

0. Introduction

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

1. 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.

1.1 Postcode Structure

Postcodes are alphanumeric references comprising 2 halves – the first half is an outward code of 2-4 characters and the second half is an inward code of 3 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. At the delivery office, the second half of the code (Inward code), combined with the building name and number, provides Royal Mail with the location of the delivery point.

Example: Postcode BT34 4AT comprises two halves:

BT34

Outward code



(Used to sort mail to Delivery Office)

4AT

Inward code



(Used to help Royal Mail identify delivery point)

The postcode is structured hierarchically, supporting 4 levels of geography unit:

- **Postcode Areas** (There are 124 in the UK and Northern Ireland comprises 1 Postcode area represented by the code BT),
- **Postcode Districts** (Northern Ireland has 82 Postcode Districts),
- **Postcode Sectors** (Northern Ireland has 268 Postcode Sectors with around 3,000 addresses per Sector), and
- **Postcode Units** (Northern Ireland has over 56,000 postcode units comprising an average of 20 addresses per unit)

The Postcode BT34 4AT is disaggregated as follows:

Geography Unit	Number in Northern Ireland (June 2005)	Postcode part (Example)
Postcode Area	1	BT
Postcode District	82	BT34
Postcode Sector	268	BT34 4
Postcode Unit	56,114	BT34 4AT

1.2 *Postcode Units (or Unit Postcodes)*

These 56,114 postcode units cover over 760,000 delivery points, which at June 2005 comprise approximately 757,000 small user and 3,000 large user postcodes. Unit postcodes are the base unit of postal geography and fall into two types:

- **Large User postcodes:** are allocated to single addresses receiving at least 500 mail items per day (e.g. 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 – 25 are more typical numbers.

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

2. *Geographic referencing*

2.1 *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.

2.2 *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 GIS, NISRA’s Central Postcode Directory (Version CPD June2005) now 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, which 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:

- Most people know their postcode so can readily supply it when asked e.g. when responding to a survey
- Postcode directories such as NISRA’s Central Postcode Directory (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 Table 2.1 below. In this case the postcode has been matched through CPD to a Local Government District, although the directory goes down to smaller geographies like the Census Output Areas (OAs) and Super Output Areas.

Table 2.1: Postcode referencing – example

Event / Attribute ↓ Postcode ↓	A birth ↓ BT34 4AT ↓
Central Postcode Directory	
↓ Area ↓ Statistics	↓ Newry & Mourne LGD ↓ 'X' births in Newry & Mourne LGD

2.3 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*

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 Central Postcode Directory allocates this postcode to one ward only. Examples of this are shown in Figures 2.1 and 2.2. below. The result is that addresses lying close to administrative boundaries are sometimes assigned to the wrong area.

Figure 2.1: Postcode with addresses wholly contained within one area.

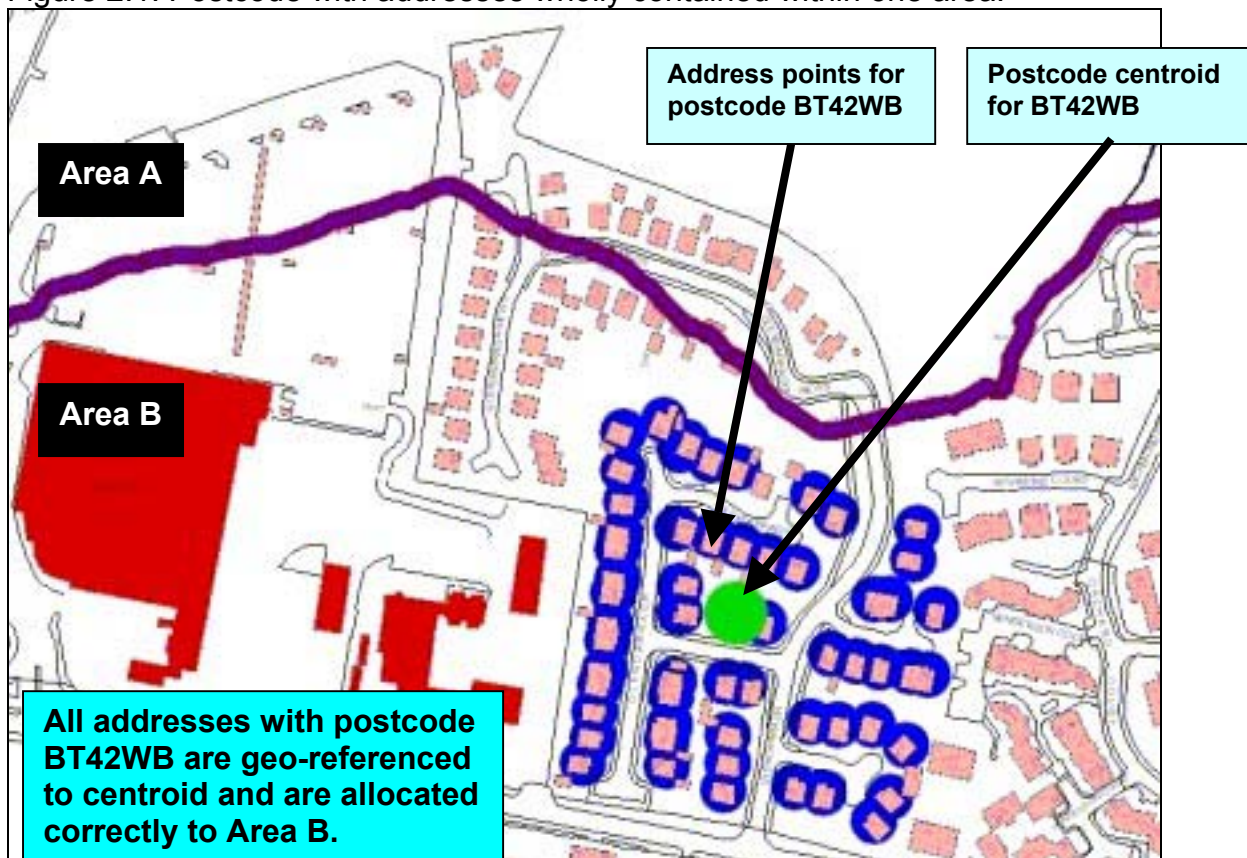
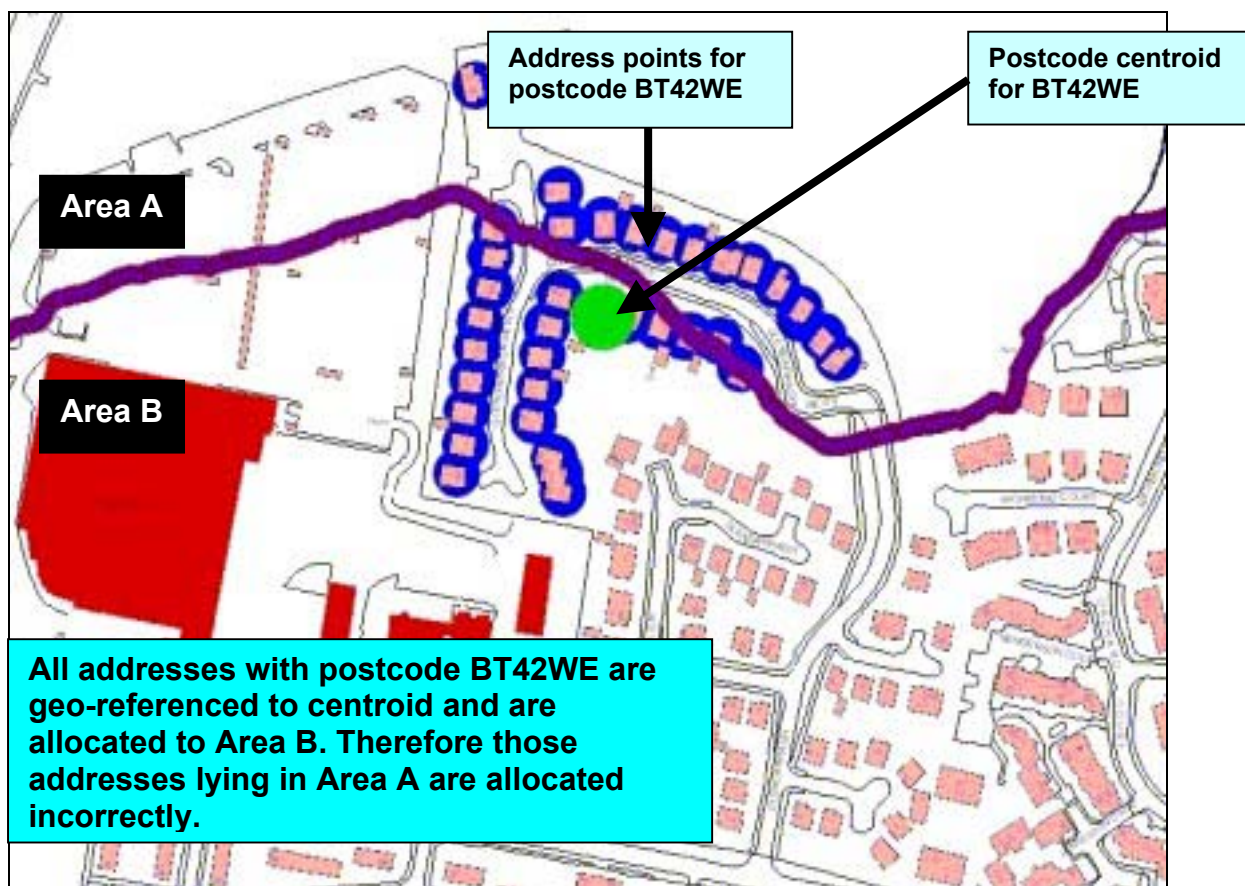


Figure 2.2: Postcode with addresses straddling two areas.



For very small areas, such as Output Areas, using postcode based geo-referencing is estimated to leave around 5% of addresses allocated to the incorrect area. This misallocation is smaller for both electoral wards, at around 2.5% of addresses and Super Output Areas at around 3%. Whilst, the size of this error is relatively small it should be considered when analysing small area data, especially when properties lie adjacent to administrative boundaries.

(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 increase, 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. Mis-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) Currency and spatial accuracy issues

The postcode geography constantly changes to reflect what is happening on the ground. Demolition, re-development 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 there are issues in respect of currency, which users need to be aware of.

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 the Ordnance Survey NI 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 currency issue 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 – e.g. the 1992 Electoral Ward Boundary Review revised the 1984 Electoral Ward boundaries. This further complicates postcode to area referencing. Take Figures 2.3-2.5 below for example. In this case the configuration of the postcodes is such that all properties have been allocated to the correct ward. Once the ward boundary has changed, then the allocation of some properties is incorrect as one of the postcodes is now straddling both wards. All properties in the split postcode will end up referenced to either Ward A or Ward B and this means that a proportion of them are bound to be wrong.

Figures 2.3 – 2.5 Postcode referencing and Boundary change

Fig 2.3

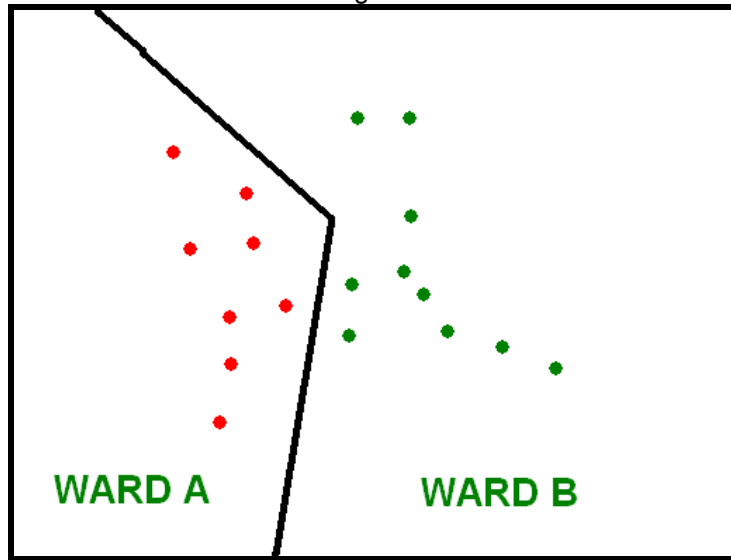


Fig 2.4

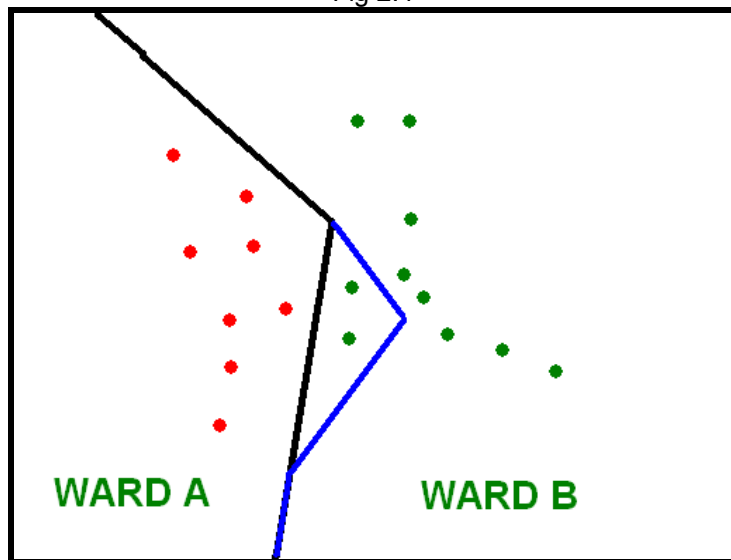
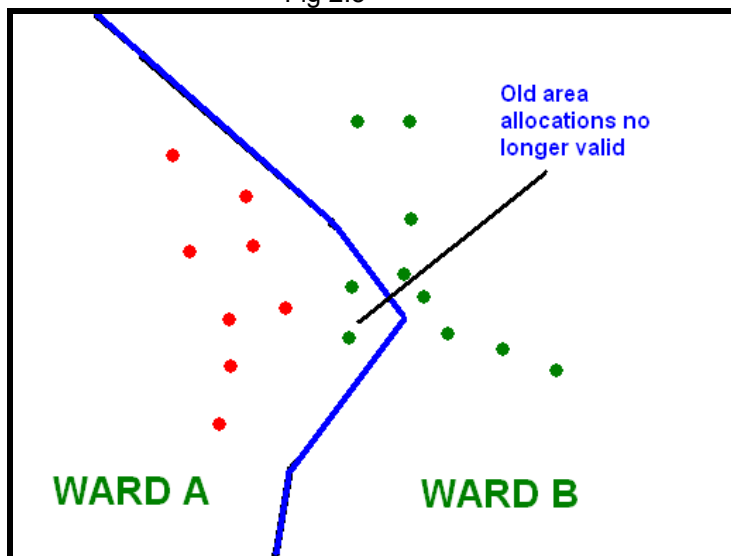


Fig 2.5



2.4 *Geo-referencing using address-level grid reference*

With the recent development and release 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:

- Not all data can be assigned to an address – the automated assignment of 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 currency problems 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.

3. 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 straddling and boundary change. 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 NISRA's Central Postcode Directory is available from:

Stephen Smith
Demography and Methodology Branch
NISRA
McAuley House
2-14 Castle Street
Belfast BT1 1SA

Tel: 028 90 348170
Fax: 028 90 348134
E-mail : Stephen.smith@dfpni.gov.uk

NISRA
December 2005