MSDIWG10-11.2B

Paper for Consideration by Marine Spatial Data Infrastructures WG

Response to paper DQWG14-10J

Submitted by:	DQWG Chair
Executive Summary:	Reponses op paper DQWG14-10J from MSDIWG: Data Integrity, marine boundaries from an MSDI perspective.
Related Documents:	ISO-TC211_N3521_ISO_FDIS_19157
Related Projects:	Development of a Minimum Standard for Data Validation

Introduction / Background

MSDIWG-Chair submitted a paper to DQWG14 (DQWG14-10J), Informative paper describing data security, data integrity, marine boundaries from a MSDI perspective. The paper was discussed at DQWG14. This paper is a written response, informative for MSDIWG10.

Analysis/Discussion

IRCC Assistant Director presented the MSDI paper at DQWG14. An example was demonstrated with a UNCLOS maritime limits and boundaries where the integrity of the result was questionable. This was discussed at DQWG14 and the recommendations from MDSI were:

The DQWG should be made aware about data integrity and using marine boundaries from a MSDI perspective is a good use case to illustrate the issues. The potential impacts and the need to provide means and mechanisms to protect the data integrity and assure the end user of the provenance of the data they are receiving should be considered further.

The DQWG was asked to note the report, discuss it and take any action as deemed necessary.

After discussion, the DQWG came to a common recommendation for the MSDIWG (see recommendations).

Post meeting, the paper was further analysed by DQWG-Chair to see if there are already existing activities delegated by the HSSC to the DQWG to assist the MSDIWG to consider the impact of incomplete, corrupted or wrongly attributed data. HSSC has tasked DQWG to develop a Minimum Standard for Data Validation in 2020. Further inspection of ISO-19157 demonstrated an ordering in data quality evaluation that may assist the MSDIWG.

Ordering in data quality evaluation:

When evaluating geographic data, one individual error may influence several data quality elements. For measurements resulting in rates (e.g. percentage rates of aspects of completeness) the use of proper denominators describing the total population is important, see figure below:

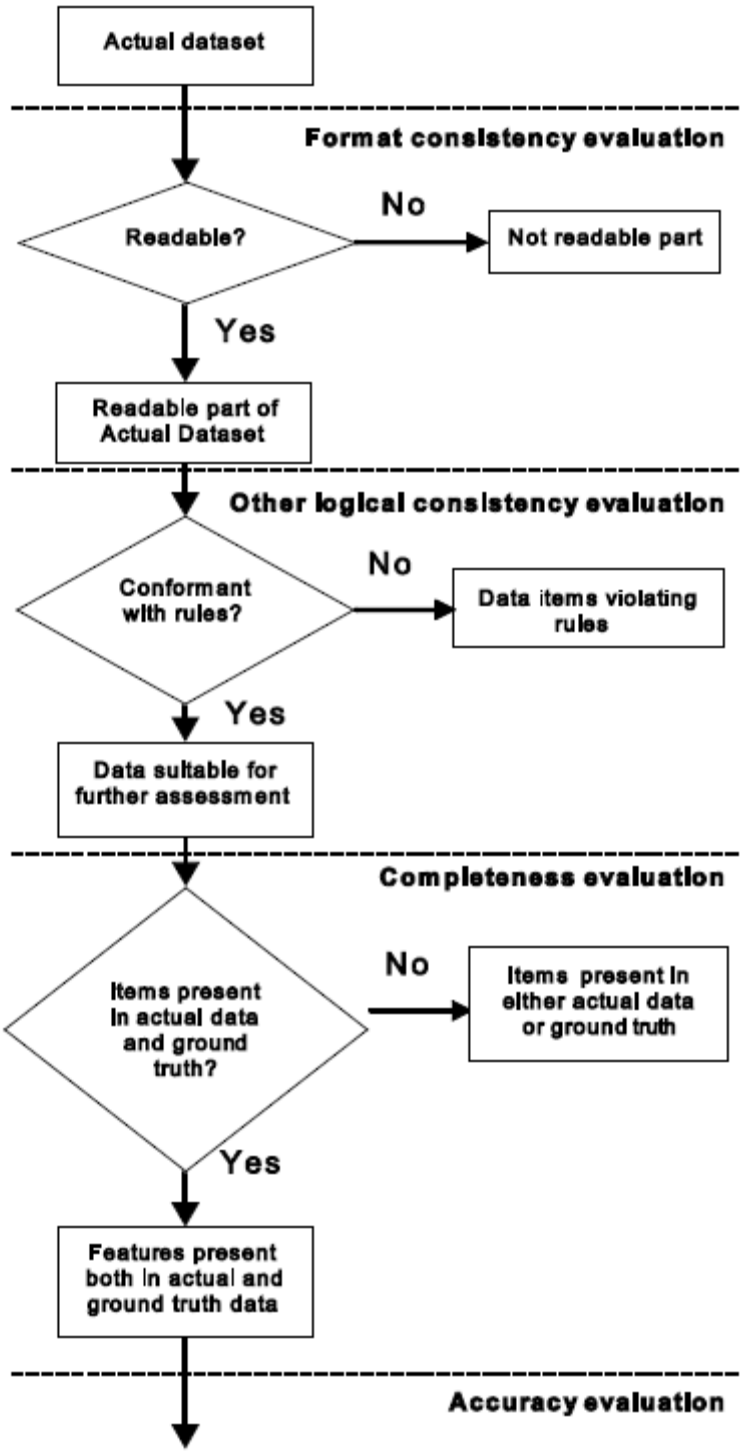

Figure 1: Ordering in data quality evaluation

When evaluating data quality, the usual ordering is:

1) Logical consistency/Format consistency: The very first to be evaluated is the readability (or interpretability) of the data to decide whether it is possible to decode/read/understand the data or not. Not interpretable data should be reported and ignored in the further evaluation. The result of the format consistency should describe which parts of the data are not readable.
2) Logical consistency: Decide if the rules set up for the dataset are followed. Parts of the dataset not conforming to the rules should be ignored in the further evaluation.
3) Completeness: The next step in the evaluation is the feature existence aspect covered by completeness. To evaluate this, the features in the actual dataset and the ground truth data are compared, and commissions and omissions reported.
4) Accuracy (positional, thematic and temporal aspects): The last step in the evaluation covers the accuracy aspect, measuring the deviation between actual and ground truth feature properties. These measurements can be based only on parts of the dataset present in both the actual dataset and the universe of discourse.

In the example of the maritime boundary, the temporal accuracy is unclear. An “old” boundary and a recent boundary are shown in one image. Without quality information about the “old” boundary (originator, source, date of creation, coordinate reference system used) it will be very difficult to draw conclusions in respect to the recent boundary.

Conclusions

DQWG is tasked to develop and maintain a data quality checklist for product specification developers. Once completed, the focus will shift towards developing a Minimum Standard for Data Validation. MSDIWG may benefit from the Data Quality Checklist (S-97 part C – Data Quality) as well as from the Data Validation Standard.

Recommendations

DQWG recommends that the process of SENC distribution that entered into force for certified (Norske Veritas procedure) value-added resellers for S-57 data should be considered (or re-considered) within the S-100 framework for S-100 based products.

Justification and Impacts

None at this time.

Action Required of Marine Spatial Data Infrastructures Working Group

The MSDI is invited to:

a. note this report;.