Creating Small Areas for Ireland – Problems and Opportunities

A Stewart Fotheringham[1]
Martin Charlton[1]
Mary O'Brien[1]
David Martin[2]

[1]National Centre for Geocomputation, National University of Ireland, Maynooth, County Kildare, Ireland
[2]Department of Geography, University of Southampton, Southampton, Hampshire SO10 1BJ

Corresponding author: A Stewart Fotheringham (stewart.fotheringham@nuim.ie)

Biography

A. Stewart Fotheringham is a Science Foundation Ireland Research Professor and Director of the newly established National Centre for Geocomputation at the National University of Ireland, Maynooth.

Tel: +353 708 6455
Fax: +353 708 6456
Web: ncg@nuim.ie
Email: stewart.fotheringham@nuim.ie

Introduction

Within the United Kingdom the opportunity was grasped before the 2001 Census of Population to create a set of output areas using consistent and automated definitions for which data would be released. These output areas have not only allowed data to be released at a more detailed spatial resolution, but they also provide the basis for more consistent reporting of data over time. The development of output areas has been described by Martin et al (1999, 2001) and incorporates earlier ideas on the creation of zoning systems by automatic means (Openshaw 1996).

The opportunity has now arisen to undertake a similar exercise in the Republic of Ireland. The exercise is being undertaken as part of the development of an Irish Spatial Data Infrastructure (ISDI). The Republic of Ireland currently has several sets of reporting units for which census data are available. These are often spatially incompatible, and in many cases, boundaries are not coterminous which creates problems in comparing data from different agencies.

Irish Spatial Units

The major administrative unit in the Republic of Ireland is the County (29) with the major urban areas having County Borough status (5). These units form a complete spatial partition of Ireland. Urban Districts and Rural Districts (212) are the next level down in the hierarchy, with Electoral Districts (3440) currently the smallest unit for which data area readily available. Within urban areas, census data can sometimes be made available at a lower level, the Enumeration Area, subject to confidentiality tests, by the Central Statistical Office (CSO) although release at this level may require a special tabulation. Within rural areas, there is another set of spatial units, Townlands, of which there are about 54000 in the Republic. Townland boundaries are not sometimes but not always coterminous with ED boundaries.

EDs themselves are large and display wide internal social heterogeneity – the average population is 1144 with a range from 55 to 24400. Maynooth, the location of the National University of Ireland, Maynooth, is amongst the larger EDs with a population of 11000.

The existence of a range of incompatible areal units is one of the reasons why, under the auspices of ISDI, a project has been started to consider the creation of a set of atomic small areas for the Republic. The idea is that these may then be combined into larger areas as necessary, but for which there will be outputs from the Census and for which there will also be boundaries in digital form. This will at least bring Ireland into line with what has been available in other countries for several decades.

Research Issues

The need for a set of atomic units raises a host of questions. How many are required? How shall they be created? What data do we have available to assist with their creation? What algorithms shall we use? What rules will we need to create a consistent definition? How will we handle exceptions and unusual cases? Into which spatial units should they nest? To examine these questions, a pilot project is being constructed in two adjacent EDs, Maynooth and Leixlip, to determine the viability of an automatic means of creating meaningful small areas from existing data sets. The familiarity of the majority of the researchers on this project with the two EDs will assist in ground truthing the results of the determination of meaningful small areas.

Some initial attempts have been made to answer these questions. The CSO have stated that they are happy to release data for atomic units of around 50 households, a vast improvement on what is currently available. If each small area is to contain an average of 50 households then around 25000 areas will be required – an ED such as Maynooth will require around 64 areas. However, there are around 170 EDs with fewer than 50 households, so the 'small area' will equate to the ED boundary in these cases. It has also been established that small areas should nest into EDs.

The question of what set of rules should be used to determine the creation of small areas is the subject of a pilot project undertaken by the authors and this paper will report on some of our experiments in this area. Within the UK it has been observed that ED boundaries have in the past often followed street centrelines. This implies that the units of social interaction in a settlement have been used as boundaries rather than as unifying features and means of correcting this in the pilot project will be described. Watercourses and railways are examples of natural boundaries to social interaction, so the question of their inclusion becomes important. Ordnance Survey Ireland, one of the partners in the project, is providing suitable data for the project, 1:1000 vector data for urban areas and 1:2500 vector data for rural areas, and the feature coding on these data makes extraction of the centrelines and boundaries possible.

There are no postcodes in Ireland (apart from 25 large areas around Dublin) but the Irish Post Office, An Post, release a regularly updated data file, GeoDirectory, which relates x and y coordinates to individual postal addresses. Each address is given a 1m spatial reference, together with a quality indicator. The data include both domestic and commercial addresses. The project will use GeoDirectory to geocode each property so that they can be allocated to centroids of newly formed output areas on the basis of proximity. To help road centrepoints be unifying rather than boundary features in the creation of these small areas, we will use road distances rather than straight line distances to measure proximity.

It will also be desirable to have areas which are socially homogenous and how one measures social homogeneity is perhaps the major research topic of this paper. We are extremely

fortunate that for this pilot project, the CSO have agreed to release the individual household census records. Consequently, we have much more accurate means of measuring the social homogeneity of the small areas that are created than say was possibly in England and Wales and Northern Ireland were output areas were built as amalgamations of unit postcodes rather than individual properties. We still have to decide on the appropriate measure of social homogeneity but we have asked the CSO to release the following data to us which we can use to measure social homogeneity:

1. One-person household (binary)
2. Socio-economic group of head of household (7 categories)
3. Flat or Apartment household (binary)
4. Year house built (8 periods)
5. Owner-occupied or not (binary)
6. Number of rooms (8 categories)
7. Number of people = 17 in household
8. Number of cars per household (0,1,2, 3+)

The division of item 8 by item 7 gives us a crude measure of cars per person for each household – a local measure of income for small areas. Thus we will have seven indicators for each household that we can use to create small areas that are internally homogeneous. One result from the project will be to investigate which of these variables is most useful so that if a nation-wide application of the small areas algorithm takes place, it can be undertaken using this single piece of information.

Another major research topic will be to define an appropriate statistical measure of homogeneity. Ideally, one wants to maximise internal homogeneity and maximise external heterogeneity. There are various measures that could be used for this and an important outcome from this project will be to recommend one particular measure.

Finally, a suitable algorithm for the automatic determination of small areas will need to be developed. It is likely that Openshaw's AZP (1996) algorithm with some modifications will provide a basis for the procedure although it might be that a location-allocation formulation is required. One useful property of the restriction that the new small areas must nest into EDs is that each of the 3421 EDs can be tackled separately, so that the computational overhead is reduced.

Consequently, this paper will report on various tests in the creation of small areas from individual units. These involve: the selection of a suitable algortithm; the selection of appropriate rules; the integration of various data bases; the selection of an appropriate variable by which to measure social homogeneity; and the selection of an appropriate statistical measure of homogeneity. Ultimately, it is hoped that the knowledge gained in these areas can then be applied to the other EDs in the Republic.

In addition, this project represents a notable collaboration between two government agencies and the academic sector with the joint ambition of creating a set of highly localised, small areas from which to report future socio-economic data. Such intensive reporting of socio-economic data will not only be of tremendous importance for national and local government decision-making but it will also be of immense value to the private sector.