure the use of the internal homepage to make learning material available to the staff. Learning material includes all information designed to support the process of acquisition of knowledge or skill by the personnel of the enterprise. This includes not only material to support training courses provided or organised by the enterprise, but also information to support self-learning. As examples of such material the item refers explicitly manuals and guides.

This item does not concern the sharing of general documents about the training strategy of the enterprise, forms to request training, training planning or sole information about the content of training courses.

e) Product or services catalogues

This item is the most specific one and the one directed to the core activity of the enterprise. It refers to lists of products or services provided by the enterprise to its clients, eventually with information on prices.

A7: In January 2008, did your enterprise use dedicated applications for employees to access human resources services (e.g. see open job positions, request annual leave, view or download payslips, or other services)?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. Tick only one; binary (Yes/No)]

The purpose of this variable is to measure the adoption of information systems for human resource management which aim to treat informational aspects concerning the relation between the enterprise and the employee.

These information systems can allow the enterprise to identify the whole of the staff, to support all the procedures of classic human resource management (recruitment, career management, job assignments, definition of statutory links and individual rights, etc.), to support the implementation of staff evaluation and promotion system among others by the formulation and publication of job descriptions, and to facilitate the matching process between jobholders and job vacancies. The systems in the scope of this question are those which implement some of the facilities just described, but not necessarily all of them.

The system can exist with a web interface in the internal homepage of the enterprise, but is not restricted to such cases. Systems based on standalone applications and not based on web pages are also included.

The question is made from the point of view of the employee in order to avoid the use of terms like human resources management which would not be understood in the same way by different respondents.

A8: In January 2008, did your enterprise have an extranet (a website or an extension of the Intranet with access restricted to business partners)?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. *Tick only one*; binary (Yes/No)]

The purpose of this question is to measure the adoption of extranets by the enterprises. An extranet is a closed network that uses Internet protocols to securely share enterprise's information with suppliers, vendors, customers or other businesses partners. It can take the form of a secure extension of an Intranet that allows external users to access some parts of the enterprise's Intranet. It can also be a private part of the enterprise's website, where business partners can navigate after being authenticated in a login page.

Online banking facilities using Internet protocols for example are generally considered as an extranet of the enterprise.

A9: Did your enterprise have in use, in January 2008, third party free or open source operating systems, such as Linux? (i.e. with its source code available, no copyright cost, and the possibility to modify and/or (re)distribute it)

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. *Tick only one*; binary (Yes/No)]

The purpose of this variable is to measure the percentage of enterprises that use open-source software at systems level in their business activity. The adoption of open source is measured in this question at the level of systems software because it seems to be more feasible to collect than at any software level, as the IT department might not be aware of the use of general open source software by users throughout the organisation.

Although "at systems level" basically means operating systems and servers, "server" is a difficult concept to use in the data collection, as it can be understood in very different ways. Therefore, only reference to "operating systems" is made in the question.

Linux is an example of system level software which is open source. The use of an example helps to clarify the question and improve the accuracy of the answer. However, it is very probable that the inclusion or non inclusion of the example in the question has a significant impact on the results. Therefore, in order to assure the comparability of the results among different countries, they all have to use the example in the question.

The reason for measuring the use of open source is on one hand because it reduces the cost of IT infrastructure of the enterprise and on the other hand because it provides interoperability. By reducing the cost of the IT infrastructure, open source is believed to promote / facilitate the adoption of e-business. For this reason, the measurement of open-source is most relevant at systems level.

Both concepts of "free software" and "open source software" are included in this measurement. Although very similar, software licenses for one of these concepts can be non compliant with the other, while the benefits just described above come from both free and open source software.

Another reason for measuring the adoption of open-source is that there is an interest of policy makers on the use of open-source as a business model, i.e. on the development of this business model as a profitable activity (selling consulting and training services on the developed open-source software, developing a more advanced version of the software which needs to be purchased, etc.). From that point of view, in order to understand the potential of the “open-source business model”, it is important to know the level of adoption of open-source. Nevertheless, the use of the information from this variable for this purpose is limited, because it would require the collection of information on the adoption of open source not only at systems level, but in general.

As the two key definitions for this measurement are those of open source and systems level software, they are explained next.

Definition of systems level software

Practical computer systems divide software into three major classes: system software, application software and programming software, although the distinction is somewhat arbitrary, and often blurred.

- **System software** helps run the computer hardware and computer system. It includes operating systems, device drivers, diagnostic tools, servers, windowing systems, utilities and more.
- **Application software** allows a user to accomplish one or more specific tasks. Typical applications include office suites, business software, educational software, databases and computer games. Most application software has a graphical user interface (GUI).
- **Programming software** usually provides some useful tools to help programmer to write computer programs and software using different programming language in a more convenient way. The tools include text editor, compiler, interpreter, linker, debugger, and so on. Integrated development environment (IDE) merges those tools in a software bundle, and programmer may not need to type a lot of commands for compiling, interpreter, debugging, tracing, and etc., because IDE mostly has a GUI.

Definition of open source software

Open source software refers to computer software under an open source license. An open-source license is a copyright license for computer software that makes the source code available under terms which allow for modification and redistribution without having to pay the original author. Such licenses may have additional restrictions such as a requirement to preserve the name of the authors and the copyright statement within the code. One popular (and sometimes considered normative) set of open source licenses are those approved by the Open Source Initiative (OSI) based on their Open Source Definition (OSD).

The **Open Source Initiative** is a non-profit organization dedicated to promoting open source software, namely through promoting its **Open Source Definition**. This Open Source Definition is used by the OSI to determine whether or not a software license can be considered open source. Under the Open Source Definition, licenses must meet ten conditions in order to be considered open source licenses:

1. Free Redistribution: the software can be freely given away or sold.
2. Source Code: the source code must either be included or freely obtainable.
3. Derived Works: redistribution of modifications must be allowed.

4. Integrity of the author's source code: licenses may require that modifications are redistributed only as patches.
5. No discrimination against persons or groups: no-one can be locked out.
6. No Discrimination against fields of endeavour: commercial users cannot be excluded.
7. Distribution of license: The rights attached to the program must apply to all to whom the program is redistributed without the need for execution of an additional license by those parties.
8. License must not be specific to a product: the program cannot be licensed only as part of a larger distribution.
9. License must not restrict other software: the license cannot insist that any other software it is distributed with must also be open source.
10. License must be technology-neutral: no click-wrap licenses or other medium-specific ways of accepting the license must be required.

It should be noted that merely opening the source code publicly does not mean that the code can be called 'Open Source', as the term is normally used as defined above. Indeed, it does not give the rights normally associated with open source software such as the freedom to adapt, test or redistribute. For example, the 'Shared Source Initiative' of Microsoft requires an agreement to be signed before gaining access to the source and furthermore the agreement does not confer any of the open source software rights.

There are also shared source licenses which have some similarities with open source, but a number of critical differences make such licenses incompatible with the Open Source Definition. Some software licenses define an open standard basis and may or may not be similar to open source, like some versions of Solaris and PGP.

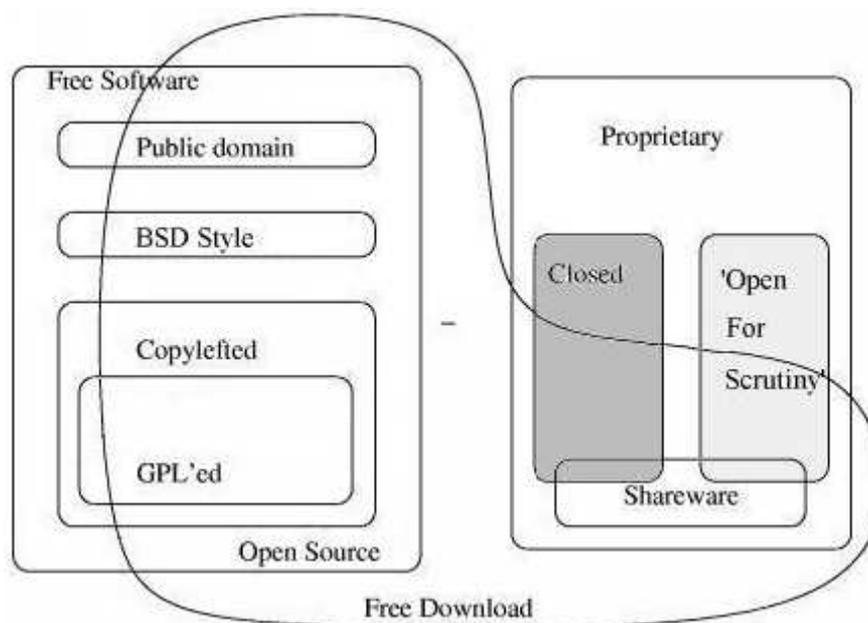
Related to the Open Source Definition is the **Free Software** definition by the Free Software Foundation, which attempts to capture what is required for a program license to qualify as being free-libre software. In practice, licenses which meet the open source definition almost always also meet the Free software definition. All licenses reported to meet the free software definition as of 2005 also meet the open source definition.

Free software is software that meets the following criteria:

1. users are free to use the program for any purpose;
2. users are free to examine the software to see how it works;
3. users are free to redistribute the program to others;
4. users are free to improve the program.

Although it is not explicitly referred to, access to the source code is a precondition for fulfilling these criteria.

A related concept is that of **Copyleft**. Copylefting licences grant the freedom of free or open source software subject to one condition: if the amended code is redistributed, it must be under the same license. This is the most controversial aspect of the GNU General Public License (GPL), but guarantees the rapid expansion of the community around the code developed under this licence. It also prevents any appropriation of code for lock-in purposes which would be possible with more permissive kind of licences. Copylefting license examples are, e.g., GPL, L-GPL and the W3C Software Notice and Licence.



Software in the public domain (that is, with no copyright license at all), meets the criteria for free or open source software as long as all source code is made available.

Free or open source software can be developed to work standalone or to be integrated with existing proprietary software.

The following list presents the software licenses which, as of 2005, were approved by OSI and therefore follow the open source definition.

- Academic Free License
- Adaptive Public License (APL)
- Apache Software License
- Apple Public Source License
- Artistic License
- Attribution Assurance Licenses
- BSD License
- Computer Associates Trusted Open Source License
- Common Development and Distribution License
- Common Public License
- End-User License Agreement (EULA)
- EU DataGrid Software License
- Eclipse Public License
- Educational Community License
- Eiffel Forum License
- Entessa Public License
- Fair License
- Frameworkx License
- GNU General Public License (GPL)
- GNU Lesser General Public License (LGPL)
- Historical Permission Notice and Disclaimer
- IBM Public License
- Intel Open Source License
- Jabber Open Source License
- Lucent Public License

- MIT License
- MITRE Collaborative Virtual Workspace License (CVW License)
- Motosoto License
- Mozilla Public License 1.0 (MPL)
- Mozilla Public License 1.1 (MPL 1.1)
- NASA Open Source Agreement
- Naumen Public License
- NetHack General Public License
- Nokia Open Source License
- OCLC Research Public License
- Open Group Test Suite License
- Open Software License
- PHP License
- Python License
- Python Software Foundation License
- Qt Public License (QPL)
- RealNetworks Public Source License
- Reciprocal Public License
- Ricoh Source Code Public License
- Sleepycat License
- Sun Industry Standards Source License (SISSL)
- Sun Public License (SPL)
- Sybase Open Watcom Public License
- University of Illinois/NCSA Open Source License
- Vovida Software License v. 1.0
- W3C License
- wxWindows Library License
- X.Net License
- zlib-libpng license
- Zope Public License

Examples of licenses which include the access to the source code but which do not follow the open source definition are:

- HESSLA
- Solaris
- PGP

I.1.6.2 MODULE B: USE OF INTERNET

B1. Did your enterprise have access to Internet, in January 2008?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. *Tick only one*; binary (Yes/No); filter question]

This variable refers to the access availability regardless of property and purposes of the connection at whatever moment during January 2008. The term "Internet access" means having an external connection to the Internet through an Internet 'service' provider (ISP).

The technologies used to connect to the Internet are physical wire based or wireless. So we can distinguish wire based connections from wireless connections. Another important distinction is between fixed connections and mobile connections. Access via fixed line refers to the connection between two stationary points, while mobile connections are performed via mobile phone networks. The distinctions wired/wireless and fixed/mobile are not the same, as wireless connections can be used to connect two fixed points, as is the case of satellite connections. Another breakdown of the Internet access is between temporary (dial-up on demand) and permanent connection (always on-line).

The Internet is a world-wide open computer mediated system (network) that transmits data by packet switching using a standardised Internet Protocol (IP) and many other related protocols (Internet protocol suit).

The Internet is made up of thousands of smaller commercial, academic, and government networks. It carries (provides access) to various information and communication services, such as electronic mail and the interlinked web pages and other documents of the World Wide Web. Internet refers to following Internet Protocol (IP) based networks and systems: WWW; Extranet over the Internet; EDI over the Internet; Internet accessed by mobile phones; Internet email.

The Internet protocol suite is the set of communications protocols that implement the protocol stack on which the Internet runs. It is sometimes called the TCP/IP protocol suite, after the two most important protocols in it: the Transmission Control Protocol (TCP) and the Internet Protocol (IP), which were also the first two to be defined.

An Internet 'service' provider (ISP) is a company that provides individuals and other companies access to the Internet and related services. An ISP has the equipment and the telecommunication line access required to have a point-of-presence on the Internet for the geographic area served. The larger ISPs have their own high-speed leased lines so that they are less dependent on the telecommunication providers and can provide better service to their customers. Most telecommunications operators are ISPs.

Generally, an ISP charges a monthly access fee to the consumer. The consumer then has access to the Internet for an unlimited number of hours, although the speed at which this data is transferred varies widely and depends on the type of the Internet connection.

B2: How many persons employed used computers connected to the World Wide Web at least once a week, in January 2008?

or

Please indicate an estimate of the percentage of the number of persons employed used computers connected to the World Wide Web at least once a week, during January 2008.

[Scope: enterprises with Internet access, i.e. B1 = Yes]

[Type: numerical, in real or percentage values]

This variable refers to persons employed who have access to the World Wide Web (WWW) from their workstation or from a computer to which he/she has free access. In principal, any person employed in an enterprise with access to the Internet can access it from a computer in the enterprise. However, here only persons employed to whom access was granted are counted. The criteria in this variable to identify Internet access is WWW access. The reason for its usage is that from the several services that can be run in the Internet the WWW is the most common one, together with e-mail. A mere access to the Internet through the e-mail service is not considered in this variable, because it's very already very common in Europe and it doesn't provide the same potentialities in terms of access to information as the WWW.

Just like with persons employed using computers (question A2), this variable can be collected in **real** or in **percentage** values. Eurostat in its model questionnaire (annex I.3.1) recommends using a combination of both, giving to the respondent the opportunity to choose to answer in real or percentage values.

Independently on how this variable is collected, in real or percentage values, the background variable "number of persons employed" (X2) is needed for grossing up the overall percentage of persons employed using computers connected to the World Wide Web.

B3. Did your enterprise have the following types of external connection to the Internet, in January 2008?

[Scope: enterprises with Internet access, i.e. B1 = Yes]

[Type: one single answer needed per item, i.e. *Tick only one per item; binary (Yes/No)*]

By external connection it is meant the type of the "last mile" connection of the enterprise (e.g. enterprise's computer-mediated network) to the network of the Internet access service provider "ISP". The last mile is the final leg of delivering communications connectivity to the enterprise.

Type of external Internet connection:

a) Traditional Modem (dial-up access over normal telephone line) or ISDN connection

Dial-up is a temporary connection to the Internet via an analogue (standard) modem and standard telephone line (public switched telephone network PSTN), which requires that the modem dial a phone number when Internet access is needed (to dial the Internet service provider's node to establish a modem-to-modem link, which is then routed to the internet).

Dial-up pertains to a telephone connection in a system of many lines shared by many users. A dial-up connection is established and maintained for limited time duration. A dial-up connection can be initiated manually or automatically by the computer's modem or other device. This type of the Internet connection is capable of carrying up to 56 kilobits per second (Kbit/s) and is considered as the typical example of the low capacity 'speed' connection (narrowband).

ISDN (Integrated Services Digital Network) connection is a temporary connection to the Internet using a type of circuit switched telephone network system (a set of CCITT/ITU standards), designed to allow digital (as opposed to analogue) transmission of voice and data over ordinary telephone copper wires (enables digital transmission over the public switched telephone network), resulting in better quality and higher speeds, than available with analogue systems.

Enterprises that install an ISDN adapter (in place of a telephone modem) receives up to 128 Kbps compared with the maximum 56 Kbps rate of an analogue (standard) modem connection. ISDN services can simultaneously transmit voice, data and video. ISDN is also considered as a low capacity 'speed' connection (narrowband).

This category includes Basic Rate Interface (BRI) consisting of two B channels, each with bandwidth of 64 Kbit/s, and one D channel with a bandwidth of 16 Kbit/s.

This category does not include Primary-Rate Interface, a type of ISDN service designed for larger organizations. PRI includes 23 B-channels (30 in Europe) and one D-Channel. PRI service is generally transmitted through an E1 line in Europe. This type of connection belongs under alternative c).

b) DSL (xDSL, ADSL, SDSL etc) connection

xDSL (Digital Subscriber Line) refers to a family of a high-bandwidth (broadband), local loop technologies that provide a digital permanent Internet connection over the copper wires of the local telephone network.

Asymmetric Digital Subscriber Line (ADSL) where more bandwidth is allocated to download than upload and High Rate Digital Subscriber Line (HDSL) are considered dominant DSL technologies. Typically, individual connections will provide from 1.544 Mbps to 512 Kbps downstream and about 128 Kbps upstream. A DSL line can carry both data and voice signals and the data part of the line is continuously connected. DSL connection is considered as one of the high capacity 'speed' permanent 'fixed' Internet connection (broadband).

c) Other fixed internet connection

Other high capacity 'speed' fixed (wire or wireless) connection includes the following types of the Internet connection: Cable modem 'cable TV network connection'; High capacity leased lines 'Frame Relay, ATM, Digital Multiplex'; Ethernet LANs connection; Optical fibre connection; Satellite connection; Wi-fi connection, other FWA connections etc..

Cable modem uses modems attached to cable television networks (cable TV lines) for permanent 'fixed' access to the Internet. The term cable internet (or simply cable) refers to the delivery of internet service over this infrastructure. A cable modem is a device that enables you to hook up your PC to a local cable TV line and receive data at about 1.5 Mbps. It is considered as one of the high capacity 'speed' permanent 'fixed' Internet connection (broadband).

High capacity leased line is a permanent telephone connection between two points set up by a telecommunications common carrier. Typically, leased lines are used by businesses to connect geographically distant offices. Unlike normal dial-up connections, a leased line is always active. Because the connection doesn't carry anybody else's communications, the carrier can assure a given level of quality. For example, a T-1 channel is a type of leased line that provides a maximum transmission speed of 1.544 Mbps. You can divide the connection into different lines for data and voice communication or use the channel for one high speed data circuit. Dividing the connection is called multiplexing. Increasingly, leased lines are being used by companies, and even individuals, for Internet access because they afford faster data transfer rates and are cost-effective if the Internet is used heavily.

A leased line is a telephone line that has been leased for private use. In some contexts, it's called a dedicated line. A leased line is usually contrasted with a switched line or dial-up line". Leased lines are usually available at speeds of 64k, 128k, 256k, 512k, 2Mb and provided to the customer on X.21 presentation. Frame relay protocol and T-1 and T-3 (in Europe called E1 and E3) lines are used for the Internet connection via leased lines. Higher speeds are available on alternative interfaces.

Frame relay is an efficient data transmission technique used to send digital information quickly and cheaply to one or many destinations from one point. It is a packet-switching protocol for connecting devices on a Wide Area Network (WAN). Frame Relay networks in the U.S. support data transfer rates at T-1 (1.544 Mbps) and T-3 (45 Mbps) speeds. In fact, you can think of Frame Relay as a way of utilizing existing T-1 and T-3 lines owned by a service provider. Most telephone companies now provide Frame Relay service for customers who want connections at 56 Kbps to T-1 speeds. In Europe, Frame Relay speeds vary from 64 Kbps to 2 Mbps. Frame relay is being displaced by ATM and native IP based products, including IP virtual private networks.

Fixed wireless Internet connection (FWA) is a technology which uses radio-frequency, infrared, microwave, or other types of electromagnetic or acoustic waves in place of wires, cables, or fibre optics to transmit signals or data (provide Internet access) between stationary (fixed) points. It includes e.g. a satellite Internet connection (long range wireless transmission) or Wi-fi (medium range wireless transmission).

Wi-Fi (or Wi-fi, WiFi, Wifi, wifi), short for "Wireless Fidelity", is a set of Ethernet standards for wireless local area networks (WLAN) currently based on the IEEE 802.11 specifications. New standards beyond the 802.11 specifications, such as 802.16 are currently in the works, they offer many enhancements, anywhere from longer range to greater transfer speeds. Wi-Fi was intended to be used for wireless devices and LANs, but is now often used for Internet access (one of the main international standards for wireless broadband Internet access and networking, with widespread use in business, homes and public spaces). It is based on radio signals with a frequency of 2.4 GHz and capable of speeds of up to 11 Mbps. It enables a person with a wireless-enabled computer or personal digital assistant to connect to the Internet when in proximity of an access point called a hotspot.

Under this category it's not included the Internet connection via high capacity 'speed' mobile phone networks, which are included in alternative d) instead.

d) Mobile connection

This includes connections to the Internet via mobile phone networks and includes connections via low capacity 'speed' mobile phone networks (HSCSD, GPRS etc) - mobile

Internet narrowband - and connections via high capacity 'speed' mobile phone networks (UMTS 'W-CDMA'; EDGE 'EGPRS'; CDMA 2000x etc) - mobile Internet broadband.

A Connection via **low capacity 'speed' mobile phone networks** is an access to the Internet using a long range wireless transmission of the mobile network technologies as High-Speed Circuit-Switched Data (HSCSD) or General Packet Radio Service (GPRS) that is sometimes called as 2,5 mobile generation technology (2,5 G).

Access to the Internet via mobile phone networks should be considered as low capacity 'speed' mobile connection (narrowband) if it is being equal to, or greater than 256 kbit/s, as the sum of the capacity in both directions (download or upload).

HSCSD is a development of Circuit Switched Data, the original data transmission mechanism of the GSM mobile phone system. As with the CSD channel the allocation is done in circuit switched mode. The difference comes from the ability to use different coding methods and even multiple time slots to increase data throughput. HSCSD is a temporary mobile connection.

2.5G is a stepping stone between 2G and 3G cellular (mobile) wireless technologies. The term "second and a half generation" is used to describe 2G-systems that have implemented a packet switched domain in addition to the circuit switched domain. While the terms "2G" and "3G" are officially defined, "2.5G" is not. 2.5G provides some of the benefits of 3G (e.g. it is packet-switched) and can use some of the existing 2G infrastructure in GSM and CDMA networks. The most commonly known 2.5G technique is GPRS.

GPRS is a 2.5G mobile standard typically adopted by GSM operators as a migration step towards 3G (W-CDMA). GPRS is based on packet-switched technology enabling high-speed data transmission.

Examples of some most common spread standards of low capacity 'speed' mobile network connection and their bandwidth (can be different between the countries and also within the same country):

	download	upload
GPRS 4+1	57.6 kbit/s	14.4 kbit/s
GPRS 3+2	43.2 kbit/s	28.8 kbit/s
CSD	9.6 kbit/s	9.6 kbit/s
HSCSD	28.8 kbit/s	14.4 kbit/s
HSCSD	43.2 kbit/s	14.4 kbit/s

A connection via **high capacity 'speed' mobile phone networks** is an access to the Internet using a long range wireless transmission of the 3rd generation (3G) mobile network technologies based on the CDMA (Code Division Multiple Access) as UMTS (Universal Mobile Telephone System - Wideband 'W'-CDMA); CDMA2000x; CDMA 2000 1x EV-DO; CDMA 2000 1x EV-DV) or some other high capacity mobile technologies based on the GPRS as EDGE (Enhanced Data rates for Global Evolution - EGPRS) etc.

3G (or 3-G) is short for third-generation mobile telephone technology. The services associated with 3G provide the ability to transfer both voice data (a telephone call) and non-voice data (such as downloading information, exchanging email, and instant messaging). It includes high-speed mobile networks (e.g., CDMA2000 1X, WCDMA, CDMA2000 1xEV-DO, etc.).

Universal Mobile Telecommunications System (UMTS) is one of the third-generation (3G) mobile phone technologies. It uses W-CDMA as the underlying standard, is standardized by the 3GPP, and represents the European answer to the ITU IMT-2000 requirements for 3G Cellular radio systems. It presently delivers packet switched data transmission speeds up to 384 kbps and up to 2 Mbps when fully implemented.

CDMA2000 1x is an IMT-2000 3G mobile network technology, based on CDMA that delivers packet switched data transmission speeds of up to 144 kbps. It is also referred to as 1XRTT.

CDMA2000 1xEV-DO is an IMT-2000 3G mobile network technology, based on CDMA that delivers packet switched data transmission speeds of up to 2.4 Mbps.

Enhanced Data rates for Global Evolution (EDGE) is an intermediate technology that brings second-generation GSM closer to third-generation capacity for handling data speeds up to 384 Kbits/s. The standard is based on the GSM standard and uses TDMA multiplexing technology.

Narrowband is defined by the type of the Internet connection and includes: 1) Dial-up (connection via standard telephone line 'PSTN using analogue modem); and 2) ISDN (Integrated Services Digital Network)

Broadband is defined by the type of the Internet connection and includes: 1) DSL (Digital Subscriber Line) technologies as ADSL, HDSL, SDSL, VDSL that form core broadband; and 2) Other high capacity 'speed' fixed (wire or wireless) connection (Cable modem 'cable TV network connection'; Leased lines 'Frame Relay, ATM, Digital Multiplex'; Ethernet LANs connection; Optical fibre connection; Satellite connection; Wi-fi connection etc) – extended fixed wire based and wireless broadband. Within the framework of this survey, mobile 'broadband' – part of item d) of this question – is not included in this definition of broadband.

B4. Did your enterprise use the Internet for the following purposes, in January 2008?

[Scope: enterprises with Internet access, i.e. B1 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

This variable intends to measure the adoption of the Internet by the enterprise as a user or consumer of several services. These services may be paid or not. This does not include the provision by the enterprise of these services in the Internet. For example, for "banking and financial services" this variable measures the use of the Internet to use the services provided by banks and not the use of the Internet by banks to provide financial services.

This variable includes the use of the Internet in general and is not restricted to the World Wide Web.

The following Internet uses are measured:

a) Banking and financial services

This includes the use of web-banking, the consultation of financial information (like exchange rates, interest rates, etc.) and the use of the Internet for automatic data

interchange between the enterprise and other financial organisations (for instance through the use of XML or EDIFACT).

b) Training and education

This includes the use of online education programs by the employees.

B5. Did your enterprise use the Internet for interaction with public authorities, during 2007?

[Scope: enterprises with Internet access, i.e. B1 = Yes]

[Type: one single answer needed, i.e. Tick only one; binary (Yes/No); filter question]

As the competence of the government or the 'public authorities' differs strongly from one country to another, the concept has been kept as wide as possible: public authorities refer to both public services and administration activities. While 'administration' refers to obligation and rights one has as a citizen or as an enterprise in activity in the country (the so-called red tape), public services can refer to non-administrative tasks or competences of government bodies, e.g. offering the public library's catalogue on-line.

The public authorities can be at national or regional level, but also at the level of cities or municipalities. They can even be 'semi-governmental', e.g. public libraries, hospitals, universities, etc.

B6. Did your enterprise interact with public authorities in the following ways, during 2007?

[Scope: enterprises having interacted with public authorities, i.e. B6 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

a) For obtaining information

Includes searching to obtain any type of information from public authority web sites. Public authorities' web sites include local or central government offering information or services. It would be helpful to provide a list of local/central government web sites to facilitate respondents to recall public authority web sites that they have used. For example respondents may not realise some public information sites are run by public authorities, e.g. a meteorological office offering weather forecast services.

b) For obtaining forms, e.g. tax forms

Includes downloading official forms for any purpose of use (e.g. for information or for requesting a service). The downloaded files could be in formats such as PDF or Microsoft Word.

c) For returning filled in forms, e.g. provision of statistical information to public authorities

Includes filled in forms sent via internet only. Forms downloaded, printed, filled in and sent by post should not be included in this category.

d) For treating an administrative procedure completely electronically

Includes only administrative procedures (e.g. registration, authorisation request) for which all steps can be treated electronically by means of the internet, including payment, signature etc if needed; procedures (partially) carried out using paper or manually typed email messages are excluded under this item.

e) For submitting a proposal in an electronic tender system (e-procurement)

This item includes the submission of proposals in Internet based systems (either based on web interfaces or any other architecture). It excludes the sending of proposals by normal manually typed email messages.

This question makes use of the conceptual framework for the levels of interaction between enterprises or citizens and public authorities, used by DG INFSO. The framework of four levels of interaction was developed to measure the level of sophistication offered by websites supplying basic public services. That way, one can distinguish websites that only provide information on a public service (1st level) from websites that allow to download forms (2nd level) or to submit information in an online form (3rd level) up to websites that allow to completely treat the public service or administrative obligation without the need for any paperwork whatsoever (4th level of sophistication).

It is expected or logical that a particular website that reaches a certain level of sophistication also includes all lower levels, for instance a website offering the possibility to download a form should certainly provide information on the particular public service that form would refer to. The concept of the four levels of sophistication was adopted for the measurement of the usage of e-government by enterprises (and the first 3 levels for citizens). However, the interpretation of the results will have to differ slightly from the interpretation done for the sophistication of the public websites. At usage level, it is natural to assume that a user (enterprise or citizen), before knowing which form to download (2nd level), will obtain information on the public service he is interested in (1st level), but he/she will only do it once or twice and not every time s/he downloads the form. Therefore, when measuring the usage every year, it is perfectly possible that there are users who have downloaded forms for obtaining information in those cases in which the user already had all the information he needed (although not very probable, as users may well constantly search for information on several different issues). In this sense, it is possible that the respondents answers Yes to item *b* without having answered Yes to *a* (or Yes to *c* without having answered Yes to *a* and/or *b*).

B7. Did your enterprise have a Web Site or Home Page, in January 2008?

[Scope: enterprises with Internet access, i.e. B1 = Yes]

[Type: one single answer needed, i.e. Tick only one; binary (Yes/No); filter question]

This observation variable doesn't refer specifically to the ownership of the website, but to the use of a website by the enterprise to present its 'business'. It includes not only the existence of a website which is located in servers which belong to the enterprise or are located at one of the enterprise's sites, but also third party web-sites (e.g. one of the group of enterprises to which it belongs).

However, it doesn't include any web presence of the enterprise. That would be too broad, as it would include the presence of the enterprise (e.g. its name or its contact information) in directories and online yellow pages. These cases are not included in this variable.

Finally, this variable includes any type of website, independently of its sophistication or services provided.

B8. Did the Web Site provide the following facilities for your enterprise, in January 2008?

[Scope: enterprises with a website, i.e. B7 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

For those enterprises using a website, this question focuses on the measurement of its specific use. Just like in the case of the specific use of the intranet (see question A6), this information is of particular importance, because the sole existence of the website provides only the *potential* for better performance by the enterprise. The purpose of this question is to *complement* the measurement of *e-business*.

a) Product catalogues or price lists

It includes the provision of lists of products or services offered by the enterprise to its clients and which might include also the characteristics of these products and their prices. The information about the products or services may be detailed or not. The information may be static or dynamic (extracted online from a database and as such always updated).

b) Possibility for visitors to customise or design the products

This item refers to the existence in the website of an interactive interface where users can choose from several possible characteristics of the product (colour, etc.) or services and see online in the website the impact, for instance, on the price. The interface might also include the possibility for the user to visualise online the appearance of the product with the options that were selected. For the financial sector, the carrying out simulations or what-if calculations for products like loans belongs here as well.

c) Online ordering or reservation or booking, e.g. shopping cart

This item refers to a facility which allows the user to order products or services with no additional contact offline or via email necessary (for the ordering). It includes also websites which allow the reservation of hotel rooms or the booking of flights. It does not include a link in the website which directs the user to an email application which requires the user to send the order via email. Payment may or may not be included in the ordering facility, e.g. payment may be made on reception of the product.

Carrying out via online banking a transaction in general does not qualify as online ordering in case the enterprise is in the banking sector; specific cases however, e.g. when opening online an account (that implies payment of a fee) or when buying shares (with a commission to be paid to the bank), qualify as online orders in this sector.

d) Online payment

In this facility the customer enters in the enterprise's website financial information, e.g. credit card, debit card or bank account information, which provides the enterprise with a transfer to its own bank account. Typically, the transfer is made with the intermediation

of another enterprise called payment gateway. The payment gateway facilitates the transfer of information between the enterprise's website and the bank.

Another method, included in this item, for the customer to make the online payment is through an intermediary which holds a balance account which can be used to make the purchases in the enterprise's website.

This item includes the cases where the form to enter the financial information is not physically in the website of the enterprise, but is for instance provided by the payment gateway. In this case the customer might not even be aware that he moved from the enterprise's website to a page in the payment gateway website to enter this information. After the payment is confirmed, the customer is returned to the enterprise's website.

However, this item does not include the case where the website requires the customer at some stage of the process to go to the online banking facility of his or her own bank to make the transfer.

A special situation arises here again for the banking sector: online payment refers here of course to payments to the enterprise/bank itself for services provided (like the commission charged for buying shares).

e) Personalised content in the website for regular/repeated visitors

This facility consists on the ability of the website to recognise the user from previous visits and adapt the content of the pages accordingly. For the financial sector one could think of the possibilities to save in a personal environment the information of (business) relations and the registration/consultation of future transactions.

f) Advertisement of open job positions or online job application

This item includes both cases where just simple information on job vacancies is provided in the website as well as those where the website provides also an online facility for candidates to apply for the jobs.

B9: Was your enterprise, in January 2008, using a digital signature in any message sent, i.e. using encryption methods that assure the authenticity and integrity of the message (uniquely linked to and capable of identifying the signatory and where any subsequent change to the message is detectable)?

[Scope: enterprises with Internet access, i.e. B1 = Yes]

[Type: one single answer needed, i.e. Tick only one; binary (Yes/No)]

This question aims at measuring the adoption of digital signatures by enterprises, also referred to as advanced electronic signature¹. One important distinction to note is between the more general concept of electronic signature and the narrower one of digital signature.

Electronic signature refers to data in electronic form which are attached to or logically associated with other electronic data and which serve as a method of authentication constituting in that manner a signature. A digital signature is an electronic signature which uses cryptographic means to add non-repudiation and message integrity. Non-repudiation

¹ Advanced electronic signature is the term used in the Directive 1999/93/EC of the European Parliament and of the Council on a Community framework for electronic signatures.

means that the message is authentic and is uniquely linked to and capable of identifying the signatory. Integrity means that any eventual subsequent change to the data will be detected if tested for using the signature.

A signature in paper which is then digitised and sent together by e-mail is not a digital signature. It is even questionable that it is an electronic signature. Even if considered an electronic signature, it isn't considered a digital signature in this manual because it can't assure that the data transmitted was really created by the signatory and it has no way to assure the integrity of the message. Similarly, a signed paper document sent by fax does not constitute the use of a digital signature.

Examples of electronic signatures which are not digital signatures are PIN codes, chip cards, retina scanners and fingerprint readers. Although these devices allow the authentication of signatories, they don't provide data integrity assurance.

Some web pages and software EULAs (end user license agreements) claim that various electronic actions are legally binding signatures, and so are instances of electronic signature. For example, a web page might announce that, by accessing the site at all, you have agreed to a certain set of terms and conditions. However, these mechanisms are not digital signatures.

It is important to understand that the use of cryptographic techniques in digital signatures are much more than an error checking technique akin to checksum algorithms, or even high reliability error detection and correction algorithms. These can offer no assurance that the text has not been changed. In addition, no message integrity protocols include error correction, for to do so would destroy the tampering detection feature.

Popular electronic signature standards include the OpenPGP standard supported by PGP and GnuPG, and some of the S/MIME standards. All current cryptographic digital signature schemes require that the recipient have a way to obtain the sender's public key with assurances of some kind that the public key and sender identity belong together, and message integrity measures which assure that neither the attestation nor the value of the public key can be surreptitiously changed. However, a secure channel is not required to transmit the data or message.

Council Directive 2001/115/EC (the "e-invoice" Directive) includes as one of the conditions to make the acceptance of e-invoices obligatory the use of procedures that assure the authenticity and integrity of its contents, of which digital signature is one of the methods. The adoption of digital signatures is also relevant in a broader sense as it is also a measure to increase security in the online environment and it applies to all electronic documents circulating electronically between enterprises.

The requirement that the e-signature has a legal binding status is not considered as a relevant feature for measurement. What is important is the adoption of this authentication mechanism.

In the scope of EU legislation the digital signature is referred to as "advanced electronic signature". However, the wording "digital signature" should be used in the question as supposedly it is the term used most often by IT professionals.

Nevertheless, for the construction of the question, the wording of the Directive 1999/93/EC, when it refers to message integrity and authenticity, should be used (article 2, point 2).

I.1.6.3 MODULE C: AUTOMATED DATA EXCHANGE

The purpose of this module is to measure the external integration of the business processes of the enterprise, i.e. the linking with other enterprises. For that, it uses the concept of automated data exchange. Conceptually, it corresponds exactly to Electronic Data Interchange (EDI). However, the term EDI needs to be avoided in the questionnaire because it is understood many times by the respondents as a synonymous of closed proprietary computer networks.

Automated data exchange between the enterprise and other ICT systems outside the enterprise means:

- exchange of messages (e.g. orders, invoices, payment transactions or description of goods)
- via the internet or other computer networks
- in an agreed format which allows its automatic processing (e.g. XML, EDIFACT etc.)
- without the individual message being manually typed.

Making reference in the clarifications of the questions to the (mass) machine-machine communication aspect of automated data exchange will help the respondent to recognise this form of external integration; the same applies to referring to national examples of automated data exchange systems.

C1. In January 2008, was your enterprise using such automated data exchange?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. *Tick only one*; binary (Yes/No); filter question]

C2. Was automated data exchange used for the following purposes?

[Scope: enterprises which use automated data exchange, i.e. C1 = Yes]

[Type: one single answer needed per item, i.e. *Tick only one per item*; binary (Yes/No)]

The items in this question are basically general purpose business messages that make part of several standards developed for the transmission of business messages. This has the advantage of facilitating the answer of the respondent when he finds a correspondence between those standard business messages and the items in this question.

The list of items allow, in one hand, the identification of the automation of specific business processes (e.g. the automation of purchases is identified with the use of ADE for sending orders to suppliers) and, on the other hand, to measure the extensiveness of the use of ADE in the enterprise (i.e. the use of more types of business messages means a more extensive use of ADE).

The list of items is the crossing of the type of message, which corresponds to the content of the message, and the interlocutor. It distinguishes as content: orders, invoices, product

information, transport documents and payment instructions; and as interlocutor: suppliers, customers, financial institutions and public authorities.

The list is not exhaustive: other purposes for the automated data exchange not included in this question but covered by question C1 are requesting a quotation, inquiring the status of an order etc.

- a) Sending orders to suppliers
- b) Receiving e-invoices
- c) Receiving orders from customers
- d) Sending e-invoices
- e) Sending or receiving product information (e.g. catalogues, price lists, etc.)
- f) Sending or receiving transport documents (e.g. consignment notes)
- g) Sending payment instructions to financial institutions

A payment instruction, also named payment order, is a business message sent by the enterprise to a bank where it has an account instructing the bank to debit its account by a certain amount and to arrange for its payment to a certain beneficiary. The beneficiary can be a supplier, an employee or any other entity.

- h) Sending or receiving data to/from public authorities (e.g. tax returns, statistical data, [national examples], etc.)

C3. Were the following formats used for the automated data exchange?

[Scope: enterprises which use automated data exchange, i.e. C1 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

This question is the more technical one of this module. The remaining questions of module C should be known by the IT decision maker in the enterprise. However, the format used to exchange the messages is probably hidden behind the software applications used. They should be known by the IT department of the enterprise, but if IT implementation is outsourced, this information might not be known by the enterprise.

However, this information is still very important because the format used provides insight about the capacity of the enterprise to integrate its business processes with those of other enterprises. The format used can follow an open standard or be a closed proprietary one agreed by the enterprise and its business partner. The use of open standards supposedly provides the enterprise with a higher capacity for external integration of business processes.

Nowadays, there seems to exist in general an interest of the policy makers to promote the adoption of open standards.

Open standards are publicly available and can be readily implemented by the enterprises. In principle, they are sufficiently general so they can be applied by many different enterprises. They have the advantage over closed propriety formats that they don't have the cost of development of the format and that the same format can be used by the enterprise with

several business partners. With a lower cost of implementation, open standards also seem to benefit the adoption of e-business by smaller enterprises. On the other hand, closed formats, not being readily accessible to the public, offer in theory higher security as it is more difficult to intercept and understand the messages exchanged. However, the use of proper encryption methods in the exchange of the messages, even when following an open standard, should offer the same level of security.

Among the open standards one can distinguish mainly two families of standards, those based on EDI and those based on XML. EDI is more ancient and usually more expensive to implement. XML, a more recent format, is usually cheaper and considered an opportunity for small enterprises to adopt e-business.

The three items in this question intend to capture these two dimensions (open – closed, 'old'/expensive – 'new'/cheap).

a) EDIFACT or similar standards (e.g. EANCOM, ANSI X12)

EDI (electronic data interchange) is a family of standards for structuring information to be exchanged electronically between and within enterprises and other organisations. The standards define structures that mimic business documents and forms, such as purchase orders and invoices, in order to allow the automation of business processes. The two main sets of EDI standards are UN/EDIFACT, mainly used in Europe, and ANSI ASC X12, mainly used in North America.

UN/EDIFACT stands for *United Nations/Electronic Data Interchange for Administration, Commerce and Transport* and is the international EDI standard developed under the United Nations. EDIFACT has been adopted by the International standards Organisation (ISO) as the ISO standard 9735.

ANSI ASC X12 stands for American National Standards Institute Accredited Standards Committee X12. It is the EDI standard developed by the US National standards body for the development and maintenance of EDI standards for the United States.

EANCOM (EAN + Communication) is a subset of the EDIFACT standard, which is used worldwide in the consumer goods industry. Originally designed for the retail sector, EANCOM has become one of the most widely used UN/EDIFACT subsets, and is now found in other industries such as health, publishing, transport and electronics.

There are also other EDIFACT subsets such as EDIKEY, EDIFURN and EDIFICE. They are all included in this item.

Example of an EDIFACT message (an Intrastat declaration)

```
UNA:+.? 'UNB+UNOC:3+7899999999+STATEC::INTRASTAT+990626:0941+AA990601'UNH+1+CUSDE
C:D:99A:UN:INSTAT'BGM+896+1233455666990AAAA000001'CST++A:176'DTM+320:199906:610'DTM+
137:19990626:102'GIS+2:105'RFF+ACD:IDEP:8.1'NAD+DT+1233455666+PSI NAME:15 STATISTICS S
TREET:CITY:123456:COUNTRY'COM+123654789:TE'COM+321654987:FX'NAD+AE+7899999999+TDP:
45 THIRD PARTY STREET:CITY:789456'COM+78888888:TE'COM+56666666:FX'NAD+DQ+STATEC+ST
ATEC (SERVICE INTRASTAT):B.P. 667 6,BD. ROYAL:L-2016LUXEMBOURG L-2449 LUXEMBOURG:TE
L.?: 478-4286 OR 478-4293:FAX?: 220371, 220381, 229158, 222065'MOA+ZZZ::LUF'MOA+39:1044444'U
NS+D'CST+1+01011100:122++11:112'LOC+35+DK'LOC+27+DZ'MEA+WT++KGM:1200'MEA+AAE++PC
E:12'TDT+2++4'MOA+38:600000'MOA+123:880000'CST+2+01039211:122++32:112'LOC+35+FR'LOC+27
+508'MEA+WT++KGM:15000'MEA+AAE++PCE:150'TDT+2++1'MOA+38:444444'MOA+123:600000'UNS+
S'CNT+2:2'CNT+18:16200'CNT+19:162'CNT+20:1044444'CNT+22:1480000'AUT+IDEP'DTM+187:199906
26:102'UNT+42+1'UNZ+1+AA990601'
```

b) XML based standards, for example ebXML, RosettaNet, UBL, papiNET

XML is the abbreviation for *eXtensible Markup Language* and it's an emerging format, newer than EDIFACT, which allows data to be encoded in a meaningful structure which

can be understood by both humans and computers. XML is a formal specification of the World Wide Web Consortium (W3C) similar to HTML (Hypertext Markup Language), the language in which web pages are stored and transmitted. However, XML is more general than HTML and it allows its extension to include user-specified or industry-specified elements. By itself, XML does not define the structure in which business messages or documents can be sent. It needs to be extended with the definition of a specific structure and data elements tailored for the transmission of business information.

Example of an XML message (a hypothetical invoice)

```
<?xml version="1.0" encoding="UTF-8"?>
<Invoice>
  <InvoiceHeader>
    <InvoiceNumber>2007-UZP998877</InvoiceNumber>
    <InvoiceDate>2007-04-30</InvoiceDate>
    <TermsPaymentInfo>
      <PercentageDiscount>5.00</PercentageDiscount>
      <DaysUntilDue>40</DaysUntilDue>
    </TermsPaymentInfo>
    <TotalAmount>8017.35</TotalAmount>
  </InvoiceHeader>
  <InvoiceDetail>
    <InvoiceLineItem>
      <LineItemNumber>713-256</LineItemNumber>
      <LineItemDescription>Black typewriter</LineItemDescription>
      <UnitPrice>129.00</UnitPrice>
      <Quantity>1</Quantity>
      <LineItemTaxes>0.00</LineItemTaxes>
      <LineItemComments>Item on promotion.</LineItemComments>
    </InvoiceLineItem>
  </InvoiceDetail>
  <InvoiceParties>
    <CustomerParty>
      <Name>Business A</Name>
      <CustomerNumber>98765</CustomerNumber>
      <Address>100 Somewhere Street</Address>
      <TaxIDNumber>9871364-45</TaxIDNumber>
      <PhoneNumber>(352) 555-11111</PhoneNumber>
    </CustomerParty>
    <VendorSupplierParty>
      <Name> Business B</Name>
      <Address>200 Anywhere Avenue</Address>
      <TaxIDNumber>8769253-45</TaxIDNumber>
      <PhoneNumber>(352) 555-22222</PhoneNumber>
    </VendorSupplierParty>
  </InvoiceParties>
</Invoice>
```

ebXML, or e-business using XML, is a set of XML based standards sponsored by the United Nations (UN/CEFACT) and OASIS (Organization for the Advancement of Structured Information Standards), for conducting business over computer networks. The objective of ebXML is to enable a global marketplace where enterprises of any size and in any place can meet to do business via the exchange of XML based business messages. As such, it is more than just a standard to exchange business messages. It also provides a standard method to conduct trade relationships and define business processes. However, this item refers only to the use of ebXML standards to exchange business messages.

RosettaNet is a non-profit consortium of private companies dedicated to the development of standard electronic commerce interfaces to align the processes of enterprises in the supply chain. Its origin is in the high technology and adjacent industries, but it starts to be used also in other sectors. It was originated in the United States and it is more spread there, but it is also used in Asia and Europe. The RosettaNet

standards are XML based. The area most addressed is the supply-chain, but other areas like manufacturing are also covered.

papiNet is an initiative of the forest and paper industries to develop, maintain and promote standards for electronic business messages amongst all parties involved in the buying, selling and distribution of forest, paper and wood products. papiNet standards are XML based and business messages defined by papiNet include purchase order, confirmation of delivery, invoice, etc.

UBL (Universal Business Language) is an OASIS sponsored royalty-free library of standard electronic XML business documents such as purchase orders, shipping notices and invoices. Its purpose is to solve the problem of the many industry specific XML based standards that have been developed, leading to the definition of many messages that are not applicable when other industries adopt the standards and to the existence of different definitions of basic documents common to most industries. UBL is a simpler and broader standard defining a set a common business messages which exist in most industries, like orders, invoices and shipment notices. UBL is built on the ebXML infrastructure. However, ebXML is more ambitious in its objectives and UBL documents are usable in other e-commerce frameworks. Nevertheless, UBL is ebXML compatible.

c) Proprietary standards agreed between you and other organisations

This item covers the "closed" formats, thus not belonging to the open formats of the items a) or b).

C4. Were the following issues reasons for the enterprise not to use automated data exchange?

[Scope: enterprises which do not use automated data exchange, i.e. C1 = No]

[Type: one single answer needed per item, i.e. Tick only one per item (Yes/No/Don't know)]

In this question, one or more reasons for not using automated data exchange can be selected. The – non-exhaustive – list of reasons that can be chosen is:

- a)** No interest in using it, because it isn't relevant for the business
- b)** Lack of expertise in-house for its implementation
- c)** Return on the investment too low or not clear
- d)** Lack of appropriate software for the specific sector/size of the enterprise
- e)** Difficulty with agreeing common standards with business partners
- f)** Uncertainty of the legal status of the messages exchanged

Frequent use of the "Don't know" answer category should be avoided. Several reasons for not using automated data exchange can be selected.

I.1.6.4 MODULE D: SHARING ELECTRONICALLY INFORMATION ON THE SUPPLY CHAIN MANAGEMENT

As the previous module, this Module intends to measure the level of integration of business processes of an enterprise with those of its suppliers and/or customers. It focuses on processes related to the supply chain management (SCM). Where in Module C the keyword is "in an agreed format" and the area of business processes for which exchange takes place is very wide, in module D the scope of business processes studied much more narrow – the coordination of availability and delivery of products or services – while excluding only manually typed e-mail messages as carrier of the electronic exchange of information.

Definition

Sharing electronically information on the supply chain management means:

- exchanging all types of information with suppliers and/or customers in order to coordinate the availability and delivery of products or services to the final consumer;
- including information on demand forecasts, inventories, production, distribution or product development;
- via computer networks, not only the Internet but also other connections between computers of different enterprises.
- It can be from you to your suppliers/customers or the other way around.

This information may be exchanged via websites or via automated data exchange (recall definition in module C), but it excludes normal e-mail messages manually written.

D1. In January 2008, was your enterprise regularly sharing electronically information on the supply chain management with your suppliers or customers?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. *Tick only one*; binary (Yes/No); filter question]

This question measures the exchange of electronic information in the context SCM with suppliers, customers or both; the answer does not say anything about the coverage of different purposes or the intensity of the exchange.

D2. Was your enterprise regularly sharing electronically the following information with its suppliers, in January 2008?

[Scope: enterprises which are sharing regularly electronically information n the supply chain management with their suppliers or customers, i.e. D1 = Yes]

[Type: one single answer needed per item, i.e. *Tick only one per item*; binary (Yes/No)]

For this question, only the sharing of information with **suppliers** is measured for the following purposes.

- a) Inventory levels, production plans or demand forecasts or (optionally)
 - a1) Demand forecasts
 - a2) Inventory levels
 - a3) Production plans
- b) Progress of deliveries (i.e. distribution of raw materials or finished products)

D3. Was your enterprise regularly sharing electronically the following information with its customers, in January 2008?

[Scope: enterprises which are sharing regularly electronically information in the supply chain management with their suppliers or customers, i.e. D1 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

For this question, only the sharing of information with **customers** is measured for the following purposes.

- a) Inventory levels, production plans or demand forecasts or (optionally)
 - a1) Demand forecasts
 - a2) Inventory levels
 - a3) Production plans
- b) Progress of deliveries (i.e. distribution of raw materials or finished products)

D4. Were the following methods used for the electronic exchange of this information, in January 2008?

[Scope: enterprises which are sharing regularly electronically information in the supply chain management with their suppliers or customers, i.e. D1 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

The items in this question allow measuring the technique used for the sharing of information electronically; but, sharing information via automated data exchange implies a further integration with systems of suppliers and/or customers (agreed format).

- a) Websites (yours, those of your business partners or web portals)
- b) Automated data exchange (XML, EDIFACT, etc.)

For the second method please refer to the explanations provided under question C3.

I.1.6.5 MODULE E: AUTOMATIC SHARE OF INFORMATION WITHIN THE ENTERPRISE

This module measures the internal integration of business functions within the enterprise; the first two questions focus on the links between specific front-office functions (sales, purchases) and specific related back-office functions (accounting, inventory, production/services and distribution); the 3rd and 4th questions intend to measure the adoption of specific software (ERP, CRM) which have the objective of integration of business processes.

Sharing information electronically and automatically between different functions of the enterprise means any of the following:

- Using one single software application to support the different functions of the enterprise;
- Data linking between the software applications that support the different functions of the enterprise;
- Using a common database or data warehouse accessed by the software applications that support the different functions of the enterprise;
- Automated data exchange between different software systems (recall definition in module C).

E1. In January 2008, when your enterprise received a sales order (either electronically or not), was the relevant information about it shared electronically and automatically with the software used for the following functions?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

- a) Your management of inventory levels
- b) Your accounting
- c) Your production or services management
- d) Your distribution management

The purpose of the questions E1 and E2 is to provide a measure of the integration of business processes within an enterprise. Implicitly, we are identifying that integration with automatic links. That is, something changes in the IT system supporting a business process and automatically (i.e. without human intervention) that change is reflected in the IT system supporting another business process. The next issue then is which business processes we are interested in (as we can't cover all business processes in the enterprises, and as they actually differ between enterprises from different economic activities). The solution chosen here is to start with two business processes that exist in every enterprise: selling (E1) and buying (E2).

The final purpose of these questions is to identify the existence of - automatic - links between different business processes.

The definition of software application is not restricted to any specific software used or its level of sophistication. The level of sophistication of a software application or a system is not directly correlated to its effectiveness or efficiency, and therefore correlated to a positive impact in the enterprise, which ultimately that's what we want to identify.

So, in conclusion, even an excel worksheet or an access database where orders are inserted manually can be considered as a software application for managing orders.

E2. In January 2008, when your enterprise sent a purchase order (either electronically or not), was the relevant information about it shared electronically and automatically with the software used for the following functions?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

- a) Your management of inventory levels
- b) Your accounting

E3. In January 2008, did your enterprise have in use an ERP software package to share information on sales and/or purchases with other internal functional areas (for example, finance, planning, marketing, etc.)?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item (Yes/No/Don't know)]

ERP stands for Enterprise Resource Planning and consists of one or of a set of software applications that integrate information and processes across the several business functions of the enterprise. Typically ERP integrates planning, procurement, sales, marketing, customer relationship, finance and human resources.

ERP systems have the following characteristics:

1. ERP systems are software designed for a client server environment, whether traditional or web/html based.
2. ERP systems integrate the majority of a business's processes.
3. ERP systems process a large majority of an organization's transactions.
4. ERP systems use an enterprise-wide database that typically stores each piece of data once.
5. ERP systems allow access to the data in real time.

ERP software can be customised software, built on purpose for or by the enterprise, but there is also package software designed to perform this function. However, this ERP package software is usually built in a modular way allowing enterprises to customise it for their specific economic activity or size, by implementing only some of those modules.

The main relevance of the measurement of ERP adoption is the integration of business processes and the consequent impact on the productivity of the enterprise.

Experience in other surveys shows that ERP is easy to ask. However, it is not necessarily easy to understand by the respondents.

It is argued that if the respondent doesn't know what ERP is, then he's not using it (just like for most ICT related questions). However, this opinion is not consensual. It depends on what one specifically means by ERP: a software package or the concept of generalised integration of business processes. In the second case it is possible that an enterprise is in such a situation, but doesn't recognise it as ERP. Mentioning of examples of nationally used ERP software packages might also help to recognise the use of ERP.

This question is intended to measure the use of ERP software packages and not of ERP as a concept of generalised integration of business processes.

As such, this question is not a summary of the questions on sharing of information within the enterprise (module E) or sharing/exchanging information with its customers and/or suppliers (modules C, D and F), although they are related.

Questions E1 and E2 respectively measure the integration of business processes (internal and external) looking at the link between the front-office (in contact with customers and suppliers respectively) and other specific business functions: management of inventory levels, accounting, production or services management and distribution management. Question E3 is intended to measure the adoption of a specific type of software packages.

The difference between these two approaches is that with an ERP package the enterprise is supposedly integrating its business processes, but it doesn't provide information about to what extent and which processes it is integrating. As an ERP package is usually made of several modules, the enterprise might be making a very restrictive use of it. On the other hand the enterprise doesn't necessarily need an ERP package to integrate. If the enterprise is small the use of normal office software can provide this integration.

The adoption of ERP software packages is also important because there are simple (free/open source) solutions which might be a good opportunity for enterprises to increase their internal integration of business processes.

Therefore, the approach here is to measure the integration of business processes within the enterprise with several questions.

The question on ERP software should follow questions E1 and E2 on the integration of business processes. The question on ERP adoption shouldn't be filtered by questions E1 or E2, because any change to the specific items of question E1 or to the question E2 would make the results non comparable between different years.

The measurement of integration of business processes should be based on transactions, i.e., from the front office to the back office. Therefore, this question refers to sales **and** purchases.

E4. In January 2008, did your enterprise have in use any software application for managing information about clients (so called CRM) that allows it to:

- a)** Capture, store and make available to other business functions the information about its clients?
- b)** Make analysis of the information about clients for marketing purposes (setting prices, make sales promotion, choose distribution channels, etc.)?

[Scope: enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed per item, i.e. Tick only one per item; binary (Yes/No)]

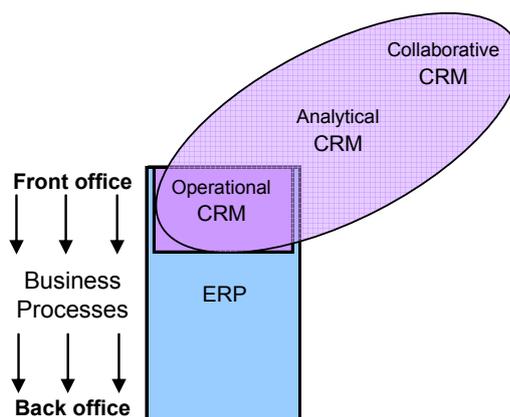
CRM stands for Customer Relationship Management and consists of a management methodology which places the customer at the centre of the business activity, based in an intensive use of information technologies to collect, integrate, process and analyse information related to the customers.

CRM can be considered to be composed of three elements:

1. Operational CRM – Integration of the front office business processes that are in contact with the customer.
2. Analytical CRM – Analysis, through data mining, of the information available in the enterprise on its customers. This aims to gather in depth knowledge of the customer and how to answer to its needs.
3. Collaborative CRM – Adoption of a set of practices aiming to facilitate interaction with the customer, such as after sales support. Collaborative CRM means customer service where the customer and the company agent communicate in real-time with the aid of ICTs. So web co-browsing solutions (where the agent and the customer browse together on the customer's desktop), chat, instant messaging, and various forms of application or desktop sharing can all be considered "Collaborative CRM".

From these three elements it is evident that CRM is broader than just ICT adoption and use. It includes also the creation of new business processes, like data mining on customer information and after sales support. It can be even understood in a broader sense including the restructuring of all the enterprise's business processes in a customer-centric way.

Therefore, CRM is broader in concept than ERP as it includes more than just integration of business processes. ERP on the other hand is broader in extension as it includes the integration of the majority of the business processes, front office and back office, for the majority of the transactions, and not only those related to customer relationship.



Although CRM is more than adoption of ICTs it is based on an intensive use of these technologies. It is believed that the adoption of CRM software improves the marketing functional area of the enterprises, by improving customer service and customer relationship. Improvement comes, for instance, from providing user-friendly mechanisms for receiving complaints, helping identify potential problems before they occur, in general, by facilitating communication with the customer and help track customer interests, purchasing habits and product use. When these technology driven improvements lead to long-term customer satisfaction, they can ensure increased customer loyalty, decreased customer turnover, decrease marketing costs, increase sales revenues and increase profit margins.

Therefore the approach adopted in this variable is to measure the adoption of CRM software tools.

The question is restricted to the first two meanings of CRM, operational and analytical CRM. Collaborative CRM seems not to be a well established term and it is relevant mostly for manufacturing and distribution sectors. Having both first two meanings assembled in one

single question would make it too complex. Therefore, it should be split in two, one focusing on operational CRM and another on analytical CRM (items a) and b) of the question).

I.1.6.6 MODULE F: E-COMMERCE

In order to assure the broadest international comparability of the enterprise ICT usage statistics, the OECD definition of e-commerce is used throughout this module. According to this definition:

An electronic transaction is the sale or purchase of goods or services, whether between businesses, households, individuals or private organisations, conducted over computer-mediated networks. The goods and services are ordered over those networks, but the payment and the ultimate delivery of the good or service may be conducted on or off-line.

Source: OECD Guide to Measuring the Information Society

Compared to the surveys for previous years covers this module not only orders and purchases which are transmitted via Internet, but also via other computer networks. One question (F3) is consecrated to the breakdown between the different types of networks to assure continuity of the series built up in the past.

One important difference in this module compared with most of the questionnaire is that the reference period is the previous calendar year, instead of January of the survey year. Flux economic variables like turnover and purchases, the main variables to measure in e-commerce, need to be measured for a period longer than 1 month. In order to keep comparability with the main business statistics a calendar year is taken as reference period.

The concept of Internet is the same as the one taken in module B. That is, it refers to the public open worldwide network known as Internet and not to the use of the Internet protocol in closed networks.

e-Commerce consists of a business process which is conducted via computer networks. The role of the computer network is very important. It doesn't include the use of this network merely as a human-to-human communication tool. For this reason conventional e-mails manually typed by humans are excluded from this definition. In this case the computer and the network are being used only as a communication device, just like a telephone and its network. However, if the e-mail system is just used for the transmission of an automatic message, i.e. computer-to-computer without human intervention, which represents an order for a product or service, then it is considered an e-commerce transaction.

Therefore orders transmitted from a human to another human in a manual manner through a computer network are not considered e-commerce. The typical ways in which e-commerce orders are placed is: 1) through a web-site, which typically characterises business-to-consumer e-commerce; 2) through automatic data interchange, i.e. using XML, which characterises business-to-business e-commerce; and 3) through B2B e-marketplaces. Nevertheless, any other automatic use of a computer network to transmit orders is included.

F1. Did your enterprise receive orders for products or services via computer networks (excluding manually typed e-mails), during 2007?

[Scope: non-financial (i.e. outside NACE section J) enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. *Tick only one*; binary (Yes/No); filter question]

F2. Please state the value of the turnover resulted from orders received electronically (in monetary terms, excluding VAT), in 2007.

or

Please indicate an estimate of the percentage of the total turnover resulted from orders received electronically, in 2007.

[Scope: non-financial (i.e. outside NACE section J) enterprises which have received orders via computer networks, i.e. F1 = Yes]

[Type: numerical]

Contrary to purchases, sales are usually centralised in enterprises. For this reason, computer network sales are collected in value (percentage or real) and not in percentage classes like it is the case for purchases (see questions F5 and F6).

Order received via computer networks collected in percentage values

Please indicate an estimate of the percentage of the total turnover that resulted from orders received electronically, in 2007.	<table border="1" style="display: inline-table;"><tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr></table> %			

Collecting electronic orders in percentage values has the advantage of allowing us to obtain an estimate from respondents which don't have in the enterprise any records which can provide this value. Therefore, unless specific records of all these transactions are kept in a centralised form, it may be difficult for the enterprise to provide the value of orders received via computer networks.

However, the disadvantage of collecting this variable in percentage values is that for the majority of enterprises this is still a small value. For most of the enterprises it is actually less than 1%. The problem is that this question doesn't allow the respondent to provide such an answer. The answer in this case would have to be either 0% or 1%.

Also, even when making educated estimates the respondents have the tendency to round it to multiples of 10% and 5%. For very small numbers there is evidence that respondents round significantly around 1%. Therefore, when collected in this form, a large proportion of the values collected are 1%.

Another disadvantage of this method is that for big enterprises a value of less than 1% can still be a significant amount. If collected in percentages, computer network sales of less than 1% can either not be taken into account if the respondent answers 0% or be over-estimated. When answers from big enterprises are involved this can bring a significant instability to the results.

Orders received via computer networks collected in real values

Please state the value of the turnover resulted from orders received electronically (in monetary terms, excluding VAT), in 2007.	
	(National Currency)

Therefore, the preferable way to collect orders received via computer networks is in real values. Unfortunately, for a large number of enterprises this value is not available and to make educated estimates, percentage values are better than real values.

Orders received via computer networks collected in real and percentage values

Please state the value of the turnover resulted from orders received electronically (in monetary terms, excluding VAT), in 2007.					
	(National Currency)				
If you can't provide this value,					
Please indicate an estimate of the percentage of the total turnover resulted from orders received electronically, in 2007.	<table border="1"><tr><td></td><td></td><td></td><td>%</td></tr></table>				%
			%		

Therefore, currently the recommended method is to give the alternative to the enterprise to answer in real values if it has the information or, in case such information is not available, ask for an estimation of its percentage in relation to the total turnover.

This intends first to reflect what the practice in several national surveys is already: some countries ask for the real value, while others ask for the percentage value. Second, it intends to provide an alternative to avoid the bias that might exist from asking in each of these ways: underestimation in the case of the real value and overestimation in the case of percentage values, by giving the enterprises the option of using the one for which they can provide more accurate answers.

The real or percentage values collected can be grossed-up to real or the overall percentage. Eurostat preferred practice is to get the grossed-up value of the percentage, i.e. the overall percentage of turnover made via computer networks.

When grossing-up, the values collected as percentages need to be weighted by the total turnover (variable X4). Please refer to chapter I.2.3.3 for more information on the weighting.

F3. Please indicate what percentage represented orders received via each one of the following ways, out of total turnover, in 2007.

[Scope: non-financial (i.e. outside NACE section J) enterprises which have received orders via computer networks, i.e. F1 = Yes]

[Type: numerical]

- a) via a website
- b) via automated data exchange (XML, EDIFACT, etc.) over the internet
- c) via automated data exchange (XML, EDIFACT, etc.) over other computer networks

This question intends to allow continuing the time series that have been built up by the previous surveys. In these surveys – as mentioned at the beginning of this module – orders received via computer networks were collected separately for orders received via Internet and for orders received via other networks. The 100% of orders received via computer

networks should in this question be broken down into three percentage values – adding up to 100%. The sum of item a) and b) can be compared to the e-commerce figures of previous years via the internet whereas item c) continues the series of e-commerce figures via other computer networks.

Like for question F2 should the results be grossed up using as weights the total turnover (variable X4).

F4. Was your enterprise using a secure protocol, such as SSL and TLS, for the reception of orders via Internet, in January 2008?

[Scope: non-financial (i.e. outside NACE section J) enterprises which have received orders via computer networks, i.e. F1 = Yes]

[Type: one single answer needed, i.e. *Tick only one*; binary (Yes/No)]

The measurement of the adoption of secure protocols is related to the measurement of the security level of the online environment for e-commerce over internet. A low level of security functions as a barrier to the further development of e-commerce.

A protocol is a set of rules for exchanging information between two or more computers over a network. Knowing the protocol and having access to the network allows in principal third parties to retrieve and read the messages exchanged. In order to provide secure communications over open networks, e.g. the internet, secure protocols use cryptographic mechanisms to assure that only the recipient of the message can read it and to authenticate the sender.

Secure Sockets Layer (SSL) and Transport Layer Security (TLS), its successor, are cryptographic protocols which provide secure communications on the Internet. SSL provides endpoint authentication and communications privacy over the Internet using cryptography. In typical use, only the server is authenticated (i.e. its identity is ensured) while the client remains unauthenticated. The protocols allow client/server applications to communicate in a way designed to prevent eavesdropping, tampering, and message forgery.

Common applications of the SSL/TLS secure protocols are https and S-HTTP (Secure hypertext transfer protocol). Web browsers use the HTTP protocol to communicate with web servers, sending and receiving information without encrypting it. For sensitive transactions, such as Internet e-commerce, the browser and server need to encrypt this information. The https and S-HTTP were both defined to address this need.

https is a scheme equivalent to the http scheme, originally intended to be used with the HTTP protocol, but with added encryption layer. Instead of using plain text communication, the session data is encrypted using either a version of the SSL (Secure Socket Layer) protocol or the TLS (Transport Layer Security) protocol, thus ensuring reasonable protection from eavesdroppers.

The relevant filter for this question is internet e-commerce, because it is only concerned with security in open networks for e-commerce. Additionally, it should be directed only to enterprises receiving orders via internet. The use of a secure protocol such as SSL/TLS requires an active role of both the computer which sends the message (e.g. order) and the one which receives it (they both need to be prepared to use the protocol). However, the "human" decision to use it lies with the enterprise receiving the orders, while the buyer, e.g. in the case of its use to receive orders via the web-site, will be simply using a browser SSL/TLS enabled, without having to do anything to establish that secure communication.

The question should make reference to internet and not the web-site, because the use of secure protocols is not restricted to the World Wide Web. Also in automated data exchange over the internet the use of secure protocols is fundamental to provide security to e-commerce transactions.

Please note that this question refers – as the only one in this module – only to the internet.

F5. Did your enterprise send orders for products or services via computer networks, during 2007 (excluding manually typed e-mails)?

[Scope: non-financial (i.e. outside NACE section J) enterprises which use computers, i.e. A1 = Yes]

[Type: one single answer needed, i.e. *Tick only one*; binary (Yes/No); filter question]

F6. Please indicate for 2007 the percentage of orders that were sent electronically in relation to the total purchases' value (in monetary terms, excluding VAT).

or

Please state the value of the purchases resulted from orders that were placed electronically (in monetary terms, excluding VAT), in 2007.

If you can't provide this value,

Please indicate an estimate of the percentage of the total purchases that resulted from orders placed electronically, in 2007.

[Scope: non-financial (i.e. outside NACE section J) enterprises which have ordered via computer networks, i.e. F5 = Yes]

[Type: numerical]

The value of the purchases made via computer networks can be collected in 3 different ways. The first one, with the lowest level of measurement, is to collect it in terms of classes of percentage of the total purchases. This is the minimum level of measurement required for this variable. The reason is that in several countries the collection of any other information on this value is not feasible. In this case, the 5 response categories used are:

- Less than 1%
- 1% or more and less than 5%
- 5% or more and less than 10%
- 10% or more and less than 25%
- 25% or more

The main disadvantage of collecting the information in this way is that it doesn't allow to aggregate the value of Internet purchases.

As the aggregated value of purchases via computer networks is one indicator of policy and general interest, alternative ways of collecting this data are recommended when such is possible. Purchases via computer networks should in alternative be collected with a numerical value, either in real terms or in percentage terms.

Purchases via computer networks collected in percentage values

Please indicate an estimate of the percentage of the total purchases that resulted from orders placed electronically, in 2007.	<table border="1"><tr><td data-bbox="1129 190 1166 241"></td><td data-bbox="1166 190 1203 241"></td><td data-bbox="1203 190 1240 241"></td><td data-bbox="1240 190 1276 241"></td><td data-bbox="1300 190 1326 241">%</td></tr></table>					%
				%		

Collecting electronic purchases in percentage values has the advantage of allowing us to obtain an approximate estimate from respondents which don't have in the enterprise any records which can provide this value. Field experience has shown that enterprise purchases are done in a decentralised way. Therefore, unless specific records of all these transactions are kept in a centralised form, it may be difficult for the enterprise to provide the value of purchases via computer networks.

However, the disadvantage of collecting this variable in percentage values is that for the majority of enterprises this is still a small value. For most of the enterprises it is actually less than 1%. The problem is that this question doesn't allow the respondent to provide such an answer. The answer in this case would have to be either 0% or 1%.

Also, even when making educated estimates the respondents have the tendency to round it to multiples of 10% and 5%. For very small numbers there is evidence that respondents round significantly around 1%. Therefore, when collected in this form, a large proportion of the values collected are 1%. This has to be taken into account when the results are tabulated in percentage classes. If 1% is at the limit of the classes used (e.g. [0% - 1% [; [1% - 5% [), then in which class the 1% is included will have a significant impact on the results.

Another disadvantage of this method is that for big enterprises a value of less than 1% can still be a significant amount. If collected in percentages, computer network purchases of less than 1% can either not be taken into account if the respondent answers 0% or be over-estimated. When answers from big enterprises are involved this can bring a significant instability to the results.

Purchases via computer networks collected in real values

Please state the value of the purchases resulted from orders that were placed electronically (in monetary terms, excluding VAT), in 2007.	<table border="1"><tr><td data-bbox="1086 1352 1382 1397">(National Currency)</td></tr></table>	(National Currency)
(National Currency)		

Therefore, the preferable way to collect purchases via computer networks is in real values. Unfortunately, for a large number of enterprises this value is not available and to make educated estimates, percentage values are better than real values.

Purchases via computer networks collected in real and percentage values

<p data-bbox="284 1688 1040 1742">Please state the value of the purchases resulted from orders that were placed electronically (in monetary terms, excluding VAT), in 2007.</p> <p data-bbox="325 1796 609 1823">If you can't provide this value,</p> <p data-bbox="284 1854 1040 1892">Please indicate an estimate of the percentage of the total purchases that resulted from orders placed electronically, in 2007.</p>	<table border="1"><tr><td data-bbox="1086 1715 1382 1760">(National Currency)</td></tr><tr><td data-bbox="1086 1850 1289 1892"><table border="1"><tr><td data-bbox="1098 1850 1134 1892"></td><td data-bbox="1134 1850 1171 1892"></td><td data-bbox="1171 1850 1208 1892"></td><td data-bbox="1208 1850 1244 1892"></td><td data-bbox="1337 1850 1362 1892">%</td></tr></table></td></tr></table>	(National Currency)	<table border="1"><tr><td data-bbox="1098 1850 1134 1892"></td><td data-bbox="1134 1850 1171 1892"></td><td data-bbox="1171 1850 1208 1892"></td><td data-bbox="1208 1850 1244 1892"></td><td data-bbox="1337 1850 1362 1892">%</td></tr></table>					%
(National Currency)								
<table border="1"><tr><td data-bbox="1098 1850 1134 1892"></td><td data-bbox="1134 1850 1171 1892"></td><td data-bbox="1171 1850 1208 1892"></td><td data-bbox="1208 1850 1244 1892"></td><td data-bbox="1337 1850 1362 1892">%</td></tr></table>					%			
				%				

Therefore, currently the recommended method is to give the alternative to the enterprise to answer in real values if it has the information or, in case such information is not available, ask for an estimation of its percentage in relation to the total purchases.

This intends first to reflect what the practice in several national surveys is already: some countries ask for the real value, while others ask for the percentage value. Second, it intends to provide an alternative to avoid the bias that might exist from asking in each of this ways: underestimation in the case of the real value and overestimation in the case of percentage values, by giving the enterprises the option of using the one for which they can provide more accurate answers.

The grossing-up of results will be different depending on how enterprise's purchases via computer networks are collected. If collected in percentage classes, then only the proportion of enterprises per class can be grossed-up. If collected in real or percentage values then real or the overall percentage can be grossed-up. Eurostat preferred practice is to get the grossed-up value of the percentage, i.e. the overall percentage of purchases made via computer networks.

When grossing-up, the values collected as percentages need to be weighted by the total purchases of goods and services (variable X3). Please refer to chapter I.2.3.3 for more information on the weighting.

I.1.6.7 **MODULE G: PERCEIVED BENEFITS OF THE USE OF ICT**

This module intends to measure the (perceived) - positive - effect of the introduction of ICT projects in the eyes of the respondent.

The **implementation of an ICT project** refers to the introduction of a new or updated ICT (e.g. a new/updated software application or a new/updated hardware) or a change in the use of an existing ICT.

Examples of ICT projects are: a new or a restructured website, a new internal homepage, the starting of using automated data exchange or the starting of receiving orders via computer networks.

G1. In January 2008, to what degree have ICT projects implemented in the last 2 years caused improvements in the following areas, compared to the previous task handling?

[**Scope:** enterprises which use computers, i.e. A1 = Yes]

[**Type:** one single answer needed per item, i.e. *Tick only one per item*; answer categories "Minor/none", "Moderate", "Significant", "Don't know/Not applicable"]

If your enterprise has not had any ICT projects, please tick all boxes 'not applicable'.

- a)** Reorganisation and simplification of work routines
- b)** Release of resources
- c)** Higher earnings for the enterprise
- d)** Development of new products and services

The resources mentioned in item b) refer to all types of resources, be it human, financial or others.

I.1.6.8 MODULE X: BACKGROUND INFORMATION

The background variables have several purposes. First, they are used to breakdown the results from the main characteristics. That's the case of the "Main economic activity of the enterprise", "Average number of persons employed" and "Location".

Second, they are needed to weight the percentages of turnover and purchases from e-commerce. The background variables "Total purchases of goods and services" and "Total turnover" are used for that. The variable "Average number of persons employed" is similarly used to weight the percentage of persons employed using computers, the percentage of persons employed using computers with access to the WWW, etc.. The number of persons employed is also used to weight the qualitative variables.²

Third, the background variables are used in the sampling design. Namely, the "Main economic activity" and the "Average number of persons employed" are used to stratify the sample.³

The background variables may be collected through the ICT survey questionnaire or obtained from alternative sources. The alternative sources are mainly the registers and one main business survey, usually used to produce the structural business statistics. It is very important that the background information is at least consistent with the structural business statistics.

X1. Main economic activity of the enterprise

[Scope: all enterprises]

[Type: categorical]

The main (or principal) economic activity is identified as the activity which contributes most to the total value added of the enterprise. The principal activity so identified does not necessarily account for 50% or more of the enterprise's total value added. The classification of principal activity is determined by reference to NACE Rev. 1.1, first at the highest level of classification and then at more detailed levels ("top-down" method).

The nomenclature NACE Rev. 1.1 is available in Eurostat's RAMON database:

<http://europa.eu.int/comm/eurostat/ramon>

The main economic activity of the enterprise should be classified by NACE Rev. 1.1 at its highest level of detail (4 digits). Nevertheless, only the following level of detail which is used in the breakdown is strictly necessary.

NACE groupings		
1	DA + DB + DC + DD + DE	Manufacture of products based on: food, beverages, tobacco, textile, leather, wood, pulp and paper; publishing and printing
2	DF + DG + DH	Manufacture of coke, refined petroleum products, chemical products, man-made fibres, rubber and plastics

² For more information on the weighting refer to chapter I.2.3.3.

³ For more information on the sampling refer to chapter I.2.3.

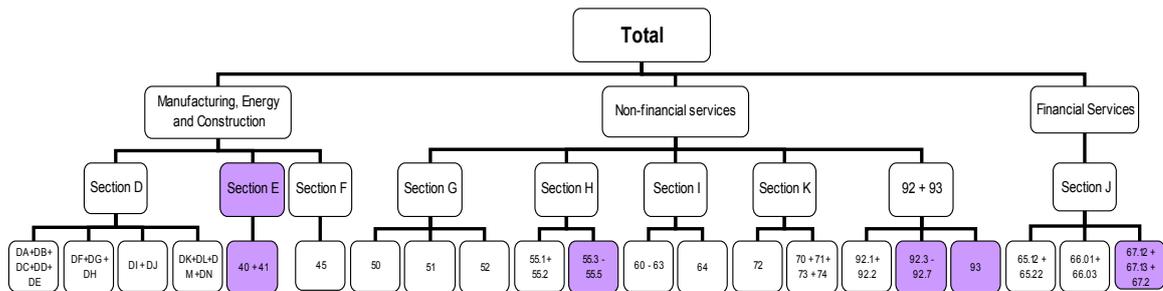
3	DI + DJ	Manufacture of other non-metallic mineral products, basic metals and fabricated metal products
4	DK + DL + DM + DN	Manufacture of machinery and electrical, optical and transport equipment and other manufacturing n.e.c.
5	45	Construction
6	50	Sale, maintenance and repair of motor vehicles and motorcycles; retail sale of automotive fuel
7	51	Wholesale trade and commission trade, except of motor vehicles and motorcycles
8	52	Retail trade and repair of personal and household goods
9	55.1 + 55.2	Hotels and other provision of short-stay accommodation
10	60 + 61 + 62 + 63	Transport and storage
11	64	Post and telecommunications
12	72	Computer and related activities
13	70 + 71 + 73 + 74	Real estate, renting and business activities except computer activities
14	92.1 + 92.2	Motion picture, video, radio and television activities
F1	65.12 + 65.22	Monetary intermediation and other credit granting, except central banking
F2	66.01 + 66.03	Insurance, except compulsory social security
Optional:		
17	22	Publishing, printing and reproduction of recorded media
18	40 + 41	Electricity, gas and water supply
19	55.3 + 55.4 + 55.5	Restaurants, bars, canteens and catering
20	92.3 to 92.7	News agencies, sporting and libraries, archives, museums and other cultural and recreational activities
21	93	Textile washing and dry-cleaning, hairdressing and other beauty treatment, funeral activities, physical well-being activities, etc.
F3	67.12 + 67.13 + 67.2	Activities auxiliary to financial intermediation, except administration of financial markets

These NACE categories are grouped together for dissemination purposes into several aggregates organised in 3 hierarchical levels. At the first level there are 3 categories distinguishing "Manufacturing, Energy and Construction", "Non-financial services" and "Financial Services". At the second level, activities are grouped at the Section level of NACE Rev. 1.1, making 9 categories. Level 3 is the most detailed one, composed of 21 NACE groupings described above.

Level 0	Level 1	Level 2	Level 3
Total	Manufacturing, Energy and Construction	Section D	DA+DB+DC+DD+DE
			DF+DG+DH
			DI+DJ
			DK+DL+DM+DN
		Section E	40 + 41
		Section F	45
	Non-financial Services	Section G	50
			51
			52
		Section H	55.1 + 55.2
			55.3 - 55.5
		Section I	60 - 63
			64
		Section K	72
			70 + 71 + 73 + 74
Section O	92.1 + 92.2		
	92.3 - 92.7		
	93		
Financial Services	Section J	65.12 + 65.22	
		66.01 + 66.03	
		67.12 + 67.13 + 67.2	

Legend:

	Optional category
	Non-optional category



Because part of the NACE categories are optional, and not all countries provide data for for (all) those optional categories, several additional groupings are added to this list excluding the optional categories to allow comparison between the countries; an example of such a grouping is the "total" non-financial business sector i.e. NACE sections D-O, excluding section E, groups 55.3-55.5, section J, groups 92.3-92.7 and division 93. See also the Transmission format in annex I.3.2.

X2. Average number of persons employed

[Scope: all enterprises]

[Type: numerical]

For purposes of general harmonisation of enterprise ICT usage statistics and the more general field of business statistics, the concept of persons employed used here is the one from the Structural Business Statistics (SBS).

The number of persons employed is defined as the total number of persons who work in the observation unit (inclusive of working proprietors, partners working regularly in the unit and unpaid family workers), as well as persons who work outside the unit who belong to it and are paid by it (e.g. sales representatives, delivery personnel, repair and maintenance teams). It includes persons absent for a short period (e.g. sick leave, paid leave or special leave), and also those on strike, but not those absent for an indefinite period. It also includes part-time workers who are regarded as such under the laws of the country concerned and who are on the payroll, as well as seasonal workers, apprentices and home workers on the payroll.

The number of persons employed excludes manpower supplied to the unit by other enterprises, persons carrying out repair and maintenance work in the enquiry unit on behalf of other enterprises, as well as those on compulsory military service.

Unpaid family workers refer to persons who live with the proprietor of the unit and work regularly for the unit, but do not have a contract of service and do not receive a fixed sum for the work they perform. This is limited to those persons who are not included on the payroll of another unit as their principal occupation.

Note: In order to check the comparability of data, it is necessary to indicate whether voluntary workers have been included under this heading or not.

The number of persons employed is a headcount and is measured as an annual average using at least data for each quarter of the year except for the statistics on the activities defined in Section 23 of Annex V, VI and VII of Regulation (EC) No XX/XX for which the calculation can be done on the basis of data with a lower frequency.

Source: Structural Business Statistics Regulation (variable 16 11 0), in process of amendment

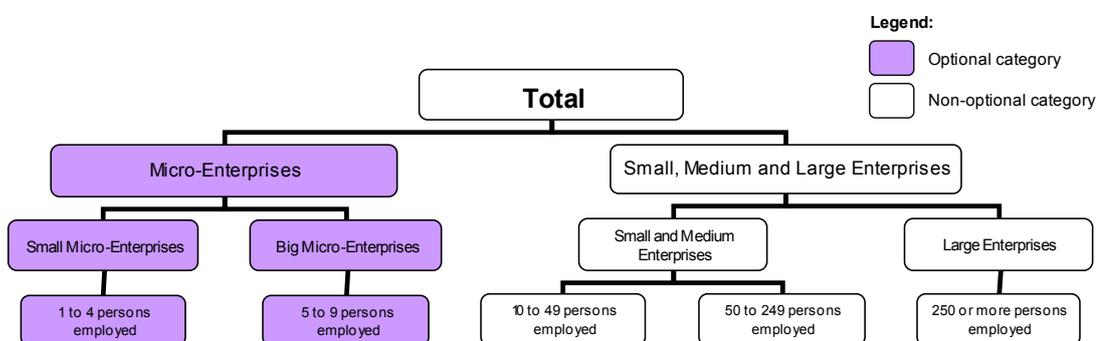
The number of persons employed should be measured as the yearly average during the previous calendar year. For comparability reasons, the number of persons employed should

not be confused with the number of employees (that excludes unpaid workers) or the number of employees in full time equivalent units (fte's).

The average number of persons employed is coded into 5 size categories. For the tabulation of the results, the size categories are grouped in a hierarchical way into 3 levels. The first level distinguishes micro-enterprises from the others. The second and third levels further breakdown these two groups.

Size Categories (persons employed)		
1	10 to 49	Small enterprises
2	50 to 249	Medium enterprises
3	250 or more	Large enterprises
Optional:		
4	Less than 5	Small micro-enterprises
5	5 to 9	Big micro-enterprises

Level 0	Level 1	Level 2	Level 3
Total	Micro-Enterprises	Small Micro-Enterprises	Small Micro - Enterprises
		Big Micro-Enterprises	Big Micro - Enterprises
	Small, Medium and Large Enterprises	Small and Medium Enterprises	Small Enterprises
			Medium Enterprises
	Large Enterprises	Large Enterprises	



X3. Total purchases of goods and services

[Scope: all enterprises, except for those in the financial sector (NACE section J)]

[Type: numerical]

This variable is needed to weight the percentage of purchases resulted from orders placed via computer networks. However, if these variables are collected in classes, instead of percentage (or real) values, then this background variable is not necessary.

As for the number of persons employed and turnover, the SBS definition is to be used:

Purchases of goods and services include the value of all goods and services purchased during the accounting period for resale or consumption in the production process, excluding capital goods the consumption of which is registered as consumption of fixed capital. The goods and services concerned may be either resold with or without further transformation, completely used up in the production process or, finally, be stocked.

Included in these purchases are the materials that enter directly into the goods produced (raw materials, intermediary products, components), plus non-capitalised small tools and equipment. Also included is the value of ancillary materials (lubricants, water, packaging, maintenance and repair materials, office materials) as well as energy products. Included in this variable are the purchases of materials made for the production of capital goods by the unit.

Services paid for during the reference period are also included regardless of whether they are industrial or non-industrial. In this figure are payments for all work carried out by third parties on behalf of the unit including current repairs and maintenance, installation work and technical studies. Amounts paid for the installation of capital goods and the value of capitalised goods are excluded.

Also included are payments made for non-industrial services such as legal and accountancy fees, patents and licence fees (where they are not capitalised), insurance premiums, costs of meetings of shareholders and governing bodies, contributions to business and professional associations, postal, telephone, electronic communication, telegraph and fax charges, transport services for goods and personnel, advertising costs, commissions (where they are not included in wages and salaries), rents, bank charges (excluding interest payments) and all other business services provided by third parties. Included are services which are transformed and capitalised by the unit as capitalised production.

Expenditure classified as financial expenditure or extraordinary expenditure in company accounts is excluded from the total purchases of goods and services.

Purchases of goods and services are valued at the purchase price, i.e. the price the purchaser actually pays for the products, including any taxes less subsidies on the products bought excluding however value added type taxes

All other taxes and duties on the products are therefore not deducted from the valuation of the purchases of goods and services. The treatment of taxes on production is not relevant in the valuation of these purchases.

For the statistics on activities defined in Section 3 of Annexes I to IV, of Regulation (EC) No XX/XX except for the enterprises with an activity classified in NACE Rev.2 Section K, expenditure classified as financial expenditure or extra-ordinary expenditure in company accounts is excluded from the total purchases of goods and services.

For the statistics on activities of NACE Rev.2 groups 65.1 and 65.2, the total purchases of goods and services is defined as the gross value of reinsurance services received plus total of commissions as referred to in Article 64 of Council Directive 91/674 plus any other external expense on goods and services (excluding personnel costs).

For the statistics on activities defined in Section 3 of Annex VI of Regulation (EC) No XX/XX, the total purchases of goods and services is defined as commissions payable plus other administrative expenses plus other operating charges.

Source: Structural Business Statistics Regulation (variable 13 11 0), in process of amendment; see also that Regulation of further details of the link to company accounts.

X4. Total turnover

[Scope: all enterprises, except for those in the financial sector (NACE section J)]

[Type: numerical]

This background variable is needed to weight the percentage of turnover resulting from orders received via computer networks.

As for the number of persons employed and purchases of goods and services, the SBS definition is to be used:

For the statistics on activities defined in Section 3 of Annexes I to IV and VIII of Regulation (EC) XX/XX except for the activities classified in NACE Rev.2 Section K turnover comprises the totals invoiced by the observation unit during the reference period, and this corresponds to market sales of goods or services supplied to third parties. The sales of goods include the goods produced by the enterprise as well as the merchandise purchased by a retailer or land and other property held for resale (if land and other property were initially purchased for investment purposes they should not be included in turnover). The rendering of services typically involves the performance by the enterprise of a contractually agreed task over an agreed period of time. The revenue of the long-term contracts (e.g. building contracts) should be recognised by reference to the stage of completion of the contract and not the finished contract method. Goods produced for own consumption or investment should be excluded from turnover.

Turnover includes all duties and taxes on the goods or services invoiced by the unit with the exception of the value added type taxes (VAT). VAT are collected in stages by the enterprise and fully borne by the final purchaser.

It also includes all other charges (transport, packaging, etc.) passed on to the customer, even if these charges are listed separately in the invoice. Reduction in prices, rebates and discounts as well as the value of returned packing must be deducted.

Income classified as other operating income, financial income and extra-ordinary income in company accounts according to the 4th Accounting Directive and revenue from the use by others of enterprise assets yielding interest, royalties and dividends and other income according to IAS/IFRS is excluded from turnover. Operating subsidies received from public authorities or the institutions of the European Union are also excluded.

For the statistics on activities defined in Section 3 of Annex V of Regulation (EC) No XX/XX, the corresponding title of this characteristic is 'Gross premiums written'. This characteristic is defined in Article 35 of Directive 91/674/EEC⁴. Note: For the layout of the profit and loss account (technical account): Article 34, I. 1(a) of Directive 91/674/EEC for non-life insurance business and Article 34, II. 1(a) of Directive 91/674/EEC for life insurance business.

For the statistics on activities defined in Section 3 of Annex VII of Regulation (EC) No XX/XX, the corresponding title of this characteristic is 'Total pension contributions'. This characteristic shall comprise all pension contributions, due during the financial year, in respect of pension contracts, such as all mandatory contributions, other regular contributions, voluntary additional contributions, incoming transfers, other contributions.

Source: Structural Business Statistics Regulation (variable 12 11 0), in process of amendment; see also that Regulation of further details of the link to company accounts.

X5. Location (Convergence/non-Convergence region)

[Scope: all enterprises]

[Type: categorical]

The rationale of the **Convergence objective** is to promote growth-enhancing conditions and factors leading to real convergence for the least-developed Member States and regions. In EU-27, this objective concerns – within 17 Member States – 84 regions with a total population of 154 million, and per capita GDP at less than 75 % of the Community average, and – on a “phasing-out” basis – another 16 regions with a total of 16.4 million inhabitants and a GDP only slightly above the threshold, due to the statistical effect of the larger EU.

⁴ OJ L 374, 31.12.1991, p. 7

Since 2007, the previous list of objective 1 regions is substituted by the new list of convergence regions (including related phasing-out regions) eligible for funding from the Structural Funds.

The NUTS 2 regions (statistical demarcation) eligible for support from the Structural Funds under the convergence objective - including phasing-out regions - has been decided for the whole period 2007-2013. No update of this list is foreseen in the Structural Funds Regulation. The list is shown below.

Please note that the list is different for the one used up till the 2006 survey.

Countries *entirely* eligible under the Convergence objective:

Bulgaria

Estonia (the whole country counts as one single region at NUTS2)

Latvia (the whole country counts as one single region at NUTS2)

Lithuania (the whole country counts as one single region at NUTS2)

Malta (the whole country counts as one single region at NUTS2)

Poland

Romania

Slovenia (the whole country counts as one single region at NUTS2)

Countries *partially* eligible under the Convergence objective:

Belgium: Hainaut

Czech Republic: Střední Čechy, Jihozápad, Severozápad, Severovýchod, Jihovýchod, Střední Morava, Moravskoslezsko (i.e. the whole country except Praha)

Germany: Brandenburg Nord-Ost, Brandenburg Süd-West, Mecklenburg-Vorpommern, Lüneburg, Chemnitz, Dresden, Leipzig, Dessau, Halle, Magdeburg, Thüringen

Greece: Anatoliki Makedonia, Thraki, Kentriki Makedonia, Dytiki Makedonia, Thessalia, Ipeiros, Ionia Nisia, Dytiki Ellada, Peloponnisos, Attiki, Voreio Aigaio, Kriti

Spain: Galicia, Principado de Asturias, Castilla-La Mancha, Extremadura, Andalucía, Región de Murcia, Ceuta, Melilla

France: Guadeloupe, Martinique, French Guyana, Réunion

Italy: Campania, Puglia, Basilicata, Calabria, Sicilia

Hungary: Közép-Dunántúl, Nyugat-Dunántúl, Dél-Dunántúl, Észak-Magyarország, Észak-Alföld, Dél-Alföld

Austria: Burgenland

Portugal: Norte, Algarve, Centro, Alentejo, R. A. Açores

Slovakia: Západné Slovensko, Stredné Slovensko, Východné Slovensko (i.e. the whole country except Bratislavský kraj)

United Kingdom: West Wales and the Valleys, Cornwall and Isles of Scilly, Highlands and Islands

Countries *entirely not* eligible under the Convergence objective:

Denmark (the whole country counts as one single region at NUTS2)

Ireland

Cyprus (the whole country counts as one single region at NUTS2)

Luxembourg (whole country counts as one single region at NUTS2)

The Netherlands

Finland

Sweden

The list of regions eligible under the Convergence objective was published in OJ L 243/44 (6.9.2006), "Commission Decision of 4 August 2006 drawing up the list of regions eligible for funding from the Structural Funds under the Convergence objective for the period 2007-2013"

This variable is not relevant for countries that are not member of the European Union. Such countries (EEA countries, candidate countries and other countries participating in the survey) are considered as ***entirely not*** eligible under the Convergence objective and transmit data accordingly.

I.2. PRODUCTION METHODOLOGY

I.2.1. Timetable – Survey period and reference period

The survey period of the Community Survey on ICT Usage and e-Commerce in Enterprises is the first quarter of calendar year t. The variables in the questionnaire are predominantly of a qualitative nature (yes-or-no-questions or tick-box-questions). The answers to these questions refer to the situation in January of year t. The answers to the quantitative questions - for example the percentage of total turnover resulting from orders received via computer networks - refer to the calendar year t-1. Due to national planning, for some countries the reference period still deviates from the one of the model questionnaire; this might affect the comparability.

In order to minimize the recall-bias of the respondent the collection should occur just after the reference period, so after January during the first quarter. However, some countries have difficulties in collecting monetary data (turnover and purchases) right in the beginning of the year. In these cases, the collection should start as soon as possible. Preferably still during the first quarter.

To maximize the level of harmonization between the national surveys in the different Member-States, it is important that this survey period is respected as much as possible.

Table: Overview of the survey period⁵ and reference period (2007 survey)

Source: Quality reports for the Community Survey on ICT Usage and e-Commerce in Enterprises (2007)

Country	Data collection period	Reference period
Belgium	9/5/2007-31/8/2007	as in model questionnaire
Bulgaria	1/6/2007-31/7/2007	as in model questionnaire
Czech Republic	1/3/2007-29/8/2007	as in model questionnaire
Denmark	26/2/2007-30/6/2007	as in model questionnaire
Germany	March 2007-June 2007	as in model questionnaire
Estonia	1/3/2007-June 2007	as in model questionnaire
Ireland	9/3/2007-20/9/2007	as in model questionnaire
Greece	March 2007-September 2007	as in model questionnaire
Spain	January 2007-July 2007	as in model questionnaire
France	6/11/2006-28/3/2007	2005 for background variables instead of 2006, last month instead of January 2007
Italy	15/3/2007-31/8/2007	as in model questionnaire
Cyprus	22/1/2007-18/5/2007	as in model questionnaire
Latvia	22/1/2007-23/3/2007	as in model questionnaire
Lithuania	January 2007-5/6/2007	as in model questionnaire
Luxembourg	26/3/2007-10/7/2007	as in model questionnaire
Hungary	January 2007-30/4/2007	as in model questionnaire
Malta	February 2007-April 2007	as in model questionnaire
Netherlands	1/3/2007-11/7/2007	2006 instead of January 2007
Austria	22/2/2007-29/5/2007	as in model questionnaire
Poland	30/3/2007-1/8/2007	as in model questionnaire

⁵ The periods mentioned refer to the general survey; for some countries differences can occur for the survey for the financial sector and/or for the micro-enterprises in case a separate survey is organised.

Portugal	24/4/2007-28/9/2007	as in model questionnaire
Romania	July 2007-August 2007	as in model questionnaire
Slovenia	12/2/2007-4/5/2007	as in model questionnaire
Slovak Republic	20/2/2007-31/5/2007	as in model questionnaire
Finland	16/2/2007-June 2007	as in model questionnaire
Sweden	19/3/2007-6/9/2007	as in model questionnaire
United Kingdom	2/1/2007-31/8/2007	December 2006 instead of January 2007
Norway	14/2/2007-31/5/2007	as in model questionnaire
Serbia	22/5/2007-5/6/2007	as in model questionnaire
Croatia	15/12/2007-28/2/2008	as in model questionnaire

I.2.2. Frame population

Normally, the enterprise ICT survey will be carried out in the form of a sample survey. The *frame population* (or *sampling population*) is the list of enterprises from which the sample will be drawn. Ideally, this list of enterprises should be equivalent to the target population as both over-coverage and under-coverage can induce bias and affect the reliability of the survey results.

E.g. over-coverage due to inclusion of enterprises which main economic activity is agriculture (outside the scope of the Community Survey) in the sampling frame: if this sector shows a below the average adoption of ICT, the population total may show a downwards bias due to inclusion of this group.

E.g. under-coverage due to the fact that some remote areas are not covered in the register: as such areas usually have a lower penetration of ICTs, the overall indicators may be relatively overestimated for such a country.

The sample for the Community Survey should be drawn from the business register in the different Member-States as defined in Council Regulation (EEC) No 2186/93. Part of this register is the activity code at the four-digit (class) level of NACE Rev.1.1, the size measured by the number of persons employed and the geographical location code (territorial units) of the enterprises.

The choice to comply with the enterprise as the appropriate statistical unit and the business register and its enterprise' characteristics as the framework population, is inspired by the wish to keep a certain level of consistency with the Structural Business Statistics. On a meso-level as well as on a micro-level results of the Structural Business Statistics will be combined with data on the ICT usage of enterprises for analysis and for weighting-purposes (for the latter: see paragraph I.2.3.3).

To allow for an assessment, it is important to include the coverage and the sampling frame when reporting metadata (see Chapters 3 and 5 in the reporting template, Annex I.3.3).

The table below presents the sampling frames used in the national statistical institutes for the 2007 survey.

Table: Overview of the sampling frames (2007 survey)

Source: Quoted from the Qualityreports for the Community survey on ICT usage and e-commerce in enterprises (2007)

Belgium	SBS frame population
Bulgaria	As frame population is used the Business Register of reported enterprises in 2005 at NSI. There were extracted all enterprises responding to the respective NACE classes and employment groups – 10 to 49; 50 to 249 and 250 and more persons employed. The total number of the frame population is 23228 enterprises. Before starting the interviews the sample was updated by the Regional statistical offices on the base of the reported enterprises for 2006.

Czech Republic	The frame population is the same as the one for the Structural Business Survey (SBS).
Denmark	The frame population corresponds with SBS.
Germany	The sampling base for the ICT-Enterprise survey is the current business register.
Estonia	The population described in the ICT-survey is drawn from the same business-register as the SBS. Differently from SBS the enterprises which may start the activity in 2007 are included and enterprises which deceased in 2006 are excluded. In addition for national needs the survey will cover active enterprises with 10 and more persons employed with NACE Rev. 1.1 activities 01, 02, 05, 10, 11, 14, 15, 80, 85, 90 65.21, 65.23, 67.11. There are no enterprises in Estonia in activities 12, 13, 16, 66.02.
Ireland	Central Business Register
Greece	The frame population is the one used for the SBS. There was not any deviation from the recommendation of the model survey.
Spain	The frame population of the ICT survey is the same as the one used for the SBS: DIRCE. The Central Businesses Directory (DIRCE) collects all Spanish businesses in a single directory. Its basic objective is to enable business-targeted sample surveys to be conducted, and consequently, it registers information such as identity data, location, main activity or number of employees. This information is obtained from administrative sources (Inland Revenue and Social Security) and complemented with data from common statistical operations. Moreover, this directory is annually updated.
France	Whole French enterprise population taken from SBS survey (for major units) and French business register (for financial enterprises).
Italy	Enterprises with 10 and more persons employed, included as "operating" in business register.
Cyprus	The frame population is the Registration of Establishments 2005 which is updated on the basis of the Employment Survey and other resources so as to take into account changes in the structure of employment by industry, occupation and size of establishment. The frame population is used for the Structural Business Statistics too.
Latvia	Frame population consists of units which are economic active at the end of 2006, belongs to defined NACE groups. Frame population consists of: <ul style="list-style-type: none"> • Individual merchants; • Commercial companies; • Peasant and fishermen's farms. Frame population used for the Structural Business Statistics is broader. Frame population used in Community Survey on ICT Usage and e-Commerce in Enterprises is subset of frame population used for the Structural Business Statistics.
Lithuania	Statistical units in the survey which are economically active enterprises as defined in the Statistical Profile Business Register. No deviation from the recommendation of the Community model survey.
Luxembourg	For the sample design, the enterprises register was used.
Hungary	The frame of the survey embraces the following economic branches. A) Agriculture and forestry, B) fishing, ..., O) Miscellaneous communal and personal services. The underlying register concerning the target population is the Register of Economic Enterprises. The observation and sampling units consist of those economic enterprises that have an obligatory data transmission status in December 2006 concerning the following surveys: "Simplified monthly report on the industrial activities", "Simplified monthly report on the construction activities", "Simplified monthly report on the agricultural, trading and service activities". All these enterprises build up the frame of the survey, i.e. the sampling frame. The observation units of the survey consist of the enterprises that belong to the chosen sample elements concerning the above mentioned surveys.
Malta	The frame population is taken from the Business Register within NSO.
Netherlands	The population described in the ICT-survey is drawn from the same business-register as the survey for the Structural Business Statistics and at the same reference date i.c. 1-12-2006.
Austria	The frame population is the number of enterprises in the National Business Register. SBS use the same source for sampling.

Poland	<p>The Statistical Unit Database (BJS) functions as a sampling frame for all surveys conducted by the Central Statistical Office (CSO) on the population of Polish economic entities and to extrapolate sample survey results.</p> <p>Frame population is the same as in the "Eurostat model for a Community survey on ICT usage and e-Commerce in enterprises 2007 ver. 3.1"</p> <p>Frame population includes: Sections D (15-22, 23-25, 26-28, 29-37), F(45), G (50, 51), H (55,1, 55.2), I (60-63, 64), J (65.12, 65.22, 66.01, 66.03) K((70,71, 74), 72, 73) and O (92.1 and 92.2).</p> <p>The frame population doesn't include:</p> <ul style="list-style-type: none"> • optional NACE divisions and groups: 22, 40, 41, from 55.3 to 55.5 from 92.3 to 92.7 and 93 • optional NACE classes and group in J section: 67.12, 67.13, 67.2 • enterprises employed less than 10 persons
Portugal	<p>The frame population is based on the reference frame population of the structural business statistics and entirely respects the recommendations of the model questionnaire regarding the economic activity, the enterprise size and the geographic scope.</p>
Romania	<p>The sample frame consists in: the active enterprises at 31 December 2005 that have as main activity included in NACE aggregation mentioned before; the new active enterprises created in 2006.</p> <p>The sampling frame has drawn from the Business Register – REGIS, which have characterised by the conditions mention above concerning activity and size class.</p>
Slovenia	<p>The sampling frame was made by using the data from the Business Register of the Republic of Slovenia and the data from some other surveys. Regarding activity, NACE sections D-O for the enterprise survey and NACE sections from 65-67 for the financial enterprises were included. Regarding the number of employees, units with at least 5 employees for the enterprise survey and units with at least 10 employees for the financial enterprises were included.</p>
Slovak Republic	<p>Frame population is based on the same Register of statistical units as in case of structural business statistics.</p>
Finland	<p>The frame population is drawn from the production database of Structural Business Statistics.</p>
Sweden	<p>The sampling frame is the Swedish Business Register.</p>
United Kingdom	<p>Whole UK economy taken from the ONS business register, based on SBS.</p>
Norway	<p>The frame population is based on the Structural Business Statistics.</p>
Serbia	<p>Frame population consist of enterprises with 10 or more employees.</p>
Croatia	<p>Sampling frame was designed by using data from the Business Register of the Central Financial Agency. All enterprises with any income are part of the register.</p>

I.2.3. Sampling design

The survey should be based on a probability sample from which results representative of the population could be derived, considering the agreed breakdowns defined in the questionnaire.

The sampling design and the resulting sample size (see below) should be appropriate for obtaining accurate, reliable and representative results on the survey characteristics and breakdowns specified in the Regulation and the model questionnaire.

This objective should be achieved for the overall proportions as well as for the proportions relating to the different subgroups of the population. The estimated standard error for any indicator/variable (be it a proportion or a total) should not exceed 2 percentage points for the overall population and should not exceed 5 percentage points for to the different subgroups of the population.

The aim of such requirement is to ensure the collection of a complete dataset – without empty, confidential or unreliable cells - for these indicators. In case this aim can not be reached, clarifications should be provided (see also I.2.7.3).

Table: Overview of the sampling design and sampling methods (2007 survey)

Source: Quoted from the Qualityreports for the Community Survey on ICT Usage and e-Commerce in Enterprises (2007)

Belgium	<p>General survey: Stratified random sample. Stratification by Activity (17 categories), Size class (4 categories) and NUTS1 region (3 categories), resulting in 204 strata. Exhaustive selection in strata with less than 5 enterprises. Oversampling of smaller regions. Sample selection = SRS algorithm in SPSS.</p> <p>Financial sector: Stratified random sample. Stratification by Activity (2 categories) and Size class (4 categories), resulting in 8 strata. Sample selection = SRS algorithm in SPSS.</p> <p>No procedures for the coordination of non-overlapping with other samples.</p>
Bulgaria	<p>The enterprises with 250+ persons employed are covered exhaustively. In order to ensure representative results with sufficient stochastic accuracy, from the group 10 - 249 persons employed was emanated one-stage proportionally stratified random sample at national level.</p> <p>The gross sample size is 4495 enterprises, of which: 1861 enterprises with 10 to 49 persons employed, 1962 enterprise with 50 to 249 persons employed and 672 enterprises with 250 or more persons employed.</p> <p>Before starting the interviews the sample was updated by our Regional statistical offices on the base of the reported enterprises for 2006.</p>
Czech Republic	<p>Czech Business Register for sampling was used.</p> <p>Stratified random sampling for enterprises with 5-249 and census for 250 and more employees with respect to Business Register was used. An intent sampling was used for enterprises with large share of e-commerce or with huge turnover according to its size.</p> <p>Number of enterprises and number of employees was used for stratification. GREG was used for sample design. The sample was designed with no reference to any other survey.</p> <p>Sample size was designed to enable accurate, reliable and representative results for breakdown into four size groups, 32 NACE aggregates groups and two territory groups and for main variables as a cross-tabulated.</p>
Denmark	<p>The sample is a stratified random sample made by use of register-based data. Strata were made by employment (10-19, 20-49, 50-99, 100+) and activity (Nace, 2-digit).</p>
Germany	<p>For the sampling a two-stage stratified random sample was used. The variables of stratification were the Federal State, the NACE-category and the number of employees (including the "0"-Stratum). The final number of strata amounted to 1,260.</p> <p>At first stage the total sample size was allocated to the variables Federal State and NACE-category using the "Methode der vergleichbaren Präzision".</p> <p>At second stage the optimal allocation by Neyman-Tschuprow was applied for the variable number of employees in each stratum (Federal State x NACE-category).</p> <p>For the „0“-Stratum a fixed number of three enterprises per stratum (Federal State x NACE-category) was chosen. For most cases this corresponded to an inclusion probability of less than 1%.</p>
Estonia	<p>The stratified simple random sampling method is used. The sample is stratified by economic activity and by number of persons employed. The enterprises were divided by number of persons employed into following size groups: 10-19, 20-49; 50-99; 100-249; 250-499; 500-999; 1000+. The number of strata is 430</p> <p>The Neyman optimal allocation was used for sample selection and determination of sample size. By number of persons employed, the survey includes all enterprises with 50 and more persons employed and 10 and more persons employed in financial sector. Stratified simple random sampling is used for enterprises 10 to 49 persons employed.</p> <p>Sample was drawn using permanent random numbers. The choice of starting point between 0 and 1 guarantee non-overlap with two large sample surveys: wages survey and SBS survey.</p>
Ireland	<p>Stratified Random Sample</p>
Greece	<p>The sampling method applied for all enterprises (micro enterprises as well as enterprises with number of employees 10+) was the single stratified random sampling, employing the enterprise as the survey unit.</p> <p>The sampling frame used was based on the Business Register of the N.S.S.G.</p> <p>The enterprises included in the survey were stratified:</p> <ul style="list-style-type: none"> • By geographic region – NUTSI • By aggregations of NACE Rev.1 subsections, as specified in the relevant contract, within each geographical region • By size class of the enterprise. In each one of the major strata (major stratum = Geography x Economic Activity), the enterprises were stratified into L=7 size classes, according to their size, determined by their average annual number of employees in the business register, as follows.

Size Class	Number of employees
Size class 1	1 – 4
Size class 2	5 - 9
Size class 3	10 - 19

Size class 4	20 - 49
Size class 5	50 - 99
Size class 6	100 - 249
Size class 7	250 +

Let h be one of the final strata (Final stratum = Geography X Economic Activity X Size Class). The final strata, that contain size class with $L=6,7$, are census strata (take-all), because they comprise the large size units and the element population variance is high. In each of the final strata, a sample of n_h enterprises was selected from the total number of enterprises N_h with equal selection probabilities, by applying the systematic sampling method.

In the final strata, the distribution of sampling units was conducted by applying the method of Neyman (optimal) allocation.

The sample size is 4.896 enterprises (sampling fraction 3,1%). 4.022 enterprises out of the 4.896 have a number of employees 10+, while the rest 874 are microenterprises. 1.430 enterprises belong to the census size classes, while the remaining 3.466 belong to the first five size classes (surveyed on a sample basis).

The whole sample size is defined so that the results to reflect the true situation of the total population units with a sufficient degree of representativeness.

Spain

According to the frame population (extracted from the DIRCE), the statistical units are stratified by crossing the following variables: size, economic activity and location. In every stratum, a systematic selection randomly started is executed, sorting out the enterprises by size and location.

The categories of the variables used to stratify are:

- size:
 - 0 to 2, 3 to 9 (or enterprises with less than 10 employees), and,
 - 10 to 19, 20 to 49, 50 to 99, 100 to 199, 200 to 499, and 500 or more (for enterprises with 10 or more persons employed)
- main activity: following the aggregations of NACE classification provided in previous sections, except for micro-enterprises (see details in the table above in section 5.3)
- region: all Spanish regions (18) are covered.

The total number of initial strata for the general questionnaire is 1.728 ($16 \times 6 \times 18$), including empty strata for enterprises with 10 or more persons employed and 360 ($10 \times 2 \times 18$) for enterprises with less than 10 persons employed.

France

The following strata are used based on reported employment, held on each statistical source.

- 10- 49, random sampling
- 50 - 249, random sampling
- 250 - 499, random sampling
- 500+ Fully enumerated strata

Coordination with the sampling reference year 2006 in order to have half of the sample of the 2006 survey.

71 firms who are well known for electronic business are introduced in the sample, with a weight equal to 1.

Italy

Stratified random sample: stratification by economic activities (24 for general survey), by four size classes (10-49; 50-99; 100-249; 250 and more) and by geographical breakdown (21 administrative regions). Total number of strata: $24 \times 4 \times 21$ (general survey).

It was used procedure for the coordination and non-overlapping with samples of other surveys.

Financial survey is a total survey.

Cyprus

The sampling method used was stratified random sampling. Two variables were used for stratification, nace group and size. There were 14 nace groups and 3 size groups.

Due to the small number of large and medium enterprises it has been decided to use all those enterprises in the sample. Half of the small enterprises were selected systematically from a list ordered in ascending order by enterprise size.

The size of the sample, i.e. 1557 enterprises, could satisfactorily serve the desirable and acceptable confidence limits of the survey results, taking always into consideration time and cost constraints.

The total number of all financial enterprises falling in the 2 NACE groups as below (F1, F2), for the size classes 10-49, 50-249, 250+ is small (120 enterprises), thus all enterprises are used.

Latvia

Stratified simple random sampling is used as sampling design. In each NACE group the "big" enterprises are defined:

- All enterprises with 50 or more persons employed;
- Biggest enterprises by turnover covering 50% of total turnover of corresponding NACE group.

"Big" enterprises are sampled with sampling fraction 100%.

Other ("small") enterprises are stratified by NACE groups, turnover marking sign (1 or 0), specify if are known turnover information and two groups defined by number of persons employed:

- 10 to 19 persons employed;
- 20 to 49 persons employed.

Sample allocation for "small" enterprises is computed by taking into account population size in strata and size of enterprises (by turnover) in strata. The sample size is chosen so that design weights are not exceeding 35. The number of strata for "small" enterprises is 43.

The procedures for the coordination are applied – minimum sample size in strata is three enterprises.

Lithuania

Stratification and in each strata simple random sampling was used. The number of employees was used to stratify the population. The final number of strata is 63.

The Neyman optimal allocation was used for sample selection and determination of the sample size. If stratum size is 30 or less then all enterprises in this stratum were included.

Luxembourg

A census is carried out.

Hungary

The sample selection is being done with a stratified sampling scheme. The strata are formed in the following way.

1. Concerning the NACE-codes all the 2-digit level branches are being treated as a different stratum. We break down some 2-digit level branches to smaller sub-branches, if there is a need for a separate treatment considering the importance or the size of these areas. This is being done by this way.

Branch	Sub-branches
14	14.21 and others
22	22.22 and others
28	28.11 and others
29	29.24 and others
36	36.14 and others
37	37.10 and others
45	45.1, 45.21, 45.25, 45.22-45.24, 45.3, 45.4, 45.5
50	50.10 and others
51	51.70 and others
55	55.1-55.2, 55.30, others from 55.3-55.5
60	60.24 and others
63	63.40 and others
80	80.42 and others

2. Concerning the number of employees the following categories are being treated separately: 30 (10-19 employees), 40 (20-49 employees), 50 (50-99 employees), 61 (100-149 employees), 62 (150-199 employees), 70 (200-249 employees), 81 (250-299 employees), 82 (300-499 employees), 91 (500-999 employees), 93 (1000-1999 employees), 94 (2000-4999 employees), 95 (5000- employees).
3. The capitol Budapest and the provinces are forming different strata.

The distribution of the sample among the above defined strata is being done by the following. By the cooperation with the informatics department we know the number of enterprises belonging to the sample frame in all the single strata. This information is available for the time periods January 2007 and January 2008 (in the latter case provisional data are available).

The final sample sizes are determined by the well-known Neumann-type optimal allocation. For this procedure we make use of the information available for the previous time-period concerning the stratum level statistics for different variables.

The process of sample selection is done by the following. A random number is being assigned to all the enterprises in the sampling frame. After that we add 1 or 2 to the value of some enterprises in order to give preference to some of them. Then the process of sample selection is being done by ordering the enterprises by the assigned values and then those possessing the biggest values are chosen.

Malta	For size class fte 10-49, 50-249 and 250 and over, all enterprises were covered. A random sample was used for micro enterprises.
Netherlands	<p>The target-variables (of year t-1) used to allocate the sample were:</p> <ul style="list-style-type: none"> - the number of enterprises receiving orders online (a qualitative variable) - the number of fte's spend on developing own account software (a quantitative variable). <p>The population/sample was divided into 48 NACE-aggregates and 10 groups of size-classes resulting in 480 cells/strata.</p> <p>Ca. 11 000 enterprises were allocated in a way that on a national level the relative confidence-intervals for the target-variables were as small as possible. Enterprises with 250 or more employed persons were integrally included in the sample. By choosing a maximum number of 11 000 enterprises to be allocated a response percentage of 50 percent is anticipated. In principle the Neyman-allocation was used to allocate the 11 000 enterprises over the different strata, meaning that the allocation is inversely proportional to the variance of the target-variable in the relevant strata.</p> <p>A rotation-fraction of .5 was used, meaning that 50 percent of the sample - if possible - exists of enterprises which also in the year t-1 were included in the sample. A system to 'spread' the administrative burden for enterprises is used to avoid - if possible - that certain enterprises receive a lot of questionnaires in year t and others none or just a few. So, the overlap with other surveys in the same period e.g. the structural business statistics survey is minimized.</p>
Austria	<p>Sampling was carried out as stratified random sampling. Two dimensions were used as stratification variables (16 x 3 = 48 strata):</p> <ul style="list-style-type: none"> a) Main economic activity (16 strata) 15-22, 23-25, 26-28, 29-37, 45, 50, 51, 52, 55.1+55.2, 60-63, 64, 65.12+65.22, 66.01+66.03, 72, 70+71+73+74, 92.1+92.2 b) Size classes (3 strata) 10-49 employed persons, 50-249 employed persons, 250 and more employed persons <p>A full census was carried out among enterprises with 250 and more employed persons. There were higher inclusion probabilities for the following NACE sections: 64, 65.12+65.22, 66.01+66.03, 92.1+92.2.</p>
Poland	<p>General sampling scheme was proportional stratified sampling design - that is one tried to include same proportion from each of 14*3 strata. However there were several exceptions. Firstly all large enterprises that is having more than 249 employees (2736 units) were included into the sample wholly. Secondly all units from the last year sample that existed in 2006 and were not present in 2004 year sample were included into the sample. There were 4788 such units. Thirdly due to non uniform distribution of units between different strata in the frame, some strata that contained small number of units were also included into the sample. There were 1060 such units. The remaining 5418 were allocated proportionally as mentioned above.</p>
Portugal	<p>Stratified sample according to the following variables and breakdowns:</p> <ol style="list-style-type: none"> 1) NACE Rev.1.1 (two digit levels except when indicated other level of aggregation) NACE Rev.1.1 1 C (10+11+12+13+14) 2 DA - DE (15+16+17+18+19+20+21+22) 3 DF - DH (23+24+25) 4 DI - DJ (26+27+28) 5 DK - DN (29+30+31+32+33+34+35+36+37) (except 30010; 30020; 31300; 32100; 32200; 32300; 33201; 33202; 33203; 33300) 6 30010 + 30020 + 31300 + 32100 + 32200 + 32300 + 33201 + 33202 + 33203 + 33300 7 E (40+41) 8 45 9 50 10 51 (except 51430; 51840; 51850; 51860; 51870) 11 51430 + 51840 + 51850 + 51860 + 51870 12 52 13 551+552 14 60+61+62+63 15 64 (except 64200) 16 64200 17 72 18 70+71+73+74 (except 71330) 19 71330 20 921+922 21 6512+6522 22 6601+6603 23 6712+6713+6720 24 65 (except 65.12 e 65.22) 25 66 (except 66.01 e 66.03) 26 67 (except 67.12, 67.13 e 67.20) 2) Number of employed persons (0; 1-4; 5-9; 10-49; 50-249; 250 and more) 3) Turnover (up to 5000000€; [500000€; 25000000€]; 25000000€ and more)

General remarks:

- sample size according to the Neyman-Allocation;
- definition of at least five enterprises by strata;
- a priori CV is 5% for the economic activity strata and for the employed persons strata;
- sample selection according to a sequential sampling;
- for large enterprises (enterprises with 250 and more employed persons or enterprises with a turnover $\geq 25000000\text{€}$) a census is adopted

Slovenia

Design of the sample was stratified systematic sample. Activity group and size class were used as stratification variables. In addition, implicit stratification by 5-digit NACE group was used. No procedures for coordinated samples were applied.

Slovak Republic

The sample is created on the base of the same methods, which are used in other surveys organized by SO SR (short-term surveys, structural business surveys). In comparison with other surveys the boundary between exhaustive and sample survey is changed in the direction upwards. In this enterprise survey the sample is organized in enterprises with 1 to 249 employees, whereas in STS and SBS the sample concerns the enterprises with 1 to 19 employees.

Stratification sampling is realised: - NACE, - NUTS3, - Number of employees

Finland

Stratified random sample (in enterprises with at least 100 employees a census). The variables for stratification were the size of the enterprise (as number of persons employed) and economic activity.

- The categories of size were 5: 5-9, 10-19, 20-49, 50-99 and 100+ (in financial sector no 5-9 enterprises).
- The categories of economic activity were 29: as NACE classes 15-19, 20, 21, 22, 23, 24-25, 26, 27-28, 29-33, 34-35, 36, 37, 40, 45, 50, 51, 52, 55.1-55.2, 55.3-55.5, 60-63, 64, 65.12+65.22, 66.01+66.03, 70-71, 72, 73-74, 92.1+92.2, 92.4+92.71, 93.

All together this is 143 strata.

All enterprises in the sample frame employing at least one hundred persons were included in the inquiry. Sampling was applied to enterprises with five to 99 employees. As a starting point in constructing the sample, one half of the enterprises employing 50 to 99 persons, one fifth of the enterprises with 20 to 49 employees and ten per cent of those employing 5 to 9 and 10 to 19 persons were selected by size category from the strata. This preliminary sample structure was still modified by strata to include more enterprises where the number of enterprises would be small and analytical purposes relating to economic activity were foreseen to need more enterprises to the sample. Also in strata where the number of enterprises was regarded large relative to the analytical needs the sample size was reduced. The quality of results by analytical unit from the previous survey were used in fine tuning the sample size by strata.

In financial sector all enterprises are covered, except in two smallest size groups of 10-19 and 20-49 employees in NACE 65 a sample is drawn.

Sweden

Stratified sampling with simple random sampling (without replacement) within strata is used. The stratification is done by economic activity (industrial strata) and size (number of employees).

Economic activity: NACE 15-22, 23-25 26-28, 29-37, 40-41, 45, 50, 51, 52, 55.1-55.2, 55.3-55.5, 60-63, 64, 65.12+65.22, 66.01+66.03, 65.11+65.21+65.23+66.02+67.11, 67.12-67.13+67.2, 70+71+73+74, 72, 92.1-92.2, 93

Size: 10-19, 20-49, 50-99, 100-199, 200-249, 250-499, 500 and more employees

Neyman allocation is used to decide the sample size in each size strata. The size strata 200-249 employees, 250-499 employees and 500- employees are completely enumerated (i.e. are "take-all" strata) and at least 10 enterprises are selected from each strata. If the stratum size is smaller than 10 then the stratum is completely enumerated. The final number of strata (containing at least one enterprise) is 145.

The sample selection is done by using Statistics Sweden's system for co-ordination of frame populations and samples from the Business Register (SAMU). The ICT survey is positively co-ordinated with the sample drawn for the Survey of ICT usage and e-Commerce in Enterprises 2006.

United Kingdom	<p>The sample was designed using a Neyman Allocation optimisation program. Neyman Allocation is a sampling methodology which minimises the variance (standard error) of an estimate by selecting an appropriate sample from a population. This process determines the optimal number of enterprises that should be selected within each cell (SIC and employment size band), to minimise the standard errors for the questions used to optimise the sample. Each ONS survey begins its permanent random number (PRN) selection at different points on the PRN line to limit overlap with other surveys as much as is possible. By knowing where each survey begins its PRN selection we know where to position new surveys on the line. The larger the business the more significant it is to all surveys and therefore the more ONS questionnaires it receives.</p> <p>To minimise burden on the enterprises in the 10-49 employment size band, we excluded any businesses selected by the Annual Business Survey.</p> <p>The following strata are used, based on reported employment held on our business register:</p> <table data-bbox="453 510 869 622"> <tr> <td>10- 49</td> <td>random sampling</td> </tr> <tr> <td>50 - 249</td> <td>random sampling</td> </tr> <tr> <td>250 - 999</td> <td>random sampling</td> </tr> <tr> <td>1000+</td> <td>fully enumerated strata</td> </tr> </table> <p>The sample is drawn at a national level. The number of units sampled in the convergence regions is very small, so estimates at convergence region level will be unreliable.</p>	10- 49	random sampling	50 - 249	random sampling	250 - 999	random sampling	1000+	fully enumerated strata
10- 49	random sampling								
50 - 249	random sampling								
250 - 999	random sampling								
1000+	fully enumerated strata								
Norway	<p>The sample was drawn from all enterprises with at least five employees in the Central Register of Establishments and Enterprises. It was stratified by industry and size of the enterprise measured by employment. Among enterprises with 5-9 employees the sample was 5 per cent, among enterprises with 10-19 employees the sample was 7, 5 per cent, among enterprises with 20-49 employees the sample was 15 per cent, among enterprises with 50-99 employees the sample was 50 per cent, among enterprises with 100-249 employees the sample was 75 per cent and among enterprises with at least 250 employees all enterprises were included in the sample. The same sample percentages were used within all of the selected industries.</p>								
Serbia	<p>Stratification number of employees: 10-49, 50-249, 250-499, 500+ territory: Belgrade, Central Serbia, Vojvodina aggregate: 16 classes (NACE aggregates) according to regulation Finally, number of strata was 166. Units with 500+ employees and units from classes F1 and F2 are census.</p> <p>Allocation Sample allocation among strata was obtained with multivariate and multi-domain method (Bethel, 1989). The auxiliary variable used for sample allocation (variable of interest) was turnover. The mean and variance were calculated using data in the frame. Maximum planned error limit was defined as 14%. Allocated sample consisted of 1023 units.</p> <p>Sample selection To every unit in the frame a permanent random number (PRNUM) in the range (0,1] was attached. SRSWOR sample from each stratum was selected using these PRNUMs.</p>								
Croatia	<p>NACE group activity and enterprise size were used as stratification variables. Random sample of units was selected for each stratum. Sample of small enterprises was additionally stratified in 3 regional strata (NUTS 2). None of the procedures for the coordination or non-overlapping with samples of other surveys was applied.</p>								

I.2.3.1 STRATIFICATION

The recommendation is to use a stratified sample of enterprises with the aim to form groups of units characterised, in terms of the variables collected in the survey, by maximum homogeneity within the group and maximum heterogeneity between the groups.

The background variables X1 - "*Main Economic Activity*" (in terms of NACE) and X2 - "*Average Number of Persons Employed*" (in terms of size classes) should be used for the stratification of the sample. The "*Location of the Enterprise*" (variable X5) should also be used as a stratification variable, either on the level of NUTS 2 (on a voluntary basis) or on a level that allows a (mandatory) breakdown between the regions eligible under the Convergence objective and those that are not eligible under that objective. These variables, according to the Council

Regulation (EEC) N° 2186/93 on business registers for statistical purposes, are presented in the sampling frame and can, therefore, be used to stratify the sample *à priori*.

The purpose of the stratification by main economic activity, size class and location is to assure *à priori*, accurate results for the breakdowns. In fact, if the sample is not stratified by these variables, the number of enterprises which casually end up in some NACE category, size class, or region might be too small to produce accurate results.

For the definition of the categories and level of detail of the stratification variables, the level of dissemination concerning NACE-aggregates and size-class has to be taken into account (please refer to the explanatory notes of the background variables in chapter I.1.6.8). The minimum stratification of the frame population has to be consistent with this level of dissemination, meaning that the most detailed level of categories of all the breakdowns foreseen have to be defined by adding the aggregated results of the different strata (cells) distinguished in the stratification.

Given the level of dissemination, the minimum level of stratification concerning NACE-aggregates and size-class is defined in the following table; please note that for countries partially eligible under the Convergence objective this table should read as one for the "Convergence regions" and another for the "non-Convergence regions". For reasons of estimation efficiency, additional stratification by size-class, NACE-aggregates and/or location may be useful. Under certain conditions the efficiency of the estimator can be improved by defining more detailed (homogeneous) groups of NACE-aggregates, size-classes or locations. Efficiency means here a higher reliability of the estimates for the same overall sample size.

In the figure below, the "x" in bold represent the minimum level of stratification for the mandatory scope of the enterprise Community Survey (48 strata or - for countries partially eligible under the Convergence objective – 96 strata). The "x", in normal text, represents the minimum level of stratification for the optional scope of the survey (an additional 57 (114) strata).

				Total				
				Micro-Enterprises		Small, Medium and Large Enterprises		
				Small Micro-Enterprises	Big Micro-Enterprises	Small and Medium Enterprises		Large Enterprises
				Small Micro-Enterprises	Big Micro-Enterprises	Small Enterprises	Medium Enterprises	Large Enterprises
				1 to 4 persons employed	5 to 9 persons employed	10 to 49 persons employed	50 to 249 persons employed	250 or more persons employed
Total	Manufacturing, Energy and Construction	Section D	DA+DB+DC+DD+DE	x	x	x	x	x
			DF+DG+DH	x	x	x	x	x
			DI+DJ	x	x	x	x	x
			DK+DL+DM+DN	x	x	x	x	x
		Section E	40 + 41	x	x	x	x	x
		Section F	45	x	x	x	x	x
	Non-financial Services	Section G	50	x	x	x	x	x
			51	x	x	x	x	x
			52	x	x	x	x	x
		Section H	55.1 + 55.2	x	x	x	x	x
			55.3 - 55.5	x	x	x	x	x
			60 - 63	x	x	x	x	x
		Section I	64	x	x	x	x	x
			72	x	x	x	x	x
		Section K	70 +71 +73 +74	x	x	x	x	x
			92.1 + 92.2	x	x	x	x	x
		Section O	92.3 - 92.7	x	x	x	x	x
			93	x	x	x	x	x
			Section J	65.12 + 65.22	x	x	x	x
		66.01 + 66.03		x	x	x	x	x
		67.12 + 67.13 + 67.2		x	x	x	x	x

I.2.3.2 SAMPLE SIZE

Calculation of sample sizes should take into account that it has to ensure representative results for all the estimates produced. In particular calculation of sample size should take into account that each statistic has to be tabulated by NACE category, size class and geographical location.

As budgets are limited, the design of samples involve making trade-offs along various dimensions. Larger samples make it possible to analyse sub-groups in depth but every interview or questionnaire treated increases survey costs.

On the basis of the previous considerations, it is suggested to adopt a mixed view, based on both cost and organisational criteria and on an evaluation of the sample errors of the main estimates on a national level and with reference to each of the territorial domains and to each of the breakdown variables of interest.

The calculation of sample sizes should be based on precision requirements (see I.2.3). On this basis countries should decide on sample design and calculate the sample sizes in order to receive estimates with sufficient accuracy and within possible budgetary constraints.

In practice, the sample size is usually calculated by applying the desirable overall reliability of the estimate to a target-variable. This target variable can be one of special relevance for the survey or one that correlates well with the majority of the variables to be collected. The resulting sample size is set by the dispersion of this target-variable. However, some times for several reasons, e.g. the administrative burden of enterprises, often a maximum number of enterprises to be surveyed is defined. This number of enterprises is allocated to the different strata in such a way that the reliability of the estimates is optimized. An efficient way to allocate a specified number of enterprises to the different strata is the so-called Neyman-allocation, meaning that the number of enterprises is allocated to the relevant strata inversely proportional with the variance of a specified target-variable in these strata.

$$n_h = n \times \frac{N_h \times S_h}{\sum N_h \times S_h}$$

Where: n_h is the number of units in the sample in stratum h ;

n is total sample size;

N_h total number of units in the frame population for stratum h ;

S_h true standard deviation in stratum h for the relevant variable.

Estimates of the variance of the target variable might come from a pilot-survey or the survey from a previous year.

Additional to the outcome of the Neyman-allocation, a minimum number of enterprises in each stratum can be specified. For larger enterprises one can decide to include them integrally in the survey. However, for qualitative questions this is less crucial than for variables like for example production value.

More advanced sampling techniques may be used as long as it is possible to calculate the normal statistical variables and it meets the output specified in this manual.

By specifying a maximum number of enterprises in the sample it is useful to anticipate - based on experience with a previous survey or another comparable survey - a response rate. If experience shows that only 50 percent of the enterprises addressed, actually respond in a usable way, the sample size should be adapted to this response rate, meaning should be doubled.

The table below gives an overview of the sample sizes reported by the national statistical institutes in the Quality Reports for the 2007 national surveys. Note that due to the adaptation of the sample size to specific national needs (e.g. production of regional estimates), the sample sizes can't always be compared; countries also include in different degrees the optional parts in terms of NACE groupings and employment size classes in their survey. Note also that the scope in terms employment size class – inclusion in the survey or not of micro-enterprises – has in the countries an impact on the – average – response rate. For most countries, the final or net sample size was between 3000 and 6000 elements. The response and non-response as well as the quality (in terms of standard errors) are discussed in more detail in the next chapters.

Table: Overview of the sample sizes and response rates (2007 survey)

Source: Qualityreports for the Community survey on ICT usage and e-commerce in enterprises (2007)

Country	Scope in terms of size classes of number of persons employed	Gross sample size (1)	Net sample size used for tabulation (2)	Response not usable for tabulation	Non-response	Net response rate (2)/(1)
Belgium	5+	7,753	5,500	107	2,146	70.9%
Bulgaria	10+	4,495	4,247	153	95	94.5%
Czech Republic	5+	10,999	8,118	545	2,336	73.8%
Denmark	10+	4,316	4,231	21	64	98.0%
Germany	1+	19,977	8,853	11,044	80	44.3%
Estonia	10+	3,335	3,012	37	286	90.3%
Ireland	1+	8,180	4,236	79	3,865	51.8%
Greece	1+	4,896	2,502	109	2,285	51.1%
Spain	1+	28,820	19,929	2,065	6,826	69.1%
France	10+	14,599	11,327	493	2,779	77.6%
Italy	10+	39,977	20,479	1,110	18,388	51.2%
Cyprus	10+	1,677	1,548	110	19	92.3%
Latvia	10+	4,941	4,733	101	107	95.8%
Lithuania	10+	2,447	2,379	35	33	97.2%
Luxembourg	10+	3,144	1,955	21	1,168	62.2%
Hungary	10+	6,246	5,272	974	0	84.4%
Malta	10+	1,319	926	44	349	70.2%
Netherlands	10+	10,953	7,044	0	3,909	64.3%
Austria	10+	4,900	2,931	78	1,891	59.8%
Poland	10+	14,930	12,738	335	1,857	85.3%
Portugal	1+	6,610	5,256	330	1,024	79.5%
Romania	10+	10,068	8,975	690	403	89.1%
Slovenia	5+	2,077	1,830	10	237	88.1%
Slovak Republic	1+	4,491	3,910	10	571	87.1%
Finland	5+	5,226	3,496	304	1,426	66.9%
Sweden	10+	4,260	2,679	70	1,511	62.9%
United Kingdom	10+	7,988	5,705	1,320	963	71.4%
Norway	5+	3,580	3,458	:	:	96.6%
Serbia	10+	1,023	942	0	81	92.1%
Croatia	10+	4,000	1,573	77	2,350	39.3%

I.2.3.3 WEIGHTING – GROSSING UP METHODS

The grossing up method, or weighting procedure, to be adopted for the production of figures for the total target population is determined in the first place by the sampling design used. The weighting factors are calculated taking into account in particular the probability of selection of each unit in the sample.

In this chapter, the explanation of the several weighting methods for the several types of variables will assume the selection of a stratified random sample, which is the method recommended in this manual for the enterprise survey. The formulas have to be adapted if a different sampling design is used in the survey.

In the second place, the grossing up method is determined by the type of variable collected and the statistics produced with those variables. In the enterprise ICT usage survey there are two types of observation variables:

- 1) Qualitative / Binary variables, e.g. *"Did your enterprise have access to the Internet?"*; and
- 2) Quantitative variables, e.g. *"Please indicate the percentage of the number of persons employed who used computers with access to the WWW."*

The enterprise ICT Community Survey is dominated by qualitative questions and most results are published as percentages of the number of enterprises (total number of enterprises, enterprises with access to the Internet, etc.). This leads to conclusions or observations like 'x percent of all enterprises have access to the Internet' or 'y percent of all enterprises with access to the internet sell online'. To produce these results the observations are **weighted by the number of enterprises** in the stratum to which they belong.

When presenting overall results, these conclusions or observations are usually dominated by the results of the small and medium-sized enterprises because they are the largest group expressed in number of enterprises. However, the majority of the labour force may well work in bigger enterprises, where ICT usage is qualitatively and quantitatively different from the others. For this reason, when looking especially at more employment related issues, the results should also be additionally **weighted by the number of persons employed**. In this case the ICT usage of larger enterprises gain influence on the overall result. Weighting by employed persons leads to conclusions or observations like 'x percent of the employed persons work at enterprises with access to the Internet'.

The quantitative questions are in general collected in percentage terms, e.g. percentage of persons employed using computers, percentage of turnover resulted from orders received via computer networks. In order to produce overall averages of these percentages for the total population they also have to be additionally **weighted** by the base variable of those percentages, i.e. by the **number of persons employed** if the variable is percentage of persons employed, by the **total turnover** if the variable is percentage of turnover, etc..

When the quantitative variables are collected in real terms and not in percentages, the variables don't need to be additionally weighted by the base variable of the percentages.

In the remaining of this chapter the two weighting methods are explained in detail.

Basic weighting by number of enterprises

Assuming that a stratified random sampling is used, the estimator of a total in the population based on the sample is:

$$Y = \sum Y_h , \quad (1)$$

$$Y_h = \frac{N_h}{n_h} \sum_{i \in h} y_{hi} , \quad (2)$$

where Y is the estimated total value of variable y for the total population

Y_h is the estimated value of variable y for the total population in stratum h ;

N_h total number of units in the frame population for stratum h ;

n_h is the number of units in the sample in stratum h ;

y_{hi} is the value of variable y of enterprise i in stratum h .

In the case of the quantitative variables in real terms (not in percentage terms), y_{hi} is the value of that variables. In the case of the qualitative variables, y_{hi} assumes the value 1 if the answer is "YES" and 0 if the answer is "NO". This way the total of this variable is the number of enterprises having answered "YES".

To compute the percentages, these totals are divided by either the total number of enterprises, in the case of the qualitative variables, or by the total of the quantitative variable for all enterprises, in the other case.

For the total population:

$$Y^p = \frac{\sum Y_h}{N} \quad (3)$$

For each stratum:

$$Y_h^p = \frac{\frac{N_h}{n_h} \sum_{i \in h} y_{hi}}{N_h} \quad (4)$$

In the grossing up, each enterprise i in strata h has the following weight

$$w_{hi} = \frac{N_h}{n_h} , \quad (5)$$

which gives how many enterprises in the population this sampled unit represents.

Because the weighting factor w_{hi} is based on the number of enterprises, this is called weighting by number of enterprises.

Additional weighting by an auxiliary variable (number of persons employed, turnover, etc.)

As mentioned before, the additional weighting by an auxiliary variable z has the purpose, in the case of the qualitative variables, of computing indicators of the type "percentage of variable z ", instead of "percentage of enterprises". For example "percentage of persons employed". In the case of the quantitative variables collected in percentage terms, this additional weighting by the base variable of the percentages is necessary to gross up the results.

In terms of real values, and not percentages, this means that the indicators computed using this additional weighting are of the type "number of persons employed", "turnover" or "purchases", instead of "number of enterprises" and the percentages are computed afterwards using the totals of these variables.

The formula for the computation of these indicators of the type "percentage of variable z " is actually the same in both types of variables, but the reasoning is slightly different. For this reason, they will be explained separately.

Qualitative / Binary variables

In this case we compute indicators of the type "percentage of variable z , for enterprises with y equal to some value" (where y is the qualitative / binary variable). For example "percentage of persons employed, for enterprises with Internet access", in which case the qualitative / binary variable "Do you have Internet access" equals "Yes". Or, in real values, "number of persons employed, for enterprises with Internet access".

One way of doing it for the real value is simply to select the enterprises for which the qualitative variable equals the value of interest (in the example above, "Yes") and gross up the number of persons employed in those enterprises:

$$Y = \sum Y_h ,$$

$$Y_h = w_{hi} \cdot \sum_{i \in h} z_{hi} y_{hi} \quad (6)$$

where Y_h is the estimated real value of persons employed in enterprises, for which the qualitative variable equals the value of interest, for the total population in stratum h ;
 N_h total number of units in the frame population for stratum h ;
 n_h is the number of units in the sample in stratum h ;
 z_{hi} is the number of persons employed in enterprise i of stratum h .
 y_{hi} is a binary variable assuming 1 for the enterprises for which the qualitative variable equals the value of interest, and 0 otherwise.
 w_{hi} is the weight of enterprise i in stratum h ., as defined in formula (5).

For the percentage, we divide this value by the grossed up total of persons employed in all enterprises:

$$Y_h^p = \frac{w_{hi} \cdot \sum_{i \in h} z_{hi} y_{hi}}{w_{hi} \cdot \sum_{i \in h} z_{hi}} \quad (7)$$

Which simplifies to:

$$Y_h^P = \frac{\sum_{i \in h} z_{hi} y_{hi}}{\sum_{i \in h} z_{hi}} \quad (8)$$

Quantitative variables

In this case the indicators computed are of the type “percentage of variable z , in relation to another variable”, where the first one is a parcel of the second one. For example “percentage of persons employed with access to Internet, in relation to the total number of persons employed”. When defined in real values, it refers to the numerator only: “number of persons employed with access to Internet”.

One way of doing it for the real value is to use the individual percentages of each enterprise (as it is collected), multiply it by the base of that percentage (in the example above, “total number of persons employed”) and gross it up. The total value for each stratum h would then be:

$$Y_h = w_{hi} \cdot \sum_{i \in h} z_{hi} y_{hi} \quad (9)$$

where Y_h is the estimated real value of persons employed with access to the Internet, for the total population in stratum h ;

N_h total number of units in the frame population for stratum h ;

n_h is the number of units in the sample in stratum h ;

z_{hi} is the number of persons employed in enterprise i of stratum h .

y_{hi} is a binary variable assuming 1 for the enterprises for which the qualitative variable equals the value of interest, and 0 otherwise.

w_{hi} is the weight of enterprise i in stratum h ., as defined in formula (5).

For the percentage, we divide this value by the grossed up total of persons employed in all enterprises of stratum h :

$$Y_h^P = \frac{w_{hi} \cdot \sum_{i \in h} z_{hi} y_{hi}}{w_{hi} \cdot \sum_{i \in h} z_{hi}} \quad (10)$$

Which simplifies to:

$$Y_h^P = \frac{\sum_{i \in h} z_{hi} y_{hi}}{\sum_{i \in h} z_{hi}} \quad (11)$$

Formulas (6) and (9) are identical, and formulas (8) and (11) as well. Therefore, the grossing up method when an additional weighting by an auxiliary variable is used is the same for qualitative variables and quantitative variables, when collected in percentage values.

I.2.4. Survey type

Data collection method

Face-to-face interviews, telephone interviews, postal surveys and web-based or electronic surveys are all possible techniques of collecting data. Next the main strengths and weaknesses of each method are presented.

Postal survey

Strengths:

They are relatively inexpensive to administer. You can send the exact same instrument to a wide number of companies. They allow the respondent to fill it out at their own convenience. Mail survey doesn't allow personal contact between the researcher and the respondent; consequently the researcher impressions can't influence the response of the respondent.

Weakness:

But there are some disadvantages as well. Response rates from mail surveys are often very low and mail questionnaires are not the best vehicles for asking for detailed written responses. They also require a strong monitoring system (of sending reminders, etc.)

Personal interviews (face-to-face)

Strengths:

Interviews are a far more personal form of research than a postal survey. In the personal interview the interviewer works directly with the respondent (controlled situation). Unlike with mail surveys the interviewer has the opportunity to probe or ask follow-up questions. Interviews are generally easier for the respondent, especially if what is sought is opinions or impressions (individual behaviour can be observed and exchange of material/information between interviewer and respondent is possible). They have a very good response rate.

Weakness:

Interviews can be very time consuming and they are resource intensive. The interviewer is considered a part of the measurement instrument and interviewers have to be well trained in how to respond to any contingency. They are very expensive and consequently inapplicable for global and big surveys.

Telephone interviews:

Strengths:

Telephone interviews enable a researcher to gather information rapidly. Like personal interviews they allow for some personal contact between the interviewer and the respondent (questions can be repeated and interpreted). They allow the interviewer to ask follow-up questions.

Weakness:

Some small companies don't have publicly-listed telephone numbers. Some don't have telephone. People often don't like the intrusion of a call to their homes. Telephone interviews have to be relatively short or people will feel imposed upon. The inability to use visual aids is also a weakness. Difficulties occur in case information needs to be collected by the respondent (e.g. from the accounts).

Electronic survey:

Strengths:

With the growth of the Internet (and particular the World Wide Web) and the expanded use of electronic mail for business communication, the electronic survey is becoming a more widely used survey method. Electronic surveys can take many forms. They can be distributed as electronic mail messages sent to potential respondents. They can be posted as World Wide Web forms on the Internet. They can be distributed via publicly available computers. In some cases electronic surveys are placed on laptops and respondents fill out a survey on a laptop computer rather than on paper. It is less expensive to send questionnaires online than to pay for postage or for interviewers. It is easier to make changes to questionnaire, to copy and sort data. Questionnaires can be delivered to recipients in seconds rather than in days as with traditional mail. Research shows that respondents may answer more honestly with electronic surveys than with paper surveys or interviews. Due to the speed of online networks participants can answer in minutes or hours and coverage can be global. Since many enterprises prefer to report information electronically to more traditional data delivery, offering electronic surveys generally increases the total response rate.

Weakness:

Population and sample is limited to those with access to computer and online network. Thus, in reality electronic surveys can only be a supplement to other survey vehicles. The use of several survey vehicles simultaneously may prove not cost efficient due to increased logistical work. The open natures of most online networks make it necessary to invest considerable time and expertise in guaranteeing acceptable anonymity and confidentiality. Constructing the format of a computer questionnaire can be more difficult the first few times due to a researcher's lack of experience. More instruction and orientation to the computer online systems may be necessary for respondents to complete the questionnaire. Electronic surveys can have a high technical problem with hardware and software.

Nevertheless, in practice almost all EU Member-States collect the data through self-administered postal survey, more and more combined with the possibility to fill-out and post the questionnaire electronically. The table below presents the information on the data collection methods used in the 2007 national surveys.

Table: Overview of the data collection methods (2007 survey)

Source: Qualityreports for the Community survey on ICT usage and e-commerce in enterprises (2007)

Country	Stand alone survey?	Survey vehicle	Mandatory participation?
Belgium	Yes	Web survey + self-administered mail survey	Yes
Bulgaria	Yes	Face to face interview	No
Czech Republic	Yes	Postal (self-administered mail out and mail or e-mail back) survey. Electronic questionnaire (computer assisted program available on the web site) for electronic completion was developed.	Yes
Denmark	Yes	Self-administered mail survey	Yes
Germany	Yes	Mail survey	No
Estonia	Yes	Mail survey and the possibility for enterprises to get the questionnaire from Statistical Office homepage or fill in the web questionnaire.	Yes
Ireland	Yes	Self administered mail survey	No
Greece	Yes	A combination of face-to-face interviews, self-administered mail and telephone interviews	Yes

Spain	Yes	Firstly, questionnaires are sent by mail to the enterprises in the sample. Moreover, enterprises may phone to a free-charge telephone number to get clarification of the questions and to be helped with the filling out of the questionnaire. In all cases there is a system of reminders, including letters and phone calls. The survey is mandatory and enterprises can be penalized in case of no response.	Yes
France	Yes	Self-administrated mail survey; (using phone for reminders and further information).	Yes
Italy	Yes	Self-compilation of mailed questionnaires Enterprises have possibility to fill on-line the questionnaire using a web site dedicated to the two surveys (a pdf questionnaire is downloadable for an off-line compilation and a sending via Internet); a login and a password is provided to enterprises with mailed questionnaires.	Yes
Cyprus	Yes	Face-to-face interview. The questionnaire was sent by mail or e-mail to the IT manager of the enterprise just before the beginning of the data collection period. The IT manager was informed, from the covering letter, that an enumerator would contact him in order to arrange a meeting for the completion of the questionnaire.	Yes
Latvia	Yes	It was a mail survey where enterprises also were able to submit the filled questionnaires electronically by e-mail and to fill questionnaires using our on-line submission system via Internet (in CSB web-site).	Yes
Lithuania	Yes	Mail survey	Yes
Luxembourg	Yes	Self-administered mail survey	Yes
Hungary	Yes	Self-administered mail survey	Yes
Malta	Yes	Face to face interview	Yes
Netherlands	Yes	The questionnaires are sent out by mail. Depending on the initial response one or two written reminders are sent out.	No
Austria	Yes	A paper questionnaire which comprised accompanying information as well as the glossary and a pre-paid addressed envelope for returning the questionnaire to Statistics Austria was sent to the respondents – together with a leaflet explaining the political background of the survey and providing basic instructions. Additionally, enterprises can report data by an electronic questionnaire via Internet. A download of the questionnaire is also possible. A telephone hot line for the respondents has been established. Two written reminders have been sent out and “motivation calls” were carried out by phone (via our in-house call centre). This information is valid for the survey among “enterprises” as well as in the “financial sector”. For the “financial sector” the same paper questionnaire was used but it contained a remark that the modules on e-commerce have not to be filled in.	No
Poland	Yes	Self-administered mail survey	Yes
Portugal	Yes	This survey combines two techniques of data collection: postal and electronic. The enterprises can send the information by post or download the web questionnaire available in the NSI site and return it electronically.	Yes
Romania	Yes	Self-administered mail survey. Telephone interview case of errors appeared.	Yes
Slovenia	Yes	Self-administered mail survey – the questionnaires sent to the manager of the enterprise by mail.	Yes
Slovak Republic	Yes	The surveys are included in the Program of state statistical surveys in 2006. The questionnaires as well as the reminders will be sent to the enterprises by post. Thereafter the respondents will be reminded also by phone.	Yes
Finland	Yes	Mail survey	Yes
Sweden	Yes	Self-administered mail survey. The enterprises were invited to fill in a web questionnaire instead of the paper version if they preferred that.	No
United Kingdom	Yes	Self-administered mail survey.	Yes

Norway	Yes	Self-administered mail survey in combination with web questionnaire. All sampled enterprises were able to report either via paper questionnaire or web questionnaire.	Yes
Serbia	Yes	Telephone interview	No
Croatia	Yes	Self-administered mail survey - questionnaires were sent by mail to all participants (enterprise managers) and were available in two forms: a) printed questionnaire b) web-based questionnaire. After receiving questionnaires, some of the missing/incomplete data was acquired via additional telephone interview.	Yes

Independent versus embedded survey

The ICT usage survey in enterprises is not a short survey. If this survey is attached to another survey it is possible that the complete questionnaire is too long. The respondent needs too much time to answer all the questions. Consequently the quality of the answers decrease and the rate of non-response increases. Arguments in favour of embedding the ICT usage survey into an existing business survey are reduction of its total cost and the possibility to cross ICT usage data with other business information.

The overview table above indicates that in 2007 all EU national ICT usage surveys were stand-alone surveys.

Mandatory survey versus voluntary survey

Voluntary surveys are usually cheaper, quicker and easier to manage. In mandatory surveys it is implied to send the questionnaire several times to the respondent, you have to write a reminder letter and sometimes you have to send an administrative offence. You have to wait a long time for all responses and your costs increase. The advantage of a mandatory survey is that your response quote is potentially 100%. But a voluntary survey can settle this argument by increasing the sample size and sending also reminder letters to the respondents.

Contact person of the survey

In most cases the IT manager is the appropriate recipient of the survey on ICT usage in enterprises. However, not all small enterprises have an IT manager; in these cases either the owner or the general administrator should answer. In bigger companies it is helpful to send the questionnaire to the general management. They have an overview of all functions within the company. Sometimes you find a contact person of the company in your business register. In this case it is useful to send the questionnaire to this person.

I.2.5. Questionnaire – Data collection tool

To enhance the comparability across countries, Eurostat provides a model questionnaire to the NSIs. It is recommended to adopt the model questionnaire for the national survey. However, for practical reasons (for instance linked to the survey vehicle), this may not be appropriate. In this case, the statistical institute should nevertheless make sure that the routing and the filtering is followed.

Eurostat designs the model questionnaire in close cooperation with the national statistical institutes, the main users and in coordination with other organisations such as the OECD. The results of previous surveys are taken into account, for instance to assess the relevance of certain questions or items, or to assess whether a variable needs to be collected on an annual basis. For reasons linked to the annually renewed implementing measures for Regulation 808/2004, the model questionnaire needs to be finalised 9 months before the survey takes place. In practice, this means that the questionnaire reaches its final status in March-April of the year (t-1) preceding the reference year (t). The obvious disadvantage is that 'last-minute' policy needs can't be included or that the results of the t-2 survey can't be fully taken into account. The advantage is that this gives the NSIs the necessary time to translate the questionnaire and implement the survey.

The model questionnaire for the Community Survey on ICT Usage and e-Commerce in Enterprises can be found in Annex I.3.1.

I.2.6. Quality control systems

Quality control systems are of course country-specific as most statistical institutes have standard procedures and guidelines for plausibility checks or logic tests of datasets.

Such controls can be executed on-line, at the moment of the data capture by the interviewer or the data entry in the statistical institute, or after the data entry process (a program checks the data and highlights the errors to be checked or corrected). On-line tests have the advantage that the errors can be corrected immediately (a lot of errors will be simple data-entry errors, for instance typing 17773 instead of 1773), a disadvantage is that one needs data-entry staff that is familiar with the concepts in the survey (in case they have to decide if a certain value will be accepted or not). The latter case of course doesn't apply with CATI or CAPI where the interviewer immediately enters the data.

Below, some of the most common errors or problems are briefly discussed.

Measurement error

There are a number of sources of measurement error: survey instruments (questionnaire), the respondent, the information system, the mode of data collection, the interviewer. This paragraph focuses on the latter, other types can be found below. The first and probably most important step in the statistical process is the data collection or field work. One can design excellent questionnaires, excellent capture tools, excellent imputation methods and excellent data analyses, but as with any process the adage 'Garbage In Garbage Out' also applies to statistics. If the interviewer is not well-trained, one can't expect input data of a good quality. A more important problem is usually the reliability of the interviewers. The interviewers may have an interest in finishing the interviews in the shortest possible time. This can lead to interviews where certain questions are skipped (e.g. the enterprise looks old-fashioned, let's fill in none of the more sophisticated e-business functions are used) or interviews that never took place (e.g. the enterprise is so small, let's put down it does not use computers). Therefore the sensibilisation of interviewers (which can include threats of non-payment or exclusion for future interview rounds) and the follow-up is extremely important. The follow-up can be a check for coherence and consistency of the answers or follow-up phone calls to verify with the respondents whether the interview has actually taken place or whether the interview was of an acceptable quality (e.g. the respondent confirms that the interview only took 4 minutes, but on the questionnaire answers to 60 questions were recorded).

Invalid response

This aspect is relatively unimportant in the ICT usage survey as most answers are limited to Yes or No. However, it is possible that several items were ticked in questions where only one answer is expected (e.g. Yes **and** No).

Relationship error

Comparing the answers across the survey can reveal inconsistencies between the answers. The routing and filtering should normally guarantee that respondents are not presented with questions they can't answer. It is nevertheless possible that e.g. an enterprise not using automated data exchange (question C1) indicates that it is sending e-invoices (in question C2) or that an enterprise with 20 persons employed (X2) indicates that the number of persons employed using computers with access to the WWW is 200 (question B2). In the first case, this is most probably an error related to the understanding of the questions by the respondent. In the latter case it is more likely that a "typing" error occurred.

Compulsory question left unanswered

Again, the routing and filtering should avoid this type of errors. CATI or CAPI programs usually improve the quality of the data capture in this respect. With traditional interviews or self-administered mail surveys, this error is more important.

Suspicious values

This can occur when the individual responses are compared with the average scores within a stratum and unexpected answers are found. In practice it will however be difficult to determine whether it concerns a mistake or an outlier.

In terms of quality of the survey *as such*, the methodology and outcomes of the survey can be benchmarked against other surveys:

Representativeness

It can be useful to do an *ex-post* check of the representativeness of the sample, e.g. does the sample have a representative size class distribution, is there some variability in the economic activities?

Year-to-year comparison at aggregate level

Comparing the results for the current year with the previous survey can also reveal quality problems where the growth is outside the range of the expected growth (e.g. the proportion of enterprises using the Internet decreases ...). In such cases, it is of course possible that the problem stems from the previous survey exercise ...

For this purpose, it can be interesting to produce some simple tabulations of the survey results.

Coherence or consistency with other surveys

The results can be compared with results from related survey or studies. However, in case inconsistent results are observed, it is not always easy to identify which survey gave the 'wrong' results.

I.2.7. Data processing

This chapter mainly discusses the treatment of non-response. Although the grossing-up methods can be considered as a part of the *data processing*, this topic is discussed above in section I.2.3.3.

I.2.7.1 MISCLASSIFICATION TREATMENT

Misclassification occurs when an enterprise is included in the survey (because it belongs to a size class and sector of activity covered by the survey, at least according to register data used for stratification) but later information indicates that it should not have been included. In other cases misclassification of enterprises means that enterprises that (according to the registers) are classified in a size class or NACE category should have been classified in another size class or NACE category. The misclassification will then possibly lead to the fact that the enterprises should belong to another strata than the one used for stratification.

Such a situation can arise due to frame population imperfections. Frame imperfections can occur when there is a time lag between the actual situation for an enterprise and the information available in the registers. It often takes a certain period of time to update register information after a change in the number of employed persons or a change of sector of activity has occurred.

This time lag in updating register information implies that there is a difference between the target population (i.e. the population that the survey intends to cover) and the frame population (i.e. the population that the survey actually covers based on information available in registers).

Recommendation in case of misclassification of enterprises

There are different possible options available in case of misclassification of enterprises. The options depend on the type of misclassification.

If it is obvious that an enterprise should not have been included in the survey for example because the number of employed persons have decreased and falls beyond the size cut-off limit during the reference period, the enterprise could be excluded from the sample, as it should have never been included.

However, such an approach could be hazardous as correction then only is made for enterprises that fall beyond the cut-off limit (e.g. 10 persons employed) and not for enterprises that had less than 10 employed persons according to register information and that during the reference period of the survey exceeded the cut-off limit. A more appropriate approach is in those cases to assume that enterprises where the number of persons employed has decreased below the cut-off limit offset the enterprises that have increased in number of employed persons and that exceed the cut-off limit. Enterprises falling below the cut-off limit are then treated as respondents and not as over-coverage.

If the misclassification means that the enterprise should belong to another strata than the one used for stratification, new strata should be built and the weights used in computations should be changed accordingly.

I.2.7.2 NON-RESPONSE TREATMENT⁶

Introduction

An important source of non-sampling error in surveys is the effect of non-response on the survey results. Non-response can be defined as the failure to obtain complete measurements on the (eligible) survey sample. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response.

The latter case occurs when the respondent refused to participate in the survey or when the questionnaire was sent back to the statistical office as returned mail (e.g. when the address to the enterprise is incorrect). This type of non-response is called unit non-response (see I.2.7.2.1): the sample unit does not provide any of the data required by the survey. Unit non-response is generally handled by adjusting the weight of the enterprises that responded to the survey to compensate for those that did not respond.

Partial non-response or item non-response (see I.2.7.2.2) occurs when the respondents did not answer all questions because they did not understand or misinterpreted a question, alternatively refused to answer a question.

Effect of non-response on the quality of the data

Non-response (unit as well as item non-response) can seriously affect the quality of the data collected in a survey. Firstly, the characteristics (or answering pattern) of the non-respondents can be different from those collected among the sample units who did provide eligible answers. If such difference is systematic, serious bias can be introduced in the survey results. Secondly, the reduction of the sample size (overall or for certain questions) will increase the variance of the estimates. Thirdly, non-response can have an impact on the total cost of a survey exercise. Not only because a larger initial sample may be necessary, but also because of higher unit costs of the last few percentages of respondents (due to multiple visits, sending of reminders, repeated telephone calls). Finally, non-response can be an indicator of poor overall quality of the survey and thus create an image or confidence problem.

Minimising non-response

As prevention is always better than cure, attention should be given to avoiding non-response rather than treating non-response. The number (and timing) of reminder letters or call backs, the length of the fieldwork period, the survey technique(s) used, the length of the survey (i.e.

⁶ References for this chapter:

- Government Statistical Service (1997), *Report of the Task Force on Imputation – June 1996*. Government Statistical Service Methodology Series No. 3. London: Office for National Statistics.
- Statistics Canada (2003), *Household Internet Use Survey, 2002 – Microdata User Guide*.
- European Commission (2003), *Household Budget Surveys in the EU: Methodology and recommendations for harmonisation 2003*. Luxembourg: Office for Official Publications of the European Communities.
- National Centre for Social Research (1999), *Item non-response* (different contributions). Survey Methods Newsletter Volume 10 No.2 1999.
- Adjusting for Missing Data, Mohadjer, L., Choudhry, H., Studies of welfare populations: Data Collection and research Issues, 2001.
- Introduction to Survey Quality, Biemer, P., Lyberg, L., Wiley series in survey methodology, 2003.
- Weighting for non-response, Lynn, P., Survey and Statistical Computing, 1996.
- Non-response in the Norwegian Business Tendency Survey, Wang, J., Statistics Norway, 2004.

the response burden), the use and structure of advance letters, the dissemination of previous results or the mandatory nature of the survey can all have an impact on the number of non-contacts or refusals. The length, design and complexity of the questionnaire, the interviewer's training or the interviewer's style can have an impact on the item non-response.

As this issue is common to all surveys, it will not be discussed in detail in this manual. This chapter is mainly focussed on the treatment of non-response in the specific context of the ICT usage survey.

1.2.7.2.1. UNIT NON-RESPONSE

Introduction

Unit non-response is defined as enterprises that are included in the sample but that have not participated in the survey and for which information consequently is missing for all the questionnaire variables.

Unit non-response can introduce bias in the survey results especially in situations in which the non-responding units are not representative of those that responded. Non-response increases both the sampling error level, by decreasing the sample size, and non-sampling error level.

Weighting adjustment for unit non-response

The principal method for unit non-response adjustment is weighting. Most strategies for weighting for non-response involve dividing the respondents into a set of comprehensive and mutually exclusive groups, referred to as weighting classes. A weight is then applied to each class.

Weighting classes

In order to implement non-response adjustments, it is required to create weighting classes. It is desirable to divide the sample in "response homogeneity groups/classes". Within these classes the response rates should be as homogeneous as possible, and the response rate should be different among the classes. Data used to form these classes must be available to both non-respondents and respondents. Usually it is possible to get information on size, economic activity, legal status, location, and other variables in the business registers.

More advanced methods for creating weighting classes are methods like classification based on a categorical search algorithm or a logistic regression model using auxiliary variables to estimate the probability of response.

Sample-Based Weighting Adjustment

In sample-based weighting adjustment the weight adjustment applied in each class, is equal to the reciprocal of the ratio of selected sample size to respondents within each of these classes (the inverse of the response rate within each class). The grossing-up factor should then be multiplied by the non-response adjustment factor.

A simple example:

Size Class	Population (I)	Sample size (II)	Respondents (III)	Respondent with characteristic (IV)	Non-response adjustment Factor (V = II / III)	Initial Grossing-up factor (VI = I / II)	Adjusted Grossing-up factor (VII = V * VI)
Small	35 141	878	764	595	1.15	40.0	46.0
Medium	5 362	882	821	795	1.07	6.1	6.5
Big	761	761	624	543	1.22	1.0	1.2
Total	41 264	2 521	2 209	1 933			

Alternative forms of sample-based weighting are that the weights are not inverse response rates, but estimated coefficients of a regression model (where survey response is the left-side variable). In this case, the weights are reciprocals of estimated response rate by the regression model.

Population-Based Weighting Adjustment

Population-based weighting adjustment requires population estimates and class membership of respondents. If there is no data available about the non-respondents, population-based adjustment still is possible since this uses external control counts for the population and not data from the sample. The method is used to correct simultaneously for both non-coverage and non-respondents. The method is used similar to the sample-based method.

In population-based adjustment (post-stratification adjustment) the classes are created based on variables, which are known both for respondents and for the population. Weights are then applied in proportion to the ratio of population to achieved sample, so that the sums of the adjusted weights are equal to population totals for certain classes of the population.

A two-step procedure of first adjusting for non-response (sample-based adjusting) and then adjusting to known population counts is a common method that is used. However, this procedure is the same as a population-based weighting adjustment if the weighting classes in the sample-based and the population-based weighting adjustment are equal.

If the strata used in the stratification are used as classes in the weighting adjustment, there is no need for the weighting adjustment. The adjusted weighting procedure is then equal to the final grossing up/weighting procedure.

1.2.7.2.2. ITEM NON-RESPONSE

As already mentioned above, there are several reasons for the data being unavailable. These include the refusal to provide an answer, the inability to provide an answer, inadequate quality of the provided answer (e.g. implausible, incomplete, inconsistent with answers to other questions, etc.). It can be caused by either the respondent (e.g. refusal) or the interviewer (e.g. failure to record the answer adequately) but also by the survey design itself (e.g. ambiguous routing or filtering).

In case a particular questionnaire shows too many errors, or if too many data are missing, it can be assumed that the enterprise in question has not co-operated satisfactorily in the survey. Here, the best solution is probably to remove the enterprise from the database for respondents (but not from the frame) and adjust the weighting coefficients for the other enterprises accordingly. In other words, sampling units with a very high item non-response can better be classified as total non-response or unit non-response. However, if the enterprise has provided answers on some key questions of the questionnaire, it should probably be preferable to keep the answers provided by the enterprise (even though many questions are left blank).

The general assumption that should be used is that item non-response means that the answer to the question is:

- "NO" if the question is a qualitative question (e.g. yes/no-questions)
- Zero if the question is a quantitative question (e.g. e-commerce value questions)

The general recommendation is that imputations should be avoided, except logical imputations.

In principle two kinds of imputations could be considered: 1) to use data from answers provided by the enterprise in the previous year or 2) to use data from the current survey and impute the answer on the basis of answers provided by similar enterprises (e.g. enterprises belonging to the same strata).

To use data from the previous year is not recommended as the use of ICT is expected to change over time. It is e.g. not obvious that an enterprise that did not sell goods or services via computer networks during year t-1 still did not do so during year t. The data provided the previous year can however be used in order to find enterprises that should be contacted to verify the change from previous year.

To use the data of the current survey to generate an imputed value for a variable is not recommended as there is a risk of overestimation. In this survey possible underestimation due to item non-response is preferred to possible overestimation due to imputation.

As **no imputation is recommended**, it is very important that item non-response is kept as low as possible in order to minimise possible underestimation.

There are different methods to attain lower item non-response rates:

- 1) Clarity of questions. Unclear questions might be more frequently "avoided" by the respondents. The content and form of the questions must be considered carefully. Pre-tests can be done before sending out the questionnaires in order to check that the questions are clear to the respondents.
- 2) After conducting the survey, item non-response can be further reduced by contacting the respondents in order to get answers to the missing questions. Special efforts should be put in order to gather answers for questions where item non-response is especially high.

For e-commerce value questions it is recommended to check for enterprises with non-response in e-commerce value questions but having large e-commerce values in previous years. These enterprises should be contacted to verify the current situation.

For e-commerce value questions it is important to compare the answer to that of previous year. Especially enterprises stating that they have e-sales but that have not provided an estimate of the level of e-sales should be contacted to have an estimate. If the enterprise has provided an estimate of e-sales value previous year it could be checked if the value that was provided then can still be of use.

Item non-response for the background variables that are used in the grossing-up process should also be avoided. Here also, contact with the enterprises is the preferred method.

I.2.7.3 COMPUTING ACCURACY MEASURES

The use of a sample survey unavoidably leads to sampling errors. To get an idea of the sampling error, it is advised to compute this error for certain indicators. Such information is not only interesting to decide whether the indicator is reliable enough to be published, but can also help to determine whether the sample size should be increased in a next edition of the survey for certain strata where a large sampling error was diagnosed.

The sampling error reflects the fact that only a particular sample was surveyed rather than the entire population. As accuracy measure the standard error (the square root of the variance) of the estimator for proportions is recommended (see also Chapter 9 of the *Reporting template*, Annex I.3.3). The estimation of the sampling variance should take into account the sampling design (e.g. the stratification).

Additional guidance for the standard errors including thresholds regarding the levels that are considered acceptable can be found in section I.2.3 where the sampling design is discussed.

The computed accuracy measures will also determine the reliability of the estimates. If the standard error for a certain proportion or absolute value is too high, the cell needs to be suppressed.

In case the computation of the standard error for all cells is not feasible, the number of respondents underlying the estimate can be used to decide whether a proportion can be published (e.g. if based on less than 10 respondents, the estimate is not published).

See also the annex I.3.2, the transmission format; this document describes how cells that do not comply with the quality standards mentioned in section I.2.3 in terms of standard errors or are for other reasons not of a sufficient quality (high item non-response, question not in line with the model questionnaire, etc.) should be "flagged" as unreliable; in the latter case notes accompanying the data should explain why the data is considered unreliable.

I.2.8. Survey execution reports

Drawing up reports after the execution of the survey is not only a tool for a self-assessment of the work carried out by the statistical office and a detection of possible points where there is space of improvement. Survey execution reports are also a tool for assessing the comparability from one year to another or from one country to another.

The methodological reports for the ICT usage surveys coordinated by Eurostat cover the following subjects:

1. Identification of the survey and the NSI.
2. General methodological information: reference period(s); survey period; survey vehicle; survey type; survey participation methodological differences compared to previous data collection exercises.
3. Statistical unit, scope and target population: statistical unit; economic activities covered; size coverage in terms of number of persons employed; territorial coverage.
4. Information on the national questionnaire: deviations with the model questionnaire; additional questions; inclusion of optional questions.
5. Sampling frame: name and description of the sampling frame or register used; known shortcomings of the sampling frame, if any; frame distribution by size class and economic activity.
6. Sampling design: sampling method; sample distribution (by size class economic activity).
7. Unit non-response, by giving information on: gross sample size; number of out-of-scope cases, number of other ineligible cases; number of elements usable for tabulation and grossing-up; non-response; net sample size; unit response rate; additionally, methods used for minimizing unit non-response; methods used for treating unit non-response; variables or items with response rates below 90%; methods used for treating item non-response.
8. Grossing-up procedures.
9. Sampling error: standard errors for a selected group of indicators or sub-indicators; the calculation method for the standard error; comments on reliability, representativeness and completeness of the data.
10. Reference to the dissemination of national results of the survey.
11. Problems encountered and lessons to be learnt.
12. The report includes the questionnaire in national language and, if available, in English.

In cases where the recommendations provided in this Methodological Manual are not followed, the report should clearly describe deviations, the impact of those on comparability, a justification for the deviation and – where possible – plans to comply with the recommendations.

The *reporting template* used for the European enterprise ICT usage survey, including some instructions, can be found in Annex I.3.3.

This template covers both the *Metadata report* (formerly known as *Interim Report*) and the *Quality report* (formerly known as *Final Report*) referred to in Article 7(4) of Regulation 808/2004 (see Annex III.1) and in Annex I, section 6 of Regulation 847/2007 (see Annex III.2).

I.3. ANNEXES

I.3.1. Model questionnaire

COMMUNITY SURVEY ON ICT USAGE AND E-COMMERCE IN ENTERPRISES

2008

Model Questionnaire (Version 3.4 of 30 April 2007)

(Questions relating to the *i2010* Benchmarking Indicators are marked with an asterisk *)

Module A: Use of computers and computer networks

A1.	Did your enterprise use computers, in January 2008? (Filter question)	Yes <input type="checkbox"/>	No <input type="checkbox"/> → Go to X1
A2.	How many persons employed used computers at least once a week, in January 2008? - <i>Optional</i> If you can't provide this value, Please indicate an estimate of the percentage of the number of persons employed who used computers at least once a week, in January 2008. - <i>Optional</i>	<div style="border: 1px solid black; width: 150px; margin: 0 auto; text-align: center; padding: 2px;">(Number)</div> <div style="display: flex; justify-content: center; align-items: center; gap: 10px;"> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> % </div>	
A3.*	Was your enterprise using an internal computer network (e.g. LAN - Local Area Network) in January 2008? (Filter question)	Yes <input type="checkbox"/>	No <input type="checkbox"/> → Go to A5
A4.*	Did your enterprise use wireless access within its internal computer network (e.g. wireless LAN), in January 2008?	Yes <input type="checkbox"/>	No <input type="checkbox"/>
A5.	Did your enterprise have in use an internal home page (Intranet), in January 2008? (Filter question)	Yes <input type="checkbox"/>	No <input type="checkbox"/> → Go to A7
A6.	In January 2008, was your enterprise using such systems for sharing the following information? <i>Optional</i>	Yes	No
	a) The general policy or strategy of the enterprise	<input type="checkbox"/>	<input type="checkbox"/>
	b) Internal company newsletters or daily news	<input type="checkbox"/>	<input type="checkbox"/>
	c) Day-to-day / working documents (e.g. for meeting)	<input type="checkbox"/>	<input type="checkbox"/>
	d) Manuals, guides or training material	<input type="checkbox"/>	<input type="checkbox"/>
	e) Product or services catalogues	<input type="checkbox"/>	<input type="checkbox"/>
A7.	In January 2008, did your enterprise use dedicated applications for employees to access human resources services (e.g. see open job positions, request annual leave, view or download payslips, or other services)?	Yes <input type="checkbox"/>	No <input type="checkbox"/>
A8.*	In January 2008, did your enterprise have an extranet (a website or an extension of the Intranet with access restricted to business partners)?	Yes <input type="checkbox"/>	No <input type="checkbox"/>
A9.*	Did your enterprise have in use, in January 2008, third party free or open source operating systems, such as Linux ? (i.e. with its source code available, no copyright cost, and the possibility to modify and/or (re)distribute it)	Yes <input type="checkbox"/>	No <input type="checkbox"/>

Module B: Access and use of the Internet		
(Scope: enterprises with Computers)		
B1.	Did your enterprise have access to the Internet, in January 2008? (Filter question)	Yes <input type="checkbox"/> No <input type="checkbox"/> → Go to C1
B2.*	How many persons employed used computers with access to the World Wide Web at least once a week, in January 2008? If you can't provide this value, Please indicate an estimate of the percentage of the number of persons employed who used computers with access to the World Wide Web at least once a week, during January 2008.	(Number) _____ _____ _____ _____ %
B3.*	Did your enterprise have the following types of external connection to the Internet, in January 2008?	Yes No
	a) Traditional Modem (dial-up access over normal telephone line) or ISDN connection	<input type="checkbox"/> <input type="checkbox"/>
	b) DSL (xDSL, ADSL, SDSL etc) connection	<input type="checkbox"/> <input type="checkbox"/>
	c) Other fixed internet connection (e.g. cable, leased line (e.g. E1 or E3 at level 1 and ATM at level 2), Frame Relay, Metro-Ethernet, PLC - Powerline communication, etc.)	<input type="checkbox"/> <input type="checkbox"/>
	d) Mobile connection (e.g. e.g. analogue mobile phone, GSM, GPRS, UMTS, EDGE, CDMA2000 1xEVDO)	<input type="checkbox"/> <input type="checkbox"/>
B4.	Did your enterprise use the Internet for the following purposes, in January 2008? (as consumer of Internet services)	Yes No
	a) Banking and financial services	<input type="checkbox"/> <input type="checkbox"/>
	b) Training and education	<input type="checkbox"/> <input type="checkbox"/>
B5.*	Did your enterprise use the Internet for interaction with public authorities, during 2007? (Filter question)	Yes <input type="checkbox"/> No <input type="checkbox"/> → Go to B7
B6.*	Did your enterprise use the Internet to interact with public authorities in the following ways, during 2007?	Yes No
	a) For obtaining information	<input type="checkbox"/> <input type="checkbox"/>
	b) For obtaining forms, e.g. tax forms	<input type="checkbox"/> <input type="checkbox"/>
	c) For returning filled in forms, e.g. provision of statistical information to public authorities	<input type="checkbox"/> <input type="checkbox"/>
	d) For treating an administrative procedure (e.g. declaration, registration, authorisation request) completely electronically without the need for additional paper work (including payment if required)	<input type="checkbox"/> <input type="checkbox"/>
	e) For submitting a proposal in a public electronic tender system (e-procurement) (in the system itself and not by email)	<input type="checkbox"/> <input type="checkbox"/>
B7.	Did your enterprise have a Website or Home Page, in January 2008? (Filter question)	Yes <input type="checkbox"/> No <input type="checkbox"/> → Go to B9
B8.	Did the Web Site provide the following facilities for your enterprise, in January 2008?	Yes No
	a) Product catalogues or price lists	<input type="checkbox"/> <input type="checkbox"/>
	b) Possibility for visitors to customise or design the products	<input type="checkbox"/> <input type="checkbox"/>
	c) Online ordering or reservation or booking, e.g. shopping cart	<input type="checkbox"/> <input type="checkbox"/>
	d) Online payment	<input type="checkbox"/> <input type="checkbox"/>
	e) Personalised content in the website for regular/repeated visitors	<input type="checkbox"/> <input type="checkbox"/>
	f) Advertisement of open job positions or online job application	<input type="checkbox"/> <input type="checkbox"/>
B9.*	Was your enterprise, in January 2008, using a digital signature in any message sent, i.e. using encryption methods that assure the authenticity and integrity of the message (uniquely linked to and capable of identifying the signatory and where any subsequent change to the message is detectable)?	Yes <input type="checkbox"/> No <input type="checkbox"/>

Module C: Automated Data Exchange

(Scope: enterprises with Computers)

Automated data exchange between the enterprise and other ICT systems outside the enterprise means:

- exchange of messages (e.g. orders, invoices, payment transactions or description of goods)
- via the internet or other computer networks
- in an agreed format which allows its automatic processing (e.g. XML, EDIFACT etc.)
- without the individual message being manually typed.

C1.* In January 2008, was your enterprise using such automated data exchange? (Filter question)	Yes <input type="checkbox"/>	No <input type="checkbox"/> → Go to C4
---	------------------------------	--

C2. Was automated data exchange used for the following purposes?	Yes	No
* a) Sending orders to suppliers	<input type="checkbox"/>	<input type="checkbox"/>
* b) Receiving e-invoices	<input type="checkbox"/>	<input type="checkbox"/>
* c) Receiving orders from customers	<input type="checkbox"/>	<input type="checkbox"/>
* d) Sending e-invoices	<input type="checkbox"/>	<input type="checkbox"/>
* e) Sending or receiving product information (e.g. catalogues, price lists, etc.)	<input type="checkbox"/>	<input type="checkbox"/>
* f) Sending or receiving transport documents (e.g. consignment notes)	<input type="checkbox"/>	<input type="checkbox"/>
g) Sending payment instructions to financial institutions	<input type="checkbox"/>	<input type="checkbox"/>
h) Sending or receiving data to/from public authorities (e.g. tax returns, statistical data, [national examples], etc.)	<input type="checkbox"/>	<input type="checkbox"/>

C3. Were the following formats used for the automated data exchange? <i>Optional</i>	Yes	No
a) EDIFACT or similar standards (e.g. EANCOM, ANSI X12)	<input type="checkbox"/>	<input type="checkbox"/>
b) XML based standards, for example ebXML, RosettaNet, UBL, papiNET	<input type="checkbox"/>	<input type="checkbox"/>
c) Proprietary standards agreed between you and other organisations	<input type="checkbox"/>	<input type="checkbox"/>

Go to question D1

C4. Were the following issues reasons for the enterprise not to use automated data exchange? <i>Optional</i>	Yes	No	Don't know
a) No interest in using it, because it isn't relevant for the business	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Lack of expertise in-house for its implementation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Return on the investment too low or not clear	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Lack of appropriate software for the specific sector/size of the enterprise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e) Difficulty with agreeing common standards with business partners	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f) Uncertainty of the legal status of the messages exchanged	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Module D: Sharing electronically information on the Supply Chain Management

(Scope: enterprises with Computers)

Sharing electronically information on the supply chain management means:

- exchanging all types of information with suppliers and/or customers in order to coordinate the availability and delivery of products or services to the final consumer;
- including information on demand forecasts, inventories, production, distribution or product development;
- via computer networks, not only the Internet but also other connections between computers of different enterprises.
- it can be from you to your suppliers/customers or the other way around.

This information may be exchanged via websites or via automated data exchange (recall definition in module C), but it excludes normal e-mail messages manually written.

D1. In January 2008, was your enterprise regularly sharing electronically information on the supply chain management with your suppliers or customers? (Filter question)	Yes <input type="checkbox"/>	No <input type="checkbox"/> → Go to E1
D2. Was your enterprise regularly sharing electronically the following information with its <u>suppliers</u>, in January 2008?	Yes	No
a) Inventory levels, production plans or demand forecasts	<input type="checkbox"/>	<input type="checkbox"/>
<i>Optionally, the 3 items may be collected separately:</i>		
a1) Demand forecasts	<input type="checkbox"/>	<input type="checkbox"/>
a2) Inventory levels	<input type="checkbox"/>	<input type="checkbox"/>
a3) Production plans	<input type="checkbox"/>	<input type="checkbox"/>
b) Progress of deliveries (i.e. distribution of raw materials or finished products)	<input type="checkbox"/>	<input type="checkbox"/>
D3. Was your enterprise regularly sharing electronically the following information with its <u>customers</u>, in January 2008?	Yes	No
a) Inventory levels, production plans or demand forecasts	<input type="checkbox"/>	<input type="checkbox"/>
<i>Optionally, the 3 items may be collected separately:</i>		
a1) Demand forecasts	<input type="checkbox"/>	<input type="checkbox"/>
a2) Inventory levels	<input type="checkbox"/>	<input type="checkbox"/>
a3) Production plans	<input type="checkbox"/>	<input type="checkbox"/>
b) Progress of deliveries (i.e. of distribution of raw materials or finished products)	<input type="checkbox"/>	<input type="checkbox"/>
D4. Were the following methods used for the electronic exchange of this information, in January 2008?	Yes	No
a) Websites (yours, those of your business partners or web portals)	<input type="checkbox"/>	<input type="checkbox"/>
b) Automated data exchange (XML, EDIFACT, etc.)	<input type="checkbox"/>	<input type="checkbox"/>

F4.*	Was your enterprise using a secure protocol, such as SSL and TLS, for the reception of orders via Internet, in January 2008?	Yes <input type="checkbox"/>	No <input type="checkbox"/>
Orders placed via computer networks (Purchases)			
F5.*	Did your enterprise send orders for products or services via computer networks, during 2007 (excluding manually typed e-mails)? (Filter question)	Yes <input type="checkbox"/>	No <input type="checkbox"/> → Go to G1
F6.*	Please indicate for 2007 the percentage of orders that were sent electronically in relation to the total purchases' value (in monetary terms, excluding VAT).	Less than 1% <input type="checkbox"/>	
		1% or more and less than 5% <input type="checkbox"/>	
		5% or more and less than 10% <input type="checkbox"/>	
		10% or more and less than 25% <input type="checkbox"/>	
		25% or more <input type="checkbox"/>	
<i>Alternative Question:</i>			
Please state the value of the purchases resulted from orders that were placed electronically (in monetary terms, excluding VAT), in 2007.			<input type="text"/> (National Currency)
If you can't provide this value, Please indicate an estimate of the percentage of the total purchases that resulted from orders placed electronically, in 2007.			<input type="text"/> %

Module G: Perceived benefits of the use of ICT

(Scope: enterprises with Computers)

The **implementation of an ICT project** refers to the introduction of a new or updated ICT (e.g. a new/updated software application or a new/updated hardware) or a change in the use of an existing ICT.

Examples of ICT projects are: a new or a restructured website, a new internal homepage, the starting of using automated data exchange or the starting of receiving orders via computer networks.

G1.	In January 2008, to what degree have ICT projects implemented in the last 2 years caused improvements in the following areas, compared to the previous task handling? <i>If your enterprise has not had any ICT projects, please tick all boxes 'not applicable'.</i> Optional	Minor / None	Moderate	Significant	Don't know / Not applicable
	a) Reorganisation and simplification of work routines	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	b) Release of resources	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	c) Higher earnings for the enterprise	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	d) Development of new products and services	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Module X: Background information

(X1-X5) available in some countries from SBS and thus not to be included; latest available information should be provided

X1.	Main economic activity of the enterprise, during 2007	<input type="text"/>
X2.	Average number of persons employed, during 2007	<input type="text"/>
X3.	Total purchases of goods and services (in value terms, excluding VAT), for 2007	<input type="text"/>
X4.	Total turnover (in value terms, excluding VAT), for 2007	<input type="text"/>
X5.	Location (Convergence/ non-Convergence region), in 2007	<input type="text"/>

I.3.2. Transmission format

General Enterprise Survey

Version 1.1, 1/8/2008.

Note: this transmission format refers to the 2008 survey.

⇒ See document ***MM2008_AnnexI.3.2_TransmissionFormatENT2008.doc***

I.3.3. Reporting template

Note: the reporting template for metadata (interim) & quality (final) report refers to the 2008 ICT enterprise survey.

⇒ See document ***MM2008_AnnexI.3.3_ReportingTemplateENT2008.doc***

Part II

Household survey

II.1. STATISTICAL PRODUCT

The statistical product is the clear and precise definition of the statistical information to be produced. It distinguishes itself from the production methodology. The production methodology is the way or method of doing, while the statistical product is its direct result. Different statistical methodologies can produce the same statistical product, being only different ways of doing the same thing. This means that as long as we guarantee that two figures concern the same statistical product, for instance for two different countries, they are comparable. This way, this distinction between the statistical product and the statistical methodology helps us to focus on those elements more important to assure comparability between the several national statistics and produce new ones at the EU level, i.e. the statistical product, while leaving to the discretion of the national statistical institutes the choice of the better statistical methodology to be applied in its own country taking into consideration its own specificities.

The elements that make up the statistical product, at an input level, are the statistical unit, the target population and the observation variables, and at the output level, the periodicity and the summary measures, aggregate variables and tabulation. Covering all the elements of the statistical product, the statistical concepts and the nomenclatures are the additional needed element to assure harmonization and comparability of statistics.

II.1.1. Statistical unit

The statistical unit is the base type of the elements of a group (also called population) that we want to observe or analyse. The basic statistical operations of classification, aggregation and ordering are done on the statistical unit.

The choice of the statistical unit is a matter of both the data collection process (namely the operational restrictions associated to collecting data from each type of statistical unit) and the conceptual framework chosen to observe and analyse the phenomenon. The statistical unit is the bearer of statistical characteristics or attributes, which we ultimately want to measure.

There are several types of statistical units, according to its usage. An *observation unit* represents an identifiable entity, about which data can be obtained. During the collection of data, this is the unit for which data is recorded. It should be noted that this may, or may not be, the same as the reporting unit. The *reporting unit* is the unit that reports to the survey authority. It reports information for the observation unit(s). In certain cases it may be different from the observation unit. A reporting unit is a unit that supplies the data for a given survey instance.

In the ICT usage survey, the following statistical units are used (depending on the variable):

- households;
- individuals.

Ideally, the data collected on the household, should be reported by 'the household'. In most cases, it is of course not feasible to put the household around the table in order to collect their common answer, this is especially the case in a telephone interview. In general, one individual in the household will answer the household related questions having the household perspective in mind. This one individual can for instance be the head of the household or the individual which has been selected for the individual questions.

As the survey relates to one's usage of computers and Internet, it is necessary that the selected individual answers the questions personally. Proxy interviews may lead to errors in the data collection.

Different survey units, i.e. households and individuals, are used in different sections of the model questionnaire. The household approach is used when information on access to different electronic devices, type of Internet connection, and barriers to use of the Internet is collected. The individual approach is used when information on use of computers, use of the Internet, e-commerce and e-skills is collected.

The reasoning behind using a household approach when trying to describe access to ICT is that households are the platform for providing access for a number of individuals, i.e. the members of households. In order to make comparisons between European countries on household level it is important to take into account the differences in household demographics. One of the reasons for this is that some countries might have bigger household sizes, and composition than other countries, which also probably is correlated with the rate of access. To be able to make fruitful comparisons it is also important that countries deploy the same household definition.

The individual approach is used in the context where the collected information to a larger extent refers to the individual's use of ICT. Also in this case it is important to take into account structural differences between countries, e.g. age demographics, in order to make fruitful comparisons. The problem is however not as large as when it comes to households since many people have a better knowledge of countries' differences in age structure than the differences in household structure.

II.1.2. Target population

A population is a collection of objects of the same class, which in statistical terms means a group of elements of the same statistical unit. There are two types of populations to be considered when producing statistics: the target population and the frame population.

The *target population* is the population of interest. It is defined by delimiting clearly the group of statistical elements for which some information concerning the all group we want to know. That delimitation is based on one or more attributes of the statistical unit. In the ICT usage survey, the target population for the different statistical units is:

- individuals: target population consists of all individuals aged 16 to 74;
- households: target population consists of all (private) households with at least one member aged 16 to 74.

The *frame population* is an operationalisation of the target population, consisting ideally in a complete list of the elements of the target population. Although a target population can be easily defined, in practice a list of all its elements is needed for its complete or partial (in case a sample is used) observation, and that can be very difficult to obtain. That list should be complete and include every element of the target population only once. However, most of the time it will suffer from both under-coverage and over-coverage. Usually files of statistical elements (registers) are maintained and updated, containing lists of statistical elements and also information on some attributes, usually used for delimiting target populations. Frame populations are usually extracted from those registers. The frame population will be further explained in chapter II.2.2, including an overview of the countries' practices for the 2005 survey.

II.1.3. Periodicity

The periodicity is annual, meaning the data is collected and compiled once per year. However, some variables can be observed with a lower frequency, e.g. variables which tend to be stable over time.

An annual survey should be a compromise or trade-off between response and collection burden and the need for relevant and recent information on a “fasting moving” study domain like the information society.

This periodicity is laid down in paragraph c) of Annex II of Regulation 808/2004 and in section 5 of Regulation 847/2007 (see also *Part III – EU Legislation*).

II.1.4. Observation variables

In the ICT usage survey, most of the observation variables are *qualitative*, i.e. the aim is not to collect information on quantities (“how many ...”), frequencies (“how often ...”) or amounts (“how much ...”) but rather to obtain non-numerical or categorical information. In most cases, the observations variables in the ICT usage survey are binary (dichotomous), meaning the respondent answers with *yes* or *no*, e.g. *Have you used the Internet for selling goods or services*. The final statistics will then mainly be proportions (e.g. the number of *yes* answers divided by the number of respondents who answered the question). Other observation variables are qualitative too, but with more than two answering categories (e.g. highest educational level).

The attributes or characteristics for the Community survey on ICT usage in households and by individuals are listed in Part III of this document (Annex II of the Commission Regulation).

The operational version of the observation variables are the questions in the survey questionnaire. A model for such questionnaire can be found in Annex II.3.1.

The definition of the observation variables or the model questions are discussed in more detail in the explanatory notes below, see *II.1.6. ff.*

II.1.5. Summary measures, aggregated variables, indicators and tabulation

Summary measures are numerical values defined by a statistical measure that is used to summarize the values for a specific variable for all statistical units in a specific group. Such measures can take the form of aggregates (e.g. total number of *yes*-answers on a specific questions) or indicators (e.g. percentage of *yes*-answers).

Aggregates can be compiled for the total population or for the different subpopulations defined by the background variables (e.g. gender or educational level) or for subpopulations defined conditional on the answers of other study variables (e.g. Internet users versus non users).

E.g.:

- number of individuals having used a computer in the last 3 months at home
- number of individuals aged 16-24 having used a computer in the last 3 months at home
- number of individuals with having used a computer in the last 3 months at home but not at work
- number of Internet users having used it for downloading official forms (= number of individuals having used the Internet and having downloaded official forms)

To obtain the indicators (proportions, percentages, etc.), the aggregates need to be divided by the total population or subpopulation.

The denominator of such ratio can for instance be:

- total population of individuals in selected strata (e.g. "proportion of persons aged 16-24 having used a computer at home" = the *number of individuals aged 16-24 having used a computer in the last 3 months at home* divided by *the total number of individuals aged 16-24*)
- total number of computer users (individuals that used a computer at least once in last 3 months)
- total number of regular computer users (individuals that used a computer at least once a week in the last 3 months)

The detailed transmission format for sending the aggregated data to Eurostat, can be found in Annex II.3.2 *Transmission format*.

II.1.6. Explanatory notes

The explanatory notes in this chapter refer to the questions in the 2008 model questionnaire (see *II.3.1 Model questionnaire*). The structure of this chapter follows the model questionnaire, i.e. the explanatory notes are grouped per *module* and per *question*. It is recommended to have the model questionnaire at hand while reading this section.

For reasons of continuity, the explanatory notes for questions that are not part of the 2008 model questionnaire (but were part of the 2007, 2006 and 2005 model questionnaires) have only been retained if necessary. More information on former questions can be found in the 2007 and 2006 manuals.

The statistical unit for Module A is the household while the individual is the statistical unit for the other modules.

II.1.6.1 MODULE A: ACCESS TO INFORMATION AND COMMUNICATION TECHNOLOGIES

→ Statistical unit: households

A1: Do you or anyone in your household have access to a computer at home?

[**Scope:** all households]

[**Type:** one single answer needed, i.e. *Tick only one* ; binary (Yes/No)]

For a definition of a household see below, questions F10 and F11 of *Module F: Socio-demographic background variables*.

Access at home refers to the dwelling unit where the household/individual has its address.

Compared to the 2006 survey asking about any devices at home, the question addresses access to a computer of any type, i.e. desktop, laptop or handheld (for further details, see question A3).

"You or *anyone in your household*" may make it clear to the respondent that the questions also concern other members (thus avoiding that the respondent excludes e.g. the case where his/her partner has a computer via work). The wording *have access to* is preferred to *have* because the latter may erroneously be interpreted in the sense of 'ownership' (while the only computer in the household may actually be the property of the employer of one of the members in the household).

A2: Does any member of the household have access to the Internet at home, regardless of whether it is used?

[**Scope:** all households]

[**Type:** one single answer needed, i.e. *Tick only one* ; binary (Yes/No) + *Don't know* ; filter question]

'Access' does not refer to the "connectability" (i.e. can connections be provided in the households' area or street), but to whether anyone in the household could use the Internet at home if desired, even if just to send an e-mail.

The clause 'regardless of whether it is used' will in most cases be irrelevant as households will normally make use of a service they have subscribed to (and are paying for). It is however possible that the connection was installed by one of the household members' employer or is a default facility in the building where the family is living (without actually being used).

The *Don't know* answering category should be avoided by asking the respondent additional questions. However, in some cases the household member selected for the interview may not be aware what other members are doing with e.g. the household's computer.

In certain countries, it is possible to access the Internet without subscription contracts, but for instance by ad hoc payments whenever the Internet is actually used. This can be in form of sending a text message to receive a short duration access code. In this situation, a lot of households have the *possibility* to access the Internet from home (of course under the

condition they have the necessary equipment, f.i. computer+modem or Internet enabled mobile phone) but will *not be using* it.

A3: On which of these devices is the Internet accessed at home?

[**Scope:** all households with Internet access at home, i.e. A2 = Yes]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

The following devices can provide access to the Internet.

a) Desktop computer

A desktop computer can be defined as a computer that usually remains fixed in one place. Normally the user is placed in front of it, behind the keyboard with the monitor placed on top.

b) Portable computer (laptop)

A portable computer (or laptop) can be defined as a computer that is small enough to carry. A portable computer (or laptop) can usually carry out the same tasks as a desktop computer. Includes notebooks but not handheld computers, although portable.

c) TV set with specific Internet device (e.g. digital TV or set top box)

The main issue is if the TV set with the specific device enables Internet access. That means the specific Internet device should allow users to access the WWW or e-mail service.

Digital TV refers to television via digital signals. This is the basic difference compared to ordinary TV transmission which is via analogue signals. Improvement of picture quality and Internet access are some of the features that characterize digital TV. However, there could be new, popular digital TV sets in some countries which more or less only allow watching TV and video on demand.

Today's analogue TV needs e.g. a set top box, which is a device that sits on top of the TV set and is the interface between the home TV and the cable TV company. The set top box changes the digital signal from the cable TV company so the home TV (which is analogue) can read the signal.

d) Games console

Some games consoles have built-in modems and thus give access to the Internet, e.g. Playstation 2 (Playstation 3 is planned to be available in Europe from March 2007). A games console is an electronic device designed to play games. Examples are Sony playstation, Nintendo GameCube and Microsoft Xbox.

e) Other means (i.e. none of the above)

Applies where the household does not have access to the above mentioned devices at home.

Item e) *Other means* in the 2007 model questionnaire lists two optional sub-items, namely:

- e1)** via Internet enabled mobile phone
- e2)** via handheld computer (palmtop, PDA)

e1) Internet enabled mobile phone

The household is asked if it is an Internet enabled mobile phone via WAP, GPRS or UMTS (definitions: see below). This should be regardless whether it is used or not. Internet can be defined as a networking infrastructure, where any computer or other device can communicate with each other as long as they are both connected to the Internet. The Internet is used for access to the World Wide Web and is also used for e-mail correspondence, instant messaging etc. During an interview, questions such as 'do you use picture messaging' or 'does your mobile phone have a built-in camera' can help to detect whether a respondent has an Internet enabled mobile phone.

The term GPRS (General Packet Radio Systems) stand for 2nd generation mobile technologies (though GPRS is sometimes referred to as 2,5 generation) while UMTS (Universal Mobile Telephone System) is the term for 3rd generation mobile technology. Both give access to the Internet but 3rd generation at a much higher speed.

Compared to 2006, the term 'WAP' (Wireless Application Protocol) has been deleted in the optional category *e1*, because WAP or i-mode are protocols enabling viewing web pages written by WAP language programs using this standard on a mobile phone screen (see also question A4 and C4).

e2) Handheld computer (palmtop)

A handheld computer is a computer that can be used while you are holding it and stored in a small bag or for instance in a pocket. Personal Digital Assistants (PDAs) are another term for handheld computers. A PDA uses a pen rather than a keyboard.

As some PDAs can also be used for mobile telephony (e.g. hi-tech mobile phones with some kind of integrated PDA, such as Sony Ericsson P800 and P900 series), there is an overlap with item b) of this question, i.e. these PDAs can be counted for as both a mobile phone (item *b*) and a handheld computer (item *f*).

These items e1) and e2) have been rendered optional as they come back in the individual part of the questionnaire (see Module C, question C4).

A4: What types of Internet connection are used?

[**Scope:** all households with Internet access at home, i.e. A2 = Yes]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

Below is given an explanation of the most commonly used technologies. If the NSI finds it more appropriate to use the brand names or company names than using the references to the

technologies (ADSL, cable, etc.), this is also a possibility (of course under the condition the brand names can be recoded to the different technologies based on a one-to-one relationship).

a) Modem (dial-up access over normal telephone line) or ISDN

Connects the household via dial-up either through an analogue modem or ISDN (Integrated Services Digital Network). Both types can be categorised as narrowband.

b) DSL (e.g. ADSL, SHDSL etc.)

DSL (Digital Subscriber Line) stands for technology that transports data at high speeds (i.e. equal to or higher than 144 kbit/s) over the existing copper network. DSL technology is a type of Internet connection which is significantly faster than dial-up access and therefore designated as broadband.

Important DSL technologies include:

ADSL: This term is used for DSL where more bandwidth is allocated to download than upload (Asymmetric DSL).

SHDSL: Single pair high-speed DSL covers symmetric high speed DSL.

c) Other broadband connection (e.g. cable, UMTS, etc.)

This question sums up all other types of Internet connection which is not provided via DSL but still can be designated as broadband.

The technologies in question are:

Cable modem: This technology allows high-speed Internet access to be delivered over a cable TV network or communal aerial systems. Like DSL technologies the data is transported over the existing copper network.

LAN (Local Area Network): High-speed Internet connections via housing networks (connections common to several households, e.g. residents in a college).

Fibre optic cable: A technology which is not based on copper wire. The technology uses lasers or light emitting diodes and can provide unlimited bandwidth potential.

Fixed wireless technologies (FWA): Technologies in this group are all wireless technologies that provide access between fixed points.

Satellite technologies: Provides wireless access and are predominantly used in remote areas not reached by cable or DSL networks.

UMTS: Relates to mobile phone over broadband. For a definition, see question A3 e1).

d) Mobile phone over narrowband (WAP, GPRS, etc.)

Broadband over the mobile phone is still limited (see UMTS/3G above). Second generation (2G) mobile technologies such as GPRS (General Packet Radio Service, see also question A3 for a definition) are designated as narrowband connection. GPRS is a low capacity speed mobile system which enables Internet access. GSM, also 2G where transmission is done via circuit switched mode, is intentionally not mentioned as example. WAP is actually a protocol and not a mobile phone system, see question A3 e1 above.

A5: What are the reasons for not having access to the Internet at home?

[**Scope:** all households without Internet access at home, i.e. A2 = No]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

The question gives an idea of the household reasons or barriers for not having access at home. Respondents should indicate all reasons that are relevant to them, not only the main reason(s).

a) Have access to the Internet elsewhere

All other places, e.g. workplace, but not at the home address.

b) Don't want Internet

May apply to people that find content on the Internet harmful or have a principle objection towards Internet use.

c) Don't need Internet

May apply to people that do not find it useful or interesting.

d) Equipment costs too high

People in this category find equipment costs too high, that is, the previously defined devices in question A1 (including the set-up costs, e.g. the cost of a modem for internet connection).

e) Access costs too high (telephone, etc.)

Access costs can be telephone charges, subscription fees for broadband use etc.

f) Lack of skills

The respondent does not have sufficient knowledge to use or to get access to the Internet. The respondent may answer that he/she perceives Internet access as too complicated. Then this item will apply since the barriers lies in a lower level of skills compared to others.

g) Physical disability

h) Privacy or security concerns

The respondent is worried about giving personal details over the Internet due to private or security reasons. Respondents might also be worried about virus attacks and similar security problems even without giving personal details over the Internet. In this respect, the spreading of spyware is a major concern (spyware is a program which has installed itself on the computer to monitor and report on the respondent's computer activity; the development of spyware has also led to numerous anti-spyware tools to prevent spyware activity on the computer).

i) None of the above, but other

This category is included to close the question. Assuming that respondents that do not have Internet at home must have a reason for not having it, the respondent is supposed to tick at least one of the options *a* to *h* or option *i*. In cases where the respondents states not to

have any particular reason (not even after having heard the options in the questionnaire) *None of the above, but other* can be ticked.

II.1.6.2 MODULE B: USE OF COMPUTERS, LOCATION AND FREQUENCY OF USE

→ Statistical unit: individuals

B1: When did you last use a computer?

[**Scope:** all individuals]

[**Type:** one single answer needed, i.e. *Tick only one* ; filter question]

A computer should be defined as multi purpose machine, a personal computer, powered by one of the major operating systems, i.e. Macintosh (Apple), Linux or Microsoft (Windows, XP or NT, Vista). PDA's (handheld computers or palmtops) should also be included. Other equipment with embedded computing abilities, e.g. cell phones, VCR's, TV sets, washing and dishing machines should be excluded.

B2: How often on average have you used a computer in the last 3 months?

[**Scope:** all individuals having used a computer in the last 3 months, i.e. 1st option in B1]

[**Type:** one single answer needed, i.e. *Tick only one*]

For definition of a computer see above, B1.

Persons who use a computer at work on a daily basis during the week but who don't use a computer at home during the weekend should tick option *a) Every day or almost every day*.

B3: Where have you used a computer in the last 3 months?

[**Scope:** all individuals having used a computer in the last 3 months, i.e. 1st option in B1]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

For definition of a computer see above, B1.

Concerning alternative *b) At place of work (other than home)*, this should be interpreted as any place where work is usually conducted excepting at home. The usual place of work can be much wider than just the premises of the employer.

Alternative *c) At place of education*, should be interpreted as the institution where the education takes place ("within the school walls"). Teachers using computers in their school should tick alternative *b) At place of work* as alternative *c)* rather refers to the point of view of pupils or students.

See also C3 (location of Internet use).

B4: When did you last take a training course (of at least 3 hours) on any aspect of computer use? - Optional -

[**Scope:** individuals who already used a computer, i.e. 1st, 2nd or 3rd option in B1]

[**Type:** one single answer needed, i.e. *Tick only one*]

The emphasis is on formal training via courses (education) and excludes informal training such as help from colleagues or friends/relatives and learning-by-doing. All kinds of courses which can relate to computer use have to be included, e.g. word processing or spreadsheet courses, programming, web-design and so forth.

An e-learning course is included if it lasted at least 3 hours, see also explanation for item C5k online course.

II.1.6.3 MODULE C: USE OF THE INTERNET

→ Statistical unit: individuals

C1: When did you last use the Internet?

[**Scope:** all individuals (even those who reported never having used a computer in B1 as they may have accessed the Internet via other means, though this may be more exceptional)]

[**Type:** one single answer needed, i.e. *Tick only one* ; filter question]

This question covers any use of the internet - whether at home, at work or from anywhere else and whether for private or professional purposes.

The time breakdowns are self explanatory: *Within the last 3 months; Between 3 months and a year ago; More than one year ago; Never used it.*

C2: On average how often did you use the Internet in the last 3 months?

[**Scope:** individuals who used the Internet in the last 3 months, i.e. 1st option (*Within the last 3 months*) in C1]

[**Type:** one single answer needed, i.e. *Tick only one*]

The response categories to this question could be slightly ambiguous. Respondents should therefore be presented with all four options and should select which one best describes their behaviour.

a) Every day or almost every day

If asked this should refer to more than 4 days each week. As in question B2, persons who use the Internet at work on a daily basis during the week but who don't use it at home during the weekend should tick option *a) Every day or almost every day*.

b) At least once a week (but not every day)

If asked this should refer to between 1 and 4 days each week

c) At least once a month but not every week

d) Less than once a month

C3: Where have you used the Internet in the last 3 months (using a computer or any other means)?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

The question is to be answered regardless of the device used or type of connection used. E.g. if a person uses a portable computer with a wireless data card (to be inserted in the card slot) at several locations, he or she should tick all those locations.

a) At home

This should be ticked if the respondent has used the Internet at home for any purpose, private or work-related.

b) At place of work (other than home)

This should be defined as usual place of work. If the respondent usually works at home then *a)* should be coded. If the respondent's work is usually based in a number of places then any of these places count as their place of work. However, use of mobile technology in places such as hotels etc. even for work related purposes, should be coded as *e) At other places*.

c) At place of education

This should refer to the respondent's own place of education. If the respondent works at an educational establishment – e.g. a teacher – then *b)* should be coded.

d) At another person's home

This should not be used if the respondent's place of work is another person's home.

e) At other places

This includes ANY places not already coded above. The further responses are a subset of places which might be included in *At other places* but is NOT a comprehensive list. On the other hand, not all items of the list may be relevant to each country.

Ideally the respondent should be presented with the further list only after they have said "other place"

of which:

e1) Public library: some public libraries offer the possibility to access the Internet, whether it be a free service or if a charge is made.

e2) Post office: see definition below.

e3) Public office, town hall, government agency: The wording to be used in each member state should reflect titles of public offices which respondents would recognise.

e4) Community or voluntary organisation: see definition below.

e5) Internet café: Commercial places that provide Internet access services under payment. There is usually a charge for using the Internet. This charge is normally calculated on the basis of the amount time spent on-line. An Internet Café may also charge for usage not related to the internet (e.g. use of word processing software).

Post offices, Public offices (town hall, government agencies), Community or voluntary organisations are all public places capable of providing access to the Internet. The access to the Internet in these public places can be free or under payment of a fee.

e6) Hotspot: The Internet can be used at a so-called hotspot with portable or handheld computers via wireless connection at hotels, airports, public places, etc.

C4: Do you use any of the following mobile devices to access the Internet?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

This question refers to the actual *use* as a means of access not to the 'theoretical' possibility of access. Question C4 is not limited to Internet access at home, but can refer to all possible locations (as is in general the case for Module C).

Via *Mobile phone* has been split into two items in order to separate UMTS (see item **b**) from GPRS (see item **a**). First, because of a high interest from users of ICT statistics in the take-up and usage of UMTS. Second, because this also enables to distinguish between mobile phone access over broadband versus mobile phone access over narrowband.

For additional information on GPRS, UMTS or *Handheld computer*, see Module A, question A3 (items e1 and e2).

Item **d**) portable computers (laptop) via wireless connection away from home or work as an important device to access the Internet remotely has been added in 2007 and is continued. To distinguish real mobile (in the sense of 'on the move') use from situations where the portable computer serves as a desktop at home or at work, the (longer) wording is supposed to exclude such cases.

C5: For which of the following activities did you use the Internet in the last 3 months for private purposes?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

This set of responses refers to personal or private use only (ie. NOT for work related purposes) but access could have been from anywhere – including place of work.

Compared to earlier Community surveys, the list of items in this question has been reduced due to transfer to the new module D in 2008 on advanced services.

Communication, information search and on-line services

a) Sending / receiving e-mails

It includes the use of e-mail for sending messages to friends or for getting information on goods/services.

b) Finding information about goods or services

Using the Internet to seek for information about any household good, for example, films, music, video-games, books, e-learning material, clothes, electronic equipment computer software or services for example banking, financial or health services. It should not include transactions, e.g. purchases of any goods or services (although one will usually look up information on a good or services before actually purchasing it).

c) Using services related to travel and accommodation

Includes using the Internet for ascertaining information or to purchase goods and services in relation to travel and accommodation, for example travel tickets, hotels or any other type of accommodation or web sites containing information for tourists.

d) Downloading software (other than games software)

Includes downloading software either free of charge or under payment. It does not include software for playing computer and video games which is addressed in item D3f.

e) Reading or downloading online newspapers / news magazines

This should include all types of online newspapers and magazines either free of charge or under payment.

A sub-item has been introduced for 2007 and is continued to distinguish subscription to news services from news consulted on-screen:

e1) This refers to subscribing to news services or news products, both for free or paid, which are then received regularly.

f) Looking for a job or sending a job application

Includes searching specific web sites for job 'hunting' or for sending an application for a job. Sending a job application should be included in this category only if it was sent on-line.

g) Seeking health-related information (e.g. injury, disease, nutrition, improving health, etc.)

This item refers to Internet use for health related activities. The scope is limited to private purpose, professional use is not to be taken into account. Private should however not be limited to own personal use, but can also include internet use for health related activities on behalf of other family members or friends.

Includes general searches via a search engine (Google, Yahoo!, ...) using keywords in one of the mentioned fields. This item also includes more specific searches on specialised websites such as the Ministry of Health, non-governmental bodies or interest groups. The respondent may have obtained the website's address from a folder or an article.

Although not yet available in most countries, this item also includes activities such as making an appointment on-line with a practitioner, requesting a prescription on-line from a practitioner or seeking medical advice on-line from a practitioner.

Banking, (optional: selling of goods or services)

h) Internet Banking

This includes electronic transactions with a bank for payment, transfers, etc. or for looking up account information.

Electronic transactions for other types of financial services are not covered by this category (purchase over the Internet of shares, financial services, insurances is covered by item *i*) of question D2, see Module D. A simple information search on e.g. shares or financial services is of course included under item *d*) *Finding information on goods and services*.

i) Selling goods or services (e.g. via auctions)

Selling goods or services on-line does not require an *electronic* payment transaction, i.e. the transaction or 'deal' is done on-line but the payment and/or delivery can take place off-line. Putting an advertisement on a website to, for example, sell a second-hand bicycle or a spare ticket for an event, should not be included here as the transaction is in general not concluded on-line in an automatic manner (but via a phone call or informal e-mail). In this case, the individual could tick item *l*) *Other information search or on-line service* as he/she is making use of a website (e.g. a forum) to offer the product or service.

Training and education

In 2007 the activities related to training and education are more specified regarding on-line activities.

j) Looking for information about education, training or course offers

Includes searching for information relating to course offers at school, college or university whilst part of the education system, but also post educational course offers including leisure courses for gaining and improving knowledge in any subject. Post graduate courses at university should be included.

k) Doing an online course (of any subject)

Doing an online course reflects learning courses distant from the location of education and training organisations or employer where courses can be attended in person (often but not necessarily done at home). Interaction with teachers, trainers and/or learning material is done via the Internet. The use of e-learning software programmes can play a role. This item applies also if a course is only partially done online.

l) Consulting the Internet with the purpose of learning

This new item for 2007 relates to using the Internet for self-directed learning.

Self-directed learning is wider than simply looking for information. It refers to looking for information with the purpose of learning something, of improving knowledge. Looking up a city map or a phone number would not be considered as learning, contrary to e.g. looking up definition of broadband technology or information on an economic concept. To help distinguish from more random finding of any information, the 'stronger' verb *to consult* is used ('consulting the Internet' in the sense of 'seeking advise on the Internet'). The words *with the purpose* is added to stress the *intention* of learning (compared to just looking up some 'daily life' information).

C6: For which of the following activities relating to interaction with public services or administrations and during which period did you use the Internet for private purpose?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

The question is based on the e-government related items in previous surveys. An additional reference period for the recent Internet users of "in the last 12 months" has been included. This is because of the expected irregular or seasonal nature of interactions with public administrations.

Interaction with public services or administrations

These three items make use of the conceptual framework for the levels of interaction between enterprises or citizens and public authorities, used by DG INFSO. The framework of four levels of interaction was developed to measure the level of sophistication offered by websites supplying basic public services. That way, one can distinguish websites that only provide information on a public service (1st level) from websites that allow to download forms (2nd level) or to submit information in an online form (3rd level) up to websites that allow to completely treat the public service or administrative obligation without the need for any paperwork whatsoever (4th level of sophistication). For the time being, only the first three levels are considered in the household survey.

It is expected or logical that a particular website that reaches a certain level of sophistication also include all lower levels, for instance a website offering the possibility to download a form should certainly provide information on the particular public service that form would refer to. The concept of the four levels of sophistication was adopted for the measurement of the usage of e-government by enterprises (and the 3 first levels for citizens). However, the interpretation of the results will have to differ slightly from the interpretation done for the sophistication of the public websites. At usage level, it is natural to assume that a user (enterprise or citizen), before knowing which form to download (2nd level), will obtain information on the public service he is interested in (1st level), but he/she will only do it once or twice and not every time it downloads the form. Therefore, when measuring the usage every year, it is perfectly possible that there are users who have downloaded forms with obtaining information in those cases in which the user already had all the information he needed (although not very probable, as users may well constantly search for information on several different issues). In this sense, it is possible that the respondents answers *Yes* to item *p* without having answered *Yes* to *o* (or *Yes* to *q* without having answered *Yes* to *o* and/or *p*).

o) Obtaining information from public authorities' websites

Includes searching to obtain any type of information from public authority web sites. Public authorities' web sites include local or central government offering information or services. It would be helpful to provide a list of local/central government web sites in the format of a showcard to facilitate respondents to recall public authority web sites that they have used. For example respondents may not realise some public information sites are run by public authorities, e.g. a meteorological office offering weather forecast services.

p) Downloading official forms

Includes downloading official forms from public authorities' websites for any purpose of use (e.g. for information or for requesting a service). The downloaded files could be in formats such as PDF or Microsoft Word.

q) Sending in filled forms

Includes filled in forms sent via internet (public authorities' websites) only. Forms downloaded, printed, filled in and sent by post should not be included in this category.

The downloading or sending or transmitting filled in forms may involve the need to "sign in" at the website and provide personal details.

II.1.6.4 MODULE D: USE OF ADVANCED SERVICES

→ Statistical unit: individuals

D1: Did you use the Internet in the past 3 months for private purpose for the following communication activities?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

a) Telephoning over the Internet

Telephoning over the Internet is a relatively inexpensive method to communicate and is often the method used by companies offering reduced cost telephone charges. Users may not be aware that they are communicating using such Internet-based telephony, Voice-over-IP or VoIP.

Next to VoIP, peer-to-peer telephony is becoming more important. The user needs to install a little programme (such as Skype) for making free calls over the internet to anyone else who also has this software. Usually, one can also make calls to normal fixed or mobile lines via a pre-paid credit.

b) Video calls (via webcam) over the Internet

The former C5b item in 2007 was not found relevant since videoconferencing is normally understood as being used by companies and for business related purposes and instead of face-to-face meetings. Often it is not based on IP network.

The new item video calls reflect more current trends in live audio and visual communication between persons. It is simply referring to web camera as additional mean and via Internet (IP based networks). Web cameras, usual small digital cameras are used. The features can be made accessible in programmes for Internet calls (such as Skype).

c) Posting messages to chat sites, newsgroups or on-line discussion forum

Chat sites can be used by two or more persons for the purpose of communication. This communication is by written word (similar to email). The identity of the person 'chatting' does not have to be divulged to other users (e.g. a username can be created). Users must be 'logged-on' to a chat site to use it. Users that are logged on to a specific chat site can see all correspondence between all users on that site. Correspondence is updated in real time.

A chat site or also called a chatroom can be defined as a virtual room, where participants have a chat session. Newsgroups are on-line discussion groups covering all kinds of interest. The member of a newsgroup view and post messages via a news server on the Internet.

Social networking sites (e.g. MySpace, Facebook) have become popular recently. Although they are mostly used for uploading and sharing of audio and video content, they also allow posting of messages and participation in forums on specific topics of interest.

d) Use of instant messaging

Instant messaging means real-time communication between people on the basis of typed text.

Compared to e-mail applications, it is possible to know if the person or persons addressed is online or away and communication is often seen as more easy and efficient.

e) Reading weblogs or blogs

A weblog or blog is a website where entries are made such in a journal or diary. A typical blog combines text, images, and links to other blogs, web pages, and other media related to the topic of interest.

f) Creating or maintaining own weblog or blog

Users of the Internet can create their own content (text, images etc) and communicate it to or share it with others via the website they have set up.

D2: Did you replace with your Internet calls the following other means of communication?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1) and who have used telephoning over the Internet and/or video calls (via webcam) over the Internet (i.e. option a or b in D1)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

This question will measure the impact on substitution by Internet calls.

- a) Mobile phone calls
- b) Use of fixed telephone line (not linked to Internet)
- c) Use of e-mail
- d) No effect on other communication means

Respondents may not have replaced any of the above a), b) or c) items. In addition, there could be other electronic means of communication, such as instant messaging, which should be taken into account. The item stands alone and cannot be derived from not having ticked any of a), b) or c).

The answers options very much, to some extent should be left to the personal assessment of the respondents. If guidelines are needed, very much would certainly apply for example if more than half of mobile phone calls were replaced. If less than half was replaced, to some extent would be applicable.

If a mobile phone, fixed telephone line or e-mail is not available or used, none of the boxes under a), b) or c) - very much, to some extent, not at all - is ticked.

D3: Did you use the Internet in the last 3 months for the following leisure activities related to obtaining and sharing audiovisual content?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

This new question is a combination of existing items from 2007 such as C5f, C5g and E4e, and new items in particular on podcasting and user created and shared content.

The term "audiovisual" content, it refers to audio content, i.e. to listen to sound, music or spoken language, and visual content, i.e. to see, watch pictures, movies, or videos.

Actual users generally know about terms like weblogs, podcasting and browser based news feeds such as RSS. Further explanations will be provided below.

a) Listening to web radios / watching web television

This covers both live streaming (real-time) and radio or TV 'on demand' (batch, i.e. the user can listen/watch programs later on). This does not result in any saving of information to, for example, the respondents computer.

b) Downloading and/or listening to music (other than via web radio)

This refers to actually downloading music for listening or using any music files at websites for direct listening, but not live streaming.

c) Downloading and/or watching movies, short films or video files (other than via web TV)

This refers to actually downloading movies, short films or videos for watching or directly using any of such files stored at websites for direct watching, but not live streaming.

d) Using peer-to-peer file sharing for exchanging movies, music, video files

Peer-to-peer (P2P) refers to a network where the communication/information is distributed to a wide variety of computers rather than residing on one central server. In this network each computer has equal capabilities which differ from a traditional client/server network. P2P is more and more used for sharing of files, movies and music.

Peer-to-peer file sharing does as such not represent an overlap with item f) downloading games or h) uploading self-created content as these are often stand-alone activities. Certain P2P activities could be illegal which can influence the reliability of answers or response rates. However, usage rights can allow sharing and is expanding and there is interest to get more information on these activities.

e) Using a podcast service to automatically receive audio or video files of interest

Podcast as term combines "iPod" and "broadcast". It is a method of distributing content files, such as audio programmes incl. music files or videos over the Internet for playback on personal computers and mobile devices.

The term "podcast" may not be always known. However, real users have normally to subscribe under such a term on websites which offer this broadcast medium for receiving audiovisual files of interest.

f) Downloading computer or video games or their updates

This refers to downloading computer games or video games or their updates for playing. Includes downloading software either free of charge or under payment.

g) Playing networked games with others

This refers to playing games with other person(s) on any devices linked via the Internet. It may not require downloading before playing. An example is online video games played on games consoles.

h) Uploading self-created content (text, images, photos, videos, music etc.) to any website to be shared

This can involve uploading of own produced content to own website or to any other website with the purpose of sharing it with others.

i) Using browser based news feeds (e.g. RSS) for reading new content on websites

Browser based news feeds relates to website use and are not e-mails containing news alerts. To use them, a feed reader is often downloaded for example for the PC at home but it is also possible to visit the related websites, so-called aggregators, from anywhere.

RSS stands for "Really Simple Syndication". It is a family of web feed formats used to publish frequently updated content such as blog entries, news headlines or podcasts. It enables to keep up to date with certain preferred websites in an automated manner rather than checking them manually.

The term "RSS" may not be always known. However, real users have normally to subscribe under such a term on websites to receive news on new content available in the area of interest.

j) None of the above

D4: On average how often did you download music and/or films in the last 3 months?

[**Scope:** individuals who used the Internet in the last 3 months, i.e. 1st option (*Within the last 3 months*) in C1 and who have used downloading or sharing of music, movies and/or video files over the Internet (i.e. option b, c, d and/or e in D3)]

[**Type:** one single answer needed, i.e. *Tick only one*]

The question will investigate the regularity of downloading content. Respondents should be presented with all options and should select which one best describes their behaviour.

i) Every day or almost every day

Similar to the question on the regular use of computers (B2) or the regular use of the Internet (C2) this should refer to downloading music or films more than 4 days each week, from home or from any other location.

ii) At least once a week (but not every day)

This should refer to between 1 and 4 days each week

iii) At least once a month (but not every week)

iv) Less than once a month

v) Not applicable (only listened to music and/or watched films)

If respondents routed to this question have not downloaded but directly listened music or watched films, this item should ticked.

D5: Did you pay in the last 3 months for online audiovisual content?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1)]

[**Type:** one single answer needed, i.e. *Tick only one* ; binary (Yes/No) ; filter question for D6, i.e. if "no" go to D6]

As explained under question D3 above, the term "audiovisual" content, it refers to audio content, i.e. to listen to sound, music or spoken language, and visual content, i.e. to see, watch pictures, movies, or videos.

D6: What would make you pay for online audiovisual content?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1) and who did not pay in the last 3 months for online audiovisual content (i.e. answered "no" in D5)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

a) Lack of free available content

Audio or video files are allowed to some extent to be downloaded for free. Should this no longer become possible for the preferred files, respondents could then be willing to pay for them.

b) Right to share legally protected content

The right to use the content (music, video) is often restricted to one person, normally the purchaser downloading the file. However, the rights can also be extended that they allow to share it with at least one other person or to make certain copies. Sharing with others (family members, friends, or any other person(s)) could be restricted in number, depending on sales conditions. It could be an incentive to pay if the content file can be shared with others.

c) More convenient payment methods

Convenience relates to saving time or less burden compared to existing methods for payment. Online-payment is the most obvious example in this regard where ordering and payment transaction possibilities via the website are integrated. It is also linked to more secure payment methods. This could mean for example that the download is directly charged to the customers' monthly telephone bill and that it is not necessary to provide account/debit or credit card details on the website.

d) More advantageous prices compared to offline content

Those who normally go to shops to buy CDs or DVDs could be inclined to buy online if prices are cheaper.

e) Better quality of paid content than free services

Compared to free available content, an incentive to buy and pay for online music, videos, films could lie in the better quality.

f) Wider range of choices, content more easily available

Similar to better quality available via website shopping, there could be a wider product range than usually on offer in shops. Easier access to content could mean for example that the user finds well organised online archives of music or video files to choose from.

g) None of the above, but other (e.g. to support artists work etc.)

This is an alternative choice if none of the above a) to f) was ticked.

h) Nothing, no willingness to pay

Instead of to options a) to f) or g), this category can be ticked.

D7: Did the use of the Internet replace your time spent with off-line media or other off-line activities?

[**Scope:** individuals who used the Internet in the last 3 months (i.e. 1st option in C1)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

The question will look at some frequent demand of on-line media replacing traditional consumption pattern and time spent with newspapers, CDs, and DVDs bought in shops/delivered by post, and using normal radio apparatus. The question also looks at replacing personal contacts with governments in form of phone calls or visits to public administrations. Although subjective, more information on the impact is needed. The matrix reflects heavy, some or no replacement at all.

The interest is if the group of recent Internet users (users in the last 3 months) have changed their behaviour and have spent less time with off-line activities or not.

a) Reading of online news instead of printed news, newspapers, magazines

b) Downloading music files instead of buying a CD

c) Downloading films and videos instead of buying/renting a DVD

d) Listening to web radio instead of listening to normal radio

e) Using online contacts instead of personal contacts with public services and administrations

This item relates to interaction between citizens and public authorities. It applies when the respondent has used the Internet and an e-government website, for example via a sub-webpage "contact", to request documents, to pose questions, to clarify procedures etc. whereas in the past these type of contacts were done by phone calls or visits.

D8: Do you use a mobile phone?

[**Scope:** all individuals]

[**Type:** one single answer needed, i.e. *Tick only one* ; binary (Yes/No) ; filter question for D9]

This question serves as filter for D9 and refers to any type of mobile phone, not only Internet enabled via GPRS or UMTS networks (also includes e.g. GSM).

'Use' refers to the actual use of this device. Whether the respondents *owns* the mobile phone is irrelevant to this question. In most cases, it will be a personal mobile phone but in some households a 'collective' mobile phone may be shared by several members – this is especially the case for households that are 'new' to mobile telephony. The mobile phone can be owned by the respondent or can e.g. be provided by the respondent's employer.

D9: For which of the following activities did you use a mobile phone for private purposes in the last 3 months (other than involving voice calls or SMS)?

[**Scope:** individuals who use a mobile phone (i.e. "yes" in D8)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

Depending on the technical features of the mobile phones, various activities can be done. For some activities related to browsing the Internet or the use and sharing of content except in text type (photos, music, video, TV etc), more advanced, Internet enabled mobile phones (e.g. with GPRS, UMTS) are required.

a) Sending photographs or video clips

b) Uploading photographs or video clips from your phones to websites

This means that the upload is done directly from the mobile phone to the website. This is possible with Internet enabled mobile phones via GPRS, UMTS/3G, WiFi. The more recent camera phones often enable direct file transfer to websites without prior storage on computers.

c) Receiving subscription-paid information services (for example news, weather forecast, sports results etc.)

d) Browsing the Internet

e) Reading your e-mails

f) Downloading and/or watching TV or video

g) Paying for goods or services (instead of cash or credit card)

This refers to micro payment practices for services such as tickets, parking etc.. Internet enabled mobile phones are not really necessary; pre-paid phone cards can be activated for additional credits.

- h)** Personal navigation (for finding location or address), use of location-aware services (e.g. to receive nearby travel, shopping, event information)

Finding a location or using of a location service is possible through a GPS chip in the mobile phone or through the mobile phone itself. I.e. it is not necessary for the mobile phone to have an in-built GPS. It can also be done through online search of the location.

- i)** None of the above

D10: Do you use pre-payment or post-payment for your mobile phone?

[**Scope:** individuals who use a mobile phone (i.e. "yes" in D8)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

The payment scheme will influence the usage patterns for advanced services related to mobile phones, including the use of websites and related services via mobile phones.

- a)** Prepaid

The phone service is paid in advance in prepaid schemes and there is typically no contract with the provider. The usage in time/minutes are prepaid per month and are carried to following months if not entirely used. Control of cost is thus easier. However, the services which can be used are usually more basic.

- b)** Post-paid

This refers paying monthly service fees for mobile phone usage. There are better opportunities with post-paid scheme to make use of advanced services and features, e.g. browsing the Internet or uploading content to websites, depending on the type of the mobile phone. Post-paid schemes for mobile phones often have a flat rate for monthly phone charges rather than billing on a per unit basis.

**D10 b1: Do you pay a flat rate for Internet access via your mobile phone?
– optional –**

[**Scope:** individuals who use a mobile phone (i.e. "yes" in D8) and who have used post-payment for their mobile phone (i.e. option b in D10) – *optional* -]

[**Type:** single answer needed, i.e. *Tick only one ; binary (Yes/No)*]

A flat rate provides often the possibility for unlimited use of the Internet, but there could be also certain limits in terms of usage time or bytes, e.g. 500 MB.

The item should be addressed in connection with existence of a post-paid scheme (although there could be cases where flat rates can apply to pre-paid phones as well).

A flat rate for Internet usage via the mobile phone may co-exist with other flat rates (e.g. for sending SMS) or just with phone and other charges on per unit basis. It is likely to correlate to increased demand for advanced services compared to pay per use type of tariffs.

II.1.6.5 MODULE E: USE OF E-COMMERCE

→ Statistical unit: individuals

E1: When did you last buy or order goods or services for private use over the Internet (excluding manually typed e-mails)?

[**Scope:** individuals who already used Internet, i.e. 1st, 2nd or 3rd option in C1]

[**Type:** one single answer needed, i.e. *Tick only one* ; filter question]

Purchases of financial investments, such as shares, should be included in this question. The ordering of goods and services should also include confirmed reservation for accommodation, participation in lotteries and betting and obtaining information services from the Internet that are directly paid for.

This question also applies to purchases that are made via Internet auctions, such as "eBay".

Only individuals that actually placed the order over the Internet should answer this question, even if the order was carried out on somebody else's behalf. Therefore, individuals that had other people ordering for them should not be included. For the time of the transaction, the date when the goods or services were ordered is relevant, not the date of delivery or payment.

Orders via manually written e-mails should - according to the EUROSTAT definition of e-commerce in enterprises - not be included.

Delivery via electronic means or payment via electronic means is not a requirement for an order via Internet.

Goods and services that were obtained via the Internet for free should be excluded. Such goods are e.g. free software ("freeware"), reservations in restaurants or any kind of information obtained via Internet for free (such as downloading pdf-files). Software that can be downloaded from the Internet and used for free for a certain period but loses functionality after this time should only be counted as a purchase when the product is finally paid for.

Orders that are placed for professional use should be excluded.

E2: What types of goods and services did you order over the Internet for private use in the last 12 months?

[**Scope:** individuals who already used Internet (1, 2 or 3 in C1) **and** who bought/ordered over the Internet in the last 12 months, i.e. 1st (Within the last 3 months) or 2nd option (Between 3 months and a year ago) in E1]

[**Type:** multiple answers allowed, i.e. *Tick all that apply* ; filter question for E3]

a) Foods / Groceries

Groceries include necessities of "daily use" such as sanitary products, tobacco products, cosmetics, pharmaceuticals or flowers. This category also includes drinks. Newspapers should be excluded from this category.

b) Household goods

Includes e.g. furniture, toys, washing machines, micro-wave ovens, dish washers, vehicles, gardening equipment, plants, antiques, art, tools, collectibles, etc.

c) Films, music

Includes DVDs, CDs, Video and music cassettes and similar devices that can store films, music or spoken word. Delivery of films and music on-line via files is also included. Only goods that are paid for should be mentioned here.

d) Books / Magazines / Newspapers / E-Learning materials

Subscription to newsletters and alike – if not free of charge – also belong to this category. E-Learning materials are documents which are used for multi-media learning in digital form, but also comprise CD-Roms that are used for educational purposes.

e) Clothes, sports goods

Includes shoes, textiles, accessories.

f) Computer software and upgrades (incl. computer and video games)

Includes video games and other computer software on any media.

g) Computer hardware

Includes any accessories to computer hardware, such as modem and printer.

h) Electronic equipment (incl. cameras)

Includes e.g. mobile phones, cameras, radios, TVs, stereos, DVD players, video recorders.

i) Share purchases / Financial Services / Insurance

Includes the purchase of securities of all kinds. Transfers from account to account and similar services via Internet where no purchase is carried out are not included.

j) Travel and holiday accommodation

Includes e.g. hotel reservation, travel tickets and vehicle hire. Only bookings of accommodation and travel tickets should be included. The mere search for those services and booking inquiries without legal binding should be excluded.

k) Tickets for events

Includes tickets for concerts, cinema, sport events, theatre, and alike.

l) Lotteries or betting

m) Other

Goods and services not mentioned above. Could include e.g. jewellery, fan articles, and a number of services ordered via Internet, e.g. information services (from databases).

E3: Were any of the products that you ordered over the Internet downloaded or accessed from websites rather than delivered by post etc.?

[**Scope:** individuals who already used Internet (1, 2 or 3 in C1) **and** who bought/ordered over the Internet in the last 12 months (1 or 2 in E1) **and** who bought films/music, books/magazines/newspaper/e-learning material or computer software over the Internet (i.e. option c, d or f in E2)]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

Digitally delivered products include those products which can be delivered on physical media (such as CD or tape) but also over the Internet in digitised form as a computer file (and thus independently of those physical media).

a) Films, music

Delivery of films and music on-line via files (e.g. a video, an MP3 file or a CD).

b) (Electronic) books, magazines, newspapers, e-learning materials

c) Computer software (incl. computer and video games and software upgrades)

E4: From whom did you buy or order goods or services for private purposes over the Internet in the last 12 months?

[**Scope:** individuals who already used Internet (1, 2 or 3 in C1) **and** who bought/ordered over the Internet in the last 12 months, i.e. 1st (Within the last 3 months) or 2nd option (Between 3 months and a year ago) in D1]

[**Type:** multiple answers allowed, i.e. *Tick all that apply*]

Buying goods and services from other Member States contributes to the EU policy objectives of achieving a single European market and a single information space. The question is asking about the country of origin/residence of the seller.

a) National sellers

b) Sellers from other EU countries

c) Sellers from the rest of the world

d) Country of origin of sellers is not known

Multi-national companies should be treated as national sellers when it is known from the website that they are registered as a company with an address in the surveyed country. The term "national sellers" includes the trade business or sales offices established in the country by foreign owners (development, production, other distribution may be located in the home country and/or globally). This approach is more feasible and relevant at the European level than asking about websites in national or foreign language.

If the seller and his sales office address is not known, item d) should be ticked.

II.1.6.6 MODULE F: SOCIO-DEMOGRAPHIC BACKGROUND VARIABLES

[**Scope:** all households / all individuals, excepting for F5 Occupation]

[**Type:** F2, F3, F4, F8, F9: one single answer needed, i.e. *Tick only one* ; F1, F10, F11, F12: quantitative ; F5, F6, F7: open question, i.e. description needed (or coded by the interviewer)]

This module does not focus on ICT-related study variables, but on background variables to put the results on the study variables in perspective. These background variables should enable to relate the outcome of the surveys to the sociologic, demographic and economical background of the observed statistical units.

In the model questionnaire, this module is inserted at the end of the survey. This was an arbitrary choice, the module can also be the starting point for the survey (in most cases, the information will partly be available from other sources, e.g. registers or the survey the ICT usage questions are embedded in).

The questions on age (F1), gender (F2), educational attainment (F3), employment situation (F4) and occupation (F5) are to be answered by the individual(s) who is (are) selected within the household.

The questions on regions of residence (F6, F7), geographical location (F8) and type of locality (F9) refer to the household, but the answer at household level should coincide with the individual level anyway as the individuals are part of the household. Usually these two variables do not need to be directed to the respondents as the NSI's can derive the information on the basis of the household's address.

The questions on the household composition (F10, F11) and household income (F12) of course refer to the household level.

F1: Age

The data transmitted to Eurostat is broken down by age group. Whether the year of birth or date of birth or the age is collected, is relatively unimportant. In a postal survey, the NSI could simply ask the respondent to tick the appropriate age group. However, in most cases, this information will be available from the sampling frame (e.g. the population register).

Whether the reference point is the age during the interview or rather the age during the reference period, is not expected to have a significant impact on the results and comparability.

F2: Gender

Self-explanatory.

F3: Educational level

In order to obtain comparable information for the different countries, the levels of education are linked to the UNESCO's International Standard Classification of Education (ISCED 1997).

A detailed description of the classification is available from the UNESCO Institute for Statistics' website (http://www.uis.unesco.org/ev.php?ID=3813_201&ID2=DO_TOPIC), but a short summary is presented below.

Conversion tables matching the national educational levels to the ISCED codes are available from the CIRCA site (information relating to the school year 1999-2000): http://forum.europa.eu.int/Public/irc/dsis/edtcs/library?l=/public/unesco_collection/programmes_isced97/school_1999_2000&vm=det

For the purpose of this survey, information on the educational level has to be transmitted following a regrouping in three broad categories:

- No formal education completed, primary or lower secondary education (corresponding to ISCED 0, 1 or 2);
- Upper secondary education (corresponding to ISCED 3 or 4);
- Tertiary education (corresponding to ISCED 5 or 6).

Only the *highest* level reached is required (not *all* the levels frequented). Only a *completed* level may be considered; 'completed' in this context is taken to imply successful termination, normally accompanied by an appropriate qualification.

Pre-primary education (ISCED level 0)

Programmes at level 0, (pre-primary) defined as the initial stage of organised instruction are designed primarily to introduce very young children to a school-type environment, i.e. to provide a bridge between the home and a school-based atmosphere. Upon completion of these programmes, children continue their education at level 1 (primary education).

Primary education or first stage of basic education (ISCED level 1)

Programmes at level 1 are normally designed on a unit or project basis to give students a sound basic education in reading, writing and mathematics along with an elementary understanding of other subjects such as history, geography, natural science, social science, art and music. The core at this level consists of education provided for children, the customary or legal age of entrance being not younger than five years or older than seven years. This level covers in principle six years of full-time schooling.

Lower secondary or second stage of basic education (ISCED level 2)

In many, if not most countries, the educational aim is to lay the foundation for lifelong learning and human development on which countries may expand, systematically, further educational opportunities. The programmes at this level are usually on a more subject-oriented pattern using more specialized teachers and more often several teachers conducting classes in their field of specialization. The full implementation of basic skills occurs at this level.

Upper secondary education (ISCED level 3)

The educational programmes included at this level typically require the completion of some 9 years of full-time education (since the beginning of level 1) for admission or a combination of education and vocational or technical experience and with as minimum entrance requirements the

completion of level 2 or demonstrable ability to handle programmes at this level. The entrance age to this level is typically 15 or 16 years. More specialization may be observed at this level than at ISCED level 2 and often teachers need to be more qualified or specialized than for ISCED level 2. The programmes can be general, technical or labour-market oriented.

Post-secondary non-tertiary education (ISCED level 4)

This item captures programmes that straddle the boundary between upper-secondary and post-secondary education from an international point of view, even though they might clearly be considered as upper-secondary or post-secondary programmes in a national context. ISCED level 4 programmes can, considering their content, not be regarded as tertiary programmes. They are often not significantly more advanced than programmes at ISCED level 3 but they serve to broaden the knowledge of participants who have already completed a programme at level 3. Typical examples are programmes designed to prepare students for studies at level 5 who, although having completed ISCED level 3, did not follow a curriculum which would allow entry to level 5.

First stage of tertiary education (ISCED level 5)

This level consists of tertiary programmes having an educational content more advanced than those offered at levels 3 and 4. Entry to these programmes normally requires the successful completion of ISCED level 3 (esp. general or technical programmes) or a similar qualification at ISCED level 4 (esp. those preparing for entry to ISCED 5). This level includes all the research programmes which are not part of a doctorate, such as any type of Master's degree.

The programmes can be of a general nature - i.e. largely theoretically based and intended to provide sufficient qualifications for gaining entry into advanced research programmes and professions with high skills requirements - or more technically oriented - i.e. practically oriented/occupationally specific and mainly designed for participants to acquire the practical skills and know-how needed for employment in a particular occupation or trade or class of occupations or trades (the successful completion of which usually provides the participants with a labour-market relevant qualification).

Second stage of tertiary education (ISCED level 6)

This level is reserved for tertiary programmes which lead to the award of an advanced research qualification, such as a doctorate. The programmes are therefore devoted to advanced study and original research and are not based on course-work only. They typically require the submission of a thesis or dissertation of publishable quality which is the product of original research and represents a significant contribution to knowledge. It prepares graduates for faculty posts in institutions offering ISCED 5 (general) programmes, as well as research posts in government, industry, etc.

F4: Employment situation

Unlike the Labour Force Survey, this survey doesn't contain a number of detailed questions to assess an individual's employment situation or labour status. This variable is more related to the LFS question on the *Main labour status* where the respondent gives a self-perception regarding his/her activity status.

The explanatory notes below should help the respondent and/or interviewer in indicating the appropriate employment situation. The concepts and definitions used are based on those used in Eurostat's labour market statistics and recommended by the International Labour Organisation (ILO).

Employee

Employees are defined as persons who work, by agreement, for a public or private employer and who receive compensation in the form of wages, salaries, fees, gratuities, payment by results or payment in kind; irrespective of the number or hours worked (full time or part time) and the duration of the contract (fixed or indefinite).

This category includes non-conscripted members of the armed forces, apprentices or trainees receiving remuneration, seasonal workers and persons on strike.

This category does not include voluntary workers and family workers who don't receive a regular monetary wage.

Persons who are simultaneously working in their own professional practice and for a public or private employer (e.g. doctors with their own cabinet and working in a hospital) should be classified according to the status where they work a more important number of hours.

Self-employed

Self-employed persons, with or without employees, are defined as persons who work in their own or jointly-owned business, professional practice or farm for the purpose of earning a profit.

This category includes own-account workers, family workers who don't receive a regular monetary wage, freelancers (although a person who has been regularly retained by a single employer for some time may also be regarded as an employee) and persons who look after one or more children that are not their own on a private basis and receiving a payment for this service.

Persons who are simultaneously working in their own professional practice and for a public or private employer (e.g. doctors with their own cabinet and working in a hospital) should be classified according to the status where they work a more important number of hours.

Unemployed

In accordance with the ILO standards used in the Labour Force Survey, unemployed persons comprise persons aged 15 to 74 who are (the three conditions must be fulfilled):

- (a) without work during the reference week (the week of the interview), i.e. neither have a job nor are at work (for one hour or more) in paid employment or self-employment;
- (b) currently available for work, i.e. are available for paid employment or self-employment before the end of the two weeks following the reference week;
- (c) actively seeking work, i.e. have taken specific steps in the four weeks period ending with the reference week to seek paid employment or self-employment or who found a job to start later, i.e. within a period of at most three months.

The following are considered as specific steps in "actively seeking work":

- having been in contact with a public employment office to find work, whoever took the initiative (renewing registration for administrative reasons only is not an active step),
- having been in contact with a private agency (temporary work agency, firm specialising in recruitment, etc.) to find work,
- applying to employers directly,
- asking among friends, relatives, unions, etc., to find work,
- placing or answering job advertisements,
- studying job advertisements,
- taking a recruitment test or examination or being interviewed,
- looking for land, premises or equipment,
- applying for permits, licences or financial resources.

Student

This category refers to individuals in formal education (in educational institutions such as school, university, etc.) and not belonging to any of the other categories of employment situation.

When strictly applying the LFS definitions, students with small jobs have to be classified as employees. However, due to the absence of the long list of questions used to determine the employment status in the LFS, students who work only a few hours per week may be classified as students.

Other not in the labour force

This category refers to persons who are not employed or unemployed. Students are considered as a separate category (see above).

"Other not in the labour force includes persons in retirement or early retirement, persons in compulsory military service, persons fulfilling domestic tasks (without being an employee or self-employed, e.g. housewives) or other inactive persons.

As some countries embed the ICT usage survey in the Labour Force Survey or other socio-economic surveys using LFS definitions, it is useful to briefly discuss the correspondence between the five above categories and the 9 categories used in the LFS variable *Main labour status* (see column 101 in the LFS handbook or in the Annex to Regulation 1575/2000).

ICT usage survey <i>Employment situation</i>	Labour Force Survey <i>Main Labour Status</i>
Employee	→ LFS col.101, code 1 <i>(carries out a job or profession, including unpaid work for a family business or holding, including an apprenticeship or paid traineeship, etc.)</i>
Self-employed	→ LFS col.101, code 1 <i>(carries out a job or profession, including unpaid work for a family business or holding, including an apprenticeship or paid traineeship, etc.)</i>
Unemployed	→ LFS col.101, code 2 <i>(unemployed)</i>
Student (not in the labour force)	→ LFS col.101, code 3 <i>(pupil, student, further training, unpaid work experience)</i>
Other not in the labour force	→ LFS col.101, codes 4, 5, 6, 7 and 8 <i>(in retirement or early retirement or has given up business (=4); permanently disabled (=5); in compulsory military service (=6); fulfilling domestic tasks (=7); other inactive person (=8).</i>

The above correspondence table shows a 1-1 relation between the breakdown in the ICT usage survey and in the Labour Force Survey (with codes 4, 5, 6, 7 and 8 combined into one single category as the subcategories are less relevant for the ICT usage survey or could be problematic in terms of low cell frequencies).

Only the additional breakdown in *employee* and *self-employed* cannot be distinguished directly from column 101 in the LFS as they are grouped in code 1. A pragmatic approach - with an acceptable risk of coding error - would be to additionally use the *Professional status* from the LFS (column 26) and file all persons (in code 1 of column 101) who appear to be *self-employed* or *family worker* (codes 1, 2 or 4 in LFS column 26) under code *self-employed* for the ICT usage survey and all others under *employees*. In other words, the auxiliary information from LFS column 26 is only used to redirect some persons into the group of *self-employed* while the rest remain in the group of *employees*. This recoding suggestion supposes a certain correlation between the professional status according to the ILO definitions (= col.26) and ones self-assessment on the labour status (= col.101).

F5: Occupation

[**Scope: individuals in employment, i.e. 1st (employee) or 2nd option (self-employed) in F4**]

The individual's occupation should be coded according to the ISCO-88 (COM) classification which is used in most of Eurostat's social statistics and which is based upon the ILO's International Standard Classification of Occupations, ISCO-88.

To be able to make such regroupings, it is necessary to code and store the occupations at a detailed level, at least at 3 digit level but preferably at 4 digit level. The major groups (1 digit) or submajor groups (2 digit) don't allow to distinguish ICT professionals from other workers.

Manual workers

This category corresponds to major groups 6 to 9 of ISCO:

- Major group 6: Skilled agricultural and fishery workers;
- Major group 7: Craft and related trades workers;
- Major group 8: Plant and machine operators and assemblers;
- Major group 9: Elementary occupations.

Non-manual workers

This category corresponds to major groups 0 to 5 of ISCO.

- Major group 1: Legislators, senior officials and managers;
- Major group 2: Professionals;
- Major group 3: Technicians and associate professionals;
- Major group 4: Clerks;
- Major group 5: Service workers and shop and market sales workers;
- Major group 0: Armed forces.

ICT professionals

The most appropriate recoding will be obtained where ISCO codes are collected at 4-digit level. However, as some countries collect (or enter) the information relating to the respondent's occupation only at 3-digit level, the discussion below also includes an alternative approach covering this case.

ISCO available at 4-digit level

The category *ICT professionals* consists of the individuals in one of the following eight ISCO Unit Groups (Unit Groups correspond to the 4-digit level):

1236	Computing services managers
2131	Computer systems designers, analysts and programmers
2139	Computing professionals not elsewhere classified
2144	Electronics and telecommunications engineers
3114	Electronics and telecommunications engineering technicians
3121	Computer assistants
3122	Computer equipment operators
3132	Broadcasting and telecommunications equipment operators

ISCO available at 3-digit level

The category *ICT professionals* consists of the individuals in one of the following two ISCO Minor Groups (Minor Groups correspond to the 3-digit level):

213	Computing professionals
312	Computer associate professionals

Non ICT professionals

The category *Non ICT professionals* consists of the individuals in all other ISCO Unit Groups (in the case of 4-digit coding) or all other ISCO Minor Groups (in the case of 3-digit coding).

The classification is available in the RAMON-database:

<http://europa.eu.int/comm/eurostat/ramon>

(An introductory note can be found at:

http://europa.eu.int/comm/eurostat/ramon/documents/isco_88_com/isco_88_com.zip)

The basis for the classification is the job and the skills. A job is defined as the set of tasks and duties to be performed. Skills are the abilities to carry out the tasks and duties of a job. Skills consist of two dimensions: skill level – which is usually related to the level of educational attainment - and domain specialisation.

The ISCO classification has been modified and the new version was adopted in 2008. However, it is sufficient for the survey year to use ISCO-88; general implementation in Eurostat is envisaged for the year 2010. For information on the new ISCO-08, please see

<http://www.ilo.org/public/english/bureau/stat/isco/index.htm>

If the new classification ISCO-08 on occupations adopted in 2008 is considered - an option mentioned in the 2007 implementing regulation - it is now possible to code and store the occupations at the more aggregated level of submajor groups (2 digit). At least the following sub-major groups should be considered for ICT occupations:

25 Information and communications technology professionals

35 Information and communications technicians

If possible, some ICT occupations classified in other groups should also be taken into account:

133 ICT service managers

7422 ICT installers and servicers

For the breakdown manual / non-manual workers, there are very minor changes in the groups:

Manual workers

This category corresponds to major groups 6 to 9 of ISCO-08:

- Major group 6: Skilled agricultural, forestry and fishery workers;
- Major group 7: Craft and related trades workers;
- Major group 8: Plant and machine operators and assemblers;
- Major group 9: Elementary occupations.

Non-manual workers

This category corresponds to major groups 0 to 5 of ISCO-08.

- Major group 1: Managers;
- Major group 2: Professionals;
- Major group 3: Technicians and associate professionals;
- Major group 4: Clerical support workers;
- Major group 5: Service and sales workers;
- Major group 0: Armed forces.

The purpose of the variable "occupation" as a background characteristic is not to collect data on ICT usage broken down by individual occupations (this would necessitate very large samples), but rather by groups of occupations. E.g. manual versus non-manual or ICT-jobs versus non ICT-jobs.

F6: Region of Residence – NUTS 1

The NUTS 1 regions have been introduced as being mandatory in 2008. Most, but not all countries are concerned. More information about NUTS 1 regions can be found at:

http://ec.europa.eu/eurostat/ramon/nuts/codelist_en.cfm

F7: Region of Residence – NUTS 2 (optional)

The NUTS 2 regions have been introduced as being optional in 2008. Most, but not all countries are concerned. More information about NUTS 2 regions can be found at:

http://ec.europa.eu/eurostat/ramon/nuts/codelist_en.cfm

F8: Geographical location

Former "Objective 1" promotes the development and structural adjustment of regions whose development is lagging behind, i.e. whose average per capita GDP is below 75% of the European Union average.

From 2007 onwards, the list of objective 1 regions (valid until end 2006) relates to the new list of **convergence regions (including related phasing out regions)**, see also: OJ L 243/44 (6.9.2006) "Commission Decision of 4 August 2006 drawing up the list of regions eligible for funding from the Structural Funds under the Convergence objective for the period 2007-2013. The list is to be extended: In EU-27, Bulgarian and Romanian NUTS regions are convergence regions. Non-convergence including related phasing out regions (i.e. all other regions) are classified as **regional competitiveness and employment region**.

For further information, see

http://ec.europa.eu/regional_policy/sources/docoffic/official/decideci_en.htm

The list of NUTS regions eligible for support from the Structural Funds under the Convergence Objective (ex-Objective 1), including phasing out Convergence regions, has been decided for the whole period 2007-2013. No update of this list is foreseen in the Structural Funds Regulation.

The list is available on the next two pages.

Conv = Convergence region

PO = Convergence objective related phasing out region (regions eligible for funding from the Structural Funds on a transitional and specific basis under the Convergence objective)

BE32	Prov. Hainaut	PO
BG00	Balgarija	Conv
CZ02	Střední Čechy	Conv
CZ03	Jihozápad	Conv
CZ04	Severozápad	Conv
CZ05	Severovýchod	Conv
CZ06	Jihovýchod	Conv
CZ07	Střední Morava	Conv
CZ08	Moravskoslezsko	Conv
DE41	Brandenburg - Nordost	Conv
DE42	Brandenburg - Südwest	PO
DE80	Mecklenburg-Vorpommern	Conv
DE93	Lüneburg	PO
DED1	Chemnitz	Conv
DED2	Dresden	Conv
DED3	Leipzig	PO
DEE1	Dessau	Conv
DEE2	Halle	PO
DEE3	Magdeburg	Conv
DEG0	Thüringen	Conv
EE00	Eesti	Conv
GR11	Anatoliki Makedonia, Thraki	Conv
GR12	Kentriki Makedonia	PO
GR13	Dytiki Makedonia	PO
GR14	Thessalia	Conv
GR21	Ipeiros	Conv
GR22	Ionia Nisia	Conv
GR23	Dytiki Ellada	Conv
GR25	Peloponnisos	Conv
GR30	Attiki	PO
GR41	Voreio Aigaio	Conv
GR43	Kriti	Conv
ES11	Galicia	Conv
ES12	Principado de Asturias	PO
ES42	Castilla-La Mancha	Conv
ES43	Extremadura	Conv
ES61	Andalucía	Conv
ES62	Región de Murcia	PO
ES63	Ciudad Autónoma de Ceuta	PO
ES64	Ciudad Autónoma de Melilla	PO
FR91	Guadeloupe	Conv
FR92	Martinique	Conv
FR93	Guyane	Conv
FR94	Réunion	Conv

ITF3	Campania	Conv
ITF4	Puglia	Conv
ITF5	Basilicata	PO
ITF6	Calabria	Conv
ITG1	Sicilia	Conv
LV00	Latvija	Conv
LT00	Lietuva	Conv
HU21	Közép-Dunántúl	Conv
HU22	Nyugat-Dunántúl	Conv
HU23	Dél-Dunántúl	Conv
HU31	Észak-Magyarország	Conv
HU32	Észak-Alföld	Conv
HU33	Dél-Alföld	Conv
MT00	Malta	Conv
AT11	Burgenland	PO
PL11	Łódzkie	Conv
PL12	Mazowieckie	Conv
PL21	Małopolskie	Conv
PL22	Śląskie	Conv
PL31	Lubelskie	Conv
PL32	Podkarpackie	Conv
PL33	Świętokrzyskie	Conv
PL34	Podlaskie	Conv
PL41	Wielkopolskie	Conv
PL42	Zachodniopomorskie	Conv
PL43	Lubuskie	Conv
PL51	Dolnośląskie	Conv
PL52	Opolskie	Conv
PL61	Kujawsko-Pomorskie	Conv
PL62	Warmińsko-Mazurskie	Conv
PL63	Pomorskie	Conv
PT11	Norte	Conv
PT15	Algarve	PO
PT16	Centro (PT)	Conv
PT18	Alentejo	Conv
PT20	Região Autónoma dos Açores	Conv
RO00	Romania	Conv
SI00	Slovenija	Conv
SK02	Západné Slovensko	Conv
SK03	Stredné Slovensko	Conv
SK04	Východné Slovensko	Conv
UKK3	Cornwall and Isles of Scilly	Conv
UKL1	West Wales and The Valleys	Conv
UKM4	Highlands and Islands	PO

F9: Type of locality

The concept of "urbanisation" has been introduced to indicate the character of the area where the respondent lives. The definition of the type of locality is based on the classification of *degrees of urbanisation* which is also used in the Labour Force Survey (LFS, column 176).

The following degrees of urbanisation have been identified:

Densely-populated area

This is a contiguous set of local areas, each of which has a density superior to 500 inhabitants per square kilometre, where the total population for the set is at least 50,000 inhabitants.

Intermediate area

This is a contiguous set of local areas, not belonging to a densely-populated area, each of which has a density superior to 100 inhabitants per square kilometre, and either with a total population for the set of at least 50,000 inhabitants or adjacent to a densely-populated area.

Thinly-populated area

This is a contiguous set of local areas belonging neither to a densely-populated nor to an intermediate area.

(a set of local areas totalling less than 100 km², not reaching the required density, but entirely enclosed within a densely-populated or intermediate area, is to be considered to form part of that area. If it is enclosed within a densely-populated area and an intermediate area it is considered to form part of the intermediate area)

In this definition of *degree of urbanisation* there is a criterion of geographical contiguity together with a population threshold. Harmonised, comparable correspondence between the *degree of urbanisation* and NUTS 5 regions are being (re-)defined on the basis of 2001 census data (for Member States and Candidate countries) in 2005. These correspondence tables enable the NSIs to easily recode the addresses or postal codes into the three categories. In principle, this conversion table should cover all 25 EU countries at NUTS 5 (or LAU 2) level. However, for some countries there may be no results at all or the results will only be at NUTS 4 (or LAU 1) level.

An "area" consists of a group of contiguous "local areas" where a "local area" corresponds to the following entities in the respective Member States :

Belgium: Gemeente / Commune

Czech Republic: Obce (6 251 in year 2000)

Denmark: Kommuner

Germany: Gemeinde

Estonia: Vald+Alev+Linn (254)

Greece: Demotiko diamerisma / Koinotiko diamerisma (after the kapodistria reform, ca. 6000 units)

Spain: Municipio

France: Commune

Ireland: DED / ward
Italy: Commune
Cyprus: Demos/Koinotites
Latvia: Pagast+ Pilsetas (560)
Lithuania: Seniunija
Luxembourg: Commune
Hungary: Telepules (3 135)
Malta: Localities
The Netherlands: Gemeente
Austria: Gemeinde
Poland: Gminy+Miasta (2 486)
Portugal: Freguesias
Slovenia: Obcinah (192 since 1 Jan 1999)
Slovakia: Obce a Mesta (2 920 in year 1999)
Finland: Kunnat
Sweden: Kommune
United Kingdom: Ward

Iceland: Sveitarfélag (165 until 1997, 124 from 1998)
Norway: Kommuner (435)

Bulgaria: Naseleni miasti
Croatia: not yet available
Romania: Comuni, Municipii, Orasi
Turkey: not yet available

F10 and F11: Household type

“Household” refers either to one person living alone or a group of people living together in the same dwelling unit with at least one person of the age of 16 – 74 years (see above: “General outline of the survey”, Scope - age limit).

Excluded are permanent members of institutions such as hospitals, prisons, etc.

The household composition (as used in the tabulation scheme) is determined using two indicators: the total number of persons in the household (including all age groups) and the number of children under 16. The presence (and the number) of children is collected separately as this is a driving force for adoption of ICT’s or Internet in a household.

F12: Household income (mandatory question in 2008)

The “income” concept used in this survey is the net monthly income of the household (not the individual!). The focus is on the household income as one’s access to and use of ICT’s will rather be related to the total income of the household he/she is part of than by his/her personal income.

For persons in paid employment, this refers to the monthly 'take-home' pay, i.e. the pay after deducting income taxes, employee's social security contributions and any voluntary contributions.

Bonuses regularly paid at the time of each payment – such as regular overtime hours, bonuses for team, night or weekend work, tips and commissions - are included.

Bonuses that are not paid at each pay period (f.i. annual profit shares) but that are by accident paid in the considered pay period are excluded as they will lead to a non-representative estimate.

Irregular or annual bonuses should be excluded. No "redistribution" of such bonuses – i.e. dividing by 12 and adding to the monthly estimate – is required as this would unnecessarily increase the recall burden on the respondent. This may however make international comparisons less relevant as the impact of such bonuses tends to differ heavily from one country to another.

Where applicable and feasible, an estimate of the monetary value for the *payments in kind* should be added.

Although the *income from self-employment* may be difficult to measure, the respondents can be asked for an estimate of their disposable income.

Income from allowances such as unemployment benefits, pensions and sickness, disability or invalidity allowances should refer to the last monthly allowance received before the interview (or the monthly average for a recent period if this is more easy to collect or estimate).

Income from investments (assets, savings, stocks, shares, etc) is to be excluded as this information is difficult to measure and/or collect.

Although undesirable, derogations from the above definition of the household income may be justified in cases where the data can be obtained from slightly different income-related questions in the survey vehicle the survey on ICT usage is linked to. Especially when using such already available data is the only way of providing information on this sensitive, optional variable.

Ideally, information should be collected by asking the respondent for an estimate of the actual amount. Where it is not desirable or feasible to collect such information, the respondent can indicate a size band. These size bands should be compatible with the breakdown used in the tabulation scheme. This tabulation scheme foresees the use of income quartiles rather than income classes. Such approach should make cross-country comparisons more meaningful. Not only can this overcome the dispersion of income levels across countries in Europe, the use of quartiles can also avoid threats to comparability caused by the use of different concepts by different countries (e.g. gross versus net earnings, whether or not including regular monthly bonuses, whether or not including 1/12th of the annual bonuses, whether or not including an estimate of the value of payments-in-kind, etc.).

For additional information, see *Annex Transmission format*.

II.2. PRODUCTION METHODOLOGY

II.2.1. Timetable – Survey period

According to Annex II of Commission Regulation (EC) No 847/2007 of 18 July 2007 implementing Regulation (EC) No 808/2004 of the European Parliament and of the Council concerning Community statistics on the information society, “the reference period for the statistics to be collected is the first quarter of 2008” (see Part III of this manual, Annex III.2). As the majority of the questions in the model questionnaire and the variables in the above mentioned legal act refer to the last three months – which means in practice: *the three months preceding the interview* – the data collection should take place immediately after the first quarter. Taking into account that the fieldwork can take several weeks or even months, the recommendation is to interview the respondents in April-May (or more general: in the second quarter). This way, a satisfying overlap with the theoretical reference period is guaranteed.

The table and figure below present an overview of the survey periods in the 2007 survey.

As mentioned in the previous paragraph, for most questions the reference period is 3 months. The choice for this 3 months’ reference period is a trade-off between recall bias and seasonal bias. Using a reference period of 12 months filters out seasonality, but a recall period of 12 months may be too demanding for the respondents. In general, priority has been given to limiting the intellectual burden for the respondents, only for the questions related to e-commerce (where an important seasonal effect can be expected in consumption patterns throughout the year, e.g. more purchases Christmas, more booking of holidays during Summer) a longer reference period is used.

Looking at the results of previous surveys, the pattern for the last 3 months is not very difficult from the last 12 months’ pattern. In terms of e.g. Internet use, this means that only few people will state they have used the Internet in the past 12 months, but not in the past 3 months. The harmonisation of the data collection periods across countries should further reduce the possible risk of seasonal bias when comparing across countries or from one year to another.

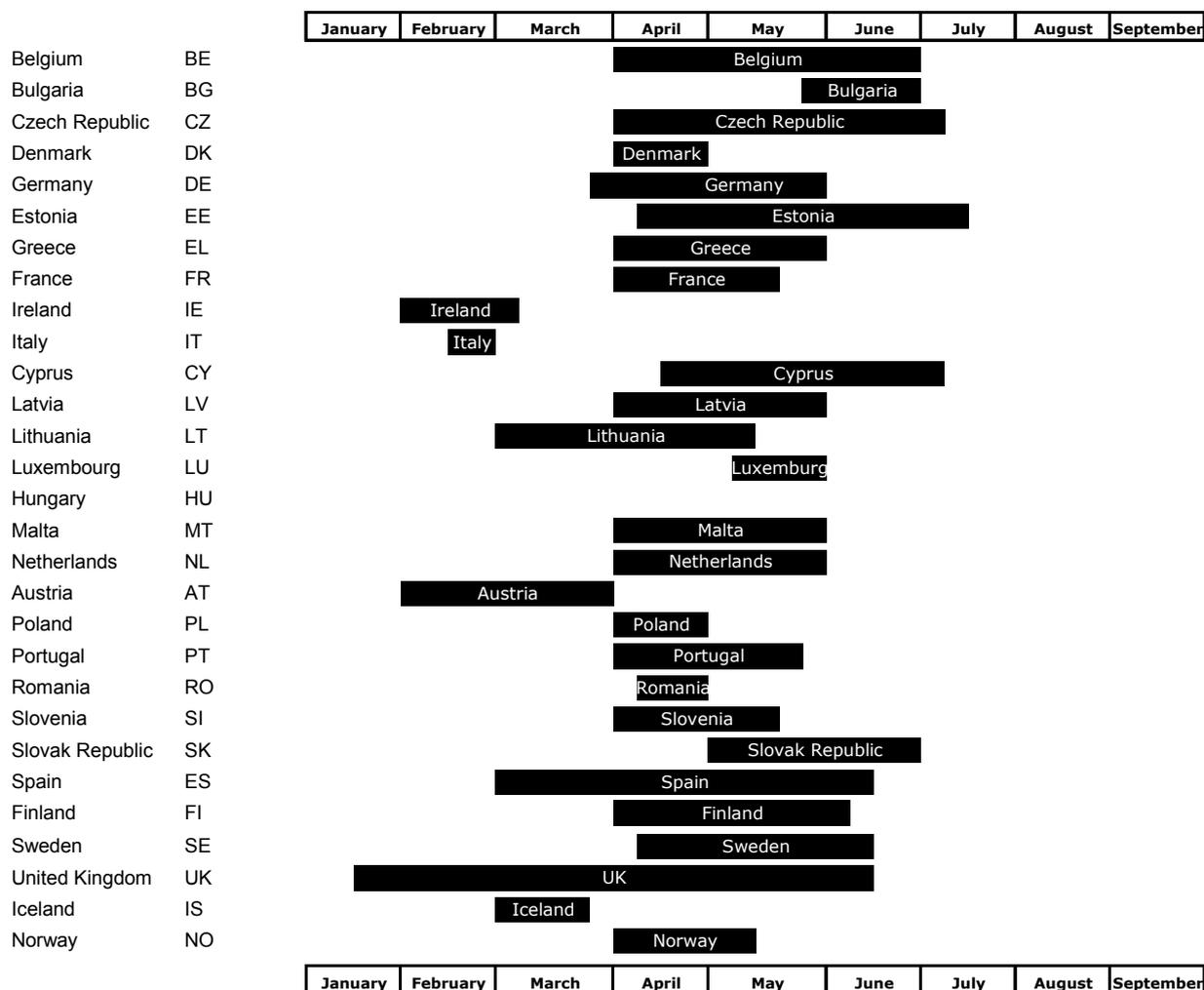
Table: Overview of the survey periods, survey vehicles and data collection method (2007 survey)

Source: Quality reports for the Community survey on ICT usage in households and by individuals (2007)

Belgium	2 nd quarter 2007
Bulgaria	17/05/2007 – 30/06/2007
Czech Republic	First week April – first week July
Denmark	02/04/2007 – 30/04/2007
Germany	End of March-end of May
Estonia	9/04/2007 – 15/07/2007
Greece	1/04/2007 – 31/05/2007
Spain	5/03/2007 – 15/06/2007
France	2/04/2007 – 19/05/2007
Ireland	5/02/2007 – 2/03/2007
Italy	15/02/2007 – 28/02/2007
Cyprus	20/04/2007 – 12/07/2007
Latvia	1/04/2007 – 31/05/2007
Lithuania	1/03/2007 – 10/05/2007
Luxembourg	7/05/2007 – 31/05/2007
Hungary	1/04/2007 – 21/04/2008
Malta	April – May 2007
Netherlands	2/04/2007 – 31/05/2007
Austria	February – March 2007
Poland	2/04/2007 – 30/04/2007
Portugal	2/04/2007 – 22/05/2007
Romania	11/04/2007 – 30/04/2007
Slovenia	1/04/2007 – 15/05/2007
Slovak Republic	May – June 2007
Finland	2/04/2007 – 4/06/2007
Sweden	9/04/2007 – 17/06/2007
United Kingdom	January – June 2007
Iceland	1/03/2007 – 21/03/2007
Norway	2/04/2007 – 9/05/2007

Figure: Overview of the survey periods (2007 survey)

Source: Quality reports for the Community survey on ICT usage in households and by individuals (2007)



II.2.2. Frame population

This issue was already discussed in chapter II.1.2 on the *target population*. The *frame population* (of *sampling population*) is the frame from which the sample will be drawn. Ideally, this list of units should be equivalent to the target population as both overcoverage and undercoverage can induce bias and affect the reliability of the survey results.

E.g. overcoverage due to inclusion of persons aged 75-79 in the sampling frame (making the simplistic assumption that the frame only has information on one's age using age bands, not the actual age or year of birth): as this age group to score relatively low on ICT usage indicators, the population total may show a downwards bias due to inclusion of this group.

E.g. undercoverage due to the fact that some remote areas are not covered in the register: as such areas usually have a lower penetration of ICTs, the overall indicators may be relatively overestimated for such country.

To allow for an assessment, it is important to include the coverage and the sampling frame when reporting metadata (see Chapters 3 and 5 in the reporting template, Annex II.3.3).

The table below presents the sampling frames used in the national statistical institutes for the 2007 survey.

The second table gives an overview of the target and non-target population. In most countries, the target population of individuals aged 16 to 74 represents approximately 75% of the total population. The target population for households covers the population relatively more (approximately 90%) as only households where *all* members are over 74 are excluded. This information has to be kept in mind when disseminating the results: the results are not representative for the total population but only for the subpopulation of persons in the age group 16 to 74. This difference can be important as the out-of-scope groups may have a different profile in terms of ICT use (heavy use among persons younger than 16 but low use among persons over 74).

Table: Overview of the sampling frames (2007 survey)

Source: Quoted from the quality reports for the Community survey on ICT usage in households and by individuals (2007)

Belgium	Population register (contains all residents of Belgium, regardless of nationality, together with information on household)
Bulgaria	The general population and administrative-territorial division by statistical districts of the settlement from which the sample for the survey is formed comprise all the households in the country. Population census 2001 register was used as sampling frame.
Czech Republic	Register of census areas
Denmark	Central Population Register (contains general personal data on individuals with residence in Denmark; individuals without registered address in the CPR or people having their name and phone number protected from use in research purposes do not enter the survey)
Germany	Quota sample – German microcensus 2005 German microcensus is a 1%-sample of all households in Germany; continuous household sample survey; mandatory
Estonia	The Population Register is used as sampling frame. Population Register is an electronic database containing the general personal data about all Estonian citizens and foreigners residing in Estonia. Population register is administered by the Estonian Ministry of the Interior.
Spain	The sampling frame is the Population Register managed by INE and updated in a continuous way.
France	The sampling frame is a phone number database, which contains phone numbers of private people who are in the white pages.
Greece	The sampling frame (areas, households, individuals) that contains the surveyed units, is based on data from the EU-SILC survey of the period 2003-2006.
Ireland	Census of Population
Italy	Public register of households (name of the components of the household, addresses of the household)
Cyprus	Population Census 2001 register. In addition, the newly constructed housing units since 2001 were also included in the frame. A list of domestic consumers of electricity that incorporates new consumers was obtained from the Electricity Authority of Cyprus (EAC). It has already been tested that there is one to one correspondence between the definition of the housing units of EAC (domestic consumer of electricity) and the statistical definition of households.
Latvia	Two sampling frames are built for each sampling stage due to two-stage sampling is used. At the first stage the list of Population census 2000 counting areas is used as sampling frame. All territory of Latvia was divided in small territories (smaller than NUTS4) during the Population census 2000. The list contains information about the number of households in each counting area. At the second stage sampling frame is built from The Population Register, statistical register of dwellings and statistical register of households.

Lithuania	<p>The Residents' Register. The Residents' Register Service under the Ministry of the Interior is the manager of Residents' Register. Data of the Residents' Register in on-line mode are submitted to Statistics Lithuania. The Residents' Register database includes data on the residents of the Republic of Lithuania: the citizens of Lithuania, the citizens of foreign countries or persons without citizenship, declaring the place of residence in Lithuania or registering any changes of the civil state in a registry office.</p> <p>Due to the absence of a sampling frame for selecting households, a sample of individuals of 16 and older from the Residents' Register is drawn. Households whose members are the selected persons are surveyed.</p>
Luxembourg	<p>It will be worked with RDD (random digital dialling) by using computer generated national phone numbers. There are a number of phone numbers randomly chosen to be called. In order to reach representative results, a quota system for the individuals to be interviewed is applied.</p>
Hungary	<p>The updated stock or dataset of the addresses of the 2001 population and housing census was used.</p>
Malta	<p>Database maintained by the NSO containing all households and individuals in Malta and Gozo.</p>
Netherlands	<p>Municipal Population Register. This register contains all the persons who are registered in Dutch municipalities. Advantage of using the Municipal Population Register is the actuality of information from the registered individuals. Every day information about birth, death, move (house), new addresses et cetera is coming in. Monthly the information is incorporated in the mentioned register. After the register is brought up to date in month t, individuals are selected for interviewing in month t+2</p>
Austria	<p>The "Zentrales Melderegister (ZMR)" is used as a sampling frame for the Continuous Labour Force Survey. A sub-sample of this sample was drawn for the ICT survey.</p>
Poland	<p>The basis for selection of Primary Sampling Units (PSU) and dwellings is the Domestic Territorial Division Register. The Register includes (among others) information about:</p> <ul style="list-style-type: none"> - enumeration statistical district (ESD) - dwelling addresses <p>The first stage sampling frame was a record of 29 thousand Area Survey Points (ASPs). They are enumeration statistical districts or sets of them which consist of at least 250 dwellings for urban area and 150 dwellings for rural area. The second stage sampling frame was based on the register of inhabited dwellings in the area survey points from the first stage. The sample frame do not cover collective households like students' hostels, social welfare homes, prisons, convents etc.</p>
Portugal	<p>This survey uses a sample selected from Master-sample (Census 2001). It constitutes the sampling basis to all households and individuals related surveys done by INE.</p>
Romania	<p>The household surveys carried out by NSI-Romania are based on the use of a master sample EMZOT. EMZOT is a Multifunctional Sample of Territorial Areas, made by the data registered from Population and Dwelling Census in 2002. It is a data base including approximate 1.500.000 dwellings, selected according to probabilistic criteria, serving as sample base for all household surveys, in 2004-2013. The sample of 780 centres of research is distributed in the whole counties and in Bucharest: 427 in urban area and 353 in rural area.</p>
Slovenia	<p>Units of observation are persons aged between 10 and 74 years at the time of interviewing and their households. The basis for the sampling frame is the Central Population Register (CRP). Target persons (persons aged between 10 and 74 years) are randomly selected from the register. For the allocation of persons, in addition to the CRP also the frame of the enumeration units is used, which serves as a sampling frame for the selection of units at the first stage in two-stage sampling designs.</p>

Slovak Republic	<p>Population and housing census is organised by So SR in 10year periodicity. The last one was realised in May 2001. On base of census results the information on whole population is accessible - household type, age, sex, education level, employment situation, location. Meanwhile SO SR organizes surveys on population increase and decrease by age and by sex.</p> <p>Census register from 2001 is used as sampling frame for survey on information and communication technologies (ICT) in households and individuals.</p> <p>Sampling frame for individuals is created on the state of population as of 31 December 2006. Negative side of the sample frame is the time difference between Household Register (May 2001) and state of population as of 31 December 2006. Missing social and demographic data (level of education, employment) are included into the state of population as of 31 December 2006 by statistical methods used in SO SR.</p>
Finland	<p>Population database maintained by Statistics Finland. It is based and updated from the official Central Population Register of the Population Register Centre. It contains the demographic information of resident population, address and residence information etc.</p>
Sweden	<p>The sampling frame is the Total Population Register (TPR) kept by Statistics Sweden. The TPR obtains notifications of births, deaths, migrations etc. daily from the Tax Authorities.</p>
United Kingdom	<p>The Omnibus Survey uses the Postcode Address File of 'small users' as its sampling frame. Private household addresses in Great Britain are included in this frame. For Northern Ireland, the sampling frame is the Valuation and Lands Agency list of addresses.</p>
Iceland	<p>The National Register of Persons is used as a sampling frame for the ICT survey. All Icelandic as well as foreign citizens domiciled in Iceland are registered in the National Register of Persons. Basic information such as gender, age and sample municipality are always found in this register. To be able to monitor every changes Statistics Iceland's survey center has unlimited access to the National Register of persons during the data collection period.</p>
Norway	<p>The Statistical Population Register (SPR) is the central demographic population database at Statistics Norway. The database is updated from the administrative Central Population Register several times a month.</p>
Serbia	<p>The sampling frame is based on the data from the Population Census conducted in 2002, which includes household living on the national territory.</p> <p>Target persons (persons aged between 16 and 74 years) are selected from the household which was chosen.</p>
Croatia	<p>Units of observation are Croatian citizens aged between 16 and 74 years at the time of interviewing. The basis for the sampling frame is the Census 2001 Population Database of individual data. In the sample selection process full name and address of citizens were used as well as information about gender, date of birth and education level. Respondents (persons aged between 16 and 74 years) were randomly selected from this database.</p> <p>For the allocation of respondents, in addition to the individual Census data, the frame of the Census Enumeration Units was also used. These units serve as a sampling frame for the selection of units at the first stage in two-stage sampling designs.</p>

Table: Overview of the target and non-target population (2007 survey)

Source: Quality reports for the Community survey on ICT usage in households and by individuals (2007)

	Individuals			Households		
	Target population	Non-target population	Target population as % of total population	Target population	Non-target population	Target population as % of total population
Belgium	7,750,000	2,700,000	74%	4,000,000	400,000	91%
Bulgaria	6,074,520	1,773,882	77%	2,723,128	198,759	93%
Czech Republic	8,040,000	2,170,000	78%	3,850,000	390,000	91%
Denmark	3,984,000	1,463,000	73%	2,281,000	350,000	87%
Germany	62,000,000	19,000,000	77%	35,000,000	3,000,000	92%
Estonia	1,028,900	313,500	77%	540,000	:	:
Greece	8,201,531	2,039,057	80%	3,689,737	407,213	90%
Spain	33,796,402	6,032,099	81%	14,692,323	1,421,728	91%
France	43,700,000	16,300,000	73%	22,600,000	3,000,000	88%
Ireland	3,170,851	1,148,453	73%	1,324,750	145,738	90%
Italy	44,270,114	14,079,874	76%	20,439,971	2,571,489	89%
Cyprus	565,099	213,585	73%	241,518	21,781	92%
Latvia	1,771,242	510,063	78%	:	:	:
Lithuania	2,575,586	827,698	76%	1,310,000	115,000	92%
Luxembourg	338,393	121,107	74%	156,451	15,484	91%
Hungary	7,680,000	2,242,000	71%	3,464,000	370,000	90%
Malta	303,990	:	:	131,558	8,025	94%
Netherlands	12,800,000	3,200,000	80%	6,500,000	600,000	92%
Austria	6,160,000	2,000,000	75%	3,180,000	310,000	91%
Poland	28,503,984	9,726,096	75%	12,592,289	744,751	94%
Portugal	8,026,077	2,570,159	76%	3,504,104	344,481	91%
Romania	16,000,000	5,500,000	74%	7,373,373	:	:
Slovenia	:	:	:	652,939	41,135	94%
Slovak Republic	4,149,644	1,239,536	77%	1,897,007	79,042	96%
Finland	3,901,000	1,300,000	75%	2,190,000	295,000	88%
Sweden	6,500,000	2,500,000	72%	:	:	:
United Kingdom	43,427,055	15,839,326	73%	22,265,149	3,214,765	87%
Iceland	216,800	93,000	70%	113,000	9,000	93%
Norway	3,355,382	1,187,976	74%	1,846,517	218,324	89%

II.2.3. Sampling design

The population of interest of this survey, i.e. the group of statistical units around which it was intended to investigate is made up of households (with the exclusion of permanent members of cohabitation) and individuals living in Member States.

As the country details in the table below show, there are great differences among national statistical institutes as regards sampling design. Some countries use a sample based on individuals as primary sampling units. For other countries the primary sampling units are represented by households registered in the public register and in general the second sampling units are all the actual components belonging to each household included in the sample.

The survey should be based on a probability sample from which results representative of the population could be derived, considering the agreed breakdowns defined in the questionnaire.

The final sample units should be the individuals but each participating country should design its sample selection according to what is most efficient to that country.

The sampling design and the resulting sample size (see below) should be appropriate for obtaining accurate, reliable and representative results on the survey characteristics and breakdowns specified in the Regulation and the model questionnaire. Taking into account the unbalanced distribution of e.g. Internet users over the different groups of society, this condition can be difficult to hold for all indicators and all breakdowns, but this requirement should at least be followed for the main indicators (e.g. the eEurope Benchmarking Indicators, listed in the Council Resolution of 18 February 2003 on the implementation of the eEurope 2005 Action Plan, Official Journal of 28/02/2003, C48, p. 2-9; since 2006: the i2010 benchmarking indicators, see http://ec.europa.eu/information_society/eeurope/i2010/docs/benchmarking/060220_i2010_Benchmarking_Framework_final_nov_2006.doc).

This objective should be achieved for the overall proportions as well as for the proportions relating to the different subgroups of the population. In line with section II.2.7.3 below on accuracy measures and the template for reporting quality in annex II.3.3, the estimated standard error (standard deviation), expressed by the square root of the estimate of the sampling variance, shall not exceed 2% of the overall proportions and shall not exceed 5% for the proportions relating to the different subgroups of the population, where these subgroups constitute at least 10% of the total population in the scope of the survey.

The aim of such requirement is to ensure the collection of a complete dataset – without empty, confidential or unreliable cells – for these indicators. In case this aim cannot be reached, clarifications should be provided (see also II.2.7.3).

Table: Overview of the sampling design and sampling methods (2007 survey)

Source: Quoted from the quality reports for the Community survey on ICT usage in households and by individuals (2007)

Austria

For the Continuous Labour Force Survey (LFS) every quarter of the year approx. 22,500 households are surveyed. Every household remains in the sample for five quarters (this means that 20% of the surveyed households are replaced every quarter).

For the ICT survey a gross sub-sample of approx. 7,000 households has been drawn for which the "reference week" of the survey are the calendar weeks 6 to 13 in the year 2007. All household members between 16 and 74 in the sampled households are surveyed.

Belgium

The LFS sampling designs defines 12 strata (corresponding to NUTS2 units, with one exception: the Liège province split into two). We use a 2 stage sampling design. Within each stratum, all (private) hh have the same selection probability, while the selection probability of an individual is (zero if <16 or >74, and otherwise) inversely proportional to the number of persons 16-74 in the hh s/he belongs to.

Initial LFS sample size 14924 households

Individual within household: Interviewer takes person aged 16-74 with most recent birthday at the date of interview.

Cyprus

The sampling frame is stratified into urban and rural strata by district.

The size of the sample is predetermined on the basis of the average size of households having at least one individual in age group 16-74. The size of the sample should satisfactorily serve the desirable and acceptable confidence limits of the survey results, taking always into consideration time and cost constraints.

The households are then allocated in each stratum according to the number of households in the population census register.

The selection of the sample in the urban areas is done by using simple systematic random sampling. A random start is selected, and by using the sampling interval N/n , the households for each district (urban areas) were selected.

The selection in rural areas is done two stages. The villages of each district are the Primary Sampling Units and the households the Ultimate Sampling Units.

All the villages of each district are listed in ascending order of size. The sample is drawn with Probability proportional to size (P.P.S.), the latter being determined from the number of individuals. Once a PPS sampling is employed the number of households in each selected village is fixed at 20. In cases of very small villages with less than 20 households, these villages are attached to other neighbouring ones and they are considered as one village.

For the selection of the households in each selected village, simple systematic random sampling is employed in each village separately.

Czech Republic

The sample is stratified and two-staged. On the first stage the census areas are sampled in every district by the method of randomised systematic sampling. The probabilities of inclusion are proportional to its size (PPS). On the second stage of the sampling 6 dwellings are sampled in every selected census area by the method of simple random sampling. All individuals aged 16+ in households living in selected dwellings are interviewed.

Germany

The ICT survey is based on a representative quota sample. The population of households (microcensus 2005) was divided into groups by the following characteristics:

Federal State (16 Bundesländer);

Household Type (4 types): 1) one person household; 2) two persons or more (>16years) without children under 16; 3) one person (>16 years) with at least one child under 16; 4) two or more persons (>16 years) with at least one child under 16;

Social status of head of the household (5 types): 1) self employed; 2) employee; 3) retired – (only for household type 1 and 2); 4) student (only for household type 1); 5) farmer (separate group regardless of household type and household net monthly income).

Household net monthly income (4 classes): 1) < 1300 EUR; 2) 1300 to <2600 EUR; 3) 2600 to < 3600 EUR; 4) > 3600 EUR;

Sampling unit is the private household at main residence. Households are selected either from an access panel or from other address databases (i.e. from household budget survey).

In the first step, households are contacted by mail and asked to send back their confirmation of participation. On that confirmation they are also asked to report the number of persons in the household and particularly the number of persons 10 years and over. Then the household questionnaire and the correct number of personal questionnaires are sent to the household. All persons 10 years and over living in the household are selected respondents. There is no second step of sampling within the households.

Denmark

The sample is formed by a random selection from the CPR. When the sample has been created, the individuals taken part are then contacted by mail, in which they are asked to participate in the survey. If it is impossible to find a telephone number, the respondent is asked to contact Statistics Denmark with a number by which he or she can be reached.

After the data has been collected, the data is grouped by gender, age, civil status and gross income.

Sexes:

- Men
- Women

Groups of ages:

- 16-19 years
- 20-39 years
- 40-59 years
- 60 years and more

Occupation:

- Students

- Blue-collar workers

White-collar workers

- Self-employed
- Outside occupation

Gross income:

- DKK 0-99,999
- DKK 100,000-399,999
- DKK 400,000 and more
- Unknown and do not know.

Estonia

The sample design is stratified systematic sampling of individuals whose households are included into sample. Different inclusion probabilities of households are taken into account during calculation design weights.

Stratification is made by place of residence. The 15 counties of Estonia and Tallinn are divided into four strata according to the population size (I – Tallinn, II – four bigger counties, III – ten smaller counties, IV – Hiiu county) and different inclusion probabilities are used in stratas, the highest being for Hiiu county.

All household members aged 15-74 are interviewed for labour force survey and for ICT.

Greece

Sampling method

The multistage stratified area sampling is applied for the survey. The primary units are the areas (one or more unified blocks) and secondary sampling units selected in each sampling area are the households containing members belonging to the target population. The final unit is one person randomly selected among the household members of age sixteen to seventy four years.

STRATIFICATION

The sampling design involves two levels of area stratification: (i) The first level is geographical stratification based on the partition of the total country area into thirteen standard administrative regions corresponding to the European NUTS II level. The two major city agglomerations of Greater Athens and Greater Thessalonica constitute separate major geographical strata. (ii) The second level of stratification involves grouping municipalities and communes within each NUTS II administrative region by degree of urbanization, i.e., according to their population size, into four categories. These categories are defined by the population size intervals 0-999, 1000-4999, 5000-29999, 30000 and over. The number of final strata in the thirteen regions, i.e., non-empty strata formed by crossing region and degree of urbanization, was 50. The two major city agglomerations were further partitioned into 31 and 9 substrata (administrative subsections), respectively, on the basis of the city blocks of the municipalities that constitute them. Thus, the total number of strata for this survey was 90.

STAGES OF PROBABILITY SAMPLING

The sample of households for the ICT survey of the year 2007 was selected of the sample used in the Greek Survey of Income and Living Conditions (EU-SILC of the years 2003, 2004, 2005 and 2006). The EU-SILC is an annual rotating household survey covering the target population of the ICT survey.

The selection probabilities of the households for the ICT survey of the year 2007 were defined suitably so that the demands of the survey to be met. The definition of selection probabilities was as follows:

1st stage: The primary unit of order i in stratum h has probability of being drawn proportional to the target

population size as follows:
$$P_{hi} = \frac{N_{hi}}{N_h}$$

N_{hi} : The updated (from EU-SILC survey) target population of households in the hi primary unit

N_h : The updated (from LFS 2007) target population of households in the h stratum

2nd stage: Out of N_{hi} households, a sample of n_{hi} households was selected with equal probabilities. Each of

n_{hi} households has the same chance to be selected, equal to:
$$\frac{n_{hi}}{N_{hi}}$$
. As the estimator of the stratum total

Y_h (for any characteristic) should be self-weighting, the n_{hi} was defined, as follows: $n_{hi} = \frac{n_h}{a_h}$, where $n_h = \sum_i n_{hi}$ and a_h is the number of primary units in the h stratum.

The initial probabilities of selection of sampling households of EU-SILC were based on the population sizes (coming from Greek General Population Census of the year 2001), which differ considerably from the new population sizes that better suit the demands of the current ICT survey. Additionally, the target populations of EU-SILC and ICT do not coincide. The measures of EU-SILC were based on all persons, but the current sample for ICT is restricted to households with individuals aged 16 to 74 years old. Thus, although the sample of households for the ICT survey of the year 2007 was selected of the sample used in the Greek Survey of Income and Living Conditions (EU-SILC of the years 2003, 2004, 2005 and 2006), the following measures were taken for improving the representativeness:

The 1st stage probabilities of selection of primary units were modified taking into consideration the updated target population size in each stratum using estimated data from Labour Force Survey with reference period the 2nd quarter of the year 2007.

The 2nd stage probabilities of selection of households were modified taking into consideration the updated register of households in the primary sampling units.

The allocation of sampling households in each separate stratum was carried out proportionally to the target population size, which was estimated from data coming from Labour Force Survey with reference period the 2nd quarter of the year 2007.

After the application of the above measures, the sampling households for the ICT have no the same probability of selections (1st and 2nd) with the sampling households for EU-SILC, after changing the selection probabilities of the EU-SILC households so that the probabilities of ICT households, to be determined on updated target population.

Spain

TYPE OF SAMPLING

An independent sample is designed in each autonomous community (NUTS 2) to represent it. The sampling type used is a three-staged stratified design. The first-stage design is the census sections; the second-stage units are the main family households and the third-stage unit is one person aged 16 or over per household selected in second-stage. The stratification criterion used is the sized of the municipalities to which the section belongs.

SIZE OF SAMPLING .AFFIXATION.

The sample was distributed among autonomous communities using a compromise between uniform and proportional affixation to the size of the community. Between strata, affixation is proportional to their size, also upholding that the number of sections per stratum in each autonomous community be a multiple of four.

SELECTION OF THE SAMPLE

The selection of first-stage units in each stratum was made using probability proportional to the size of each section. In a second stage, the households were selected by means of systematic sampling with random start and equal selection probabilities for each household in the section; so this procedure leads us to obtain self-weighted samples of households in each stratum.

In a third-stage, and within each household, a person aged 16 or over was chosen with equal probability using the Kish random method, as well children aged from 10 to 15.

Hungary

The ICT sample in 2005 was a sub-sample of the sample of the micro-census. The micro-census sample was designed to provide reliable estimates of the main demographic indicators for the 176 General Electoral Districts (GEDs) of the country. The GEDs were roughly of the same size, the average being 24,000 in terms of dwellings. Each GED has a 2 % sample of its own, resulting in a self-weighting 2 % overall sample of the country. Within each GED localities were stratified by size in terms of the number of dwellings. Some GEDs are towns or segments of major cities (type 1), other GEDs consist of a number of small localities (type 2) GEDs of type 1 have 2 % systematic sample of dwellings, those of type 2 have two-stage stratified samples of dwellings; the PSUs (primary sampling units) are localities, selected with pps.

When selecting the ICT sample from the micro-census sample, an effort was made to keep it close to a self-weighting sample with 10000 dwellings in 374 strata. In the selected dwellings each household and one individual per household were observed.

In 2006 2267 new dwellings were selected completing the respondent dwellings (7734) in 2005. According to a rotational design, in 2007 roughly 1100 households were dropped out and 3400 new dwellings were selected in the sample.

Additional measures taken at the time of sampling design to improve representativeness

If any, and if not covered under §6.1. E.g. corrections for sampling frame undercoverage, etc.

When selecting the sample of localities within the GEDs in 2005, we were able to compare population counts estimated from the sample with the corresponding actual values of 1 January 2004. In the case of extremely large sampling errors, the selection procedure was repeated.

Ireland

For the QNHS (national title given to the continuous quarterly Labour Force Survey), a two-stage sample design is used. This comprises a first stage sample of 2,600 blocks (or small areas) selected at county level to proportionately represent eight strata reflecting population density. Each block was selected to contain, on average, 75 dwellings and the sample of blocks is fixed for a period of about five years. In the second stage of sampling, 15 households are surveyed in each block to give a total quarterly sample of 39,000 households.

Households are asked to take part in the survey for five consecutive quarters and are then replaced by other households in the same block. Thus, one fifth of the households in the survey are replaced each quarter and the QNHS sample involves an overlap of 80% between consecutive quarters and 20% between the same quarter in consecutive years. The ICT household survey was conducted on a sub-sample (20%) of the main QNHS sample.

Italy

The sampling pattern is a complex type and makes use of two different sampling plans. Municipalities are divided into two sub-groups on the basis of the resident population:

The group of Self-Representative municipalities (which since now we'll referred as AR municipalities) made up of larger demographic size of municipalities;

The group of Non Self Representative municipalities (or NAR) made up of the remaining municipalities.

Within the group of AR municipalities, each municipality is considered as an independent layer and a pattern known as cluster sampling is used. The primary sampling units are represented by households registered in the Public Registry, taken systematically from the Public Register of the municipality itself; for each registered household included in the sample, the characteristics that are the subject of the survey for all the actual components that belong to the same household are extracted.

Within the group of NAR municipalities a two stage pattern with layering of the primary units is adopted. The Primary Units (UP) are the municipalities, the Secondary Units are the households registered in the Public Registry; for each registered household included in the sample, the characteristics that are the subject of the survey for all the actual components that belong to the same household are extracted

The municipalities are selected with probabilities that are proportional to their demographic size and without reintroduction, and the households are extracted with probabilities that are equal and without reintroduction.

Stratification

During the survey in question, the municipalities were stratified according to their demographic size and observing the following conditions:

self-deliberation of the sample on a regional level;

selection of a sample municipality within each layer defined on the municipalities of the NAR group;

choice of a minimum number of households to be interviewed in each sample municipality; for this survey, this number was set at 24;

Formation of layers with a more or less constant size in terms of resident population.

Latvia

Stratified two-stage sampling (systematic sampling with inclusion probabilities proportional to unit size at the first stage, simple random sampling at the second stage) was used.

The stratification was made depending on degree of urbanisation of area. Riga (the capital city), six other largest cities, other cities and rural areas forms four strata. The code of administrative territories was used to stratify. The population census counting areas were used as primary sampling units (PSU) at the first stage. PSUs were selected by systematic sampling with inclusion probabilities proportional to population size (number of households) of PSUs.

Households were used as secondary sampling units (SSU). Simple random sampling was used to select SSUs in each sampled PSU.

All individuals aged 15-74 will be selected in each sampled household.

The total sample size (number of households) was approximately defined depending on the resources available and desirable result of the survey. The initial sample size was proportional to population size in each stratum. The initial sample size was adjusted according to response rates in each stratum to get the final sample size in each stratum.

Lithuania

Stratified two-stage sample design. The biggest towns of Lithuania's counties, small and medium towns and rural area were divided into separate strata. A simple random sample of individuals of 16 and older was drawn from the Residents' Register in the biggest towns of the counties. A household, which lives at the selected person's address, was surveyed. A two-stage sample design was used in small and medium towns. The Pareto sample with probability proportional to the clusters size was used at the first stage. Each town was a cluster. A simple random sample of persons aged 16 and older was drawn from the Residents' Register in the already selected cluster at the second stage. In rural areas, the similar sample design as in medium and small towns was used. In this case, cluster was a rural territorial unit.

Luxembourg

The national sample is constructed in 2 stages: 1st level: selection of households by RDD. 2nd level: selection of respondent in the household; 1 individual per household is selected.

Quota sampling: age, sex, region, nationality and professional activity. In other words: there is a stratification on age (16-18 years, 19-24 years, 25-34 years, 35-49 years, 50-64 years and 65-74 years), sex (male, female), region (city, centre, south, north (including west), east), nationality (Luxembourgish, Portuguese, Italian, other) and professional activity (active, not active).

A weighting procedure for these quota variables is applied (including sex by age) after the fieldwork. The reference figures are taken from STATEC.

Malta

A sample of 1500 individuals was selected using systematic random sampling from a database maintained by the NSO on individuals of the whole Maltese population.

The Netherlands

Sampling method: For the ICT survey a stratified two-stage sample is used. In the first stage municipalities are selected. Large municipalities being selected with a probability of 1. All the other municipalities are selected at random, taking into account the size of the municipalities. In the second stage the number of individuals, as determined for every municipality, are selected. Individuals aged 12 to 74 year are selected.

Except the usual telephone register of KPN an additional register (SNT) with fixed and mobile telephone numbers is used. In this way for about 75 percent of the sample phone numbers are available.

Poland

A two-stage sampling scheme was adopted with stratification on the first stage. ASPs have been stratified within each of 16 voivodships in Poland by type of locality. Big cities usually constitute separate strata, whereas the strata in rural areas are composed of ASPs comprising neighbouring poviats in voivodship. There are 96 strata (of which 31 covering rural areas).

The number of the ASPs selected to the sample from a given stratum was proportional to the number of dwellings in that stratum. The probability of ASP selection was proportional to the estimated number of dwellings in the ASP.

Beginning from 2006 a scheme of rotation has been applied in sampling design in order to improve comparability of data over time. Six dwellings from each of 675 ASPs selected for 2006 survey round constituted a half of sample for 2007 survey, which gave 4050 dwellings.

In order to chose the second half of the sample 675 "new" ASP were selected by means of procedure of systematic sampling after random ordering of the units (ASPs selected with PPS using Hartley-Rao method).

On the second stage 6 dwellings were selected from each "new" ASP by means of simple random sampling method.

Gross sample size for ICT usage survey 2007 counts: $(6 \times 675 \text{ "old" ASP}) + (6 \times 675 \text{ "new" ASP}) = 8100$ dwellings.

Portugal

The survey on ICT Usage in Households and by Individuals is a subsample of the Master Sample (MS). The MS was designed and selected using the information of the Census of Population and Housing 2001 and it is constituted only by private dwellings (collective households and institutions were excluded). This sampling frame follows a stratified one-stage cluster sample. In each stratum the clusters were selected systematically with proportional probability to size (number of private dwellings of usual residence). The stratification was done at NUTS III level and the clusters

are geographical areas constituted by one or more contiguous census enumeration areas. The MS has 1408 clusters and almost 750.000 private dwellings (535.000 of which are as usual residence, the remaining are vacant and for secondary use).

The ICT is representative at NUTS I level: Portugal (Mainland), Azores and Madeira. For this propose, 539 geographical areas were selected from the MS. Ten dwellings were allocated to each area in the Mainland, 16 dwellings to each area in Azores and 17 dwellings to each area selected in Madeira. The total sample size is 6026 dwellings.

A rotation system comprising four waves is used. Dwellings are kept in the sample for four consecutive years before being replaced by an identical number of dwellings in the same cluster. One-fourth of the sample is replaced each year.

Romania

TIC sampling design is founded on a two-stage sampling technique.

In the first stage, a stratified random sample of 780 areas, Primary Sampling Units (PSU's), was designed after the 2002 census, using as stratification criteria the residence area and county and selected based on a systematic selection algorithm. The primary sampling unit, corresponding to the selection of the master sample, is a group of census section. The including probabilities from the first stage were proportional calculated with the size, expressed in number of permanent dwellings.

In the second stage, dwellings are systematically selected from the initial sample of PSU's: 12 from urban PSU's and 16 from rural PSU's. The secondary (ultimate) sampling unit, corresponding to the selection of the survey sample, has been the dwellings. All households within each sampling unit are included. The final sample consists of 10 472 dwelling units.

Stratification concerns only the first stage sampling. There are 88 strata and the criteria used being the area where a certain PSU is located (urban or rural area) and county (NUTS 3 level).

Slovenia

Two-stage, stratified sample design was employed in the survey. The sample is implicitly stratified according to the size and type of settlements (6 classes) and according to the statistical region (12 regions).

At the first stage 313 sampling units are selected with the probability proportional to size (PPS with replacement), and at the second stage 8 persons aged between 10 and 74 years are selected. Thus 2504 persons were selected. The person who is supposed to be interviewed is defined by the selected person.

The probability of selection of a person is calculated by the formula:

$$\frac{N_a}{\sum_{Slo} N_a} \cdot \frac{8}{N_a},$$

where

N_a is the number of persons aged between 10 and 74 years in the sampling unit.

The selected person in the household is interviewed.

Slovak Republic

The sample was realized on the base of the same methods, which are used in other surveys organized by SO SR (census, inquiry methods face-to-face)

Sample contains 4500 households including households with members at the age of 12 to 74 and more. The breakdown of households was following:

- With 1 member
- With 2 members
- With 3 members
- With 4 members
- With 5 and more members.

Stratification sampling was realised:

Number of persons in households

Location

The households are selected inside of each strata applying random sampling.

Finland

Systematic random sampling is used to extract the sample from population database. The sorting system of the sampling frame is based on geographical population density. The target area is the whole country, and the respondents represent 16-74 years old population according to age, gender, province, and native language.

Sample size is based on the knowledge of response rates in respective surveys earlier.

After drawing the sample from the population database the telephone numbers are searched for the respondents of whom approximately 93 - 95 per cent are within reach of the telephone. Fixed telephone, if available, will be used as the primary way of communication. However, roughly 80 per cent of the respondents are reached by mobile phone. A couple of weeks before the contact by phone, Statistics Finland sends out an advanced letter informing the respondent briefly about the contents and the meaning of the survey. The sample population without a telephone number will receive a letter by post, in which they are asked kindly to contact Statistics Finland.

Sweden

The LFS-sample is drawn at the end of the year to cover the coming year's need of sample persons. When the sample is drawn it is stratified according to county, sex and age. In this way 144 strata are constructed and from each stratum a simple random sample is drawn. The sample size is directly proportional to the size of the stratum.

The LFS sample consists of three separate samples, one for each month in the quarter. Each of the samples, which consists of about 20 000 people, is rotated in such a way that an eighth leaves the sample between two survey cycles. For every sample this occurs every three months. Persons in sample are interviewed once a quarter with a total of eight interviews during a two-year period, after which they leave the sample.

The sample for the Survey on ICT use is made by selecting individuals in waves other than the first and the last. The ICT survey is conducted at the end of the interview, i.e. after the LFS questions. A sub sample consisting of retired persons 65 years of age or older gets only the ICT questions.

To improve the coverage of persons aged 65 or over an extra sample has been drawn among persons 65-74 who are no longer employed or otherwise active on the labour market.

United Kingdom

Northern Ireland

In May 2007, a sample of 2200 addresses was drawn from the Valuation and Lands Agency list of addresses and stratified by region. Further selection stages were used to convert the listing of addresses to a listing of individuals from which one person (the selected respondent) was chosen to complete the questionnaire. If addresses contained more than one household, the interviewer used a standard procedure to randomly select just one household. In the Omnibus survey, of the total co-operating households, one in three was asked the Internet Access module.

Great Britain – January, February and March 2007

A new sample of 67 postal sectors were selected and stratified by: region, the proportion of households renting from local authorities; and the proportion in which the household reference person is in Socio-Economic Group 1-5 or 13 (i.e. a professional, employer or manager). The postal sectors were selected with probability proportionate to size and, within each sector, 30 addresses (delivery points) were selected randomly. If addresses contained more than one household, the interviewer used a standard ONS procedure to randomly select just one household. Within households with more than one adult member, just one person aged 16 or over was selected with the use of random number tables. The interviewers endeavoured to interview that person – proxy interviews were not taken.

At the time of sampling, the age of the respondent is not known, nor the ages of the other members of the household.

Iceland

Stratified random sampling without replacement is used to select individuals. Population is divided into two strata by residence. Within households with more than one member in the age of 16-74 years, one person is randomly selected, i.e. each individual represents one household.

Norway

The individuals are randomly selected from SPR – the statistical demographic population database, classified by municipality, family and individual.

Statistics Norway draws a master sample of between 4500 – 5000 individuals where all age groups are represented. This sample also includes family members. Then the individuals' age by 31 December 2006 are calculated.

The final sample consists of 2000 individuals drawn from the master sample. In some cases the final sample are adjusted due to gender and age groups. This in order to be representative for the Norwegian population.

Croatia

Two-stage, stratified sample design was employed in the survey. The sample is implicitly stratified according to the size of settlements and according to the region (21 counties).

At the first stage 500 sampling units are selected with the probability proportional to size (PPS with replacement), and at the second stage 10 persons aged between 16 and 74 years are selected. All of them were from different households. Thus 2000 persons were selected.

The probability of selection of a person is calculated by the formula:

$$\frac{N_a}{\sum_{slo} N_a} \cdot \frac{10}{N_a}$$

Where N_a is the number of persons aged between 16 and 74 years in the sampling unit.

Republic of Serbia

The sample is based on two-stage stratified random sampling design. Primary sampling units are enumeration districts and second stage units are households.

Stratification of enumeration districts was done according to type of settlement (urban and rural) and territory (14 territory regions).

Enumeration districts are stratified proportionally to the number of households by regions as well as urban and rural areas.

Enumeration districts are selected with probabilities proportional to the number of households, and from each enumeration district 5 households are randomly selected.

II.2.3.1 STRATIFICATION

The recommendation is to use a stratified sample of individuals or households with the aim to form groups (or layers) of units characterised, in relation to the variables subject of the survey, by maximum homogeneity within the layers and maximum heterogeneity between the layers. Achieving this goal in statistical terms means precision of estimates, or a reduction in sampling errors on a part with the sample quantity.

Each country should use the stratification variables according to what is most efficient to that country with particular attention to the demographic size of the localities.

II.2.3.2 SAMPLE SIZE

Calculation of sample sizes should take into account that this is a survey with multiple objectives. It has to ensure representative results for all the estimates produced. In particular calculation of sample size should take into account that each statistics have to be tabulated by age, sex, education level, employment situation, geographical location and type of locality.

As budgets are limited, the design of study involves making trade-offs along various dimensions. Larger samples make it possible to analyse sub-groups in depth but every interview increases survey costs.

On the basis of the previous considerations, it is suggested to adopt a mixed view, based on both cost and organisational criteria and on an evaluation of the sample errors of the main estimates on a national level and with reference to each of the territorial domains and to each of the breakdown variables of interest.

The calculation of sample sizes should be based on precision requirements (see above). On this basis countries should decide on sample design and calculate the sample sizes in order to receive estimates with sufficient quality and within possible budgetary constraints.

The table below gives an overview of the sample sizes reported by the national statistical institutes in the Final Reports i.e. Quality Reports for the 2007 surveys. Note that, due to national programs (e.g. production of regional estimates) and different units (household versus individual), the countries can't always be compared. For most EU countries, the final or net sample size was between 3000 and 6000 elements. The response and non-response as well as the quality (in terms of standard error) are discussed in more detail in the next chapters.

Table: Overview of the sample sizes and response rates (2007 survey)

Source: Quality reports for the Community survey on ICT usage in households and by individuals (2007)

	Unit	Gross sample size	Ineligible: out-of-scope	Other ineligible	Number of eligible elements	Non-contact	Refusal	Inability to respond	Rejected interviews	Other non-response	Net sample size	Unit response rate
Belgium	IND	14,924	255	0	14,669	1,707	454	0	0	2,087	10,421	71.0%
Bulgaria	IND	10,163	0	184	9,867	418	445	15	0	6	8,983	91.0%
Czech Republic	HH	7,059	531	395	6,133	271	1,380	6	136	0	4,340	70.1%
Denmark	IND	5,044	0	549	4,495	528	86	0	847	0	3,034	67.0%
Germany	HH	12,000	500	0	11,500	:	:	:	:	:	10,516	91.4%
Estonia	IND	4,772	0	0	4,772	14	35	0	0	0	4,723	99.0%
Greece	IND	6,267	10	0	6,257	688	435	0	0	0	5,134	82.1%
Spain	IND	27,690	2,097	2,545	23,782	2,084	2,407	160	0	0	19,131	80.4%
France	HH	7,100	746	538	5,816	698	832	57	2	0	4,227	72.7%
Ireland	IND	6,167	587	0	5,580	0	615	0	0	0	4,965	89.0%
Cyprus	IND	5,709	1,120	824	3,765	175	3	3	0	0	3,584	95.0%
Italy	IND	51,815	6,513	1,059	45,302	3,922	3,320	0	0	1,582	36,478	80.5%
Latvia	HH	5,495	285	357	4,853	993	314	29	89	288	3,140	65.0%
Lithuania	HH	7,500	601	221	6,678	333	653	33	0	81	5,578	83.5%
Luxembourg	IND	12,363	1,656	1,995	8,712	1,678	5,034	368	0	123	1,509	17.3%
Hungary	HH	10,089	809	418	8,862	615	766	40	0	152	7,289	82.2%
Malta	HH	1,500	0	0	1,500	99	97	14	23	0	1,267	84.5%
Netherlands	IND	9,480	0	3,304	6,176	215	1,051	137	8	742	4,023	65.1%
Austria	IND	15,712	0	0	15,712	0	7,881	303	0	0	11,311	47.9%
Poland	IND	17,026	3,652	0	13,374	199	135	16	0	42	12,982	97.0%
Portugal	IND	13,463	1,110	1,937	10,417	999	357	0	0	439	8,622	82.8%
Romania	HH	10,115	819	473	8,823	221	140	0	0	179	8,283	93.9%
Slovenia	IND	2,315	23	12	2,280	203	288	29	0	158	1,602	70.3%
Slovak Republic	IND	4,092	0	0	4,092	0	0	0	0	0	4,092	100.0%
Finland	IND	4,273	0	16	4,257	268	522	49	4	464	2,950	69.3%
Sweden	IND	6,847	3	0	6,844	627	1,281	111	7	869	3,949	57.7%
United Kingdom	IND	6,296	0	778	5,518	593	1,538	0	81	0	3,306	59.9%
Iceland	IND	2,020	0	31	1,989	211	188	29	0	1	1,560	78.4%
Norway	IND	1,930	0	14	1,916	224	440	64	0	14	1,174	61.3%
Croatia	IND	5,000	76	1,018	3,906	397	301	77	2	51	3,078	79.0%
Serbia	IND	2,000	0	0	1,939	41	14	4	2	0	1,939	97.0%

Regarding the 2007 survey and EU-27, the net sample sizes of households and individuals according to Eurostat's database is 151171 households and 229340 individuals.

II.2.3.3 WEIGHTING – GROSSING UP METHODS

The weighting factors are to be calculated taking into account in particular the probability of selection and external data relating to the distribution of the population being surveyed, where such external data are held to be sufficiently reliable.

As the sampling design used differs strongly across countries, it is difficult to present 'fit-all' guidelines. Moreover, the weighting procedures / grossing up methods are usually determined by the sampling design used. The discussion is more of a theoretical nature and goes beyond the scope of this manual.

Where more advanced methods for dealing with unit non-response are not feasible, it is advised to correct for unit non-response by adjusting the grossing up weights. Ideally, auxiliary information such as socio-economic differences should be taken into account.

II.2.4. Survey type

Data collection method

Face-to-face interviews, telephone interviews and postal surveys are all possible techniques of collecting data. The socio-demographic characteristics which can be found in registers need not be collected in the survey.

Face-to-face interviews are recommended especially where telephone penetration is low. Interviews face-to-face or by telephone are preferred to postal surveys because interaction between the respondent and the interviewer can be important for the understanding of the questions and the answers but care should be taken to ensure a low rate of proxy response. When using telephone interviews it is important to include mobile and ex-directory users as well as fixed line users in the population sampled. Postal surveys might be used to lower survey costs when budgets are limited.

The table below indicates that most countries collect the data via a face-to-face interview, although telephone interviewing is also a common technique. Postal surveys are used by only two countries.

Table: Overview of the data collection methods (2007 survey)

Source: Metadata reports for the Community survey on ICT usage in households and by individuals (2007)

Belgium	Face-to-face (for persons aged 65+: face-to-face or telephone)
Bulgaria	Face-to-face interviews
Czech Republic	Face-to-face - CAPI
Denmark	Telephone interviews - CATI
Germany	Self-administered mail survey
Estonia	Face-to-face - CAPI
Greece	Telephone interviews
Spain	Face-to-face CAPI and telephone interviews CATI
France	Telephone interviews
Ireland	Face-to-face - CAPI
Italy	Face-to-face, PAPI, self-administrate questionnaire.
Cyprus	Face-to-face using CAPI
Latvia	Face-to-face
Lithuania	Face-to-face
Luxembourg	Telephone interviews - CATI
Hungary	Face-to-face
Malta	Face-to-face
Netherlands	Telephone interviews - CATI
Austria	Telephone interviews - CATI
Poland	Face-to-face - CAPI
Portugal	Face-to-face
Romania	Face-to-face
Slovenia	Face-to-face and telephone interviews
Slovak Republic	Face-to-face
Finland	Telephone interviews - CATI
Sweden	Telephone interviews
United Kingdom	Face-to-face
Iceland	Telephone interviews
Norway	Telephone interviews

Independent versus embedded survey

For practical reasons, an important number of countries have embedded the ICT usage survey into an existing social survey. The main advantages of such approach are a cost-reduction (e.g. travel time for interviewers) and the fact that certain variables need to be collected only

once (this is especially the case for background characteristics such as educational level or employment situation). Further, linking the survey to an existing survey may give additional analytical possibilities (e.g. when embedded in the EU-SILC, the ICT usage pattern can be linked to the individual's or the household's living conditions).

However, although the filtering and routing limits the length of the ICT usage survey, it can be very burdensome for the respondents to be expected to give 50 yes/no answers following an already long interview for the survey vehicle, e.g. the Labour Force Survey. As the ICT usage questions will usually be inserted after the questions of the main survey, this can have unwanted negative effects on the response rates and the reliability of the answers.

The overview table indicates that a majority of the countries has a separate survey, although an important number of countries link the ICT usage survey to an existing survey. In this case, the ICT usage survey is usually embedded in another social survey such as the Labour Force Survey or general-purpose surveys (omnibus, micro-census, etc.).

Table: Overview of the survey vehicles (2007 survey)

Source: Quality reports for the Community survey on ICT usage in households and by individuals (2007)

Belgium	Labour Force Survey
Bulgaria	Stand-alone survey
Czech Republic	Labour Force Survey
Denmark	Stand-alone survey
Germany	Stand-alone survey
Estonia	Labour Force Survey
Greece	Stand-alone survey
Spain	Stand-alone survey
France	Stand-alone survey
Ireland	Labour Force Survey
Italy	Embedded in multipurpose social survey
Cyprus	Stand-alone survey
Latvia	Stand-alone survey
Lithuania	Stand-alone survey
Luxembourg	Stand-alone survey
Hungary	Stand-alone survey
Malta	Stand-alone survey
Netherlands	Stand-alone survey
Austria	Labour Force Survey
Poland	Stand-alone survey
Portugal	Stand-alone survey
Romania	Stand-alone survey
Slovenia	Stand-alone survey
Slovak Republic	Stand-alone survey
Finland	Stand-alone survey
Sweden	Labour Force Survey
United Kingdom	Embedded in Omnibus Survey
Iceland	Stand-alone survey
Norway	Embedded in Omnibus survey

Mandatory versus voluntary survey

Voluntary surveys are usually cheaper, quicker and easier to manage. A mandatory survey implies to make several attempts to contact the respondent or to send several reminders. This process usually makes the collection period longer as one need to wait a long time for all responses. The advantage of a mandatory survey is that your response rate is much higher, reducing the risk of having serious non-response bias (see also below, 0).

But a voluntary survey can settle this argument by increasing the sample size and sending reminder letters to the respondents.

In practice, the ICT usage surveys will be mandatory as of 2006 in all EU countries following the legal acts (Regulation 808/2004 and Regulation 1099/2005).

II.2.5. Questionnaire – Data collection tool

To enhance the comparability across countries, Eurostat provides a model questionnaire to the NSIs. It is recommended to adopt the model questionnaire for the national survey. However, for practical reasons (for instance linked to the survey vehicle), this may not be appropriate. In this case, the statistical institute should nevertheless make sure that the routing and the filtering is followed.

Eurostat designs the model questionnaire in close cooperation with the national statistical institutes, the main users and in coordination with other organisations such as the OECD. The results of previous surveys are taken into account, for instance to assess the relevance of certain questions or items, or to assess whether a variable needs to be collected on an annual basis. For reasons linked to the annually renewed implementing measures for Regulation 808/2004, the model questionnaire needs to be finalised 9 months before the survey takes place. In practice, this means that the questionnaire reaches its final status in March-April of the year preceding (T-1) the reference year (T). The obvious disadvantage is that 'last-minute' policy needs can't be included or that the results of the T-2 survey can't be fully taken into account. The advantage is that this gives the NSIs the necessary time to translate the questionnaire and implement the survey.

The model questionnaire for the survey on ICT usage in households and by individuals can be found in Annex II.3.1.

For a more general discussion of questionnaire design, the reader can consult the *Australian Bureau of Statistics Forms Design Standards Manual*:

<http://www.sch.abs.gov.au/SCH/A1610103.NSF/Survey+Design?OpenView>

II.2.6. Quality control systems

Quality control systems are of course country-specific as most statistical institutes have standard procedures and guidelines for plausibility checks or logic tests of datasets.

Such controls can be executed on-line, at the moment of the data capture by the interviewer or the data entry in the statistical institute, or after the data entry process (a program checks the data and prints the errors to be checked or corrected). On-line tests have the advantage that the errors can be corrected immediately (a lot of errors will be simple data-entry errors, f.i. typing 17773 instead of 1773), a disadvantage is that one needs data-entry staff that is familiar with the concepts in the survey (in case they have to decide if a certain value will be accepted or not). The latter case of course doesn't apply with CATI or CAPI where the interviewer immediately enters the data.

Below, some of the most common errors or problems are briefly discussed.

Measurement error

There are a number of sources of measurement error: survey instruments (questionnaire), the respondent, the information system, the mode of data collection, the interviewer. This paragraph focuses on the latter, other types can be found below. The first and probably most important step in the statistical process is the data collection or field work. One can design excellent questionnaires, excellent capture tools, excellent imputation methods and excellent data analyses, but as with any process the adage 'Garbage In Garbage Out' also applies to statistics. If the interviewer is not well-trained, one can't expect input data of a good quality. A more important problem is usually the reliability of the interviewers. The interviewers may have an interest in finishing the interviews in the shortest possible time. This can lead to interviews where certain questions are skipped (e.g. the person looks poor, let's fill in he's unemployed) or interviews that never took place (e.g. the person is 70 years old, let's put down he never used a computer nor Internet). Therefore the sensibilisation of interviewers (which can include threats of non-payment or exclusion for future interview rounds) and the follow-up is extremely important. The follow-up can be a check for coherence and consistency of the answers or follow-up phone calls to verify with the respondents whether the interview has actually taken place or whether the interview was of an acceptable quality (e.g. the respondent confirms that the interview only took 4 minutes, but on the questionnaire answers to 60 questions were recorded).

Invalid response

Relatively unimportant in the ICT usage survey as most answers are limited to Yes or No. However, it is possible that several items were ticked in questions where only one answer is expected.

Relationship error

Comparing the answers across the survey can reveal inconsistencies between the answers. The routing and filtering should normally guarantee that respondents are not presented with questions they can't answer. It is nevertheless possible that e.g. an individual aged 18 indicates higher educational level or that an unemployed person indicates 'place of work' as a location where the Internet has been accessed. In the first case, this is most probably an absolute error. In the latter case the combination is possible if the respondent lost his job only very recently (meaning he could have accessed the Internet at a place of work during the last 3 months).

Compulsory question left unanswered

Again, the routing and filtering should avoid this type of errors. CATI or CAPI programs usually improve the quality of the data capture in this respect. With traditional interviews or self-administered mail surveys, this error is more important.

Suspicious values

This can occur when the individual responses are compared with the average scores within a stratum and unexpected answers are found. In practice it will however be difficult to determine whether it concerns a mistake or an outlier.

In terms of quality of the survey *as such*, the methodology and outcomes of the survey can be benchmarked against other surveys:

Representativeness

It can be useful to do an *ex-post* check of the representativeness of the sample, e.g. does the sample have a representative age distribution, is there some variability in the occupational and educational codes?

Year-to-year comparison at aggregate level

Comparing the results for the current year with the previous survey can also reveal quality problems where the growth is outside the range of the expected growth (e.g. the proportion of individuals using the Internet decreases ...). In such cases, it is of course possible that the problem stems from the previous survey exercise ...

For this purpose, it can be interesting to produce some simple tabulations of the survey results.

Coherence or consistency with other surveys

The results can be compared with results from related survey or studies. However, in case inconsistent results are observed, it is not always easy to identify which survey gave the 'wrong' results.

II.2.7. Data processing

This chapter mainly discusses the treatment of non-response. Although the grossing-up methods can be considered as a part of the *data processing*, this topic is discussed above in section II.2.3.3.

II.2.7.1 MISCLASSIFICATION TREATMENT

Not applicable to the household survey.

II.2.7.2 **NON-RESPONSE TREATMENT**⁷

Introduction

An important source of non-sampling error in surveys is the effect of non-response on the survey results. Non-response can be defined as the failure to obtain complete measurements on the (eligible) survey sample. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response.

The latter case occurs when the interviewer was either unable to contact the respondent, no member of the household was able to provide the information, the respondent refused to participate in the survey or not enough information was collected in the interview (i.e. the response is too incomplete to be useful). This type of non-response is called unit non-response (see II.2.7.2.1): the sample unit does not provide any of the data required by the survey. Unit non-response is generally handled by adjusting the weight of the households and/or individuals that responded to the survey to compensate for those that did not respond.

Partial non-response or item non-response (see II.2.7.2.2) occurs when the respondents did not answer all questions because they did not understand or misinterpreted a question, refused to answer a question or could not recall the requested information. Item non-response is generally dealt with by imputation.

Effect of non-response on the quality of the data

Non-response – unit as well as item non-response – can seriously affect the quality of the data collected in a survey. Firstly, the characteristics (or answering pattern) of the non-respondents can be different from those collected among the sample units who did provide eligible answers. If such difference is systematic, serious bias can be introduced in the survey results. Secondly, the reduction of the sample size (overall or for certain questions) will increase the variance of the estimates. Thirdly, non-response can have an impact on the total cost of a survey exercise. Not only because a larger initial sample may be necessary, but also because of higher unit costs of the last few percentages of respondents (due to multiple visits). Finally, non-response can be an indicator of poor overall quality of the survey and thus create an image or confidence problem.

⁷ References for this chapter:

- Government Statistical Service (1997), *Report of the Task Force on Imputation – June 1996*. Government Statistical Service Methodology Series No. 3. London: Office for National Statistics.
- Statistics Canada (2003), *Household Internet Use Survey, 2002 – Microdata User Guide*.
- European Commission (2003), *Household Budget Surveys in the EU: Methodology and recommendations for harmonisation 2003*. Luxembourg: Office for Official Publications of the European Communities.
- National Centre for Social Research (1999), *Item non-response* (different contributions). Survey Methods Newsletter Volume 10 No.2 1999.
- Adjusting for Missing Data, Mohadjer, L., Choudhry, H., Studies of welfare populations: Data Collection and research Issues, 2001.
- Introduction to Survey Quality, Biemer, P., Lyberg, L., Wiley series in survey methodology, 2003.
- Weighting for non-response, Lynn, P., Survey and Statistical Computing, 1996.
- Non-response in the Norwegian Business Tendency Survey, Wang, J., Statistics Norway, 2004.

Minimising non-response

As prevention is always better than cure, attention should be given to avoiding non-response rather than treating non-response. The number (and timing) of reminder letters or call backs, the length of the fieldwork period, the survey technique(s) used, the length of the survey (i.e. the response burden), the use and structure of advance letters, the dissemination of previous results or the mandatory nature of the survey can all have an impact on the number of non-contacts or refusals. The length, design and complexity of the questionnaire, the interviewer's training or the interviewer's style can have an impact on the item non-response.

As this issue is common to all surveys, it will not be discussed in detail in this manual. This chapter is mainly focussed on the treatment of non-response in the specific context of the ICT usage survey.

II.2.7.2.1 UNIT NON-RESPONSE

Introduction

Unit non-response is defined as households/persons that are included in the sample but that have not participated to the survey and for which information consequently is missing for all the questionnaire variables.

Types of non-respondents include:

- Non-contact
- Refusals
- Inability to respond
- Rejected interviews
- Ineligible: out-of-scope
- Other ineligible
- Other non-response

Unit non-response can introduce bias in the survey results especially in situations in which the non-responding units are not representative of those that responded. Non-response increases both the sampling error, by decreasing the sample size, and non-sampling errors.

An overview of the non-response patterns in the 2004 survey (based on the 2004 Final Reports) can be found above (II.2.3.2, Sample size). It is clear that in most countries the unit non-response is important with rates of 25% and more, with *non-contacts* and *refusals* as the main types of non-response.

Weighting adjustment for unit non-response

The principal method for unit non-response adjustment is weighting. Most strategies for weighting for non-response involve dividing the respondents into a set of comprehensive and mutually exclusive groups, referred to as weighting classes. A weight is then applied to each class.

Weighting classes

In order to implement non-response adjustments, it is required to create weighting classes. It is desirable to divide the sample in "response homogeneity groups/classes". Within these classes the response rates should be as homogeneous as possible, and the response rate should be different among the classes. Data used to form these classes must be available to both non-respondents and respondents. Usually it is possible to get information about demographical (age, gender, ethnicity), geographical (urban/rural, zip code) or socioeconomical (employment, income) variables from administrative data.

More advanced methods for creating weighting classes are methods like classification based on a categorical search algorithm or a logistic regression model using auxiliary variables to estimate the probability of response.

Sample-Based Weighting Adjustment

In sample-based weighting adjustment the weight adjustment applied in each class, is equal to the reciprocal of the ratio of selected sample size to respondents within each of these classes (the inverse of the response rate within each class). This non-response adjustment factor should be multiplied with the initial base weight.

A simple example:

	Population (I)	Sample size (II)	Respondents (III)	Respondent with characteristic (IV)	Non-response adjustment Factor (V = II / III)	Initial Base Weight (I / II = VI)	Adjusted Base Weight (V*VI=VII)	Adjusted population estimate (=VIII)
Male	8 820 000	2 100	1 600	1 000	1.31	4 200	5 502	5 502 000
Female	9 020 000	2 200	1 750	1 200	1.26	4 100	5 166	6 199 200
Total	17 840 000	4 300	3 350	2 200				11 701 200

Alternative forms of sample-based weighting are that the weights are not inverse response rates, but estimated coefficients of a regression model (where survey response is the left-side variable). In this case, the weights are reciprocals of estimated response rate by the regression model.

Population-Based Weighting Adjustment

Population-based weighting adjustment requires population estimates and class membership of respondents. If there is no data available about the non-respondents, population-based adjustment still is possible since this uses external control counts for the population and not data from the sample. The method is used to correct simultaneously for both non-coverage and non-respondents. The method is used similar to the sample-based method.

In population-based adjustment (poststratification adjustment) the classes are created based on variables, which are known both for respondents and for the population. Weights are then applied in proportion to the ratio of population to achieved sample, so that the sums of the adjusted weights are equal to population totals for certain classes of the population.

A two-step procedure of first adjusting for non-response (sample-based adjusting) and then adjusting to known population counts is a common method that is used. However, this procedure is the same as a population-based weighting adjustment if the weighting classes in the sample-based and the population-based weighting adjustment are equal.

If the strata used in the stratification are used as classes in the weighting adjustment, there is no need for the weighting adjustment. The adjusted weighting procedure is then equal to the final grossing up/weighting procedure.

II.2.7.2.2 ITEM NON-RESPONSE

Introduction

As already mentioned above, there are several reasons for the data being unavailable. These include the refusal to provide an answer, the inability to provide an answer, inadequate quality of the provided answer (e.g. implausible, incomplete, inconsistent with answers to other questions, etc.). It can be caused by either the respondent (e.g. refusal) or the interviewer (e.g. failure to record the answer adequately) but also by the survey design itself (e.g. ambiguous routing or filtering).

In case a particular questionnaire shows too many errors, or if too many data are missing, it can be assumed that the household/individual in question has not co-operated satisfactorily in the survey. Here, the best solution is probably to remove the household/individual from the database and adjust the weighting coefficients for the other households accordingly. In other words, sampling units with a very high item non-response can better be classified as total non-response or unit non-response (cf. the category rejected interviews in the reporting template in §2.8). It is however difficult to define a threshold as not all questions are equally important (i.e. having missing data on four crucial indicators can possibly be worse than missing data on eight questions of secondary importance).

In other cases where the household/individual has supplied high-quality information for most variables but for which data on other variables are missing, the missing data can be estimated by using appropriate imputation techniques. The imputed values are supplied in such a way as to preserve the underlying structure of the data and to ensure that the resulting records will pass all required edits or plausibility checks. In other words, the objective is not to reproduce the true microdata value, but rather to establish internally consistent data records that yield good aggregate estimates.

Where, for a specific variable, the proportion of missing data in relation to the total number of households exceeds a certain threshold, it must be asked whether or not imputation is still appropriate as this variable can probably not be used for analytical purposes. High item non-response (e.g. more than 20% non-response) on a specific question across several countries, can be a good reason to consider dropping or modifying the question or variable.

It should be stressed that the detection and imputation of missing data is to be done by the national statistical offices as Eurostat only receives aggregated tables which do not allow imputations.

Why treating non-response in the survey on ICT usage by households and individuals?

Most methods of compensating for missing items implicitly or explicitly make the assumption that data are missing at random. That is, the probability of an item being missing does not depend of the value of the missing item.

In our survey on households' usage of information and communication technologies there are without any doubt systematic patterns in the occurrence of non-response. It is obvious that non-response may be higher among older respondents or lower educated respondents as they are more at risk of not understanding the questions. We can take this into account by imputing within strata or classes. But the risk of wrongly imputing the data of ICT users (who feel concerned and "happily" answer the questions) to non ICT users (who drop out because they consider themselves not concerned by the survey) remains when it is the research variable itself (e.g. internet use) which may be the critical factor for the willingness or ability to provide an answer.

The logical solution to this problem would be not to impute at all. However not imputing does not exist as there is always an implicit imputation.

This is obviously the case where a respondent is dropped: the status moves from item non-response to unit non-response, which will normally be solved by changing the extrapolation weight of the other respondents (and this basically comes down to a mean imputation ...).

Where the statistician decides to preserve the collected data instead of imputing, the user will in most cases do make an implicit imputation of the "non-stated" cases. Let's take the following numerical example ("did you use internet during the past 3 months"):

Answering category	Number of respondents (grossed up)	Percentage
Yes	924 000	21,0%
No	1 980 000	45,0%
Non-stated (missing)	1 496 000	34,0%
Total	4 400 000	100,0%

The proportion of persons using the internet can be calculated in two ways:

- Number of Yes answers compared to total population: $924\,000 / 4\,400\,000 = 21,0\%$
- Number of Yes answers compared to total respondents (Yes + No): $924\,000 / 2\,904\,000 = 31,8\%$

In a realistic setting, the end-user will only be interested in the proportion of users, not in the percentage of non-users and definitely not in the number of Non-stated. If these Non-stated are mentioned separately, they will most probably be overlooked by the user or interpreted in different ways by different users – by accident or on purpose (which is not a good thing in terms of transparency and relevance of the official statistics).

Both results above suffer from silent or implicit imputation. In the first case, the Non-stated cases are treated as 'No' answers. This bears the risk of overestimating the number of non-users: whether one uses the internet is in any case not the only factor affecting response rates in household surveys.

In the second case, there is a risk of overestimating the number of users: the underlying distribution of 'Yes' and 'No' answers in the 'Non-stated' group is implicitly supposed to be identical to the distribution in the group of respondents while the incidence of non-use is expected to be higher among the non-respondents.

The true value (or better: most suitable estimate) is somewhere in between, but unknown.

The easy to implement mean imputation (see below) will normally give exactly the same proportion of internet users as the case where we only take into account the respondents, but it should be remembered that this method is the 'worst imputation scenario' where no additional information is taken into account. More intelligent methods would give more accurate estimates. E.g. in a postal survey the respective response pattern of the consecutive reception waves can help fine-tuning the imputation: if response to our questionnaire is related to ICT use, the questionnaires received after a first or second reminder letter will show lower internet use figures. Such degressive pattern could be introduced in the imputation process if we assume that non-response is the extreme case of a late reply.

How to treat non-response in the survey on ICT usage by households and individuals?

The choice of the imputation method is at the discretion of the national statistical institutes, but a number of options are briefly discussed in this section. As experiences with the application of different methods are still missing, putting forward one particular method is not desirable at this stage. A more detailed discussion of these and other methods is beyond the scope of this manual, but can be found in the extensive literature on the issue of dealing with non-response.

Deductive methods

These methods are rather related to heuristics than to modelling. They try to deduct the most logical answer using the available information for the household or individual. In general, such procedures will be part of the validation checks and not of the non-response treatment.

Example: the respondent did not state whether s/he uses internet but from his other answers, we know s/he doesn't use a computer. In this case, the most obvious value to impute for internet use would be No. Wrong imputations will occur if this specific person only uses internet by means of a mobile phone, TV, etc.

Imputing the mean or mode

This method consists of imputing missing values by the mean observed in the group of respondents in case of numerical variables or the mode in case of categorical or binary variables.

Instead of imputing the overall mean or mode, usually, the imputation will be done taking into account some background characteristics of the household: before imputing, the

respondents are grouped into different classes according to background variables such as age group, gender and educational level; the breakdown variables can however also include study variables (e.g. computer users versus non users). Within each group, the class mean or class mode is imputed to cases with missing values. The classes may be different for each variable to be imputed.

The big advantage of this method is that it is very easy to implement and to explain. The main drawback is that it may compress the distribution of the survey variables (as the less popular items will probably never serve as a donor).

Example: in the class 'age group 45-54' x 'lower educational level' x 'male', 20% of the respondents appear to use internet (while 80% don't). The most popular answer – in other words, the mode – is No. Therefore, all non-respondents will be imputed with a No value (which means approximately 20% of those may be misclassified – to avoid this problem one can try the next technique).

Hot deck imputation

For each respondent with a missing value for a specific variable, this value will be imputed with the corresponding value from the previous respondent in the database with a valid response on this variable. Usually, this method is applied within classes in order to improve the quality of the imputations.

After fixing a starting value for each item and each class, each case is processed sequentially. If the case has a missing item, this is replaced by the imputation value from the relevant class. If the item is not missing, it replaces the stored initial imputation value for its class, and can be used for imputation of subsequent missing items.

Closely related to this sequential hot deck imputation, is hierarchical hot deck imputation. Here, a lot more imputation classes can be considered as the boundaries of the classes are not fixed. When no suitable donor is found at the finest level of the classification, classes can be collapsed into broader groups until a donor is found. Taking into account the relatively small sample sizes used in our ICT survey, the hierarchical hot deck imputation may be less appropriate than the sequential variant (there are probably not enough cases available to fill the high number of classes).

These methods are relatively simple to execute, assign real/existing values to a non-respondent and better respect the underlying distribution than the mean or mode imputation. On the other hand, the algorithms may be more difficult to program (especially in the case of hierarchical hot deck) and there is a risk that one donor value is used several times (in case successive cases all show a missing value).

Nearest-neighbour imputation

This method relies on being able to identify the distance between any two units based on some suitable distance measure.

Regression imputation

Imputing by regression methods is usually carried out one variable at the time. The methods require that the values of one or more auxiliary variables are known for both the complete cases on which the variable of interest is recorded and for the missing cases. A regression model is fitted that relates the variable of interest to the set of auxiliary variables. For categorical variables, particularly binary variables, logistic regression is usually used. In this case, the imputed value will usually be the value with the higher predicted probability. For categorical variables with many categories, this method is less suited. For binary variables, more complex methods such as discriminant analysis could also be used.

Indicating an optimal imputation method is difficult. The best method will depend on the data available, the nature of the survey and of course the nature of the question (binary, numerical, etc.).

A general rule of thumb is however that the more simple methods such as mode imputation within classes or sequential hot deck do not necessarily give poorer results than the advanced methods using modelling. This argument will even be stronger when making a cost/benefit analysis (in terms of processing and programming time).

II.2.7.3 COMPUTING ACCURACY MEASURES

The use of a sample survey unavoidably leads to sampling error. To get an idea of the sampling error, it is advised to compute this error for certain indicators. Such information is not only interesting to decide whether the indicator is reliable enough to be published, but can also help to determine whether the sample size should be increased in a next edition of the survey for certain strata where a large sampling error was diagnosed.

The sampling error reflects the fact that only a particular sample was surveyed rather than the entire population.

As accuracy measure, the standard error (the square root of the variance) of the estimator for proportions is recommended (see also Chapter 9 of the *Reporting template*, Annex II.3.3).⁸

The estimated standard error (standard deviation), expressed by the square root of the estimate of the sampling variance, shall not exceed 2% of the overall proportions and shall not exceed 5% for the proportions relating to the different subgroups of the population, where these subgroups constitute at least 10% of the total population in the scope of the survey.

The estimation of the sampling variance should ideally take into account the sampling design (e.g. the stratification).

⁸ As mentioned in the original 2006 version of the manual, the (estimated) relative standard error – or (estimated) coefficient of variation (CV) – is the ratio of the square root of the variance of the estimator for the proportion to the expected value of the proportion. It is estimated by the ratio of the square root of the estimate of the sampling variance to the estimated value. To avoid that the level of the estimate when using the coefficient of variation has an extreme impact on the accuracy measure (e.g. a proportion of 1% leads to a denominator of 0,01 when calculating the CV, in other words, the standard deviation is multiplied by 100), the standard deviation (i.e. the square root of the sampling variance) should be used as a reference measure.

Some guidelines for interpretation and suggested use of the standard deviation can also be found in section II.2.3 above where the sampling design is discussed.

The computed accuracy measures will also determine the reliability of the estimates. If the standard error for a certain proportion or absolute value is too high, the cell needs to be suppressed.

In case the computation of the standard error for all cells is not feasible, the number of respondents underlying the estimate can be used to decide whether a proportion can be published (e.g. if based on less than 10 respondents, the estimate is not published).

See also Annex II.3.2, the Transmission Format: Cells that do not comply with the quality standards in terms of standard errors or are for other reasons not of a sufficient quality (high item non-response, question not in line with the model questionnaire, etc.) should be "flagged" as unreliable; in the latter case notes accompanying the data should explain why the data is considered unreliable.

Parameters that affect the sampling error are the estimate i.e. the actual proportion and the sample size. Standard errors reported in the last surveys have shown the overall proportions, for example for all individuals, are relatively accurate. But the situation is different when looking at the accuracy for the breakdowns.

The breakdown between men and women doesn't really affect the quality of the estimates, because of the absence of a strong relation between Internet use and gender and hence relatively equal sample sizes for the two groups.

For the breakdown by age group, the situation is more problematic. In the older age groups, the sample Internet users is becoming very small (because of the filter in the questionnaire), which leads to unreliable estimates in most cases. These conclusions can be taken into account when designing the sample, e.g. by including relatively more respondents in the sample for strata where poor quality in terms of accuracy can be expected (such overrepresentation should of course be compensated for in the weighting). Unfortunately, cost restrictions may not always allow for larger samples and reducing the sample size (and accuracy?) in other strata may not be the optimal trade-off either.

II.2.8. Survey execution reports

Drawing up reports after the execution of the survey is not only a tool for a self-assessment of the work carried out by the statistical office and a detection of possible points where there is space of improvement. Survey execution reports are also a tool for assessing the comparability from one year to another or from one country to another.

The methodological reports for the ICT usage surveys coordinated by Eurostat should cover the following subjects:

13. General methodological information: reference period(s); survey period; survey vehicle, where applicable; survey type; pre-tests; methodological differences compared to previous data collection exercises.
14. Statistical unit(s), scope and target population: statistical unit; age groups covered; territorial coverage; target population for households and for individuals; non-target population for households and for individuals.

15. Information on the national questionnaire.
16. Sampling frame: name and description of the sampling frame or register used; known shortcomings of the sampling frame, if any.
17. Sampling design: sampling method; additional measures taken at the time of sampling design to improve representativeness.
18. Unit non-response at household and at individual level, by giving information on: gross sample size; number of out-of-scope cases, number of other ineligible cases; number of eligible elements; number of non-contacts; number of refusals; number of cases unable to respond; number of rejected interviews; other non-response; net sample size; unit response rate. Additionally, the report shall discuss the methods used for minimizing non-response as well as the methods used for dealing with unit non-response.
19. Item non-response: variables or items with item response rates below 90%; methods used for dealing with item non-response.
20. Grossing-up procedures for households.
21. Grossing-up procedures for individuals.
22. Sampling error: standard errors for a selected group of indicators or sub-indicators listed in the template to be provided by the Commission; the calculation method for the standard error; comments on reliability, representativeness and completeness of the data.
23. Problems encountered and lessons to be learnt.
24. The report should include the questionnaire in national language and, if available, in English. The statistical unit is the base type of the elements of a group (also called population) that we want to observe or analyse. The basic statistical operations of classification, aggregation and ordering are done on the statistical unit.

In cases where the recommendations provided in this methodological Manual are not followed, the report should clearly describe deviations, the impact of those on comparability a justification for the deviation and – where possible – plans to comply with the recommendations.

The *Reporting template* used for the European ICT usage surveys, including some instructions, can be found in Annex II.3.3.

The template covers both the Metadata report (Interim Report) and the Quality report (Final Report) referred to in Article 7(4) of Regulation 808/2004 (see Part III, Annexes III.1 and III.2).

II.3. ANNEXES

II.3.1. Model questionnaire

2008 / Version 3.1.

Community survey on ICT usage in households and by individuals 2008

Eurostat Model Questionnaire (v3.1)

Module A

Access to Information and Communication Technologies

Note: this module is directed to the household

A1 Do you or anyone in your household have access to a computer at home?

(any type: desktop, laptop, palmtop)

Yes

No

[-> go to A2]

A2 Do you or anyone in your household have access to the Internet at home, regardless of whether it is used?

Yes [-> go to A3]

No [-> go to A5]

Don't know..... [-> go to B1]

A3 On which of the following devices is the Internet accessed at home?

(tick all that apply)

a) Desktop computer

b) Portable computer (laptop)

c) Other mobile devices

of which: (optional) c1) via Internet enabled mobile phone (GPRS, UMTS, etc.).....

c2) via handheld computer (palmtop, PDA)

d) TV set with specific Internet device (e.g. digital TV or set top box)

e) Games console

f) Don't know

[-> go to A4]

A4 What types of Internet connection are used?

(tick all that apply)

- a) Modem (dial-up access over normal telephone line) or ISDN
- b) DSL (e.g. ADSL, SHDSL, etc.)
- c) Other broadband connection (e.g. cable, UMTS, etc)
- d) Mobile phone over narrowband (GPRS, etc.)

[-> go to B1]

A5 What are the reasons for not having access to the Internet at home?

(tick all that apply)

- a) Have access to Internet elsewhere
- b) Don't want Internet (because content harmful, etc.)
- c) Don't need Internet (because not useful, not interesting, etc.)
- d) Equipment costs too high
- e) Access costs too high (telephone, etc.)
- f) Lack of skills
- g) Privacy or security concerns
- h) Physical disability
- i) None of the above, but other

[-> go to B1]

Module B: Use of computers

Note: this and the following modules are directed to the individual selected within the household

B1 When did you last use a computer? (filter question)

- Within the last 3 months [-> go to B2]
Between 3 months and a year ago [-> go to B4]
More than 1 year ago [-> go to B4]
Never used one [-> go to C1]

B2 How often on average have you used a computer in the last 3 months?

(tick one)

- Every day or almost every day
At least once a week (but not every day)
At least once a month (but not every week)
Less than once a month

[-> go to B3]

B3 Where have you used a computer in the last 3 months?

(tick all that apply)

- a) At home
b) At place of work (other than home)
c) At place of education
d) At another person's home
e) Other (e.g. public library, hotel, airport, internet café, etc.)

[-> go to B4]

B4 When did you last take a training course (of at least 3 hours) on any aspect of computer use?

(for respondents who didn't answer "Never used one" in question B1)

OPTIONAL QUESTION

- Within the last 3 months
Between 3 months and a year ago
Between 1 and 3 years ago
More than 3 years ago
Never taken one

[-> go to C1]

C1 When did you last use the Internet? (filter question)

- Within the last 3 months [-> go to C2]
- Between 3 months and a year ago [-> go to D8]
- More than 1 year ago [-> go to D8]
- Never used it [-> go to D8]

C2 On average how often did you use the Internet in the last 3 months?

(tick one)

- Every day or almost every day
- At least once a week (but not every day)
- At least once a month (but not every week)
- Less than once a month

[-> go to C3]

C3 Where have you used the Internet in the last 3 months (using a computer or any other means)?

(tick all that apply)

- a) At home
 - b) At place of work (other than home)
 - c) At place of education
 - d) At another person's home
 - e) At other places
- of which: (optional)
- e1) Public Library
 - e2) Post office
 - e3) Public office, town hall, government agency
 - e4) Community or voluntary organisation
 - e5) Internet Café
 - e6) Hotspot (at hotels, airports, public places etc.).....

[-> go to C4]

C4 Do you use any of the following mobile devices to access the Internet?

(tick all that apply)

- a) Mobile phone via GPRS
- b) Mobile phone via UMTS (3G)
- c) Handheld computer (palmtop, PDA)
- d) Portable computer (laptop) via wireless connection away from home or work
- e) None of the above

[-> go to C5]

C5 For which of the following activities did you use the Internet in the last 3 months for private purpose?

(tick all that apply)

Communication, information search and on-line services

- a) Sending / receiving e-mails
- b) Finding information about goods or services
- c) Using services related to travel and accommodation
- d) Downloading software (other than games software).....
- e) Reading or downloading online news / newspapers / news magazines
- if yes to e) e1) Have you subscribed to news services or products to receive them regularly Yes
- No
- f) Looking for a job or sending a job application
- g) Seeking health-related information (e.g. injury, disease, nutrition, improving health, etc)

Banking, (optional: selling of goods or services)

- h) Internet Banking
- i) Selling of goods or services, e.g. via auctions (optional)

Training and education

- j) Looking for information about education, training or course offers
- k) Doing an online course (in any subject)
- l) Consulting the Internet with the purpose of learning

[-> go to C6]

C6 For which of the following activities relating to interaction with public services or administrations and during which period did you use the Internet for private purpose?

(tick all that apply)

- | | in the last
3 months | in the last
12 months |
|---|--------------------------|--------------------------|
| a) Obtaining information from public authorities' web sites | <input type="checkbox"/> | <input type="checkbox"/> |
| b) Downloading official forms | <input type="checkbox"/> | <input type="checkbox"/> |
| c) Sending filled in forms | <input type="checkbox"/> | <input type="checkbox"/> |

[-> go to D1]

D1 Did you use the Internet in the past 3 months for private purposes for the following communication activities?

(tick all that apply)

- a) Telephoning over the Internet.....
- b) Video calls (via webcam) over the Internet.....
- c) Posting messages to chat sites, newsgroups or on-line discussion forum.....
- d) Use of instant messaging (real-time communication with others by typed text).....
- e) Reading weblogs or blogs.....
- f) Creating or maintaining own weblog or blog.....
- g) None of the above.....

[If 'yes' to categories a) and b), go to question D2; otherwise go to question D3]

D2 Did you replace with your Internet calls the following other means of communication?

(if 'yes' to a and/or b in question D1)

(tick all that apply)

- | | Very much | To some extent | Not at all |
|--|--------------------------|--------------------------|--------------------------|
| a) Mobile phone calls..... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| b) Use of fixed telephone line (not linked to Internet)..... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| c) Use of e-mail..... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| d) No effect on other communication means..... | <input type="checkbox"/> | | |

[-> go to D3]

D3 Did you use the Internet in the last 3 months for the following leisure activities related to obtaining and sharing audiovisual content? *(filter question)*

(tick all that apply)

- a) Listening to web radios and/or watching web television.....
- b) Downloading and/or listening to music (other than via web radio).....
- c) Downloading and/or watching movies, short films or video files (other than via web TV)..
- d) Using peer-to-peer file sharing for exchanging movies, music, video files.....
- e) Using podcast service to automatically receive audio or video files of interest?.....
- f) Downloading computer or video games or their updates.....
- g) Playing networked games with others.....
- h) Uploading self-created content (text, images, photos, videos, music etc.).....
to any website to be shared
- i) Using browser based news feeds (e.g. RSS) for reading new content on websites.....
- j) None of the above

[-> go to D7]

[If 'yes' to categories b), c), d) and/or e), go to question D4 ; otherwise go to question D5]

D4 On average how often did you download music and/or films in the last 3 months?

(if 'yes' to b, c, d, and/or e in question D3)

(tick one)

- Everyday or almost every day.....
- At least once a week (but not every day).....
- At least once a month (but not every week).....
- Less than once a month.....
- Not applicable (only listened to music and/or watched films)....

[-> go to D5]

D5 Did you pay in the last 3 months for online audiovisual content?

(filter question)

- Yes [-> go to D7]
- No [-> go to D6]

D6 What would make you pay for online audiovisual content?

(tick all that apply)

- a) Lack of free available content.....
- b) Right to share legally protected content.....
- c) More convenient payment methods.....
- d) More advantageous prices compared to offline content.....
- e) Better quality of paid content than free services.....
- f) Wider range of choices, content more easily available.....
- g) None of the above, but other (e.g. to support artists work etc.).....
- h) Nothing, no willingness to pay.....

[-> go to D7]

D7 Did the use of the Internet replace your time spent with off-line media or other off-line activities?

(tick all that apply)

- | | Very much | To some extent | Not at all |
|---|--------------------------|--------------------------|--------------------------|
| a) Reading of online news instead of printed news, newspapers, magazines | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| b) Downloading music files instead of buying a CD..... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| c) Downloading films and videos instead of buying/renting a DVD..... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| d) Listening to web radio instead of listening to normal radio..... | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| e) Using online contacts instead of personal contacts with
public services and administrations | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

[-> go to D8]

D8 Do you use a mobile phone? (filter question)

(tick only one)

Yes [-> go to D9]

No [-> go to E1]

D9 For which of the following activities did you use a mobile phone for private purposes in the last 3 months (other than involving voice calls or SMS)?

(tick all that apply)

- a) Sending photographs or video clips.....
- b) Uploading photographs or video clips from your phones to websites.....
- c) Receiving subscription-paid information services.....
(for example news, weather forecast, sports results etc.)
- d) Browsing the Internet.....
- e) Reading your e-mails.....
- f) Downloading and/or watching TV or video.....
- g) Paying for goods or services (instead of cash or credit card).....
- h) Personal navigation (for finding location or address), use of location-aware.....
services (e.g. to receive nearby travel, shopping, event information)
- i) None of the above.....

[-> go to D10]

D10 Do you use pre-payment or post-payment for your mobile phone?

(tick all that apply)

a) Pre-paid.....

b) Post-paid.....

(Optional) if yes to b)

b1) Do you pay a flat rate for Internet access via your mobile phone?

Yes...

No.....

[-> go to E1]

(for respondents who didn't answer "Never used it" in question C1)

E1 When did you last buy or order goods or services for private use over the Internet (excluding manually typed e-mails) ?
(filter question)

Within the last 3 months [-> go to E2]
 Between 3 months and a year ago [-> go to E2]
 More than 1 year ago [-> go to F1]
 Never bought or ordered [-> go to F1]

E2 What types of goods or services did you buy or order over the Internet for private use in the last 12 months?
(tick all that apply)

a) Food or groceries
 b) Household goods (e.g. furniture, toys, etc)
 c) Films, music
 d) Books, magazines, newspapers or e-learning material
 e) Clothes, sports goods
 f) Computer software and upgrades (incl. computer and video games)
 g) Computer hardware
 h) Electronic equipment (incl. cameras)
 i) Share purchases, financial services or insurances
 j) Travel or holiday accommodation
 k) Tickets for events
 l) Lotteries or betting
 m) Other

[If 'yes' to categories c), d) or f), go to question E3; otherwise go to question E4]

E3 Were any of the following products that you bought or ordered over the Internet downloaded or accessed from websites rather than delivered by post etc.?
(if 'yes' to c, d or f in question E2)
(tick all that apply)

Films, music
 (Electronic) books, magazines, newspapers, e-learning material
 Computer software (incl. computer and video games and software upgrades)

[-> go to E4]

E4 From whom did you buy or order goods or services for private purpose over the Internet in the last 12 months?
(tick all that apply)

a) National sellers.....
 b) Sellers from other EU countries.....
 c) Sellers from the rest of the world.....
 d) Country of origin of sellers is not known.....

[-> go to F1]

Socio-demographic background characteristics

F1 Age:

F2 Sex: Man
 Woman

F3 Educational level: Primary or lower secondary education, no formal educat. [ISCED 0, 1 or 2]
(tick only one) Upper secondary education [ISCED 3 or 4]
 Tertiary education [ISCED 5 or 6]

F4 Employment situation: Employee or self-employed (incl. family workers)
(tick only one) Unemployed
 Student (not in the labour force)
 Other not in the labour force
(retired, inactive, in compulsory military service, etc.)

F5 Occupation: < description >
< to be recoded into at least 3-digit ISCO categories >

F6 Region of Residence < description NUTS 1 >

F7 Region of Residence < description NUTS 2 > OPTIONAL QUESTION

F8 Geographical location: "Convergence" Region
 "Regional Competitiveness and Employment" Region

F9 Type of locality: Densely-populated area
 Intermediate area
 Thinly-populated area

F10 Number of members in the household:

F11 of which, number of children under 16:

F12 Household income: _____ <currency>
(average net monthly income) < to be recoded at least into size bands compatible with income quartiles >

II.3.2. Transmission format

The transmission of aggregate data is mandatory according to current EU legislation:

⇒ See document ***MM2008_AnnexII.3.2_TransmissionFormatHH2008.doc***

The additional transmission of microdata is optional.

⇒ See document ***MM2008_AnnexII.3.2_TransmissionFormatHH2008-microdata.doc***

II.3.3. Reporting template

Note: the reporting template for the interim (metadata) & final (quality) reports refers to the 2008 ICT household survey.

⇒ See document ***MM2008_AnnexII.3.3_ReportingTemplateHH2008***

Part III

EU Legislation

III.1. REGULATION (EC) No 808/2004 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL OF 21 APRIL 2004 CONCERNING COMMUNITY STATISTICS ON THE INFORMATION SOCIETY

Official Journal of the European Union, 30.04.2004, L143.

⇒ See document ***MM2008_AnnexIII.1_Reg808-2004.pdf*** (7 pages)

Versions in other languages can be found on the EU's legislation server

<http://eur-lex.europa.eu/en/index.htm>

1. click 'Official Journal'
2. select 'Year' (=2004) and 'OJ Number' (=143)
3. select language (upper right, only the 11 official languages before 01/05/2004)
4. scroll down in the table of contents to the link to page 49.

III.2. COMMISSION REGULATION (EC) No 847/2007 OF 18 JULY 2007 IMPLEMENTING REGULATION (EC) NO 808/2004 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL CONCERNING COMMUNITY STATISTICS ON THE INFORMATION SOCIETY

Official Journal of the European Union, 19.07.2007, L187.

⇒ See document ***MM2008_AnnexIII.2_Reg847-2007.pdf***

Versions in other languages can be found on the EU's legislation server

<http://eur-lex.europa.eu/en/index.htm>

1. click 'Official Journal'
2. select 'Year' (=2007) and 'OJ Number' (=187)
3. select language (upper right)
4. scroll down in the table of contents to the link to page 5.