Jeppesen Marine, Inc.
55 Inverness Drive East
Englewood, CO 80112-5498
Tel.: +1.303.328.6547

Updating Publication Datasets in XML

By Raphael Malyankar
Eivind Mong
Jeppesen

Introduction

Encapsulations in an XML format such as Geography Markup Language (GML) or another XML-based format are among those under consideration for nautical publications information in the S-100 framework. This paper discusses issues arising in updating nautical publications datasets encoded in an XML-based format. It attempts to outline the fundamental issues and problems which arise and general strategies and utilities or off-the-shelf solution components which apply to solving the general problems.

Definitions and Abbreviations

GML Geography Markup Language
GRIB GRIdded Binary. A data format used in meteorology.
WFS Web Feature Service

Factors

Management of data streams

The S-57 standard defines an update mechanism and format using the ISO 8211 encapsulation. Adding a different data stream format complicates the system and the processing required of both the hydrographic agencies and OEMs. They also know that anything XML is verbose. There will also be the inertia to overcome in the OEMs to introduce a second or third data input flow into the SENC. Some already add GRIB format information to their systems.

Another consideration is reconciling the diverse needs of the data end user population taken as a whole, and the constraints on producers. There are different data consumer groups who use the same information but whose needs are different, as regards content, update frequency, update delivery formats, delivery mechanisms, etc. Commercial shipping vs. pleasure boating is a case in point. At the production end of the chain, on the other hand, it may be more convenient for a hydrographic office to focus its resources on collection and preparation, e.g., on data quality, timeliness, and the introduction of new kinds of data, and produce the resultant data and updates in a single comprehensive transfer set format and high update frequency. For example, updates might be published on an “as-available” basis. Highly demanding data consumers (those whose need for data is time-critical) might choose to utilize a direct feed, and service organizations digest, re-package and integrate the information into the NP equivalent of a SENC, apply it to current NP datasets at discrete weekly intervals, post-process it into forms suitable for different consumer populations, etc., for consumers whose needs are less time-critical or who need the information in other forms, or integrated with other information.

Update volume and frequency

The current frequency and volume of incremental updates for nautical publications are low both in absolute terms and compared to updates to charted information (in particular, to the status of

Jeppesen Marine, Inc.
55 Inverness Drive East
Englewood, CO 80112-5498
Tel.: +1.303.328.6547

navigation aids). Notices to Mariners are the publication which is issued the most frequently (every week or two weeks). One year's data from BSH indicate that the publications which were updated most frequently were the List of Lights and the Radio Signals. Updates to sailing directions were issued at the rate of 10 to 40 per volume (ratio of pages changed to total pages is approximately 5% to 20%). New editions of sailing directions are generally published at multi-year intervals.

Expectations for keeping publications up to date are likely to change as publications are issued in digital form, with more updates being issued more frequently. Some hydrographic offices may not want to have to reissue datasets each week or month, preferring an update mechanism that only "cuts and pastes" the changed data, while leaving all the rest in place. Others may prefer to do exactly that (i.e., reissue datasets), especially if the size of a datasets is small enough. Yet another possibility is that of issuing updates on an as-available basis (see the discussion in the previous section).

More data and more analysis of the questions of expectations, volume, and frequency of updates are required.

Photographs and diagrams

In comparison to text, the largest contributors to volume are graphics, e.g., photographs, diagrams, and chart blocks. Some hydrographic offices (e.g., UKHO) use technological means to transmit differences in diagrams instead of transmitting the whole diagrams, but even so, new diagrams which are quite large (500KB or more) sometimes need to be transmitted.

It may be possible to introduce a convention that gives users the option to download certain graphics immediately or postpone download of the file to a later time. For example, and by analogy to ECDIS, the categories "base" and "standard" can be introduced for publications information items, with "base" referring to items that are carriage-required and "standard" to all elements in the "base" category plus additional information. As an example, a photograph of a specific harbour entrance may be in the "base" category if it shows a hazard which is difficult to communicate effectively by means of words alone; though in general photographs of harbour entrances may be in the "standard" category but not part of the "base" set. Or the base set might contain a verbal description of a significant landmark, while the standard set contains a photograph as well as the verbal description.

To ensure that carriage requirements are met and navigation safety is not compromised, it will be necessary to: (a) define means in the publications data model for marking individual items (e.g., specific photographs or diagrams), or collections of items, as essential (carriage-required / "base") or optional/recommended, (b) define standards for what should be in the "base" and "standard" categories, and (c) get Port State inspectors' acceptance of the concept that parts of an information product or overlay may be carriage-required and other parts not carriage-required.

Datasets and cells

Nautical publication data is expected to be scale-independent which should simplify the data management task.

Management of Updates

Methods and tools for creating updates

Exporting changes, testing for compliance, and other change management functions are standard features of ENC tool suites; adaptation to S100 will of course be needed, and furthermore, convenience enhancements to the UI may be necessary, e.g., for complex attributes and for attribute values which

Jeppesen Marine, Inc.
55 Inverness Drive East
Englewood, CO 80112-5498
Tel.: +1.303.328.6547

are relatively longer text fragments than in ENCs. Adaptations to export GML or another XML format is another situation where development effort to extend tools may be needed.

At the implementation level, tool support depends on the tools used for maintaining data. Data management (HO-side) systems which use relational-database systems, XML databases, and/or XML content management systems should all be capable of generating updates. Given the current state of the art, all these data management systems will need some software development effort for data management, version control, and generating updates in the proper format. The basic problem of creating XML from relational database or XML databases is easy, the management and revision control aspects, and integrating it with hydrographic databases will take more development effort.

Revision management, file differencing, and applying updates to files are common problems in software development. Utilities for creating and applying updates ("patches") to files are standard software components on general-purpose computers running variants of the Unix and Linux families of operating systems. Versions are available for Windows-based operating systems under open-source and proprietary licenses; if not already installed on ECDIS/ECS computers as part of the operating system, they would need to be added to new versions of the ECDIS/ECS or accompany the update itself (which is practical only for updates shipped on physical media). Again, some development effort to integrate and adapt the basic building blocks will be needed.

We have not yet found any tools specifically intended for updating GML files.

Strategies for applying updates

Strategies applicable for ENC data can be used for nautical publications information in XML:

Format of update

The GML standard [GML] does not itself define an update mechanism, but the Web Feature Service (WFS) standard [WFS] (which is another OpenGIS specification) does define an XML-based update format which can be used for updating GML. WFS is intended to be used for Web services but there is an XML schema which defines the format of the update "payload". As an aside, this suggests that updates can be distributed via Web servers which implement WFS. Implementation of such servers and the necessary adaptations for S-100 format data will need a certain amount of software development.

Delivery

Delivery can be via physical media or electronic means and utilize channels already in use for ENC data delivery. Adaptations of delivery channels and business processes may be needed due to the possibility of optional components (see the discussion of "base" and "standard" components earlier in this paper), the possibility of "jumbo updates" (e.g., an update containing a new large graphic file), multiple product types (e.g, MPA and pilotage data products) over the same channel, and management of complex subscriptions.

Conclusion

This paper has attempted to identify the issue and summarize the current state of affairs in updating of XML information as it pertains to nautical publications information. More analysis of the expectations for digital delivery of updates to digital NP information will be required.

The basic building blocks for managing updates already exist, but development effort will be needed for supporting tools (either new software, or extensions to existing hydro office tools).
The same is true at the other end of the chain. OEMs will need to integrate existing relatively low-level software utilities into a robust update application and management scheme at the user end.

Jeppesen Marine, Inc.
55 Inverness Drive East
Englewood, CO 80112-5498
Tel.: +1.303.328.6547

Recommendations

1. SNPWG should consider whether (S100-style) product specifications for nautical publications information should define “base” and “standard” information content within each product.
2. SNPWG should consider defining a model that is capable of providing the same graphical information at different resolutions, and the consequences of such differing resolutions on applications and on navigation safety.
3. Efforts should be made to integrate graphics formats which lead to smaller volumes than TIFF images.
4. Ongoing developments in software and standards for XML compression and efficient XML should be monitored.
5. More analysis of expectations and constraints on updates of digital data in particular is needed.

Acknowledgments

Much of this paper is based on discussions with Mr. Acland (UKHO). We thank BSH and Capt. Schröder-Förstenberg for the information about update frequency. The opinions, conclusions and recommendations in this paper are those of the authors and not necessarily of Mr. Acland, the UKHO, or BSH.

References

[GML] OpenGIS Geography Markup Language (GML) Encoding Standard, OGC 07-036, v. 3.2.1, ed. C. Portele, Open Geospatial Consortium Inc., 2007. Also ISO 19136:2007 Geographic Information – Geography Markup Language, International Organization for Standardization, Geneva, Switzerland.

[WFS] OpenGIS Web Feature Service 2.0 Interface Standard, OGC 09-025r1, ed. P. A. Vretanos, Open Geospatial Consortium Inc., 2010. Also ISO 19142, International Organization for Standardization, Geneva, Switzerland.

[S100] Universal Hydrographic Data Model, IHO Publication S-100, Edition 1.0.0, International Hydrographic Bureau, Monaco, January 2010.