2nd S-100WG MEETING
Genoa, Italy – 15-18 March 2017

Paper for Consideration by the S-100 Working Group (S-100WG)

Use of Codelists

Submitted by:	IHO Secretariat (TSSO).
Executive Summary:	More detailed guidance is required on the use of Codelist type attributes in S-100 based Product Specifications, particularly so as to distinguish their use from enumerated attribute types.
Related Documents:	1. IHO Publication S-100 – *Universal Hydrographic Data Model*: Appendix 11C – *Guidance on Codelists*. 2. European Commission – INSPIRE Registry (http://inspire.ec.europa.eu/registry).
Related Projects:	Development of the IHO Geospatial Information Registry; S-100 based Product Specification development.

Introduction / Background

Throughout S-100, there are numerous references to attribute type codelist – in particular S-100 Appendix 11C - *Guidance on Codelists*. Despite this information included in S-100, there is still little understanding of how codelists are to be implemented in the IHO GI Registry and S-100 based Product Specifications, and in particular when to distinguish between application of a codelist type or an enumerated attribute type in Product Specification (Feature Catalogue) modelling.

Analysis / Discussion

Since the IHO GI Registry became operational in October 2016, there have been numerous discussions between the FCD Register Manager (and other IHO Secretariat staff), Submitting Organizations and Domain Control Body representatives as to when a registered item should be modelled as an enumerated attribute type or a codelist attribute type. Codelist type is a new attribute type introduced in S-100 that does not exist in the S-57 Standard, therefore there is little knowledge to date of its application in data modelling within the IHO community (and in particular for the FCD Register Manager). Despite consulting the numerous references to codelists included in S-100, including Appendix 11C - *Guidance on Codelists* (see Annex to this paper), there appears to remain considerable confusion in this regard.

For instance, a proposal was submitted to the FCD Register to include a new enumerated attribute **dayOfWeek**, with enumerate values **monday** ….. **sunday**. After discussion between the FCD Register Manager and other IHO Secretariat staff; and consultation with the Submitting Organization, it was decided to register **dayOfWeek** as a codelist type. This decision was based principally on the following factors:

- The days of the week are a universally recognised convention;
- There is no requirement to fully "define" the attribute or its values in regard to evaluation and maintenance (as "hydrographically relevant" concepts) within the Registry;
- The concept "day of the week" and its values are fixed and extremely unlikely to change;
- The concept would be similarly recognised (and likely utilized) across all Domains within the FCD Register.

In addition to the above, it was noted that a similar concept **dayTypeCode** has been included in the INSPIRE Codelist Register (http://inspire.ec.europa.eu/codelist/DayTypeCode). However, on completion of registration of this concept as a codelist type in the FCD Register, there were concerns raised by some other FCD Register Domain representatives that this should remain as an enumerated attribute type.

In order to better inform data modellers and all participants in the Registry proposal process, it is suggested that more concise guidance, including worked scenarios demonstrating clear distinctions between the application of enumerated and codelist attribute types, is required. Such guidance should be included as an extension to the guidance already included at Annex 11C of S-100, and/or in guidelines/conventions for proposals to the IHO GI Registry (see S-100WG2-09.4A) While the FCD Register Manager is prepared to take the lead in the development of this guidance, as acknowledged

above his knowledge on the application of codelists is limited. Therefore it is requested that input is required from anyone who has some knowledge on this subject, or who may be interested in contributing to such discussion, on a volunteer basis.

Conclusions

IHO knowledge of the use of codelist type attributes is limited. While there is some guidance on the use of codelists in S-100, there is still a considerable amount of confusion as to when to implement a concept as an enumerated or codelist type attribute. Additional guidance is required to better inform data modellers and all participators in the Registry proposal process on the use of codelists.

Recommendations

1. S-100WG to agree to the development of enhanced guidelines for Submitting Organizations, Domain Control Bodies, Register Manager(s) and data modellers on the use of the codelist attribute type, particularly provision of a clearer distinction between codelist and enumerated attribute types.

2. If approved, S-100WG to agree that these guidelines are to be published as an extension to S-100 Annex 11C. Additionally, it is recommended that guidance also be included in "Guidelines for Proposals to the IHO GI Registry".

3. S-100WG to approve the establishment of a small focus group, coordinated by the FCD Register Manager (TSSO of the IHO Secretariat), to do this work; and provide the result to the S-100WG for approval.

Justification and Impacts

The recommendations included in this paper are the result of the observations of the IHO Secretariat (TSSO) since the IHO GI Registry became operational and the TSSO position was activated in October 2016; in conjunction with discussions with ADDT, the S-100WG Chair, and participants in the FCD Register proposal process (including data modellers). To date there has been no clear consensus as to the application of codelist type attributes resulting from these discussions.

Clearer guidance as to the use of codelists and distinctions from the use of enumerated attributes will contribute to more consistent data modelling in S-100 based Product Specifications, and more consistent Data Dictionary content within the IHO GI Registry.

Action required of NCWG

The S-100WG is invited to:

    a.   **Note** this paper.

    b.   **Approve** the development of enhanced guidance for the implementation and use of codelist type attributes.

    c.   **Approve** the establishment of small focus group to perform this task, operating by correspondence, under the S-100WG.

Appendix 11-C
Guidance on Codelists
(informative)

11-C-1.   Introduction to Codelists

Product specifications should balance all relevant considerations, for example implementation costs, application operational environment, cross-domain reuse, and reduction of maintenance and distribution efforts, when deciding which approach to use for any particular attribute.

11-C-2.   Modelling

When deciding between using a codelist and enumeration, consider the completeness, stability, source, reuse, and application dependencies of the list of values.
- If the set of allowed values is fixed and reasonably short (say, fewer than 20 values?), an enumeration must be used.
- If the list is fixed but long, an enumeration is preferred but a "dictionary model" codelist may be used.
- If only the likely values of an enumeration are known, or the list may be extended by data producers or the user community, a codelist must be used. Whether the "dictionary" or "open" form is preferable depends on who might add values – if it is maintained by an organization, the dictionary form is preferable, if user communities or data producers may add values, the "open" form is preferable.
- If the allowed values change frequently and the list should be updated without major revisions of the product specification, a codelist may be used. The "dictionary" form may be preferable under these circumstances.
- If application logic or portrayal rules depend on values, an enumeration is preferred but a codelist may be used if the logic/rules can be written to cover all possible values (for example, using wildcards or defaults), or otherwise allow graceful recovery from unanticipated values.
- Collections which have internal structure (e.g., types and subtypes of vessels) should be modelled as "dictionary" codelists, pending discussion of the matter by ISO TC211.

11-C-2.1.   Hierarchies of codelists

A codelist may also be used as a super-type for more specific codelists. The vocabulary of the super-type is the union of the vocabularies of its sub-types[1]. If additional values are permitted the super-type must be an "open enumeration" or "open dictionary" codelist. Practically, this allows vocabularies developed by different domain expert groups or organizations to be merged.

11-C-3.   Codelists maintained by external organizations

If there is an existing well-established codelist maintained by a responsible source, it can be referenced in an application schema. The codelist should meet the following requirements[2]:
- It must be managed by a responsible source – an official national or international standards body, long-established user community, group, or consortium;
- The codelist and its values must be identified by persistent HTTP URIs;
- The list should be well-maintained, meaning all its values must remain available forever, even if they have been deprecated, retired or superseded;
- The list should be in a dictionary language accepted for use in S-10x product specifications.

The IHO may be requested to arrange for the translation, reproduction, and maintenance of codelists meeting only some of the above requirements. Note that this may necessitate a discussion between the IHO and the source.

11-C-4.   Data formats of codelist typed attributes

The codelist model in S-100 is designed to be flexible by decoupling application schema from data format to some extent. Data formats may use "codelist extractions" created by extracting codes or values from a codelist dictionary and treat them as ordinary enumerations. The effect is to allow data formats to use either an external dictionary or

---

1 Note that the super-type cannot augment the union set with additional definitions. This conforms to the INSPIRE usage but it may be reconsidered if such augmentation is required at a later time.
2 Adapted from INSPIRE guidelines.

ordinary enumerations. For example, an XML data format might convert an *ISO3166CountryCodes* codelist maintained by IHO into an XML Schema type:

```
<xs:simpleType name="ISO3166CountryCodesType">
  <xs:restriction base="xs:string">
    <xs:enumeration value="EN"/>
    <xs:enumeration value="FR"/>
    ... other country codes ...
```

As far as implementations using that schema are concerned, it is indistinguishable from an ordinary enumeration. The decision as to which alternative(s) to use in any particular product specification should depend on the circumstances of the data product and its use environment. The decision should be made by the product specification authors when developing the data format. Obviously allowing different data formats to use different representations introduces additional maintenance requirements relating to some data formats, these would be limited to the formats which use "closed" representations (i.e., convert the codelist to an ordinary enumeration).

11-C-4.1. GML and other XML data formats

Enumeration with pattern: The data format in XML schemas must conform to ISO 19136 E.2.4.9, i.e., a union of an enumeration and a pattern of the form:

```
other: [a-zA-Z0-9]+( [a-zA-Z0-9]+)*
```

Examples of use (assuming a codelist which explicitly lists "Norwegian" but not Nynorsk and Bokmål):

```
<language>nor</language>      <!-- Norwegian is an enumerated value -->
<language>other: nno<language> <!-- Norwegian Nynorsk is not an enumerated value -
->
```

External Dictionary: The data format in XML schemas must be the XML Schema built-in types *anyURI*. The use of spaces is discouraged.

Example: (UN/LOCODEs, United Nations Code for Trade and Transport Locations)

In XML schema: Type definition:

```
<xs:simpleType name="unLoCodeType" type="xs:anyURI">
```

and later (in the feature definition):

```
<xs:element name="unLoCode" type="unLoCodeType"/>
```

In a dataset:

```
<unLoCode xlink:href="http://registry.iho.int/codelists/locode/2013/1/USNYC"/>
```

for New York City, identified by code "US NYC" in the UN/LOCodelist version 2013-1 (published July 2013).

11-C-4.2. ISO 8211 encodings

Enumeration with pattern: To accommodate producer-defined values ("other: xyz") this can be encoded either as a "text" type (character string) or as a complex attribute with an integer sub-attribute (for the listed allowed values) and a text sub-attribute (the "other:..." values).

External Dictionary: This can be encoded in two ways:
1. A URI data type with value a URI constructed by combining the URI for the vocabulary (dictionary) and the item code. For example:
   `http://registry.iho.int/codelists/locode/2013/1/USNYC` for New York City (in the July 2013 edition of UN/LOCODEs list).
2. A complex attribute with two sub-attributes: Vocabulary location (URI) and item code (text). To use the same example: sub-attributes are *vocabulary*=
   `http://registry.iho.int/codelists/locode/2013/1/` and *itemCode*=USNYC.

The first method is recommended as it reduces data complexity.

11-C-5. Dictionary formats

Use of GML dictionary or SKOS format is recommended. Other formats may be considered under compelling circumstances or after the development of standards in ISO or elsewhere.

11-C-6. Dictionary distribution and discovery

In order to remove dependence on Internet connectivity for interpreting codelist values, codelist dictionaries may be distributed as support files in exchange sets. For the purposes of distribution, discovery, management of updates, and version control, such local dictionary files can be treated as ordinary support files. Discovery metadata for support files is described in Part 4a (see class S100_SupportFileDiscoveryMetadata).

11-C-6.1. Entity resolution with local dictionary files

If mappings from namespaces to dictionary files are needed for a data product, the use of a catalogue file is suggested in which case the product specification may specify the catalogue file name and format. The catalogue file itself can be treated as another support file, having a fixed filename and location in the exchange set which are stated in the product specification.

EXAMPLE   A product specification uses XML catalogues for resolving codelist namespaces to local dictionary files. It specifies that the catalogue file shall conform to the OASIS standard for XML catalogues ("XML Catalogs V. 1.1"),

URL: https://www.oasis-open.org/standards#xmlcatalogsv1.1). The product specification standardizes the name of the catalogue file as CODELSTCAT.XML.