Aristotle User Guide: How to record a controlled vocabulary

Key Terms: **Controlled Vocabulary, Classification, Reference Data**

User story: **I have a series of terms used to classify data that I want to record in my metadata registry, so I can inform people of what data I have and what it contains.**

My controlled vocabulary uses codes and categories to classify data. These codes should be able to be used for validating data, so I need to be able to access codes from an API. I need to be able to map between different codes when they record the same meaning.

If the codes in my vocabulary have changed over time, I need to record when different codes were valid for use in data. If my vocabulary is endorsed, I want to record who and when this controlled vocabulary was approved for use.

What metadata objects should I use?

Classifications, Conceptual Domains and Value Domains

Controlled Vocabularies are recorded in the Aristotle Metadata Registry using the **Classification**, **Conceptual Domain** and **Value Domain** metadata types. These metadata types can be linked to other metadata within the registry to make others aware of how data in the registry is coded or classified.

- A **Value Domain** is description of the possible values data can take and can include an ordered collection of codes with descriptive meanings.
- A **Conceptual Domain** is a description of how data is categorised and can include an ordered collection of named categories that describes how data is categorised.
- A **Classification** is an ordered collection of codes and categories for classifying data based on a particular topic. Classifications can be hierarchical to store complex categorisations of data.

One way to understand why controlled vocabularies and classifications are useful is to consider how why people classify things, such as clothing or books in a department store:

When stocking products in a store, department stores will group similar items together to help shoppers quickly find the products they want – for example, products could be grouped as cookware, toys, or clothes.

A child's cooking apron could be grouped in either one of these categories within a store, however shoppers will expect to be able to find the same product in the same place no matter which store they visit. So the company will set a strict classification for products so that users can easily find the products they want to buy.

How products are classified may also change based on how the clothing is processed. When on display, goods will be grouped by shoppers needs, when in a warehouse products may be grouped by manufacturer, and when products are shipped they may be grouped by weight.
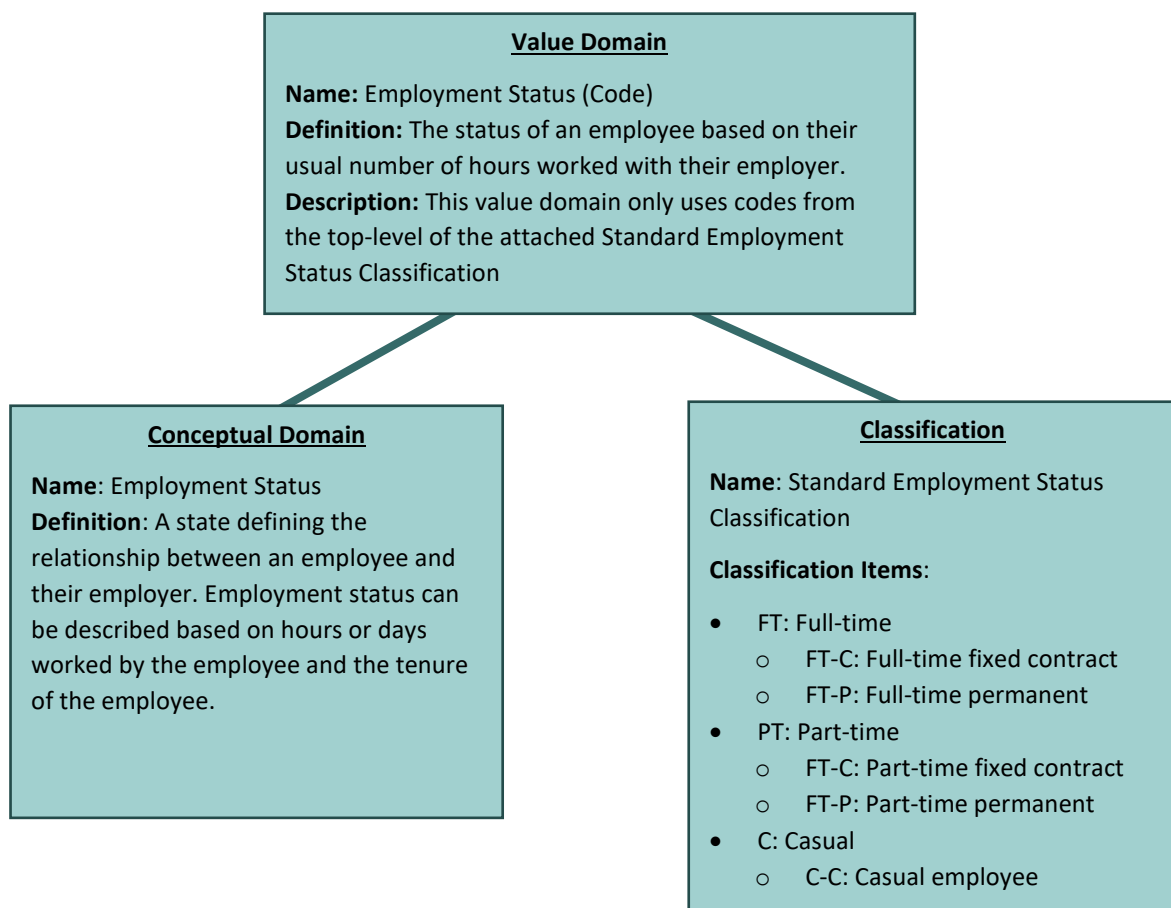
Having these classifications recorded and centrally managed means employees can easily find the information to do their job.

How Value Domains, Conceptual Domains and Classifications relate

Within the Aristotle Metadata Registry, a **Classification** is way of defining how your data is grouped together, including the standard codes and categories used to describe data. A Classification may record groupings for product data, employment data or health data and can be used to describe the appropriate use of these categories.

A **Value Domain** describes which values, codes and categories are used when recording data. When data is recorded using a fixed list of codes, these are recorded in a **Value Domain** as **Permissible Values**. When these values are reused at different levels in many different **Value Domains**, they can be stored as items in a Classification

Different business areas may use different codes when recording their data, due to different data practises or recording data in different language. By storing the basic definition of codes used in data in a **Value Domain**, users of the data will be able to continue to understand it.

A **Conceptual Domain** records the basic ideas used when categorising data for use. A **Conceptual Domain** can be purely descriptive, or it can also include the different lists of categories used when breaking down or aggregating data.

Value Domain

Name: Employment Status (Code)
Definition: The status of an employee based on their usual number of hours worked with their employer.
Description: This value domain only uses codes from the top-level of the attached Standard Employment Status Classification

Conceptual Domain

Name: Employment Status
Definition: A state defining the relationship between an employee and their employer. Employment status can be described based on hours or days worked by the employee and the tenure of the employee.

Classification

Name: Standard Employment Status Classification

Classification Items:

- FT: Full-time
  - FT-C: Full-time fixed contract
  - FT-P: Full-time permanent
- PT: Part-time
  - FT-C: Part-time fixed contract
  - FT-P: Part-time permanent
- C: Casual
  - C-C: Casual employee

Examples of Controlled Vocabularies and Reference Data in Aristotle

Below are some examples of how controlled vocabularies and reference data can be registered in the Aristotle Metadata Registry. Follow any of the links in the descriptions to see a complete example on the Aristotle.Cloud registry.

Storing simple code lists using Value Domains

When data is recorded using coded values it is important that codes are recorded in a registry so that values can be interpreted during later analysis. In this example, numbers are used to record a persons' employment status in a company – with 1 meaning "employee", 2 meaning "employer" and so on.

When this code list is attached to **Data Elements**, data in specific dataset can be properly interpreted.

View the example 'Status in employment code N' Value Domain online at:
https://aristotle.cloud/item/270622/valuedomain/status-in-employment-code-n
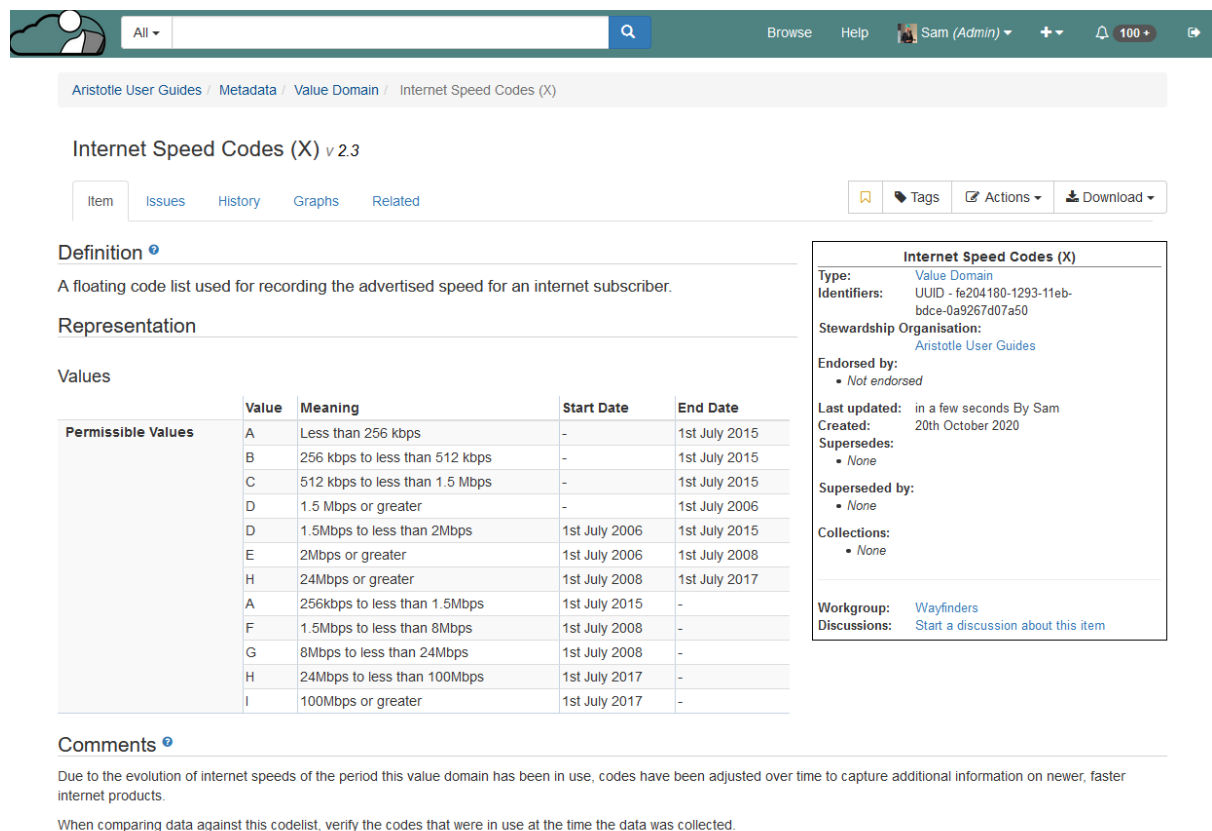
Storing code lists that change over time

As data can be collected over a range of years, metadata systems must be prepared to handle scenarios where data definitions and codes can change over time. This idea of tracking data as it changes over time is often called 'longitudinal data'.

Within the Aristotle Metadata Registry, code lists can be updated over time to ensure that users can accurately understand data as it was collected at different points in time.

In the example below, we have constructed a code list for recording Internet speeds provided to subscribers. Due to the evolving nature of the internet, internet speeds have increased dramatically over the past 10 years. Fortunately, the Aristotle Metadata Registry supports the recording of dates that codes were valid when managing code lists as Value Domains.

Because of this, definitions for Value Domains and Data Elements can be updated, without having to be recreated, which simplifies the management of 'longitudinal data' over time.

View the example 'Internet Speed Codes' Value Domain online at:
https://aristotle.cloud/item/605097/valuedomain/internet-speed-codes-x

Storing concept lists using Conceptual Domains

Where code meanings have are complex or are reused in different code lists, they can be recorded separately in a **Conceptual Domain**. A Conceptual Domain allows users to record the different ways that data can be categorised, without specifying how these categories are recorded in data. This means that different data sets that use different codes, can still link to the same conceptual meanings, which makes comparison of different but related data easier.

This example, the Longitudinal Employment Study **Dataset**, records employment data collected over a ten-year period. In this **Dataset**, each **Distribution** records a data file for each year so that changes to data across time can be recorded.

This example also shows how the formats used to store the data have change over time, so users can understand how the data can be access and analysed.

View the example 2009-2019 Longitudinal Employment Study Data record online at:
https://aristotle.cloud/item/605082/conceptualdomain/status-in-employment-concepts

Using a Classification to store hierarchical codes and categories

A **Classification** can be used to record complex and deep data classifications.

This example **Classification** is based on an official classification of occupations – the Australia and New Zealand Standard Classification of Occupations. This classification stores over 1000 job categories at 5 levels

View the Australia and New Zealand Standard Classification of Occupations online at:
https://aristotle.cloud/item/605070/classificationviewer

Because this classification is recorded as metadata within the registry, all **Value Domains** that link to this can be directly accessed through the registry as well.

View metadata related to the Australia and New Zealand Standard Classification of Occupations online at: https://aristotle.cloud/item/605070/related/value_domains

References and Links

To ensure consistency between user stories across different organisations, the Aristotle Metadata Registry uses standard terminology when describing metadata types. Using standard terms makes it easier for people communicate by giving them a common vocabulary for describing how they manage their data and metadata.

The terms **Value Domain** and **Conceptual Domain** are defined in the internationally recognised ISO/IEC 11179 standard for metadata registries. The term **Classification** comes from the UN Statistics Division Classifications Model.

Links to these standards are included below.

ISO/IEC 11179: http://metadata-standards.org/11179/

UNSD Classification Model:
https://unstats.un.org/unsd/classifications/expertgroup/egm2015/ac289-22.PDF

When browsing or creating metadata within the Aristotle Metadata Registry, **Value Domains** and **Conceptual Domains** are included under "**Basic Metadata Registry Objects**" section.

When browsing or creating metadata within the Aristotle Metadata Registry, **Classifications** are included under "**Classifications Management**" section.

For more help on creating metadata in the Aristotle Metadata Registry, visit our Knowledge base: https://desk.zoho.com.au/portal/aristotlecloudservicesaustralia/kb/articles/creating-metadata-items

For more information on Aristotle.Cloud or the Aristotle Metadata Registry, visit us online at www.aristotlemetadata.com, contact the Aristotle team at hello@aristotlemetadata.com.