



STANDARDIZATION of NAUTICAL PUBLICATION WORKING GROUP (SNPWG)

[A Working Group of the Hydrographic Services and Standards Committee (HSSC)]

Chairman: Jens SCHRÖDER-
FÜRSTENBERG jens.schroeder-fuerstenberg@bsh.de

Vice-Chairman: Thomas LOEPER thomas.loeper@noaa.gov

Secretary: vacant

SNPWG Letter: 04/2014

To SNPWG Members

Date 28 August 2014

Dear Colleagues,

Subject: Simplification of the SNPWG data model on textual information

References: <http://www.fuerstenberg-dhg.de/mediawiki/index.php/TXTCON>

The development of the MPA and Radio Signal Product Specifications and the initial check of the S-101 base line version came up with the need to simplify the current SNPWG data model for textual information.

The current SNPWG data model of textual information employs the following separate complex attributes:

information (<http://www.fuerstenberg-dhg.de/mediawiki/index.php/INFORM>)

textualDescription (file reference) (<http://www.fuerstenberg-dhg.de/mediawiki/index.php/TXTDSC>)

textContent (<http://www.fuerstenberg-dhg.de/mediawiki/index.php/TXTCON>)

The first 2 complex attributes are taken from the S-101 data model.

The third attribute is being used to cover all kinds of textual information. It contains the same sub-attributes as the first two attributes and adds other sub-attributes which provide more information about the text content.

While the first two attributes might be sufficient to encode text information which is present on paper charts, nautical publications datasets often need more expressive modeling of text. On the other hand, allowing three different ways to model text may be unnecessarily complicated. The simplified SNPWG approach would employ only the third method for all kinds of textual information. In other words, features (and information types) would bind only the textContent attribute instead of information, textualDescription, or textContent. The full UML diagram and the revised definitions are provided in the Annex.

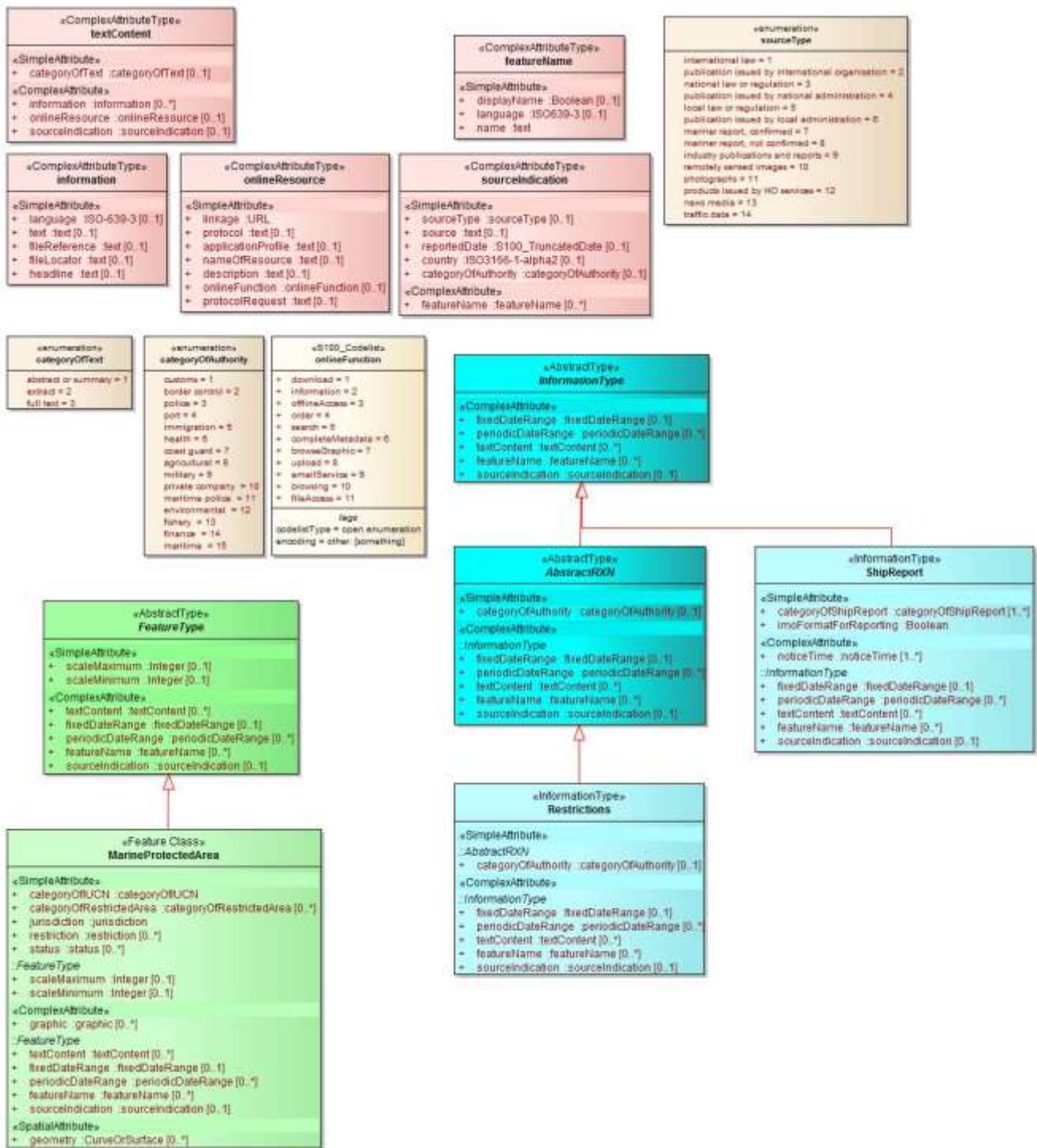
We know that, if “textContent” is used everywhere, each instance of “textContent” in a dataset results in some extra overhead for data volume and data retrieval, and perhaps a little more work for the human data encoder. There might also be added complexity for portrayal rules and implementing portrayal. However, it is considered worthwhile to be tested.

We expect comments, suggestions, improvements and any other contributions on the file at the latest by 30 September, 2014.

Yours sincerely,
Jens Schröder-Fürstenberg, Chairman

Annex

Full textContent UML data model, with examples of its use in feature and information types.



Enterprise Architect report detailing selected items:

textContent

«ComplexAttributeType»

Alias:

Package: TXTCON_INFORM_Proposal

Subclass Of:

Notes:

Remarks:

- Exactly one of sub-attributes **onlineResource** or **information** must be completed in one instance of **textContent**.
- Product specifications may restrict the use or content of **onlineResource** for security. For example, a product specification may forbid populating **onlineResource**.
- Product specification authors must consider whether applications using the data product may be prevented from accessing off-system resources by security policies.

Inherited Attributes: (none).

Constraint:

count(information + onlineresource) > 0

Attributes (for feature and information types) OR Listed Values (for enumerations and codelists)

Multiplicity and Type columns should be ignored for enumerations and codelists.

Attr. Name / Enum. Label	Mult.	Type	Code	Notes / Definition
categoryOfText Alias: CATTXT	[0..1] Seq:False	categoryOfText		
information Alias:	[0..*] Seq:False	information		
onlineResource Alias:	[0..1] Seq:False	onlineResource		
sourceIndication Alias:	[0..1] Seq:False	sourceIndication		

information

«ComplexAttributeType»

Alias:

Package: TXTCON_INFORM_Proposal

Subclass Of:

Notes:

Provides textual information that cannot be provided using other allowable attributes for the feature, in a defined language. The information may be provided as a string in sub-attribute **text**, or by encoding the file name of a single external text file that contains the text in sub-attribute **file reference**.

Remarks:

- The sub-attribute **text** should be used, for example, to hold the information that is shown on paper charts by cautionary and explanatory notes. No formatting of text is possible within the sub-attribute **text**. If formatted text is required then an associated text file referenced by the sub-attribute **file reference** must be used.
- The sub-attribute **file reference** is generally used for long text strings or those that require formatting, however there is no restriction on the type of text (except for lexical level) that can be held in files referenced by sub-attribute **file reference**.

Inherited Attributes: .

Constraint:

count(text + fileReference) > 0

Attributes (for feature and information types) OR Listed Values (for enumerations and codelists)

Multiplicity and Type columns should be ignored for enumerations and codelists.

Attr. Name / Enum. Label	Mult.	Type	Code	Notes / Definition
language <i>Alias:</i>	[0..1] Seq:False	ISO-639-3		ISO 639-3 value
text <i>Alias: INFORM</i>	[0..1] Seq:False	text		A non-formatted digital text string. Remarks: The attribute should be used, for example, to hold the information that is shown on paper charts by short cautionary and explanatory notes. Therefore text populated in text must not exceed 300 characters. Text may be in English or in a national language defined by the attribute language . No formatting of text is possible within the sub-attribute text . If formatted text, or text strings exceeding 300 characters, is required, then the attribute file reference must be used.
fileReference <i>Alias: FILREF</i>	[0..1] Seq:False	text		The string encodes the file name of a single external text file that contains the text. Remarks: The attribute file reference is generally used for long text strings or those that require formatting, however there is no restriction on the type of text (except for lexical level) that can be held in files referenced by sub-attribute file reference .
fileLocator <i>Alias: FILLOC</i>	[0..1] Seq:False	text		The string encodes the location of a fragment of text or other information in a support file. Remarks: <ul style="list-style-type: none">• Application schemas must describe how the associated file is identified. The associated file will commonly be named in a file reference co-attribute of the same complex attribute.• Each DCEG must specify requirements for the format of the associated file and the semantics of file locator. For example, the value of file locator may be an HTML ID in an HTML file, line number in a text file) or a bookmark in a PDF file.
headline <i>Alias: HEADLN</i>	[0..1] Seq:False	text		Words set at the head of a passage or page to introduce or categorize

categoryOfText

«enumeration»

*Alias:**Package:* MPADomain*Subclass Of:**Notes:***Inherited Attributes:** .**Attributes (for feature and information types) OR Listed Values (for enumerations and codelists)***Multiplicity and Type columns should be ignored for enumerations and codelists.*

Attr. Name / Enum. Label	Mult.	Type	Code	Notes / Definition
abstract or summary <i>Alias:</i>	[1] Seq:False		1	A statement summarizing the important points of a text.
extract <i>Alias:</i>	[1] Seq:False		2	An excerpt or excerpts from a text.
full text <i>Alias:</i>	[1] Seq:False		3	The whole text