Rationale for Signification Pattern

One of the powerful ideas behind metadata management is reuse. Sometimes one part of the description (the metadata) for an object may be reused to describe another, because there is something similar between them. For instance, different variables for gender might share the same value domain, say,

Gender Codes - {<m, male>; <f, female>}.

Different value domains might share the same categories, say,

Gender Codes 1 - {<m, male>; <f, female>} and

Gender Codes 2 - {<0, male>; <1, female>}.

The idea of reuse also applies to the DDI-4 model itself. There are sections in the model that require the same structure as in other sections. For instance, code lists, category sets, and classification schemes can be modeled in similar ways. To prevent the proliferation of incompatible structures for solving the same problem, we develop similar sections of the model wherever needed through constructions called patterns. There are currently four patterns used in DDI-4 now, and this paper contains a rationale for the Signification Pattern.

Signification is the result of assigning some name or label to an object. This outcome arises in many forms. The most common ones, with a short description of each, are:

* Label: linguistic[[1]](#footnote-1) alphanumeric string assigned to some object
* Name: nonlinguistic alphanumeric string assigned to some object
* Identifier: label or name intended for dereferencing[[2]](#footnote-2)
* Locator: identifier with a dereferencing mechanism included[[3]](#footnote-3)
* Designation: label or name assigned to some concept, where a concept is a kind of object
* Code: non-linguistic designation
* Term: linguistic designation
* Numeral: numeric designation, typically used to designate numbers

The purpose in providing each of the descriptions listed above is to illustrate the common feature throughout. This common feature is the assignment of a handle (some way of referring) to an object. The object is often referred to as the referent, or in our case it is called the Signified. The string being assigned is called the Signifier. The combination, the assignment, is called the Sign.

The Signifier does not have to be a string. It could be some symbol, such as an emoticon, a traffic light, or a road sign. However, for the purposes of DDI, it is simpler to restrict Signifiers to strings.

There are several advantages to addressing the needs of signification in a uniform way, but first it is probably better to speak of the disadvantages. There is one big one – it is complicated. We need three classes in the pattern model: Signifier, Signified, and Sign. Why can’t we just assign a label or name to an object through an attribute and be done with it?

Anyone who has searched on the Web through a search engine such as Google or Bing knows the answer –

* Search words often uncover irrelevant pages and sites (homonyms)
* Searches don’t find pages indexed by synonyms of search words (synonyms)

The Signification Pattern provides the means to manage homonyms and synonyms. Since the pattern provides the means to separate the labels and names from the objects they tag, it is simple to find all objects labelled or named the same way (homonyms) and to find all labels or names for a particular object (synonyms).

Finally, the DDI-4 model is bound to XML and RDF for ease of implementation. The resulting bindings will look the same anywhere there is a label or name assigned to an object due to the use of the Signification Pattern. Ultimately, this makes it easier to manage and use the metadata (in this case, labels and names) collected.

1. A string is linguistic if the characters spell a word in natural language or some technical special language. It is non-linguistic otherwise. [↑](#footnote-ref-1)
2. Dereferencing is a term from computer science. It refers to retrieving an object from some storage facility. For example, every time a web page is displayed in the browser, it is dereferenced from the server on which it is stored then transferred to that browser. [↑](#footnote-ref-2)
3. The URL (Uniform Resource Locator), which is the address or identifier for a web page, uses the HTTP protocol to be able to display the identified page. HTTP is the dereferencing mechanism in this case. [↑](#footnote-ref-3)