JRC TECHNICAL REPORTS

Sampling system for LUCAS 2015

Javier Gallego, JRC
Alessandra Palmieri, Eurostat
Helena Ramos, Eurostat

2015

Joint
Research
Centre

EUR 27596 EN

Sampling system for LUCAS 2015

JRC Science Hub
https://ec.europa.eu/jrc

JRC98827

EUR 27596 EN

ISBN 978-92-79-54001-1 (PDF)

ISSN 1831-9424 (online)

doi:10.2788/963 (online)

How to cite: Gallego F.J., Palmieri A., Ramos H., 2015, Sampling system for LUCAS 2015, EUR 27595 EN; doi: 10.2788/963

CONTENTS

Abstract

The sampling design of LUCAS 2015 took into account experience from previous campaigns. While remaining a Two-phase sampling scheme, efforts were focused in reducing the set of excluded points from the second phase sample, which are placed in difficult and inaccessible areas. This action was implemented both through:

1. *the improvement of the first phase sample (LUCAS master sample) by updating each point of the grid with the most recent available information (NUTS borders, road network, elevation, Corine Land Cover results (CLC2006))*
2. *the fine-tuning of the past rule for eligibility of the point introducing an additional new indicator of accessibility based on CLC; the final criteria combine all the auxiliary information [Elevation, Road distance, Slope, CLC indicator].*

However excluded points will remain a likely source of bias which has to be treated separately from the field survey. The corresponding excluded area needs to be covered with a complementary photo-interpretation operation, including an ex-post photo-interpretation also for 2012 and 2009 surveys. Such a photo-interpretation task, although important and linked to the sampling exercise, is not addressed in this document but represents next priority for the future work.

1. Introduction and Background: Two-phase sampling scheme in the previous LUCAS surveys

After the end of the Pilot phase, which lasted until the 2006 field-survey, the LUCAS (Land Use and Cover Area frame) survey design can be considered stable: two phase sampling with stratification aiming at producing estimates at NUTS2 region level.

However in each round some improvements and fine-tuning have been added on the base of the experience developed and the feedbacks from stakeholders. The challenge is represented by the trade-off between introducing improvements and keeping comparability with previous surveys.

First phase (Base and Master)

A two-phase sampling design has been adopted in the LUCAS survey since 2006.

The **base -**was obtained using the 1 km^2 grid resulting from the INSPIRE (INfrastructure for SPatial InfoRmation in Europe) recommendations; it included around 4,000,000 points in the entire European Union territory.

The projection used is the Lambert Azimuthal Equal Area coordinate reference system (ETRS 1989 LAEA)[1]. The grid is squared, with origin: 4,321,000 m West of centre point of the projection (52N, 10 E), and 3,210,000 m South of the projection centre point (52N 10E) and orientation: South – North, West – East. Each point has been given a unique numeric code going sequentially from South-West to North-East direction.

The **LUCAS first phase** sample or **LUCAS master** is a subset of the base file corresponding to a systematic 2-km grid in the LAEA coordinates. After excluding points located on small islands, it includes a total of 1,097,607 points for 28 EU countries. Each point of this master sample has been photo-interpreted for stratification with a simple classification of 7 classes[2] leading to 7 strata. Most of the points were photo-interpreted in 2005[3] on images that could not be kept because of copyright limitations. This photo-interpretation was based on the most recent ortho-photos or, where ortho-photos were not available, on satellite imagery (the "Image2000" images, obtained from Landsat7 ETM+, which were also used to produce CORINE Land Cover 2000). There is a general awareness that the stratification should be reviewed.

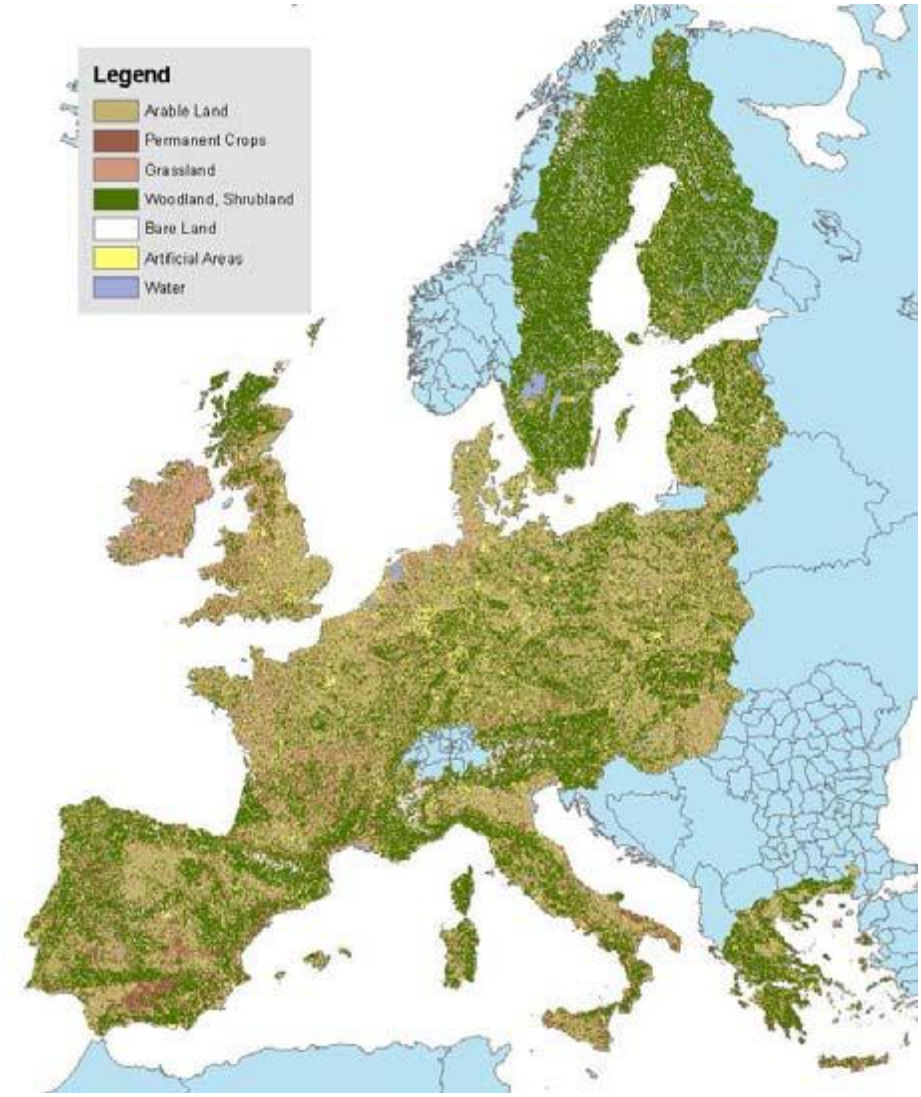
Results of the stratification on 23 countries are reported in Figure 1 (data source: LUCAS).

[1] INSPIRE-recommended

[2] 1. Arable land, 2. Permanent crops, 3. Grassland, 4. Woodland and shrubland, 5. Bareland, 6. Artificial, 7. Water and wetlands.

[3] Latvia and Lithuania were photo-interpreted in 2004; Bulgaria, Cyprus, Malta and Romania were not covered by stratification of 2005.

Figure 1: Land cover distribution in 23 EU Member States according to ortho-photo interpretation (2004/5)



Second phase sample or field sample – Overview of main pillars of sampling design in 2006/2009/2012 LUCAS survey

From the stratified first phase sample, a second phase sample of points, namely the field sample, is extracted to be classified during field visit according to the full land classification[4].

[4] *Two levels classification for Land Cover – 54 classes, and three levels classification for Land Use – 33 classes.*

This field sample is subsampled with a method that allows tuning the sample size per stratum and ensures a certain spatial homogeneity at the same time.

LUCAS 2006

The 2006 survey was carried out in 11 Member States (Luxembourg, Belgium, Czech Republic, Germany, Spain, Poland, Italy, France, the Netherlands, Hungary and Slovakia) to test the methodology at EU level with a restricted budget. The focus of the survey was agricultural land with a sampling rate of 50% for arable land and permanent crops, and of 40% for grassland, (non-agricultural strata are covered with a sampling rate of 10% each). Points above 1200 m were excluded because the crop area above this threshold is marginal.

The main drawbacks of the sampling design adopted until 2006 were the imbalance of the strata size (the agricultural strata were over-represented) and the geographical detail focused only at EU level.

LUCAS 2009

In 2009, being the first official round of the LUCAS survey covering all the EU countries (except Malta and Cyprus), the focus of the survey changed from a merely agricultural to a broadly agro-environmental one. Taking into account also the users' needs for more geographically detailed figures, the sampling rates were tuned in function of targets of coefficient of variation (CV) for each major land cover class (level 1) and NUTS 2 region, using as much as possible all the available auxiliary information[5].

LUCAS 2006 sample points were included as much as possible in order to collect longitudinal data on land cover and land use (panel approach). Points belonging to islands not connected to main land and points above 1000 m were considered inaccessible and excluded from the second phase sample.

LUCAS 2012

In 2012 a redistribution of points has taken place on the basis on the 2009 survey results and the users` needs.

LUCAS is a field survey and the added value of such approach, if compared to mapping approach for monitoring land cover/use such as Corine Land Cover (CLC), is represented by the field observation of features (land cover and land use) which cannot be detected with more detail otherwise. Unfortunately access to points can be difficult in absence of adequate road network, for the landscape characteristics, or simply due to the denied access from land owners or because point is placed in forbidden areas. In LUCAS 2009 out of 234.000 points of the sample, around 59.000 points could not be

[5] For details see 2009 ITACOSM09 3First Italian Conference on Survey Methodology – 10-12 June 2009, Siena, – Martino, Gallego, Palmieri "Use of auxiliary information in the sampling strategy of a European area frame agro-environmental survey"

visited and were photo interpreted[6] (29.000 in the field and 29.000 in the office[7]) (Table 1).

As the information recorded in LUCAS are used in the production of reference data (e.g. for the validation and calibration of CLC or Copernicus HRL (High Resolution Layers)) in the selection strategy applied for 2012, in agreement with the data users, it was given priority to accessibility of the points.

A large number of points (around 30% of the first phase sample) were considered difficult to reach and were excluded from the second phase sampling of 2012. The criteria for the exclusion were:

- Points selected in the second phase sample of 2009 that were assessed too difficult to be accessed and were photo interpreted ex ante; the reason for this was that Eurostat in 2009 didn't have at its disposal adequate data sources to assess the accessibility of the points beforehand (small scale elevation models, detailed road network data etc.). The photo-interpretation for these points was done using the most recent available ortho-photos. It would be very unlikely that the land cover and or use have changed due to the remoteness of these areas and hence it would not be economically justifiable to repeat this exercise.
- Points that were far from roads or with a strong elevation change compared to the closest road. The distance and slope thresholds were tuned per NUTS0.

In addition in order to improve the precision of estimates, approximately 40,000 points were added to the field sample. The more varied the land cover is, the more points are needed for precise estimates. The variety was measured by using the Shannon Index (SI) of the transect data in 2009 survey. All the countries which had the SI above the EU average were allocated more points. For Malta and Cyprus all points in the master were surveyed as no stratification results are available for those countries. Most of the points visited in the field in 2008/2009 survey were surveyed also in 2012 survey in order to maintain the panel approach.

The points which were photo-interpreted in the field in 2008/2009 survey due to inaccessibility discovered in the field were replaced by keeping the points in the same strata and NUTS2 area. The accessibility (detailed TeleAtlas road network and slope) were considered in the replacing process to make sure that the points were as accessible as possible.

Points which were verified as inaccessible in 2009 have been excluded from the sample in 2012 and, with the use of auxiliary information related to elevation, slope and closest road distance, other points were included. In addition, the level of elevation, above which points were excluded[8] from field visit, was raised from 1000m to 1500m (i.e. in 2012 the altitude threshold changed to 1500 m).

The table below shows the improvement in the number of observed and photo-interpreted points in 2009 and 2012.

[6] Surveyors are provided with Ground Documents including recent ortho-photos

[7] in the 2009 survey it was agreed to photo- interpret 29,902 points before the field survey due to the accessibility problems (25 % of the points in FI and SE due to the large not easily accessible northern forests and lakes, 10 % of the points in other countries, mainly points in the mountains and large forest areas).

[8] The choice of introducing an elevation threshold was motivated by reducing the costs of point reaching

Table 1: Survey 2012 / 2009: surveyors and points (PI =photo-interpreted)

	Total	Percentage of points per type of observation		
		Ex-ante PI	Observed in the field	PI in the field
Total 2009	234622	12,7	74,6	12,7
Total 2012	270272	0,2	90,1	9,7

However the different thresholds applied and allocation of the sample might have been at the origin of the difficulties to compare estimates from 2009 and 2012. Currently Eurostat is optimising the comparability of the LUCAS 2009 and 2012 results improving the LUCAS 2009 estimation and producing the LUCAS 2012 estimation by proposing proper imputation models for the areas not covered.

2. Updated master grid (1st phase sample)

For the 2015 sampling exercise Eurostat has corrected a number of anomalies in the previous 2-km grid that constitutes the master sample, including points out of the EU area and missing points inside different countries.

In the reviewed list of points some of them have a blank value for the NUTS region in the table of attributes. An overlay on the NUTS 2010 GIS layer at the 1:100k scale confirms that they are indeed out of the EU area and therefore they are removed from the set for the second phase sample. Another set of 1942 points appears in the attribute table of the master grid as belonging to a NUTSX region, but falling outside the NUTSX regions in the NUTS 2010 shape file. These points mainly correspond to **transitional waters** (estuaries, intertidal areas, coastal lagoons, etc.). There has been some discussion on the possible exclusion of these points from the set to be sampled for the field survey. Reasons for exclusion could be:

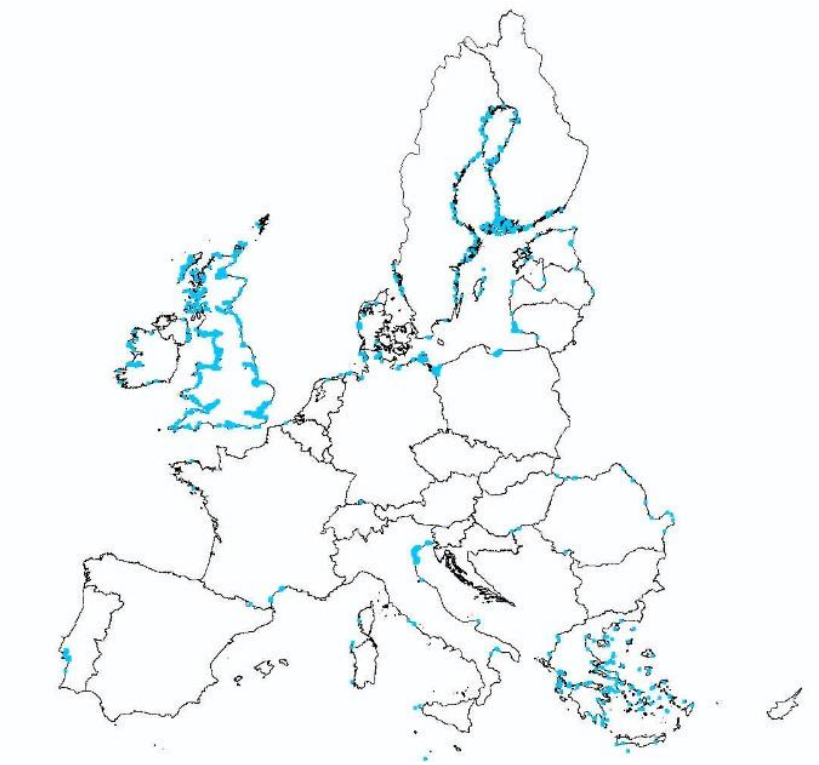
- The combination of different data sources to determine which points are part of transitional waters and/or are associated to a NUTS2 region produces ambiguous results. In fact, by definition, all points in transitional waters should be excluded from the NUTS limits, but in practice this does not occur due to usage of data sources coming from different providers.
- The observation of the distribution of the 1942 points mentioned above on a map (Figure 2) shows that a large number of points in transitional waters appear in the UK, Ireland, Scandinavian countries and Greece, and very few points in Spain, France, Portugal, Italy and Croatia.
- It would be better that the area estimates refer to an officially accepted definition of the territory. This can be the NUTS 2010 boundaries or a further version.
- In general surveyors will not reach the points in transitional waters (except some times in intertidal areas).

The main reason for keeping in the sample points suspected to belong to transitional waters (therefore not belonging inside the NUTSX boundaries by definition) is that there is a request of field information on this category for the validation of maps that include off-boundaries areas. The suggested compromise is that off-boundaries transitional waters are included for the second phase sampling (they will be generally observed from a certain distance), but their weight for the extrapolation should be zero. It might be good to include transitional waters in a photo-interpretation operation for points classified as "non-eligible" because they are difficult to reach. Remains to clarify: which information is pertinent for users, what can be derived by photo-interpretation, and which criteria would be used to define transitional waters outside the NUTS regions, but associated to them. For the final sample 259 points outside the boundaries have been selected.

Table 2: Number of points in the first phase sample.

	Number of points
Total initial	1097607
Allocated to a NUTS region in the attribute table	1093834
Inside NUTS 2010 regions	1091892

Figure 2. Location of the 1942 points that belong to a NUTS region according to the initial attribute table, but fall outside the NUTS 2010 boundaries.



Points excluded for the second phase sample.

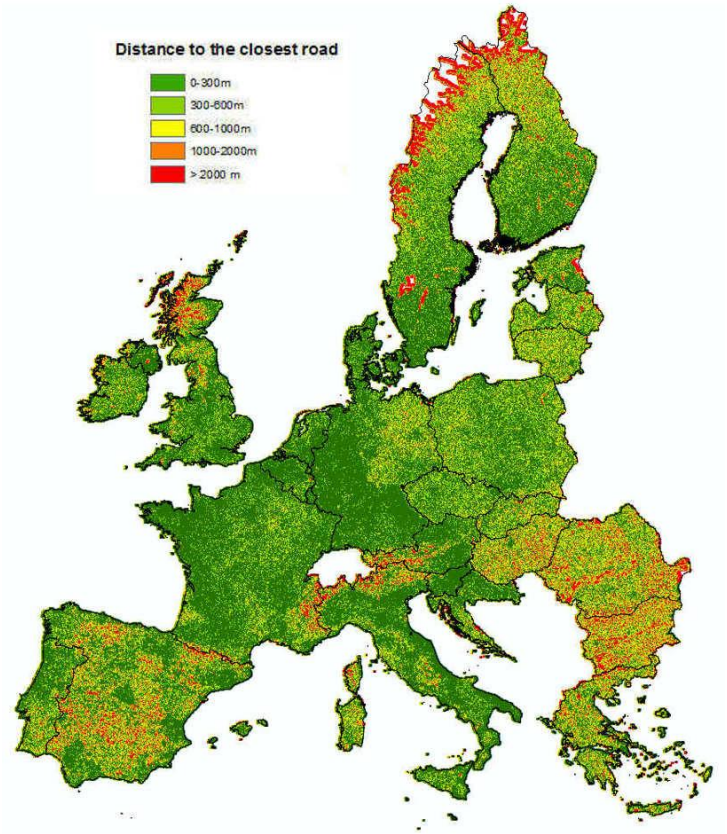
Altitude

A specific category of points that are difficult to reach is the set of points above 1500 m. We have around 22500 points in this class. One way to deal with these points is defining separate strata with them. In 2009 a lower threshold had been used, but the experience in 2012 suggests that points between 1000 and 1500 do not present specific problems to access unless other circumstances appear (to be considered in the next paragraphs). The strata defined by this altitude threshold should be mainly surveyed by photo-interpretation. However points that fall at less than 600 m from the closest road and an altitude difference of less than 100 m are included in the second stage sampling for the field visits. It is clear that these points do not constitute a valid sample for area estimation, but they can provide valuable information for thematic studies.

Distance to roads

The second criterion is the distance to the closest road. The distance has been computed on the basis of Tele-atlas. A visual inspection at the behaviour of this parameter (figure 3) indicates the different density of roads across the EU.

Figure 3: Distance from the first-phase sample to the closest road.



The road network generally excludes rural dirt roads used for the access to agricultural fields, usually good enough to allow the access of enumerators by car. We assume that all points in agricultural landscapes are reachable thanks to these dirt roads, although other obstacles may appear, such as private property delimited by fences.

Accessibility indicator from CORINE Land Cover (CLC).

We define here a very simple indicator that assumes that agricultural areas are rich in drivable dirt roads, in particular where there is a low density of paved roads. There is also an implicit assumption that the density of drivable dirt roads is much lower in other landscape types: forest, shrub, wetland, etc…

We have grouped CLC into two categories: potentially easy and difficult accessibility. Difficult accessibility includes forest, scrub, non-agricultural bare land, wetland and water. We consider that a point in the master frame is potentially difficult to access if all CLC classes 600 m around are in the above mentioned categories. Figure 4 indicates the potentially CLC-derived difficult access areas.

Figure 4: CLC-derived potentially difficult access (in black).

A rule for eligiblility.

We propose a more homogeneous rule for eligibility for the second phase sampling in 2015. The rule combines CLC-based accessibility with distance to roads and altitude. The thresholds proposed are obviously non eligible points are:

a) Points above 1500 m (around 22500 points) and distant>600 m from the closest roads or with an elevation change >100 m from the closest road. 18361points are considered non-eligible with this rule, but almost 90% of them are also non-eligible using the CLC-related rule. Merging both criteria could be a better way to present the eligibility rules.
b) Points below 1500 m with a land cover type neighbourhood (600 m circle) classified as potentially problematic accessibility (forest, shrub, water, wetland) and distant>600 m from the closest roads or with an elevation change >100 m from the closest road. 124191 points in this category.
c) Small islands. At the moment this criterion is not considered because the field (No_island), recovered from the master sample 2012 does not correspond to the set of islands that is considered difficult to reach for the 2015 survey. Points in islands are included in the second

phase sampling. It should be a task of the contractor to propose and agree with Eurostat which points to photo-interpret because the access to the island is too difficult: no regular ferries or too long trip for a small number of points.

d) Points that would have been eligible with the general rules, but could not be reached in 2009 (OBS_TYPE=3 or 4) and were considered non-eligible in 2012.

Categories a) and b) can be merged in a set of strata to be treated with photo-interpretation. Category c) can be added to these strata. A shape file of islands difficult to reach might be asked from contractors on the basis of their experience on the field. Category d) can be treated as missing data in the regular strata until a photo-interpretation is carried out.

With these rules, the master sample is split into approx. 927000 eligible points and 166900 non-eligible points (regardless of the altitude). For the non-eligible points a photo-interpretation operation should be launched. The photo-interpretation rules, including land use and land cover nomenclatures, should be simplified to make them compatible with photo-interpretation.

Table 3: LUCAS 2015 Sample: eligible and non-eligible points for the second phase.

	N points
Eligible	927566
altitude < 1500, CLC-difficult access dist to road > 600 m, difference altitude road> 100 m	124191
altitude < 1500, dist to road > 600 m, difference altitude road> 100 m	18361
Other non-eligible not reached in 2009 (OBS_type 2009 =3) and excluded for this reason in 2012	23716

3 Sampling rates

The sample size per country had been previously published in the call for tender. It would have been possible to tune differently the sample size, but the practical inconveniencies of doing so would have been larger than the hypothetical improvements.

The subsampling method used is a systematic procedure with multiple ranked replicates that ensure a certain spatial homogeneity in the distribution. The sampling rate could have been adjusted separately per domain (NUTS2 x Stratum), but the accuracy targets are rather arbitrary. A reasonable criterion is requesting a CV inversely proportional to the square root of the abundance of each class. This criterion is approximately optimized with a homogeneous sampling rate that has the advantage of simplicity for users (minor impact if users do not use the extrapolation weights).

The only exception made to the homogeneous sampling rate per country is the rule of having a minimum of 2 sample points per stratum in each NUTS 2, unless there are not enough points in the master sample. This rule has introduced 151 points that would not have been selected with the general rule.

4 Some additional modifications to the sample.

Some minor modifications have been introduced on the standard sampling procedure:

- for Cyprus and Malta the full first-stage sample was selected in 2012 and is kept as eligible and sampled for 2015.
- The soil bureau had sampled 24026 points, most of them already surveyed in 2009. This sample includes 919 points that would not have been sampled for the field survey with the general rule, including 317 points that were classified as difficult to reach in the procedure described above. These 919 points have been in the 2015 sample and should receive extrapolation weight=1 in the second phase.

Table 4: LUCAS 2015 - Final sample size and eligible points per country

Country	Total Master sample	Field sample	Eligible	Sampling rate %
AT	20979	6679	15704	42,5
BE	7682	2412	7232	33,4
BG	27741	6623	22696	29,2
CY	2311	1442	1442	100,0
CZ	19718	5492	19195	28,6
DE	89501	24900	85300	29,2
DK	10825	3447	10334	33,4
EE	11354	2255	9594	23,5
EL	33045	7852	24915	31,5
ES	124613	35231	106524	33,1
FI	84542	13407	60302	22,2
FR	137306	38417	125042	30,7
HR	14141	3533	12727	27,8
HU	23271	4626	21429	21,6
IE	17557	3470	15429	22,5
IT	75335	20931	62273	33,6
LT	16334	3873	14875	26,0
LU	646	206	642	32,1
LV	16145	4498	14248	31,6
MT	80	79	80	98,8
NL	8864	2219	8454	26,2
PL	78141	21721	73671	29,5
PT	22261	7318	20542	35,6
RO	59610	14233	51369	27,7
SE	112494	22340	76830	29,1
SI	5067	1614	4705	34,3
SK	12263	2438	10680	22,8
UK	62008	12144	51332	23,7
total EU28	**1093834**	**273400**	**927566**	**29,5**

Figures

Tables

Europe Direct is a service to help you find answers to your questions about the European Union
Free phone number (*): 00 800 6 7 8 9 10 11
(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server http://europa.eu

How to obtain EU publications

Our publications are available from EU Bookshop (http://bookshop.europa.eu),
where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents.
You can obtain their contact details by sending a fax to (352) 29 29-42758.

LB-NA-27596-EN-N

JRC Mission

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new methods, tools and standards, and sharing its know-how with the Member States, the scientific community and international partners.

Serving society
Stimulating innovation
Supporting legislation

doi:10.2788/963

ISBN 978-92-79-54001-1