A population grid for the Republic of Ireland:

Making use of national databases and local geography

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

This paper outlines the process behind producing a population grid map for the Republic of Ireland using Census 2006 data. The grid map produced is a combination of aggregation and disaggregation techniques. The aggregation approach is applied predominately in urban areas, and is the result of a project which attempted to geocode the address on the Census 2006 Forms against the Irish address database (called the GeoDirectory). The disaggregation approach is applied in rural and peri-urban areas using the levels of geography available on the GeoDirectory and the Census 2006 file.

The methodology and results of this exercise are presented in this paper. The plans for producing a population grid dataset from the 2011 Census are also discussed, as are the opportunities and challenges from an Irish perspective in relation to dissemination of grid data and contributing to the GEOSTAT initiative.

1. INTRODUCTION

The Central Statistics Office (CSO) was established in 1949 as an independent office with the remit of producing detailed official economic and demographic statistics for the Republic of Ireland. Since its establishment, the CSO has been given the responsibility for conducting the Irish population and housing Census which, from 1951 onwards, has normally occurred every five years[1].

Similar to other National Statistical Institutes, the CSO has needed to respond to the developments in the area of geographical and spatial information systems. The concept of owning and maintaining spatially referenced statistical data is relatively new to the CSO, and

[1] There are two exceptions. The 1976 Census was cancelled for economic reasons and the 2001 Census was postponed to 2002 because of the foot and mouth outbreak.

developments in this area have been driven mainly by the availability of a national address database (called the GeoDirectory), and the arrival of a new level of geography (called the Small Area).

2. THE GEODIRECTORY AND SMALL AREA BOUNDARIES

The GeoDirectory is a database created in 1998 following a collaborative approach between the Irish postal service (An Post) and Ordnance Survey Ireland (OSi) [2] as a means to cater for:

- The automation of mail processing in An Post and the requirement for a database of addresses,
- The desire by OSi to develop and grow the GIS market by producing a geocoded buildings database,
- The market opportunity to develop a new revenue stream (Keogh 2011).

The GeoDirectory staff release an updated version of the database every three months and the most recent version contains the address points and spatial coordinates of approximately 2 million buildings throughout the State. Attributes associated with each building are also provided, such as its date of creation on the GeoDirectory, the geographical location of each building and whether it is residential, commercial or mixed residential/commercial, vacant, under construction or derelict. Approximately 40% of addresses on the GeoDirectory are non-unique and these generally relate to buildings in rural areas, small towns and villages. Since the State does not have postcodes, this means that the delivery of post in these areas is heavily dependent on the postman knowing who lives where.

The Small Area boundaries were created in order to meet user demands for geographical areas which are more homogenous in size and composition than the lowest available administrative unit available for the dissemination of Census data (the Electoral Division) [3]. In 2007, the National Institute for Regional and Spatial Analysis (NIRSA), at the request of OSi, developed an algorithm to produce a series of Small Area boundaries for the State. The algorithm took into account the location of residential address points on the GeoDirectory, road centrelines, natural boundaries (such as rivers, railway lines etc), as well as ensuring that all Small Areas nest within the boundaries of the Electoral Divisions.

Approximately 19,000 Small Areas have been generated across the State, and most of these contain, on average, 75-150 residential address points. Following their release in 2009, the Small Areas have already provided a number of benefits to the CSO:

[2] Ordnance Survey Ireland is the national mapping agency for the Republic of Ireland.
[3] There are 3,441 Electoral Divisions varying in size from 4 to 11,558 households according to the Census 2011 preliminary results.

- The Enumeration Areas for Census 2011 were created by combining 3 or 4 Small Areas together.
- The Small Area is facilitating the timely publication of Census 2011 results, as well as providing detailed Census data at neighbourhood level.
- The Small Area is currently being used by the CSO for the selection of sample areas for the Quarterly National Household Survey (QNHS), the Household Budget Survey (HBS), the Programme for the International Assessment of Adult Competencies (PIAAC), and the EU-SILC survey.

3. GEOCODING OF CENSUS 2006 AGAINST THE GEODIRECTORY

The CSO has been keen to exploit the obvious statistical benefits to the office and the Irish Census offered by the development of a spatially referenced Irish address database. The first significant trial of the GeoDirectory occurred during the 2004 Census Pilot, where pre-assigned dwelling numbers, which were linked to each GeoDirectory address point selected for enumeration, were labelled on the enumerator maps. Each enumerator was supplied with an address list which contained the pre-assigned dwelling numbers and addresses of the GeoDirectory points displayed on the map. The intention behind this exercise was to link Census data to the address points and coordinates on the GeoDirectory.

However, the overall conclusion following the pilot was that "the GeoDirectory contained flaws, which at the present time prevent it from being of practical use in a full Census" (CSO 2005 p.7). It was felt within the CSO that the gaps in coverage on the GeoDirectory which were identified during the pilot would pose a real threat to the enumeration process, and the extent of such gaps were unknown in light of significant housing development that was occurring in Ireland at that time [4]. The results of the pilot meant that for the 2006 Census, it was deemed necessary to revert to the traditional methods of enumeration that had served the CSO well in previous Censuses where enumerators carried out a full visual enumeration and coded the location of the dwelling down to the level of street and Townland.

The developments in the Small Areas project prompted the CSO to take the decision in 2007 to retrospectively geocode the address of each dwelling enumerated in the 2006 Census to its corresponding address on the GeoDirectory, with a view to publishing Census 2006 data at Small Area level once the boundaries were finalised. In order to geocode the addresses of dwellings on the Census 2006 file, an application was designed by the Census IT department. This application allowed the user to examine the scanned image of the front panel of the Census 2006 Form for a particular dwelling and to search the address as written on the

[4] The Department of Department of the Environment, Community and Local Government has estimated that the housing stock in the State increased from 1.5M dwellings in 2002 to over 1.9M dwellings in 2007, an average increase of 80,000 dwellings per year. For more information see http://www.environ.ie/en/Publications/StatisticsandRegularPublications/HousingStatistics/

scanned image against the GeoDirectory. Where the address on the Census Form could be exactly matched to an identical address on the GeoDirectory the user selected the 'Exact Match' option, and in cases where an exact match was not possible, the 'Nearest Match' option was selected.

A total of 1.485 million occupied dwellings on the Census 2006 file were geocoded against the GeoDirectory. However in areas where the addresses on the GeoDirectory are non-unique, the geocoding exercise had limited success, and approximately half of all the geocoded dwellings were recorded as 'Nearest match'. Different approaches were then applied to review and correct these records:

- Dwellings coded as 'Nearest matches' in the Dublin NUTS 3 region or the administrative cities of Cork, Limerick, Galway and Waterford were digitally 'pulled' to their correct position in a GIS according to the location of the dwellings on the enumerator map.
- Dwellings coded as 'Nearest matches' in peripheral urban areas, towns and villages were geocoded to the nearest GeoDirectory building in the same locality.
- Dwellings coded as 'Nearest matches' in rural areas required an alternative approach. The enumeration of Irish Censuses in rural areas has traditionally been carried out using the lowest geographical unit, called the Townland. There are approximately 51,000 Townlands throughout the State, and a digitised set of Townland boundaries are available from OSi. During the consultation phase concerning the creation of Small Area boundaries, it was agreed that in rural areas, each Small Area was to consist of an aggregation of complete Townlands. The decision was taken therefore to allocate each rural dwelling on the Census 2006 file the centroid coordinates of the Townland in which the dwelling was located.

The work on the coding of rural dwellings ensured that every dwelling on the Census 2006 file had spatial coordinates. Upon the receipt of the final version of the Small Area boundaries in 2009, the spatially referenced geocoded Census 2006 file was imported into GIS along with the boundaries, and each dwelling was allocated a Small Area code.

4. PRODUCING THE GRID MAP

The final results of the geocoding project revealed that a total of 57% of the 1.485 million occupied dwellings on the Census 2006 file had been coded as an 'Exact match' in the geocoding application or had been pulled to their correct location in a GIS according to the enumerator maps. An aggregated approach can be applied to these dwellings. The remaining 43% of dwellings have either the centroid coordinates of a Townland or the coordinates of a GeoDirectory building in the same locality, and Census 2006 data for these dwellings need to

be disaggregated into grid cells. Therefore mapping the population of Ireland on 1 Km² grids using Census 2006 data involves a combination of aggregation and disaggregated methods.

Before these methods can be applied a 1 Km² vector grid was created using a coded script on the ESRI website called 'Fishnet' (ESRI 2003). This application enables the user to set the coordinate reference system in which the grid is to be created, and also provides the option to establish the size of each grid cell (in metres) as well as the total number of grids to be created. The Fishnet application was imported into ArcGIS, and the Irish Grid coordinate reference system was chosen for the generation of the 1 Km² grid cells. Finally each grid generated by Fishnet was allocated a unique identifier code (called Grid_ID), which is in a ten digit character format (e.g. IR_GR00001).

Aggregated grid data

Two techniques were applied to produce aggregated grid data from the geocoded 2006 Census file. As a result of the geocoding exercise every dwelling on the 2006 Census file had a Small Area code. An analysis of the Small Area boundaries using GIS showed that 62% of Small Areas are less than 1 Km² in size. Therefore, the first aggregation technique involved identifying Small Areas which are not split across grid cells, and automatically allocating a Grid_ID code to dwellings on the Census 2006 file located in these Small Areas using SAS.

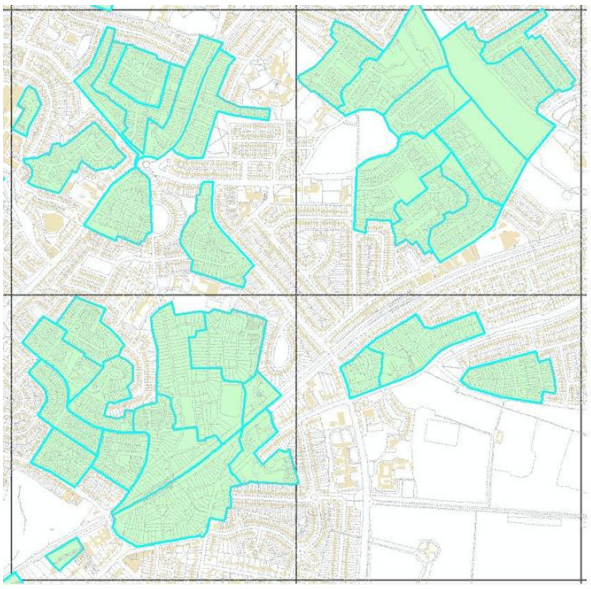

Figure 1: Small Areas not split across 1 Km² grids.

The second aggregation technique involved identifying the remaining dwellings which were coded as an 'Exact match' in the geocoding exercise or had been pulled to their correct location in a GIS according to the enumerator maps. These dwellings, which were not part of the first aggregation technique, were imported into GIS and allocated the Grid_ID of the grid cell in which the dwellings were located.

Disaggregated approach

The proposed methodology to disaggregate Census 2006 data involves the use of the July 2006 version of the GeoDirectory. The reason for the choice of the July 2006 GeoDirectory is that it is the first version of the database following Census night[5] and would therefore reflect the location and coverage of residential address points that would have been enumerated during the 2006 Census.

In rural areas the intention is to disaggregate the population related to the Townland on the Census 2006 file into grids according to the number of GeoDirectory residential address points located within that Townland as of July 2006. Any GeoDirectory residential address points which were geocoded exactly to an occupied dwelling on the Census 2006 file were excluded from the disaggregation techniques.

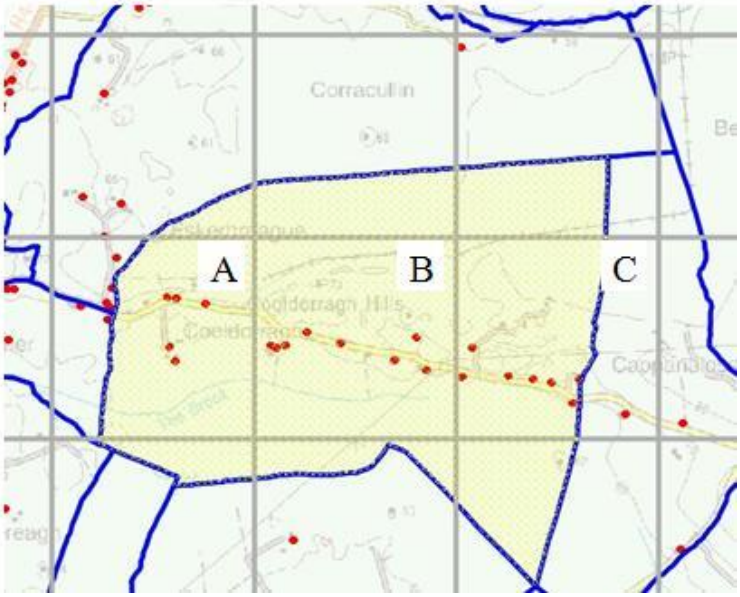

Figure 2: Example of using the disaggregation approach in rural areas.

Figure 2 provides an example of a Townland (with boundary shown in blue) in a rural area, along with the boundary of the 1 Km² grids (shown in grey). The Census 2006 file shows that 52 persons were enumerated in this Townland in 2006. According to the GeoDirectory from July 2006, there are 20 residential address points (shown as red points in Figure 2) in this particular Townland, with five points located in grid A, eight points in grid B and seven points in grid C.

The methodology for calculating the population of these grids is to distribute the population of this Townland depending on the percentage breakdown of residential GeoDirectory points within each Townland that reside in a particular grid. In the example above, the calculation of

[5] The 2006 Census was taken on the night of 23rd April 2006

the population for grid B involves dividing the total GeoDirectory residential points of the Townland that reside in grid B over the total GeoDirectory residential points for the entire Townland (8/20). This calculation produces the result that 40% of the Townland's residential address points fall within grid B and therefore the population of this grid is 40% of the Townland's population (52 x 0.4 =20.8 persons).

In the case of the other grids 25% (5/20) of the Townland's population will form part of the population total for grid A, and 35% (7/20) of the Townland's population will contribute to the population total of grid C. The calculation of the population for a particular grid cell (G_j) in rural areas can be represented by the following equation:

$$G_j = \sum_{i=1}^{N} T_{ij}$$

Where:

$$T_{ij} = \left\{ \frac{Number\ \ of\ GeoDirectory\ \ points\ \ of\ Townland\ \ i\ in\ grid\ \ j}{Number\ \ of\ GeoDirectory\ \ points\ \ in\ Townland\ \ i} \text{ X } population\ of\ Townland\ i \right\}$$

N= Number of Townlands in grid j

A similar methodology is applied to Census 2006 dwellings in peripheral urban areas, towns and villages. Dwellings in these areas do not have a Townland code on the Census 2006 file (instead they have a street code), and therefore the lowest geographical unit available is the Small Area. The methodology involves the disaggregation of the population of the Small Area into grids, relative to the location the GeoDirectory buildings located within the Small Area as of July 2006. Again any exact matches between the Census 2006 file and the GeoDirectory within these Small Areas were excluded.

The calculation of the population for a particular grid cell (G_j) in peripheral urban areas, towns and villages can be represented by the following equation:

$$G_j = \sum_{i=1}^{N} S_{ij}$$

Where:

$$S_{ij} = \left\{ \frac{Number\ \ of\ GeoDirectory\ \ points\ \ of\ Small\ \ Area\ \ i\ in\ grid\ \ j}{Number\ \ of\ GeoDirectory\ \ points\ \ in\ Small\ \ Area\ \ i} \text{ X } population\ of\ Small\ Area\ i \right\}$$

N= Number of Small Areas in grid j

The combined aggregation and disaggregation techniques produced the population grid map as shown in Figure 3.

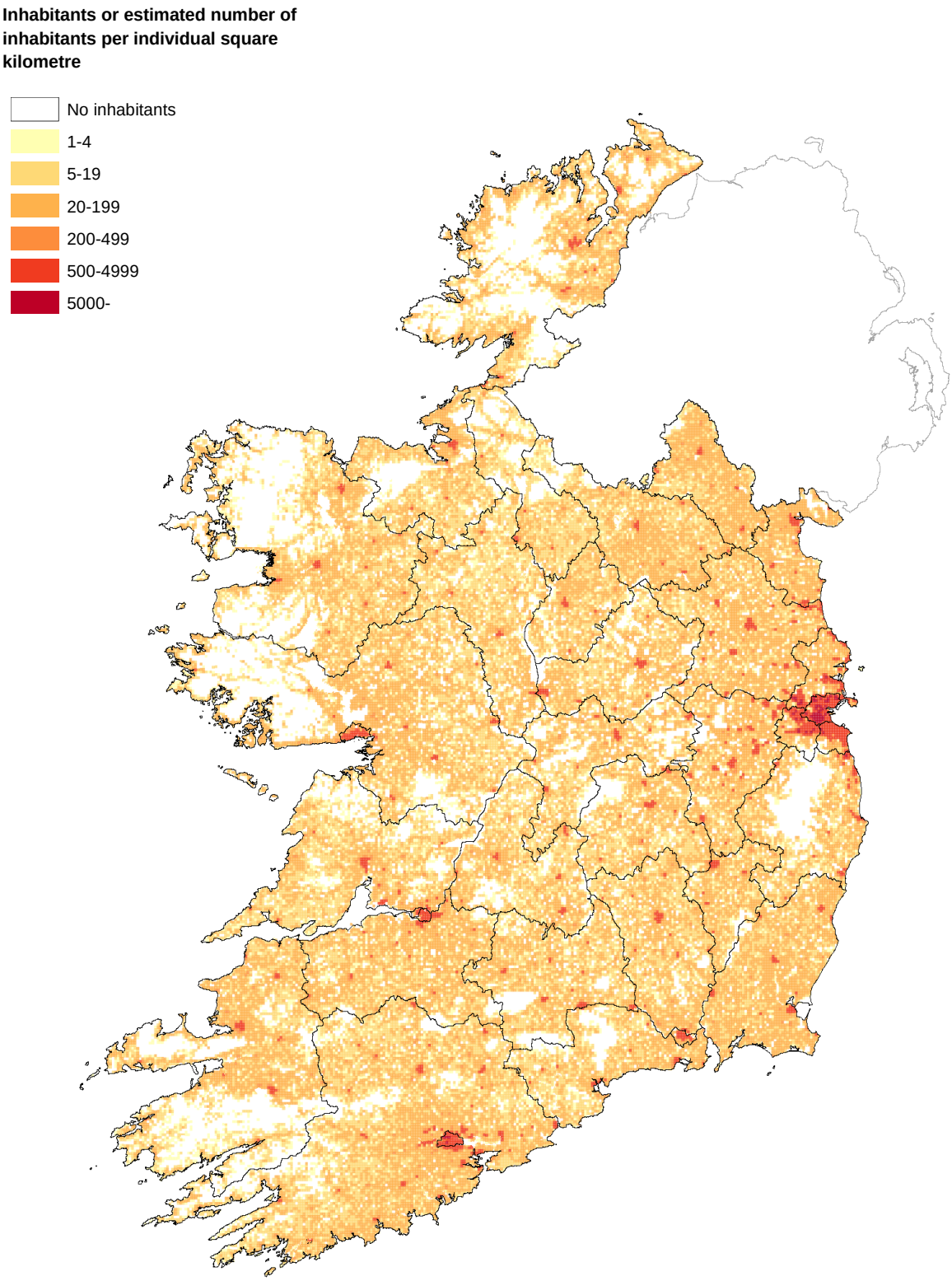

Figure 3: Population of the Republic of Ireland on 1 Km² grid cells using Census 2006 data.

5. RESULTS

Table 1 shows the household and population breakdown according to the type of aggregation approach used to produce the grid map. Close to 60% of the population of the Irish Republic in 2006 were allocated a Grid_ID code using one of the aggregation methods.

	Households	% Households	Population	% Population
Aggregated method	887231	59.74%	2471534	58.29%
Disaggregated method	597837	40.26%	1768314	41.71%
	1485068	100.00%	4239848	100.00%

Table 1: Breakdown of Census 2006 Household and population according to the method used to produce grid data.

A total of 83,121 grids cover the surface area of the Republic of Ireland occupying a land area of 68,977 Km². Table 2 shows that 66% of the population reside in 1 Km² grid cells occupying 3% of the total land area in the State, while the remaining 33% are in 1 Km² grid cells covering 77% of the land area[6].

Population per grid	Number of grids	% Number of Grids	Total Land Area in the grid cells- Km²	% Land Area	Grid population	% Grid population
No inhabitants	25843	31.09%	13722	19.89%	0	0.00%
1-4	6940	8.35%	6794	9.85%	20383	0.48%
5-19	21559	25.94%	20973	30.41%	254005	5.99%
20-199	26249	31.58%	25242	36.59%	1169243	27.58%
200-499	1096	1.32%	981	1.42%	305797	7.21%
500-4999	1345	1.62%	1194	1.73%	2043673	48.20%
5000+	89	0.11%	71	0.10%	446816	10.54%
Total	83121	100.00%	68977	100.00%	4239917	100.00%

Table 2: Breakdown of the population per grid.

Almost 20% of the total land area comprises of 1 Km² grid cells which are uninhabited, and Table 3 provides a comparison of the inhabited land area against other Countries which have produced population using 1 Km² grids. Please note that the figures in Table 3 were taken from the 'GEOSTAT Population Map 2010', which is available on the EFGS website.

[6] The land area was calculated by importing the shapefiles of all the rivers and lakes in the State into GIS and subtracting the area of these waterbodies from the total area covered by the grid cells.

Country	Inhabited 1 Km² Grid cells as % of total land area
Republic of Ireland	80
Austria	52
Denmark	91
Estonia	51
Finland	34
Kosovo	51
The Netherlands	89
Northern Ireland	77
Norway	18
Slovenia	65
Sweden	28
Switzerland	48

Table 3: Comparison of inhabited grid cells as % of total land area.

6. CENSUS 2011 AND FUTURE PLANS

The 2009 Census Pilot once again tested the concept of using the GeoDirectory to underpin Census fieldwork operations. The pilot results found that the database had made significant improvements in the intervening years, and that "using the GeoDirectory....does not pose an under-coverage threat to the Census" (CSO 2010, p.5).

As part of Census 2011 preparations, an Enumeration Register was constructed on the basis of the April 2010 version of the GeoDirectory and was subsequently updated in September 2010. The register included all addresses attached to residential buildings. Provision was made for addition of about 4,700 commercial address points where the address string suggested the existence of a communal establishment or a place of accommodation for homeless persons.

The resulting 2,017,297 residential address points identified from the GeoDirectory were each assigned a unique dwelling number within the Enumeration Area (EA). The dwelling numbers were labelled on all of the enumerator maps. The register was then used as the basis for printing the enumerator address lists which contained the pre‐assigned dwelling numbers and addresses of the GeoDirectory points displayed on the map. Enumerators were then given specific instructions to record the address, EA code and dwelling number on the front panel of the Census Form before handing it over to the respondent, thereby achieving linkage between the spatial coordinates on the GeoDirectory and the data on the Census Form.

Furthermore, dwellings found by the enumerators which are not on the GeoDirectory are to be digitised by Census Geography staff using the enumerator maps. Therefore all dwellings on the Census 2011 file will be 100% spatially referenced and that an aggregated or 'bottom-up'

grid dataset can be produced. Such a dataset provides a number of opportunities to the CSO such as:

- Facilitating the identification of new Census towns and extensions to existing towns and cities.
- Supplying data to Eurostat for the GEOSTAT and 'Degree of Urbanisation' projects (Poelman 2010).
- Allowing for the possibility of an all-Ireland grid based dataset

However before Irish grid based statistics can be disseminated the following considerations will need to be explored:

- Experiences faced by other Countries which have already released statistical grid based data.
- The Irish legislative and CSO policy framework regarding disclosure of information and data confidentiality.

The 2011 Census was taken on April 10th this year, and Census Forms are currently being processed with a view to publishing the first definitive results in March 2012.

CONCLUSION

Given the proximity regarding the availability of Census 2011 results, it is not intended to release the grid population data from Census 2006. There are also quality issues with the Census 2006 grid dataset due to processing errors in the geocoding project and missing buildings on the July 2006 GeoDirectory. Instead the CSO intend to supply grid based population data from the 2011 Census to both Eurostat and the EFGS, which will be completely aggregated.

However the work in producing a hybrid grid dataset using 2006 Census data has proved informative in building up the required knowledge within the CSO on grid based statistics. In addition, it provides an example that local knowledge of the geographical areas and spatial databases available within a particular Country is a crucial element if the aims of producing a pan-European grid dataset are to be achieved.

REFERENCES

Central Statistics Office 2005, *Report of 2004 Census Pilot Survey*, <http://www.cso.ie/census/documents/2004_census_pilot_test_report.pdf>

Central Statistics Office 2010, *Report of Census Pilot Survey carried out in April 2009*, <http://www.cso.ie/census/documents/Report_%20of_%20Census_%20Pilot_%20of_%20April_%202009.pdf>

Department of the Environment, Community and Local Government 2011, *Latest House Building and Private Rented Statistics*, <http://www.environ.ie/en/Publications/StatisticsandRegularPublications/HousingStatistics/>

ESRI Support Center 2003, *Create a grid polygon shapefile (FISHNET)*, <http://arcscripts.esri.com/details.asp?dbid=12807>

European Forum for GeoStatistics website, *ESSnet GEOSTAT project*, <http://www.efgs.ssb.no/ >.

European Forum for GeoStatistics website, *GEOSTAT population map 2010,* <http://www.efgs.info/data/GEOSTAT_Population_Map_2010.pdf/view>.

Keogh D (Dara.GeoDirectory@anpost.ie), 26 August 2011. GeoDirectory data for study. Email to D Corcoran (Dermot.Corcoran@cso.ie).

Poelman, H. (2010). A new urban-rural typology, developed using grid data. e-Proceedings of European Forum for Geostatistics Conference, Tallinn, Estonia. 5-7 October, 2010.