**Constructing households (CHH) code flow**

Run order:

1. setup\_views\_and\_tables\_individual\_level.sql
2. setup\_views\_and\_tables\_household\_level.sql
3. data\_prep\_notifications.R
4. data\_prep\_validation.R
5. individual\_address\_analysis.R
6. group\_address\_analysis.R
7. validation\_suite.R
8. final\_table\_to\_snz\_format.sql

The supporting files:

* utility\_functions.R
* individual\_address\_analysis\_functions.R
* group\_address\_analysis\_functions.R
* validation\_suite\_functions.R

Are called during the main run order.

The code prefix for all IDI\_Sandpit tables and all IDI\_UserCode views is “chh\_”

In the diagrams below:

Boxes (with four straight sides) for process/action

Documents (with one curved side) for tables in the database

**Data flow**

## Setup

Setup views and tables individual level

Notification view—one view per source dataset

Gathered data

Data prep notifications

Setup views and tables household level

Notification view—one view per validation dataset

Household validation

Data prep validation

## Individual refinement

Gathered data (phase 1)

Keep only address changes

Remove remaining simultaneous notifications (phase 6)

Remove unsupported notifications in gaps (phase 5)

Remove low quality notifications (phase 4)

Remove unsupported simultaneous notifications (phase 3)

Remove notifications outside window (phase 2)

Replaced notifications

Individual refined

Address changes (phase 7)

## Household refinement

Address changes (phase 8—household start)

Ensure consistent handovers

Sync children with parents (phase 10)

Sync within household members (phase 9)

Replaced household notifications

Household refined (phase 11)

## Validation suite

Compare

Gathered data

Household validation

Reconstruct notifications present at each phase

Individual matched (for performance individual)

Replaced household notifications

Household refined

Address changes

Replaced notifications

Report accuracy

Performance household