

**DQWG 4th Meeting  
Helsinki, 14 - 17 June, 2011**

**Paper for Consideration by DQWG**

**S-101 Data Quality Section**

<b>Submitted by:</b>	S-101 Work Item Leader
<b>Executive Summary:</b>	This paper seeks summarizes preliminary work done by DQWG and TSMAD on data quality for S-101
<b>Related Documents:</b>	S-101 Product Specification
<b>Related Projects:</b>	N/A

**Introduction / Background**

1. At TSMAD 22 a brief discussion was held regarding the way forward for the data quality section in S-101. TSMAD noted that the DQWG has issued a comprehensive survey regarding data quality and hope that that will allow for a refinement in current data quality indicators that are used in S-57 for S-101.

**Analysis/Discussion**

2. As part of the discussion the following data quality definitions were discussed.
3. Data Quality is normally defined as:
  - a. Data Quality refers to the degree of excellence exhibited by the data in relation to the portrayal of the actual scenario.
  - b. The state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use – Government of British Columbia.
  - c. RTCA/DO-200A defines data quality by the following criteria:
    - **Accuracy** – The degree of conformance between the estimated or measured value and its true value.
    - **Resolution** – The smallest difference between two adjacent values that can be represented in a data storage, display, or transfer system.
    - **Assurance Level** – Quantifiable value that communicates clearly what level of trust a user can place on the assessed data.
    - **Traceability** – The degree to which a system or a data product can provide a record of the changes made to that product and thereby enable an audit trail to be followed from the end-user to the data originator.
    - **Timeliness** – The degree of confidence that the data is applicable to the period of its intended use.
    - **Completeness** – The degree of confidence that all of the data, needed to support the intended use, has been provided.
    - **Format** – The process of translating, arranging, packaging, and compressing a selected set of data for distribution to a specific target system. A result of this process is a data structure that fulfills the characteristics of data quality.
4. Notably – TSMAD felt that as part of data quality that the indicator for Assurance and the timespan associated with the assurance level should also be considered.
5. In addition, there has been discussion regarding the visualization of M\_QUAL and the usefulness of M\_QUAL on small scale ENCs, in fact, it is generally accepted that with the change in definitions to S-57 in supplement 2, M\_QUAL covers bathymetry fairly well. However, there has not been substantial discussion as to what other data quality indicators are needed for an S-101 ENC.

6. Currently, S-101 has utilized the S-57 Use of the Object Catalogue sections on Data Quality as a placeholder for S-101. This includes guidance on M\_SREL and M\_QUAL. According to S-100 the data quality section for a product specification shall use the following guidance:

The data product specification shall identify the data quality requirements for each scope within the data product in accordance with S-100 Part 3. For every data quality scope it is necessary to list all the data quality elements and data quality sub-elements defined in S-100 Part 3, even if only to state that a specific data quality element or data quality sub-element is not applicable for this data quality scope.

Each product specification shall describe the data quality requirements. One aspect is the “data quality overview element” which should allow a user to decide whether this dataset is the one they want. The other aspect is the metadata allowed for specific feature collections, features and attributes within the dataset.

The data quality overview element should include at least the intended purpose and statement of quality or lineage. Other data quality elements cover: completeness, logical consistency, positional accuracy, temporal accuracy, thematic accuracy, and anything specifically required for the product being specified.

The product specification should comment on which of these are to be used and how, including a description of (or reference to) conformance tests. For example, should data only be published if it passes a particular test, or is it allowable to publish the data with a quality statement which indicates non-conformance? The product specification shall describe how each quality element is to be populated, for example, stating the mechanism to reference the quality evaluation procedure, and allowable values for the quality results.

The application schema shall indicate how the data quality elements will be related to the data items, for example whether a particular dataset should have homogeneous quality, or whether quality elements can be related to feature collections, individual feature objects or attributes.

Finally, the encoding description (clause 15) shall indicate how the quality elements will be encoded.

7. In addition, S-100 states that the following must be included in the product specification for data quality. See S-100 Part 4C Annex B (ISO lines 78 and 79)

	Item name	Definition	Obligation	Maximum occurrence	Data type	Domain
1	dataQuality	required level of data quality	M	N	DQ_DataQuality	see ISO 19115
2	role: qualityScope	scope for the quality information	M	1	DPS_ScopeInformation	see Annex D

8. Another factor to consider when defining the scope of data quality is the interoperability between existing S-57 elements and S-101. Hydrographic Offices have only just begun to take M\_QUAL seriously and populate CATZOC with meaningful values. There is concern that this data would be lost in the transition from S-57 to S-101.
9. Lastly, the biggest complaint regards to data quality is that the portrayal on an ECDIS does not provide meaningful information. At the last stakeholders meeting there was some preliminary discussion regarding how to better portray this information.
  - a. Potential Idea's were the following for route monitoring:
    - i. Eliminate the existing portrayal and utilize a source diagram approach in the margins of the monitor that is colour coded red, yellow, green.
    - ii. Have the ECDIS utilize the underlying meta information and provide a colour coded track line (red, yellow, green).

## Conclusions

10. TSMAD would like to note that Data Quality can fall in three different areas of the product specification:
  - a. Discovery Metadata can contain data quality indicators

- b. There can be data quality indicators within the dataset such as M\_QUAL and M\_SREL
- c. Portrayal of Data Quality information

**Action Required of DQWG**

11. Prepare a draft Data Quality Section for S-101 taking into account the guidance given in S-100 (Clause 6 of this paper) and in the S-100 metadata section – specifically Part 4C for the next TSMAD meeting in January 2012
  
12. Propose potential portrayal scenarios of Data Quality information contained in M\_QUAL for the next joint TSMAD/DIPWG meeting in May 2012