

SIMS 202 PHONE PROJECT – Metadata Creation Guidelines

October 8, 2003

0. DOCUMENT SUMMARY

1. INTRODUCTION

2. METADATA CREATION GUIDELINES

1. INTRODUCTION

This document contains a set of guidelines for creating metadata frameworks that resulted from the metadata feedback sessions.

2. METADATA CREATION GUIDELINES

1. Format

The Excel file containing your metadata system will be used as an input to the Flamenco application. Treat this document like code. The syntax has to be right or it won't compile. The structure and formatting of your spreadsheet must comply with the formatting guidelines specified in Assignment 4 (Part 1) sections 2 and 3.

2. Metadata Verification

Using your application scenario as a guide, evaluate whether your metadata structure is sufficiently developed in its breadth and depth. Could your metadata framework organize the data that you need to implement your application scenario?

3. Monohierarchical Rule

The monohierarchical rule is violated if 2 descriptors under the same facet have same name.

For example,

People>Individual>**Name**

People>Group>**Name**

uses the Name descriptor twice under the People facet. The monohierarchical requirement means that a particular descriptor has only one parent in the hierarchy.

4. Front End vs. Back End

Remember that your metadata framework is part of your application back end and as such has to support a complete set of descriptors. Your

application front end will contain the logic that retrieves and presents only the information that is specifically necessary for your application.

5. Specificity Level

If descriptors exist on the same level in the hierarchy, they should denote the same level of specificity. In the example below,

Place>Type>Building>**Theater**

Place>Type>Building>**Room**

Theater and Room do not have the same level of specificity because Room is part of a building, while Theater is a type of building.

Another thing to look out for as you determine whether subfacets at the same level are also the same degree of specificity is whether all descriptors at a particular level are also the same part of speech. A facet sublevel that contains a mix of descriptors like Strenuous, Light, Working, Playing probably needs to be rearranged so the adjectives and gerunds are in the right facets. Strenuous and Light describe activities while Working and Playing are activities.

6. Combining Descriptors

Recognize that a descriptor may be created by combining descriptors from more than one facet. For example, the incorrect descriptor path

Activities>Tasting>Wine

can be corrected by using an Activities facet that contains Tasting and an Object facet that contains Wine. When you go to annotate a photo depicting wine tasting you use one descriptor from the Activities facet (Activities>Tasting) and one descriptor from the Object facet (Object>Drink>Wine). So each facet should be targeted at categorizing broad types of concepts, e.g. Things, Places, People. These general facets can be combined to describe many particular items or individuals.

Since top level facets occupy the highest level in the metadata hierarchy, they should be general enough to support applications other than your own. Avoid names that are vague or too specific. For example, a facet called Culture is too vague because it is unclear what Culture refers too. On the other hand, a facet called Equipment is too specific because not all objects can be classified as Equipment.

7. Quantifiers

Descriptors that convey quantity can be applied to a variety of entities and should be grouped together in a quantifier facet. For example, the quantifiers Small, Medium, and Large can be used to quantify either a Person or Object.

8. Use of Slashes

Avoid using slashes in descriptor names. Slashes are a good indication that a facet should be broken down into more levels of increasing specificity, or that you need to decide on the best of several synonymous words.

9. Mutual Exclusivity

Descriptors on the same hierarchical level should be mutually exclusive to eliminate ambiguity. For example, the descriptors

Activities>**Creative Arts**>

Activities>**Entertainment**>

are not mutually exclusive because Dancing can be classified under either of them.

Questions? Email: is202-ta@sims.berkeley.edu