# Master and Golden History Duplicate Fix

## Problem:

On Geldmaat, they received a full data load (not delta) from their suppliers on a daily basis.

This causes new records to be created in the Golden History and Master History tables without there being any changes in the actual data.

## Solution:

The solution requires 3 additional components in Semarchy:

* A lookup enricher (**DetectDuplicates**) to execute some SQL which finds whether the exact same duplicate record exists.  This effectively concatenates all the fields in the database and concatenates all the fields from Semarchy and if they match, then it returns so, otherwise it doesn’t change the record
* A new field to hold the output of the enricher (**Duplicate**).  Effectively this will hold ‘t’ if the enricher found a duplicate and be null otherwise
* A semQL validation (**PreventDuplicates**) that checks if the Duplicate field is ‘t’.  If it is not, then the validation passes (there is some up-side-down logic here because the validations must return true for a positive result)

### The way this works is:

1. When a source record is loaded it undergoes enrichment first
2. The DetectDuplicates compares the incoming, concatenated data (delimited with underscores to prevent multiple nulls from skewing things) with the existing data in the MD table – that is, it compares the incoming record against the current version of itself
3. If it is the exact same, then it sets Duplicate=’t’
4. After the enrichment is complete, the validation takes place
5. If Duplicate=’t’ then the validation fails and the source record never makes it to the MD table and therefore not the GD table.  You still get a record of the insert in the source error table (SE) but not the golden history

### Some limitations of this:

* It is time consuming to set up, and a pain in the neck to maintain, as each individual field needs to be added here manually
* While the **idea** will work for authored data in Semarchy, this is specifically set up to only work on source data.  Therefore it will only work on data coming via GIP not that which is authored in Semarchy
* I noticed when testing that the golden data override prevents some changes coming from source data, if it has been overridden within the UI.  This isn’t a limitation, it was just an observation, as I couldn’t see my changes during testing reflecting on the GD table because of this.

## Video: