The role of national address database in adding value to Irish statistics

Corcoran, Dermot

Statistician, Census Geography Section, Central Statistics Office,

Swords Business Campus, Balheary Road, Swords, Co. Dublin, Ireland

E-mail: dermot.corcoran@cso.ie

ABSTRACT

The GeoDirectory is the name for the national address database of Ireland and is the result of a joint effort between the Irish postal service (An Post) and Ordnance Survey Ireland (OSi) to link all residential and commercial addresses to a geospatial location.

This database has already improved the operations and outputs of the Irish statistical office through its role in underpinning the 2011 census, and in helping to create a new level of geography. However the recent announcement of a new postcode for Ireland, which is unique at address point level and will be generated from the GeoDirectory, means that the opportunity exists to add new value to Irish statistical outputs.

1. INTRODUCTION

At the 2013 European Forum for GeoStatistics conference, the call was made for the greater integration of NMCA (National Mapping and Cadastral Agency) and NSI (National Statistical Institute) data to serve public authority information.

This paper looks at the partnership between the Irish NSI (Central Statistics Office or CSO) and NMCA (OSi) in the context of the Irish address database, called the GeoDirectory[1]. This cooperation is shown in three case studies, all of which have directly affected the Irish statistical system.

2. SMALL AREA BOUNDARIES

In 2005, the Small Area Working Group, established as part of the Government's Irish Spatial Data Infrastructure project, discussed the benefits of creating a new level of geography (called the Small Area) for the Republic of Ireland, the operational advantages of using Small Areas and how to introduce the Small Area into Irish administrative records.

[1] The history and composition of the GeoDirectory has been discussed at the 2011 EFGS conference. For further information see http://www.efgs.info/workshops/efgs-2011-lisboa-portugal/efgs-2011-conference-in-lisboa/dermot_corcoran_paper_submitted_to_the_european_form_of_geostatistics.pdf

As a key member of this group the CSO committed itself to the collection and dissemination of data at Small Area level for the census in 2011, provided definite Small Area boundaries were created for the whole State. Arising from the group's recommendations, the National Centre for Geocomputation, at the request of OSi, devised an algorithm to produce a series of Small Area boundaries for the State.

It was recognised that the GeoDirectory was considered essential for this algorithm. The number of residential address points within the GeoDirectory was chosen as the basis for setting the minimum and maximum threshold values of dwellings within each Small Area. Polygons were then generated around the coordinates of residential buildings from the GeoDirectory using road centreline segments as the boundary outline. The digital boundaries of legacy geography (EDs, Townlands) and natural divisions (rivers, canals and railway lines) were also taken into consideration during the creation process.

Staff from both CSO and OSi liaised strongly on reviewing the output produced by the algorithm, which involved each Small Area polygon being checked in order to ensure that they were practical and navigable. Other issues such as the Small Area codes and ownership rights were agreed by the two agencies prior to their release in 2009. By using the GeoDirectory as the source for their creation, around 92% of the 18,488 Small Areas have a range of between 75 to 150 dwellings.

So far the CSO have been the biggest user of the boundaries, contributing towards the preparation and dissemination of the 2011 census as well as providing the sample blocks for the mandatory social surveys operated by the office.

3. THE USE OF THE GEODIRECTORY IN CENSUS 2011

The benefits of the GeoDirectory for the Irish census were apparent from the outset. Initially its use for the censuses of 2002 and 2006 was restricted to helping to identify the possible location of newly constructed residential buildings in order to plan enumerator workloads. The reason for this restriction was that the database was struggling to keep abreast with the surge in housing construction that occurred in the State between 2002 and 2007. For the census in 2011, the GeoDirectory was deemed to have the necessary quality to be used as the lynchpin of the census operations.

The primary intention of its use in 2011 was to link census data to the addresses and coordinates contained within the GeoDirectory. The methodology involved assigning a unique census ID (a nine digit code representing the county, EA and the dwelling number) to each GeoDirectory address selected for enumeration. The dwelling numbers were labelled on the enumerator maps, with the GeoDirectory address along with the dwelling number being pre-printed on the address lists supplied to each enumerator. When the enumerator visited the dwelling they then used their map to locate the relevant dwelling number on their address list,

and then transferred the contents (address, census ID) onto the census form before handing it to the householder.

In addition to its use in the fieldwork operations, the GeoDirectory was used to geocode addresses relating to place of usual residence and place of work/school/college during the census processing stage. This work facilitated the release of a Research Microdata File called POWSCAR (Place of Work, School or College Census of Anonymised Records), which is made available to bona fide researchers for the purposes of analysis on commuting patterns.

As part of the quality assurance process for future censuses, the CSO supplied feedback to the GeoDirectory staff based on perceived shortcomings observed by the enumerators. This included the estimated location of approximately 40,000 dwellings found in the field that were not originally on the address database, residential address points that were recorded by the enumerators as commercial only, not existing or derelict, as well as identifying buildings where the coordinates were in the wrong location. The GeoDirectory team then consulted with the postal staff to confirm or disprove the findings of the enumerators.

Overall the linkage between the GeoDirectory and the census has been mutually beneficial to CSO, OSi and An Post. It meant that the 2011 Irish census file was 100% georeferenced, enabling the CSO to have complete flexibility in their geographical outputs. Significant savings also occurred in relation to certain post-census work activities such as the examination of extensions to the suburbs of Irish towns and cities. From a GeoDirectory perspective the census was the first time that the database had undergone a third party review of its quality and a number of changes were made resulting from the enumerator's observations.

4. THE NEW IRISH POSTCODE

Ireland is the last country in the EU or OECD without a national postcode system. In 2009 the Government stated its intention to introduce postcodes throughout the State, and following a tendering process it was announced in 2013 and the postcode design will consist of a unique identifier for every residential and commercial address. From a situation where around 35% of addresses are non-unique, causing confusion with regards to the delivery of post and services, Ireland will soon have one of the most specific and detailed postcode system in the world.

The new postcode will be known as Eircode, and the format will be similar to A65 R2AF. The first part of the postcode will represent the post towns, of which there are 139 throughout the State, while the second part will be assigned randomly from a carefully selected set of letters and numbers[2]. By mid-2015, all 2.2 million addresses in the country will have been

[2] A 25 character set will be used: 10 numbers and 15 letters. B,G,I,J,L,M,O,Q,U,S and Z will not be used.

notified of their Eircode. At that time, it is intended that the major state bodies will be using the postcodes in correspondence and in their dealings with the public.

In order to implement such a system, the GeoDirectory will be used as the seed for the creation of the Eircode Address Database. Each address point on the GeoDirectory will be allocated a postcode. The linkage between Eircodes and the GeoDirectory means that the Irish postcode will have a geospatial component. In addition, to prepare for the launch of the Eircodes, a number of large databases held by the main Government departments are being postcoded in advance by geocoding the addresses against the GeoDirectory. However only unique addresses will be geocoded, and for non-unique addresses the onus will be on individual departments to obtain the postcode through contact with the public.

The successful adoption of the postcode will be contingent on a number of issues, the key one being whether the Eircode will take hold in the public psyche. In order for this to happen it is vital that the positive aspects of the postcode are emphasised during the publicity campaign (easier for ambulances to locate rural households, swifter delivery of postal and courier services etc.). Furthermore, there has to be a strong desire from Government bodies, utility companies and commercial businesses to compel households to supply their Eircode if they want to use their services. Another issue deals with addressing data protection concerns, and the Irish Data Protection Commissioner has called for the privacy impacts of the new postcode to be assessed prior to its launch.

The type of postcode that has been chosen by Government will lead to significant statistical and, by extension, policy benefits across the public sector in Ireland. The principal improvement will ensue from having Irish administrative data linked to a spatial location once postcodes are included in the address fields. Other benefits will cover the whole range of statistical tasks such as location finding for interviewers, register management, administrative data files matching, the presentation of survey results, and the smooth aggregation of data to defined geographical areas. The postcode also offers the chance to introduce GIS into the CSO's statistical production process, and make it an important software tool in several sections of the office.

CONCLUSION

This paper has attempted to demonstrate how geospatial information has brought added value to the Irish statistical system through three major developments; the creation of a new geographical area, the georeferencing of the Irish census and the impending introduction of a new geospatial postcode. In all of these examples the GeoDirectory has been the common denominator to their implementation, and demonstrates how important the database has been in benefitting Irish statistics. As a result the collaboration between the CSO, OSi and An Post using the GeoDirectory can be seen as an example for showcasing the potential of integrating geography with statistics.