

Paper for Consideration by NIPWG

Creation and storage of S-127 for US waters

Submitted by:	University of New Hampshire
Executive Summary:	Description of the creation process of S-127 (pilot services only) and feedback.
Related Documents:	S-127 Product Specification, NIPWG_Letter_2018_07
Related Projects:	S-100, S-127, MRN

Introduction / Background

The development of the S-127 (Marine Traffic Management) S-100 compliant product specification was initiated by HSSC5/5 in 2013. Based on HSSC Action item 9/25, IHO Program of Work 2017 as approved by IHO Member States and Task 2.6.2 (Ref: IHO CL 65/2016 dated 9 December 2016), the development of the S-127 Product Specification was contracted out to Portolan Sciences LLC. Version 1.0.0 of the product specification was delivered as scheduled by the end of November 2018 (Work Plan section F, HSSC 10 Action item 10/37). In Dec 2018 all files were uploaded ready for download and testing by the NIPWG. It is expected that the S-127 Ed. 2.0.0 would be submitted to IHO MS approval in 2021 at the earliest. Starting in 2019 UNH began the process to create data for the Pilot Services portion of the S-127 from the Coast Pilot.

The release of the first version of the S-127 allowed UNH to use a pre-existing approved data model to test their concept of nautical textual feature-centric data storage and confirm the results as a GML-based NIPWG product. This exercise was done in the hopes others would also be able to model their own data storage systems in a similar fashion allowing them to begin the transition to storing their data in a more versatile and vector-based way, as were the requests from the stakeholders forum at NIPWG6 in Germany, 2019.

Analysis/Discussion

GML product – manual generation

The first step was to generate the actual GML file from the Coast Pilot (only part of book 3 was used for testing). All the information was manually selected and collated from the Coast Pilot that related to Pilot Services. Using the Application Schema from Appendix B of the S-127 UML as a visual guide, the feature and information types were then identified and color-coded within the associated text. Once all textual information had a colour assigned to it, they were then manually inserted into the GML file to build a well-formed and valid S-127 product. Once this product was built it was sent to Korea for review and returned with a stamp of approval.

The following data model components were used in this exercise:

- Feature Type
 - Pilotage District (1)
 - Pilot Boarding Place (7)
 - Pilot Service (6)
- Information Type
 - Authority (7)
 - Contact Details (13)
 - Service Hours (2)
 - Applicability (8)
 - Piloted Vessel (4)
 - Pilot Vessel (4)
 - Permissions (8)
 - Piloted Vessel (4)
 - Pilot Vessel (4)

The following issues were noted during the encoding process and submitted as comments for Edition 2.X of S-127:

Note: FOR REASONS OF ECONOMY, DELEGATES ARE KINDLY REQUESTED TO BRING THEIR OWN COPIES OF THE DOCUMENTS TO THE MEETING

1. Pilot Boarding Place

- a. The destination is a text field -> should be a link to an identifier.
It is defined as the place or "general direction" to which a vessel is going or directed. See below for reasons to separate destination and general direction into two different attributes.
 - i. this would be the perfect place to link to a feature MRN to maintain relationships of textual nautical information to the ENC.
 - ii. It would avoid the need to have to do a textual search on something that could easily be misspelled or misnamed.
 - iii. It would be useful to be able to collate all items linked to the same item for auto generated services / destinations lists...allowing the portrayal and filtration of the services to be more enhanced/dynamic.
 - iv. Additional Attribute suggestions:
It is NOTED from the NIPWG wiki that text string could include phrases such as "This pilot boarding place is for inbound vessels" or that mention "The north-west" or "upriver"...HOWEVER...these are really vessel directions (general or cardinal) and the applicability of such travelling vessels to the boarding place (all terms boil down to simply "inbound" or "outbound" traffic, which should be an enumeration). It is recommended to give this information its own attribute name "applicable vessel direction". Encoding them as such would allow for easy computer filtration of boarding places that apply to the inquiring vessel.
- b. The destination should have multiplicity of [0..*]
 - i. In the US and other countries there are multiple destinations for some Pilot Boarding Places.
 - ii. The suggestions with text string is to comma separate the list...it would be preferred for ease of programming as well as fewer mistakes made to have them in their own data elements.
- c. NOTE: It is recommended that the Pilot Boarding Place *not* contain the destination attribute and instead the Pilot Services should (see Pilot Services section for more). However, the recommended changes to the destination outlined here still apply.

2. Pilot Services

- a. Should have destination instead of Pilot Boarding Place
 - i. This would keep the place of boarding responsible only for the specific details related to it.
 - ii. This would eliminate the suggestion to create multiple identical Pilot Boarding Places that have all the same information but different destination values.

3. Pilot Vessel

- a. No place for description / applicability
 - i. There are many descriptions of the Pilot Vessel in the Coast Pilot text. The Applicability Information type is the closest feature related to describing a vessel. In this example, however there was an attempt to use it for Pilot Vessels as well. (Outside of the S-127 proposed use, yet still managing to be a valid S-127 product)
 - ii. This would also be beneficial for reusability of vessel descriptions.
 - iii. NOTE: it was discussed earlier within NIPWG and decided *not* to encode this information since the Pilot Vessel can be replaced with no notice. (The source of this information is said to be the critical component.)
- b. Add more attributes to expand the applicability attributes OR create a way to represent more vessel details...if it was felt to be *value-added*:
 - i. Example text from the Coast Pilot where this would be useful/necessary:
(60) Quoddy Pilots boat MEDRIC II is a 48-foot long black-hulled fishing vessel, and when on station it displays International Code flag H by day and a white light over red light at night. The pilot boat and station monitor VHF-FM channels 14 and 16 and work on channel 65A. The pilot boat cruising and boarding station is about 1 mile eastward of East Quoddy Head at the north end of Campobello Island. Vessels are requested to locate the pilot boarding ladder on the port side at the water's edge and make the best possible lee for pilot boarding. Foul weather and sea conditions may affect boarding time. A 24-hour estimated time of arrival is requested.

- ii. NOTE: this example includes a use case where weather and sea conditions interact with traffic services. It might add value to either tie into S-126 weather descriptions in the physical environment or the S-412 weather overlay as well.

These comments as well as others were compiled into the feedback template and submitted to the NIPWG.

GML to Database

After the generation of the S-127 GML product, UNH began to model relational database tables to accurately reflect the associations and multiplicities. A generic method was adopted and followed that is outlined below.

Enumerations/Code lists – these can be easily stored in a database table and used in an interface to fill in attributes of the associated tables. (Helping to avoid spelling errors, ensuring normalization on terms)

Simple Attributes –

Singles - these become elements of a main feature or complex type table as long as the multiplicity is [0..1].

Multiples - When the multiplicity includes [0..*] then a new one-to-many or many-to-many table needs to be created that can link the infinite list to its owner table.

Complex Attributes – these become tables with single simple attributes and more links to one-to-many or many-to-many tables. Linking to other tables based on main id or MRN.

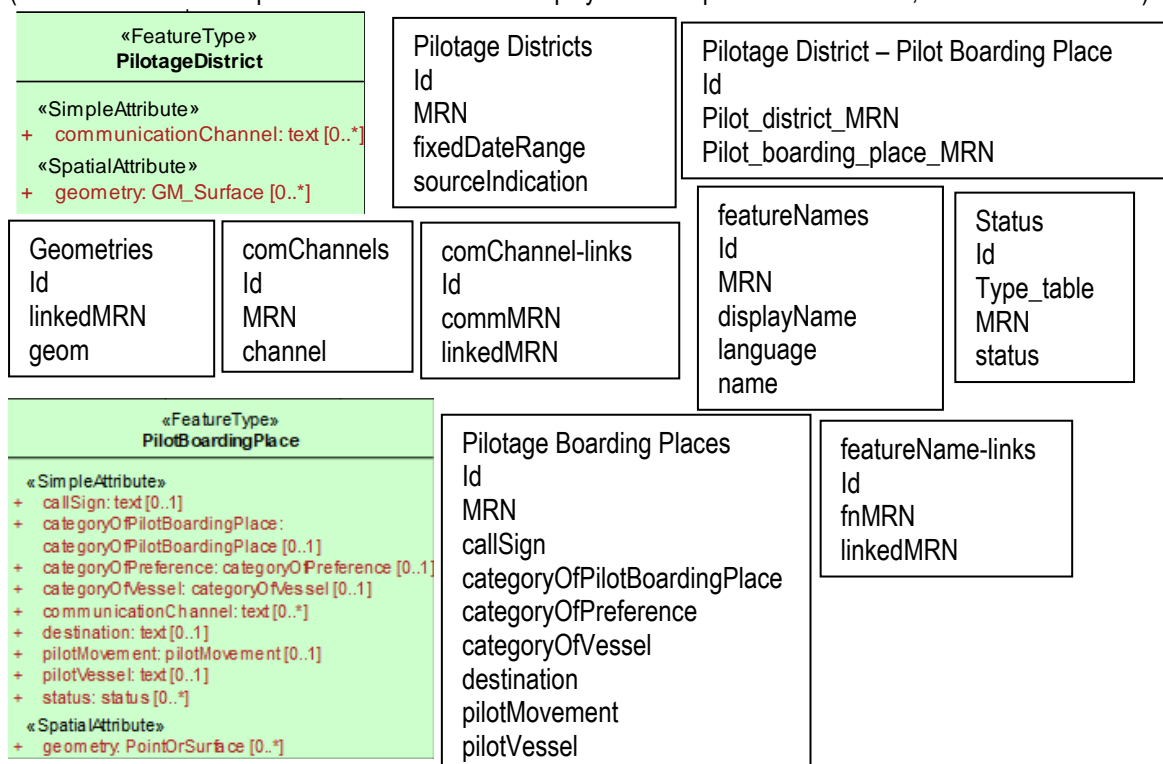
Information Types – these tables follow the same pattern of complex attributes.

Feature Types – these tables will be like the Information Type tables with the addition of a linked geometries table.

Identifiers - The ids for all of these tables followed an MRN style of identification based on sample GML files included with the S-127 v1.0.0 zip files. The method for creating the id's followed this pattern: waterway initials, item acronym, type name, feature name initials. So, for example, the Authority the Delaware Bay Launch would have the following identifier assigned: CP.AUTORI.LAUNCH.DBL (CP – Chesapeake Bay, AUTORI – Authority, LAUNCH – launch, DBL – Delaware Bay Launch). It is up for debate the best way to standardize the creation of these items, however, this was an acceptable starting point.

Example of UML table to Relational Database table

(not all attributes are represented in an effort to simplify the example. GREEN – UML, WHITE – DB Tables)



Note: FOR REASONS OF ECONOMY, DELEGATES ARE KINDLY REQUESTED TO BRING THEIR OWN COPIES OF THE DOCUMENTS TO THE MEETING

GML product – automated generation

The *Oracle* database was for this example and *Python* was leveraged to access the data inside of *Oracle* using *cx_Oracle*. *Anaconda* was used to set up the *Python* environment. A simple script opened up a connection to the *Oracle* database and cursors were used to expedite working with the db connection. *Lxml etree* was used to produce the output XML/GML. *Python* procedures were set up to incrementally create sections of the S-127 product. This allowed for parallel work on the database and the *Python* code as well as made any updates from the iteration process easier to fix. The goal was to use the S-127 GML product that was manually generated as a guide for the correctness of the automated version. Once it was set up the only thing left to do was to continue to populate the tables in the database. This is still in development, however, the sections that have been completed are working fine as expected....the remaining database tables and *Python* sections will be completed before the next meeting.

The next steps will be to use the database and or the GML output to generate overlays on the ENC in order to test the portrayal on the chart as well as combined with other relevant data.

Conclusions

This exercise proved that translating the UML data model schema into a relational database schema can be done with some generic methods that all can use. It allows the storage for nautical textual information to be reused in different formats (as well as platform agnostic as per the requests of the stakeholder's forum at NIPWG6) for different domains/purposes. Other products developed by the NIPWG will also benefit from following the principles outlined in the analysis section of this document.

Recommendations

Recommendations, as stated above, have been made to the NIPWG for the next version of S-127 related to Pilot Services. It is also a recommendation that other HO's that do not have a plan for transitioning their publications look to this as a guide and inspiration when they make the decision to move forward to do the work (or hire a contractor out to do the work).

Justification and Impacts

This S-127 dataset is available and can be used. Testing the dataset by others would be helpful. UNH is now in a position to advise others who might consider transitioning to this database approach for storing NIPWG related products. This database structure can be used for implementing other GML based S-100 compliant product specifications.

Action Required of NIPWG

The NIPWG is invited to:

- a. note this paper
- b. provide input if appropriate.