POSTAL GEOGRAPHY AND GEO-REFERENCING

August 2024

Northern Ireland Statistics and Research Agency
Gníomhaireacht Thuaisceart Éireann um Staitisticí agus Taighde

1. Introduction

This document provides background information on postal geography, geographic referencing and summarises the issues on using postal geography for geographic referencing.

2. Background to Postal Geography

The Royal Mail maintains a UK-wide system of postcodes to identify postal delivery areas. Most people know their postcode, thus allowing statistics to be created based on their postcode as a main geographic reference. This reference can be related to geographic units used for statistical production such as a Local Government Districts or Electoral Wards. Postal Geography is thus very valuable.

2.1 Postcode Structure

Postcodes are alphanumeric references comprising of two halves – the first half is an outward code of 2-4 characters and the second half is an inward code of three characters. When mail is collected and taken to a Mail Centre, the first half of the postcode (Outward code) tells Royal Mail which delivery office to transport the mail to. There are 39 Mail Centres in the UK, but only one in Northern Ireland (Mallusk). At the delivery office, the second half of the code (Inward code), combined with the building name and/or number, provides Royal Mail with the location of the delivery point.

The postcode is structured hierarchically, supporting four levels of geography unit:

- **Postcode Areas:** There are 124 in the UK and Northern Ireland comprises one Postcode area represented by the code BT;
- **Postcode Districts:** Northern Ireland has 81 Postcode Districts;
- **Postcode Sectors:** Northern Ireland has 322 Postcode Sectors with around 3,000 addresses per Sector; and
- **Postcode Units:** Northern Ireland has 62,462 postcode units comprising an average of 21 addresses per unit.

Table 2.1 shows how an example postcode, BT9 5RR, is disaggregated.

Table 2.1. Levels within postcode hierarchy using example of BT9 5RR.

Geography Unit	Number in Northern Ireland	Postcode part (Example)
Postcode Area	1	BT
Postcode District	81	BT9
Postcode Sector	322	BT9 5
Postcode Unit	62,462	BT9 5RR

2.2. Postcode Units (or Unit Postcodes)

These 62,462 postcode units cover around 919,012 delivery points, which at July 2024 comprise approximately 917,697 small user and 1,315 large user postcodes. Unit postcodes are the base unit of postal geography and fall into two types:

- **Large User postcodes:** allocated to single addresses receiving at least 500 mail items per day (for example, business addresses); and
- **Small User postcodes:** are collections of usually adjacent addresses. A single small user postcode may contain up to 100 addresses but 15 to 25 are more typical numbers.

It is possible for large buildings with many separate delivery points (for example, a block of flats) to have more than one unit postcode within the building.

3. Geographic referencing

3.1 General

The production of statistics involves the collection, processing, and output of statistical data. Most data events can be referenced to a known location, and this means that most statistics can be output to or allocated to geographic classifications. For example, we might produce statistics of unemployment by Electoral Ward, or birth statistics for each Local Government District.

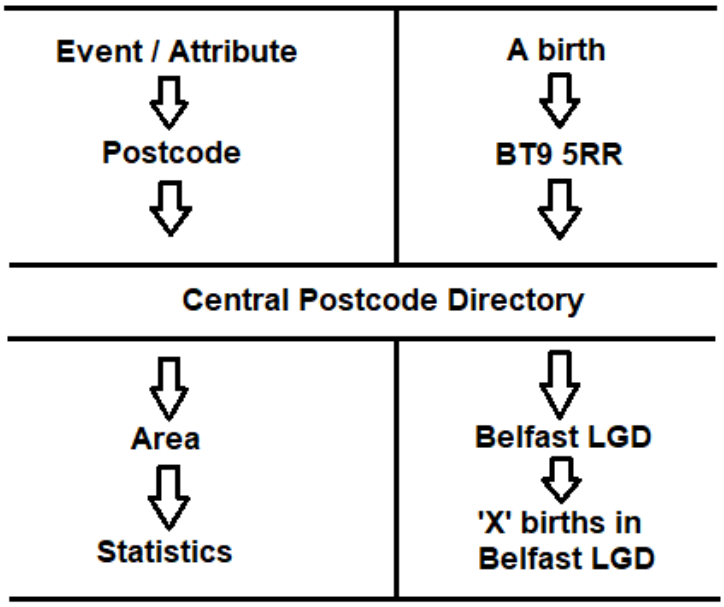
3.2. Postcode geo-referencing

Geographic referencing (or 'geo-referencing') is an increasingly important process in the production of statistics, allowing greater data accuracy and facilitating the sharing and aggregation of data. In recent times the approach to geographic referencing has been to use the postcode.

Through the use of Geographic Information System (GIS), NISRA's Central Postcode Directory (CPD) provides the grid reference of the property closest to the average grid reference for all the properties within the postcode, i.e. the address-weighted centre of the postcode. This is commonly known as the postcode centroid (more information on how individual postcode centroids are allocated to postcodes on the CPD is provided in the CPD Guidance Notes). This is a good start and may be the most accurate reference possible, as we may not have any more detailed locational information for the data event. The traditional method of referencing data to an event postcode has a number of advantages.

First, most people know their postcode so can readily supply it when asked, for example, when responding to a survey. Secondly, postcode directories such as NISRA's CPD can be used as a ready means of matching each postcode to a range of geographic areas. An example of this is shown in Figure 3.1 below. In this case, the postcode has been matched through the CPD to a Local Government District, although the directory goes down to smaller geographies like the Census 2021 Data Zones and Super Data Zones.

Figure 3.0. Postcode referencing example.



3.3. Limitations of using postcodes as a geographic reference

Postcodes form a compact geographic reference with which the public and businesses are familiar. Although postcode referencing is relatively straightforward, it has some limitations. A number of these are outlined below.

(i) Postcodes do not map directly to other geographic areas

The postcode geography does not take account of administrative boundaries – postal geography overlaps or straddles other boundaries. This straddling of boundaries means that many postcodes can only be assigned uniquely to administrative areas on a 'best fit' basis. If a unit postcode straddles a ward boundary, the CPD allocates this postcode to one ward only.

Figure 3.1 shows an example where all addresses (blue) with postcode BT4 2WB are on the same side of the purple boundary. The centroid of these addresses is represented by a green point. Figure 3.2 shows all addresses with postcode BT4 2WE, which are located on either side of the purple boundary. The green centroid of these addresses lies just below that boundary, hence all addresses with this postcode will be allocated to that area. The result is that addresses lying close to administrative boundaries are sometimes assigned to the wrong area.

Figure 3.1. Postcode with addresses wholly contained within one area.

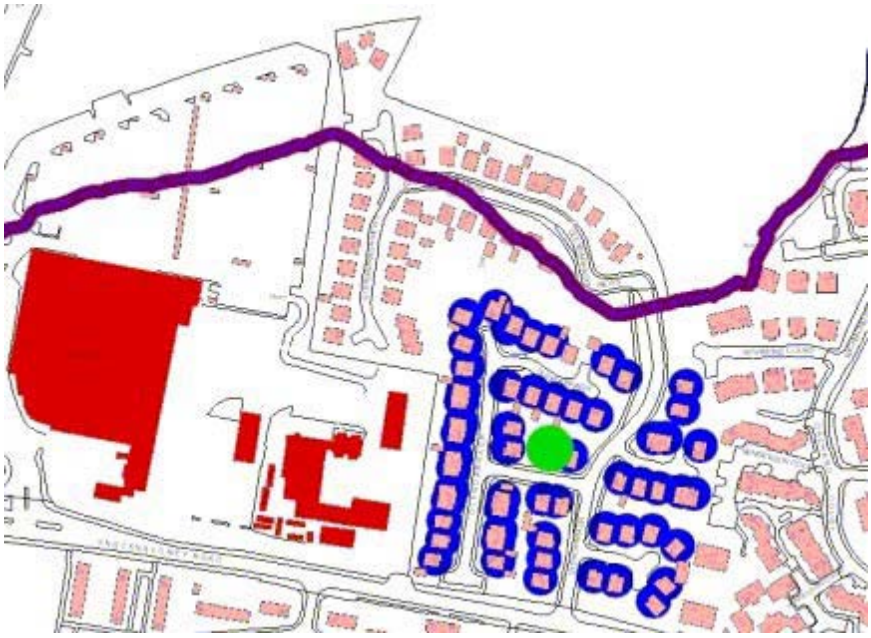
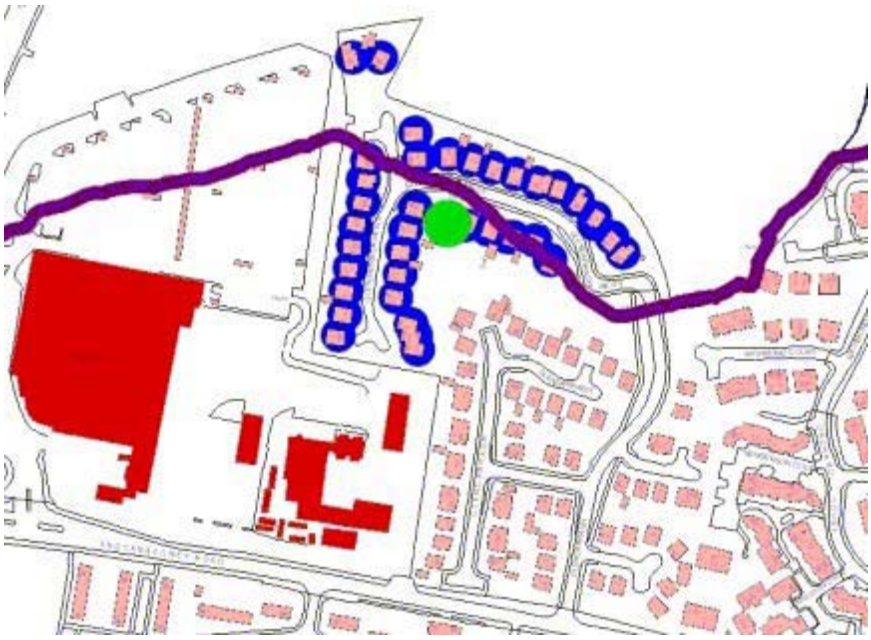


Figure 3.2. Postcode with addresses straddling two areas.



This issue becomes more apparent for very small areas, as a larger proportion of all postcodes have addresses that straddle boundaries. This issue can be quantified by comparing either Census 2021 or Census 2011 statistics by postcode in combination with the Central Postcode Directory (July 2024), and published statistics.

Tables 3.2a and 3.2b show some key indicators of this exercise for several administrative and statistical geographies. The tables are split according to the basis of the accuracy assessment; the first six geographies use Census 2021 usual resident population and household counts, whereas the three geographies in the bottom section use the corresponding counts from Census 2011. As an example, for half of the 850 Super Data Zones (SDZ2021) the difference between the published population count and that based on aggregating the populations of postcodes assigned to that area is less than 2.1 per cent (see median figure highlighted). For one in ten Super Data Zones (9th decile), this difference is greater than 6.5 per cent.

For the larger administrative geographies – Local Government Districts (LGD2014), Assembly Areas/Parliamentary Constituencies (AA2024) and District Electoral Areas (DEA2014) – the postcode-based lookup returns estimates that are within one per cent of the published figure for almost all areas. As areas become smaller, the postcode-based lookup becomes less accurate, due to a larger proportion of postcodes expected to straddle area boundaries. Note that the published Census 2021 usual resident population and household counts for AA2024 are themselves approximated (see Aggregating Data Zone to produce statistics for higher-

level geographies); as a result, despite being a larger geography, the postcode approximation of AA2024 is less accurate than for DEA2014. For the smallest Census 2021 statistical geography of Data Zone, postcode approximations of the population for circa 3,020 out of a total of 3,780 areas (80 per cent) are within 16 per cent of the published figure, whilst for half of the Data Zones the approximation is broadly within seven per cent.

Table 3.2. Accuracy of Central Postcode Directory (CPDjul24) in estimating Census 2021 or 2011 usual resident (a) population and (b) households.

(a)

Geography	Number of areas	Average	Median	8th decile	9th decile
LGD2014[1]	11	*0.0%*	*0.0%*	*0.1%*	*0.1%*
AA2024[1,3]	18	*0.3%*	*0.3%*	*0.6%*	*0.8%*
DEA2014[1]	80	*0.3%*	*0.2%*	*0.5%*	*0.6%*
WARD2014[1,4]	462	*1.5%*	*1.2%*	*2.4%*	*3.2%*
SDZ2021[1]	850	*2.9%*	*2.2%*	*4.7%*	*6.4%*
DZ2021[1]	3,780	*9.7%*	*7.1%*	*16.0%*	*21.8%*
WARD1992[2]	582	*1.8%*	*1.3%*	*2.7%*	*3.9%*
SOA2001[2]	890	*2.3%*	*1.7%*	*3.5%*	*5.3%*
SA2011[2]	4,537	*5.9%*	*3.3%*	*9.9%*	*15.2%*

(b)

Geography	Number of areas	Average	Median	8th decile	9th decile
LGD2014[1]	11	*0.0%*	*0.0%*	*0.0%*	*0.1%*
AA2024[1,3]	18	*0.4%*	*0.3%*	*0.6%*	*0.8%*
DEA2014[1]	80	*0.3%*	*0.2%*	*0.4%*	*0.6%*
WARD2014[1,4]	462	*1.4%*	*1.1%*	*2.4%*	*3.2%*
SDZ2021[1]	850	*2.7%*	*1.9%*	*4.3%*	*6.0%*
DZ2021[1]	3,780	*9.3%*	*6.8%*	*15.1%*	*20.9%*
WARD1992[2]	582	*1.7%*	*1.2%*	*2.7%*	*3.6%*
SOA2001[2]	890	*2.2%*	*1.6%*	*3.5%*	*4.9%*
SA2011[2]	4,537	*5.5%*	*3.1%*	*9.4%*	*14.0%*

[1] Accuracy assessment based on Census 2021 usual resident population/household counts
[2] Accuracy assessment based on Census 2011 usual resident population/household counts
[3] Census 2021 usual resident population/household counts approximated via Data Zone aggregation
[4] Census 2021 usual resident population/household counts approximated via grid square aggregation

The difference between the published Census 2021/2011 usual resident population and households counts and those approximated by using postcodes for each area of the geographies presented in Tables 3.2a and 3.3b is available on the NISRA website. This enables users of the CPD to assess whether the approximation by postcodes is acceptable for their purposes.

Whilst the size of this error is relatively small it should be considered when analysing data for small geographic areas, especially when properties lie adjacent to administrative boundaries. Also, as demonstrated in the two tables, the reported accuracy depends on the choice of statistics, i.e. population or households. This is particularly important when looking at relatively rare events or spatial concentrations.

(ii) Postcodes can move around and are subject to change

Royal Mail assigns postcodes to address locations for the sole purpose of providing an efficient mail delivery service. Postcode boundaries are subject to continuous change due to new addresses, single addresses acquiring large user postcodes as mail volume increases,

and the need to restrict the number of addresses per unit to less than 100. Areas can also be re-coded and in some instances terminated codes can be re-used in a different place after just two years. Wrongful allocations can occur and continuous monitoring is therefore required to minimise this.

Although NISRA's CPD retains discarded postcodes (terminated postcodes), this cannot be regarded by itself to provide an accurate locational reference. Royal Mail may occasionally decide to re-use these discarded postcodes in another part of the same postcode sector and thus the physical location of the postcode may shift. This could cause data to be assigned to the wrong area unless care is taken to use the relevant postcode directory.

(iii) Accuracy issues

The postcode geography constantly changes to reflect what is happening on the ground. Demolition, redevelopment and more acutely the development of new properties means that Royal Mail are constantly adjusting the postcode geography by creating new postcodes, revising the content of existing postcodes and as we have seen even terminating postcodes. Although NISRA routinely receives and publishes information about changes to the postcode geography, users need to be aware of issues around the time lag between changes on the ground and the updating of spatial information.

When a new postcode is created it takes considerable time for Royal Mail to gather all the accurate spatial data for that postcode. For example, the individual grid references of the constituent addresses, which enable calculation of the postcode centroids, will not be available to Royal Mail until Ordnance Survey NI (OSNI) has surveyed those areas and updated the Pointer Address database. This process, in some cases, can take many months.

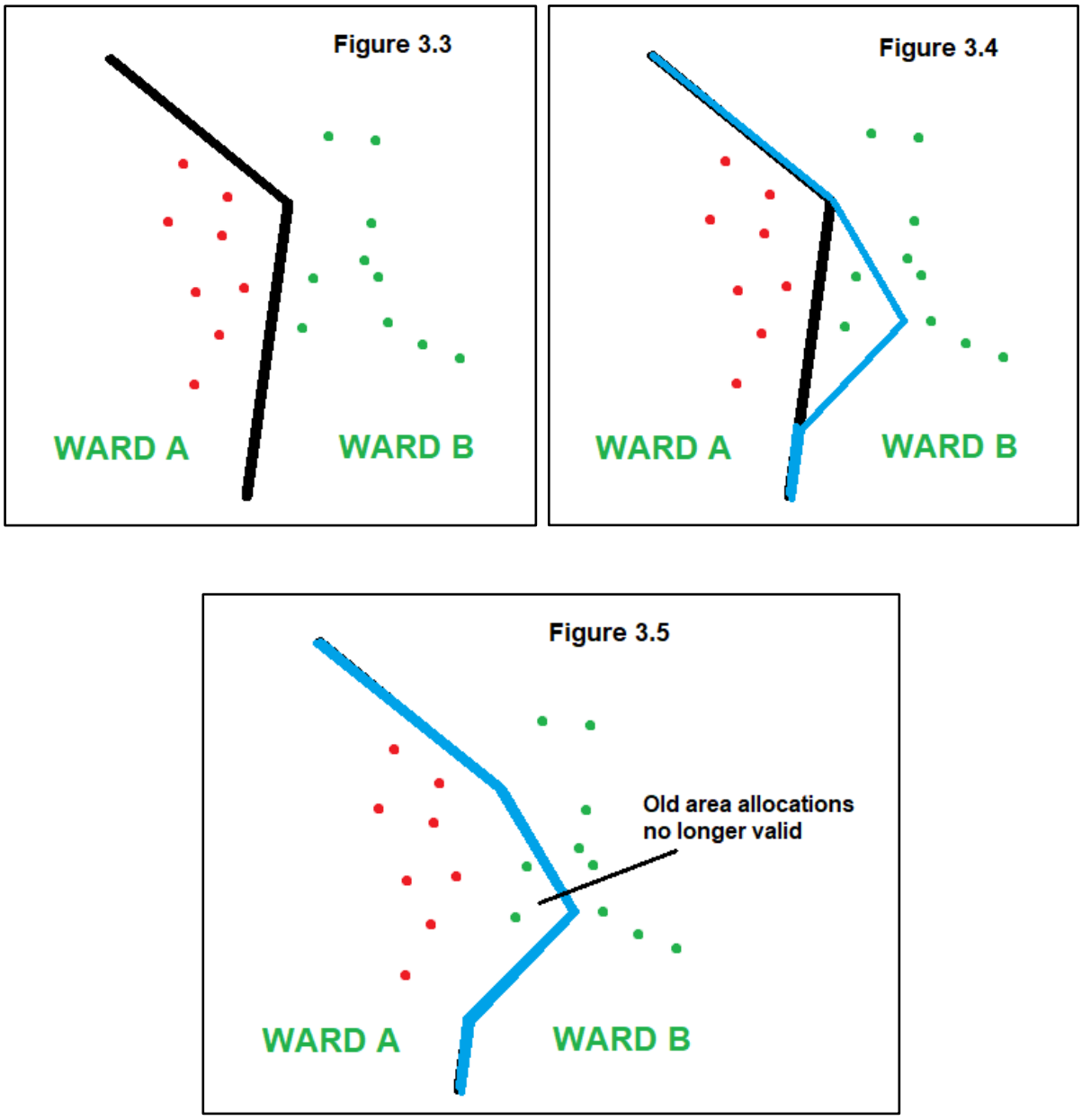
However, to get around this issue newly introduced postcodes will initially have a temporary grid reference attached to them. This imputed temporary grid reference is usually allocated on a nearest-neighbour basis. The method used is such that the closest postcode numerically/alphabetically is used to assign a grid reference to the new postcode. In due course this temporary grid reference will be replaced by the improved Pointer based grid references from data supplied by OSNI.

Users need to take cognisance of this time lag and understand that the grid references temporarily assigned for some postcodes, particularly new postcodes, may be spatially inaccurate. Given that the imputation process is not an exact science, the result is that, in

some cases, the temporary grid referencing may assign postcodes to the wrong administrative, statutory or statistical areas.

(iv) Statutory boundaries change

Statutory and administrative boundaries in Northern Ireland periodically change; for example, the 2014 Electoral Ward Boundary Review revised the 1992 Electoral Ward boundaries. This further complicates postcode to area referencing. Take Figures 3.3-3.5 below for example.



Figures 3.3-3.5. Postcode referencing and boundary change.

In Figure 3.3 the black line is the boundary between Ward A and Ward B, with the red and green points showing the properties in two different postcodes. The configuration of the postcodes is such that all properties have been allocated to the correct ward, i.e. all red properties lie within Ward A and all green properties within Ward B. Once the ward boundary has changed as shown by the blue line overlaid on the original boundary in Figure 3.4, then the allocation of some green properties is incorrect as one of the postcodes is now straddling both wards. All of the green properties in the split postcode arising from the new ward boundary (Figure 3.5) are referenced to one ward based on the location of the postcode centroid; consequently, some of the green properties are assigned to the incorrect ward.

3.4. Geo-referencing using address-level grid reference

With the availability of the OSNI Pointer Address database the move to address level grid referencing may be facilitated. This is extremely powerful and accurate. Whereas the postcode centroid gives an approximate location of the data event, the address level grid reference describes precisely where it occurs. This has several advantages:

- Straddling is no longer an issue, as addresses rather than postcode are used to allocate geographical identifiers. By using the address grid reference event, addresses are allocated to the correct geographic areas.
- Dealing with boundary change is easier. New boundary datasets can be loaded into GIS platforms and, knowing the events are precisely located, we can, very quickly, produce accurate statistics for the new boundaries.
- Outputs and analysis can be more flexible and can be produced for virtually any geography (with due cognisance taken of confidentiality issues).

However, although address level grid referencing is powerful, it does have its limitations:

- The automated allocation of addresses to grid references is more difficult than it is for postcodes. This is because, unlike postcodes, addresses can be lengthy, complicated and inconsistent. For example, the first line of an address may be the building number and street name, the number of a flat within a building, or the name of a property. However, this could be overcome if data collectors used a standard address dataset and the OSNI Pointer address database is an appropriate tool to facilitate this.
- As noted for postcode-based methods, there are also time lags in recording changes on the ground, which can lead to lack of completeness of an address database and thus inaccuracy.

- As data relates to individual addresses, greater security precautions may be required to protect the confidentiality of individuals.

4. Conclusion / Further information

The approach of using postcodes to reference geographic data has been a valuable tool but is subject to a number of limitations, especially when trying to produce statistics for very small areas. The move towards geographic referencing based on the postcode centroid offers many advantages in terms of facilitating event linkage, data visualisation and data analysis. However, it does not eliminate the problems caused by postcodes straddling around boundaries and indeed boundary changes. If a reference can be given at address-level however, the potential is even greater, allowing for detailed and accurate small-area statistics. Different types of data will of course require different types of referencing, and issues such as ensuring confidentiality are crucial.

Further information on geo-referencing and the Central Postcode Directory is available on the NISRA website. Alternatively, you can get in touch with NISRA's Geography team:

NISRA Census Office
Colby House
Stranmillis Court
Belfast
BT9 5RR

E-mail: geography@nisra.gov.uk