Guidelines for Labour Market Area delineation process: from definition to dissemination

Version 1.0

Released: August 2017

Authors: Luisa Franconi¹, Daniela Ichim², Michele D’Alò³ and Sandro Cruciani¹

¹ Istat, Italian National Institute of Statistics, Via Balbo 16, 00184, Rome, Italy
² Directorate for environmental and territorial statistics
³ Department for statistical production
³ Directorate for methodology and statistical process design
## Contents

Preface ................................................................................................................................. 3  
1. Introduction .................................................................................................................... 4  
2. Definitions of Labour Market Areas (LMAs) ................................................................. 5  
3. Needs and objectives of LMAs production ...................................................................... 6  
4. Identification of data input: the commuting matrix ........................................................... 8  
5. Analysis of available input data and explorative data analysis ......................................... 9  
6. Identification of method and tools to be used ................................................................. 9  
7. Operational process of LMA development ..................................................................... 10  
8. Dissemination .................................................................................................................. 20  
9. Maintenance of the geography ....................................................................................... 21  
10. Conclusions .................................................................................................................... 21  
References ............................................................................................................................ 22
Preface

These guidelines stem from the experience gained by Istat and the members of the Task Force on “Labour Market Areas Harmonisation” as well as the collaborative work performed during several meetings held between 2016 – 2017. Such collaborative work is one of the most important results achieved by the system of Eurostat grants awarded on the theme of Labour Market Areas (LMAs) both to investigate methodological aspects of LMA design and to produce LMA geographies of participating Members States.

Purpose of this document: This document gives recommendations on the whole process of Labour Market Areas (LMA) design and production. It intends to provide some operational advice on the whole conceptual process of LMAs definition and concrete delineation. The aim of these guidelines is to facilitate the delineation of LMAs in Europe and to foster harmonisation.

Rationale: The process of LMAs design, production and dissemination has been divided into stages. The first part is dedicated to the design process i.e. the conceptual steps to be undertaken before considering any methodological or operational phase of construction. The second part describes the process that uses commuting data to derive the Labour Market Areas and disseminate the geography, as well as statistics and indicators at this level of geography. For each stage, concrete actions are identified and proposed to the users.

Updates: This guidelines reflect the knowledge acquired so far in the LMA development process in the EU National Statistical Institutes. In order to provide coherent, relevant, accurate and accessible information on LMA delineation practices these guidelines will be updated to reflect significant achievements in knowledge and understanding.

Acknowledgements

These guidelines is one of the output of the Eurostat Grant “EU-TTWA: method: improvements, documentation and sharing knowledge activities” awarded to Istat; the authors acknowledge support by Eurostat grant No 08141.2015.001-2015.501. The authors would like to thank Per Kropp, Barbara Schwengler (IAB, Germany), Thomas Thorsen (Statistics Denmark, Denmark), Stephanie Mas (DARES, France), David Levy and Luc Brière (Insee, France), Andy Bates (ONS, UK), Elsa Soares (INE - Statistics Portugal), Pawel Stopiński, Maciej Ryczkowski and Agnes Piorek (Statistics Poland) Valeriya Angelova-Tosheva (Eurostat) for their contributions as well as the other colleagues who have been awarded the other Eurostat grants on this subject contributing in the development of further evidence and specific aspects of Labour Market Areas: Johan Van Der Valk (Statistics Netherlands) for cross border LMAs, Reni Petkova (Statistics Bulgaria) and Yrjö Palttila (Statistics Finland) for the SAS code.

Roma, Italy

L. Franconi
D. Ichim
M. D’Alò
S. Cruciani

---

1 For more information on the grants and the results obtained see the devoted Cros portal Web page: https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en.
1. Introduction

This document is one of the deliverables of the Eurostat Grant “EU-TTWA: method: improvements, documentation and sharing knowledge activities” awarded to Istat; however the recommendations are based on the experiences gained by several Member States between 2014 and 2017 that have performed the delineation of this geography based on the data from their own country. During the meetings held in the last few years methods proposed by other institutions have also been examined and compared leading to a better focus on characteristics needed to define and manage the entire process of LMA delineation and dissemination.

For the derivation of LMAs geographies National Statistical Institutes involved in this system of grants used the method described by Eurostat “Task Force on Harmonised Labour Market Areas” (Eurostat, 2015). Such method is based on Coombes et al. (1986) as modified in Coombes and Bond (2008) and has been used following the recommendation provided in the final report by Coombes et al. (2014).

The implementation of the method into a software was carried out by Istat based on an initial Java code provided by Eurostat and an R (R Development Core Team, 2016) script developed by Statistics Netherlands. Due to the grant, the open source R package LabourMarketAreas has been produced and is now freely available on Cran https://CRAN.R-project.org/package=LabourMarketAreas. Besides the implementation of the rule-based method, for details see Franconi et al. (2016), the R package includes many utilities aiming at helping users to derive and deal with complex objects such as LMAs. As part of the LMA system of grants a SAS interface has also been developed and will be released in the LMA Cros portal Web page: https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en.

Other institutions from Germany, Great Britain and France made experiments using the method and comparing it with other methodologies or software (Kropp and Schwengler, 2016, Coombes and Bond, 2008, Durieux, 2012, Insee, Dares and Datar 2010, Semecurbe and Timoteo, 2013); the NSI of Portugal made comparisons with its previously used method (see for example Soares et al. 2017).

This document deals with the whole LMA delineation process from definition to dissemination adopting an operational point of view. We start by providing definitions and principles for LMAs construction (Section 2) and then tackle the proper production process. Such process is based on the standard business process model adopted in official statistics, the GSBPM (Generic Statistical Business Process Model). This means that the operations start by analysing needs and objectives (Section 3). For LMAs these may vary from simple statistical reporting to more elaborate instruments for placed based policies (Franconi et al., 2017). This stage of the analysis is followed by the identification of the input data (Section 4) available in the country and their characteristics. A preliminary data analysis is necessary to understand the structure of the data and possible issues/problems (Section 5). When the objectives, limitations, data and their characteristics are clear it is possible to select the method and the instruments to be used (Section 6). The operational process to derive the LMA geography is divided into five stages (Section 7). The dissemination phase (Section 8) and the maintenance of the geography (Section 9) end the process.

The proper production process is, as much as possible, written in a general form; the proper operational phase refers to the method proposed by the Task Force.
2. Definitions of Labour Market Areas (LMAs)

Labour Market Areas (LMAs) are generally accepted as areas built on the basis of commuting to work data so that the majority of the labour force lives and works within their boundaries. In order to gain different perspectives on this multi-faceted and complex partition of the territory several definitions of Labour Market Areas are made available below.

1. Labour market areas (LMAs) are sub-regional geographical areas where the bulk of the labour force lives and works, and where establishments can find the largest amount of the labour force necessary to occupy the offered jobs. They respond to the need for meaningfully comparable sub-regional labour market areas for the reporting and analysis of analysis. LMAs are defined on a functional basis, the key criterion being the proportion of commuters who cross the LMA boundary on their way to work.

2. A Labour Market Area is a functional geographic area or region beyond the administrative boundaries defined for purposes of compiling, reporting and evaluating employment, unemployment, workforce availability and related topics. It is an economically integrated spatial unit within which residents can find jobs within a reasonable commuting distance or can change their employment without changing their place of residence.

3. Economically integrated regions within which residents can find jobs in a reasonable commuting distance, or can change their employment without changing their place of residence. They are designed to allow new analyses of socio-demographic characteristics and to overcome the traditional constraints of regional statistics, as administrative boundaries often result from historical circumstances rather than real world social and economic realities.

4. The concept of the labour market area is of an area within which demand and supply for labour meet and fix a price for labour.

5. Labour Market Areas are geographic units of analysis intended to more closely reflect the local economy where people live and work.

We would like to focus attention on some features shared by the above definitions:

- **Sub-regional/local** geographical areas: the local nature of this geography is repeatedly underlined. To achieve meaningful and representative local units, the need to define a minimum size for these areas becomes the main issue. Preferably, such minimum levels should be harmonised.

- **Labour force**: there is the need to clearly define, and possibly harmonise, the labour force. The labour force is represented by the commuters for working reasons. If the source of the data is the census, the data on commuting flows are probably available stemming from the strict definition of commuter provided by the census. On the contrary, if administrative data are used, all the

---


3 Cros Portal: [https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en](https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en)


7 Usually the definition of commuter provided by the census implies going out of the usual residence to reach the same place of work every day and going back the same residence every night. Such definition excludes several types of jobs and provides a specific labour force formed by commuters.
employees/working people are considered. In the latter case the commuting flows need to be derived by linking data present in population registers with data in available in registers of employers with information on employees. It becomes soon obvious that the labour force definition depends on the type of available data source and the constraints imposed.

- **The bulk of the labour force lives and works there**: there is the need to measure the bulk of labour force that remain (live and work) in a given area. The labour force that works in the area may be resident in the area or simply be working there. To measure the closeness of the boundaries w.r.t. such travel to work flows the concept of self-containment is used. Such self-containment is defined both for incoming and outgoing flows; in what follows we refer to the minimum between the two sides. At national level there is the need to agree on the minimum level of self-containment acceptable for each area. Such levels depend on the structure of the flows in the country and may vary a lot from country to country.

- **Economically integrated**: the area is a collection of smaller administrative territories that show some form of internal cohesiveness between them (in terms of flows). This means that the internal flows need to be consistent.

- **The areas are beyond administrative boundaries**: LMAs are functional areas. They are intended to overcome some known drawbacks of administrative regions (mainly based on the limited flexibility over the time of their historically defined border).

- **The areas cover the whole of the country**: LMAs are a partition of the territory of the country or at least of the inhabited part of this territory.

A more structured and comprehensive analysis of these common elements is made by mean of the principles for the construction of LMAs that are thoroughly explained in Coombes et al. (2012). Here we simply report them, see Table 1, with a suggestion of possible actions. For the rest of the guidelines the adopted definition will be the one proposed by Eurostat (3).

### 3. Needs and objectives of LMAs production

The identification and selection of decisional and analytical needs is carried out with internal and external stakeholders. Needs may range from the simple description of the territorial objects to more complex and policy oriented disseminations. In fact, LMAs have long being recognized as relevant for assessing the effectiveness of local policy decisions. Possible objectives when drawing LMAs may range from:

- Identify the territorial boundaries of labour market for planning and analysis purposes;

- Define an economic geography to support policy making (place-based policy);
Table 1: Principles for LMA delineation.

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1. Purpose      | To be statistically defined areas  
• Action: need to be based on sound methodology.                                                                                      |
| 2. Relevance    | Each area to be an identifiable labour market  
• Action: need to reflect what is a labour market in own country (minimum size in terms of commuters living in the area/employees). |
| 3. Comparability| The areas stem from the same set of parameters  
• Action: unique set of parameters for the whole country  
• a replicable process.                                                                                                           |
| 4. Partition    | Every elementary territorial unit to be in one and only one area  
• Action: use a method that generates a partition of the country (not relevant e.g. scarcely populated areas as in the extreme northern part of Europe might be ignored). |
| 5. Contiguity   | Each LMA to be a single contiguous territory  
• Action: necessary to be considered as proper area/local unit; need to check for contiguity of all areas (fine tuning)  
• Definition of minimum contact point  
• Islands: to which territory are to be assigned? According to the situation the best option amongs harbours, bridges, connections (raft, boats), etc. |
| 6. Autonomy     | Self-containment of flows maximised  
• Action: elementary units are assigned to the area with highest level of link. Need to define the linking function. |
| 7. Homogeneity  | LMA’s size not too large in terms of commuters living in the area or in terms of extension of territory  
• Action: essential in order to interpret the territory: maximum size of resident employees not too large. |
| 8. Coherence    | Boundaries to be recognisable  
• Action: minimum administrative units are not broken down.                                                                 |
| 9. Conformity   | Alignment with administrative boundaries is preferable  
• Action: being a functional area LMAs should not respect regional administrative boundaries. An important example is the growing interest in cross-border LMAs. |

Different objectives lead to different dissemination products to be initialized. The delimitation of the LMAs is itself a product to be released that can be helpful and useful for several stakeholders such as ministries, researchers, economists, etc.. In this case the information is directly provided from the geographical description of the partition. If this geography aims also at improving policy actions and decisions, indicators should be produced in order to allow a correct monitoring and assessment of the results of the
actions made. The geography can also be used by the NSIs to investigate socio-economic phenomena or the structure of the dependencies imposed by commuting on different localities.

In this phase, it is essential to plan a systematic

a) updating of this geography and of the related socio-economic indicators
b) monitoring of possible changes in the way commuting data are collected,

Moreover, the usage of other sources of data (registers, big data) together with a better employment of auxiliary information might generate significant quality improvements and changes.

- **Action: Identify needs and define the main objectives. Assign priorities. Divide into short, medium and long-term achievable goals.**

4. **Identification of data input: the commuting matrix**

The LMA is an aggregation of elementary territorial units called communities (LAU2, province, census output areas, etc.). Depending on the level of detail of the communities the granularity of the LMAs will change.

The process of aggregation is made on the basis of travel to work data. Generally travel to work data are either commuting flows stemming from the census data or derived from other typology of source data (e.g. administrative data). In any case the input data is a commuting matrix.

Depending on the algorithm, the matrix might need to be symmetrized or not (i.e. the flows from and to two given communities are summed up and divided by two in order to make them equal). Moreover, it should be specified whether the lack of commuting flows (zero flows) should be included or not into the matrix. The clarification phase should include the following steps:

- Definition and identification of the territorial level of the community
- Definition and identification of the commuters (living in families, coming back home every day, going to the same place every day, etc.);
- Derivation of the commuting matrix taking into account the choices to be made to manage a symmetric matrix, null and negligible flows.

Metadata are needed together with the type of source of the data. Eligibility depends on data availability in different countries.

- **Action: document definitions and the choices made.**
5. Analysis of available input data and explorative data analysis

Currently most MSs still have commuting matrices stemming from traditional censuses. In the future, the implementation of rolling censuses and the more intensive use of administrative data will change the data treatment scenario (uses of weights, changing definitions, etc.). Investigations on these aspects would be appropriate. An initial descriptive analysis of the flows and their sizes depending on the region is to be made routinely. Explorative data analysis of the flows, structure of the territory (distribution of sizes of minimum territorial units), profiling the territory, outliers (travelling times/distances), etc.), identification of specific structures (administrative islands, concentric patterns, etc.) would be targeted to specific characteristics and therefore helpful in understanding features of the LMAs to be built.

- **Action: report on the findings of the explorative analysis of the input data.**

6. Identification of method and tools to be used

In this phase of the process the user on the basis of the objectives and users/stakeholders needs and depending on input data, resources and constraints, choose the method to be adopted and identify the tool to be used. To this end it has to be noticed that the initial report by Coombes et al. (2012) after a long investigation on pros and cons of various methods and after the practical comparisons of the two most promising algorithms recommended the use of the method by Coombes and Bond (2008). Eurostat Task force on LMA proved it possible to use the algorithm as described in the final report (Eurostat, 2015). To this end the R package LabourMarketAreas has been developed and is now freely available on Cran. The package may be downloaded from [https://CRAN.R-project.org/package=LabourMarketAreas](https://CRAN.R-project.org/package=LabourMarketAreas).

During the time of the system of Eurostat grants launched on LMAs several MSs have used their commuting data to derive, by means of the R package, the boundaries of this geography. The method has proved to be flexible and to provide sound results.

For MSs which normally use the SAS language, the translation into SAS is now available on [https://CRAN.R-project.org/package=LabourMarketAreas](https://CRAN.R-project.org/package=LabourMarketAreas). The implementation of the algorithm which is realized in the package LabourMarketAreas is described in details in Ichim et al. (2017) and it is not reported here.

- **Action: Explore the method and the tools to be used for LMAs delineation.**
7. Operational process of LMA development

The proper operational process of LMA delineation is divided into five stages:

1. Prepare the input: standardise input data for the tool and choose parameter values;

2. Run the algorithm: apply the algorithm to the selected data;

3. Validate the solution: select the parameters of the algorithm to produce a solution that is in accordance with expectations, knowledge of specific domains and production of areas with acceptable characteristics;

4. Finalise the output: perform fine-tuning adjustment of the output in order to meet the requirements set at the starting of the process (principles);

5. Dissemination: select and compute the structural statistics on labour market areas. Derive socio-economic indicators to be disseminated at LMA level. The derivation of the latter indicators depends heavily on both the conceptual definition of LMA and on the phenomena under study.

Such general stages are valid for any algorithm; for the scope of these guidelines, we provide indications for the specific case of the algorithm reported in Eurostat (2015) and implemented as described in Franconi et al. (2016). In these guidelines only a broad description of such stages is provided. A more comprehensive and detailed account of the stages and the proper R functions used to carry out the actions needed is the scope of a further document that will be released in the near future.

7.1. Prepare the input

The input of the algorithm is the commuting matrix and the parameters related to size, measured in employment, and level of self-containment of the LMA. It is important to notice that no other input is necessary (e.g. number of inhabitants); so for example if a basic territorial units does not present any commuters it will not be considered in the algorithm.

In this document, self-containment refers to the minimum between supply side and demand side self-containment. To be considered a proper LMA a cluster of communities must have some minimum characteristics in terms of size and self-containment. These features are determined by the researchers in charge with LMA derivation and according to the user needs. Different countries may have acquired different concepts regarding the minimum size of a labour market. Consequently, depending on the density of their population, the territorial morphology and the structure of the commuting, different national contexts may actually need different levels of self-containment to achieve meaningful results.

The algorithm adds the flexibility of allowing the level of self-containment to change according to the size of the cluster so that it can be considered a LMA. This trade-off between size and self-containment is expressed by four parameters, see Table 2 and Figure 1.
Table 2: Parameters that have to be chosen in order to define the main characteristics of the LMA: size in terms of number of commuters and self-containment (intended as the minimum between supply side and demand side self-containment).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>minSZ</td>
<td>Minimum size a cluster in order for it to be considered a LMA.</td>
</tr>
<tr>
<td>tarSC</td>
<td>Level of self-containment which is necessary for a cluster with minimum size in order to be considered a LMA</td>
</tr>
<tr>
<td>tarSZ</td>
<td>Size of a cluster for which the minimum level of self-containment (minSC) is adequate for the cluster to be considered a LMA</td>
</tr>
<tr>
<td>minSC</td>
<td>Minimum level of self-containment for a cluster that has size at least tarSZ to be considered a LMA</td>
</tr>
</tbody>
</table>

Figure 1: Plot of clusters in terms of their size (in thousands of commuters living in the cluster) and minimum self-containment that are considered acceptable LMAs when minSC=0.75, tarSC=0.60, minSZ=1000 and tarSZ=10000.

The data input of the algorithm is the commuting matrix which needs to be prepared in an origin-destination – number of commuters form. The commuting matrix should not be symmetric. In order to reduce the computational time; flows equal to zero are eliminated from the commuting matrix. The R package LabourMarketAreas can deal with community labels which are numeric or alphanumeric. The variables included in the input data may be provided in any order. Instead, their labels are mandatory: “community_live” for the territorial unit representing the origin, “community_work” for the territorial unit representing the destination and “amount” for the numerical quantity representing the number of commuting flows between the origin and the destination. An example of input file is shown in Figure 2.
7.2 Values of the parameters and their interpretation

First of all it is necessary to reflect, for labour market areas in the given country, on the minimum conceivable size in terms of number of resident commuters (i.e. the number of commuters living in the area which represents the resident workforce size). Given substantial differences in the structure of rural versus urban areas, the minimum size has a stronger impact on the former rather than the latter. The minimum size needs also to take into account the size of the building blocks of the LMAs i.e. the communities. If these basic territorial units for which commuting data are available are large areas in terms of resident population or workforce, e.g. the province level, then the minimum size will necessarily be related to the size of the province. The minimum size needs to take into account the minimum size of the community area: there are of course differences whether the input are provinces or census output areas.

Some general comments are always valid: given a commuting matrix, the higher the values of self-containment the larger the LMAs are. In the algorithm implemented in the R package LabourMarketAreas, the self-containment is measured through two parameters tarSC and minSC. Firstly, the parameter tarSC (target self-containment) needs to be greater than minSC (minimum self-containment). Common values for tarSC range between 0.75 and 0.8 but in specific situations, such as in countries with large rural areas and sparse population, higher values maybe needed otherwise each single community would be a LMA. For the parameter minSC possible values usually range between 0.6 and 0.7.

Self-containment values such as 0.75 or 0.6667 can easily be imagined in terms of commuters that remain within the area: for example the former corresponds, broadly speaking, to the fact that the workforce that lives and works in the area is more than 75% of the commuters who only lives in the area and more than 75% of those who only works there.
To analyse the effects of changing one parameter at a time on the resulting dimensions of the LMAs - size in terms of commuters living in the cluster of communities and self-containment - a graphical investigation is proposed, see Figure 3.

**Figure 3: Variation of the parameters of the algorithm: minimum size (minSZ), target self-containment (tarSC), target size (tarSZ) and minimum self-containment (minSC)**

The parameter corresponding to the minimum size of the LMA defines the minimum dimension in terms of number of resident commuters that the aggregation of communities should have in order to be considered an acceptable LMA. In this sense, this parameter has a conceptual value, as it really puts a threshold on the definition of the phenomenon. Its effects are relevant when the size of the cluster is relatively small.

Variation of the parameter corresponding to target self-containment has a strong impact on cluster of communities with relatively small sizes. Its effects tend to disappear when the size of the cluster increases. These first two types of variations have strong impact on small size clusters therefore those cluster that are present in rural or sparsely populated areas.

Variations of the parameter related to the target size allow to accept cluster with the minimum self-containment at the defined value. This means in a sense that the number of commuters living in an area can be associated to a “large” or large enough LMA. Its impact on relatively large size cluster is significant.
The value of the minimum self-containment has a strategic role. It defines the threshold under which a cluster cannot be considered a LMA as it does not retain a large enough self-containment which is its defining characteristics. The variations of this parameter affect mainly large size clusters. This parameter allows for the presence of LMAs around metropolis or large cities that have very high commuting rates and therefore possibly show a network type of structure with nearby communities (for an example see Figure 4 around the area of Milano and large industrialised towns in the north of Italy and Naples). As these last two parameters have an influence on relatively large clusters they are more related to urban areas.

**Figure 4: Demand side (left) and supply side self-containment (right) for LMAs in Italy. Low level of self-containment are mainly present in areas with high commuting rates such as near large towns in the Nord of Italy and around Naples.**

This visual investigation of the effects of the variations of single parameters on the dimension of the LMAs is a simplified one as interactions between parameters are evident. Anyway this kind of analysis gives hint on size and self-containment of clusters that could be classified as LMAs.

In conclusion, the feasibility of some sets of parameters depends on

a) the available data

b) the territorial detail of the communities

c) population characteristics, by rural and urban areas

d) a rough classification/description of the acceptable LMAs

To identify possible values of the input parameters, researchers in charge of the derivation of LMAs might apply some functionalities of the R package LabourMarketAreas: i) apply the clustering algorithm on
smaller datasets that show specific characteristics ii) apply the clustering algorithm using extreme parameter values, etc. These special applications of the algorithm should mainly exclude some parameters value due to their lack of appropriateness with respect to the phenomenon under study.

• Action: Prepare the input data and identify possible values of input parameters to run the algorithm for LMAs delineation.

7.3 Run the algorithm

The definition of the needs and objectives as well as the type of commuting structure present in a country has a significant impact on the range of parameter values that will be investigated in each country. The preparation of a “design of experiment” for parameter selection should be aimed. However, to simplify/sped up the process, it is advisable to identify an initial set of parameters within a defined range of values and run the algorithm by changing only these (and not the whole set of values).

In order to simplify the process, another option is the selection of particular territories that have specific characteristics (eg. extremely dynamic areas, rural areas etc.) and try to optimize the parameters for those areas and set these parameters on the rest of the country to see results. (The third principle on comparability states that a unique set of parameters should be applied for the whole country).

In the R package LabourMarketAreas, the rule-based clustering algorithm should be run using the function “findClusters”. An example of how to apply this function in an R environment is given in the help/documentation included in the R package.

• Action: Run the algorithm for LMAs delineation and get first results.

7.4 Validate the result

Depending on available time and resources, a parameter selection process based on sensitivity analysis is advisable. The quality statistics available from the R package LabourMarketAreas coupled with knowledge of territories and side information on other phenomena (presence of large enterprises, intensity of commuting) would drive the identification of an optimal set of parameters. The translation of the principles listed in Section 2 into statistics and indicators guides the choice by identifying useful characteristics that the final partition should have. Table 3 presents a selection of statistics implemented in the R package and the relationships they have with the corresponding principle. Quality statistics and indicators were derived from the scientific literature. The collaboration of the grant partners was crucial for this development stage.
### Table 3: Principles for LMA delineation.

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description / Action (Quality statistics available in the R package)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Purpose</td>
<td>To be statistically defined areas</td>
</tr>
<tr>
<td></td>
<td>• Action: In order to be a proper LMA the validity needs to be greater than 1. A possible strategy would then be: minimise the number of clusters having a validity less than 1. This information is given by the variable <code>NbClustersValidLess1</code>.</td>
</tr>
<tr>
<td>2. Relevance</td>
<td>Each area to be an identifiable labour market</td>
</tr>
<tr>
<td></td>
<td>• Action: # unique LMA – minimize the number of LMA composed by a unique community. This information is given by the variable <code>NbClusterUniqueCom</code></td>
</tr>
<tr>
<td></td>
<td>• Number of LMA with no Central LMA - minimise the number of clusters with no central community. This information is given by the variable <code>NbClustersNoCentralCom</code></td>
</tr>
<tr>
<td>3. Comparability</td>
<td>The areas stem from the same set of parameters</td>
</tr>
<tr>
<td></td>
<td>• Action: unique set of parameters for the whole country. Do not combine results obtained using different sets of parameters.</td>
</tr>
<tr>
<td>4. Partition</td>
<td>Every elementary territorial unit to be in one and only one area</td>
</tr>
<tr>
<td></td>
<td>• Action: number of LMA check the existence of communities belonging to more than one cluster. The number of clusters (variable <code>NbClusters</code>) should be less than the number of communities. Compute the cross-classification communities – clusters. The corresponding frequencies should all equal one. This information is available through the <code>clusterList</code> component of the output of the <code>findClusters</code> function.</td>
</tr>
<tr>
<td>5. Contiguity</td>
<td>Each LMA to be a single contiguous territory</td>
</tr>
<tr>
<td></td>
<td>• Action: identify the number of non-contiguous territories and analyze them. This analysis is available through the function <code>FindIsolated</code>.</td>
</tr>
<tr>
<td>6. Autonomy</td>
<td>Self-containment of flows maximised</td>
</tr>
<tr>
<td></td>
<td>• Action: analyse the demand-side and supply-side self-containment of the obtained partitions. This analysis is available through descriptive statistics (e.g. mean and standard deviation) and via the internal cohesion flows statistics.</td>
</tr>
<tr>
<td>7. Homogeneity</td>
<td>LMA’s size not too large in terms of resident commuters in the area or in terms of extension of territory</td>
</tr>
<tr>
<td></td>
<td>• Action: compare the descriptive statistics (e.g. min, mean, quantiles) on the residents, workers/job, and resident workers.</td>
</tr>
<tr>
<td>8. Coherence</td>
<td>Boundaries to be recognisable</td>
</tr>
<tr>
<td></td>
<td>• Action: analyse the statistics on the Internal Cohesion Link</td>
</tr>
<tr>
<td>9. Conformity</td>
<td>Alignment with administrative boundaries is preferable</td>
</tr>
<tr>
<td></td>
<td>• Action: count the of LMAs that cross boundaries. This analysis is possible only if information of administrative boundaries is available, too. Generally, this information is available and a frequency table representing, for each LMA, the number of administrative region they belong to performs the desired analysis.</td>
</tr>
</tbody>
</table>
As parameter selection is a complex process. A stepwise procedure can be a possible and feasible way to proceed. The aim is to lead to a shortlist of candidate solutions and to analyse them in full details in order to choose the most promising one.

1) A set of mandatory statistics are calculated for each solution based on what are considered the most important principles;
2) Based on the results of such statistics select a subset of solution to be further analysed;
3) A set of recommended statistics can then be evaluated for each selected solution. Based on the results of this second round a shortlist of solution is determined.

The validation of the solution, i.e. the process that brings to the identification of a set of parameters which is considered optimum for the set of data available is a complex multidimensional problem and many studies are currently being undertaken in order to find the best way to reach this solution. This section of the guidelines will undertake major revision when the results of the studies will be presented. A promising area for improvements is the integration of cohesion measures into the rule that is at the base of the algorithm as suggested in June 2017 at the Workshop in Rome. Further experiments will be implemented.

• Action: Analyse the results obtained and validate the solution that will be adopted. The function CompareLMAStat in the R package LabourMarketAreas provides the input to such analysis

7.5 Finalise the output

The finalisation of the output implies a series of actions that are necessary to disseminate the result. First LMAs need to be labelled. In the R package LabourMarketAreas, LMAs are labelled after the name of the community which gets the highest number of incoming commuters. The function AssignLmaName is useful for this action.

The analysis of the output of the package could reveal clusters which do not reach the minimum size required to define a cluster as a proper LMAs or communities belonging to the reserve list (see Franconi et al. 2016 for details) that where not assigned by the algorithms. All these cases can be classified under the heading of “Self-contained cluster” i.e. a cluster which is completely self-contained (no flows inward and/or outward). This could be the case of a small island or groups of islands that does not reach the given threshold on the size or of remote communities. A manual assignment resolves this type of situations (e.g. islands are assigned to the LMA where the connection with the mainland exists). In the R package this situation is simulated through the function AssignSingleComToSingleLma.

The algorithm by Coombes and Bond (2008) does not take into account in its production process any territorial contiguity principle. Therefore, areas that are non-contiguous might belong to the same LMA. These need to be treated in order to create proper areas. This treatment is called fine-tuning; it treats the non-contiguity by assigning part of the territory chosen by the user to one of the other nearby LMA based on the function of cohesion used (see Franconi et al. 2016). As this process is extremely time consuming it should not be performed on the whole set of possible solutions obtained but only on the final one (or a reduced subset if there is the need to know how different solutions react in different areas).
There are different causes that may create non-contiguities, but some of them cannot be treated as they present structural characteristics of the territory and cannot be solved via algorithms/fine-tuning of the result. An analysis of the various types of non-contiguity that may occur coupled with the corresponding type of treatment is necessary. Table 4 presents an overview of the different types encountered till now and the suggested treatment; Table 5 shows concrete examples of some of these types of non-contiguities as seen in the results obtained by the package LabourMarketAreas run on the demo data Sardinia (available in the R package LabourMarketAreas).

Table 4: Type of non-contiguity and corresponding treatment

<table>
<thead>
<tr>
<th>Type of non-contiguity</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enclave: a community which is not contiguous to the LMA it belongs to</td>
<td>Fine-tuning; assignment of the community to the neighbouring LMA that maximises the link (i.e. the cohesion between the community and the cluster).</td>
</tr>
<tr>
<td>Multiple enclave: a small cluster of communities which is not contiguous to their belonging LMA</td>
<td>Fine-tuning; assignment of the communities to the neighbouring LMA that maximises the link. Currently this will be an iterative process of one enclave at the time starting from the smallest community.</td>
</tr>
<tr>
<td>Exclave: community that is separate from its LMA because it stands in a territory out of the scope of the LMA. This is the case of communities that stand in the territory of another country (e.g. extra-territorial units, e.g. “Campione”, an Italian community that is inside Swiss territory).</td>
<td>This non-contiguity cannot be resolved. No action can be taken as this is a structural enclave.</td>
</tr>
<tr>
<td>Singleton: an LMA that comprises of a single community</td>
<td>A decision needs to be taken whereas this type of LMA is a proper LMA. As a labour market is generally seen as a network of communities, fine tuning is applied and the single community is assigned to another LMA among those that are contiguous (by maximising the cohesion function).</td>
</tr>
<tr>
<td>Administrative island: Part/parts of the territory of a community which are not contiguous to the territory of the community itself.</td>
<td>This non-contiguity cannot be resolved. No action can be taken as this is a structural enclave.</td>
</tr>
</tbody>
</table>
Table 5: Examples of types of non-contiguity

**Enclave**: The left hand side of the figure shows the communities that constitute the LMA of Carbonia. Community 92007 is separate from the rest of its LMA; it is highlighted in yellow in the figure on the right.

**Multiple Enclave**: the LMA of Bono presents two communities (90020 and 90028) that are non contiguous. These communities need to be reassigned by fine tuning.

**Administrative island**: the figure on the right shows the communities and the other territories (administrative islands) that belong to the municipalities in the LMA Lanusei although they are not contiguous to their mainland. No action can be taken in this case.

**Administrative island**: even very small islands (therefore not contiguous to their main territory of the community) can be assimilated to the case of administrative island (as in the LMA of Oristano).
In the R package LabourMarketAreas, the fine tuning process is implemented via three functions: FindIsolated, FindContig and FineTuning.

- **Action:** Check for non-contiguity in LMAs and check the assignments of all communities to proper LMAs.

8. **Dissemination**

The dissemination manages the release of the statistical products to customers according to the objectives set at the origin of the process. We do not discuss here the routine operations (formatting data and metadata, loading them into output databases etc.) rather the specific types of outputs for LMAs to meet user needs.

The basic standard dissemination of LMAs implies the identification of their territorial boundaries. This is done by releasing the list of their composition in terms of elementary administrative units, along with the geographical shape files needed for cartographic and graphical purposes. In the R package, the LMA geographical shape files may be derived using the function CreateLmaShape. A graphical example is given in Figure 5.

**Figure 5:** Communities and LMA obtained using Sardinia data in the R package LabourMarketAreas. Fine tuning not performed for sake of illustration.

As for any other territorial object, statistics on LMA main characteristics in terms of population, density of inhabitants, employees, number of enterprises etc. should also be foreseen.

The release of all the evaluated statistics (quality statistics which are the output of the LabourMarketAreas package) of the delineation process is also advisable, as such statistics characterize the structure of this geography.
For transparency reasons the release all the information on the LMA considered not acceptable and reallocated and the communities involved in the fine tuning process is also advisable.

A more complex dissemination is connected with a more sophisticated use of LMAs. This involves the definition and estimation of specific labour related indicators for policy definition and assessment purposes. An advanced release may involve designing systems of indicators with their calendar and specific web visualization and query tools to reach a large number of users and stakeholders.

Also internal use might be investigated; LMA are territorial aggregation that might be useful when defining domain in sample surveys than the more common administrative areas.

The answer to customer queries and requests for services might indicate new needs emerging from users in front of a new geography.

Comparisons with previous instances of the geography maintaining the same parameters might give insights on regions where changes are concentrated and the general trend of the commuting phenomenon.

**Action:** Produce the products identified in the objectives and monitor user needs.

9. Maintenance of the geography

Changes of the basic territorial units, the communities, such as unions, cancellations, changes in the territory of the municipalities may sometime happen. In cases of geographical changes of the basic territorial units for LMAs, for continuation of LMA statistical production, the affected LMAs may need to be updated. Update is necessary every time that a geographical change affects the principle that states that every community belongs only to one LMA. For instance it can happen that a newly created municipality is divided into two different LMAs due to an administrative boundaries' change. Then a procedure is put in place (a sort of fine tuning) in order to assign the new community to one of the two LMAs by means of the same process that has led to the definition of the current setting. This means that any update of the geography is based on the commuting flows.

**Action:** Update the geography of LMAs; in case of severe changes (new municipalities belong to two different LMAs) reassign them according to the linking function adopted.

10. Conclusions

The harmonisation process for LMAs delineation does not mean sharing the same parameters rather sharing a common way to proceed and harmonised methods. Further studies will provide answers on the characteristics for LMAs to be used at EU level and on ways to combine geographies from different countries such as the case of regions with significant cross border commuting. Initial steps have been made so far which brought to concrete cross border LMAs between The Netherlands, Germany and Belgium; such experiments have highlighted possible improvements to be made.

Further studies are also needed to improve the homogeneity of large capital cities when there is the need to divide a large LMA them into smaller pieces that need to maintain specific functional characteristics.
existence of a group of active researchers that continue to investigate these matters to provide solutions shows that there is the concrete possibility to create an LMA geography that covers the whole of the EU.

References


