

On Source Information in S-1xx Exchange Sets

Paper for consideration by SNPWG15

Submitted by: Jeppesen

Executive Summary

Source information is an important indicator of data quality for nautical publications and Nautical Publications information modeling. This paper outlines the case for source information attributes in nautical publications information encoded in S-100 compliant formats and describes the source information model proposed for nautical publications information datasets in the S-100 family.

1. Introduction

DQWG developed a model of data quality for S-101 in 2011-12. SNPWG discussed data quality for nautical publications information at length before and at SNPWG 14, and part of this discussion concerned the adaptation of the DQWG model for NPUBS information. The DQWG model was subsequently adapted for SNPWG's Marine Protected Areas dataset. TSMAD subsequently made some changes to classes and attributes in the DQWG proposal for the purposes of ENC information (S-101). Separately, TSMAD originally considered defining a complex attribute to replace S-57's source indication, but as of the writing of this paper the source indication attribute is unlikely to become part of S-101, except that attributes for survey date and report date are being discussed and likely to be included.

While source information may be less important as a component of the data which gets to the ENC end-user, for nautical publications source information is an important indicator of the reliability and quality of information and is often passed on to the mariner in the form of mariners' notes, annotations or remarks associated with other information.

2. Background - source information in ISO standards

The data quality overview elements defined in ISO 19113 include the temporal accuracy of information. This is further divided into 3 sub-elements as follows:

ISO sub-element	definition
accuracy of a time measurement	correctness of the temporal references of an item (reporting of error in time measurement)
temporal consistency	correctness of ordered events or sequences, if reported
temporal validity	validity of data with respect to time

ISO 19115 defines DQ_TemporalValidity as the validity of data with respect to time and this is included in S-100.

Also, ISO metadata includes **lineage**, defined as the history of a dataset and the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form. Lineage may contain two unique components:

- source information: The parentage of a dataset,
- process step or history: A record of events or transformations in the life of a dataset, including the process used to maintain the dataset whether continuous or periodic, and the lead time.

3. The case for defining a source-of-information model

For ENC's passing on source information to the end-user (the mariner) is not as important, because quality assessments are nearly all done as part of the production process and the results provided to the end-user (mariner).

- Much of the data in ENC's is numeric (coordinates and depth/height values). The quality of the data is important rather than the source, and quality assessments are communicated to the mariner in the form of confidence parameters (S-100 data quality model) or, in S-57, categorical attributes (CATZOC, QUAPOS) and accuracy estimates (POSACC, SOUACC).
- Data quality can be objectively evaluated (because the data is numeric and objectively observed in surveys).
- Given all the above, in a small number of features there may still be quality notes such as "reported" annotation associated with a significant feature which has not been surveyed, or reported dates (for unverified obstructions, and for construction).

Nautical publications, on the other hand, are supposed to contain information which cannot be easily represented in charts. Often this is non-numeric data. Quantification of quality assessments is more difficult especially as it pertains to individual chunks of information, and passing on source information to the end user is therefore an important part of indicating data quality. The factors in the assessment of publications data quality are described below.

1. Age may be a factor for some kinds of information. Meta-information such as the date of the publication is often used to evaluate reliability. If the age of specific information matters, the information may be qualified by a note saying when the information was last obtained/confirmed. Date is sometimes included in the text. Statements like "Development in progress (2008)" might suggest that the new facility is complete and already in service in 2011.
2. Regulatory changes may change the information – providing the source allows the user to check most recent information.
3. Publications may remark on or mention the presence of major navigation aids, landmarks, radar reflectors, etc. Whether these are correctly represented is theoretically a factor in the quality of NPUBS information. Sometimes there are discrepancies between the ENC and the publication and the ENC is not necessarily the authoritative source.
4. For information obtained from port authorities or extracted from port information guides, mentioning the source may be a useful indicator.
5. Sometimes mariners' reports are the source of information. Questions arise when information provided by mariners is used, such as: How is it verified? How reliable is it? How does an HO indicate its reliability? UKHO rule of thumb is that the word of a qualified mariner is not normally disputed, particular if he or she has taken the trouble to make the report. However in all cases official support is sought for the initial report. If it arrives, it can be used; if not, a decision has to be made based on the merit of the report and consequences of not publishing. If the information cannot be verified before publication it will be noted as "reported" and usually qualified by a date.
6. Regulations are issued by many authorities. There may be a question about which should prevail. The authority issuing the regulation should be considered as an indicator (perhaps given in the publication) of the reliability.
7. The age of information is also a factor in determining whether it should be reviewed in any particular maintenance cycle. Will data producers find it more convenient to maintain age (i.e., source date) separately from the published dataset, or would it be a better idea to include it in the data set? Again, note that the utility of providing age may be different for ENC and non-ENC information.

4. Source information model

The proposal is to define a complex attribute to replace S-57 SORIND (Source indication), with sub-attributes shown in the figure below and defined in the table which follows it.

Figure 1 illustrates the complex attribute *sourceIndication* and its sub-attributes, and their use in the data quality model for nautical publications. The source information complex attribute is bound to the *QualityOfNPInformation* geographic feature class, which in turn can be qualified by its “*hierarchyLevel*” attribute to apply to various levels of data. (The details of *hierarchyLevel* as applicable to S-100 data will be addressed separately as part of the NPUBS data quality model.)

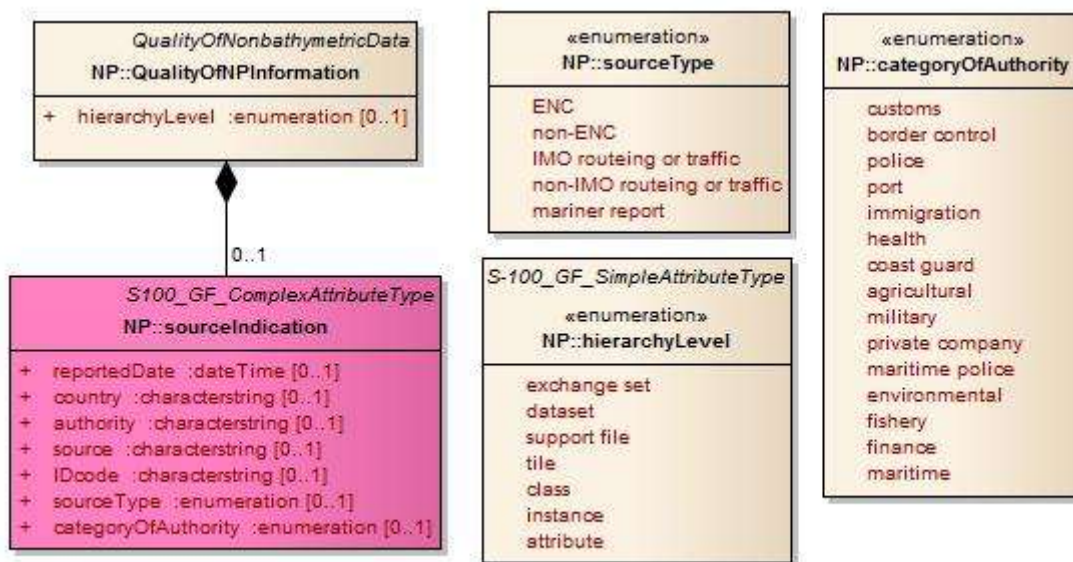


Figure 1. Structure and use of source indication and its sub-attributes

Sub-attribute	Definition and Domain	Notes
country	The name of a nation characterString	SORIND in S-57 has a <i>country</i> sub-field for the 2-letter ISO 3166 country code. The proposal is to use SNPWG attribute “country” already defined on the SNPWG Wiki which has format characterString, for the name of a country.
authority	The name of a governmental or other organization with the power to issue or enforce laws or regulations. characterString	SORIND in S-57 has an <i>authority</i> sub-field, which is intended to specify an ENC producer. This is replaced by an attribute which can describe a wider scope of authorities, most of which will not be ENC producers. In the context of <i>sourceIndication</i> it identifies the authority responsible for the information, e.g., by originating or officially publishing the information
source	A description of the source of the information in question, e.g., “graphic”, “report”, CFR, Coast Pilot, SeeSchiffO. characterString	S-57 defined <i>source</i> as a sub-field of SORIND, with suggested values “graph” for graphic and “reprt” for report (e.g., wreck report). The proposal is to extend the definition. The value should be the name of the publication (or a description, e.g., “port guide”) instead of the name of the authority.
IDCode (or citation)	Identifying code of the source, e.g., ID Code of paper chart characterString	Sub-field of S-57 SORIND. Retain. Consider replacement with the new “citation” attribute being developed by TSMAD.

reported date	date of report	New. Proposed by TSMAD. Replaces sourceDate (SORDAT in S-57). Discuss with TSMAD changing type to allow “reduced” representations, e.g., year only.
	dateTime	
source type	the type of source	New. This enumeration can be regarded as a refinement of the “reported” annotation currently used in ENCs. The categories are defined from the perspective of helping the mariner assess the information quality.
	ENC non-ENC IMO routing or traffic non-IMO routing or traffic mariner report	
category of authority	(SNPWG CATAUT from Wiki)	New. Describes the type of authority identified in the “authority” sub-attribute.
	Enumeration. Categories already defined by SNPWG.	

Justifications for individual sub-attributes:

- Attribute *reportedDate* addresses the “age” factors (1, 5, and 7) in the list in Section 3.
- Attributes *authority*, *source*, *category of authority*, and *source type* together address factors 2-6 in the list in Section 3.
- Attribute *ID Code* addresses factors 2 and 3 in the list in Section 3.
- Using the sub-attribute *country* addresses factors 2 and 6 in the list.

Guideline for data capture and encoding: The general guideline is that source information should be provided in the dataset only where necessary, for example, if it helps the mariner to determine the quality¹ of information. Product specifications should provide detailed guidelines as appropriate.

5. Action requested

SNPWG is invited to:

- Extend the NPUBS data model with the proposed complex attribute *sourceIndication*, and its sub-attributes.
- Include the proposed attributes *reportedDate* and *IDCode/citation* in current efforts to reconcile the NPUBS and S-101 data models.

6. References

- [ISO19113] ISO 19113:2002, Geographic Information – Quality Principles.
 [ISO 19114] ISO 19114, Geographic Information – Data Quality Procedures.
 [ISO19115] ISO 19115:2003, Geographic Information – Metadata. Corrected by Technical Corrigendum 1.
 [ISO19138] ISO 19138:2006, Geographic Information – Data Quality Measures.
 [ISO19157] ISO19157, Geographic Information - Data Quality.
 [S100] S-100 – Universal Hydrographic Data Model, part 4c. Ed. 1.0.0, January 2010.
 [SNPWG14-9] Data Quality for Nautical Publications. Paper at SNPWG14.
 [TSMAD23-4.5.13] S-101 Data Quality. Paper for consideration by TSMAD 23, January 2012. See also TSMAD23-4.5.13A, Annex A - Draft of S-101 Chapter 6, Data Quality; TSMAD23-4.5.13B, Annex B: Proposal for Data Quality additions to the Hydro Register. URL: http://www.iho.int/mtg_docs/com_wg/TSMAD/TSMAD23/TSMAD23Docs.htm

¹ “Quality” is determined by reference to the ISO 19113 elements and sub-elements for data quality, taking into account the differences between publications and ENCs. SNPWG 14-9 lists them in more detail, as do ISO 19113 and ISO 19115.