

Notes on Scottish Postcode Directory Quality Checking

Gerald Leung

Public Health Scotland

Draft version: March 2, 2022

Contents

1 Introduction	1
2 Initial Quality Checks	2
3 Lookup Files	3
A Appendix: Data Dictionary	4

1 Introduction

In this short document, we summarise the process of the Scottish Postcode Directory (SPD) file checks from the National Records of Scotland (NRS) by the Geospatial Team at Public Health Scotland (PHS). In short, NRS updates SPD files twice a year, around March and September¹. The Geospatial Team will then conduct quality checks of the updated files and compare any changes with the previous version. They will then produce lookup files in various formats, such as `.csv`. This document focuses on summarising the processes of quality checking (QC) and lookup files production using R. Here we also reproduce a brief summary of the data, which can be found in Appendix A. Full details can be found through the [data dictionary](#) provided by the NRS, or accessed through

```
//PHIBCS/PHI/Referencing & Standards/GPD/1_Geography  
/Scottish Postcode Directory/Source Data/2021_2  
/ISD Data Dictionary_2021-2
```

for the **2021-2** version.

¹<https://www.nrscotland.gov.uk/statistics-and-data/geography/nrs-postcode-extract>

2 Initial Quality Checks

The R scripts are located in

```
//PHIBCS/PHI/Referencing & Standards/GPD/5_GitHub/GPD/Geography  
/Scottish Postcode Directory
```

and the Standard Operating Procedures (SOPs) are located in

```
//PHIBCS/PHI/Referencing & Standards/GPD/1_Geography  
/Scottish Postcode Directory/SOPs.
```

The source files can be found in

```
//PHIBCS/PHI/Referencing & Standards/GPD/1_Geography  
/Scottish Postcode Directory/Source Data.
```

They can be accessed through the organisation's VPN. The first QC is done through the R script `1_Check NRS SPD.R`.

In short, this script conducts a general initial QC of the newest version of SPD files and compares changes with the previous version. Figure 1 shows roughly the structure of the script and the QC process.

To start with, all empty rows contained in the files are removed and the files are combined. The user will then record the total number of records and postcode types. Sanity checks are conducted to make sure there are no blank entries for Health Board and Council Areas etc. Checks are then carried out to make sure columns are aligned correctly. For example, between Health Board Area Code and Council Area Code columns. This is generally done by assigning a number to the case when a particular column does not have an expected code that should match with the other column. During 2019, eight postcodes moved from Glasgow City Council to North Lanarkshire Council. Extensive checks are then carried out to make sure there are no blank fields. Checks are also carried out to make sure there are no English voting codes included. Finally, the script makes sure that mappings are done correctly, such as Data Zones and Health Boards.

The script also makes comparisons with previous SPD files. For instance, it checks for changes in postcodes (some have been deleted), different data zones and health boards etc.

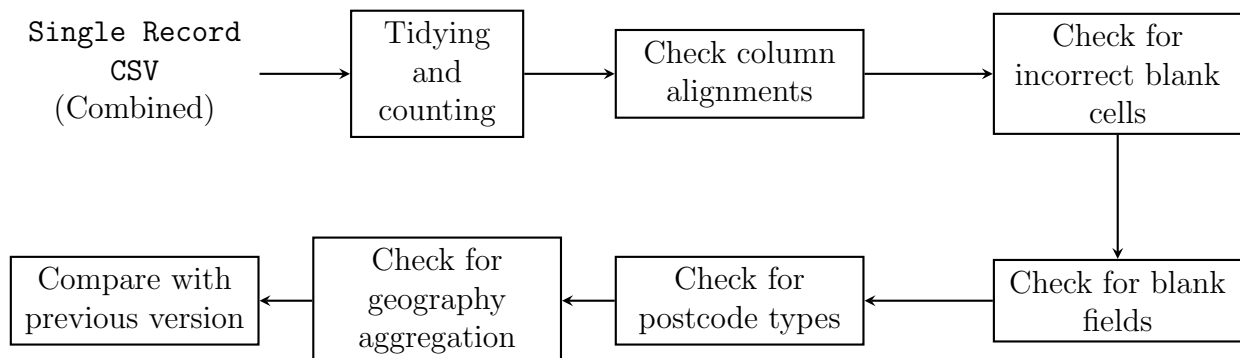


Figure 1: Schematic diagram of the initial QC process.

3 Lookup Files

The next step of the process is to produce lookup files for other uses and analysis within the organisation.

A Appendix: Data Dictionary

Name	Type	Notes
Postcode	Character	The Royal Mail postcode. Consists of area, district, sector and unit. E.g. AB1 0AA.
SplitChar	Character	A, B, C
PostcodeDistrict	Character	Postcode District. E.g. AB1.
PostcodeSector	Character	Postcode Sector. E.g. AB1 0.
DateOfIntroduction	Date	Postcode introduction date. Currently shown as #### instead of a date.
DateOfDeletion	Date	Postcode deletion date. Similarly to date of introduction.
PostcodeType	Character	Small (S) or Large (L).
LinkedSmallUserPostcode	Character	Small User postcode that contains the grid reference of the Large User postcode.
LinkedSmallUserPostcodeSplitChar	Character	A, B, C
Imputed	Character	Whether postcode's higher area values have been imputed.
DeliveryPointCount	Integer	The Royal Mail delivery point count.
DeliveryPointCountNonResidential	Integer	Non-residential delivery point count.
HouseholdCount	Integer	The Royal Mail Household count.
GridReferenceEasting	Character	Grid reference easting.
GridReferenceNorthing	Character	Grid reference northing.
Latitude	Numeric	Coordinates in degrees.
Longitude	Numeric	Coordinates in degrees.
SplitIndicator	Character	Split postcode indicator. Y/N.
CouncilArea2019Code	Character	Identifies a 2019 Council Area.
UKParliamentaryConstituency2005Code	Character	2005 UK Parliamentary Constituency.

ScottishParliamentaryRegion2021Code	Character	2021 Scottish Parliamentary Region.
-------------------------------------	-----------	-------------------------------------

Table 1: Data dictionary of the SPD file. Reproduced from NRS.