



Copyright CSW Informatics Ltd 2003

SUMMER
XML **CSW** 2003
SCHOOL

Code Lists

- A code is a character string that represents a definitive value
- Code lists are valuable as unambiguous taxonomies
- In many cases, such as product classifications, code lists are big business
 - Some code list owners charge for their use

Colors

Pick one:

01=white 02=blue

03=red ...

Countries

Pick one:

AW=Aruba CA=Canada

FR=France ...

- Often they are merely maintained in text documents
- But formal encodings are extremely useful, for example:
 - RDF ontologies
 - The ebXML Registry Information Model's `<ClassificationScheme>` markup
 - XSD (such as enumerated simple types)
- You could develop different representations for different purposes

- Schema validation can do code checking “for free”
- This step usually occurs early in the processing pipeline
- This encoding benefits from tool availability
 - And could even be generated from a more-primary XML representation
- These all support UBL’s “leverage XML technology” goal

- Many code lists are too large (~10K codes) or dynamic (~daily) to take advantage of XSD
 - But one study showed more than one-third of legacy code lists to be variants of Yes/No!
- Validation through schemas will never be complete for some applications
 - Such as codes that become dynamically invalid depending on previous code choices

- But re-coding a code list over and over in different schemas is costly and prone to error
- Better to help code list owners produce their own code list schema modules

UBL elements...
UBL types...

Colors

Pick one:

01=white 02=blue
03=red ...

UBL elements...
UBL types...

Colors

Pick one:

01=white 02=blue
03=red ...

- A code list owner can choose to conform to the rules by producing a reusable schema module that defines a code list datatype
- The level of validation is entirely up to them
 - Enumeration
 - Regular expression
 - No constraints
- The “normative status” of the module is also up to them
- They just need to provide enough metadata to uniquely identify the meaning of each code
- We're working with a number of groups to help them do this

- UBL elements would be bound to a foreign type defined by a code list owner
 - This would be done in the “code list adapter module”
- The metadata attributes could be defaulted, or even fixed

```
<ubl:CountryID  
  xsi:type="unece:ISO3166CountryCodeType"  
  various metadata attributes...>  
FR  
</ubl:CountryID>
```


- If all goes well, we could see the following benefits:
 - Less duplication of work in XML vocabulary development
 - Wider application support for well-known code lists
 - Earlier validation of code values
 - Standardization of more code lists, and even formally described subsets and extensions
 - Greater “semantic clarity” through identifying standard code list metadata