# Population Estimates Comparison Project

### Data Sources Review

**Published on 20 January 2014**

# EXECUTIVE SUMMARY

This report documents and reviews the data sources that have contributed to the work of the National Records of Scotland's (NRS) Population Estimates Comparison Project. A further aspect of the report is to identify additional sources of information that could potentially contribute to the work of the project. As the Population Estimates Comparison Project is also a pathfinder for the Scottish Beyond 2011 Programme, the findings from this report will also feed into the research phase of the Beyond 2011 Programme.

The following data sources have been reviewed within this report:

- Census
- Mid Year Estimates (MYE)
- Community Health Index (CHI)
- NHS Central Register (NHSCR)
- One Scotland Gazetteer
- NRS Address Data
- The Citizen's Account (CS)
- Glasgow City Council Tax Data

The background and purpose of each data source has been documented, alongside a review of the quality, accuracy, coverage and inputs/outputs of each data source where appropriate. The report includes data sources that have already featured in the work of the Population Estimates Comparison Project and those data sources that are being considered for future inclusion in the work of the project.

This report will be further expanded as additional data sources are examined for inclusion within the work of the Population Estimates Comparison Project.

# Contents

# 1. Introduction

## 1.1 Background

NHS funding in Scotland is allocated across the fourteen NHS Boards (information available on the NHS Scotland website) using a resource allocation formula. This NHS Scotland Resource Allocation Committee (NRAC) formula which is available on the TAGRA website uses population information derived from the Mid Year Estimates (MYEs) (available on the National Records of Scotland (NRS) website) to help determine the share of resources each NHS Board should receive. The population share has the greatest influence over the final allocation of resources and very small changes to this share can potentially lead to large changes in funding allocations.

The Community Health Index (CHI) is a register of NHS patients in Scotland and can provide an additional count of the resident population to the MYE. Due to different recording practices and purposes, it is recognised that the CHI suffers from over-inflation compared to the MYE due to 'ghosts' and duplicates in the system and how short-term migrants are recorded. However, a particularly large disparity has been noted between the population estimate taken from the CHI and the MYEs for NHS Greater Glasgow and Clyde (NHS GG&C). For example, NHS GG&C have proposed that if the CHI were to be used in the resource allocation formula instead of the MYEs, then NHS GG&C would gain an additional £80m per annum.

Due to the funding implications, NHS GG&C are keen to investigate the disparity between the two data sources, particularly as there is a concern that this over-inflation could be due to a considerable number of unknown people accessing health care services who have not been included in the census (which acts as the base for the MYEs). As a result, NHS GG&C have asked National Records of Scotland (NRS) to lead an investigation into the differences between the population estimates derived from the CHI and the MYEs, resulting in the creation of the Population Estimates Comparison Project ('the Project').

Phase 1 of the Project will focus primarily on the population of NHS GG&C which has the largest disparity between the CHI and the MYE. However, a wide range of other learning opportunities are envisaged and any relevant findings will be investigated and applied across Scotland as a whole in Phase 2. The Project will also act as a pathfinder for the comparison of population estimates taken from different administrative data sources. This is an important aspect of the Beyond 2011 Programme which is concerned with the future provision of population and socio-demographic information. Further information about the programme can be found on the NRS website.

## 1.2 Purpose of Report

By comparing only aggregate population estimates from the CHI and the MYEs, it is not possible to say with any certainty which source is more accurate. It is feasible that the CHI is overestimating the population or that the MYE is underestimating it. It is therefore necessary to look at other available data sources to help formulate a better understanding of the differences between population estimates as calculated by NRS (the census and the MYEs) and as derived from administrative data sources (i.e. the CHI, NHSCR, etc).

The purpose of this report is to review and assess both the potential and tangible contributions different data sources have brought to the Project. The document will outline the background and methodology behind each data source, including a description of any differences between the various sources or any relationships and inter-dependencies that exist between certain datasets.

The following list of data sources have already been included in the work of the Project:

- Census
- Mid Year Estimates (MYE)
- Community Health Index (CHI)
- NHS Central Register (NHSCR)

The secondary list below identifies additional sources of information that could potentially contribute to the work of the project, such as the possibility of conducting a data linkage exercise using council administrative data sources in order to create a population count through record linkage:

- One Scotland Gazetteer
- National Records of Scotland (NRS) Address Data
- The Citizen's Account
- Glasgow City Council Tax Data

The information in this report was derived from a series of meetings that took place between the NRS Data Access and Communications Team and a range of groups and individuals relevant to the Project. The section that has been composed on each data source has been reviewed by the relevant area of responsibility to ensure that the information that has been provided is accurate. This is a living document and will be updated as additional data sources are examined throughout the lifecycle of the Project.

## 1.3  Supplementary Information

A range of administrative data source reviews have been completed in recent years. Relevant links to additional information which supplement and enhance the findings in this report are provided below:

- The Office for National Statistics (ONS) have produced a series of PDF publications on their websites which review different administrative data sources as part of the England and Wales Beyond 2011 Programme:
  - A Conceptual Framework for UK Population and Migration Statistics (885 kb)

- National Records of Scotland (NRS) have carried out a number of data reviews. These papers have been published on the NRS website in PDF format and can be found in the links below:
  - Overview of Administrative Comparator Data Used in 2011 Census Quality Assurance (82kb);
  - Assessing Administrative Data: A Comparison of Population Counts from Aggregated Administrative Data and the Mid-Year Population Estimates (171kb); further information on:
  - Local Area Migration can be found within the Migration section of their website

## 2. Census

### 2.1 Overview

The census collects information on the number and characteristics of people and households in Scotland. It is used by central and local government, health authorities and many other organisations to allocate resources, plan services, develop policies, meet demands for population related information and meet international statistical obligations.

Scotland's Census published a paper on their website looking at how the 2011 Census population estimates were obtained (PDF document 99.78kb) provides a comprehensive review of the background, methodology, coverage, availability and outputs of the census.

## 3. Mid-Year Estimates (MYEs)

### 3.1 Overview

The MYEs are produced by the Population and Migration Statistics branch in NRS. They produce an annual population estimate of the usually resident population for Scotland and its constituent NHS Board and Council areas. As the MYEs are used in the NRAC (NHS Scotland Resource Allocation Committee) formula for calculating the resource allocation of NHS funding, they are fundamental to the work of the Project.

The Mid-Year Population Estimates for Scotland: Methodology Guide available withinin the Population Estimates section of the NRS website) provides a comprehensive review of the background, methodology, coverage, availability, outputs and future developments of the MYEs.

## 4. Community Health Index (CHI)

### 4.1 Overview

Information in this section is derived from a range of documents on the CHI and discussions with the Practitioner Services Division (PSD)[1] (www.psd.scot.nhs.uk/). Information is also included from discussion with other individuals from the Information Services Division (ISD) (www.isdscotland.org/).

### 4.2 Background

The CHI is an index system used throughout NHS Scotland. It contains patient demographic data; supports General Practitioner (GP) registration and records transfer and tracking; provides data for GP payments; manages and issues prepayment and exemption certificates; and supports links to many other healthcare systems including healthcare screening and surveillance programmes and hospital systems. There are many feeds from the CHI to other health systems, including record seeding exercises and daily CHI broadcast files that allow the

**Footnote**

1) The Practitioner Services Division has three Regional Offices. The Aberdeen Regional Office processes work on behalf of Tayside, Grampian, Highland, Orkney and Shetland. The Edinburgh Regional Office processes work on behalf of Lothian, Borders, Fife and Forth Valley. The Glasgow Regional Office processes work on behalf of Greater Glasgow & Clyde, Lanarkshire, Ayrshire & Arran, Dumfries & Galloway and Western Isles.

population and updating of local NHS systems with the CHI number and demographic data. The CHI number is used as the patient identifier throughout NHS Scotland and is important to the implementation of 'eHealth' Electronic Health Records (EHR) and other Information and Communications Technology (ICT) strategy in Scotland. Each patient is assigned a unique ten-digit CHI number, which helps to preserve confidentiality by reducing the use of patient identifying information.

The CHI was created in the mid 1970s in Tayside, and rolled out to the rest of Scotland in the 1980s with eight CHI databases areas (consortia) that cover the 15 (now 14) territorial health board areas. Each consortium had its own allocation of CHI numbers which meant that when patients registered in different consortia with GPs or as temporary residents, a new CHI number was generated resulting in some patients having more than one CHI number. In 1997 the Unique Patient Identifier (UPI) was introduced which allowed each patient to be identified by one unique CHI number which would be used throughout NHS Scotland. All other CHI records for a patient were marked as historical and linked to this one UPI. A UPI database/index was also created enabling all the CHI databases to be linked together and allowing the UPI number to be transferable between different CHI databases.

Authorised primary and secondary care users can perform patient search routines via direct access to CHI (e.g. PSDWebCHI; CHI 24) or interfaced Accredited Business Systems (ABS) (e.g. patient administration systems; screening systems), using the UPI number (or any CHI number). Alternatively, searches may be carried out using patient demographic details. Where a patient record is recorded in an ABS, a contact for each ABS is recorded in the CHI record. This allows patient updates to be sent to the appropriate ABS in daily broadcast files.

The records on CHI include those for patients: currently and previously registered with a GP; deceased; and temporary residents (never registered with a GP and only demographic details recorded). Only PSD can register patients with GPs. Records for non-GP registered patients may be created by other NHS Scotland services, e.g. screening system or health board/secondary care users, to enable a CHI record/number to be generated (where one does not already exist) for inclusion in their systems and for linking investigative procedures/results and episodes of care.

No single body has overall responsibility for CHI; the data controllers for CHI are the 14 NHS Boards. Decisions on access and use have traditionally been taken by the appropriate Directors of Public Health (DPH) in their role as CHI data custodians. Since 2005, the CHI Advisory Group (CHIAG) has provided guidance/recommendations to boards and developed policy for the use of and access to CHI data. Additionally it will approve access to CHI data on a national basis. Locally, decisions are taken by the appropriate DPH/Caldicott Guardian. National operational and data quality matters are the responsibility of the CHI PPI Steering Group.

## 4.3 Metadata

Each patient registered is allocated a unique CHI number (UPI) with which he/she can be accessed. Each person is uniquely identified on the CHI by a 10 digit CHI Number. The first six digits of this number consist of the person's date of birth (day, month number and year respectively). The last four digits are allocated by the CHI system when the person is first registered on it. The first three of these are

allocated by sex (even numbers for females, odd numbers for males). The last digit is simply a check digit, which enables the CHI Number to be validated (to some extent) away from the CHI system itself (e.g. for batch keying of amendments on key-to-disk equipment or for entry on other, 'stand alone', systems that use the CHI number).

As a person moves around Scotland the CHI number should move with them from Health Board to Health Board. The UPI Index allows users to access any CHI database seamlessly to locate a patient anywhere in Scotland no matter which CHI database they are recorded.

CHI records and numbers can be erased from the CHI database but is only carried out in rare and exceptional circumstances, e.g. record created in error. However, where a more current record already exists, an erroneously created record will be linked to the current record.

There are around 11.6 million records on the CHI database at the moment. The CHI database includes individuals who:

- Are currently registered with a GP;
- Have previously been registered with a GP, but:
  - have now transferred off as moved outwith Scotland, but elsewhere in the UK;
  - whereabouts are unknown;
  - have emigrated; or
  - are not currently registered with a GP, but are still active patients in the NHS Board area (i.e. have been compulsorily removed by GP, or elected not to be registered with a GP)
- Have died;
- Are with the Armed Forces or are Service Dependents;
- Who have attended A&E in Secondary Care;
- Have a screening history.

Where a patient is GP-registered, each person's CHI record comprises demographic data which includes; CHI number, date of birth, name(s), address, GP registration details, NHSiS contact data and an NHS number (where known). The record may also contain previous address, name(s) and previous/new NHS Board/Health Authority area code.

## 4.4 Inputs

The following data systems update the CHI system:

- NHS Central Register (NHSCR):
  - Migration from Scotland to rest of UK - the NHSCR are notified when an individual moves to another part of the UK. The trigger for this notification is when the patient registers with a GP in their new area - the driver being to transfer the patient's medical records. Currently these notifications are received by Practitioner Services Division (PSD) within same month/following month of the new GP registration.
  - Deaths – the NRS Vital Events team send information to the NHSCR registering deaths. NHSCR then update the Community Health Index (CHI) system. These are automatically marked as deceased on CHI

- Births – the NRS Vital Events team also send birth information to the NHSCR.

- Practitioner Services Division (PSD):
    - GPs send updates of addresses to PSD if patients register with a new GP or update their addresses/name. PSD updates the CHI with the new address. The feed between CHI and the GP System is via electronic link (PARTNERS) and feeds are sent/processed a number of times each day. Over 98% of GP registration transactions are processed by PSD within two working days.

## 4.5 Outputs

The following outputs are extracted from the CHI:

- Mid-Year Estimate (MYE):
    - As the CHI contains the postcode of the patient's address, this enables migration to be estimated for councils, and for smaller areas. The approach used for estimating council-level migration involves matching CHI patient records extracted from a database which reflects the 'live' CHI system[2] on two occasions one year apart.
    - Currently, migration data derived from the NHSCR is considered to be the most reliable data available at NHS Board level, so estimates from the CHI are controlled to ensure that they are consistent with the NHSCR data for moves across a NHS Board boundary by origin, destination, age and sex[3].
    - The version of CHI that is sent to NRS from Information Services Division (ISD) to assist in the calculation of the MYE is a snapshot of data from three months ago.
    - In addition to the use of the CHI in the estimation process of the MYE, it is also used by NRS for life expectancy and household projections. There are also many other statistical and non-statistical publications and reports which NRS population estimates feed as well as per capita calculations.

- NHS Central Register (NHSCR):
    - Information from the CHI is used to update the NHSCR system. The postcode on the NHSCR is updated from the CHI system. Patients transferring into Scotland from elsewhere in the UK, or from abroad and movements within Scotland are used to update the NHSCR system. The NHSCR is then used to update a number of systems outlined below.

**Footnotes**
2) This CHI extract is sent from ISD to NRS.
3) MYE formula has used CHI records since 2002. Previously it used information from the Electoral Register, which was not as complete as less people register to vote than register for health care.

## 4.6 Coverage

Examples of those groups not registered with an NHS GP include individuals who choose not to be registered (e.g. through personal choice or through provision of private healthcare users, armed forces personnel, and patients in long stay hospitals. Prisoners with sentences over six months should now be registered with the appropriate prison health centre on the CHI, although it is understood by PSD that some prison establishments still have to complete this exercise. It is likely that in the near future armed forces personnel and their dependents will be registered with the appropriate forces health centre on CHI. In many cases these patients may have a CHI record, but it will be currently de-registered, or may never have been GP registered.

Patients should only be registered with a GP where their stay in the local area is for more than three months. If the stay is for less than three months and they require treatment, the GP should see them as a temporary resident.

Temporary residents treated by GPs are not registered on the CHI. If they are subsequently referred to hospital or tests are sent to labs, a CHI record/number may be generated by the hospital where necessary and where they have the facility to do so as not all boards have CHI registration functionality yet.

Where a patient moves within the UK, the transfer off from their old GP is normally triggered by their registration with a new GP. If they fail to register promptly with a new GP, there can be a delay in the transfer off from their old GP. PSD do not receive any systematic feed from other health or public sector systems when patients leave the UK. When PSD do transfer patients off their GP it is usually from advice from the patient, or advice from the GP Practice (either voluntarily provided or from anti inflation work carried out by PSD).

It has been suggested that GP registration rates are higher among women due to the higher likelihood of ongoing healthcare needs, particularly around contraception, cervical smear screening, and pregnancy.

Similarly, it is likely that there is also a relationship between health and coverage in the CHI. Healthy individuals are less likely to register with a doctor as they do not immediately need the service. Consequently, as there is a strong association between health and age, young people might be less likely to register promptly with a GP.

## 4.7 Accuracy

The accuracy of addresses in the CHI is of great interest to the project as it relates to the geographical distribution of population estimates in the MYE. As mentioned above, often when individuals move internally and internationally they do not always de-register or register with a new GP immediately. This has an impact on the accuracy of the CHI and the data systems it feeds into.

It is acknowledged that the extent of any list inflation depends on the demographic profile of each area. Seven population sub-groups can be identified as being potential sources of list inflation:

- Students
- Pupils in residential schools
- Multi-occupancy dwellings
- Deceased patients
- Immigrants
- Patients who go abroad for more than three months
- Elderly patients over 100 years old

Due to medical screening there are some demographic groups with higher address accuracy in the CHI, as they are contacted more regularly by their GP and their address is more likely to be kept up-to-date. This mainly applies to the following groups:

- Children – due to health visitors, child surveillance and immunisation;
- Women – potentially due to regular visits for contraception and cervical screening; and
- Older men - due to bowel and prostate screening.

PSD also take advantage of various other communications between NHS Scotland and patients in order to maintain the accuracy of the system. Where correspondence is sent to patients and returned undelivered, this can be passed to PSD (e.g. screening invites, flu mailers etc). This is then raised with the GP practice in order to establish whether a new address is available, or whether the patient should be de-registered. Note: secondary care cannot amend CHI records; any demographic changes identified must be routed via PSD or the GP.

### 4.7.1 Data Cleaning

There are also a number of tools that the Practitioner Services Division (PSD) data quality team use or have previously used to improve the accuracy of the data.

- The PSD use Patient Information Comparison Test (PICT) – a reconciliation tool. This takes the GP system data and runs it against the CHI system (but this does not de-register patients who appear consistently on both systems).
- Student registrations are coded and after five years (after they should have graduated) they can be targeted for confirmation with the GP Practice/subject to letters to establish if still resident. However, it is a costly exercise and revised if covered in another source. The identification of student registrations can be difficult where no dedicated practice serves the university/college, and where addresses are not easily identifiable as student accommodation.
- Immigrant registrations are coded and subsequently subject to mail shot/GP confirmation to determine whether still resident.

### 4.7.2 Improvements to Accuracy

PSD continue to work to improve the data quality on CHI, (e.g. through the use of software such a QAS both in batch matching and transaction processing and through improved operating procedures to reduce the likelihood of duplicate

registrations and patient mis-associations. These help to ensure that data quality is not compromised during peak times of registration activity (e.g. during university intake) and when the data supplied from GP Practices may be less reliable.

### 4.7.3 Differences between NHS Greater Glasgow & Clyde (GGC) and the rest of Scotland

Rules about how to register births and deaths are the same across all three PSDs. However, in the west of Scotland there are more single handed (one doctor) GP practices than the rest of Scotland. They are more common in rural areas and urban deprived areas. The west of Scotland has more than the rest of Scotland. Lothian only has one single handed GP practice. Only around 10% of GP practices are single handed GP practices in the UK. Single handed GPs do not have the same level of administrative support as other GPs and could possibly not update their administrative systems as often or to the same standard as the rest of the UK.

Anti-inflation work regarding students may be easier to undertake in some areas than in others, e.g. where practices solely provide services to students and where the initial intake of students can be easily identified. Although Glasgow University has such a practice, there is a much smaller proportion of its students who use it.

Another explanation given was that council boundaries are close together and difficult to distinguish between in NHS GGC. NHS GGC caters for six whole councils (Glasgow, East Dunbartonshire, West Dunbartonshire, East Renfrewshire, Renfrewshire, and Inverclyde) as well as parts of North Lanarkshire and South Lanarkshire. Lothian provides for Edinburgh, East Lothian, Mid Lothian and West Lothian only, which are all distinguishable.

## 5.    NHS Central Register (NHSCR)

### 5.1  Overview

The NHSCR contains basic demographic details of everyone who was born or died in Scotland plus anyone else who is (or has been) on the list of a General Medical Practitioner (GP) in Scotland. The Register exists mainly to allow the smooth transfer of patients who move between Health Board areas (or across borders within the UK). The Register holds little information about people's health and only for the purpose of medical research, especially reducing the incidence of cancer.

The NHSCR operates under Section 57 of the Local Electoral Administration and Registration Services (Scotland) Act 2006 (information can be found on the www.legislation.gov.uk website)

### 5.2  Background

On the formation of the NHS in 1948, there was no central index of patients registered with a GP. This resulted in the rapid inflation of GP lists as people moving about the country appeared on more than one list. The NHSCR was set up as an index to control this inflation in the early 1950s. It has acted ever since as an index to NHS patients.

The principle purposes of the NHSCR are to:

- Manage and maintain the NHSCR database;
- Record change of life events, e.g. births, deaths, etc;
- Issue NHS Numbers;
- Instigate the movement of patients Medical Record Envelopes and record GP registrations across UK Borders;
- Provide Unique Citizen Reference Numbers (UCRNs) and other updates to the Improvement Service to maintain the Local Government Citizens Account;
- Manage a Medical Research flagging, linking, tracing and reporting system;
- Supply information to inform NRS Population and Migration Statistics; and
- Trace people resident in the UK who have lost contact with their families.

## 5.3 Inputs

There are over one million changes made to the NHSCR each year. The following data systems are used to update the system:

- Forward Electronic Register (FER) system – the NHSCR receive a nightly feed from the FER system of all Scottish birth and death registrations.
- Community Health Index (CHI) system – the NHSCR review a nightly feed of the CHI system of all demographic amendments, GP registrations and postcode update changes. The NHSCR is then used to update a number of systems outlined below.
- NHS data systems (rest of UK) – the NHSCR receive a nightly feed from England, Wales and Northern Ireland of cross border moves and the NHSCR is updated accordingly.
- Citizens Account – the NHSCR receive regular updates from local government covering property reference numbers and postcodes.

## 5.4 Outputs

The following outputs are extracted from the NHSCR.

- CHI system – a nightly feed is sent from the NHSCR to update the CHI system of all demographic updates, civilly registered deaths and GP registrations out with scotland.
- NHS data systems (rest of UK) – the NHSCR send a nightly feed to England, Wales and a weekly feed to Northern Ireland of cross border moves.
- Citizens Account – the Register ensures that the Improvement Service can update the Citizen's Account in the following ways:
  - o update citizen details;
  - o create, correct and delete UCRN;
  - o notification of death and notification of death correction.
- NRS monthly download – NRS receive a monthly download of the NHSCR for statistical purposes.

## 5.5 Coverage

The NHSCR is a list of individuals who have been born and died in Scotland and not registered with a GP, and those who have registered with a GP. Individuals intending to stay in the UK for longer than three months can register with an NHS

GP. This will mean that the register will include some individuals who would be excluded from the 2011 Census definition of usual residence (individuals who intended to stay for longer than six months).

It is not possible to distinguish between short-term migrants (those who intend to stay for less than six months) and usual residents on the register, nor is it possible to identify whether all short-term migrants were removed from the register when they departed.

### 5.6 Accuracy

The accuracy of the NHSCR depends on patients registering with an NHS GP when they move into or within Scotland. Accuracy of the register is known to vary across Scotland as individual Health Boards are responsible for its maintenance.

List inflation exists on the register where individuals are no longer resident at the address at which they are on the register. Estimated list inflation is around 8 per cent on average for Health Boards in Scotland but is known to vary in some areas. It is acknowledged that the extent of any list inflation depends on the demographic profile of each area. Seven population sub-groups can be identified as being potential sources of list inflation:

- Students
- Pupils in residential schools
- Multi-occupancy dwellings
- Deceased patients
- Immigrants
- Patients who go abroad for more than three months
- Elderly patients over 100 years old

## 6.    Citizen's Accounts

### 6.1 Overview

The Citizen's Account (CAS) (information available on the Improvement Service website) is run by Scottish Local Authorities and, at the moment, primarily supports the National Entitlement Card scheme. Customer First deal with the administration, monitoring and security of the system. The corrections are made to the CAS by the Improvement Service (IS). The Local Authorities are the data controllers for the system.

### 6.2 Background

The CAS provides a secure environment for citizens to register for and have access to services, which also helps councils keep accurate, up-to-date records of their customers.

Births, deaths and marriages and other changes in circumstance may not be picked up automatically by the disparate local government systems traditionally used to store information about individuals. The CAS changes all that by creating a single online record that customers can access securely and update themselves, using online authentication. This allows councils to maintain a definitive electronic record of all their citizens. It also allows citizens to personalise content and manage their own services online, without the need to contact the council for further information.

The purpose of the CAS is to:

- Provide an online account where customers can store personal details which they own and control;
- Support access to local and national services:
    - registration and enrolment facilities;
    - validation of person and address;
    - secure messaging to handle changes of circumstance; and
    - on-line Authentication mechanism.
- Support improved data quality through person and address validation and the exchange of change information.

## 6.3 Metadata

The following information is included on the CAS system:

- Names, date of birth and gender (UCRN);
- Postal address, Unique Property Reference Number (UPRN);
- Email address and mobile phone number (if provided);
- Links to any services that have been enrolled for; and
- A username, password and security phrase (if asked for access to on-line services).

The CAS system does not hold any transactional data, for example council tax payments, etc.

## 6.4 Inputs

The following data systems update the CAS system (Annex A):

- NHSCR: (updated by CHI and NRS data on births and deaths) acts as a registrar and validator for person data:
    - create, correct and delete UCRN
    - update citizen details
    - notification of death and notification of death correction
    - swaps
    - merges

- One Scotland Gazetteer: (feed from local authorities) acts as a registrar and validator for address data:
    - create, correct, update UPRN address
    - update street and locality identifiers
    - create new street

- My Diabetes My Way (MDMW): acts as a requester and informer of person and address data:
    - registers patients for Citizen's Accounts to support on-line access to MDMW

- National Entitlement Card: (also updated by SEEMIS) acts as a subscriber, requester and informer of person and address data.
  - registers citizens as part of the National Entitlement card application
  - exchanges change of name and address with CAS
  - receives Death Notifications from CAS

## 6.5 Outputs

Address and other changes can be shared with the NHS Central Register (NHSCR) in order to ensure that the data is current and valid.

Local Authorities can act as a subscriber and requester of person and address data, e.g. death notifications are provided by NHSCR and Citizen Account (CAS) makes them available for all Local Authorities to access. The main partner for CAS services is the National Entitlement Card scheme.

## 6.6 Coverage

CAS data and processes are available to support registration of 100% of the Scottish population.

Currently just over two million CAS registrations have been completed.

## 6.7 Accuracy

The figures provided are subject to revision due to a data quality project currently taking place. Figures are based on data collated in March 2013.

- Level of accuracy of CAS data:
  - of 1,985,573 CAS records, 1,958,064 have a UCRN match. This represents a match rate of 98.6% for person data.
  - of 1,985,573 CAS records, 1,937,034 have a UPRN match. This represents a match rate of 97.5% for address data.

Improvement Service (IS) are currently scoping and implementing processes to maintain the match rate and to try and address the remaining non-matched records.

## 7.    One Scotland Gazetteer

### 7.1 Overview

The One Scotland Gazetteer (www.onescotlandgazetteer.org.uk) is an address database made up of all 32 individual local authority gazetteers in Scotland. All addresses are created in accordance with the national standard for addressing and provide spatially referenced address records and property lifecycle details.

### 7.2 Background

All 32 Scottish Local Authorities are responsible for maintaining a Council Address Gazetteer (CAG); a list of geo-referenced property records for their area. The data is derived from the councils Street Naming and Numbering function with amendments, deletions and additions gathered through planning applications submitted to the council.

The One Scotland Gazetteer (OSG) is an amalgamation of these 32 CAGs. This provides a property index with a high level of accuracy and is the most up-to-date source of address information in Scotland.

The OSG falls under the remit of the Customer First Programme (further information is available on the Improvement Services website), led by the Improvement Service. This is a partnership between the Scottish Government, the Improvement Service and the 32 Local Authorities and aims to create a more efficient way of delivering services to the Scottish population, especially those entitled to certain services (e.g. National Entitlement Card).

## 7.3 Metadata

The One Scotland Gazetteer (OSG) may be accessed in two different formats, either through a web service or as an export on Compact Disc. NRS obtain a quarterly download of the OSG data through the web service.

As the OSG complies with the addressing conventions stated in British Standard 7666:2006, it contains the:

- Basic Land and Property Unit (BLPU) – grid references and quality indicator, status of property, Unique Property Reference Number (UPRN)
- Land and Property Indicator (LPI) – property name, number, postcode
- Street name, town, locality, Unique Street Reference Number (USRN)
- Organisation or trading name
- Classification of property

## 7.4 Inputs

Local Data Custodians from each authority provide frequent uploads to the OSG portal. The data is then validated using the Scottish Gazetteer Conventions.

The One Scotland Gazetteer portal has a function which allows users of the OSG data to report errors to the data custodians. These are investigated further and any necessary changes made to the CAG which feeds into the OSG.

## 7.5 Outputs

The OSG data is used to update a number of administrative data sources:

- Citizen Account
- NRS Scottish Address Register
- Ordnance Survey AddressBase products

## 7.6 Coverage

The One Scotland Gazetteer covers the whole of Scotland.

## 7.7 Accuracy

As the OSG is created using data derived from those directly responsible for maintaining property information, there is a high level of accuracy to the data.

## 8. National Records of Scotland (NRS) Scottish Address Register

### 8.1 Overview

The NRS Geography branch maintains a Scottish Address Register based on information received from the Royal Mail Postcode Address File (PAF). It is then matched to the OSG and unique record identifiers are added from both datasets (the Unique Delivery Point Reference Number (UDPRN) from PAF and the Unique Property Reference Number (UPRN) from OSG).

Although the Address Register was initially developed to support NRS Geography and Beyond 2011, it has become clear that the need for a robust address register is really an NRS wide requirement and other projects such as e-Registration, Census Variant Design and Vital Events will all rely on good addressing either now or in the future.

### 8.2 Background

A new version of the Address Register is created quarterly, in February, May, August and October of each year.

With the development of the new Ordnance Survey product AddressBase, Geography are investigating whether it could be used to create the Scottish Address Register in place of the current method of using PAF.

AddressBase is effectively the OSG matched to PAF and although similar to the matching work of Geography, AddressBase is primarily matched using an automated system. As Geography have used both automated and manual matching processes there is a slight difference in the matches obtained, although 94% of matches are in agreement.

Development work has started on creating a system using AddressBase as the input and the two systems will be run in parallel until Geography are satisfied that Addressbase meets all NRS requirements.

### 8.3 Metadata

Currently the address database is maintained in Access but will move to Sequel Server. Other formats of the address database can be provided.

Table 1 provides a list of the fields currently held on the Scottish Address Register. The majority are populated using formatted PAF data but to assist searching or matching of the database, the SearchSubBuilding, SearchBulding and SearchBuildingNumber are populated using raw PAF data.

## Table 1: NRS Address Register Variables

| Field name | Description |
|---|---|
| Postcode | The Royal Mail (RM) postcode (e.g. EH10 4SU) |
| Split indicator | Y or N to indicate whether the postcode crosses a council boundary, island/mainland, island/island |
| Postcode type | Large user or small user |
| Non residential | Tick if the address is deemed to be non-residential (based on Geography criteria) |
| Household count | Total number of households as recorded by RM (takes account of multi occupancy) |
| Organisation | Organisation name including Department name |
| Property | Property name (Building and Sub-Building) |
| Street | Street. (This provides the dependent street and street field concatenated together as appropriate to provide a complete street field. Any building number will also be with the street) |
| Locality | Locality. (This provides the double dependant locality and dependant locality fields concatenated together as appropriate. The building number may also be with this if there is no street) |
| Town | Town |
| UDPRN | Unique Delivery Point Reference Number from RM PAF |
| UPRN | Unique property reference number from One Scotland Gazetteer (OSG) |
| USRN | Unique street reference number from OSG |
| Grid reference easting | Grid reference easting position |
| Grid reference northing | Grid reference northing position |
| SearchSubBuilding | Sub Building name required for searching/matching purposes only |
| SearchBuilding | Building Name required for searching/matching purposes only |
| SearchBuildingNumber | Building Number required for searching/matching purposes only |
| Date last changed | Is populated with a date only if the addresses differs to that held in the previous quarter. |

## 8.4 Inputs

The following data systems are used to update the Address Register:

- Royal Mail and OSG:
  The register contains the address information from PAF, including Unique Delivery Point Reference Number (UDPRN), and the Unique Property Reference Number (UPRN) and Unique Street Reference Number (USRN) from OSG, where records from the two datasets have been matched.

  Geography also apply certain criteria to addresses to help identify non-residential properties. These are then marked as such on the address register.

- Ordnance Survey:
  The grid reference on the register is taken from Ordnance Survey (OS) AddressPoint product. Only grid references deemed high quality by OS are included. This means that the address record has been matched to PAF by OS, and the property has been surveyed by an OS surveyor. Around 3% of address records will not hold a grid reference. The grid reference enables address records to be geo-referenced at address level should that prove necessary.

- AddressBase:
  Geography have been investigating the use of AddressBase as the input for the Scottish Address Register. As the new system would still be using the PAF information as the base, there would be a shortfall in records compared to the existing system. This is because PAF data only appears on AddressBase if it has been matched to the OSG.

  It is anticipated that matching rates between the two datasets will increase over time until a point is reached where differences are largely down to timing differences in the creation of the two datasets.

  It is also likely that Geography will switch to Addressbase as the main input file supplemented by PAF and other information. This will greatly facilitate sharing the Scottish Address Register with others.

## 8.5 Outputs

The Address Register is provided for use by NRS staff every quarter. The register also supports the Scottish Government's Office of the Chief Statistician (OCS) Survey Methodology and Co-ordination activity. Geography has sent regular address extracts to facilitate OCS running three key Scottish surveys since 2011. The surveys are The Scottish Household Survey, The Scottish Crime and Justice Survey and The Scottish Health Survey all of which can be found on the Scottish Government website.

Geography matching data is being used in a joint project with Edinburgh and Glasgow Councils, www.geoplace.co.uk (a sub-set of OS) and the Improvement Service. This will provide more accurate results for harder to match addresses, such as flats, as it will involve checking addresses on the ground.

## 8.6 Coverage

The Scottish Address Register covers the whole of Scotland.

## 8.7 Accuracy

As the Address Register derives from PAF it has a high degree of accuracy.

Due to the fact that PAF only contains addresses which receive mail, there are a percentage of valid addresses which do not appear on the Address Register. With further development it is hoped such addresses can be incorporated into the register.

However, as the Address Register is benchmarked to the Census 2011 address list, work has been undertaken to evaluate the coverage of the address list based on information returned from the field. As the number of addresses returned, which did not appear on the census list, was minimal (less than 1%) and most have since become valid entries on PAF, it can be concluded that the PAF has a high percentage of coverage for Scottish addresses.

## 9. Glasgow City Council (GCC) Council Tax Data

### 9.1 Overview

Each individual council is charged with establishing liability for council tax due on domestic dwellings in their area and for the billing and recovery of resultant charges. In order to manage that process they use one of a number of off-the-shelf package software solutions. In GCC's case this is supplied by Capita Software Services. The same product is used in Aberdeen City, Stirling and East Lothian Councils. GCC's Information Technology (IT) infrastructure is managed by a joint-venture between the council and Serco called ACCESS. Although the system is owned and managed by ACCESS the data belongs to the council.

### 9.2 Background

The council tax system is live and is updated continuously. Council tax is a property based method of raising local taxation. Who is liable to pay council tax at a particular property is usually, but not necessarily, based on occupancy. Because billing links a liable party to a property, data is not at an individual level but is at account level which links a liable party to a property. There is no mechanism in the relevant legislation to allow the council to collect information about anyone other than the liable parties and no obligation for the liable party to provide details of other members of the household. As a result the account will have details of the lead liable person(s) only in some cases, with additional information on other liable individuals only recorded if provided voluntarily by the liable person or to support a discount or exemption application. As a result council tax data may not provide the number of residents in a property in every case.

The Council tax data also includes event data (recording the changes in accounts key to combining data) and transactional data (relating to balances and payment methods).

GGC Council Tax methodology is specific to the council (although the data structure will be the same as other Capita systems); this is important to keep in mind for Beyond 2011 work as some of the definitions and practices may be different between councils.

### 9.3 Metadata

Dwellings liable for council tax in Scotland are registered by local Assessors and information is available from the Scottish Assessors Association Portal webite.

The data is active and the system is updated by users. When an individual moves they can contact the council via letter, email, phone or they can update their details online. Landlords and previous owners often notify the council of the names of the new owners or residents.

The accounts are then linked or preferably 'consolidated'. Consolidating the accounts is preferable as it pools together the account history.

Disregarded persons (people who are exempt from paying council tax) are classified into a number of categories. They include:

- Carers
- Students (there are lots of different types of student discounts)
- Severely mentally impaired
- Apprentice
- Child benefit
- Diplomatic immunity
- Spouse or dependents of student
- In hospital
- Hostel (including homeless and women's shelters)
- Foreign language student
- Student nurse
- Detained
- School leaver
- Visiting forces
- Youth training

## 9.4 Validation

A number of data sets are used to validate the council tax system.

- Updates of death registrations are within a few days.
- Assessors address data and the PAF. The accuracy and consistency of addresses has improved greatly within the last few years. They provide feedback of incorrect address data to the other data sources.
- Experian data is used to check single occupancy every two years. Experian address data can, however, be quite inaccurate.
- Benefits data is linked from the Department of Work and Pensions (DWP). Although the benefit fraud checks are separate from the council tax system they can be used to improve the accuracy of the data.

## 9.5 Coverage

The coverage of the council tax system for individuals is limited. Properties with multiple occupancy have poor accuracy concerning the number of individuals within the household and their identities as this is not a purpose of the system. The council sends out a survey only when they do not know who lives at the address. They previously sent this to all households when the account was moved or closed (if moving outside of Glasgow City Council (GCC)) but this was not found to be cost-effective as they had a very small number of returns.

The coverage of properties is excellent in the council tax system. The local authority updates the system when a property is extended or altered such that planning permission is required. However, there may be some cases where properties are converted or merged without legal planning permission, causing inaccuracies to the data.

The census definition of a household and communal establishment may differ from the local authorities' definition of a dwelling in some cases. For example, the census will list one university student hall as a communal establishment, whereas the data may list each flat within the student hall as a separate dwelling. Council tax

data identifies where dwellings are part of a student communal establishment as the address will have a schedule code attached to it. This is when an organisation, such as Strathclyde or other private student accommodation such as Fresh, is trusted to provide accurate information and the balance is kept at zero.

## 9.6 Accuracy

The council tax system use a range of sources to validate the data. Validation of exemptions is commonly done by the individual and landlords to say who is moving in and out.

Experian credit information data is also used. This costs £1 a search and £10 for an investigation.

The Land Registry can also be used to find out about ownership of the property. Universities provide details of all students every two weeks to the council. Glasgow Caledonian University, University of Strathclyde, Glasgow School of Art, Cardonald College all provide the data. Staff then verify whether or not an individual is a student by checking the list. Glasgow and Edinburgh are unique in this, as other UK university students have to provide documentation in order to be given an exemption.

Single Occupancy is verified through a signed declaration. Every two years they run the Single Occupancy households through Experian to check if they have other residents. If Single Occupancy households are found to have other residents they are contacted and asked if they want to continue claiming Single Occupancy discount. If a date for another individual moving in to the property is not given then they are assigned a managerial date. Therefore the move in dates may not be entirely reliable.

There have been complications with multi-level buildings as a postcode has been found to change at the higher floors. This may happen, for example, when old properties are split into flats or temporary postcodes are assigned to new build areas before the new postcode is officially assigned. The accuracy of address data has improved greatly in the last few years.

Indicators for rented properties are not always reliable as some landlords or tenants might not notify that it is a rented property. GCC matched their data against the Private Property Register and the council found that they have additional records of private landlords. This could be due to private landlords not being aware of the Private Property Register or not being part of the scheme.

GCC has some problems with slightly different naming conventions of flats. This can cause problems when matching addresses and updating the system. When asked about duplicate property records it was noted that this was a possibility. However, due to validation processes and the fact that one household would be billed twice for their council tax payments it was believed that it would be very likely that the household would notify the council of this problem and that it would be quickly resolved.

**Annex A – Citizen's Account Functional Model**



CAS Future Logical Model August 2011