

## **4. Annex I - Interim Research Report**

## TABLE OF CONTENTS

EXECUTIVE SUMMARY .....	70
INTRODUCTION .....	72
CHAPTER 1: REVIEW OF THE STATE-OF-ART OF APPLIED SCIENCES IN THE FIELD OF LMAS .....	75
1.1 Policy Context.....	75
1.2 Scientific context .....	76
1.3 Geographic foundations.....	78
1.4 Moving forward .....	81
CHAPTER 2: OVERVIEW OF THE INFORMATION COLLECTED .....	88
2.1 NSIs' position towards an LMA geography and their expectations from this.....	88
2.2 Inputs for the delineation of LMAs and characterisation of LMAs .....	91
CHAPTER 3: CROSS-NATIONAL EVALUATION OF EXISTING CONCEPTS OF LMAs .....	95
3.1 Principles in the definition of LMAs .....	95
3.2 Cross-national evaluation of national methods of defining LMAs .....	102
3.3 Towards an empirical test of 'best practice' in defining LMAs .....	107
ANNEX 1: COUNTRY NOTES ON DELINEATION METHODOLOGY .....	110
ANNEX 2: TEMPLATE QUESTIONNAIRE ON LMAs.....	139

## EXECUTIVE SUMMARY

This is the report on progress at an approximate mid-way point in a study to consider the potential for identifying a common European way of defining the boundaries of Local Labour Market Areas (LMAs).

The study has so far carried out four main activities.

- Outline the state-of-the-art in applied social science research in the definition of LMAs
- Compiled a preliminary inventory of necessary information and of information sources
- Compared national methods of defining LMAs, based on a survey of all Member States (MSs)
- Considered the implications in terms of establishing best practice in the definition of LMAs

The key purpose of this Interim Research Report is to inform stake-holders in all MSs of progress in the study and also to provide an early sight of emerging best practice in the definition of LMAs. It is hoped that as a result of this information sharing, MS stake-holders will respond to the invitation for feed-back on the report, and perhaps engage with a later stage of the study (as outlined at the end of the report).

The first substantive research activity was a comprehensive desk-based overview of relevant literature on appropriate methods for defining the boundaries of policy-relevant LMAs. The diversity of academic work which had to be reviewed hinted at the lack of a very strong convergence on agreed best practice. There were also new approaches being developed, as part of the emerging field of spatial economics and its analysis of large datasets.

All the same, there was no evidence any serious contradiction of the principles of LMA definitions that will be policy-relevant, as set out in “*Study on employment zones*” (EUROSTAT Document E/LOC/20).

As a result, the state of the art review was able to build on these principles, combining them with other concerns evident arising from the newer and/or more academic literature. This led to enumerating a final total of 14 issues against which methods of LMA definitions can be compared.

In addition to the review of LMAs in the applied social sciences, this report also considers the information provided by national statistics institutes (NSIs) in reply to a tailor-designed questionnaire on LMAs, which was aimed to compile homogeneous information on this issue. The research team is extremely grateful for the information that was provided primarily by the respondents at the NSIs and, in some cases, by the academics that co-operated with them in those tasks. The answers received show that despite an obvious disparity in the national approaches to the issue of delineating LMAs, a vast majority of the MSs rely on the use of data on commuting flows from a variety of statistical sources. The questionnaire investigated the availability of such data in all EU MSs, including those countries where LMAs have not been defined until now. As a preliminary conclusion, the decennial Census of Population and Housing seems to be the preferable source of commuting datasets. The responses to the questionnaire confirmed for the majority of countries that after the 2011 wave it will be possible to

gather the commuting data that would permit the definition of LMAs in the diverse parts of the EU. This dataset will be available at LAU2 level for most MSs, and this could allow the characterisation of the sets of LMAs built over them through aggregation of data, something that the questionnaire confirmed as common practice in those countries where LMAs are in use.

The last section of this report compares existing national LMA definitions against the 14 relevant issues established in the course of the first research activity. This examination embraced both official and academic approaches included in the NSIs' responses to the questionnaire, and aimed at selecting the methods that could be then explored in the empirical analyses that will be conducted in the last stage of the project. The choice was guided first by recognising that a small number of issues are of primary concern. It was also argued that there were some rather different types of method, with the proposal emerging that it would be valuable to test one method that was based on the initial selection of centres, and another one that explores the whole dataset with an 'open system' approach that avoids imposing such a structure. The other factor taken into account was that methods whose results were adopted by the respective NSIs can be seen as thereby having had a positive endorsement. The combination of these factors led to the selection of the **UK** and **SE** methods as the recommended candidates' for the empirical research as respectively a more 'open system' method and one heavily based on a preliminary step of identifying core [urban] areas. There is also a case for a more 'academic' stochastic procedure (which performs simultaneously local and global optimisation of a within-region interaction index) as an appropriate comparator to the **SE** and **UK** methods already mentioned.

This forward look towards the empirical research activity later in the study was completed by some thoughts of quantitative indicators for the evaluation of the results of different methods applied to data for several MSs. A remaining task is to select MSs whose datasets it will be most valuable to analyse, and some factors relevant to this selection are suggested.

The remaining stages of the study will move on from the four essentially retrospective tasks described here to look forward at [a] the possible value added by having consistent EU-wide LMA definitions of LMAs, and also [b] the potential for identifying a single harmonised definition method that can be seen to delimit suitable LMA boundaries in diverse parts of the EU.

## INTRODUCTION

### Overview of the project

In 2007, when the implementation of the NUTS Regulation was reviewed by Commission Regulation (EC) No 105/2007, National Statistical Institutes (NSIs) proposed an investigation of alternative classifications to the administrative levels below NUTS for the EU management of territory. As a result, Eurostat committed itself to exploring functional regions and the possible application of that concept to the entire EU. The first step was then to investigate with the help of the research community the potential value-added and feasibility of, and best practice for, a consistent EU-wide definition of labour market areas.

The report that follows is the intermediate output from research activities undertaken by, and on behalf of, Eurostat in this direction. It is a result of the work of Eurostat Unit E4, coordinated by Mr. Oliver Heiden, and the external contractor DevStat – Servicios de Consultoría Estadística with its associated researchers, Prof. Mike Coombes from Newcastle University, and Prof. José Manuel Casado and Dr. Lucas Martínez from the University of Alicante.

Labour Market Areas (LMAs) represent a territorial breakdown which is seen as a valuable alternative to local and regional administrative areas for some statistical purposes, and for the design, implementation and monitoring of labour market and other public policies. These functional territories have gained importance for policy-makers, mostly during the last decade, because they provide a territorial grid with boundaries that do not derive from historical events or geographical factors, but they reflect the organisation of social and economic relations in each specific area.

Currently there is no common European definition of LMAs, although in several EU countries they have been defined and are in use. The purpose of the “Study on comparable Labour Market Areas” is to explore the possibility of a consistent statistical classification of the whole EU territory, defined on a functional basis. To be specific, the objectives are to: (1) outline the state-of-art of applied sciences in the field of LMAs; (2) compare the LMA concepts recognised and implemented in each Member State; (3) draw conclusions on relevant best practice; (4) explore the added value of a common definition for the entire EU; and (5) identify possible ways and means of harmonising LMA definitions across the EU.

What makes this study different from the previous ones<sup>1</sup> is the intention to cover all the territory of EU 27 Member States (MSs) while also empirically testing a proposed method for the delineation of LMAs so as to formulate proposals for a possible EU wide harmonised grid of comparable LMAs. This empirical research is to reflect the assessment of LMAs’ potential applications in the policy fields of the EC, but

---

<sup>1</sup> The topic was previously investigated by other international organisations and DGs of the European Commission (EC). Two decades ago, Eurostat and Newcastle University carried out a study of principles upon which to base definitions of LMAs to be used in a policy context. In 2001, OECD undertook a cross-national survey in order to examine the relevance of functional delineation of regions on the basis of travel-to-work. OECD’s study offers a clear view of different existing concepts of LMA in different states, although it does not include all EU countries. More recently, DG REGIO expressed a strong interest in LMA statistics and Eurostat carried out a survey of the LMA definitions in Member States.

will also rely on the involvement of MSs in the collection of comparable information and the evaluation of intermediate results of the study.

### **Purpose of the Interim Research Report**

This interim report details the progress of the work undertaken for the possible definition of a common EU way of defining LMAs. It is intended to provide the main stake-holders, Eurostat and NSIs, with a comparable view of the situation of LMAs within the EU.

The intention to undertake this research was communicated to NSIs at the Working Group on Regional Statistics and Rural Development (4-5 October 2011), when NSIs agreed to provide information for the study.

The report considers LMAs from both theoretical and practical perspectives by reviewing the state-of-art in applied social sciences, but also comparing existing national LMA definitions. It is intended to provide both experience and non-experienced users with the overall result of the research undertaken so far. In so doing, it aims to inform NSIs about the possibilities of a consistent LMA geography and to seek their views.

### **Structure of the Interim Research Report**

Apart from this introductory section, the report includes the following chapters:

- **Chapter 1:** Overview of the state-of-art of applied sciences in the field of LMA

This chapter identifies the key features of best practice emerging from the review of academic work on methods for defining LMAs and ends by proposing a set of principles to consider for the evaluation of LMA definition methods and their results. The 14 principles drafted here derive from the definitional criteria proposed in the earlier study undertaken by Eurostat, Coombes (1992), extending their scope so as to cover in more detail the technicalities of the methods of delineation and to enable not only the comparison but also the evaluation of existing LMA definition methods.

- **Chapter 2:** Overview of the information collected

This chapter gives an overview of the information collected through the 2012 questionnaire on Labour Market Areas sent out to EU Member States in January - February 2012.

- **Chapter 3:** Cross-national evaluation of existing concepts of LMA

This chapter analyses the existing national methods of LMAs definitions as implemented in Member States, and provides a detailed comparison of the methods using the principles identified in Chapter 1. The cross-national evaluation of existing concepts of LMA ends by drawing some conclusions in terms of identifiable best practices in the definition of LMAs.

- **Annexes**

The report is accompanied by the detailed description of the national methods analysed and compared in Chapter 3 (for Belgium, Czech Republic, Estonia, Greece, Finland, France, Italy, Portugal, Sweden, Slovenia, Slovak Republic, and United Kingdom) and by the standard format of the 2012 questionnaire on LMAs.

## CHAPTER 1: REVIEW OF THE STATE-OF-ART OF APPLIED SCIENCES IN THE FIELD OF LMAS

This study was motivated by the fact that there is no common European way of defining Local Labour Market Areas (LMAs). In this Activity there has been a review, based on desk study, of the present state of the art in the applied social sciences in methods of defining LMAs. The concept of the labour market area relates to the ‘law of one price’ in that it is an area within which demand and supply for labour meet and fix a price for labour. Although a few labour market segments have national or even international markets, for most forms of labour national markets are fragmented into different LMAs. This spatiality of labour markets is part of the functional reality within modern economies, and hence LMAs are a specific aspect of the functional regions that are increasingly recognised across Europe.

The first section of this chapter outlines the scope of the desk study reported here. The two main sections of this chapter then identify in the applied sciences *key features of best practice* emerging from the review of academic work on methods for defining LMAs. Finally there is a section that looks forward to the next Activities of this study, ending with suggested issues to consider in evaluating LMA definition methods and their results.

### 1.1 Policy Context

It has become increasingly common in European and other developed countries for LMAs to be defined for policy delivery or analysis, or the publication of data by national statistical institutes (Cattan 2001). This trend is notable because there is an inherent tendency to use local or regional administrative areas for official purposes, so considerable benefits must result from using LMAs for them to be used instead. The fundamental reasons lie in the key policy focus on sub-national contrasts in economic geography, with the recognition that addressing these contrasts coherently calls for analyses that are comparing *areas which represent labour markets*, because this is one of the key concepts in economic geography (Combes et al 2005).

Administrative boundaries do not very frequently match the functional realities of LMAs due to being the product of tradition and/or topographic factors that have become less relevant (eg. rivers that once were natural barriers but, when bridged, become the focus for economic development). Even when functional realities had informed the drawing of an administrative boundary (Andersen 2002), this boundary may have remained unchanged for so long a period that it no longer relates to changed economic patterns.

Although some administrative areas may approximate LMAs, this is not consistently true either within countries or, more obviously, between them (Forstall et al 2009). The importance of this inconsistency stems from the fact that *policy analysis requires all the LMAs to be defined in a comparable way* so that data for the areas can be used with a minimum of anxiety about how far the way the boundaries were drawn affects comparisons between areas (eg. when identifying areas in most need of policy support).



In short, the choice of areas in analyses raises the familiar risk of ‘comparing apples with pears’ but the general issue is in fact unavoidable: in spatial analyses it is termed the modifiable areal unit problem (Openshaw & Taylor 1981) and is a key concern when analysing economic data for administrative areas (Mitchell & Watts 2010).

European integration has led to the need for functionally coherent LMAs that are comparable not only within but also *between* EU member countries. This challenge calls for comparable data for comparable ‘building block’ areas to be analysed, ideally delineating LMAs in a single process across all countries. One advance would be an empirical evaluation of LMA definition methods by applying them to data from different countries: if such analyses prove possible they can be attempted later in this study. This chapter is a first step in that direction, because it aims to establish some best practice guidance that can inform the search for appropriate methods to devise ‘European standard’ LMA boundary definitions.

## 1.2 Scientific context

Scientific research into LMA definitions has not led to consensus on best practice (Schubert *et al* 1987), in part because of the need for an appropriate method for evaluating different methods. This need will be addressed in the last section of the present chapter, but the focus here is first on teasing out lessons of best practice emerging from the academic LMA definition literature (nb. a later Activity of the study will examine official definitions of LMAs in the countries of Europe).

Delineating sets of LMA boundaries is a specific form of partition (if done ‘top down’) or of grouping (‘bottom up’), and hence is a type of taxonomic analysis. In principle, LMAs could be defined ‘top down’ but nearly all methods in scientific literatures are *based on grouping building-block areas (‘bottom up’)* by analysis of relevant data. The fact that flows are key defining features of functional regions has resulted in most *definition methods analysing patterns of commuting* because they are flows that are an aspect of the labour market through which the strength of the links between building-block areas can be measured. There are some alternative data sources relevant to the labour market which can show spatial patterns of workplace and home locations but the coverage of such data is limited and liable to bias: for example the movement of people into employment – ‘hirings’ – only covers new employees and so has a bias towards young people among others.

This study is concerned with methods to identify LMAs covering the whole European space and this can be seen as a form of taxonomy. Some taxonomic analysis methods such as cluster analysis are quite familiar but are unsuited for LMA definition because they do not readily handle the matrices on flows between areas that are understood to characterise LMAs. In addition, they do not readily handle issues of area topology and the particular *need to avoid non-contiguous LMAs* (LMAs comprising two or more bounded areas that are separated by other LMAs) as illustrated by results of Hensen & Cörvers (2003). In geographic research, cluster analyses readily produce what are termed *formal regions* (eg. groupings of textile towns), whilst the contiguous groupings which include LMAs are termed *functional regions* (Spence & Taylor 1970).

Before reviewing the literature specifically on LMA definitions it is important to recognise that other scientific literature can be at least indirectly relevant here. There is increasing interest in spatial issues within economics, aided by the diffusion of geographic information system techniques which make large spatially-referenced datasets more readily interrogated. In the growing field of spatial economics the ‘discovery’ of computer-assisted geographic analyses has led to exploratory work which is in effect looking for LMAs in data ‘attribute space’ (eg. Marcon & Puech 2003; Duranton & Overman 2005). These new approaches can be compared with those used when taxonomic definitions were first being computerised and methods were derived from general principles in mathematics and related fields (Sokal & Sneath 1963). For example, there was early work using graph theory Nystuen & Dacey (1961) which is now echoed by recent exploratory adaptations of social network analysis (Newman & Girvan 2004; Green 2007).

Many past studies defined LMAs with the argument that commuting patterns can be ‘proxies’ of the other forms of spatial interactions which would be part of a more ideal definition of functional regions (Sohn 2005). A question for this study is whether this logic might work in reverse: if there is no up-to-date commuting dataset available then might LMAs be defined with other interaction data? The dataset which is the nearest equivalent to commuting and is quite often available to the necessary level of reliability for all small building-block areas is migration. A relatively high proportion of the early functional region definitions relied on migration data (eg. Hemmasi 1980) but it has been less used recently and some work in England suggests that migration patterns in many regions differ markedly from those of commuting (Hincks & Wong 2010) so are not a plausible basis for defining LMAs.

Still at the experimental stage are studies attempting to define what are, in effect, functional regions based on analyses of mobile phone traffic (eg. Candia et al 2008) or even surname incidence data (Longley et al 2011). The former approach is analysing a form of interaction that may – but may not – mirror commuting patterns, whereas the latter uses non-flow data to surmise a pattern of interaction (viz. longer-term migration). The latter approach is slightly more familiar in modelling strategies used when there is no interaction dataset but some data on transport infrastructure or services. Hugo (2001) uses road networks to measure the relative ease by which more rural areas can access urban centres, rather as Green (1950) analysed bus service information to identify urban centres and their hinterlands, an approach which may still be relevant in countries where public transport is still key to mobility. In the absence of even these datasets it may be possible to predict patterns of commuting from data on jobs and employed residents: for example, Glover & Openshaw (1995) offer a version of the well established gravity model, while a simpler option assuming no “wasteful commuting” (Small & Song, 1992) has been illustrated by Coombes (2004).

With computational power continually growing there is also interest in the transferability of techniques such as location-allocation analysis to region definition (eg. Lolonis & Armstrong 1993), along with the automatic zoning program of Openshaw & Rao (1995). In recent years the challenge of distilling the

patterns from huge datasets is an increasingly active research frontier due to the ever growing volumes of information available to scientific enquiry. Indeed new methods of taxonomy may emerge in computer science, for example, which are transferable to the definition of LMAs. One example of a more general-purpose computational technique with possible application to LMA delineation is simulated annealing (Kirkpatrick et al 1984), which can be set in the broader field of evolutionary computation (Fogel 2006). Exploratory research (Flórez et al 2008) suggests that such methods can be adapted to delineate LMAs by formulating objective functions to, for example, maximise cohesiveness in terms of commuting flows while requiring that minimum levels of self-containment and size are respected.

The potential scientific advantages of these approaches – such as their inherent rigour, potential self-optimisation and replicability – tends to be off-set by their rather abstract nature and risk of being seen as a ‘black box’ when policy-makers prefer readily understood definitions. This means that *transparency* in the definition process is important, so that it can be seen why any particular area has been given the boundary the analysis selected (nb. one problem with some computationally-intensive methods is that they are not deterministic, which means that different ‘runs’ of the same method on the same data can produce different results).

### **1.3 Geographic foundations**

Defining functional regions such as LMAs builds on earlier scientific analyses in economic geography, with fundamental concepts dating back over a century. Christaller (1933) elaborated the *central place* model that saw whole territories orientated around urban centres, with southern Germany used as the possible exemplar. In the UK a more empirical tradition saw Mackinder (1902) identify economically integrated urbanised regions that were later termed *city regions* by Geddes (1915). Later technical change and growing prosperity made people more mobile, allowing more distant places to be linked with cities in particular, thus creating wider *urban systems* (see for example Pahl 1965).

In the closely-spaced urban systems found in much of northern Europe especially there are numerous examples of previously distinctive local economies being characterised as emerging *polycentric regions* (eg Burger et al 2011). In fact van der Laan & Schalke (2001) argued that LMA definitions would more fully reflect the complexities of modern patterns if they were *over-lapping* in many cases, but there are very few such definitions. In fact over-lapping LMAs would be outside the concerns of this study whose policy focus requires a set of LMAs covering the whole European space but, at the same time, *each building-block area should be in one, and only one, of the defined LMAs.*

Goodman (1970) provided foundations for subsequent research on defining LMAs by recognising the value of commuting flows to LMA definitions and identifying as the two essential requirements of LMAs:

- (1) their boundaries are crossed by few journeys to work (ie. they are relatively self-contained)
- (2) a relatively high level of intra-market movement results from the LMAs being as integrated as possible.

These two key factors have remained key foundations for LMA definitions subsequently. All the same, the lengthening and diffusion<sup>2</sup> of commuting flows makes meeting both requirements more difficult. Goodman (1970) warned against “the danger of seeking external perfection at the expense of losing the essentially *local* character of the market” – which implies priority of the second requirement over the first – but the lengthening of average commuting trips in the 40 years since then has left some areas as not locally integrated at all (see for example, van Nuffel 2007).

Turning to existing LMA definitional practice, there is in fact no off-the-shelf measure of integration, whereas the self-containment measure is well established. As a result, best practice methods are characterised by the analysis requiring a minimum level of commuting self-containment for each LMA. Although there may well also be an aspiration to maximise integration, few if any definition methods set a specific level of integration as a constraint.

Smart (1974) set out to “produce a systematic definition of areas by which the main relationships between homes and workplaces could be indicated” and his pioneering efforts were followed by a novel computerised approach that brought scientific rigour to the definitions of TTWAs (Travel-to-Work Areas), the official UK LMAs (Coombes et al 1986). As often, this innovation relied on newly accessible relevant data being available for sufficiently small areas.

A key argument for analysing commuting data to define LMA boundaries is that the LMAs are where “jobs are sought and job decisions are made... in terms of ability to commute” (Wilcock & Sobel 1958). A focus on commuting may seem old-fashioned in a world where new means of communication allow many jobs to be done without the need for people to travel, but in fact *teleworking* displaced relatively few physical work trips so that “spatial patterns of commuting are more complex today than in previous decades, but no less important” (Arbuckle 1998).

In fact a secondary but crucial advantage of the usual dependence on commuting data in defining LMAs is that the ‘friction of distance’ that restricts people’s patterns of movement causes most of the strongest interactions to be between nearby areas. This means that contiguous groupings of areas are *inherently* the most likely to be produced from commuting data: for example, less than 1% of building-block areas were allocated non-contiguously in an application of the unconstrained TTWA method to New

---

<sup>2</sup> The strong tendency for personal mobility to increase was enabled by several factors, most notably

- rising average earnings, making greater commuting costs more affordable
- increased car use, enabling more diffused and distended commuting patterns, and also
- decreasing real cost of car use.

Zealand (Papps & Newell 2002). By contrast, any methods that use an explicit contiguity constraint at every step of the analysis will see their results shaped by “irregular base areas” (Spence & Taylor 1970).

Contiguity constraints were once valued for vastly speeding up analyses – by greatly restricting the options considered as building-block areas are being grouped – but now vast numbers of permutations can be evaluated quickly. Thus little is gained from if a contiguity constraint is imposed throughout the analysis but, if one is imposed, it will reduce the options available to the analyses and this will inevitably risk creating some sub-optimal boundary definitions (Roca & Moix 2005). The clear implication is that methods whose processes are not limited by contiguity constraints are preferable in most cases.

Another distinction which can be drawn is between those methods that use one ‘rule’ from start to finish, and those which apply different rules at different stages of the process. The former type proceeds until all the LMAs satisfy a criterion that decides when the procedure stops (e.g. the population size of the smallest remaining region is large enough). Several such methods were developed in the 1970s at the time when computerised matrices first became available (eg. Slater & Winchester 1978), while some may still be sometimes used, as for example INTRAMAX which was created by Masser & Scheurwater (1980) and used recently by Mitchell & Watts (2010). The new ‘black box’ methods mentioned earlier mostly involve multiple applications of a single rule.

Methods with numerous rules, by contrast, often identify LMAs in accordance with a geographical model (eg. a first rule uses a size criterion to find urban centres, then a later step analyses commuting patterns to group non-central areas as part of the LMAs around centres). As this implies, a single or multiple step method is usually adopted as a direct consequence of the *general* approach being taken towards the definition of LMAs, so past research does not readily indicate whether a single or a multiple rule method is the better practice in general. That said, if all ‘other things are equal’ then single step options gain from their advantage of having fewer parameters to justify and to then potentially have to put through sensitivity testing (cf. Papps & Newell 2002).

The single rule approach often repeatedly applies its single criterion within a hierarchical process and this can create sub-optimal results at larger scales. This is because the area groupings made at the early stages of the analysis inherently restrict the options available<sup>3</sup> at later stages. The solution is for methods to ‘escape’ from being hierarchical, although no ‘universal’ procedures exist to achieve this. The method to define TTWAs has since the 1980s used a procedure that aimed at self-optimisation

---

*These processes have reinforced each other over recent decades, increasing the proportion of people who commute longer distances and thus increasing the integration of previously separate local areas.*

<sup>3</sup> For example, a grouping of European countries might be expected to link Luxembourg with Belgium and then with the Netherlands in its early stages; in a hierarchical procedure, these early groupings would then prevent the later stages creating what may be the more optimal broader groupings in which, for example, the Netherlands could be linked with other northern countries which speak Germanic languages whilst Belgium and Luxembourg were grouped with France and more southern countries.



(Casado-Díaz 2000): this procedure became the basis of the entire process in the latest version of this LMA definition method (Coombes 2010).

An important distinction that van der Laan & Schalke (2001) made among LMA definition methods was between *deductive* and *inductive* based approaches. In practice, deductive methods begin with the identification of urban centres around which the LMAs are constructed; by contrast inductive methods have no such pre-conceived structure to their definitions. Most deductive methods rely on the familiar notion of a single urban centre which attracts commuters from surrounding areas (eg. Arbuckle 1998), and this is readily understood by non-experts. However this model pre-dates the growth of polycentric urban systems, and decentralising employment to less urban centres. In fact many versions of the inductive approach exclude more remote areas from their results (eg. Cheshire & Gornostaeva 2002), an approach running counter to the all-inclusive objective of this study. The familiar deductive approach of metropolitan definitions has been seen as an ‘urban bias’ by those interested in rural development (Killian & Tolbert 1993).

The inductive approach, not surprisingly, has the converse strengths and weaknesses; it copes flexibly with diverse commuting patterns in different times and places, but this same flexibility means it does not have a simple form which is easily recognisable. The process is one of gradual integration, so that each step in the process only makes a small contribution to the final result. It reflects a less rigid conception of LMAs as clusters of commuter flows, within a wider ‘space of flows’ (cf. Castells 1989). In effect then, the choice between deductive and inductive approaches depends upon the evaluation criteria adopted. The search here is for LMA definition methods to produce adequate results in many different conditions across the European space and this leads to an emphasis on flexibility and so the inductive approach may be favoured. By contrast, the intuitive appeal of the deductive approach has the advantage that its urban centred model offers the policy relevant benefit of greater simplicity and transparency of method.

#### **1.4 Moving forward**

There is an emerging need for establishing systematic criteria allowing comparison among methods. One proposal is for a straightforward sensitivity analysis of methods by making small changes to some parameters in their definition procedures to assess the scale of the impact on the boundaries produced (see for example the tests conducted by Papps & Newell 2002). Such analyses are examples of what can be termed *intrinsic testing*: the test is of how well a method meets its intended objectives. One form of intrinsic test in the LMA context was provided by Feng (2009), with a membership function measuring how connected each building-block area is to others in the LMAs to which it was allocated. Casado-Díaz et al (2010) provide an example in which a more complete set of related indicators is tested.

Extrinsic testing strategies, on the other hand, assess how far LMAs are appropriate for spatial economic analyses because of their properties on variables *other than* those used in their definitions.

Examples include Barkley et al (1995) who computed three indicators of spatial association to test the core-hinterland dependence in per capita income change, Cörvers et al (2009) who analyse variations in income levels, housing prices, and employment and unemployment rates, and others ranging from Baumann et al (1983) through to Maza & Villaverde (2011) who explore regional economic growth and compare the results from using functional rather than administrative areas.

It is noteworthy that these tests for internal homogeneity or homogeneous behaviour of the areas are not so relevant here because LMAs can be internally heterogeneous but still have the required intensity of flows between their constituent areas. If homogeneity in local attributes was more important to LMAs than cohesiveness in terms of flows then homogeneity would have been the more appropriate objective to use in the definitions. It should also be noted that these analyses are all very problem-dependent and likely to have results which are at least partly determined by the scale of the areas analysed.

These evaluation methods are approaches that can be used to compare different sets of LMA definitions covering the same territory (nb. it would be necessary to ensure that the sets of definitions are in key respects – for example, the number of LMAs that they divided that territory into – very similar, otherwise those more basic factors could dominate the results). The aspiration here is for a consistent set of LMA definitions covering the whole European territory but at present LMAs are either defined for one county only or they do cover many countries but only for selected parts (mainly around large cities).

It is a possible task for a later stage of the present study to create the situation needed to allow rigorous methods of evaluation to be applied. The intermediate step would be to apply selected methods to data for one or more countries: an earlier example of this strategy is reported in Eurostat & Coombes (1992). This strategy requires the selection of LMA definition methods whose results will be evaluated, and this presupposes that the range of plausible methods has been evaluated to make that selection. The first step then has to be decide on the basis for this evaluation of methods; the best practice conclusions derived from the review of academic work here provides the necessary foundations (Table 1).

**Table 1.1: List of the best practice conclusions drawn from the academic literature review**

- [A] areas which represent labour markets
- [B] policy analysis requires all the LMAs to be defined in a comparable way
- [C] based on grouping building-block areas ('bottom up')
- [D] definition methods analysing patterns of commuting
- [E] need to avoid non-contiguous LMAs
- [F] policy preference for readily understood definition processes
- [G] analysis requiring a minimum level of commuting self-containment
- [H] each building-block area should be in one, and only one, of the defined LMAs

- [I] methods whose processes are not limited by contiguity constraints
- [J] advantage of having fewer parameters to justify
- [K] methods to produce adequate results in many different conditions

To avoid the risk of ‘reinventing the wheel’ here it is very valuable to set the items in Table 1 against the “Principles” for LMA definition that were outlined in Eurostat and Coombes (1992), as shown in Table 2. There are two over-riding Principles that establish what makes a set of LMA boundaries fit for purpose: that they should be the product of a rigorous method of definition and reflect labour market geography. These two objectives are echoed in Table 1 by its first two findings on best practice (but note that [A] relates to Principle 2 while finding [B] relates to Principle 1). In addition finding [D] clearly also links with Principle 1 because its emphasis on commuting was derived from the concept of local labour markets.

**Table 1.2: Principles to guide local labour market area definitions**

Principle	Practice
OBJECTIVES	
1. Purpose	to be statistically-defined areas appropriate for policy
2. Relevance	each area to be an identifiable labour market
CONSTRAINTS	
3. Partition	every building block to be allocated to 1 and only 1 area
4. Contiguity	each LMA to be a single contiguous territory
CRITERIA <i>in descending priority</i>	
5. Autonomy	self-containment of flows to be maximised
6. Homogeneity	LMAs’ size range to be minimised (eg. within fixed limits)
7. Coherence	boundaries to be reasonably recognisable
8. Conformity	alignment with administrative boundaries is preferable
SUMMARY	
9. Flexibility	method must perform well in very different regions

Turning to the next Principles which establish what should be constraints in definition methods there are again parallels in Table 1: Principle 3 is directly echoed by finding [H] and similarly Principle 4 has its equivalent in finding [E]. The remaining Principles have a lower priority than the first four, which are the primary objectives and constraints. In practice, there is room for much debate over the extent to which any one of these should over-ride others or, as is perhaps more likely, which of them can be ‘traded-off’ against others. All the same, Table 2 does make clear that Principle 5 has a degree of precedence and it



is echoed in Table 1 in the form of finding [G]. The following three Principles are less clearly reflected in the findings from the academic literature, although the policy-driven Principle 8 that administrative areas should be followed if there are no strong countervailing arguments is supported by a recent study (Cörvers et al 2009). The last of the Principles is echoed by finding [K].

Looking back at Table 1 then reveals that there are four findings which are not obvious echoes of the Principles previously identified. These are all more specifically about the technicalities of methods, although finding [F] is in fact a warning that in the policy field there is preference for fewer technicalities to maximise the transparency of the way the results were produced from the data. There is little problem in this regard with finding [C] because building LMAs ‘bottom up’ is more intuitive, and can involve more simple methods, than a ‘top down’ method (of which there are few). Equally sympathetic with the policy aim for transparency is finding [J] because methods with few separate parameters will often be simpler. This leaves finding [I] as perhaps the ‘odd one out’ because it will call for greater computational intensity in order to achieve more optimal results.

The conclusion of this chapter involves deriving from this final section the issues by which to compare, and then also evaluate, potentially relevant existing LMA definition methods. Table 3 lists these issues (nb. it will also be necessary to consider issues such as the average size of LMAs the method defined).

**Table 1.3: List of issues by which to compare/evaluate existing LMA definition methods**

- does the method produce adequate results in different conditions (eg. metropolitan/peripheral)?
- is every building-block area in one, and only one, of the defined LMAs?
- are all the areas explicitly defined as labour markets?
- how consistently have the areas been defined so as to be comparable?
- are there any non-contiguous LMAs?
- how closely aligned are the LMAs to administrative areas (and was this an explicit constraint)?
- how readily understandable/transparent is the definition process?
- was the process a grouping of building-block areas or a subdivision of the whole territory?
- did the process analyse patterns of commuting and/or any other flow data?
- did the analysis explicitly require a minimum level of commuting self-containment?
- did the analysis explicitly require a minimum of population size or of any other dimension?
- did the definition processes have a contiguity constraint throughout?
- how many parameters are there which need to be justified?
- were parameter values set on a deductive basis or arrived at inductively (so readily modifiable)?

## **References**

- Andersen A (2002) “Are commuting areas relevant for the delimitation of administrative regions in Denmark?” Regional Studies **36** 833-844
- Arbuckle D (1998) “Alternative approaches to defining metropolitan and nonmetropolitan areas” Federal Register **63** 70526-70561
- Barkley D, Henry M, Bao S & Brooks K (1995) “How functional are economic areas? Tests for intra-regional spatial association using spatial data analysis” Papers in Regional Science **74** 297-316
- Baumann J, Fischer M & Schubert U (1983) “A multiregional labour supply model for Austria: the effects of different regionalisations in multiregional labour market modelling” Papers in Regional Science **52** 53-83
- Burger M, de Goei B, van der Laan L & Huisman F (2011) “Heterogeneous development of metropolitan spatial structure: evidence from commuting patterns in English and Welsh city region, 1981-2001” Cities **28**, 160-170
- Candia J, Gonzalez M, Wang P, Schoenharl T, Madey G. & Barabási A-L (2008) “Uncovering individual and collective human dynamics from mobile phone records” Journal of Physics A: Mathematical and Theoretical **41**, 1-11
- Casado-Díaz JM (2000) “Local labour market areas in Spain: a case study” Regional Studies **34** 843-856
- Casado-Díaz JM, Martínez-Bernabeu L & Flórez F (2010) “Los mercados locales de trabajo españoles: una aplicación del nuevo procedimiento británico” JM Albertos & JM Feria (eds) La ciudad metropolitana en España: procesos urbanos en los inicios del siglo XXI Thomson-Civitas, Madrid
- Castells M (1989) The informational city: information technology, economic restructuring and the urban-regional process Blackwell, Oxford
- Cattan N (2001) Functional regions: a summary of definitions and usage in OECD countries. DT/TDPC/TI(2001)6 OECD, Paris
- Cheshire P & Gornostaeva G (2002) “Cities and regions: comparable measures require comparable territories” Cahiers de L’Aurif **135** 13-21
- Christaller W (1933) Central places in central Germany [English translation] Prentice-Hall, Eaglewood Cliffs
- Combes P-P, Duranton G & Overman H (2005) “Agglomeration and the adjustment of the spatial economy” Papers in Regional Science **84** 311-349
- Coombes M (2004) “Multiple dimensions of settlement systems: coping with complexity” in A Champion and G Hugo (ed) New Forms of Urbanization: Beyond the Urban-Rural Dichotomy Ashgate, Aldershot
- Coombes M (2010) “Defining labour market areas by analysing commuting data: innovative methods in the 2007 review of Travel-to-Work Areas” pp227-241 in J Stillwell, O Duke-Williams & A Dennett (eds.) Technologies for migration and population analysis: spatial interaction data applications IGI Global, Hershey
- Coombes M, Green A & Openshaw S (1986) “An efficient algorithm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain” Journal of the Operational Research Society **37** 943-953
- Cörvers F, Hensen M & Bongaerts D (2009) “Delimitation and coherence of functional and administrative regions” Regional Studies **34** 19-31
- Duranton G & Overman H (2005) “Testing for localization using micro-geographic data” Review of Economic Studies **72** 1077-1106
- EUROSTAT and Coombes M (1992) Study on employment zones (E/LOC/20) Office for Official Publications of the European Communities, Luxembourg

- Forstall R, Greene R & Pick J (2009) “Which are the largest? Why lists of major urban areas vary so greatly” Tijdschrift voor Economische en Sociale Geografie **100** 277-297
- Feng Z (2009) “The fuzziness of Travel-to-Work Areas” Regional Studies **43** 707-720
- Flórez-Revuelta F, Casado-Díaz J, Martínez-Bernabeu-Bernabeu L (2008) “An evolutionary approach to the delineation of functional areas based on travel-to-work flows” International Journal of Automation and Computing **5** 10-21
- Fogel D (2006) Evolutionary computation: toward a new philosophy of machine intelligence (3rd edition) Wiley, Hoboken
- Geddes P (1915) Cities in evolution: an introduction to the town planning movement and to the study of civics Williams and Norgate, London
- Glover T and Openshaw S (1995) “The AGW spatial interaction modelling workstation” Available at <http://www.googleedsacuk/people/sopenshaw/w3agwhtml>
- Goodman J (1970) “The definition and analysis of local labour markets: some empirical problems” British Journal of Industrial Relations **8** 176-96
- Green F (1950) “Urban hinterlands in England and Wales: an analysis of bus services” Geographical Journal **116** 64-89
- Green N (2007) “Functional polycentricity: a formal definition in terms of social network analysis” Urban Studies **44**. 2077–2103
- Hemmasi M (1980) “The identification of functional regions based on lifetime migration data: a case study of Iran” Economic Geography **56** 223-233
- Hensen M and Cörvers, F (2003) “The regionalisation of labour markets by modelling commuting behaviour” Paper presented at the 43th European Regional Science Association Congress <http://www.roa.unimaas.nl/seminars/pdf2005/M.Hensen.pdf>
- Hincks S & Wong C (2010) “The spatial interaction of housing and labour markets: commuting flow analysis of North West England” Urban Studies **47** 620-649
- Hugo G (2001) “Addressing social and community planning issues with spatial information” Australian Geographer **32** 269-294
- Killian M & Tolbert C (1993) “Mapping social and economic space: the delineation of local labor markets in the United States” in J Singlemann & F Deseran (eds) Inequality in local labor markets Westview Press, Boulder
- Kirkpatrick S, Gelatt C & Vecchi M (1983) "Optimization by simulated annealing" Science **220-4598** 671-680
- Laan L van der & Schalke R (2001) “Reality versus policy: the delineation and testing of local labour market and spatial policy areas” European Planning Studies **9** 201-221
- Lolonis P & Armstrong M (1993) “Location-allocation models as decision aids in delineating administrative regions” Computers, Environment and Urban Systems **17** 153-174
- Longley P, Cheshire J, Mateos P (2011) “Creating a regional geography of Britain through the spatial analysis of surnames” Geoforum **42** 506–516
- Mackinder H (1902) Britain and the British seas Appleton, New York
- Marcon E & Puech F (2003) “Evaluating the geographic concentration of industries using distance-based measures” Journal of Economic Geography **3** 409-428
- Masser I & Scheurwater J (1980) “Functional regionalisation of spatial interaction data: an evaluation of some suggested strategies” Environment and Planning A **12** 1357-1382
- Maza A & Villaverde J (2011) “EU regional convergence and policy: does the concept of region matter?” Journal of Policy Modeling **33** 889-900

- Mitchell W & Watts M (2010) “Identifying functional regions in Australia using hierarchical aggregation techniques” Geographical Research **48** 24-41
- Newman M & Girvan M (2004) “Finding and evaluating community structure in networks” Physical Review E **69** 026113:1-15
- Nuffel N van (2007) “Determination of the number of significant flows in origin-destination specific analyses: the case of commuting in Flanders” Regional Studies **41** 509-524
- Nystuen J & Dacey M (1961) “A graph theory interpretation of nodal regions” Papers of the Regional Science Association **7** 29-42
- Openshaw S & Rao L (1995) “Algorithms for re-engineering 1991 Census geography” Environment & Planning A **27** 425-446
- Openshaw S & Taylor P (1981) “The modifiable areal unit problem” in N Wrigley, R J Bennett (eds) Quantitative Geography Routledge, London
- Pahl R (1965) Urbs in rure: the metropolitan fringe in Hertfordshire (Geographical Papers 2) LSE, London
- Papps K & Newell J (2002) “Identifying functional labour market areas in New Zealand: a reconnaissance study using travel-to-work data” Discussion Paper 443 Institute for the Study of Labor, Bonn
- Roca J & Moix M (2005) “The interaction value: its scope and limits as an instrument for delimiting urban systems” Regional Studies **39** 357-373
- Schubert U, Gerking S, Isserman A & Taylor C (1987) “Regional labour market modelling: a state of the art review” in Fisher M and Nijkamp P (eds) Regional labour markets: analytical contributions and cross-national comparisons Elsevier, Amsterdam
- Slater P & Winchester H (1978) “Clustering and scaling of transaction flow tables: a French inter-Departmental migration example” IEEE Transactions on Systems, Man and Cybernetics SMC **8** 635-640
- Small K & Song S (1992) “‘Wasteful’ commuting: a resolution” Journal of Political Economy **100** 888-898
- Smart M (1974) “Labour market areas: uses and definition” Progress in Planning **2** 238-353
- Sohn J (2005) “Are commuting patterns a good indicator of urban spatial structure?” Journal of Transport Geography **13** 306-317
- Sokal R & Sneath P (1963) Principles of numerical taxonomy Freeman, San Francisco
- Spence N & Taylor P (1970) “Quantitative methods in regional taxonomy” Progress in Geography **2** 1-64
- Wilcock R & Sobel I (1958) Small city job markets: the labor market behavior of firms and workers Institute of Labor and Industrial Relations, University of Illinois, Urbana

## CHAPTER 2: OVERVIEW OF THE INFORMATION COLLECTED

In addition to the review of LMAs in the applied social sciences, this report also considers the information provided by NSIs in reply to a tailor-made questionnaire on LMAs (from now on called “the 2012 questionnaire” or Q-2012). The questionnaire was sent to NSIs in January – February 2012 and was aimed to compile the fundamental evidence needed for the comparison of LMA concepts recognised and implemented in each Member State. The questionnaire was customised for each of the EU MS: it included pre-filled questions, where the answers were retrieved from NSIs replies to a previous related questionnaire sent by Eurostat in 2007 or from other recent work carried out by Eurostat. Apart from the pre-filled answers some of the national questionnaires included methodological notes on the delineation method used in each specific country.

The standard template of the questionnaire included questions that were aimed at retrieving from all EU MS homogenous information covering four informational dimensions, specifically:

- a. Characterisation of the existing delineation methods
- b. NSIs’ position towards an LMA geography and their expectations from this
- c. Inputs for the delineation of LMAs
- d. Possible characterisation of LMAs

According to the answers to the 2012 questionnaire, nine countries have official LMAs in use (Belgium, Germany, Estonia, Finland, France, Italy, Netherlands, Sweden and the United Kingdom), and seven NSIs reported on LMAs definitions which were not used officially (Cyprus, Czech Republic, Denmark, Greece, Portugal, Slovenia and Slovak Republic). While Chapter 3 of this report investigates the specific methodology (ie. characteristics of the existing delineation methods), this chapter, instead, gives an overview of the remaining information provided by MSs in reply to the 2012 questionnaire.

The overview information that follows covers all EU27 MS except for Malta. All the information was retrieved through the 2012 questionnaire, except in the case of Belgium, where the information is a result of the previous questionnaire sent in 2007 by Eurostat.

### 2.1 NSIs’ position towards an LMA geography and their expectations from this

One of the goals of the dissemination of this Interim Research Report is to demonstrate the usefulness of an exercise of delineating LMAs, that is why it was considered relevant to gather information that could guide the process from the first steps. In this direction, the 2012 questionnaire included questions related to NSIs position towards an LMA geography and their expectations from this. This information unfolds in a mixture of qualitative and quantitative aspects.

An overview of the quantitative information that could reflect NSIs’ position towards an LMA geography is presented in the table below. This type of information is mostly related to the issue of scale of LMAs, which is measured in terms of n° of LMAs, minimum/maximum/average area and/or

population. The table below summarises this information for those countries where, according to the questionnaire, LMAs have been defined either officially or as a result of academic exercises (in the latter case the table only includes the cases for which the information available was sufficiently detailed).

**Table 2.1: Scale of existing LMAs**

Country	Total population 2011	Area 2009 (km <sup>2</sup> )	n° NUTS 3	n° LAU 1	n° LAU 2	National denomination	Produced by	Year when first produced	In official use	Basic building blocks	Current n° of LMA	Min. LMA population	Max. LMA population	Average LMA population	Min. LMA area (km <sup>2</sup> )	Max. LMA area (km <sup>2</sup> )	Average LMA area (km <sup>2</sup> )	Population density
BE*	10.951.266	30.666	44	-	589	bassins d'emploi	Stat. Office Wallonie	2000	yes	LAU 2	47	NA	NA	233.006	NA	NA	652	357
CZ	10.532.770	78.870	14	77	6.251	-	Charles University Prague	1961	no	LAU 2	184	10.304	1.442.991	57.243	101	2.719	429	134
DE*	81.751.602	357.114	412	1.351	12.066	Arbeitsmarktregionen	Statistisches Bundesamt	1972	yes	NUTS 3	258	60.560	3.442.675	316.867	75	6.999	1.384	229
EE*	1.340.194	43.452	5	15	226	Tööjõuareaal	University of Tartu	2001	no	LAU 2	15	10.168	522.147	89.346	1.023	4.807	2.897	31
EL	11.309.885	120.167	51	1.035	6.130	topikes agores ergasias	KEPE	2001	no	LAU 1	667	-	3.887.000	16.956	-	3.607	180	94
FR*	65.048.412	548.763	100	3.785	36.680	zones d'emploi	INSEE	1984	yes	LAU 2	321	10.044	5.851.493	202.643	119	40.945	1.710	119
IT*	60.626.442	301.392	110	-	8.094	Sistemi locali del lavoro	ISTAT	1981	yes	LAU 2	686	3.318	3.808.731	88.377	10	3.666	439	201
NL*	16.655.799	37.357	40	-	418	Regionale Platforms voor de Arbeidsmarkt	CBS	1991	yes	LAU 2	34	85.406	1.826.562	489.876	148	6.561	1.099	446
PT	10.636.979	88.840	30	308	4.260	-	INE	2007	no but planned	LAU 1	18	44.996	2.987.044	590.943	3.393	6.385	4.936	120
SK	5.435.273	49.013	8	79	2.928	zóny zamestnanosti	INFOSAT	2002	no	LAU 2	51	10.392	737.484	106.574	292	3.581	961	111
FI*	5.375.276	335.765	20	70	336	Työssäkäyntialueet	Statistics Finland	1990	yes	LAU 2	58	4.150	1.414.693	92.677	NK	NK	NK	NK
SE*	9.415.570	449.159	21	-	290	Lokala arbetsmarknader	Statistics Sweden	1991	yes	LAU 2	76	2.878	2.435.363	123.889	883	27.410	5.910	21
UK*	62.435.709	244.436	139	380	10.310	Travel-To-Work-Areas	ONS & Newcastle Univ.	1998	yes	LSOA	243	7.907	8.952.972	256.937	53	5.174	1.006	255

**\*official LMAs**

**Note :** In Finland the municipalities with weak commuting figures were not allocated to LMAs.

**Source:** 2012 questionnaire on LMAs and Eurostat Chronos Database (own calculations)



Qualitative information on NSIs positions towards an LMA geography and their expectations from this (such as the purpose and the reasons which stood for the creation of LMAs, the main concerns of the exercise of defining a new statistical classification based on LMAs, and other issues about the statistical characteristics that should be met by LMAs if an EU-wide grid comparable areas would be proposed) was also collected. Although not summarised here, this information will be analysed and considered for the following activities of the project (ie. Elaboration of recommendations on the establishment of an EU-wide harmonised grid of comparable LMA).

## 2.2 Inputs for the delineation of LMAs and characterisation of LMAs

Despite an obvious disparity in the empirical and theoretical approaches to the issue of delineating LMAs, a vast majority of the MS rely on the use of data on commuting flows. The following table gives an overview of the sources of datasets used for delineating existing LMAs.

**Table 2.2: Sources of datasets of existing LMAs**

Country	Sources of datasets
BE*	Population Census 1991
CZ	Population Census 2001
DE*	Federal Employment Agency 1998 - 2011
EE*	Population Census 2001
EL	Population Census 2001
FR*	Population Census 1982,1984, 1999, 2006-2008
IT*	Population Census 1991, 2001
NL*	Population Census 2001, StatLine since 2006
PT	Population Census 1991, 2001
SI	SRDAP 2000-2010, Population Census 2002
SK	Population Census 2001
FI*	Register-based population statistics 1996-2011 discon.
SE*	Employment register 1993-2010
UK*	Population Census 1991,2001

\*official LMAs

The table above leads to the conclusion that a variety of statistical sources for information on commuting flows exist, although the most traditional and still the most widely-used are the decennial Census of Population and Housing (for Belgium, Czech Republic, Estonia, Greece, France, Italy, Portugal, Slovak Republic, United Kingdom).



In addition to identifying the sources of the commuting datasets used for delineating existing LMAs, through the 2012 questionnaire, the existence of such data in all EU MS, was also investigated in order to accomplish the purpose of this exercise aimed at defining an EU-wide grid of comparable LMA. As a preliminary conclusion, the decennial Census of Population and Housing seems to be the preferable source of commuting datasets, especially after the 2011 Census, being this the first one legally regulated for all EU27 MS.

The “2011 European Census Programme” establishes the technical and legal framework required for census output harmonisation, leaving the Member States free to decide how to conduct the census, and to select the data sources, the methodology and technology applicable in each national context. The main requirements established at EU level refer to census topics, reference period, breakdowns, metadata, quality reporting and data transmission<sup>4</sup>.

The breakdowns for geographical area in the 2011 round of censuses, as established by Commission Regulation (EC) N° 1201/2009, identify geographical areas from a high level of detail (LAU 2) to the national level.

**Table 2.3: Breakdowns for geographical area in Population and Housing Census 2011**

Geographical area		GEO.N.	GEO.L.	GEO.M.	GEO.H.
o.	Total (in the territory of the Member State)	o.	o.	o.	o.
x.	All NUTS 1 regions in the Member State		x.	x.	x.
x.x.	All NUTS 2 regions in the Member State		x.x.	x.x.	x.x.
x.x.x.	All NUTS 3 regions in the Member State			x.x.x.	x.x.x.
x.x.x.x.	All LAU 2 regions in the Member State				x.x.x.x.

**Source:** Regulation (EC) N° 1201/2009

**Note:** The codes ‘x.’, ‘x.x.’ and ‘x.x.x.’ depend on the NUTS classification, the code ‘x.x.x.x.’ on the LAU classification, valid for the Member States on 1 January 2011. The annotation ‘N’ identifies the breakdown that refers to the national level.

Regulation (EC) N° 763/2008 of the European Parliament and of the Council establishes the topics to be covered by the round of 2011 censuses by geographical levels: for NUTS 3 and LAU 2, and for NUTS 1 and NUTS 2. While the “place of residence” is ordinarily included among population census topics at the highest level of detail (LAU 2), the Census Regulation includes the “location of place of work” as one of the obligatory topics to be covered, but at a lower level of detail (NUTS 2)<sup>5</sup>. However, this only applies to the data to be submitted to Eurostat, and not actually to the collection of data.

<sup>4</sup> EU legislation on the 2011 Population and Housing Censuses: Explanatory Notes, EUROSTAT, 2011 edition

[http://epp.eurostat.ec.europa.eu/portal/page/portal/product\\_details/publication?p\\_product\\_code=KS-RA-11-006](http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-11-006)

<sup>5</sup> Commission Regulation (EC) N° 1201/2009 on population and housing censuses as regards the technical specifications of the topics and of their breakdowns

The replies to the 2012 questionnaire showed that a majority of Member States collected through the 2011 Population Census the variable “location of place of work” at LAU 2 level. **Table 2.4** gives an overview of the availability of commuting datasets after 2011 Census and their source, as resulting from MSs’ replies.

**Table 2.4: Availability of commuting datasets in EU27 MS after 2011 Population Census**

Country	Available	Source	Collection method	Level
BE	NK	NK	NK	NK
BG	YES	Population Census 2011	directly collected, exhaustive	LAU 1
CZ	YES	Population Census 2011	directly collected, exhaustive	LAU 2
DK	YES	Register-Based Labour Force statistics	register based	LAU 1
DE	YES	Federal Employment Agency	register based	LAU 2
EE	YES	Population Census 2011	e-census, directly collected, exhaustive	LAU 2
IE	YES	Population Census 2011	directly collected, exhaustive	LAU 2
EL	NK	NK	NK	NK
ES	YES	Population Census 2011	directly collected, 12 % sample	LAU 2
FR	YES	Population Census 2011	register based	LAU 2
IT	YES	Population Census 2011	directly collected, exhaustive	LAU 2
CY	YES	Population Census 2011	directly collected, exhaustive	LAU 2
LV	YES	Population Census 2011	e-census, directly collected, exhaustive	LAU 2
LT*	YES	Population Census 2011	e-census and directly collected, exhaustive	LAU 2
LU	YES	Population Census 2011	directly collected, e-census, exhaustive	LAU 2
HU	YES	Population Census 2011	e-census, partially register based, directly collected, exhaustive	LAU 2
MT	NK	NK	NK	NK
NL	YES	StatLine Database	register based combined with sample survey	LAU 2
AT	YES	Population Census 2011 & Register based	Register-based full enumeration	LAU 2
PL**	YES	Population Census 2011	register based, e-census and directly collected, exhaustive	LAU 1
PT	YES	Population Census 2011	e-census and directly collected, exhaustive	LAU 1
RO*	YES	Population Census 2011	e-census and directly collected, exhaustive	LAU 2
SI	YES	Population Census 2011 & SRDAP	directly collected, exhaustive & register-based	LAU 2
SK	YES	Population Census 2011	directly collected, exhaustive	LAU 2
FI	YES	Register-Based Population Statistics	register based	LAU 2
SE	YES	SCB Sweden database	register based labour statistics	LAU 2
UK	YES	Population Census 2011	directly collected, exhaustive	LSOA

**Source:** 2012 questionnaires on LMAs & Population Census 2011 official webpages

NK – not known

\*according to the questions included in the 2011 Population Census questionnaire

\*\* Works on the possibility of presenting data on commuting flows at LAU 2 in Poland are ongoing.

Finally, in what regards the availability of statistical variables that characterise LMAs, in the majority of the countries where LMAs have been defined, specific data for LMAs are not collected. The information sources for the characterisation of LMAs derive from the statistics available at the basic building block level (LAU 1, LAU 2 or other level) and are constructed through aggregates. However, some exceptions still exist: for Germany, where four regional indicators (ie. unemployment rate, averaged over four years, annual infrastructure indicator, gross annual wages per employee, and employment forecast) are collected for the Joint action for improving regional economic structures, and for United Kingdom, where some specific statistics (unemployment claimant counts, job centre vacancies) are published.

## CHAPTER 3: CROSS-NATIONAL EVALUATION OF EXISTING CONCEPTS OF LMAs

This chapter is considering the potential for identifying a common European way of defining Local Labour Market Areas (LMAs). An initial research reviewed the state of the art in the applied social sciences in methods of defining LMAs. Subsequent to that, information was sought from each member state (MS) on existing national definitions of LMAs: the research team is extremely grateful for the information that was provided, primarily by the national statistical institutes (NSIs). This report includes in its Annex extracts from the supplied information, converted into a standard format so that the different national methods can be directly compared. The next section of this chapter summarises the principles against which the national methods of LMA definitions will be compared. The followed section is the core of this report, which provides these comparisons of methods in detail. The chapter ends by drawing some conclusions in terms of identifiable best practices in the definition of LMAs.

### 3.1 Principles in the definition of LMAs

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. LMAs are a specific form of the functional regions which are increasingly recognised across Europe as the appropriate units for economic research and policy analyses which need comparable data for comparable areas. In most countries, the availability of data on labour market flows between areas is limited to commuting between home and workplace. The result is that in practice most official and academic approaches to defining LMAs focus on patterns in matrices of commuting flows between areas. Their objective is typically to draw boundaries which delimit LMAs that are highly self-contained and cohesive in terms of travel-to-work.

Looking back at Chapter 1 provides the basis for the assessment of the national methods for defining LMAs. That chapter drew on its review of best practice in the applied social sciences to set out the issues by which to compare and/or evaluate existing LMA definition methods in its Table 1.3:

1. does the method produce adequate results in different conditions (eg. metropolitan/peripheral)?
2. is every building-block area in one, and only one, of the defined LMAs?
3. are all the areas explicitly defined as labour markets?
4. how consistently have the areas been defined so as to be comparable?
5. are there any non-contiguous LMAs?
6. how closely aligned are the LMAs to administrative areas (and was this an explicit constraint)?
7. how readily understandable/transparent is the definition process?
8. was the process a grouping of building-block areas or a subdivision of the whole territory?

9. did the process analyse patterns of commuting and/or any other data?
10. did the analysis explicitly require a minimum level of commuting self-containment?
11. did the analysis explicitly require a minimum of population size or of any other dimension?
12. did the definition processes have a contiguity constraint throughout?
13. how many parameters are there which need to be justified?
14. were parameter values set on a deductive basis or arrived at inductively (so readily modifiable)?

Not all the issues listed above are equally relevant. The primary issues are those that should determine which methods to take forward to empirical analyses of the commuting datasets of different countries. These principal objectives, constraints and criteria are those that make it most likely that a method will produce appropriate LMA definitions for policy analysis across the whole of the EU:

- issue 2 (regions are exhaustive & non-overlapping)
- issue 4 (consistency: same method and parameters formally applied in a replicable process)
- issue 5 (regions are contiguous)
- issue 10 (minimum self-containment required).

By the same token, the other issues relate to more secondary concerns which a method may diverge from without contravening the basic objectives or constraints for this research:

- issue 1 (non-core-based)
- issue 3 (explicitly defined labour markets)
- issue 6 (respects administrative boundaries)
- issue 7 (clarity simplicity)
- issue 8 (aggregative)
- issue 9 (not based on other data apart from commuting)
- issue 11 (minimum size)
- issue 12 (not contiguity constrained)
- issue 13 (few parameters)
- issue 14 (analytically adjusted parameters)

The evidence base for assessing each national method in relation to each of the issues is rather varied. The principal source has been the set of responses to Q-2012 together with the further responses to the follow-up queries where necessary. In several of these responses there were links to further information and this has been drawn upon too as appropriate, along with previous knowledge of the researchers which is based in part on academic research. Despite this range of evidence, there are numerous cases of issues for which a robust assessment cannot be provided of numerous national methods: these cases are indicated by the essentially provisional assessment of “?” (Tables 3.1 and 3.2). In such cases it may

well be appropriate to update the information in Tables 3.1 and 3.2 if further information becomes available to the researchers.

The following notes summarise how the evidence has been used to reach a rather general assessment of each national method in relation to each issue. These assessments are reported in the form of the set of symbols used in Tables 1 and 2 where the 14 issues are considered in turn. The essential arguments supporting these assessments were outlined in the report of Activity 1: for example, the basic objective of this research – to seek a common European method of defining LMAs – was interpreted as meaning that any such method needs to produce adequate results in a range of different geographical conditions (issue 1 here). In this report, these rather general guide-lines have been made more specific so that the different methods can be more clearly contrasted. Thus for example this need to define adequate LMAs in all the very contrasting geographical circumstances across Europe is seen to ‘count against’ a method which has pre-suppose that all LMAs will take a particular geographical form. For the definition methods under review here, this leads to a negative assessment of method based on an initial step to identify urban cores, because this cannot be expected to be so appropriate in peripheral less urbanised regions.

1. Does the method produce adequate results in different conditions (eg. metropolitan/peripheral)?

All the methods analysed for this report have been either defined and/or utilised by NSIs or developed and tested in the academic sphere so their appropriateness to the territory they were designed for is not in question here. At the same time, to establish a cross-national definition of LMAs it is preferable that the method does not rely heavily on the definitions of foci or urban centres, otherwise it will be less likely that adequate results in a wide range of different geographical circumstances.

☑ = not core based

★ = cores are used but then reconsidered during the process

? = unclear evidence

☒ = core based

2. Is every building-block area in one, and only one, of the defined LMAs?

This issue too would be ideally resolved by a full set of maps, with the alternative source here being the information on methods provided in response to Q-2012 which is summarised in the Annex.

☑ = map evidence or the answer to Q-2012 supports the answer Yes

☒ = No

3. Are all the areas explicitly defined as labour markets?

For almost all countries the evidence is limited to the answer to Q-2012: in some cases the reported LMAs were defined more generally as functional or city/metropolitan regions and in these cases some scepticism about their validity as labour market areas is warranted.

☑ = Yes they are explicitly defined as LMAs

★ = whilst not explicitly defined as LMAs, they probably are reasonable approximations of LMAs

4. How consistently have the areas been defined so as to be comparable?

For this issue too the evidence on almost all countries is limited to the answer to Q-2012 and it should be acknowledged that there may be some cases where the definitions were reported to be entirely consistent when in practice some adjustments – which cannot be identified from the available evidence – had taken place to improve the results. This issue has been assessed in three<sup>6</sup> separate items.

4a. Consistency in applying the same method to all the territory

4b. Consistency in applying the same parameters to all the territory

4c. Consistency in absence of ‘manual’ (non-formalised) adjustments

Each of the items has been assessed according to the following categories:

☑ = evidence of a high level of consistency

★ = probably consistent (but the evidence is not definitive)

? = there are certain grounds for doubt

☒ = inconsistency clear (eg. between regions)

5. Are there any non-contiguous LMAs?

The evidence on this issue would ideally have been a full set of maps of the LMAs produced by the method concerned but where they were unavailable the responses to Q-2012 were drawn upon.

☑ = map evidence or the answer to Q-2012 supports the answer No

☒ = Yes presumably (answer to Q-2012 says so)<sup>7</sup>

6. How closely aligned are the LMAs to administrative areas (and was this an explicit constraint)?

---

<sup>6</sup> This division of the issue allows a more complete assessment of the characteristics of each method (nb. the same strategy has been applied to issues 7, 10 and 11).

<sup>7</sup> Even if contiguity was not a constraint upon the definition process (see issue 12), it is typically then enforced through a final manual adjustment: as a result it may be that if the answer to issue 5 is Yes a similar adjustment to the results could be applied subsequently.

This issue too would be ideally resolved by a full set of maps, with the alternative source here being the information on methods provided in response to Q-2012 that is summarised in the Annex. As almost all countries use as ‘building block areas’ small areas from the administrative hierarchy (most often LAU2), so the issue relates to alignment of the LMAs with higher order<sup>8</sup> administrative areas.

- ☑ = map evidence suggests close alignment
- ★ = non-trivial adjustments to increase alignment
- ? = uncertain
- ☒ = evidence suggests the answer is No

#### 7. How readily understandable/transparent is the definition process?

The evidence on this issue is in the answers to Q-2012 and the assessment then relies on past experiences of policy-makers (mostly in Britain) viewing a certain level of complexity to be problematic. This issue is subdivided into two items, with the latter – “transparency” – giving a negative assessment to methods that rely upon informal judgements, because these mean their results may not be replicable.

##### 7a. Clarity

- ☑ = clear and readily understandable
- ★ = understandable enough, based on the use of the results by the NSI
- ? = there is some lack of clarity
- ☒ = probably too complicated

##### 7b. Transparency

- ☑ = complete
- ★ = non-formalised final step which makes minor changes based on local knowledge
- ? = uncertain
- ☒ = frequent reliance upon informal judgements

#### 8. Was the process a grouping of building-block areas or a subdivision of the whole territory?

This issue recognises that although most of the familiar methods start with individual ‘building blocks’ (eg. LAU2 areas) which are then grouped by the algorithm, there have been some methods which begin with the whole territory and then sub-divide it into LMAs. Some research evidence suggests that the latter type is more likely to produce sub-optimal results, while their infrequent use also indicates that they are less intuitively understandable. The evidence for assessing methods on this issue is in the responses to Q-2012.

- ☑ = Yes the process was an aggregative grouping
- ☒ = No the process proceeded by dividing up the whole territory

---

<sup>8</sup> It is worth noting that a close fit to administrative areas could be related to the nature of those areas in the particular country analysed; as a result, the most positive assessment of methods on this issue is here reserved for any method that includes increased alignment with administrative areas as a formal secondary criterion within the definition procedure.



9. Did the process analyse patterns of commuting and/or other related data?

This issue too was directly addressed in Q-2012 (and the relevant aspects of the responses have been summarised in the Annex). Given the focus here on finding a method potentially applicable to the whole EU there is a preference for a lower number of inputs because there are few relevant datasets which are available in every country.

- = no data other than commuting flows
- ★ = one additional dataset (eg. distance/time)
- ? = uncertain
- = more than one additional dataset

10. Did the analysis explicitly require a minimum level of commuting self-containment?

This issue also was directly addressed in Q-2012 along with the subsequent requests for clarifications, from where all the information necessary for the assessment has been extracted. The issue has been subdivided to allow a twofold assessment.

10a. Is reaching a certain degree of self-containment for all defined LMAs a criterion in the method?

- = Yes
- ★ = it is initially but it is not rigidly adhered to in the end
- ? = uncertain
- = No

10b. Does functional dependence (eg. self-containment) act as the primary basis for grouping areas?

- = Yes
- ★ = no but it will be tend to be supported by the grouping based on the largest commuting flow
- ? = uncertain
- = No

11. Did the analysis explicitly include a minimum of population size or of any other dimension?

This is another issue directly addressed in Q-2012 but the responses have suggested that this issue should be subdivided into two items. A minimum population size can often be valued by policy-makers (for reasons such as a reduction on the volatility of statistical trends for the areas); at the same time, there are some policy-makers who value a restriction on the physical extent of areas.

11a. Minimum population

- = Yes a minimum population size criterion is used
- ★ = Yes, but it is not respected in final steps (eg. when assigning residual areas)
- ? = unproven
- = No there is no such explicit criterion

#### 11b. Maximum area

- = Yes a maximum area is imposed, either directly or indirectly (eg. by time/distance commuted)
- ★ = Yes it is considered but it is not an absolute restriction (eg. when assigning residual areas)
- ? = uncertain
- = No there is no such explicit criterion

#### 12. Did the definition processes have a contiguity constraint throughout?

This is another issue directly addressed in Q-2012: the responses are evaluated in line with the findings in the report on Activity 1 that a contiguity constraint leads to sub-optimal results in some cases, notably when the methods are very simple and consist on only a few deterministic steps.

- = No the process is not so constrained
- ★ = it is only imposed as a final step
- ? = uncertain
- = Yes the groupings are contiguity constrained during the whole process

#### 13. How many parameters are there which need to be justified?

The evidence on this issue is in the answers to Q-2012 but in some cases this may have been less fully completed than in others so the assessment here may not be as accurate as for some other issues.

- = only one or two parameters were set
- ★ = three to five parameters are necessary
- ? = uncertain
- = more than five parameters

#### 14. Were parameter values set on a deductive basis or arrived at inductively (so readily modifiable)?

The evidence on this issue in answers to Q-2012 requires a degree of interpretation, because for many countries the research has not had access to a description of the reasoning behind the set parameters, so the comparative assessments here may not be as accurate as for some other issues.

- = parameters based on analytical processes
- ★ = parameters very largely based on analytical process
- ? = uncertain
- = arbitrarily fixed (need sensitivity testing and local knowledge)

### 3.2 Cross-national evaluation of national methods of defining LMAs

According to the responses to Q-2012 the following classification of countries in relation to the definition of LMAs can be made.

**A.** Countries that have official LMAs in use: BE<sup>9</sup> DE EE FI FR IT NL SE UK.

A1. Countries that have an official analytical LMA delineation methodology: BE DE EE FI FR IT SE UK.

A2. Countries that have no analytical LMA delineation methodology: NL.

**B.** Countries whose response to Q-2012 reported<sup>10</sup> on LMAs definitions not used officially: CY<sup>11</sup> CZ DK<sup>12</sup> EL PT SI SK.

B1. Countries that are developing/considering an LMA delineation method, which is not official yet: PT.

B2. Countries where LMA definition research was done but is not officially used: CY CZ EL DK SI SK.

**C.** Countries reporting no officially or academically defined LMAs: AT BG ES HU IE<sup>13</sup> LT LU LV PL RO.

C1. Countries reporting that there is no official definition of LMAs, and not reporting any academic study: AT BG HU IE LT LU LV RO.

C2. Countries reporting that they are considering a LMA delineation method, but is not official yet: PL.

**D.** Countries that have never answered the questionnaires (1): MT.

For evaluation purposes the analytical methodologies (official or academic) reported in responses to the Q-2012 are assessed in the report in two Tables.

Table 3.1 covers both (A1) official methods in use: BE DE EE FI FR IT SE UK  
and (B1) academic methods under official consideration: PT

Table 3.2 covers academic methods not currently under official consideration (B2): CZ EL SI SK

---

<sup>9</sup> The information about **BE** is based on the response to Q-2007 since the answer to Q-2012 was not received.



<sup>10</sup> It is important to note here that this classification is exclusively based on the information provided by the NSI (in fact in several other countries, like **ES**, some academic exercises have been conducted that were not reported by the respondent to Q-2012).

<sup>11</sup> The academic study that was mentioned by **CY** in their answer to Q-2012 refers to the same method described in **EL** applied over the **CY** dataset, therefore the assessment conducted here refer to both cases.

<sup>12</sup> None of the methods described in the answer to Q-2012 have been accepted by the NSI. Of them the one that would fit better this study consists in applying the official method **SE**; therefore no specific description is included here for **DK**.

<sup>13</sup> **IE** answered that NUTS 3 level regions are used for all purposes for labour market analysis, but it is likely that this also the case in the countries listed in the rest of the categories B and C.

It is not appropriate here to describe in detail the assessment, on each issue here, of all these methods. Tables 3.1 and 3.2 now summarise these assessment, and set out the basic grounds for selecting some methods to evaluate empirically, by using them to analyse commuting data in several different countries. Selecting the most suitable methods involves examining the evidence in Tables 3.1 and 3.2 so the approach taken here is to describe that evidence by proceeding through the 14 issues in a sequence which aims to progressively identify the methods that are less suitable<sup>14</sup> than others. To express this another way: some methods may be ideally suited to the specific territories they were designed for, but a method that can produce adequate results across the whole EU method needs to avoid being highly unsuitable for any country, rather than ideal for just one or even a few countries. One factor relevant for the selection of methods to evaluate further is that the method produced areas used by that NSI, because this implies a practical value rather than an ‘optimality’ that may be purely academic. The implication here is that there will be a preference here for the eight methods shown in Table 3.1 over those in Table 3.2.

The results in Table 3.1 indicate that at the outset it is possible to set aside issues 2, 3 and 8 since they do not strongly discriminate between the eight methods because they all have  in these columns. The same is true of issue 14, although in this case it is because of the lack of any positive assessments. The remaining issues of primary concern are then 5 along with 4 and 10 (with the latter two issues now subdivided into three and two separable issues respectively). The method with a negative assessment on issue 5 (contiguous final results) is that of **FI** but in fact it is possible for an ‘extra’ stage to be added to the method to resolve any problems of non-contiguity: this is the approach used by several methods (eg. **IT** and **UK**, leading to their  assessments for issue 12). It is also true that while some methods have no non-contiguous areas in the results on their ‘home’ countries the same outcome may not occur if the method was applied in other countries (especially those with smaller LAU2 areas to analyse).

---

<sup>14</sup> It must of course be recognised that all the methods reviewed here could be reasonable options: they have been found to be appropriate sets of LMAs for the countries where they have been used by the NSI concerned and/or by academics.

**Table 3.1: Summary evaluation of official national methods against the issues identified**


MS:	BE <sup>15</sup>	DE <sup>16</sup>	EE	FI	FR	IT	SE	UK
1 (non-core-based)	✓	?	✗	✗	✓	★	✗	✓
2 (exhaustive & non-overlapping)	✓	✓	✓	✓	✓	✓	✓	✓
3 (explicitly defined labour markets)	✓	✓	✓	✓	✓	✓	✓	✓
4a (consistency – method)	✓	✓	✓	✓	★	✓	✓	✓
4b (consistency – parameters)	✓	✓	✓	✓	✗	✓	✓	✓
4c (consistency – analytical)	✗	✗	✓	✓	★	★	✓	★
5 (contiguous LMAs)	✓	?	✓	✗	✓	✓	✓	✓
6 (aligned to adm. boundaries)	✗	★	✗	✗	★	✗	✗	✗
7a (clarity/simplicity)	★	?	✓	✓	★	★	✓	✓
7b (transparency/reproducible)	✗	★	✓	✓	✓	★	✓	★
8 (aggregative)	✓	✓	✓	✓	✓	✓	✓	✓
9 (based on additional datasets)	✓	★	★	✓	✓	✓	✓	✓
10a (minimum self-containment)	✗	✓	✗	✗	★	✓	✗	✓
10b (self-containment guides the process)	✓	?	✗	✗	✓	✓	★	✓
11a (minimum population)	✗	✓	✗	✗	?	✓	✗	✓
11b (maximum area/distance/time)	✗	✓	✗	✗	✓	✗	✗	✗
12 (not contiguity constrained)	✓	✓	✗	✓	✗	✓	✓	✓
13 (few parameters)	✗	?	✓	✓	★	✗	✓	★
14 (analytically set parameters)	✗	?	✗	✗	✗	✗	✗	✗

Table 3.1 makes it clear that there are some uncertainties about the precise operation of the **DE** method, although other aspects, such as the information needs and the criteria all valid LMAs must fulfil are clear. One of these is the fact that the procedure includes a maximum commuting time within any valid LMA. The remaining uncertainties and the lack of readily usable information regarding this last point seriously hamper the possibility of considering this method for the empirical analyses later in this research study.

There is only one ? marked in the **FR** column but issue 4b also has a negative assessment because there is use of local knowledge, and this could not be replicated across the whole of the EU.

<sup>15</sup> See Note 9.

<sup>16</sup> See the annex, where the DE method is reviewed.

On a similar basis the **BE** method can be discarded because of its assessment on issue 4c (as well as the fact that the evidence here is limited to the Q-2007 response and so involves some uncertainty).

Of the remaining five methods, there are two (**IT** and **UK**) that are more ‘open system’ methods in that they are not heavily based on a preliminary identification of core areas. In fact the **IT** method is basically the method applied in the UK several decades ago which has since been refined as a result of findings from applying it to several other countries around the world (Casado-Díaz et al 2010). As a result of this experience it is justifiable to select the **UK** method in preference to that of **IT** for the analyses of data from other countries in the EU.

Table 3.1 then has three remaining methods (**EE**, **FI**, and **SE**) and these would be potentially interesting contrasts to that of the **UK** because they all begin by identifying employment centres or foci. All the three methods lack an explicit minimum self-containment (giving a negative evaluation in issue 10a), in the **SE** method there is a clustering analysis that is less restricted to foci, producing a more positive assessment on the primary issue 10b. This suggests that the **SE** method is the most appropriate comparator to the **UK** method for the cross-national analysis in the next stage of this research study.

The case for choosing **SE** method as one to test empirically is also supported by the fact that it is, together with **UK** method, the one that has been applied more often beyond its own national boundaries. Examples of other applications of the **SE** method were cited in the responses to Q-1012 from fellow Scandinavian countries **DK**, **FI** and **NO** as well as **SI** (Drobne et al 2009).

Some of the responses to Q-2012 included reference to official methods of which the NSI was aware but whose resulting LMAs had not been adopted for any official purposes. Table 3.2 has evaluations of these methods on the same basis as that applied to the methods which produce official national sets of LMAs. As noted earlier, the lack of any adoption of the results of the method by the NSI concerned makes these methods less suitable candidates for the analyses of data on several countries later in this study.

**Table 3.2: Summary evaluation of other national methods against the issues identified**

issue:	MS:	PT	CZ	EL <sup>17</sup>	SI <sup>18</sup>	SK
1 (non-core-based)		✓	✗	✓	✓	✗
2 (exhaustive & non-overlapping)		✓	✓	✓	✓	✓
3 (explicitly defined labour markets)		✓	✓	✓	✓	✓
4a (consistency – method)		✓	✓	✓	✓	✓
4b (consistency – parameters)		?	✓	✓	✓	✓
4c (consistency – analytical)		?	✗	✓	✓	?
5 (contiguous LMAs)		✓	✓	✗	?	✓
6 (aligned to administrative boundaries)		✗	✗	✗	✗	✗
7a (clarity/simplicity)		★	★	✓	✓	?
7b (transparency/reproducible)		✓	★	✓	✓	?
8 (aggregative)		✓	✓	✓	✓	✓
9 (based on additional datasets)		✓	✓	✓	✓	✓
10a (minimum self-containment)		★	✗	✗	✗	✓
10b (SC guides the process)		✓	★	✓	✓	★
11a (minimum population)		✗	✓	✗	✗	✗
11b (maximum area/distance/time)		★	✗	✗	✗	✗
12 (non-contiguity-constrained)		✗	✓	✓	✓	✗
13 (few parameters)		✓	★	✓	✓	✗
14 (analytically set parameters)		✗	✗	✗	✗	✗

**CZ** and **SK** fall in the core-based category and can therefore be compared with the selected **SE** method. The latter has a better assessment for simplicity and clarity. Doubts about the **SK** method description (noted in the Annex) result in a better assessment for the **SE** method on primary issue 4c. On the other primary issues **SK** does have a better assessment than **SE** on issue 10a but this is probably due to the inherent size of LAU2 areas in **SE** which made a minimum size less necessary: the effect of adding this constraint to the **SE** method could be a valuable additional experiment to test in the empirical analyses. Table 3.2 also has three more ‘open system’ methods in the academic procedures cited by **PT**, **EL** and **SI**. The larger number of uncertainties about the **PT** method may be due it being under consideration by the NSI at the present time; until such uncertainties are resolved, it cannot be considered a strong candidate

<sup>17</sup> See Note 11. This same procedure, originally proposed by Prodromídís (2010) has been tested in **CY** (and as so was mentioned in the response to Q-2012).

<sup>18</sup> Of the are six papers cited, the one evaluated here is Konjar et al (2010) because it is the most distinct and is an interpretation of commuting zones in the US

for the empirical analyses later in this study. Both the other two methods (**EL** and **SI**) have at least one negative assessment in relation to the issues of primary concern here (2, 4, 5 and 10) and this then adds weight to the negative factor – in common with all the methods in Table 2 – that their results have not been adopted by their respective NSIs.

### 3.3 Towards an empirical test of ‘best practice’ in defining LMAs

While accepting that to some extent the assessments detailed above are provisional – because of the incompleteness of the available evidence – it is necessary to sketch the conclusions here on this basis. The immediate requirement is to propose the nature of the remaining research activity in this study.

The core activity will be the analysis of data for several countries by alternative methods to define LMAs. This basic outline raises the question of how the alternative methods will be evaluated. In the short term this can be simply outlined by identifying the need for a combination of simple and more sophisticated quantitative indicators such as:

- measures of the self-containment of the resulting LMAs (eg. how many are under a given threshold; how much variation across the different LMAs)
- measures of the cohesion of the resulting LMAs (nb. there is less consensus on how the integration of the constituent areas of LMAs are compared)
- more statistical indicators (eg. proportion of LAU2 areas not grouped ‘optimally’ in some sense).

Another element of the design of the final stage of the research is the choice of countries whose data will be analysed by selected LMA definition methods. There remains some uncertainty about which countries have all the necessary data available, so the recommendations at this stage can only be very general in nature. There is a strong case for choosing case studies which between them provide a good sample of the variation in territorial terms found across the EU: for example from heavily urbanised parts of the continent to more rural and perhaps peripheral circumstances (which might include island areas). On a more technical note, it may well be useful to include countries with a variety of LAU2 area sizes, because this factor can be influential in the likely effectiveness of methods. Another consideration may be to prioritise countries without existing official definitions, to maximise the new findings from the study.

The previous section of this chapter provided the basis for answering the next question in designing the next stage of this research study: which methods to apply to data for several countries. The choice was guided first by recognising that a small number of issues are of primary concern. It was also argued that there were some rather different types of method, with the proposal emerging that it would be valuable to test one method that was based on the initial selection of centres, and another one that



explores the whole dataset with an ‘open system’ approach that avoids imposing such a structure. The other factor taken into account was that methods whose results were adopted by the respective NSIs can be seen as thereby having had a positive endorsement. The combination of these factors led to the selection of the **UK** and **SE** methods as the recommended candidates’ for the empirical research.

The remaining question is whether there are alternative methods which were not cited in any of the responses to Q-2012 which should be evaluated in the remaining stage of this study. At the very least, such alternatives may provide a ‘benchmark’ against which to assess the established national methods. From this more detached viewpoint, it can be suggested that all the methods cited in Tables 3.1 and 3.2 have at least one key feature in common: they are deterministic. Thus the case for assessing a very different method from the academic research sphere rests on the value of identifying a suitable method which does not share this characteristic.

Deterministic procedures are those that systematically produce the same output for given input data, together with a fixed set of parameter values. Contrasting with deterministic methods are stochastic procedures which introduce random variables to allow sub-optimal choices to be made in the short term, as a way of reaching the global optimum solution by the end of the procedure. (It is recognised that most sets of results will fall short of true ‘optimality’ but it is expected that adopting this strategy will produce results which are closer to optimality than those of a deterministic method.) Based on substantial and very recent research in this area (Watts, 2009) and related fields (Fortunato, 2010), it is proposed that the candidate stochastic method for evaluation in the final stage of this study is a grouping evolutionary algorithm (Flórez-Revuelta et al 1998, Martínez-Bernabeu et al 2012). Such a method conducts a stochastic search procedure to perform simultaneously local and global optimisation of a within-region interaction index (based on that used in the **UK** method in fact). It is therefore proposed that this method provides an appropriate comparator to the **SE** and **UK** methods that were identified above to represent centre-based and ‘open system’ approaches respectively.

## References

- Casado-Díaz, J.M.; Martínez-Bernabeu, L. and Flórez-Revuelta, F. (2010). “Los mercados locales de trabajo españoles. Una aplicación del nuevo procedimiento británico” [“Spanish local labour markets. An application of the new British procedure”], in *La ciudad metropolitana en España: procesos urbanos en los inicios del siglo XXI* (Albertos, J.M. and Fera, J.M., ed.). Madrid. Thomson-Civitas, 275-313.
- Drobne, S., Konjar, M., and Lisec, A. (2009). “Delimitation of Functional Regions using Labour Market Approach”. Proceedings of the 10th International Symposium on Operational Research SOR '09 in Slovenia, Nova Gorica, September 23-25, 2009.
- Flórez-Revuelta, F., Casado-Díaz, J.M. and Martínez-Bernabeu, L. (2008). "An evolutionary approach to the delineation of functional areas based on travel-to-work flows". *International Journal of Automation and Computing*, vol. 5 (1), pp. 10–21.
- Fortunato, S. (2010). “Community detection in graphs”. *Physics Reports*, vol. 486, n. 3-5, pp. 75-174, Elsevier.
- Martínez-Bernabeu, L., Flórez-Revuelta, F. and Casado-Díaz, J.M. (2012). "Grouping genetic operators for the delineation of functional areas based on spatial interaction". *Expert Systems with Applications*, vol. 39 (8), pp. 6754–6766.
- Prodromídís, P. (2010). “Identifying spatial labor markets in Greece from the 2001 travel-to-work patterns”. *South-Eastern Europe Journal of Economics* 1, pp. 111-128.
- Watts, M. (2009). “Rules versus Hierarchy: An Application of Fuzzy Set Theory to the Assessment of Spatial Grouping Techniques” M. Kolehmainen et al. (Eds.): ICANNGA 2009, LNCS 5495, Springer-Verlag Berlin Heidelberg, pp. 517–526.

## ANNEX 1: COUNTRY NOTES ON DELINEATION METHODOLOGY

### Belgium's method for LMAs delineation [provisional description]

This description of the method used in Belgium is extracted from the description provided in answer to Q-2007 and from the description of the methodology attached to it.

#### Input data (bold) and indices:

- Commuting matrix  $T$ :  $T_{ij}$  is the flow (number) of workers that reside in municipality  $i$  and work in municipality  $j$ .
- $R_i$  is the summation of flows with origin in municipality  $i$  (total number of workers residing in  $i$  – it includes  $T_{ii}$ ). Similarly,  $R_A$  is the summation of flows with origin in any of the constituent municipalities of LMA  $A$ .
- $SSA_A$  is the supply-side self-containment of  $A$ , equal to  $T_{AA}/R_A$ .
- $SSD_{AB}$  is the supply-side dependence of  $A$  on  $B$ , equal to  $T_{AB}/R_A$ .

#### Parameters:

There are several values mentioned in the algorithm that could be considered parameters, namely the thresholds to classify municipalities into low (<30%), medium and high (>40%) self-containment and the ones to classify the dependencies into high (>5%) or low.

#### Algorithm:

1. Start considering each municipality as an LMA.

##### First phase:

2. Repeat:
  - 2.1. Find the pair of areas  $A$  and  $B$  that maximises  $SSD_{AB}$ .
  - 2.2. If  $A$  is a single municipality join  $A$  and  $B$ , otherwise start second phase.

##### Second phase:

3. For every remaining single-municipality area  $A$  with  $SSA_A < 30\%$  do:
  - 3.1. Find area  $B$  that maximises  $SSD_{AB}$  and area  $C$  that has the second highest  $SSD_{AC}$ .
  - 3.2. If  $B$  is a local hub (cluster of municipalities), merge  $A$  and  $B$ ,
  - 3.3. else, if  $B$  is liable to become local hub ("would have been able to access the ranks of primary hubs, had they not found themselves 'encircled' by other hubs larger than them"), wait for step 4 to see if  $B$  becomes a hub,
  - 3.4. else, merge  $A$  and  $C$  [**even if  $C$  is not a primary hub nor liable to become one?**].
  - 3.5. if maximum dependence of  $A$  is located abroad wait for similar cases [**there is only**

**one case and it is not mentioned what to do then; see step 4.e].**

4. For every remaining single-municipality area A with  $SSA_A > 40\%$  do:
  - 4.1. Find area B that maximises  $SSD_{AB}$ .
  - 4.2. For every area A that has  $SSD_{AB} > 10\%$  and  $SSD_{AX} < 10\%$  for any other area X, merge A and B [**B is a primary hub in all cases, it is unsure what would have been done otherwise**].
  - 4.3. For the remaining areas A, find the number of municipalities X that have  $SSD_{XA} > 10\%$ ; “the higher the number of strongly dependent communes, the greater our chances of having a local hub” [**no threshold value is specified**].
  - 4.4. If  $SSD_{AB} < 5\%$  but A is not a local hub “it is necessary to weigh up in a specific fashion the hub (primary or secondary) to which its workers will tend to go and from which hub the persons performing a professional activity on its territory will tend to come” [**this step is not fully explained**].
  - 4.5. If the maximum dependence of A is directed abroad, it is necessary to look for a national hub that can serve as a national anchor point for it, or decide to set up a hub oriented towards another country (only in one case).
5. For every remaining single-municipality area A with  $30\% \leq SSA_A \leq 40\%$ , perform a case by case examination of the flows and attachment to other primary or local hubs or identification as a local hub following the previous criteria [**it is not explained how to apply those criteria**].

Third phase:

6. Ad hoc comparison of the obtained delimitation with the one that would be obtained in the first phase if not stopped (full dendrogram) [**it is not specified how to perform this step**].

#### Notes:

- A. This method begins with the partial use of a regular open hierarchical clustering procedure that stops before merging two clusters (groups of communes), then the regionalisation is completed through a rule-based procedure.
- B. Several questions about the methodology description in Q-2007 where addressed in Q-2012 but no answer has been received. Therefore, we cannot be sure of the interpretation in the previous algorithm. Moreover, the methodology was described as an ongoing work and without more recent input we cannot consider this method as definitive or official.
- C. The third phase is not properly described in the methodology and it is assumed that it is a manual procedure. The same can be true for some parts of the second phase.
- D. The self-containment levels of LMAs are not directly considered so a minimum level is not enforced, although the mergers are driven by dependency.
- E. Contiguity is not considered in the algorithm, but no non-contiguous LMAs arise in the exercise.

## Germany's method for LMAs delineation

The following description of the method used in Germany is based on the answer to Q-2012 and to the subsequent request for clarifications, part of which is literally quoted below:

*To define the labour market areas, the indicator method was used. To apply the indicator method, the matrix of out-commuters and in-commuters is set up as a first step. For every Kreis (administrative district) of the area examined, the number of out-commuters into any other Kreis of the area examined is to be indicated. Based on that matrix, two matrices can be set up. The out-commuter matrix indicates for every Kreis the share of out-commuters to any other Kreis as a percentage of the persons in employment living in the Kreis concerned. The in-commuter matrix indicates the share of in-commuters from any other Kreis as a percentage of the persons employed in the Kreis concerned. Consequently, there are always four cells for any two Kreise, indicating the commuter links between the Kreise. In the indicator method, two Kreise form a labour market area if at least one of the four possible links exceeds the given critical value of commuter shares (e.g. 0.1 or 0.2).*

*To define the LMAs as part of GRW (joint task of the Federation and the Länder to improve the regional economic structure), a proposal for definition was indicated for each of two given different critical values. The first proposal for definition groups all Kreise for which one of the commuter shares exceeds the value of 0.2. The second proposal for definition is less restrictive as the critical value is reduced to 0.1. Subsequent to that analysis, the resulting definition was checked for whether it meets the following constraints:*

- *The definition of LMA borders must be identical to Kreis borders because, first, major statistical data are available only at Kreis level and, second, this enhances the political-administrative enforcement and implementation of GRW promotion activities. This also applies to the constraint that labour markets generally must not cross Land borders and that the entire territory must be covered without overlaps.*
- *The self-containment of labour markets is covered by the fact that the area's self-supply in terms of jobs should be at least 65%. At the same time, this means that not more than 35% of the persons in employment living in the area commute out to other areas.*
- *Another criterium ensuring the self-containment of the labour markets is that at least 65% of an area's persons in employment live in that area. Consequently, not more than 35% of an area's jobs may be filled by in-commuters from other areas.*
- *To ensure a certain relevance of the labour market areas, the minimum number of inhabitants is set at 100,000 people.*
- *Within a labour market area, the acceptable commuting time of 45 minutes per journey, that is,*

90 minutes to work and back, must not be exceeded.

*In cases where not all criteria can be met at the same time, a discretionary decision was taken regarding the definition.*

This information describes in detail the constraints that a valid LMA should fulfill and the general criteria that guide the aggregation process, but the operation of the aggregation process itself is not fully described.

### Input data (bold) and indices:

- Commuting matrix T:  $T_{ij}$  is the flow (number) of workers that reside in municipality i and work in municipality j.
- Distance matrix D:  $D_{ij}$  is the distance (in meters or minutes) between municipalities i and j.
- $R_i$  is the summation of flows with origin in municipality i (the total number of workers residing in i – it includes  $T_{ii}$ ).
- $J_i$  is the summation of flows with destination in municipality i (the total number of jobs in i – it includes  $T_{ii}$ ).
- $SSD_{AB} = T_{AB}/R_A$  is the supply-side dependence of A on B (proportion of residents of A employed in B).
- $DSD_{AB} = T_{AB}/J_A$  is the demand-side dependence of A on B (proportion of jobs in A hold by residents of B).
- $SSA_A = T_{AA}/R_A$  is the supply-side self-containment of A.

### Parameters:

1. minDependence: minimum supply- or demand-side dependence of a kreiss on another one to justify their merge, set to 10% or 20% (it is not stated which one is finally used, or the different uses of the LMAs resulting from both values).
2. minSC: minimum (supply- and demand-side) self-containment a LMA to be considered as valid, set to 65%.
3. minSize: minimum population size (residents) of a valid LMA, set to 100000 (according to data some official LMAs have less population).
4. maxDist: maximum distance (in time or space) between two kreiss in a LMA to be considered as valid, set to 45 minutes.

### Notes:

- It is not possible to describe the algorithm itself using the information that has been made

available. More specifically, and among other aspects, no information is provided regarding the sequence in which the kreiss are considered. They could be for example simultaneously considered for potential mergers that would affect all pairs of kreiss reaching the minimum dependence without any recalculation of the dependences at that stage, or the process could be hierarchical, and based on the merger of the pair of municipalities that maximise the dependence indicator, and a subsequent recalculation of dependences between kreiss-kreiss or kreiss-group before considering the next pair of areas.

- The description of the method mention that a “discretionary decision” is taken in the cases were not all criteria can be met. As there is no explicit criteria to perform those decisions it can be considered as a manual final stage.

## Czech Republic's method for LMAs' delineation

This description of the method used in Czech Republic is based on the answer to Q-2012 and the reply to a request for clarifications.

### Input data (bold) and indices:

- Commuting matrix T:  $T_{ij}$  is the flow (number) of workers that reside in municipality i and work in municipality j.
- Municipalities' population (inhabitants) P:  $P_i$  is the number of residents (active or not) in municipality i. Similarly,  $P_A$  is the aggregated population of LMA A.
- $N_A$  is the number of municipalities in LMA A.

### Parameters:

- minPopLMA: minimum inhabitants in the LMA, 10000.
- minPopHinter: minimum inhabitants in the hinterland of an LMA: 4000.
- minMunLMA: minimum number of municipalities per LMA: 4.

### Algorithm:

1. Identification of the dominating flow of each municipality i ( $T_{ij} \geq T_{ix}$  for every municipality x).
2. Identification of work centres: municipalities that are the destination of the dominating flow of other municipalities.
3. Assignment of non-central municipalities to their dominating work centre (forming its hinterland).
4. Identification of non-valid LMAs: areas that do not count with (a) at least 10000 inhabitants ( $P_A \geq \text{minPopLMA}$ ), (b) at least three municipalities assigned to its work centre ( $N_A \geq \text{minMunLMA}$ ) and (c) at least 4000 inhabitants in its hinterland ( $P_A - P_{\text{workcentre}(A)} \geq \text{minPopHint}$ ) are non-valid LMAs.
5. Each non-valid LMA is assigned to its own work centre's dominating work centre. In case of balance (when the number of commuters from the non-valid centre towards its diverse potential centres are of similar relevance), the decision is made considering also the flows from the non-valid centre towards its potential centres' hinterlands.
6. Application of local knowledge to resolve incoherence in the borders working over a map.

### Notes:

- A. The procedure could be described as a form of (hierarchical) clustering with a very simple linkage criterion: only the links between central elements in each cluster are considered.
- B. All assignments in each phase are performed simultaneously.



- C. The threshold values are not strict.
- D. Contiguity is not included as a requisite, although it did not appear during the empirical testing of the procedure.
- E. According to the description, the procedure is not purely computerized and it is performed manually over spreadsheets and maps.
- F. LMAs with very low self-containment levels could be created, as the aggregated flows are not considered. For the same reason, municipalities could be assigned to LMAs to which few workers commute (compared to the total number of outcommuters), despite having stronger aggregated dependence to other neighbouring LMAs (i.e. to the set of municipalities that conform them); phase 5 is intended to mitigate that, but this stage is not based on formal rules and cannot be codified.
- G. Since the assignments are performed simultaneously, a rare case not considered in the procedure could cause ambiguity in phase 4: if there are two invalid LMAs A and B, with centres  $A_c$  and  $B_c$ , where dominating flow of  $A_c$  is towards  $B_c$ , and its second dominating flow is towards another work centre  $C_c$ , while the dominating flow of  $B_c$  is towards another work centre  $D_c$ . Would area A be assigned to  $D_c$  along with area B or would it be assigned to  $C_c$ ?
- H. The values of minimum number of inhabitants in core and hinterland and number of subordinated municipalities (the criteria of validity of a LMA) could and should be treated as parameters of the algorithm.

## Estonia's method for the delineation of functional regions

The method used in Estonia is described in the partial answer to Q-2012 and a second piece of information by Dr. T. Tammaru.

Two concepts of core-based areas are considered in these documents:

- I. Spheres of influence comprise county seats (cores) and the municipalities around them that send to the core more than 15% and 30% (two delimitations) of their workers. They do not allow overlapping (apparently this was possible in previous delimitations).
- II. Labour-catchment areas comprise county seats and the municipalities that send at least 10% of their workers to the core or, if the municipality does send less than 10% to any core, to the core to which it sends more workers (regardless of any minimum threshold) or, if it is not contiguous, to the core of the municipalities that surround the one under consideration.

### Input data (bold) and indices:

- Commuting matrix  $T$ :  $T_{ij}$  is the flow (number) of workers that reside in municipality  $i$  and work in municipality  $j$ .
- Contiguity matrix  $C$ :  $C_{ij}$  is the contiguity between municipalities  $i$  and  $j$  (1 contiguous, 0 non-contiguous). Similarly,  $C_{AB}$  is the contiguity between areas  $A$  and  $B$  (1 if any municipality in  $A$  is contiguous to any municipality in  $B$ ).
- $R_i$  is the summation of flows with origin in municipality  $i$  (total number of workers that reside in  $i$ , it includes  $T_{ii}$ ).
- $SSD_{ic} = T_{ic}/R_i$  is the supply-side dependence of municipality  $i$  on core  $c$ .

### Algorithm (labour-catchment areas):

1. Consider as core every county seat (previously they used  $jobratio > 1$ ).
2. For each free municipality identify its core, the one to which it sends more workers (maximise  $SSD_{ic}$ ).
3. Attach each free municipality  $i$  to its core  $c$  if  $SSD_{ic} > 10\%$ .
4. For each remaining free municipality  $i$  (those where  $SSD_{ic} < 10\%$  for any core  $c$ ) attach it to its core if it is contiguous, or attach it to the core of the surrounding municipalities otherwise.

### Algorithm (spheres of influence):

1. Consider as core every county seat (in previous versions municipalities for which  $jobratio > 1$  were considered as cores).
2. For each free municipality identify its core, the one to which it sends more workers.
3. For each free municipality that sends more than 15% (30%) of its workers to its core, attach to

it.

**Notes:**

- A. It is a core-based clustering procedure. A core-identification step was previously used (municipalities with job ratio > 1) but lately being one of the 15 county seats is the condition to qualify as a core.
- B. Self-containment of the areas is not evaluated.
- C. All municipalities must be directly connected through commuting flows to the core of the area they are assigned (no multilink is allowed).
- D. Contiguity is enforced but the authors state it was correcting for it was rather necessary, and it is not clear if the algorithm itself considers the contiguity information in any way or if that restriction is ensured later, manually, while working with the plotted maps.
- E. The spheres of influence regionalisation is **not exhaustive**: depending on the thresholds used, many municipalities can end up unassigned.
- F. Although the maps attached to the answer to Q-2007 seemed to indicate that overlapping was permitted, the full answer to Q-2012 clarifies that overlapping is not allowed.

## Greece's unofficial method for LMAs delineation

The description of this unofficial method used in an academic exercise in Greece is based on the response to Q-2012 which included a country note. The answer was prepared by Dr. Prodromos-Ioannis Prodromidis, the author of this academic study.

### Input data (bold) and indices:

- Commuting matrix  $T$ :  $T_{ij}$  is the flow (number) of workers that reside in municipality  $i$  and work in municipality  $j$ . Similarly,  $T_{AB}$  is the summation of flows with origin in any municipality of area  $A$  and destination in any municipality of area  $B$ .
- $R_i$  is the summation of flows with origin in municipality  $i$  (the total number of workers residing in  $i$ , it includes  $T_{ii}$ ). Similarly,  $R_A$  is the summation of flows with origin in any area (group of municipalities)  $A$ .
- $J_i$  is the summation of flows with destination in municipality  $i$  (the total number of jobs in  $i$  – it includes  $T_{ii}$ ). Similarly,  $J_A$  is the summation of flows with destination in any area  $A$ .
- $SSD(A,B) = T_{AB}/R_A$  is the demand-side dependence of area  $A$  on area  $B$ .
- $DSD(A,B) = T_{BA}/J_A$  is the demand-side dependence of area  $A$  on area  $B$ .
- $MD(A) = \min(SSD(A,B), DSD(A,B))$  is the minimum dependence of area  $A$  on area  $B$ .

### Parameters:

- minDep: minimum dependence of an area onto another in order to allow their aggregation, 15% (levels 10% and 20% were also tested).

### Algorithm:

1. Start considering every territorial unit as a LMA.
2. Repeat:
  - 2.1. Find the pair of areas  $A$  and  $B$  that maximise  $MD_{AB}$ .
  - 2.2. If  $MD_{AB} \geq \text{minDep}$  merge  $A$  and  $B$  (and recalculate flows between areas); otherwise terminate.

### Notes:

- A. The method can be described as an iterative hierarchical clustering procedure.
- B. Contiguity is not imposed, but only three cases of non-contiguous LMAs appeared in their exercise.
- C. Self-containment of the defined LMAs is not directly considered. Therefore, LMAs with very low self-containment levels could arise depending on the relation between the distribution of flows in the region and the parameter minDep (for example, in a region composed of territorial units

with low autonomy levels and many relatively significant out-commuting flows it will be necessary a higher minDep to reach the same self-containment level than in the opposite scenario).

- D. The answer to Q-2012 mentions another methodology recently applied in an academic exercise. The description of this method can be found in the academic paper Kallioras, D. and Kandyliis, Y. and Kromydakis, N. and Pantazis, P. (2011) “Definition of Local Labor Market Areas in Greece on the Basis of Travel-to-Work Flows”, Univ. of Thessaly and National Centre for Social Research. The concept proposed there allows the overlapping of LMAs (one municipality can be attached to many poles).

## Finland's method of LMAs' delineation

This description of the method used in Finland derives from the answer to Q-2012.

### Input data (bold) and indices:

- Commuting matrix  $T$ :  $T_{ij}$  is the flow (number) of workers that reside in municipality  $i$  and work in municipality  $j$ .
- $R_i$  is the summation of flows with origin in  $i$ .

### Parameters:

- minAut: minimum supply-side ( $T_{ii}/R_i$ ) self-containment, 75% (for identification of central municipalities).
- maxDep: maximum supply-side dependence ( $T_{ij}/R_i$ ) from a single municipality, 10% (for identification of central municipalities).

### Algorithm:

1. Identification of central municipalities: municipality  $i$  is central if  $T_{ii}/R_i > \text{minAut}$  and  $T_{ij}/R_i < \text{maxDep}$  for every municipality  $j$ .
2. While there are non-central municipalities unassigned to a central municipality:
  - 2.1. Find the non-central municipality  $i$  that maximises  $T_{ij}/R_i$  for every central municipality  $j$ .
  - 2.2. Assign  $i$  to  $j$  (that is, assign  $i$  to its dominating central municipality).

### Notes:

- A. It is a core-based hierarchical clustering procedure with a very simple linkage criterion: only the links between central elements of each cluster are considered.
- B. Contiguity is not a restriction and non-contiguous LMAs are accepted (only a few cases in practice).
- C. LMAs do not have to fulfil any criteria apart from having a central municipality and a hinterland (no minimum/maximum levels for self-containment, population or area).
- D. LMAs with very low self-containment could be created, as the aggregated flows are not considered. For the same reason, municipalities could be assigned to LMAs to which few total workers commute, despite having stronger aggregated dependence to other neighbouring LMAs.

## France's method for LMAs delineation

This description of the method used in France is based on the answer to Q-2012, the description of the method attached to the questionnaire, the answer to a request of clarifications, and another method's description as a diagram (it will be referred to as “Diagram” in this text).

Some doubts about the method still remain (see numbered notes).

### Input data (bold) and indices:

- Commuting matrix T:  $T_{ij}$  is the flow (number) of workers that reside in municipality  $i$  and work in municipality  $j$ . Similarly,  $T_{AB}$  is the aggregation of flows with origin in any of the constituent municipalities of area A and destination in a municipality of area B.
- Adjacency matrix C:  $C_{ij}$  is the contiguity between municipalities  $i$  and  $j$  (1 contiguous, 0 non-contiguous). Similarly,  $C_{AB}$  is the contiguity between areas A and B (1 if any municipality in A is contiguous to any municipality in B).
- Distance matrix D:  $D_{ij}$  is the distance (in meters or minutes) between municipalities  $i$  and  $j$ . Similarly,  $D_{AB}$  is the average distance between municipalities in areas A and B.
- $R_i$  is the summation of flows with origin in municipality  $i$  (the total number of workers residing in  $i$  – it includes  $T_{ii}$ ). Similarly,  $R_A$  is the summation of flows with origin in any of the constituent municipalities of LMA A.
- $J_i$  is the summation of flows with destination in municipality  $i$  (the total number of jobs in  $i$  – it includes  $T_{ii}$ ). Similarly,  $J_A$  is the summation of flows with destination in any municipality of LMA A.
- $SSD_{AB}$  = supply-side dependence  $T_{AB}/R_A$  (proportion of residents of A employed in B).

### Parameters: (see note A)

- minLink: minimum proportion of residents in the satellite area that work in the pole area needed in order to allow its aggregation (usually less than 1%).
- minSize: minimum population size (jobs) of a valid LMA (2).
- maxDist: maximum distance (in time or space) between two areas (measured as the average between municipalities in one area and municipalities in the other) to allow its aggregation [this restriction appears in the Diagram, but not in the texts that describe the method].

### Algorithm:

1. Start considering every single municipality as a LMA.
2. Identify the pair of areas A and B with the highest  $SSD_{AB}$ , where A is a non-pole area ( $[J_A < \text{minSize}]$  (see note B)) and B is any of its adjacent ( $C_{AB}=1$ ), close enough ( $D_{AB} < \text{maxDist}$ ), neighbouring [pole] (see note C) areas.

3. If  $\text{link}_{AB} > \text{minLink}$ , merge areas A and B (recalculation of T, C and D) and go to step 2.  
Otherwise finish.

**Notes:**

- A. There is a general case algorithm, and major adaptations are made to fit diverse territorial realities (*grandes communes, territoires proches des frontières, zones se situant sur plusieurs régions, zones enclaves, poles isolés*) with the aim of for example avoiding too wide or too small LMAs. The associated criteria/values are not well specified; a general rule on how the decision of adjusting them is taken based on quantifiable territorial data is not made explicit.
- B. There is a contradiction about the variable used to measure minSize in the different descriptions of the methodology:
  - a. From description attached to the answer to Q-2012: “A chaque étape d’agrégation, on vérifie, en outre, la taille (en **nombre d’emplois**) de chaque unité (commune ou groupe de commune). Si celle-ci dépasse un seuil (appelé paramètre d’isolation et comptabilisé en termes d’**emplois**), elle devient pôle”.
  - b. From description in the Diagram: “Paramètre d’isolation: toute zone ayant une **population active au lieu de résidence** supérieur à ce seuil devient automatiquement pôle et ne pourra ainsi pas devenir satellite d’une autre zone”.
  - c. “**Population active au lieu de résidence**” would be  $R_A$ , while “**emplois**” is  $J_A$ .
- C. The answers to the request of clarifications state that a pole area cannot be attracted by another pole area, but in the texts describing the procedure it is not specified that a non-pole area cannot attract another non-pole area when searching for the more intense link. However, at the end of the Diagram it is stated that both poles and satellite areas can “if possible” be outputs of the search of the highest link, so it is not clear whether this is a core-based procedure (only areas formed by municipalities identified as poles from the beginning can attract other municipalities) or not (aggregation of non-pole areas is possible and they can reach the status of pole area, and consequently become an independent LMA). The first interpretation is more likely after analysing all the information available. The notes in the Diagram about importance of a proper adjustment of the parameter minSize to avoid the creation of too small or too big areas reinforce that assumption.
- D. Taking as correct the previously mentioned assumptions, this procedure could be described as a hierarchical core-based clustering procedure, with contiguity and minimum interaction restrictions for the aggregations and where the cores are determined by their size in jobs.
- E. A final phase of manual adjustments to the delimitation can be performed after a round of consultations to the local authorities.



## Italy's method for LMAs delineation

This description of the method used in Italy is extracted from the description provided in answer to Q-2007, confirmed in answer to Q-2012.

### Input data (bold) and indices:

- Commuting matrix T:  $T_{ij}$  is the flow (number) of workers that reside in municipality  $i$  and work in municipality  $j$ .
- $R_i$  is the summation of flows with origin in municipality  $i$  (the total number of workers residing in  $i$ , it includes  $T_{ii}$ ). Similarly,  $R_A$  is the summation of flows with origin in any municipality of LMA  $A$ .
- $J_i$  is the summation of flows with destination in municipality  $i$  (the total number of jobs in  $i$  – it includes  $T_{ii}$ ). Similarly,  $J_A$  is the summation of flows with destination in any municipality of LMA  $A$ .
- $SSA_A = T_{AA}/R_A$ , is the supply-side self-containment of area  $A$ .
- $DSA_A = T_{AA}/J_A$ , is the demand-side self-containment of area  $A$ .
- $MSC_A$  is the minimum self-containment of an area  $A$ , equal to  $\min(SSA_A, DSA_A)$
- $SSD_{AB} = T_{AB}/R_A$  is the supply-side dependence of area  $A$  on area  $B$ .
- $\text{job ratio}_A = (J_A - T_{AA}) / (R_A - T_{AA})$ .
- $\text{interaction}_{iA} = T_{iA}^2 / (R_i + J_A) + T_A^2 / (R_A + J_i)$
- $\text{validity}_A$  is the equation that measures how close is an area to be a valid LMA. It is calculated as  $\min(MSC_A / \text{minSC}, 1) * \min(J_A / \text{minJob}, 1)$ .

### Parameters:

- **minSCcores**: Minimum self-containment of the proto-cores, 50%.
- **minSC**: Minimum self-containment of the LMAs, 75%.
- **minJob**: Minimum size in number of jobs, 1000.
- **minDep1**: Minimum dependence of a municipality on a proto-core to consider its merger in step 2: 10%.
- **minDep2**: Minimum dependence of a proto-core on a municipality to consider its merger in step 2: 1%.
- **minInteraction**: Minimum interaction 0.002 (referred to as 0.2% in the source description).

### Algorithm:

Start considering each municipality as an independent area.

1. Calculate *job ratio* and SSA for every municipality and consider as proto-cores the 20% of municipalities with higher *job ratio* and the 20% of municipalities with higher SSA.
2. For each proto-core A, in decreasing order of  $J_A-T_{AA}$ , do:
  - 2.3. While  $MSC_A < 50\%$  do:
    - 2.1.1 Find all the non-core municipalities i for which ( $SSD_{iA} > \text{minDep1}$ ) and ( $SSD_{Ai} > \text{minDep2}$ ) and ( $\text{interaction}_{iA} > \text{minInteraction}$ ).
    - 2.1.2 Merge A and the municipality i from previous step (if any, otherwise end this loop) that maximises  $\text{interaction}_{iA}$  and recalculate  $MSC_A$ .
3. For every locality A (free municipality, core or multi-core), in decreasing order of  $\text{validity}_A$ , do:
  - 3.1 While  $\text{validity}_A < 1$  do:
    - 3.1.1 Find all the localities X for which  $SSD_{XA} > \text{minDep1}$ .
    - 3.1.2 Merge A and municipality X from previous step (if any, otherwise end this loop) that maximises  $\text{interaction}_{XA}$ .
4. For every municipality A where  $\text{validity}_A < 1$ , in decreasing order of  $J_A-T_{AA}$ , do:
  - 4.1. Merge A with the proto-LMA X (locality where  $\text{validity}_{XA} \geq 1$ , if any) that maximises  $\text{interaction}_{AX}$ .
5. For every proto-LMA A, in increasing order of  $MSC_A$ , do:
  - 5.1. If  $MSC_A < \text{minSC}$ , dismember A into its constituent municipalities and do:
    - 5.1.1. While there are any other free municipalities do:
      - 5.1.1.1. For each free municipality i, in order of  $J_i$ , assign i to the proto-LMA A that maximises  $\text{interaction}_{iA}$ .
6. (Optional) fine-tuning of the borders of the resulting LMAs.

### Notes:

- A. It is a core-based non-hierarchical agglomerative clustering procedure (also called rule-based) with an elaborated linkage criteria.
- B. A minimum level of self-containment ( $\text{minSC}$ ) is enforced for all the resulting LMAs.
- C. Contiguity is not considered in the algorithm, but in the answer to the questionnaire it is stated that there is a treatment for non-contiguous areas, although it is not described and the **related question included in the country note of Q-2012 was not answered**. It could be assumed that non-contiguous LMAs are altered in the final (**manual?**) phase to produce fully contiguous LMAs that still meet the statistical criteria.
- D. The algorithmic structure of this methodology is not fully represented in the source description. Although there are no doubts in the rest of the previous steps, the loop structure in step 5 of the

previous algorithm is deducted, based on previous experience. The chosen structure is intended to cover the case where a dismembered municipality  $j$ , that only has interaction with other municipalities resulting from the dismembering of the former proto-LMA to which all of them belonged, is evaluated to find its (new) attracting proto-LMA before any of its functional neighbours have been assigned, therefore  $j$  cannot find an attracting LMA. As it is written, those residual municipalities are re-evaluated once the list of municipalities resulting from dismembering has been exhausted, and this sub-step is repeated until no residual municipalities remain.

## Portugal's unofficial method for LMAs delineation

This method is used in an academic exercise in Portugal (by Mr. Pereira) and was presented to its NSI for its consideration (but still not official). Its description is included in the response to Q-2012 and in the answer to a request for clarifications.

### Input data (bold) and indices:

- Commuting matrix  $T$ :  $T_{ij}$  is the flow (number) of workers that reside in municipality  $i$  and work in municipality  $j$ . Similarly,  $T_{AB}$  is the summation of flows with origin in any municipality of area (aggregation of municipalities)  $A$  and destination in any municipality of area  $B$ .
- Adjacency matrix  $C$ :  $C_{ij}$  describes the contiguity between municipalities  $i$  and  $j$  (1 contiguous, 0 non-contiguous). Similarly,  $C_{AB}$  is the contiguity between areas  $A$  and  $B$  (1 if any municipality in  $A$  is contiguous to any municipality in  $B$ , 0 otherwise).
- $R_i$  (residents) is the summation of flows with origin in  $i$  (including  $T_{ii}$ ). Similarly,  $R_A$  is the summation of flows with origin in any municipality of area  $A$ .
- $J_i$  (jobs) is the summation of flows with destination in  $i$  (including  $T_{ii}$ ). Similarly,  $J_A$  is the summation of flows with destination in any municipality of area  $A$ .
- $S_i$  is the surface area of municipality  $i$ . Similarly,  $S_A$  is the summation of surfaces of all the municipalities that integrate area  $A$ .
- $\text{interaction}_{AB}$  is equal to the ratio of the (aggregated flow between areas  $A$  and  $B$ ) and (the summation of both resident working populations),  
$$\text{interaction}_{AB} = (T_{AB} + T_{BA}) / (R_A + R_B) = \text{interaction}_{BA}$$
- $\text{SSA}_A$  is the supply-side self-containment of area  $A$ , calculated as  $T_{AA} / R_A$ .
- $\text{DSA}_A$  is the demand-side self-containment of area  $A$ , calculated as  $T_{AA} / J_A$ .

### Parameters:

- **minSC**: minimum self-containment (supply- or demand-side) of the resulting LMA required to accept a merger: 85% (other values were also tested: 70%, 75% and 80%).
- **maxSurface**: maximum surface area of the resulting LMA required to accept a merger: 6,000km<sup>2</sup>

### Algorithm:

1. Start considering every single municipality as a LMA.
2. Repeat until no mergers are possible:
  - 2.1. Find the pair of contiguous LMAs  $A$  and  $B$  (so that  $C_{AB}=1$ ) that maximise  $\text{interaction}_{AB}$ .

- 2.2. Merge LMAs A and B if (a) the combined surface area is not over the established threshold ( $S_A + S_B < \text{maxSurface}$ ) and (b) self-containment is over the established threshold ( $\max(SS_{AA}, DS_{AA}) \geq \text{minSC}$ ).
3. Repeat until no isolated municipalities remain:
  - 3.1. Find the pair of isolated municipality  $i$  and LMA A, contiguous ( $C_{iA}=1$ ) that maximise interaction $_{iA}$ .
  - 3.2. Merge  $i$  and A.

**Notes:**

- A. The method can be described as an open (not core-based) iterative hierarchical clustering procedure.
- B. Contiguity is enforced during the whole process.
- C. A minimum level of self-containment of the defined LMAs is directly considered a condition during the main aggregation phase, but it is not enforced in the last phase of the algorithm. Therefore, areas with self-containment lower than the parameter threshold could arise.
- D. A maximum threshold of surface area is specified during the main aggregation phase, but it is not enforced in the last stage of the algorithm. Therefore, LMAs with greater surfaces could arise.
- E. It seems controversial that the merger of two LMAs is not allowed if the joint self-containment does not surpass a minimum, regardless of their intensity of interaction and the self-containment of the separated areas. This would imply, in a territory composed by many municipalities with low self-containment levels, that the phase of aggregation of areas (step 2) would be ignored. Therefore all the clustering would be performed in the last phase, where the conditions of maximum surface area and minimum self-containment levels are not considered, so the resulting LMAs' characteristics could potentially be far from the ones envisaged.
- F. A different version of this algorithm exists in a previous communication by the authors: "Para uma discussão de Regiões Urbanas Funcionais em Portugal". In that document the conditions are slightly different.
- G. These methods are based on the previous work by Pereira (1997) "Bacias de Emprego em Portugal Continental", *Revista de Estatística (INE)*, No 4, 1<sup>st</sup> semester.

## Sweden's method of LMA delineation

This description of the method used in Sweden is extracted from the answer to Q-2012.

### Input data (bold) and indices:

- Commuting matrix **T**:  $T_{ij}$  is the flow (number) of workers that reside in municipality *i* and work in municipality *j*.
- $R_i$  is the summation of flows with origin in *i*.

### Parameters:

- **minAut**: minimum supply-side self-containment ( $T_{ii}/R_i$ ), 80% (for being considered as a central municipality).
- **maxDep**: maximum (supply-side) dependence in relation to a single municipality ( $T_{ij}/R_i$ ), 7.5% (for being considered as a central municipality).

### Algorithm:

1. *Identification of central municipalities*: municipality *i* is central if  $T_{ii}/R_i \geq \text{minAut}$  and  $T_{ij}/R_i < \text{maxDep}$  for every municipality *j*.
2. Two municipalities *i* and *j* that do not fulfill the conditions to be central but have their largest out-commuting flow directed to each other (that is,  $T_{ij} \geq T_{ix}$  for every municipality *x* and  $T_{ji} \geq T_{jy}$  for every municipality *y*) give place to a combined central locality.
3. Every non-central locality is assigned to the municipality that acts as the main destination of its largest out-commuting flow. Municipalities are then classified in the following classes:
  - Type-11: central localities.
  - Type-20: municipalities whose largest out-commuting flow is directed to a type-11 locality.
  - Type-30: municipalities whose largest out-commuting flow is directed to a type-20 locality.
  - Type-50: municipalities whose largest out-commuting flow is directed to a type-30 locality.
4. Every type-11 locality and its assigned municipalities form a LMA.

**Notes:**

- A. It is a core-based hierarchical clustering procedure with a very simple linkage criterion: only the flows between central elements of each cluster are considered.
- B. Although a contiguity restriction is not considered it does not appear in their exercise.
- C. Each LMA that has type-30 (or type-50) municipalities or a type-11 combined locality is considered polycentric (the extra centres are the type-20 localities in the LMA with type-30 localities assigned to them).
- D. LMAs do not have to fulfill any criteria apart from having a central municipality (no minimum/maximum self-containment, population or area levels).
- E. LMAs characterized by very low self-containment levels could be created in the process, as the aggregated flows are not considered. For the same reason, municipalities could be assigned to LMAs towards which few (compared to the total number of commuters) workers commute, despite having stronger aggregated dependence to other neighbouring LMAs (if the main destination municipality for out-commuters from municipality A is in LMA X and the rest of commuters go to municipalities belonging to LMA Y, municipality A would be assigned to X although it is likely that the total number of commuters to Y exceeds the number of commuters to X).



## **Slovenia's unofficial method for LMAs' delineation (Drobne and others)**

In Slovenia there is not an official delineation of functional areas or LMAs. However in the national answer to Q-2012 a set of recent academic studies that deal with these issues was attached. What follows is a short review of those papers, ordered by year of publication

### **S. Drobne, M. Konjar and A. Lisec (2009) Delimitation of Functional Regions using Labour Market Approach.**

The so called Labour Market approach is explained in this paper. It is taken from Karlsson and Olsson (2006), who did not mention any other author for the LM approach but quote the Swedish Institute of Statistics SCB's method (SCB, 1992) as their inspiration (this is the same method reported by SCB in their answer to Q-2012). What follows is a brief description of the algorithm.

#### **Algorithm:**

1. Identifications of local centres (minimum supply-side self-containment and minimum number of jobs)
2. Assignment of non-centres to centres:
  - 2.1. non-centres which have their largest outgoing flow towards a centre are assigned to them
  - 2.2. non-centres which have their largest outgoing flow towards a non-centre already assigned to a centre are assigned to that center
  - 2.3. pairs of non-centres whose largest outgoing flow are directed to each other are assigned to their (**combined?**) second largest flow.

Therefore, it is a core-based hierarchical clustering procedure that results in a map of non-overlapping LMAs that exhausts the whole territory.

### **M. Konjar, A. Lisec and S. Drobne (2010) Methods for delineation of functional regions using data on commuters.**

Three approaches, so called, LM, CZ and the suggested Commuting Aggregation, are explained in this paper.

For LM approach check comments on previous paper.

The CZ (commuting zones) approach described in this work also cites Karlsson and Olsson (2006) as the source. It must be noted that the procedure described in Konjar et al. (2010) is not the same that is referred to in Karlsson and Olsson (2006), where it is not fully described since they forward to the source (Killian and Tolbert, 1993). The algorithm that Konjar et al. describe as CZ does not exactly correspond to the original<sup>19</sup> commuting zones approach by Tolbert and Sizer (1987, 1996) and Sizer and Tolbert (1993). Instead, it is extremely similar to the already described LM approach: it uses the same core identification criteria, both are hierarchical clustering (the dismembering and reassignment of the constituent municipalities of a proto-LMA is not considered) and the measure of interaction that drives the aggregation process is calculated over non-aggregated flows (from the free municipality under consideration to the core of the candidate LMA, but not to the whole LMA). The only difference between the first and the second approaches described in this article (LM and so-called CZ) refers to the interaction index used to measure the links between two nodes: in the so-called CZ approach flows in both are considered ( $T_{ij}+T_{ji}$ ) but only the working population of the smallest municipality is considered in the denominator ( $\min(R_i,R_j)$ ). In contrast, the original CZ methodology from Tolbert and Sizer is an open system (not-core-based) and uses average linkage (instead of satellite-core linkage).

Finally, the so-called Commuting Aggregation approach proposed here as a new contribution is a variation of the previous one (called CZ in this article), but with an open (not core-based) system and considering aggregated flows. That is, it is more close to the original Tolbert and Sizer's original CZs approach than the previous method, although it does not use the same linkage criterion. It is also very similar to the academic exercises by Podromidis in Greece and Pereira in Portugal, with the difference in the interaction index employed (the former consider the sizes of both areas in the denominator).

#### **S. Drobne and M. Bogataj (2012, accepted) A Method to define the number of functional regions: an application tu NUTS 2 and NUTS 3 levels in Slovenia .**

They apply Intramax method (“Flowmap” software more exactly) to Slovenia, and analyse 29 alternative maps resulting from dividing the territory into (2-30) functional regions for each year and level of aggregation (NUTS 2 and 3). That analysis is then used to develop criteria that could guide the election of the number of functional regions, considering (a) the demographic criterion of the EU guidelines for the size of the region at NUTS 2 and NUTS 3 levels, and (b) the criterion on economic homogeneity of regions.

---

<sup>19</sup> This approach is fully described in the summary of Karlsson and Olsson (2006).

They use a weighted equation of the coefficient of variation of average monthly gross earnings per capita between functional regions and the coefficient of deviation of population in the region regarding the EU guidelines, measured for each (year/NUTS' level) regionalisation. According to their results more stress on homogeneity results in a greater number of (smaller) regions compared to the alternative considered: giving more importance to be in the centre of the NUTS population orientative thresholds.

The procedure used to produce the maps of LMAs is Intramax (Masser and Brown, 1975, 1977), and it is not described in this paper. It is a hierarchical aggregation procedure that focuses on the relative strength of interactions, once the effect of size variation in row and columns totals is removed, through the comparison of the observed flow and the expected value that is derived from the multiplication of the corresponding row and column totals, when the matrix is standardised to sum to unity.

Intramax's algorithm:

1. Transform the commuting matrix T into the standardised one S:

$$S_{ii} = \frac{T_{ii}}{\sum_j R_j}$$

2. Start considering every single municipality as an area.
3. Repeat:
  - 3.1. Find the pair of areas A and B that maximise linkage<sub>AB</sub>.
  - 3.2. If linkage<sub>AB</sub> ≥ minThreshold merge together A and B; otherwise terminate.

In this case, linkage<sub>AB</sub> is calculated as follows:

$$linkage_{AB} = \frac{S_{AB}}{\sum_P S_{PB} \sum_Q S_{AQ}} \square \frac{S_{BA}}{\sum_P S_{PA} \sum_Q S_{BQ}}$$

**S. Drobne, A. Lisec, M. Konjar, A. Zavodnik Lamovšek and A. Pogačnik (2009) Functional vs. Administrative Regions – Case of Slovenia**

This paper presents a discussion about three proposals of administrative division in 2, 6 and 8 provinces based on the LM approach.

**S. Drobne, M. Konjar, A. Lisec, N. Pichler Milanović and A. Zavodnik Lamovšek (2010) Functional Regions Defined by Urban Centres of (Inter)National Importance – The Case of Slovenia**

This paper presents a general discussion on urban and functional regions (using LM approach) in Slovenia.

## Republic of Slovakia's method for LMAs delineation

This description of the method used in the Republic of Slovakia is extracted from the answer to Q-2012 and from the answer to a request for clarifications.

A number of ambiguities remain (in the answer to the request for clarification it is stated that “Mentioned material ‘LMA’ was elaborated ten years ago. The work on LMA was definitively finalized at that time. Person working on it does not work for our Office. We are not able to answer your questions”).

### Input data (bold) and indices:

- Commuting matrix  $T$ :  $T_{ij}$  is the flow (number) of workers that reside in municipality  $i$  and work in municipality  $j$ .
- Adjacency matrix  $C$ :  $C_{ij}$  is the contiguity between municipalities  $i$  and  $j$  (1 contiguous, 0 non-contiguous). Similarly,  $C_{AB}$  is the contiguity between areas  $A$  and  $B$  (1 if any municipality in  $A$  is contiguous to any municipality in  $B$ ).
- $R_i$  is the summation of flows with origin in municipality  $i$  (the total number of workers that reside in  $i$  – it includes  $T_{ii}$ ). Similarly,  $R_A$  is the summation of flows with origin in any municipality of LMA  $A$ .
- $J_i$  is the summation of flows with destination in municipality  $i$  (the total number of jobs in  $i$  – it includes  $T_{ii}$ ). Similarly,  $J_A$  is the summation of flows with destination in any municipality of LMA  $A$ .
- $core_A$  is the core municipality of area  $A$ .
- $hinter_A$  is the combination of non-core (subordinated) municipalities of area  $A$ .
- A municipality  $i$  is subordinated to core municipality  $j$  if  $i$  sends to  $j$  more workers than to any other core.
- $NSM_A$  is the number of subordinated municipalities (non-core) of area  $A$ .
- $SSA_A = T_{AA}/R_A$  is the supply-side self-containment of  $A$ .
- $jobratio_A = J_A/R_A$ .
- $extwork_A = (J_A - T_{AA})/R_A$  is the proportion of jobs held by residents from other areas.
- $closeness_{SA} = W_A/(O_A + I_A)$

where:  $W_A$  is the summation of flows with origin in any municipality of  $A$  and destination in its core;  $O_A$  is the summation of flows with origin in any municipality of  $A$  and destination in any core outside  $A$ ;  $I_A$  is the summation of flows with origin in any municipality outside  $A$  and destination in the core of  $A$ .

### Parameters:

- minSSAcore: minimum supply-side self-containment level that a municipality must meet to become core.
- minJR: minimum job ratio that a municipality must meet to become a core.
- minNSM: minimum number of subordinated municipalities that a municipality needs to become a core.
- minJF: minimum ratio of foreign job for a municipality to be a core.
- minSSA: minimum supply-side self-containment level for a valid LMA.
- minCloseness: minimum closeness for a valid LMA.

### Algorithm:

1. Consider as core every municipality  $i$  that fulfills the following conditions:
  - 1.1.  $jobratio_i > minJR$  and ( $SSA_i > minSSA$  or  $extwork_i > minJF$ )
2. For each non-core municipality identify its core (the core that receives the largest outgoing flow from that municipality to any core).
3. While there are unassigned non-core municipalities, do:
  - 3.1. Assign all the non-core municipalities to their cores if the non-core municipality is contiguous to the core's area.
4. For each LMA  $A$  where  $NSM_A < minNSM$ , merge  $A$  with the other core that exerts more attraction over  $A$ . **[sic]**. Non-contiguities are solved individually by opinion of experts on the basis of cadastral maps **[sic]**.
5. For each LMA  $A$  where  $closeness_A < minCloseness$  or  $SSA_A < minSSA$  do:
  - 5.1. Dismember  $A$  into its constituent municipalities and identify their (new) cores.
  - 5.2. While there are unassigned municipalities, assign all the non-core municipalities to their cores if adjacent to the core's area.

### Notes:

- A. The procedure can be described as non-hierarchical rule-based clustering, with some similitude with the one used in the TTWAs' revision based on the 1991 Census and the one currently in use in Italy.
- B. A minimum level of self-containment is enforced for all the LMAs defined.
- C. Contiguity is enforced during the whole process, but not necessarily in a systematic way (see notes below).
- D. There are several issues that remain unclear, mainly:
  - a. We assume that the assignments (steps 3.a and 5.b.1) are performed in parallel, since no order for a sequential assignment is specified (an explicit question on this issue was included in both Q-2012 and the request for clarifications).

- b. Step 4 of the previous algorithm (merger of LMAs with less than six municipalities) is only mentioned in answer to the last request for clarifications. In the initial description in answer to Q-2012, that condition (the number of subordinated municipalities of an LMA) was mentioned along with the conditions to identify cores, and no merger of whole LMAs (instead of dismemberment and reassignment of separated municipalities) was mentioned. Moreover, in the answer to the request for clarifications it is stated that minNSM is not a condition for core identification.
- c. Step 4 is not properly described and it cannot be reproduced without ambiguity: How is the attraction between a pair of LMAs measured? Alternative options are: from the core of the invalid LMA to the cores of contiguous LMAs, but also from the aggregated LMA to the neighbouring cores or aggregated LMAs. How is discontinuity treated?
- d. The last two steps described in the answer to Q-2012 have been removed in the answer to the request for clarifications (it is stated there that “Points 4 and 5 were not used in the framework of regionalisation”). From those two steps, the first one seemed useless, as the criteria stated for it was the same as in the previous step, so it could be arguable that no unassigned municipalities would reach step 5. Step 5 is similar to what it is done in step 3 (dismemberment and reassignment). Maybe it was a misunderstanding and someone labelled as steps 4 and 5 what were extended explanations of step 3; but, as mentioned in the first note, the five-steps version of the algorithm is very similar to the five-step algorithm used in the TTWAs’ revision after Census 1991 and the one used in Italy.

## United Kingdom's method for LMAs (TTWAs) delineation

This description of the method used in United Kingdom is extracted from the description provided in answer to Q-2007, confirmed in answer to Q-2012 and from the description of the methodology available online.

### Input data (bold) and indices:

- Commuting matrix T:  $T_{ij}$  is the flow (number) of workers that reside in municipality i and work in municipality j.
- $R_i$  is the summation of flows with origin in municipality i (the total number of workers residing in i, it includes  $T_{ii}$ ). Similarly,  $R_A$  is the summation of flows with origin in any of the constituent municipalities of LMA A.
- $J_i$  is the summation of flows with destination in municipality i (the total number of jobs in i – it includes  $T_{ii}$ ). Similarly,  $J_A$  is the summation of flows with destination in any of the constituent municipalities of LMA A.
- $SSA_A = T_{AA}/R_A$ , is the supply-side self-containment of area A.
- $DSA_A = T_{AA}/J_A$ , is the demand-side self-containment of area A.
- $MSC_A$  is the minimum self-containment level of an area A, equal to  $\min(T_{AA}/R_A, T_{AA}/J_A)$
- $interaction_{iA} = T_{iA}^2/(R_i+J_A) + T_{Ai}^2/(R_A+J_i)$
- $validity_A$  is the equation that measures how close is an area of being a valid LMA. It is calculated as follows:
  - $validity_A=1$  if  $R_A \geq \minSize1$  and  $MSC_A \geq \minSC1$  and  $MSC_A \geq M * tam + N$
  - $validity_A = \min(R_A/\minSize2, 1) * \min(MSC_A/\minSC2, 1)$  otherwise,
 where  $M = (\minSC2 - \minSC1) / (\minSize1 - \minSize2)$ ; and  $N = (\minSC1 * \minSize1 - \minSC2 * \minSize2) / (\minSize1 - \minSize2)$

### Parameters:

- $\minSC1$  and  $\minSC2$ : relaxed and strict minimum (supply- and demand-side self-containments) of a valid LMA,  $\minSC1 < \minSC2$ .
- $\minSize1$  and  $\minSize2$ : relaxed and strict minimum population size (number of workers residing in i) of a valid LMA,  $\minSize1 < \minSize2$ .



**Algorithm:**

1. Start considering every municipality as a LMA and calculate its validity.
2. While there are invalid LMAs ( $\text{validity}_A < 1$ ) do:
  - 2.1. Find the LMA **A** with lowest validity index and dismember it into its constituent municipalities.
  - 2.2. While there are dismembered municipalities unassigned do:
    - 2.2.1. For each municipality **i** in **A** find the LMA  $X_i$  that maximises interaction  $i_X$ .
    - 2.2.2. Assign simultaneously every municipality **i** to its dominating LMA  $X_i$ .
  - 2.3. Recalculate validity for every modified LMA.

**Notes:**

- A. A minimum level of self-containment is enforced for all the resulting LMAs.
- B. Contiguity is not considered in the algorithm but non-contiguous LMAs are altered in a final (manual) phase to produce fully contiguous LMAs that still meet the statistical criteria.

## ANNEX 2: TEMPLATE QUESTIONNAIRE ON LMAs

*The brief questionnaire that follows was an Action agreed by Member States at the Working Party on Regional Statistics and Rural Development (4-5 October 2011): please complete it on behalf of your Member State.*

### Introduction

Many countries now delineate labour market areas (LMAs). These sets of functional areas are seen as an alternative to local and regional administrative areas for some statistical purposes, and for the design, implementation and monitoring of labour market and other public policies. What makes LMAs attractive to policy-makers, economists and researchers is that they reflect the spatial pattern of economic and social relations in the area.

To summarise: LMAs group neighbouring ‘building block’ areas (eg. at LAU2 level) that are:

5. defined to be self-contained, so that little human interaction crosses their boundaries;
6. delineated so that commuting between homes and workplaces is mostly internalised;
7. defined for the whole country, which is covered by LMAs with no ‘gaps’ or overlaps.

Two decades ago Eurostat and Newcastle University carried out a study of principles upon which to base definitions of LMAs to be used in a policy context. Then ten years ago OECD undertook a review of the LMA definitions in its member countries. Five years ago DG REGIO expressed a strong interest in LMA statistics and Eurostat carried out a survey of the LMA definitions of Member States that same year (2007). Now this new questionnaire will provide evidence in a study to explore the feasibility of common definitions of LMAs for the entire EU.

This questionnaire is part of a larger study whose objectives are to:

- (1) outline the state-of-art in applied research on LMAs;
- (2) compare the LMA concepts recognised, and implemented, in each Member State;
- (3) draw conclusions on relevant best practice;
- (4) explore the added value of a common definition of LMA for the entire EU; and finally
- (5) identify possible ways and means of harmonising LMA definitions across the EU.

The questionnaire aims at compile the fundamental evidence needed to meet objective (2). Member States are asked to kindly collaborate by completing this questionnaire by supplying the requested information on LMA geography in their country.

### Guidance on completing the questionnaire

As mentioned above, Eurostat sent a related questionnaire in 2007 to Member States NSOs. This new and revised Questionnaire (hereafter Q-2012) derives from its predecessor (hereafter Q-2007), and its objectives are to:

- identify the progresses made at national level in the definition of LMAs (if any),
- gather more information on the availability of commuting data (used for the delineation of LMAs) at national level, and on NSOs positions towards an LMA geography across the entire EU territory.

The respondents are invited to kindly fill-in the questionnaire even if the concept of LMA is not defined in their country, nor they have plans to create LMAs.

For some questions, specific “notes” are formulated, which may also be interpreted as (partially) prefilled-in answers. These notes are based on the information already identified by Eurostat. Please consider these notes before answering the questions.

We hope that this approach will help you to complete the questionnaire quickly and without much difficulty, but if necessary you can contact the Eurostat officer responsible for the study:

**Mr. Oliver HEIDEN**

Eurostat, UNIT E4

Regional Statistics and Geographical Information

BECH A3/049

[Oliver.Heiden@ec.europa.eu](mailto:Oliver.Heiden@ec.europa.eu)

Tel: (+352) 43 01 31 961

Fax: (+352) 43 01 34 029

**THANK YOU VERY MUCH FOR YOUR COLLABORATION**

## 2012 Questionnaire on Labour Market Areas in EU Member States

**Country:**

**Name of respondent Q-2007:**

**Date of response Q-2007:**

**Name of respondent Q-2012:**

**Date of response Q-2012:**

### **Section A - Existing delineation methods**

**A - Q.1** In your country, do you use the concept of Labour Market Areas, i.e. have LMAs been defined (either by your office or by another organisation in your country)?

**A - Q.2** If the new answer is no, do you have any plans in the future to create LMAs, or are there existing boundaries, e.g. of administrative areas, that you consider adequately represent economic regions? If so, what are they?

**A - Q.3** Which is the national concept (i.e. sistemi locale del lavoro (IT), zones d'emploi (FR), etc.) of these areas and what definition do you apply? If the definition is available on Internet, please send the link.

**A - Q.4** There may be several alternative sets of LMAs definitions in your country; if so, please answer the questionnaire in relation to the set that is in the highest of the following list:

1. LMAs used for official statistics
2. LMAs used for some other government purpose (e.g. planning)
3. LMAs used only for academic or other purposes

Which of the above categories do the LMAs you describe in this questionnaire fall into?

**A - Q.5** When were these LMAs defined?

### **General questions about the definitions of LMAs**

**A - Q.6** What main information sources were used in defining the boundaries of the LMAs (e.g. commuting or migration flows, or local knowledge)?

**A - Q.7** Did the definition process consider the whole country, or only certain parts?

**A - Q.8** Was there a single unified definition process, or were regions analysed separately (e.g. so that the boundaries of the regions could not be crossed by the LMAs)?

**A - Q.9** Did the definition process end by allocating all the areas considered in LMAs, or could some areas be left unallocated (e.g. remote or island areas)?

**A - Q.10** Please describe in detail the method you used to define the LMAs currently operational in your country, in a way that could allow it to be used to define LMAs in other countries.

*Most probably the description of your methodology will be a longer text, to be attached separately to this questionnaire. We would prefer a text in English, but if it is only available in your national language, this is fine as well. We will translate it. If the text comes from a different institution, please organise the contact to that organisation, but send us the result of your research from your RESCO desk.*

**The following questions summarise issues which the description should cover; please provide summary answers here:**

**A - Q.11** What are the basic building block areas?

**A - Q.12** Is there a criterion to decide which areas to group together (e.g. a minimum rate of commuting)?

**A - Q.13** Is there a restriction on grouping non-contiguous zones?

**A - Q.14** Are multiple “step” links allowed (e.g. will chains of cities, with strong commuting between each pair of cities, be grouped into a single LMA, or will the chain be broken)?

**A - Q.15** Are all the initially defined groupings accepted as LMAs, or do they have to meet some other criteria (e.g. a minimum level of self-containment or population, or maximum size of area)?

**A - Q.16** Is there a consideration of possible transnational LMAs?

- A - Q.17** Please send us a detailed description of the boundaries of your LMAs, either through geo-reference layers or with the aid of a list of the administrative units contained in each LMA. Preferably use the LAU list that you have sent to Eurostat for the reference year concerned. A map of the LMAs in your country would also be very much appreciated.

[Please indicate the name of the file attached and the reference year]

- A - Q.18** Please send us a file with the most recent area and population in each LMA.

[Please indicate the name of the file attached and the reference year]

### **Section C – Data for the delineation of LMAs**

**Please, note that with only a couple of exceptions, all EU-27 Member States collect data that may enable the delineation of LMAs at LAU2 or LAU 1 levels. This data mainly results from Population Censuses, but could be also retrieved from specific surveys, register based surveys, or other themes generally NSOs produce data for (e.g. Employment Statistics, Local or Regional Statistics, Territorial data, Labour migrations, etc.).**

- C - Q.19** Is a travel-to-work dataset available?

Please note that we specifically refer to data on commuting flows between geographical units in your country - number of persons that commute between each pair of units considered as origin and destinations of the flow. This information can be visualised as a matrix where all the units appear as rows and columns so that each cell  $T_{ij}$  depicts the number of persons commuting from unit  $i$  to unit  $j$ .

- C - Q.20** If such data on commuting is not available, please indicate it and state whether they will be available in the future (for example after the Census of Population 2011), and give details on this.

- C - Q.21** If this type of data does not exist for your whole country, please indicate for which part of the country do they exist and proceed to the following questions.

- C - Q.22** With regards to **the most recent data currently available:**

C - Which is the **lowest territorial level** for which commuting data are available:

Q.22.

a

LAU 2

LAU 1

Other levels [Please, indicate nomenclature and number of units that constitute this level]

C - Please indicate the reference **year**:

Q.22.

b

C -

Q.22. Please, indicate the **source** (Census of Population, other, etc.)

c

C -

Q.22. **Organisation** in charge of producing the data:

d

C - Are data publicly **available**? If so indicate how these could be accessed:

Q.22.

e

C -

Q.22. When will an **update** of these data be available?

f

*Indicate year, source and periodicity of the data (bear in mind that we always refer to data on commuting flows between all geographical units of a certain level in your country). Feel free to add comments.*

C - Please specify any relevant **particularities** of the data

Q.22.

g

Do the data provide **disaggregated** information by gender / section of activity / other?

Do the data include information about **cross border incoming and outgoing commuting**? If so, which is the detail? Will it be possible to know the destination and origin of cross border commuters by geographical unit (LAU2, LAU1, other) of origin and destination? Please add your comments on this issue.

Is/Are any specific sub-group(s) of workers (such as self-employed) excluded? If this is the case, please provide details on the percentage they mean of total employment.

Are the data based on direct information from all the population or extrapolated from samples? Please give details of the sampling strategy if this was the case.

Are data aggregated for more than one geographical unit due to confidentiality concerns caused by small population? If this is the case, how is this process guided?

Other comments (e.g. are the unemployed considered in any way?)

C - Please indicate for which of the **previous years** is it possible to obtain similar information and  
Q.22. feel free to add your comments on this issue.  
h

C - With regards to the 2011 round of Population census and referring to your country, will it be  
Q.23 possible to construct a matrix of commuting flows between geographical units of LAU 2 level?  
Which is the expected date of availability of such data?

C - The study of which this questionnaire is part is researching commuting data for all EU Member  
Q.24 States: can you facilitate our access to the most recent data on commuting flows between  
small areas in your country (ie. number of persons that commute between each pair of units  
considered as origin and destinations in a matrix)? If you are not the right contact point, please  
provide an internet link or email address where the data should be sought from.

#### **Section D - Sources of information -possible characterisation of LMAs**

D - What socio-economic variables do you specifically collect for your LMAs?, where LMAs are sets  
Q.25 of other basic geographical units (i.e. LAU 2). Which are the main indicators available for such  
units?

D - How often are these variables / indicators updated?  
Q.26

D - If you have published the definition and / or the variables / indicators for LMAs on the Internet,  
Q.27 please send us the web link.



## **Section B - NSOs' positions towards an LMA geography and / or their expectations from this**

**B -** Do you find LMAs useful? Which purposes should they be fitted to?

**Q.28**

**B -** Which were the reasons for the creation of LMAs?

**Q.29**

**B -** Which are your main concerns regarding such an exercise (i.e. defining a new statistical classification)?

**Q.30**

**B -** Please share your opinion / comments about the statistical characteristics that should be met by LMAs if an EU-wide grid of comparable areas would be proposed (e.g. How a standard objective in terms of area / population, commuting self-containment or other variables should be established?, etc.)

**Q.31**