Paper for consideration by TSMAD

XML Encoding of Marine Protected Areas Information

Submitted by:	Jeppesen/SNPWG
Executive Summary:	An example of XML/GML encoding of S-100-compatible information
Related Documents:	(1) S-100 Ed. 1.0.0 (2) S-101 Product Specification Phase 4
Related Projects:	(1) S-100 (2) S-101

Introduction / Background

SNPWG recently worked on an exercise in mapping and converting data on marine protected areas (MPAs) to the SNPWG data model and encoding the result in an XML/GML format. The data model, schemas, example datasets are available on the SNPWG Wiki and in SNPWG 15 documents at the IHO SNPWG15 web site.

Terms and Abbreviations

GML	Geography Markup Language
MPA	Marine Protected Area
OGC	Open Geospatial Consortium
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

References

S-100: Universal Hydrographic Data Model, Edition 1.0.0, January 2010.

S-101: Electronic Navigational Chart Product Specification; Phase 4, June 2012.

WIKI: SNPWG Wiki at http://www.fuerstenberg-dhg.de/mediawiki/index.php/Main_Page

Discussion

The main domain model portion of the MPA application schema prepared for the SNPWG 15 meeting in November 2012 is given by the UML diagram in Figure 1. The model includes feature classes taken from S-57, as well as feature and information classes defined by SNPWG as part of the NPUBS information model and described on the SNPWG Wiki [WIKI]. A subsidiary model of data quality was also developed derived from the data quality model proposed by DQWG and TSMAD.

XML schemas were developed for the MPA domain and components required by those schemas and described in S-100 and related work by TSMAD and DQWG. The MPA schemas use the GML 3.2.1 XML schemas published by the OGC as well as related schemas used by GML (i.e., W3C *xlink* and ISO 19139 schemas). Since the MPA XML encoding schemas are derived from S-100, and also designed for compatibility with the then-current draft S-101 product specification, it was necessary to prepare XML schemas defining generic, metadata and data quality elements and attributes defined in S-100 and S-101. Uncertainties in S-100/S-101 specifications were resolved by using reasonable solutions. These schemas may be of interest to TSMAD, including for the identification of gaps and ambiguities in S-100 and S-101. Metadata types for discovery metadata and the exchange catalogue are shown in Figures 2-4. Examples of feature and information classes in the MPA domain are shown in Figure 5 and Figure 6.[1]

Sample XML transfer sets were developed for SNPWG and are available on the SNPWG 15 documents at the IHO SNPWG15 web page. These examples include MPA information encoded as XML datasets, support files, data quality features, and metadata in a catalogue file.

---

[1] JPGs are included in the accompanying package for expanded viewing.

The package accompanying this paper includes XML schemas for MPA datasets, metadata, and data quality. It also includes a sample transfer set with data for the US Hawaii MPA. This transfer set contains instances of geographic and information classes, feature and information associations, and a data quality object. It also includes an HTML support file containing text extracts from the relevant sections of the US Code of Federal Regulations and a catalogue file. The modeling approach (e.g., references in the dataset to specific places in the support file) can be understood by referring to the SNPWG Wiki.

The contents of the folders are described in the README_SCHEMAS.txt file in the package.

Refinements and extensions are necessary and anticipated, such as:

1. Consideration of development and use of profiles of GML to provide constructs more closely tailored to the needs of the specific domain, for example, GML *CircleByCenterPoint* (corresponding to GM_Circle in ISO 19107) to describe a protected area in terms of center and radius (a common legal definition) and/or excluding irrelevant GML constructs (e.g., *PolyhedralSurface* and *TriangulatedSurface* as spatial primitives that are substitutable[2] for *Surface*).
2. Addition of dataset general information, structure, identification, CRS information.
3. Definition of update format.

Conclusion

The XML schemas described here are a work in progress, an exercise in exploring the necessities and issues in constructing an XML/GML encoding of information conforming to the S-100 framework. Work on additions, refinements, and extensions, is underway. TSMAD may be interested in the subset of schemas dealing with requirements common to multiple domains, such as metadata.

Action Required of TSMAD

TSMAD is invited to:

- note this paper

---

[2] *PolyhedralSurface* and *TriangulatedSurface* are spatial types defined by the OGC GML 3.2.1 schemas, whose design allows them to be used in place of *Surface*. Their presence introduces an unnecessary inconvenience while using off-the-shelf XML editing tools and irrelevant clutter in diagrams.

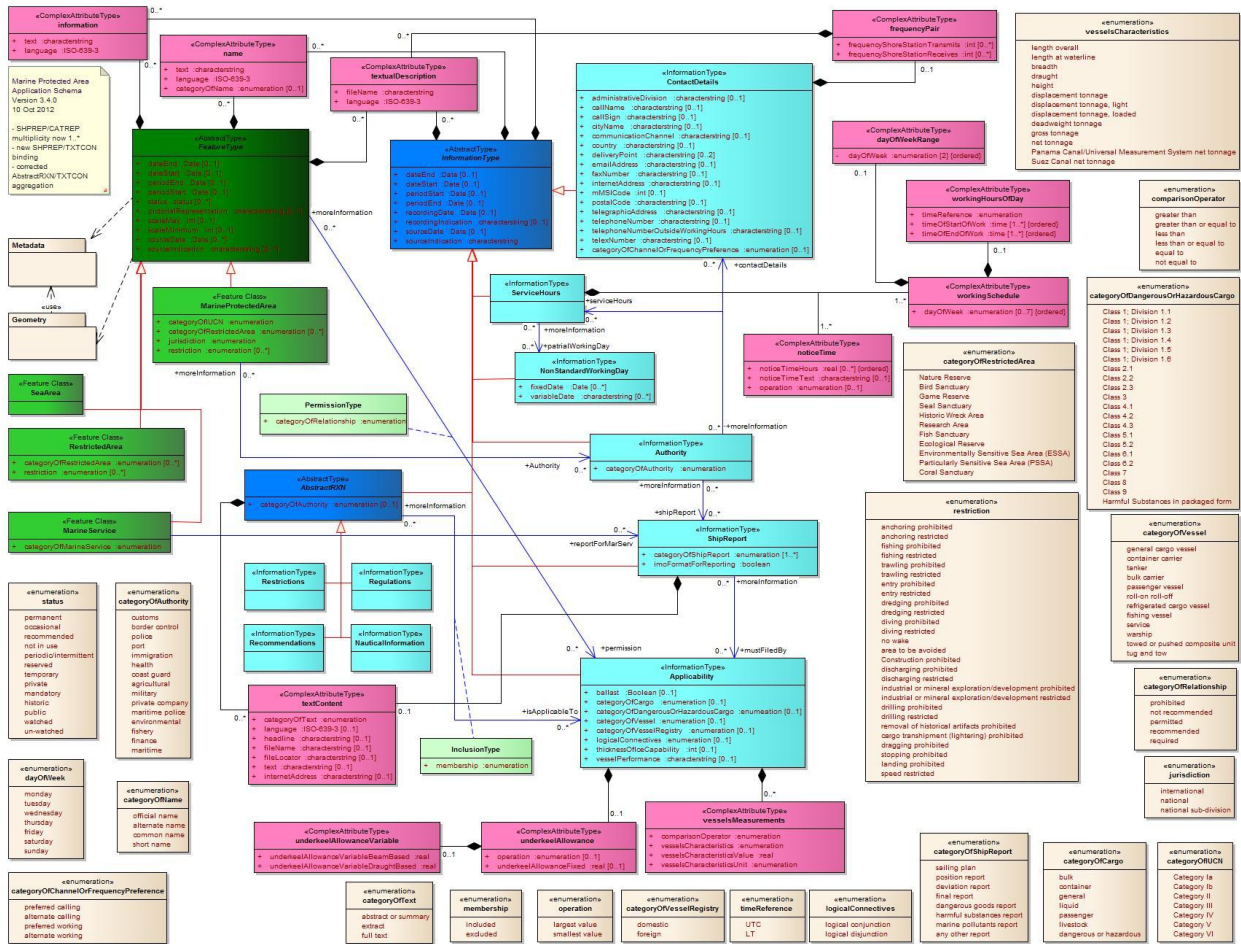# Annex A. MPA domain model and XML structure examples
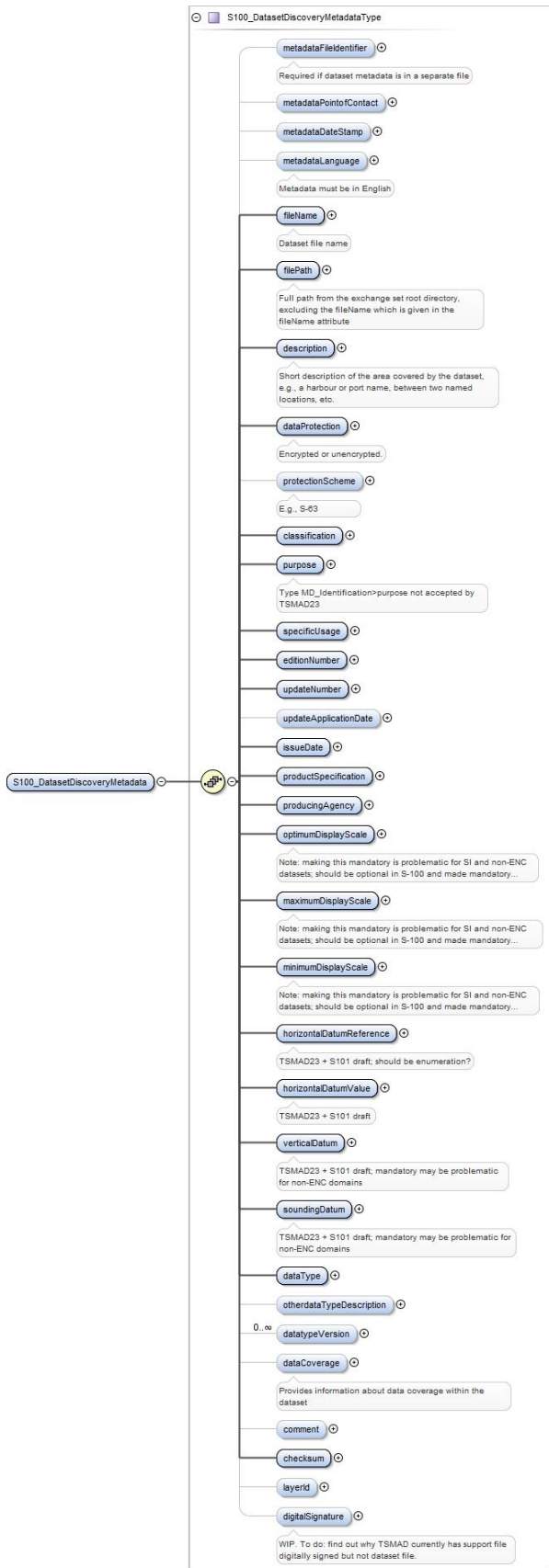

Figure 1. MPA domain model
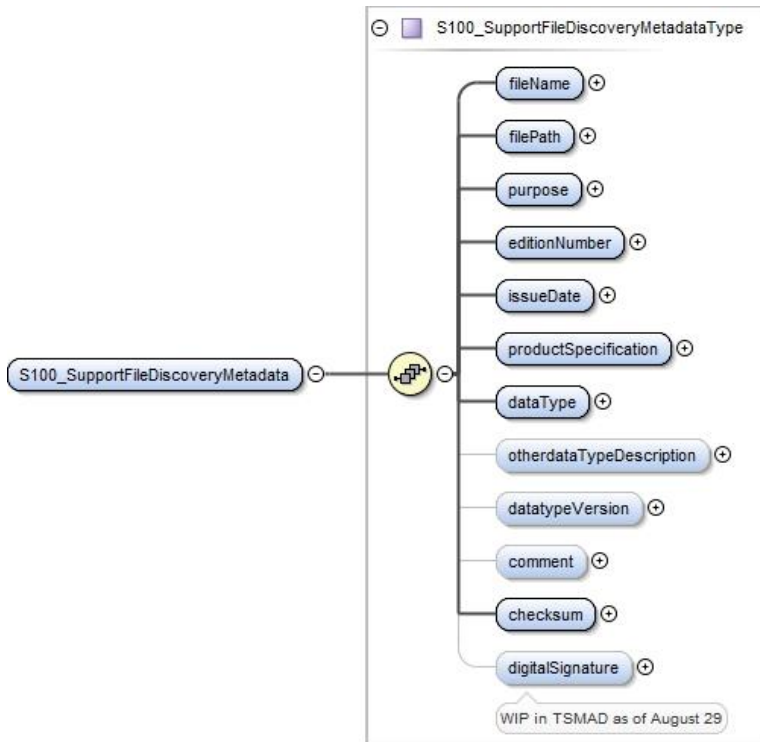
Figure 2. Dataset discovery metadata

S100_DatasetDiscoveryMetadataType

- metadataFileIdentifier
  - Required if dataset metadata is in a separate file
- metadataPointofContact
- metadataDateStamp
- metadataLanguage
  - Metadata must be in English
- fileName
  - Dataset file name
- filePath
  - Full path from the exchange set root directory, excluding the fileName which is given in the fileName attribute
- description
  - Short description of the area covered by the dataset, e.g., a harbour or port name, between two named locations, etc.
- dataProtection
  - Encrypted or unencrypted.
- protectionScheme
  - E.g., S-63
- classification
- purpose
  - Type MD_Identification>purpose not accepted by TSMAD23
- specificUsage
- editionNumber
- updateNumber
- updateApplicationDate
- issueDate
- productSpecification
- producingAgency
- optimumDisplayScale
  - Note: making this mandatory is problematic for SI and non-ENC datasets; should be optional in S-100 and made mandatory...
- maximumDisplayScale
  - Note: making this mandatory is problematic for SI and non-ENC datasets; should be optional in S-100 and made mandatory...
- minimumDisplayScale
  - Note: making this mandatory is problematic for SI and non-ENC datasets; should be optional in S-100 and made mandatory...
- horizontalDatumReference
  - TSMAD23 + S101 draft; should be enumeration?
- horizontalDatumValue
  - TSMAD23 + S101 draft
- verticalDatum
  - TSMAD23 + S101 draft; mandatory may be problematic for non-ENC domains
- soundingDatum
  - TSMAD23 + S101 draft; mandatory may be problematic for non-ENC domains
- dataType
- otherdataTypeDescription
- datatypeVersion (0..∞)
- dataCoverage
  - Provides information about data coverage within the dataset
- comment
- checksum
- layerId
- digitalSignature
  - WIP. To do: find out why TSMAD currently has support file digitally signed but not dataset file.

S100_DatasetDiscoveryMetadata
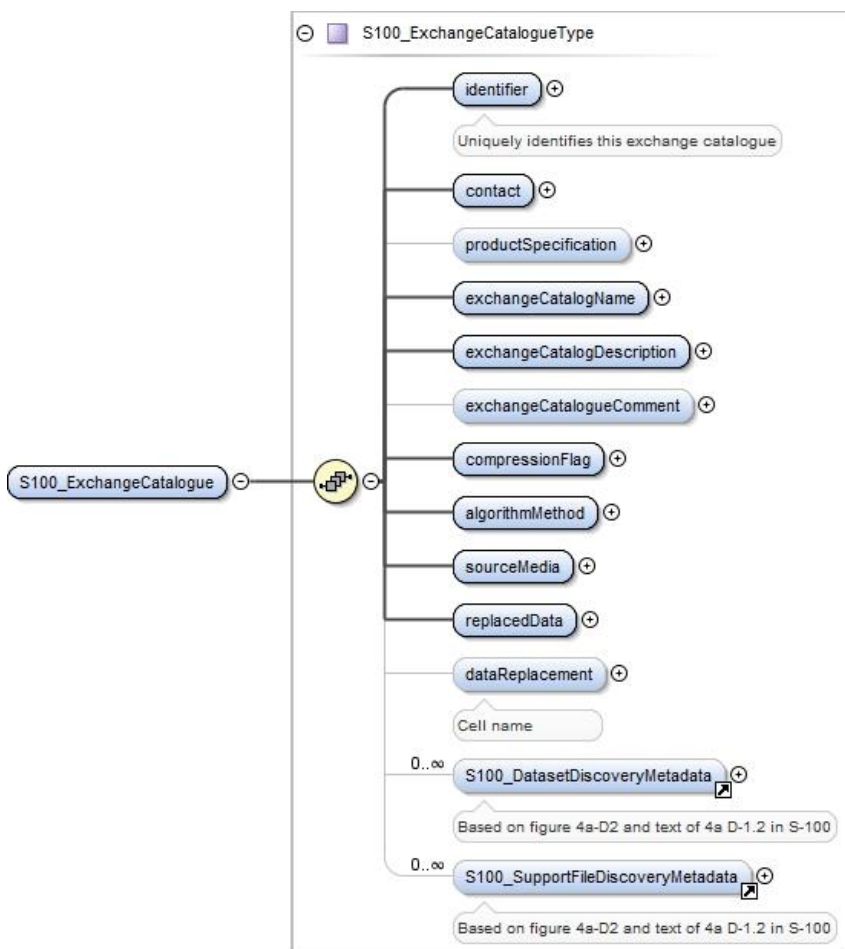
Figure 3. Support file discovery metadata

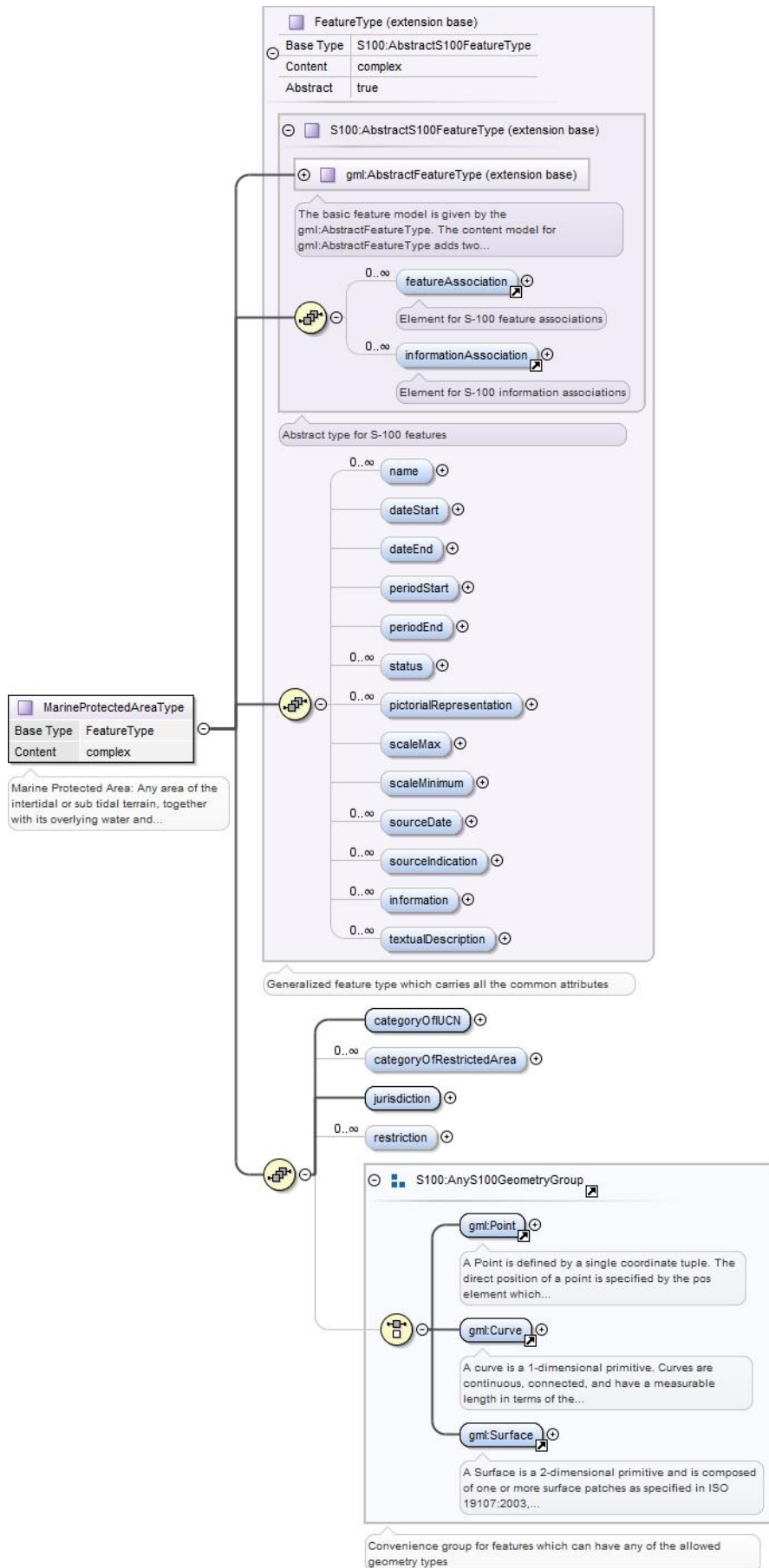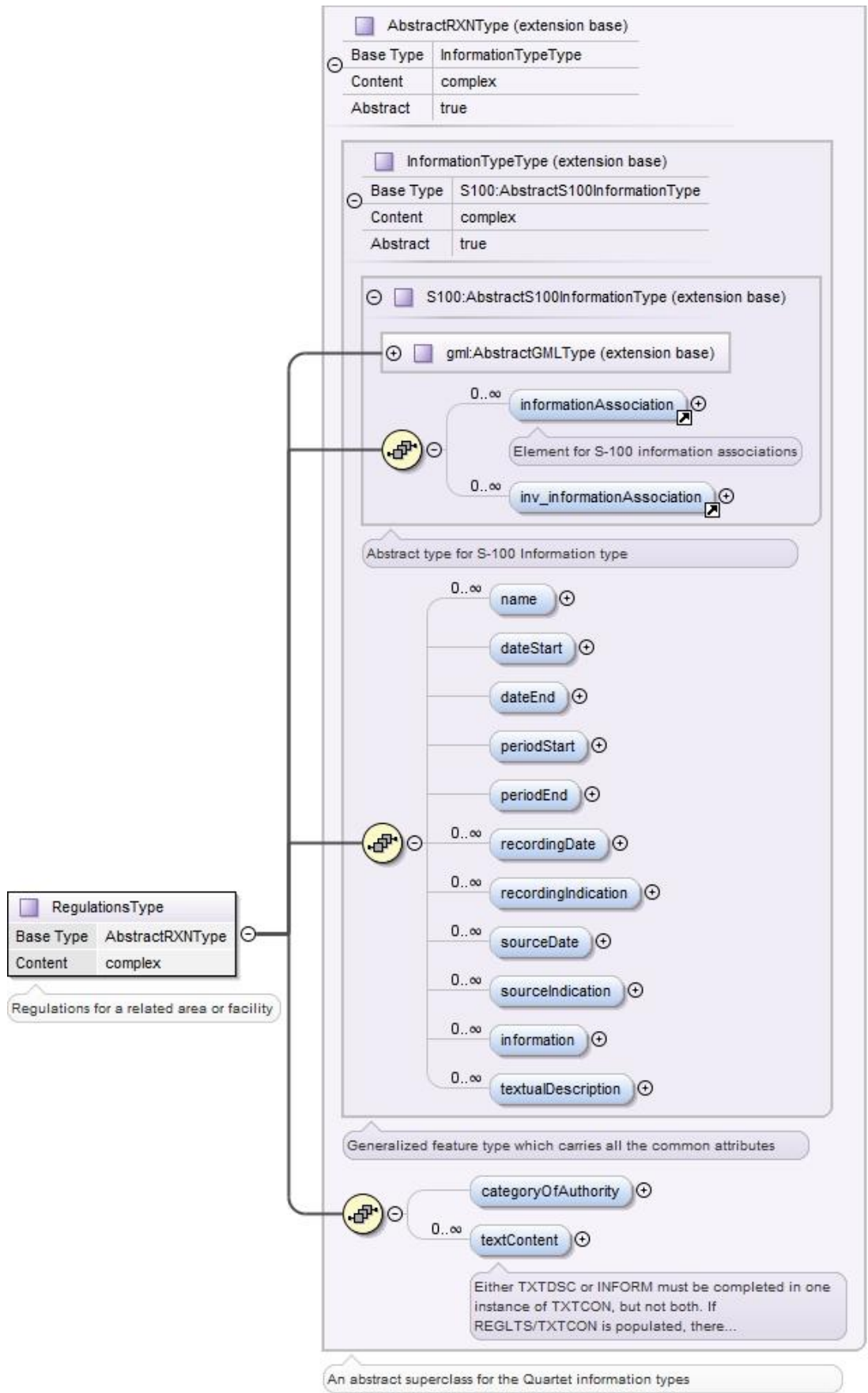**Figure 4. Exchange catalogue**

Figure 5. Example feature type (Marine Protected Area)

Figure 6. Example Information class (Regulations)