Draft 2019-09-30

### Key-Value Stores

A key-value store represents data as a set of pairs, a key and its associated value. The DDI4 model is shown in figure 12. A key is a unique value that allows look-up of its linked value. The DDI4 model includes a KVStore which contains KVPairs. Each KVPair links a Key to a Datum. The KVStore also contains a set of ConceptualKeys. ConceptualKeys and associated RepresentationRules define how a key is generated from its components. Looking at the data in figure 6, the value “3.3.1992” could be associated with a key “Marie-Born” generated from the unit identifier and the variable name. In this example the Representation rule is simple hyphen separated concatenation, but a different algorithm might produce a more compact unique key.

Figure 12 Key-value representation of the Long-skinny data

|  |  |
| --- | --- |
| **Key** | **Value** |
| Marie-Sex | Female |
| Marie-Born | 3.3.1932 |
| Marie-Died | 12.1.2005 |
| Marie-RefArea | Newport |
| Marie-Longevity | 73.7 |
| Henry-Born | 8.1.1929 |
| Henry-Died | 6.2.2008 |
| Etc. |   |

## Key-value stores

### Key-Value Stores and DataPoints’

The DDI4 model for a basic key-value store of DataPoints is shown below.



Figure 15 The DDI4 Key-value model for DataPoints

At its heart the Key value model is simple. A Key identifies a value. Both the key and the value can be stored in a DataPoint. A Key is also associated with some Unit. Those DataPoints are stored in a KeyValueDataStore.

The Key itself is a single value, but it is produced from a number of components. One set of components serves as a primary identifier. A separate set is secondary. The values making up the primary key would identify the observed unit in the case of unit level data or a population in the case of an aggregate cube. The primary key would identify a structural cell in a dimensional model.

Secondary keys describe the meaning of the value. They would include Measure components (the value represents height in centimeters) or attribute components (measured in stocking feet etc.).

Secondary Keys might also include a time annotation. (This is the January revision of the per-capita income value).

The model includes specific categories of the Primary and Secondary key members.

IdentifierComponents are values that uniquely identify a Unit. Their uniqueness depends on some context. The ContextualCompnent describes that context. An organization, for example, might have its own schema for identifiers. That schema would serve as the ContextualComponent.

A unique identifier might also be generated programmatically via some protocol like a Universally Unique Identifier (UUID). This would be a SyntheticComponent.

In a Dimensional model the Dimensional Component would be a set of values, one from each of the dimensions. That set would identify one DataPoint in the dimensional structure.

For the SecondaryKey, the TimeAnnotation would be based on a TimeComponent. The TimeComponent might be made up of a set of descriptions of specific times, or time periods. The TimeComponent might also be given a type description (like “Valid”).