

Scotland's Census 2022

Form Reconciliation Overview

December 2022

Table of Contents

1. Introduction.....	3
2. Form Reconciliation steps	4
3. Conclusion.....	12

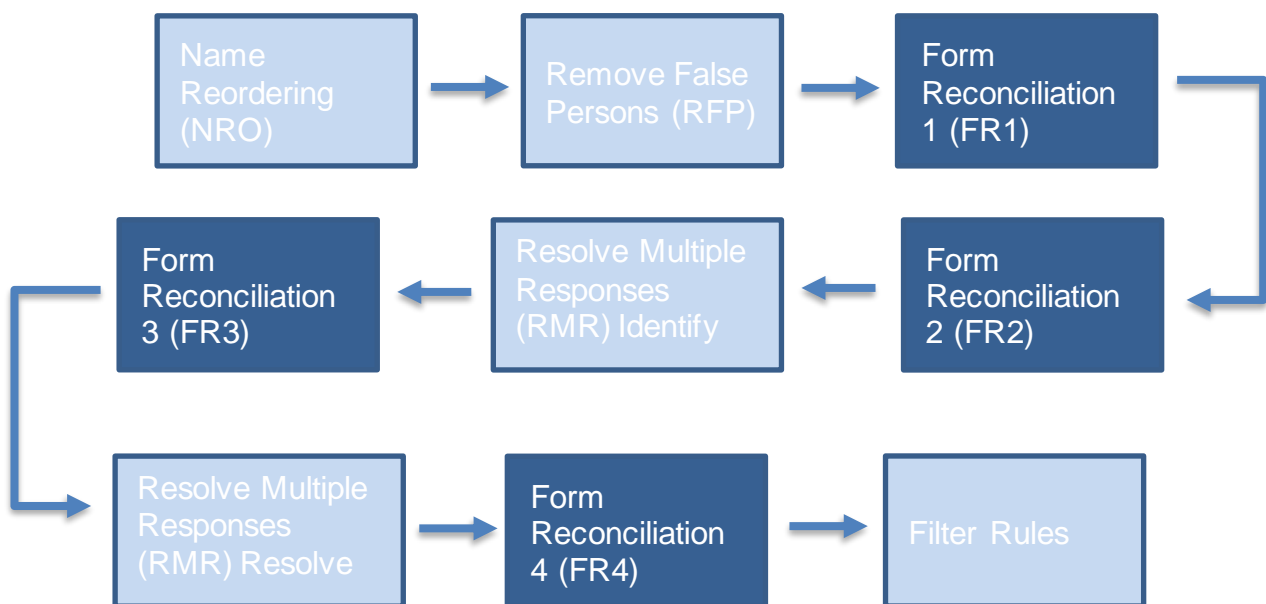
1. Introduction

1.1 What is Form Reconciliation?

Form reconciliation (FR) is a group of processes that work around/ in parallel to the main data cleansing processes.

We need to go from working with the data in the format it was collected (forms/questionnaires/returns), to working with it in the format it will be published (households and communal establishments (CEs)) – and we need to do this before the data cleansing step Resolve Multiple Responses (RMR)¹, as RMR looks to resolve multiple responses within a household or CE.

There are four FR processes. They occur at the following stages of data processing:



¹ More information on RMR can be found at [PMP014: Resolve multiple responses - identify duplicates | Scotland's Census \(scotlandscensus.gov.uk\)](#)

1.2 Creation of abode_id variable

To allow us to associate communal establishment (CE) individuals or individuals on a continuation form with their correct CE or household, we have to create a new variable in the dataset, called abode_id.

The abode_id variable is derived as part of form reconciliation and it becomes the standard household/ CE identifier for later processing steps.

For individuals, abode_id tells you which household or CE that person is associated with, and for households and CEs it serves to distinguish them. Creating a new ID allows us to retain all of the collected information and not disrupt it.

We define abode_id as being the same as return_id for all household and CE manager forms. For household residents not on a continuation form, abode_id = return_id also. For household residents on a continuation form, abode_id is the return_id of their parent household, and for CE residents it's the return_id of the associated CE manager.

2. Form Reconciliation steps

2.1 FR1 - CE manager and CE resident forms

This form reconciliation step deals with CE manager and CE resident forms and reconciling them together – this is where the term 'form reconciliation' comes from.

For CEs, this means having a single identifier that associates the correct individuals with the correct establishment.

There is no automated way to associate communal establishment residents (i.e. CI forms) with their establishment manager form (CE form). Their return identifiers are

totally distinct (and it's beneficial to preserve return identifiers since they link directly back to data collection – e.g. scanned images).

For CEs, each individual gets their own distinct sub-address (`sub_address_id`). This makes it tricky to match the CE residents to the CE based on address/ sub-address information, as each CE individual has a unique `sub_address_id`.

We need to either match on `address_id` (if this is available), which acts as a truncated `sub_address_id`, or truncate `sub_address_id` ourselves, so that everyone in the same establishment has an identifier in common, and the establishment also has that identifier in common.

This is where we create the `abode_id` variable. Each CE resident is given an `abode_id`, which is the `return_id` of the associated CE manager form.

We would match the CE individuals and CE managers on address/sub-address and then apply the new `abode_id`. For cases where there is more than one CE manager at that address/sub-address, then assign the CE individual to one CE manager at random. The CE managers will be resolved at the RMR stage, as RMR will combine these CE managers as duplicates. For any cases where there is a duplicate CE manager form, these would be reviewed at the RMR stage to ensure we are happy that they have been resolved appropriately.

If there is no CE manager at the address by the time data collection is over, consider the CE individual an Orphan form and deal with as described in section 2.5.

2.2 FR2 – Continuation forms

This step is similar to FR1 but it deals with continuation forms and reconciling them with their parent household.

It involves updating the continuation person records so that they are associated with their parent household.

Similar to FR1, there is no inbuilt way to automatically associate continuation form residents and their parent household form, as their return identifiers are totally distinct. The only information to match these up is address and sub-address.

This sub-address information is ideal for matching continuation forms and parent households, as it is common between the continuation people and parent household.

Similar to FR1, this is where we create the `abode_id` variable. We match the continuation people and parent households on address/sub-address and then apply the new `abode_id`. Each continuation person is then given an `abode_id`, which is the `return_id` of their parent household.

For cases where there is more than one parent household, we can manually inspect the forms and chose the most appropriate parent household to associate the continuation form with. If the two parent households forms overlap in terms of people then they'll be combined into one household by RMR; if they don't overlap in terms of people then they won't be resolved into one household, and we may have no grounds to pick one over the other to associate the continuation form with. The form would then be associated at random.

If there is no parent household at the subaddress by the time data collection is over, then consider the continuation form an Orphan form and deal with as described in section 2.5.

2.3 FR3 – Individual forms

Similar to FR 1 and 2, we need to ensure that there are no household individual forms that are unallocated to a household. We need to either assign them to a household or discard them.

RMR will resolve many individual forms, by resolving the individual with the minimal information back to the household form. More information on the RMR methodology can be found at [PMP014: Resolve multiple responses - identify duplicates | Scotland's Census \(scotlandscensus.gov.uk\)](#)

We associate Household Individual forms to a household by matching the person on the Individual form with the associated person on the Household form in RMR. If more than one match is found then the whole group are resolved together into one record.

During RMR, all people within a postcode are compared to identify matches. There are a number of ways a Household Individual may not match to a Household Person, with different outcomes. If a Household Individual doesn't match any household person then either:

- there is no corresponding household
- there is a corresponding household but no corresponding household person
- they failed to match to the corresponding person due to lack of evidence for a match (i.e. very incomplete data in matching variables)
- they failed to match to the corresponding person due to evidence against a match (i.e. contradictory information in matching variables).

There are another couple of considerations here:

- Household persons may indicate that they are 'decoupled', i.e. have requested an individual form that the householder knows about. These household persons will provide minimal information (name and relationship within the household) and then be routed to the end of the questionnaire.
- Individuals who provide a Paper individual response can indicate which Household person number they correspond with (i.e. "I am person 3 in the household") so that they can be associated with their household information. The response to this question may be missing or incorrect.

FR3 deals with any individual forms that RMR can't find a match for. There are four outcomes when doing this.

Outcome 1 – match individual to household resident:

Use the Household Individual return to overwrite one of the Household Persons, as determined by clerical review. In general the approach is as follows –

1. If the individual has provided a person number, could they be matched to a person with that person number at that address
2. Otherwise, could they be matched to a person at that address who has indicated they'll be providing an individual response
3. Otherwise, is there anyone at that address who the individual could match with

If we can match the individual to a household resident at that address (and we can explain why RMR identify missed this) then we will match them and resolve in RMR resolve by adding as a cluster.

Examples of why RMR might miss a match are:

- in a 'decoupled' case where the Household Person record is of poor quality (e.g. the householder recorded an inaccurate name which cannot be matched to the Household Individual)
- the Household Individual provides information which is different from the Household Person return but has accurately indicated which Household Person they correspond with (e.g. a respondent who indicated that they have a trans status may provide a different sex and forename on a Household Individual response but the same surname, date of birth and other information).

For outcome 1, 'overwriting' is done according to RMR rules, i.e. the Household Individual response gets priority, and any missing or invalid mandatory questions are filled using responses from the non-priority return.

This outcome requires time for clerical review, and is potentially open to bias, since judging here whether two responses represent the same person is more subjective than within the RMR process. In RMR, where we consider a potential duplicate and records are not identical, we're largely concerned with accidental differences due to capture or respondent error; in form reconciliation these differences may be intentional.

Outcome 2 – add household individual to household as distinct person:

Add the Household Individual return to the household as a distinct person. This option is preferred where we think that the Household Individual does not correspond with any of the existing household persons.

For example, if the householder requested an individual rather than a continuation form to capture their 6-person household (and so the Household Individual indicates a person number which isn't already in use), or wanted to add an additional person to their household after submitting the household return and did this through a Household Individual return.

If we can say the individual doesn't match any household resident at that address (e.g. non-missing, correctly captured dates of birth which are different), then retain the individual as an additional household resident.

Again this option requires clerical review to decide how to proceed, and in this case we add an additional person to the household without knowing their household relationships, which must be imputed during the Edit and Imputation (E&I) process². Clerical review is also important as it minimises the chance of counting the person twice and leading to overcount. We expect there to be very few of these cases so there should be minimal impact on overcount.

² More information on edit and Imputation methodology can be found at [PMP012: Overview of edit and imputation for Scotland's Census 2022 | Scotland's Census \(scotlandscensus.gov.uk\)](#)

Outcome 3 – discard the household individual:

If we can do neither outcome 1 or 2 (e.g. because there is too much missing information) then we must discard the individual return – we cannot correctly match it, so cannot fulfil the respondent's intentions, and to retain the individual as a distinct person would guarantee overcount in a way which our methods cannot fix (since overcount correction³ relies on consistent name and date of birth for matching).

Outcome 4 – treat as Orphan form:

Finally, if we have a Household Individual return at a subaddress (household address) where there is no household, then we treat as an Orphan form and deal with as described in section 2.5.

2.4 FR4 – Person Numbering

The earlier cleansing processes may have affected person numbering within a household or CE, so this FR4 step fixes person numbering and the relationship matrix within a household or CE. This is done after RMR, before going into filter rules and E&I.

E&I in particular relies on individuals within a household or CE being numbered sensibly (i.e. if there are N people, they're numbered 1, 2, 3...N) – any gaps (e.g. those created by Remove False Persons⁴) will cause problems for E&I.

Similarly, although person numbers will be unique to each return, they won't be unique to a household (i.e. an abode) – if we've combined two five-person households and the people don't overlap then we'll have two person 1, two person 2 etc.

³ More information on the overcount correction methodology can be found at [PMP015: Census to census linking and overcount correction | Scotland's Census \(scotlandscensus.gov.uk\)](#)

⁴ More information on the Remove False Persons methodology can be found at [PMP011: Remove false persons - methodology | Scotland's Census \(scotlandscensus.gov.uk\)](#)

The FR4 process involves arranging the associated people within a household or CE in a reasonable order, then renumbering them within the household or CE. We then move the values from the relationship matrix onto the person table in the correct order.

Note here that we won't have complete relationship information going into E&I – for example, FR4 won't try to backfill relationships between two households we've combined in RMR. We will know, though, which relationships should be imputed or left as not required based on the number of people in the household.

2.5 Orphan Forms

It is a requirement for subsequent processes and eventual outputs that each person is associated with either a communal establishment (i.e. a CE manager return) or a household (i.e. a household return).

As set out above, there are ways we can end up with unassigned CE, continuation or household individual people and no household or CE to correctly assign them to – these are called orphan forms.

In these cases there must be at least an address that the return is associated with, and that address must have been considered either a CE (in the case of CE individuals) or a household (in the case of continuation or household individuals).

Accordingly we should create CE and household records at each necessary address (CE) and subaddress (household) and associate these orphan forms with the created CEs and households.

There are two points to note:

1. It is not expected that we will see many of these cases for CE, since they have their own specific enumeration arrangements, and the associated

address must have been considered a CE for enumeration. These cases should all be reviewed manually. Here we can also populate any CE characteristics (e.g. establishment type) from the CE register, since we won't impute CE characteristics during E&I, and geographic information for the CE should be added at this stage.

2. For households, we do not necessarily need to review each case and we do not need to provide household characteristics (e.g. accommodation type, tenure) since these will be imputed during E&I. Geographic information for the created household should be added at this stage.

3. Conclusion

This paper outlines the need for, and process of carrying out, four form reconciliation steps in the census data processing journey. These steps allow for household individuals and CE residents to be associated with their parent household or CE, and to correct the person numbering within a household or CE that may have been affected by earlier cleansing processes. This is required for the later data processing steps, in particular Edit and Imputation.