# HRE – SUBSETS – OVERVIEW

**Revision history**

| | | |
|---|---|---|
| 2018-06-03 | Robin Lamacraft | Original draft |

## SCOPE

A subset is a persistent list of IDs (PIDs) which can be used to limit the scope of an HRE action to those entries. By default there are full Subset list every PID for a unique object type and object sub-type combination –usually referred to as "ALL <sub-type name>" like "ALL PERSONS", etc. There are several variants of Subsets in HRE which can be used for various research techniques. The number of Subsets is not limited, nor is size of a subset explicitly limited.

Subsets can be created by:

1. Applying a filter to some list of objects and selecting the ones that satisfy a condition
2. The user manually adding or deleting from a subset list.

The current subset types are:

- **SINGLE** – a list of PIDs for a nominated object type and object sub-type combination. These are the most common types of subset
- **PAIR** – a list of pairs of PIDs for a nominated object type and object sub-type combination. These are lists that can be used to notify that (say) 2 Persons with similar properties are considered different persons (when checking for potential duplicates)
- **MIXED** – a list of pairs of mixed PIDs. The first column of PIDs is for a nominated object type and object sub-type combination. The second column of PIDs is for any object type and object sub-type combination. These are lists that can be used to identify all links between a Source to any other object-type object sub-type combination (e.g. where and how was Source used)
- **BIO_NETWORK** – a list of pairs of Biological object PIDs of a nominated sub-type (say Persons) that includes other columns that categorise the relation between them.

## SUBSET DEFINITIONS

HRE table **351 SUBSET_DEFNS** has one record per defined subset. Table **352 SUBSET_TYPE_DEFNS** is a lookup list for the type of subset. Then a collection of records are created in the following tables depending on the type of subset:

1. **SINGLE:** table **353 SUBSET_SINGLE_DEFNS** has one record for each PID in the subset
2. **PAIR:** table **354 SUBSET_PAIR_DEFNS** has one record for each PID pair in the subset
3. **MIXED:** table **355 SUBSET_MIX_DEFNS** has one record for each PID pair in the subset. This records the object type and object sub-type of each PID
4. **BIO_NETWORK:** table **363 SUBSET_BIO_NETWORK_DEFNS** has one record for each relationship between 2 Biological objects. It has a number of other fields to classify the relationship. It is used to assist in the presentation of the degree of relationships

- The user can create new entries in the correct sequence in the Labelled Name Element Type dictionary
- Once this Name Style Data Entry list has been defined, the display of appropriate records can be re-ordered to match the entry window's preferred element entry order
- HRE will provide an initial set of labelled Name Element Types for each use of a Name Style for an object type and object sub-type combination
- Each labelled Name Element Type has a data type and possibly a formatting template.