(This example should be read in conjunction with an understanding of the MEP IDE document)

The diagram below shows the physical design of our Input Data Warehouse (IDE), included in the diagram is a mapping to the conceptual model created by Dan Gillman[[1]](#footnote-1)

SBR Concept Library

(Instance variable)

Instrument Map

Fact

(Represented variable)

Instrument

Collection Instance Instrument

Change History

Unit of Interest

“What”

“When”

“Who”

Includes “How”

Collection Instance

“Why”

Collection

Variable Library

(Dataset)

(Universe)

(Datum + Data Point)

(Observation)

The primary purpose of this implementation is to store response[[2]](#footnote-2) information for the purpose of supporting purpose statistical processing, such as editing, data derivation and imputation.

Example: Simple information collected about a person, such as:

* Name
* Age
* Sex
* Marital Status
* Ethnicity (major)
* Ethnicity (minor)

The follow tables set up the variables, both instance and conceptual (represented) that are being “collected”.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Library** |   |  | **Sbr\_concept\_library** |
| v\_key | var\_name | sbr\_code |  | sbr\_code | sbr\_concept\_name |
| 101 | Name | 50 |  | 50 | person name |
| 102 | Age | 51 |  | 51 | person age |
| 103 | Sex | 52 |  | 52 | gender |
| 104 | Marital Status | 53 |  | 53 | person marital status |
| 105 | Major Ethnicity | 54 |  | 54 | ethnicity |
| 106 | Minor Ethnicity | 54 |  |   |   |

The instrument table provides that grouping structure which brings the instance variables together to form a dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **instrument** |   |  | **instrument\_map** |   |   |
| i\_key | name\_text | instrument\_code |  | im\_key | i\_key | v\_key |
| 10 | My example survey set | mes\_2015\_set |  | 1001 | 10 | 101 |
|  |  |  |  | 1002 | 10 | 102 |
|  |  |  |  | 1003 | 10 | 103 |
|  |  |  |  | 1004 | 10 | 104 |
|  |  |  |  | 1005 | 10 | 105 |
|  |  |  |  | 1006 | 10 | 106 |

The following three tables form the structure metadata, bringing together the variables that are being collected for a period.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Collection** |   |  |  | **collection\_instance\_instrument** |
| c\_key | name\_text | frequency\_type\_key |  |  | cii\_key | ci\_key | i\_key |
| 7 | My example survey | annual |  |  | 70 | 501 | 10 |
|  |  |  |  |  |  |  |  |
| **Collection Instance** |   |   |  |  |  |
| ci\_key | c\_key | collection\_instance\_code | reference\_period start\_date | reference\_period start\_date |  |  |  |
| 501 | 7 | mes\_2015 | 1/01/2015 | 31/12/2015 |  |  |  |

The unit of interest table provides the population (sample) that is being collected.

|  |
| --- |
| **unit\_of\_interest** |
| uoi\_key | stat\_unit\_id |
| 2501 | person\_01 |
| 2502 | person\_02 |
| 2503 | person\_03 |
| 2504 | person\_04 |

As data is introduced to the fact structures supported by the IDE, a change history row is created to record (group) the set of the changes that were made.

|  |
| --- |
| **change\_history** |
| change\_history\_key | create\_date |
| 20001 | 16/01/2015 |
| 20002 | 19/01/2015 |

Each row in the fact table forms a cell (datum). The data point, fact\_value, is held in this table also – although this is an implementation choice.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **fact** |   |   |   |   |  |  |  |  |
| cii\_key | im\_key | uoi\_key | change\_history\_key | fact\_value |  |  |  |  |
| 70 | 1001 | 2501 | 20001 | Fred |  |  |  |  |
| 70 | 1002 | 2501 | 20001 | 50 |  |  |  |  |
| 70 | 1003 | 2501 | 20001 | Male |  |  |  |  |
| 70 | 1004 | 2501 | 20001 | Married |  |  |  |  |
| 70 | 1005 | 2501 | 20001 | European |  |  |  |  |
| 70 | 1006 | 2501 | 20001 | Maori |  |  |  |  |
| 70 | 1001 | 2502 | 20001 | Jane |  |  |  |  |
| 70 | 1002 | 2502 | 20001 | 25 |  |  |  |  |
| 70 | 1003 | 2502 | 20001 | Female |  |  |  |  |
| 70 | 1004 | 2502 | 20001 | Single |  |  |  |  |
| 70 | 1005 | 2502 | 20001 | European |  |  |  |  |
| **70** | **1001** | **2503** | **20001** | **Bob** |  |  |  |  |
| **70** | **1002** | **2503** | **20001** | **41** | These six rows form the response for a given unit. Spares responses are allowed, as shown for Unit 2505. |
| **70** | **1003** | **2503** | **20001** | **Male** |
| **70** | **1004** | **2503** | **20001** | **Divorced** |
| **70** | **1005** | **2503** | **20001** | **European** |
| **70** | **1006** | **2503** | **20001** | **Spanish** |
| 70 | 1001 | 5204 | 20001 | May |  |  |  |
| 70 | 1002 | 5204 | 20001 | 32 |  |  |  |  |
| 70 | 1003 | 5204 | 20001 | Female |  |  |  |  |
| 70 | 1004 | 5204 | 20001 | Married |  |  |  |  |
| 70 | 1005 | 5204 | 20001 | Asian |  |  |  |  |
| 70 | 1006 | 5204 | 20001 | European |  |  |  |  |

The fact structure(s) record all changes to data, deletion is prohibited[[3]](#footnote-3). Therefore, the most recent “version” of a fact is found through the use of the latest change\_history record. Ie

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **fact** |   |   |   |   |
| cii\_key | im\_key | uoi\_key | change\_history\_key | fact\_value |
| 70 | 1005 | 5205 | 20002 | Pacific Islander |
| 70 | 1006 | 5205 | 20002 | European |

Resulting in:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **fact** |   |   |   |   |  |  |
| cii\_key | im\_key | uoi\_key | change\_history\_key | fact\_value |  |  |
| 70 | 1001 | 2501 | 20001 | Fred |  |  |
| 70 | 1002 | 2501 | 20001 | 50 |  |  |
| 70 | 1003 | 2501 | 20001 | Male |  |  |
| 70 | 1004 | 2501 | 20001 | Married |  |  |
| 70 | 1005 | 2501 | 20001 | European |  |  |
| 70 | 1006 | 2501 | 20001 | Maori |  |  |
| 70 | 1001 | 2502 | 20001 | Jane |  |  |
| 70 | 1002 | 2502 | 20001 | 25 |  |  |
| 70 | 1003 | 2502 | 20001 | Female |  |  |
| 70 | 1004 | 2502 | 20001 | Single |  |  |
| **70** | **1005** | **2502** | **20002** | **Pacific Islander** | <-- | Logical update |
| **70** | **1006** | **2502** | **20002** | **European** | <-- | Logical Insert |
| 70 | 1001 | 2503 | 20001 | Bob |  |  |
| 70 | 1002 | 2503 | 20001 | 41 |  |  |
| 70 | 1003 | 2503 | 20001 | Male |  |  |
| 70 | 1004 | 2503 | 20001 | Divorced |  |  |
| 70 | 1005 | 2503 | 20001 | European |  |  |
| 70 | 1006 | 2503 | 20001 | Spanish |  |  |
| 70 | 1001 | 5204 | 20001 | May |  |  |
| 70 | 1002 | 5204 | 20001 | 32 |  |  |
| 70 | 1003 | 5204 | 20001 | Female |  |  |
| 70 | 1004 | 5204 | 20001 | Married |  |  |
| 70 | 1005 | 5204 | 20001 | Asian |  |  |
| 70 | 1006 | 5204 | 20001 | European |  |  |

1. The relationships has been altered such that they best represent the concepts under discussion. The result is alignment with the second “view” discussed at the Data Description meeting held on the 16/1/2015, where Instance Variable is related to Observation. [↑](#footnote-ref-1)
2. A “Response” is any data collected by us. It could be from any survey method or administrative data. [↑](#footnote-ref-2)
3. Logical deletes are enabled through the use of the fact\_life\_cycle\_key column, which exists on the fact table. This column has been left out for readability. [↑](#footnote-ref-3)