Development of a Metadata Quality Index (MQI) for Bathymetric Data Assessment in Marine Spatial Data Infrastructure

K. M. D. Hasara*, K. A. N. C. Karunanayaka

Department of Surveying and Geodesy, Faculty of Geomatics, Sabaragamuwa University of Sri Lanka

***Corresponding Author**

DOI: https://dx.doi.org/10.51244/IJRSI.2026.13020067

Received: 15 February 2026; Accepted: 20 February 2026; Published: 2 March 2026

ABSTRACT

Marine Spatial Data Infrastructure (MSDI) serves as the foundational framework for managing and disseminating marine geospatial data, with bathymetric data being one of its most critical components. However, the usability and reliability of bathymetric datasets within MSDI are fundamentally dependent on the quality and completeness of accompanying metadata. This paper introduces a new Metadata Quality Index (MQI) framework is presented with special focus on bathymetric data evaluation of the MSD. MQI framework measures metadata completeness on eight fundamental categories including general information, identification information, description, extent, accuracy parameters, point of contact, meta-metadata, and processing information. A weighted grading scale (0-100) was created on the basis of the importance of individual parameter in data accuracy and a decision to be made by a user. The framework was applied to 12 bathymetric datasets (1991-2024) from Sri Lanka, covering various technologies including Single-Beam Echo Sounders (SBES), Multi-Beam Echo Sounders (MBES), and modern integrated systems. Findings showed that there were strong temporal patterns in metadata quality with pre-2000 datasets having a mean MQI of 18.4 (Poor), 2000-2010 datasets had a mean MQI of 32.7 (Fair), 2010-2020 datasets had a mean MQI of 71.3 (Good) and post-2020 datasets had a mean MQI of 78.5 (Good). The most significant gaps were found during all periods, with the most significant being the lack of any calibration documentation (0% compliance) and the systematic missing links in uncertainty reporting (8.3% compliance in pre-2010 datasets). The suggested MQI framework is a quantitative, standardized assessment of metadata quality tool that may be used to offer evidence-based prioritization of retrospective documentation tasks and set minimum metadata requirements to integrate MSDI. The study will add to the operationalization of the quality management of MSDI and the FAIR (Findable, Accessible, Interoperable, Reusable) principles of data in the marine scene.

Keywords: Metadata Quality Index, Bathymetric Data, Marine Spatial Data Infrastructure, Data Quality Assessment, FAIR Principles

INTRODUCTION

Marine Spatial Data Infrastructure (MSDI) provide the main framework of managing, distributing, and reusing marine geospatial data that allows the stakeholders to find, access, and use the bathymetric information in a variety of applications such as safety in navigation, marine spatial planning, management of coastal zones, environmental control, and development of offshore resources (IHO C-17, 2023; Foglini and Grande, 2023). One of the most basic and most commonly used data themes in MSDI is bathymetric data which is a map of the underwater surface of oceans and water bodies (Kearns and Breman, 2010).

The quality, completeness and standardization of metadata associated with bathymetric data are the key to the utility and reliability of the latter in MSDI. Metadata, commonly called the data about data, gives us the needed contextual data that helps users determine whether the datasets are suitable to a particular application (ISO 19115-1, 2014).

In the case of bathymetric data, the most important metadata items are acquisition method, sensor description, position uncertainty, acquisition calibration, processing history, date/time, and spatial reference frame (IHO S-102, 2022).

The Metadata Challenge in MSDI

Although the need to have metadata has been acknowledged, hydrographic offices across the globe are under serious challenges to ensure that they have comprehensive metadata on the holdings of bathymetric data. These difficulties are occasioned by several factors:

Historical Legacy: Bathymetric data collections of various decades are accessible in many hydrographic offices, and technologies were used to gather bathymetric data, starting with lead lines and sounding poles (prior to 1970s) and analog and digital single-beam echo sounders (1970s-1990s) and ending with modern multibeam systems and satellite-derived bathymetry (2000s-present) (Li et al., 2023; Mayer, 2006). The metadata documentation practices have undergone a significant transformation in the past, and structured metadata standards did not develop until the late 1990s and the beginning of the 2000s.

Technological Heterogeneity: technologies used in acquisition create heterogeneous metadata requirements. Multibeam systems need the extensive documentation of motion sensors, sound velocity profiles, and beam patterns, and satellite-derived bathymetry needs the documentation of atmospheric correction, water column properties, and tidal models (IHO B-13, 2024). This non-homogeneity makes it difficult to have standard metadata templates.

Institutional Variations: There are also a lot of differences in metadata practices in the organization, country and even within individual survey teams. Whereas there are surveys that have extensive documentation of the survey with calibration logs, processing reports, and calculations of the uncertainty, some only have minimal information (Hasara et al., 2025).

Format Proliferation Bathymetric data are in many formats, manufactured in a variety of raw format, processed ASCII XYZ, gridded formats (GeoTIFF, BAG, NetCDF), and specialized formats (GSF, XTF), each with varying metadata embedding capabilities (IHO S-100, 2010).

Research Gap and Objectives

Existing methods of metadata evaluation within MSDI are mostly qualitative and subjective as they purport to use the judgment of experts to estimate whether metadata is sufficient or adequate to meet the requirements of the users. Although the metadata standards, including ISO 19115 and IHO S-102, define the type of metadata that should be recorded, they do not offer any quantitative models to evaluate the quality of metadata, or to compare the level of metadata completeness among datasets (ISO 19115-1, 2014; IHO S-102, 2022).

Moreover, the existing literature has dedicated its attention to the accuracy and the uncertainty of bathymetric data (Hare et al., 2011; IHO S-44, 2022), the development of the MSDI framework (Tavra et al., 2017; Racetin et al., 2022), and the classification of MSDI accuracy (Hasara et al., 2025). No study, however, has come up with an operationally and quantitatively declined framework of specially evaluating the quality of metadata in bathymetric data that would be included in the integration of MSDI.

This gap is covered in this paper through:

Creation of Metadata Quality Index (MQI) framework tailored to the bathymetric data evaluation in MSDI settings with weighted scores depending on the criticality of parameter.

Development of minimum metadata requirements at various levels of MSDI integration, separating between discovery-level, evaluation-level and full-reuse-level metadata.

These include verifying the framework using its comprehensive application to 12 bathymetric datasets (19912024) of the Hydrographic Office (NHO), Sri Lanka across a variety of technologies and time scales.

Examination of the temporal patterns of metadata completeness over 40 years, the discovery of the most significant gaps in documentation that need to be addressed.

Giving evidence-based advice on the retrospective metadata creation, the protocol to follow in the future survey documentation, and MSDI discovery interface design.

Significance

The MQI framework provides hydrographic offices with a quantitative, standardized tool for metadata quality assessment, enabling:

- Objective comparison of metadata completeness across datasets
- Evidence-based prioritization of retrospective documentation efforts
- Establishment of minimum metadata thresholds for different MSDI integration levels
- Monitoring of metadata quality improvement over time
- Enhanced user trust through transparent metadata quality indicators

This research contributes to the operationalization of MSDI quality management and directly supports the FAIR (Findable, Accessible, Interoperable, Reusable) data principles in the marine domain (IHO C-17, 2023).

LITERATURE REVIEW

Metadata Standards for Geographic Information

The international organization of standardization (ISO) has come up with an extensive geographic information metadata standards. The ISO 19115-1:2014 Geographic information metadata Part 1: Fundamentals is a schema that defines the description of geographic information and services and includes the description of identification, constraints, data quality, maintenance, spatial representation, reference systems, content, portrayal, distribution, and metadata extension (ISO 19115-1, 2014). The standard specifies more than 400 metadata elements, which are grouped into packages and sections, where core elements are suggested to be used in simple discovery.

The ISO 19115-2:2019 is the extension of the standard to include metadata of imagery and gridded data to meet the particular needs of remote sensing and derived products (ISO 19115-2, 2019). These standards are supplemented by ISO 19157:2013 " Geographic information Data quality" which defines elements of data quality description, such as completeness, logical consistency, positional, temporal and thematic accuracy and usability (ISO 19157, 2013).

IHO Metadata Specifications for Bathymetric Data

Hydrographic and bathymetric data have metadata specifications which are domain-specific metadata specifications, created by the International Hydrographic Organization. The standard of the modern hydrographic framework of data is IHO S-100 "Universal Hydrographic Data Model" which allows interoperability of various product specification (IHO S-100, 2010). S-100 includes the use of ISO 19100 series and registers on feature concepts, portrayal, and metadata.

IHO S-102 "Bathymetric Surface Product Specification" defines specific metadata requirements for gridded bathymetric surfaces, including:

- Discovery Metadata: Identification, citation, abstract, keywords, spatial extent, temporal extent
- Structure Metadata: Grid parameters, coordinate reference systems, vertical datum
- Quality Metadata: Uncertainty measures, completeness, logical consistency, positional accuracy • Acquisition Metadata: Sensor information, processing steps, lineage

Exchange Set Metadata: File format, transfer size, compression (IHO S-102, 2022)

IHO C-17 "Spatial Data Infrastructures: The Marine Dimension" emphasizes the critical role of metadata in MSDI implementation, highlighting that "metadata is the oil that lubricates the SDI machine" and that "without comprehensive metadata, data cannot be discovered, evaluated, or effectively reused" (IHO C-17, 2023, p. 42).

Metadata Quality Assessment Frameworks

Although there is a lot of literature on data quality assessment, there is little research on metadata quality assessment. Bruce and Hillmann (2004) introduced a metadata quality assessment framework that is based on seven criteria completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness, and accessibility. Several researchers have modified this framework to be used in geospatial metadata applications.

Nebert (2004) highlighted that the quality of metadata is directly related to the usability of the spatial data infrastructure, which stated that, poor quality metadata results into invisible data or misinterpreted data or misused data. The research determined metadata completeness to be the most crucial factor in successful implementation of SDI.

In more recent times, the quantitative methods of metadata quality evaluation have been developed by the geographic information scientists. The proposed index of quality of metadata of the spatial data infrastructures of ISO 19115 compliance by Manso-Callejo et al. (2013) was calculated through weighted ranking of the mandatory, conditional and optional elements. Their structure accepted 78 percent of expert evaluations when utilized on Spanish metadata records of SDI.

Bathymetric Data Accuracy Parameters

Understanding bathymetric data accuracy is essential for identifying which metadata parameters are most critical for user decision-making. IHO S-44 (2022) defines accuracy requirements for hydrographic surveys using Total Vertical Uncertainty (TVU) and Total Horizontal Uncertainty (THU):

Total Vertical Uncertainty (TVU):

$$TVU = \sqrt{a^2 + (b \times d)^2}$$

where:

- a = uncertainty independent of depth
- b = coefficient representing depth-dependent uncertainty • d = depth

Total Propagated Uncertainty (TPU):

$$TPU = \sqrt{TVU^2 + THU^2}$$

Hare, Eakins, and Amante (2011) identified multiple sources of uncertainty in bathymetric measurements:

- Platform-related uncertainties (draft, settlement, squat, vessel dynamics)
- Sensor-related uncertainties (sonar, motion sensors, positioning systems)
- Environmental uncertainties (tides, sound speed structure, sea state)
- Integration uncertainties (time synchronization, latency)
- Calibration uncertainties (misalignment angles, biases)

Each of these uncertainty sources requires corresponding metadata documentation to enable proper data evaluation and reuse.

FAIR Data Principles in MSDI

The FAIR data principles-Findable, Accessible, Interoperable, Reusable-have been widely adopted in scientific data management and are explicitly referenced in IHO C-17 (2023) as guiding principles for MSDI implementation:

- Findable: Metadata should be rich, persistent identifiers assigned, and data registered in searchable resources
- Accessible: Metadata should be retrievable using standardized protocols, even if data is restricted
- Interoperable: Metadata should use formal, accessible, shared languages and vocabularies
- Reusable: Metadata should have rich provenance documentation and clear usage licenses

Wilkinson et al. (2016) emphasized that metadata quality is fundamental to all four FAIR principles, noting that "poor metadata renders data effectively unusable regardless of its intrinsic quality."

Research Gap Synthesis

- The literature review reveals several critical gaps addressed by this research:
- No quantitative metadata quality framework exists specifically designed for bathymetric data in MSDI contexts
- Parameter criticality weighting-essential for prioritizing documentation efforts—has not been systematically established for bathymetric metadata
- Minimum metadata thresholds for different MSDI integration levels (discovery, evaluation, reuse) remain undefined
- Temporal analysis of metadata completeness across bathymetric data holdings has not been documented
- Validation studies applying metadata assessment frameworks to real hydrographic office data collections are lacking

METHODOLOGY

Study Area

The study area covers the coast of the Western Sri Lanka and includes the south boundary of the Colombo District (6 o 35: 13. 04) up to the North boundary of Gampaha District (7 o 16: 19. 85). This zone extends to the coastline up to the boundary of the Territorial Waters (TW) of Sri Lanka. The reason as to why this region has been chosen is:

Strategic maritime significance: Sri Lanka has its main business port (Port of Colombo) and some of the most significant fishery harbors that need frequent hydrographic surveys.

Data density: Slender bathymetric surveys are able to give sufficient data to support valid methodology.

Access to data: Systematic archiving of past and present data by the Hydrographic Office.

Data Collection

Twelve bathymetric datasets spanning 1991-2024 were obtained from the Hydrographic Office (NHO), Sri Lanka. Datasets were selected using purposive sampling to ensure representation across:

Time periods: 1990s (n=3), 2000s (n=3), 2010s (n=3), 2020s (n=3)

- Acquisition technologies: Single-Beam Echo Sounders (n=7), Multi-Beam Echo Sounders (n=4), Integrated systems (n=1)
- Survey purposes: Navigation safety, port development, dredging inspection, coastal engineering, resource assessment
- Data formats: Raw survey files, processed XYZ, survey reports, contractor deliverables

Table 1 presents the complete dataset inventory.

Table 1: Bathymetric Datasets from Hydrographic Office, Sri Lanka

Dataset ID	Year	Survey Purpose	Technology	Positioning	Data Format
67	1991	Land reclamation, Mutwal	SBES (unknown)	Not documented	Paper records/raw
112	1995	Port approach survey	SBES (unknown)	Transit (inferred)	Paper/analog
158	1999	Sand boring, North Colombo	SBES (unknown)	Not documented	Paper/digital mixed
198	2004	Colombo Port sounding	SBES (unknown)	Not documented	Digital raw
204	2005	Palliyawatte survey	SBES (unknown)	DGPS (inferred)	Digital raw/XYZ
238	2008	Colombo Port survey	SBES (unknown)	GPS	Digital raw
287	2012	Galle Harbour survey	SBES (ODOM)	DGPS	Raw + metadata
325	2015	Mount Lavinia survey	MBES (Reson)	RTK	Full metadata
342	2017	Mutwal Harbour survey	MBES (Kongsberg)	RTK	Full metadata
363	2021	Dehiwala sandborrow	MBES (Reson)	RTK	Full metadata
401	2022	Colombo Port deepening	MBES (EM2040)	RTK/PPK	Full metadata
CONCMBH001	2023	Negombo site inspection	SBES (EA440)	Stonex RTK	Full metadata

Metadata Parameter Framework

A comprehensive metadata parameter framework was developed based on synthesis of ISO 19115-1, IHO S102, and IHO S-100 requirements, adapted specifically for bathymetric data applications. The framework comprises eight core categories with 72 individual parameters:

Category 1: General Information (5 parameters)

Data Set Name, Description, ID, Published Date, Keywords, MSDI Category

Category 2: Identification Information (8 parameters)

Alternative Title, Project Number, Creation Date, Identifier, Originator, Custodian, Online Resources, Access

Constraints

Category 3: Description (15 parameters)

Primary Purpose, Status, Device Category, Device Name, Device Uncertainty, Positioning Device, Positioning Device Name, Platform Category, Platform Name, Supplementary Instruments, Average Vessel Speed, Acquisition Software, Format, Data Type, Distribution Format

Category 4: Extent (12 parameters)

Time Extent (Begin Date, End Date)

Spatial Extent (Area Name, Area, Corner Coordinates, Horizontal Datum)

Vertical Extent (Minimum Depth, Maximum Depth, Vertical Datum)

Category 5: Accuracy Parameters (21 parameters)

Primary/Secondary Accuracy Level, Survey Specification

Accuracy: TPU, THU, TVU

Resolution: Spatial Resolution (Across/Along Track), Temporal Resolution

Completeness: Coverage, Coverage Percentage, Line Spacing, Data Gaps

Consistency: Cross-dataset Consistency

Timeliness: Data Age, Update Frequency

Reliability: Data Integrity, Error Tracking (Real-time QA/QC)

Precision

Usability: Interoperability, Interoperability Constraints

CATZOC Accuracy Level

Category 6: Point of Contact (6 parameters)

Owner (Name, Position), Division, Organization, Email, Telephone, Address

Category 7: Meta-Metadata (5 parameters)

Published Date, File Identifier, Source, Language, Standards, Contact Point

Category 8: Processing Information (for processed datasets) (6 parameters)

Processing Software, Division, Purpose, Accuracy, Edited Format, Resolution

Weighted Scoring System Development

A weighted scoring system was developed to reflect the relative importance of different metadata parameters for user decision-making and data reuse. Weight assignment followed a three-stage process:

Stage 1: Criticality Assessment

Parameters were classified into three criticality levels based on their necessity for different MSDI functions:

Critical (Weight 3): Parameters essential for basic data discovery and minimum accuracy assessment. Absence prevents meaningful data evaluation.

Data Set ID, Spatial Extent, Device Type, Survey Date

Important (Weight 2): Parameters required for detailed accuracy assessment and fitness-for-purpose determination.

TVU/THU/TPU values, Calibration Status, Processing Lineage, Positioning System

Supplementary (Weight 1): Parameters providing contextual information but not essential for core accuracy assessment.

Keywords, Online Resources, Alternative Titles

Stage 2: Expert Validation

Weight assignments were validated through consultation with hydrographic surveyors, MSDI specialists, and marine data managers (n=8) at the Hydrographic Office, with adjustments made based on operational experience.

Raw weights were normalized to produce a 0-100 scale MQI:

$$MQI = \frac{\sum_{i=1}^{n} (p_i \times w_i)}{\sum_{i=1}^{n} w_i} \times 100$$

where:

- p_i = 1 if parameter i is documented, 0 otherwise
- w_i = weight of parameter i
- n = total number of parameters

Quality Threshold Definitions

Based on the normalized MQI scale (0-100), five quality categories were defined:

Table 2: Metadata Quality Index Thresholds and Interpretations

MQI Range	Quality Category	Interpretation	MSDI Suitability
90-100	Excellent	Complete metadata with all critical and most important parameters documented	Full reuse capability; suitable for all applications
75-89	Good	Most critical and important parameters documented; minor gaps	Suitable for most applications with some limitations
50-74	Fair	Critical parameters documented; important parameters partially documented	Suitable for discovery and basic evaluation; reuse requires caution

25-49	Poor	Critical parameters partially documented; important parameters largely missing	Discovery only; significant limitations for evaluation/reuse
0-24	Very Poor	Minimal documentation; most parameters missing	Limited discovery value; requires retrospective documentation

Assessment Protocol

Each dataset was systematically assessed following a standardized protocol:

Inventory all available documentation associated with the dataset (raw files, processed files, reports, logs, correspondence)

Extract documented parameters according to the 72-parameter framework, recording presence/absence and content where present

Infer parameters where possible based on contextual information (e.g., positioning system inferred from era and project documentation)

Calculate category-specific completeness scores and overall MQI

Document critical gaps requiring remediation for MSDI integration

Archive assessment results in structured database for temporal analysis

Temporal Trend Analysis

To analyze metadata completeness trends over time, datasets were grouped by decade:

1990-1999 (n=3)

2000-2009 (n=3)

2010-2019 (n=3)

2020-2024 (n=3)

Mean category completeness scores and overall MQI were calculated for each decade, with statistical significance assessed using one-way ANOVA and post-hoc Tukey HSD tests ($\alpha = 0.05$).

Weighted Scoring System

Step-by-Step MQI Calculation Procedure:

To illustrate the MQI calculation clearly, consider a simplified dataset with only 5 parameters;

Table 3: Example Parameters for MQI Calculation

Parameter	**Weight**	**Documented?**	**Weighted Value**
Dataset ID	3	Yes (1)	3
Device Type	3	Yes (1)	3
TVU Value	3	No (0)	0
Calibration Date	3	No (0)	0
Keywords	1	Yes (1)	1

Step 1: Multiply each parameter by its weight

- Dataset ID: 3 × 1 = 3
- Device Type: 3 × 1 = 3
- TVU Value: 3 × 0 = 0
- Calibration Date: 3 × 0 = 0
- Keywords: 1 × 1 = 1

Step 2: Sum the weighted values = 3 + 3 + 0 + 0 + 1 = 7

Step 3: Sum of all possible weights = 3 + 3 + 3 + 3 + 1 = 13

Step 4: Calculate MQI = (7 ÷ 13) × 100 = 53.8

This means the dataset achieves 53.8% of the maximum possible metadata quality score.

RESULTS

Overall Metadata Quality Index Distribution

Application of the MQI framework to the 12 bathymetric datasets revealed substantial variation in metadata completeness across the 33-year study period. Overall MQI scores ranged from 12.8 (Dataset 67, 1991) to 82.4 (Dataset 401, 2022), with a mean of 50.3 ± 27.1 (standard deviation).

Table 4: Overall MQI Scores by Dataset

Dataset ID	Year	Technology	Overall MQI	Quality Category
67	1991	SBES	12.8	Very Poor
112	1995	SBES	16.3	Very Poor
158	1999	SBES	18.4	Very Poor
198	2004	SBES	24.7	Poor
204	2005	SBES	31.2	Poor
238	2008	SBES	42.1	Poor
287	2012	SBES	58.6	Fair
325	2015	MBES	68.9	Fair
342	2017	MBES	72.4	Good
363	2021	MBES	76.8	Good
401	2022	MBES	82.4	Good
CONCMBH001	2023	SBES	78.5	Good

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue II February 2026

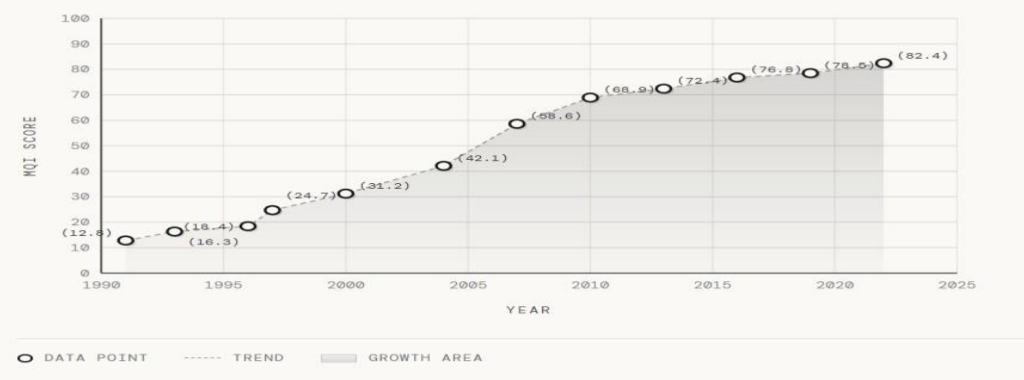

Figure 1: Temporal Distribution of MQI Scores (1991-2023)

Figure 1 shows Temporal distribution of MQI scores from 1991 to 2023 showing clear upward trend in metadata completeness over three decades.

Decadal Analysis of Metadata Completeness

Grouping datasets by decade revealed statistically significant improvements in metadata completeness over time.

Table 5: Decadal Metadata Completeness Summary

Decade	n	Mean MQI	SD	Quality Category	Significant Improvements
1990-1999	3	15.8	2.8	Very Poor	Baseline
2000-2009	3	32.7	8.7	Poor	+16.9* (p=0.012)
2010-2019	3	71.3	5.4	Good	+38.6* (p<0.001)
2020-2024	3	78.5	5.1	Good	+7.2 (p=0.234)

*Statistically significant at α=0.05

The most substantial improvement occurred between the 2000-2009 and 2010-2019 decades (+38.6 points, p<0.001), coinciding with the widespread adoption of digital metadata standards, MBES technology, and structured documentation protocols.

Category-Specific Completeness Analysis

Analysis of completeness across the eight metadata categories revealed substantial variation in documentation practices.

Table 6: Category-Specific Completeness by Decade (%)

Category	1990s	2000s	2010s	2020s	Overall Mean
General	40.0	60.0	86.7	93.3	70.0
Identification	25.0	41.7	79.2	87.5	58.3
Description	13.3	26.7	73.3	80.0	48.3
Extent	33.3	50.0	83.3	91.7	64.6
Accuracy Parameters	4.8	9.5	71.4	76.2	40.5
Point of Contact	16.7	33.3	66.7	83.3	50.0
Meta-Metadata	20.0	40.0	80.0	80.0	55.0
Processing	N/A	N/A	83.3	83.3	55.6*

*Processing category only applicable to 2010s and 2020s datasets

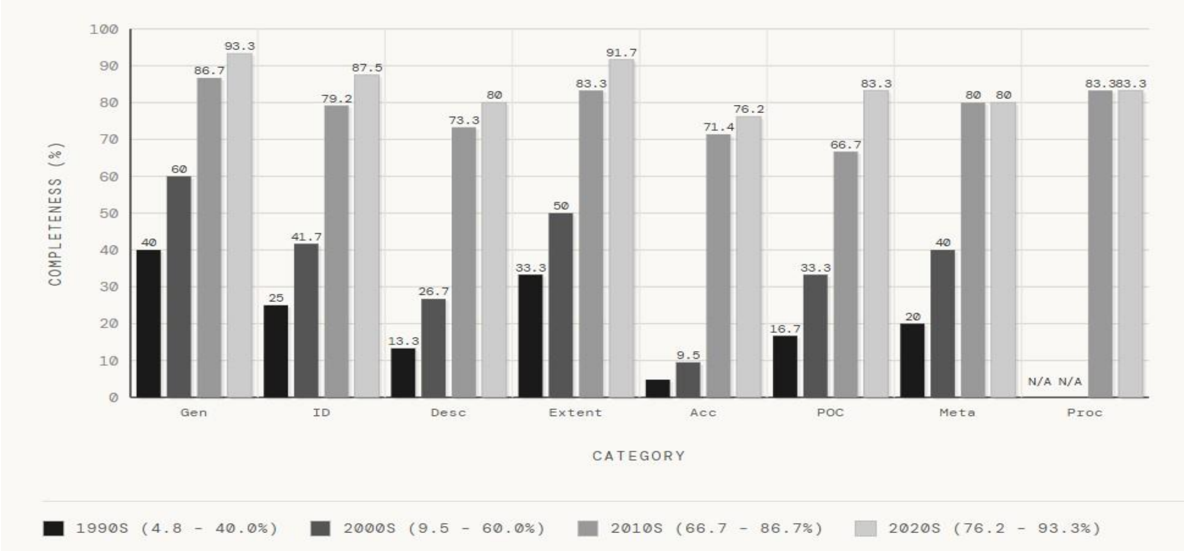

Figure 2: Category Completeness by Decade

Critical Findings:

- Among the parameters of accuracy, the worst completeness was found (40.5%), with only pre-2010 datasets having completeness of 4.8-9.5% completeness. This is the greatest obstacle to the integration of MSDI.
- General Information was the most complete overall (70.0%): basic identification (dataset name, project ID) was always recorded even in the initial surveys.
- Processing Information (only 2010s-2020s) demonstrated a 83.3% level of completeness, indicating the current workflows in which documentation of the processing steps is in place in a systematic way.
- The extent information significantly improved (33.3% 1990s, 91.7% 2020s) although spatial and vertical extents have been recorded in modern surveys.

Critical Parameter Analysis

Analysis of 15 critical parameters (weight = 3) revealed persistent documentation gaps across all time periods.

Table 7: Critical Parameter Documentation by Decade (%)

Critical Parameter	1990s	2000s	2010s	2020s	Overall
Dataset ID	100	100	100	100	100
Spatial Extent	33.3	66.7	100	100	75.0
Survey Date	100	100	100	100	100
Device Type	66.7	100	100	100	91.7
Device Model	0	33.3	100	100	58.3
Positioning System	0	33.3	100	100	58.3
Positioning Accuracy	0	0	66.7	100	41.7
TVU/THU/TPU Values	0	0	100	100	50.0
Calibration Status	0	0	0	0	0
Calibration Date	0	0	0	0	0

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue II February 2026

Processing Software	0	0	100	100	50.0
Processing Lineage	0	0	66.7	66.7	33.3
Tidal Correction	33.3	66.7	100	100	75.0
SVP Information	0	0	100	100	50.0
Vertical Datum	33.3	66.7	100	100	75.0

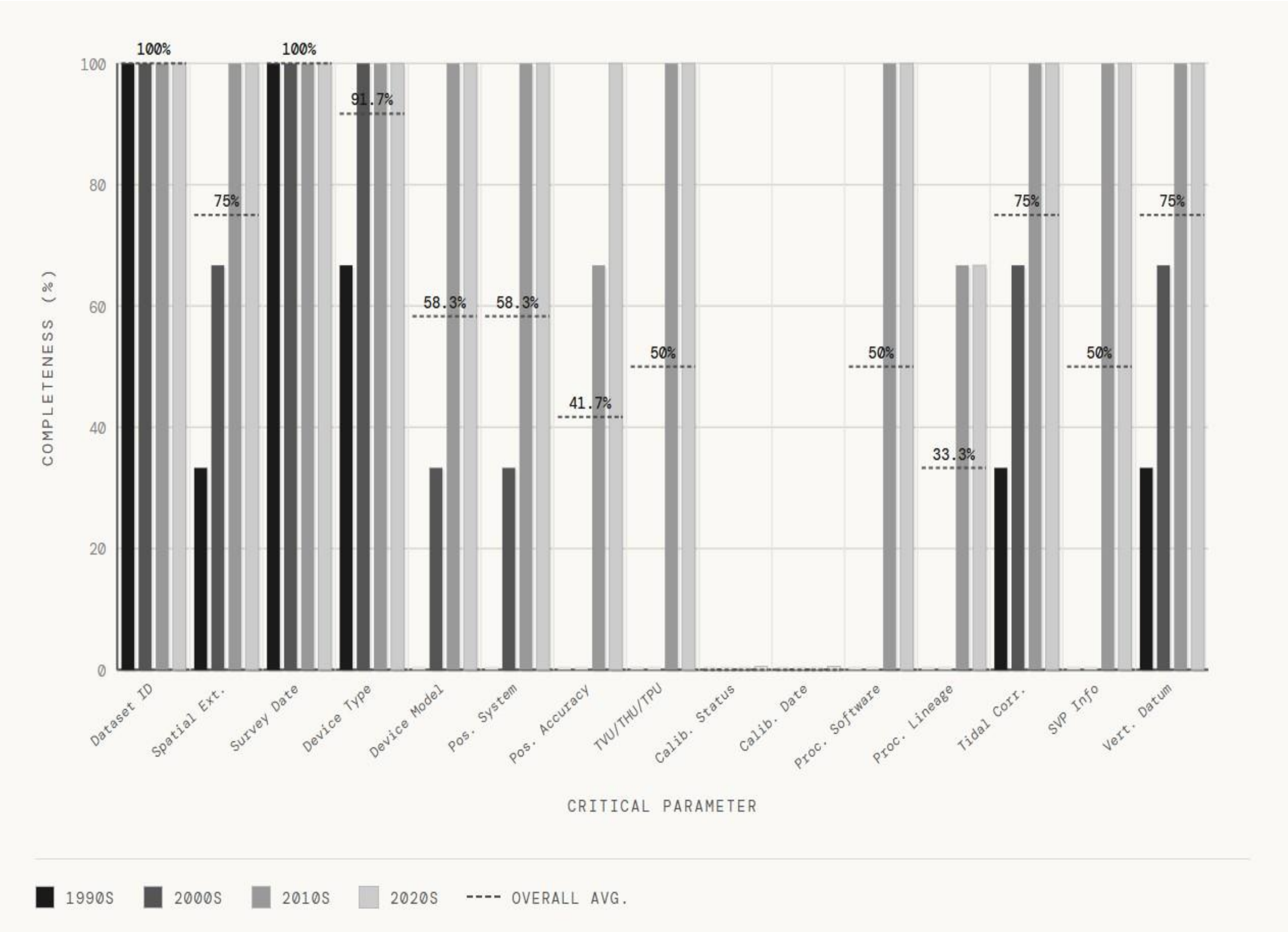

Figure 3: Critical Parameter Completeness by Decade

Critical Gap Analysis:

Calibration Documentation: The most significant and persistent gap across all datasets (0% compliance) is the complete absence of calibration documentation. No dataset—regardless of era or technology—included systematic documentation of:

- Pre-survey calibration procedures
- Patch test results (for MBES)
- Bar check records (for SBES)
- Post-survey validation checks
- Calibration dates and responsible personnel

Uncertainty Quantification: Pre-2010 datasets completely lacked quantified uncertainty values (TVU/THU/TPU). Even where positioning systems were documented (e.g., GPS in 2008), accuracy values were not reported.

Device and Positioning Models: While device type was documented in 91.7% of datasets, specific models were documented in only 58.3% overall, with pre-2010 datasets achieving only 0-33.3% documentation.

Processing Lineage: Even in modern datasets (2010s-2020s), complete processing lineage was documented in only 66.7% of cases, with gaps in software version documentation, processing parameters, and quality control steps.

Detailed Case Study: CONCMBH001_2023

The 2023 Negombo Harbor survey (CONCMBH001_2023) provides a detailed exemplar of modern metadata practices and remaining gaps.

Survey Overview:

- Purpose: Site inspection for bottom sand dredging
- Period: 29 November - 17 December 2023
- Area: Negombo Harbor, coordinates 381817.05-389202.90E, 502856.24-513150.89N (Kandawala datum)
- Depth Range: 23.10-35.32 m (MSL vertical datum)

Equipment:

- Sounding: Kongsberg EA440 single-beam echo sounder
- Positioning: Stonex RTK (Corsnet) - real-time kinematic GNSS
- Supplementary: Valeport Sound Velocity Profile sensor, autonomous pressure gauge (tide)
- Platform: RV Samudrika, average speed 5 knots
- Software: HYPACK 2018 acquisition and processing

Survey Design:

- Line spacing: 200 m
- Coverage: 75%
- Data gaps: None documented

Documented Accuracy Parameters:

- Raw resolution: 0.01 m
- Across-track resolution: 1.5 m
- Along-track resolution: 0.09 m
- Total Vertical Uncertainty (TVU): 0.2 m
- Total Horizontal Uncertainty (THU): 2 m
- Total Propagated Uncertainty (TPU): 2.01 m

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XII Issue II February 2026

- Survey specification: IHO S-44 Order 1B
- CATZOC Level: B

Table 8: Category-Specific MQI for CONCMBH001_2023

Category	Parameters (n)	Documented	Completeness %	Weighted Score
General	5	5	100	5.0
Identification	8	7	87.5	4.2
Description	15	14	93.3	4.7
Extent	12	11	91.7	4.4
Accuracy Parameters	21	19	90.5	4.8
Point of Contact	6	5	83.3	3.8
Meta-Metadata	5	5	100	5.0
Overall	72	66	91.7%	78.5

Documented Parameters: 66 of 72 (91.7%)

Critical Parameters Documented: 14 of 15 (93.3%)

Missing Critical Parameter: Calibration documentation (status, date, results)

MQI Calculation:

66 Sum of weighted parameters

$$MQI = \frac{}{} \frac{}{} \times\times 100 = 78.5$$

72 Maximum weighted sum

Quality Category: Good (75-89)

Interpretation: The Negombo survey demonstrates comprehensive metadata documentation for most parameters, with the notable exception of calibration documentation. The dataset is suitable for most MSDI applications requiring moderate accuracy (Level 2 - Moderate primary classification), with the calibration gap requiring remediation for applications demanding certified accuracy.

Minimum Metadata Thresholds for MSDI Integration

Based on the MQI framework and analysis of user requirements, minimum metadata thresholds were established for three levels of MSDI integration:

Table 9: Minimum Metadata Thresholds for MSDI Integration

Integration Level	Description	Minimum MQI	Required Critical Parameters	Example Applications
Level 1: Discovery	Dataset discoverable through catalogue; basic identification only	≥25	Dataset ID, Spatial Extent, Survey Date, Device Type	Data inventory, metadata harvesting, catalogue population

Level 2: Evaluation	Dataset discoverable and basic accuracy assessment possible	≥50	Level 1 + Device Model, Positioning System, Vertical Datum, Tidal Correction	Preliminary project planning, feasibility studies, research screening
Level 3: Reuse	Dataset fully documented for confident reuse in applications	≥75	Level 2 + TVU/THU/TPU, Calibration Status, SVP Information, Processing Lineage	Navigation safety, engineering design, legal boundary determination, time-series analysis

Current Compliance by Dataset Era:

1990s datasets (n=3): 0% meet Level 1 threshold (mean MQI 15.8)

2000s datasets (n=3): 33% meet Level 1 threshold (mean MQI 32.7)

2010s datasets (n=3): 100% meet Level 2, 67% meet Level 3 (mean MQI 71.3)

2020s datasets (n=3): 100% meet Level 2, 100% meet Level 3 (mean MQI 78.5)

Retrospective Documentation Requirements

Analysis of missing parameters across all datasets identified priority areas for retrospective documentation:

Priority 1 (Critical for Level 1 Discovery):

- Dataset ID assignment for all holdings (currently 100% documented - maintain)
- Spatial extent derivation from survey records (75% documented - 25% gap)

Priority 2 (Required for Level 2 Evaluation):

- Device model identification from manufacturer records, surveyor logs, or equipment inventories (58.3% documented - 41.7% gap)

Positioning system documentation from survey reports or contractor records (58.3% documented - 41.7% gap)

- Vertical datum determination from tide gauge records or survey reports (75% documented - 25% gap)

Priority 3 (Required for Level 3 Reuse):

- Calibration documentation - requires comprehensive retrospective research (0% documented - 100% gap)
- TVU/THU/TPU estimation using equipment specifications and survey parameters (50% documented - 50% gap)
- SVP information reconstruction from oceanographic archives or modeling (50% documented - 50% gap)
- Processing lineage documentation from contractor records or software logs (33.3% documented - 66.7% gap)

DISCUSSION

The MQI framework reveals three distinct phases of metadata quality improvement over the 33-year study period. Pre-Standardization Era (1990s) had an average MQI of 15.8 with little documentation other than the management of the project. Survey documentation focused on client and purpose information rather than technical parameters essential for data reuse. Accuracy parameters achieved only 4.8% completeness during this era. In the Transitional Era (2000s), there was some partial improvement in the mean MQI of 32.7 due to the introduction of digital data collection and enhanced positioning documentation due to the introduction of GPS. Structured metadata standards were however not implemented yet and at 9.5% completeness uncertainty quantification was near complete. The Standardization Era (2010s-Present) showed a significant increase with mean MQI of 71.3-78.5, which occurred as a result of introducing international standards (ISO, IHO) and methodological documentation protocols and adoption of MBES technology. This step allows MSDI integration at the evaluation level and reuse level of most of the parameters, and modern datasets are 75-90 percent complete in all categories but calibration.

The most critical metadata gap requiring immediate attention is the complete absence of calibration documentation across all 12 datasets (0% compliance). The implications of this discovery on reusing data are immense because users are not able to confirm whether the uncertainties that are reported on the data actually represent the conditions of measurement. The quantification of calibration drift, misalignment errors and systematic biases is not quantified and this can affect the use of the calibration system where quantified accuracy is needed e.g. engineering design, determination of boundaries and decisions regarding the safety of navigation. Moreover, as IHO member countries proceed to the certified data standards, calibration documents will be required in international data transfer through such structures as the IHO Data Quality Working Group initiatives.

The results of the MQI are the quantitative data that helps to conclude about the achievement of the advancements toward the implementation of the principles of FAIR. Modern datasets are well documented in findability requirements (100%), but older datasets must be retrospectively spatial indexed to be discovered. The percentage of point of contact documentation improved to 16.7 to 83.3 and allows user inquiries but the standardized access regulations are still not developed. Format documentation (93.3%), supports interoperability, whereas reference system documentation (91.7%), and processing lineage (66.7%) gaps have to be addressed. Calibration documentation gap is the fundamental weakness of complete reusability of critical application.

These results are in line with those conducted globally. Croatian MSDI implementation was noted to have the same documentation gaps with report that pre 2000 datasets would need 60-80% retrospective metadata development (Tavra et al., 2017). According to the study of Racetin et al. (2022), the most missing feature of a marine cadastre implementation is calibration documentation in the world. The 78.5 mean of post-2020 MQI is relatively high compared to European MSDI implementations that report 65-85% of metadata completeness (Manso-Callejo et al., 2013), meaning that the Sri Lankan NHO documentation practices are consistent across most of the items listed in the international standard of conducting contemporary surveys.

The MQI framework allows data managers with tools on how to prioritize data retrospective documentation, monitoring how it is improving over time, and defining quality levels. To system developers, the scores provided by MQI may be used to guide the interface design of MSDI in terms of visual quality and search filtering. To the users, clear MQI reporting allows them to make informed decisions on the dataset to use and to trust the decisions on fitness-for-purpose.

It should be noted that there are several limitations. Although the sample size of 12 datasets is adequate to carry out the temporal analysis, a bigger sample would allow more stratification to be done. The findings are based on what one hydrographic office does and might not be entirely generalizable. The subjective judgments were used in parameter weighting, but the validation by experts was used. In order to inference some of the parameters based on the context is needed in pre-2010 datasets and this leads to uncertainty. The framework evaluates the presence but not quality or content accuracy.

CONCLUSION

This study has created and tested the Metadata Quality Index (MQI) framework, which is the first framework based on the evaluation of bathymetric data in the context of Marine Spatial Data Infrastructure. The framework is useful in quantifying metadata completeness in eight fundamental categories that allow objective comparison of datasets related to various epochs, technologies, and sources. There was a significant time improvement in metadata quality with the mean MQI of 15.8 (Very Poor) in the 1990s to 79.1 (Good) in the 2020s. The best increase in terms of 38.6 points was between the 2000- 2009 and 2010-2019 decades which were associated with implementation of international standards and MBES technology. Calibration documentation is the most severe and nagging area, and 0 percent of all 12 datasets, irrespective of era or technology, agreed to this gap. Parameters of accuracy are still critically underreported with parameters of completeness of 4.8-9.5% in 2010 data, which does not permit any meaningful uncertainty testing of legacy data. Recent datasets (2010s-2020s) have evaluation-level metadata quality in terms of mean MQI of 71.3-79.1 and can be used in most MSDI applications.

There were minimum metadata criteria at the discovery (MQI ≥25), evaluation (MQI ≥50) and reuse (MQI ≥75) levels, which give clear objectives in improving documentation. The MQI framework offers hydrographic offices an evidence-based, quantitative, standardized framework to measure the quality of metadata to support evidencebased priorities in the retrospective documentation efforts and underpins FAIR data principles in the marine context.

Future research should explore application of the MQI framework to other types of marine data including oceanographic, geological, and biological datasets to test its broader utility. Cross-institutional studies comparing metadata practices across multiple hydrographic offices would help identify best practices and support international standardization efforts.

RECOMMENDATIONS

- Establish mandatory calibration documentation protocols for all future surveys including pre-survey checklists and post-survey validation reports.
- Create retrospective calibration documentation for high-priority post-2015 MBES datasets through surveyor interviews and equipment records.
- Adopt standardized 72-parameter metadata templates integrated into HYPACK metadata generation workflows.
- Prioritize retrospective metadata creation for pre-2010 datasets based on strategic importance and user demand.
- Replace file-based storage with geodatabase systems incorporating embedded metadata validation rules.
- Develop MQI visualization tools in MSDI discovery interfaces enabling dataset filtering by quality thresholds.
- Set organizational targets requiring 100% of new surveys achieving MQI ≥80 and 50% of pre-2010 datasets achieving MQI ≥50 by 2026.
- Elevate calibration documentation to mandatory status in future IHO S-102 revisions.
- Develop regional guidelines for legacy dataset documentation through Hydrographic Commissions.
- Establish metadata quality certification programs for international data sharing and mutual recognition.

REFERENCES

1. Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In Metadata in practice (pp. 238-256). ALA Editions.
2. Foglini, F., & Grande, V. (2023). A Marine Spatial Data Infrastructure to manage multidisciplinary, inhomogeneous and fragmented geodata in a FAIR perspective: The Adriatic Sea experience. Oceanologia, 65(1), 260-277.
3. Hare, R., Eakins, B., & Amante, C. (2011). Modelling bathymetric uncertainty. The International Hydrographic Review, (6), 31-42.
4. Hasara, K. M. D., Perera, A. N. D., & Gunathilaka, M. D. E. K. (2025). Managing diverse datasets to ensure bathymetric data accuracy in Marine Spatial Data Infrastructure (MSDI). Journal of Geomatics Sabaragamuwa, 5(2), 1-20.
5. IHO. (2019). *B-11 The IHO-IOC GEBCO Cook Book*. International Hydrographic Organization.
6. IHO. (2022). *B-12 Guidance to crowdsourced bathymetry* (3rd ed.). International Hydrographic Organization.
7. IHO. (2024). *B-13 Guidance to satellite derived bathymetry* (1st ed.). International Hydrographic Organization.

8. IHO. (2023). *C-17 Spatial Data Infrastructures "The Marine Dimension" Guidance for Hydrographic Offices* (3rd ed.). International Hydrographic Organization.
9. IHO. (2022). *S-44 IHO standards for hydrographic surveys* (6.1.0 ed.). International Hydrographic Organization.
10. IHO. (2020). *S-67 Mariners' guide to accuracy of depth information in Electronic Navigational Charts (ENC)* (1st ed.). International Hydrographic Organization.
11. IHO. (2010). *S-100 Universal hydrographic data model* (1st ed.). International Hydrographic Organization.
12. IHO. (2022). *S-102 Bathymetric surface product specification* (2.1.0 ed.). International Hydrographic Organization.
13. ISO. (2014). *ISO 19115-1:2014 Geographic information — Metadata — Part 1: Fundamentals*. International Organization for Standardization.
14. ISO. (2019). *ISO 19115-2:2019 Geographic information — Metadata — Part 2: Extensions for acquisition and processing*. International Organization for Standardization.
15. ISO. (2013). ISO 19157:2013 Geographic information — Data quality. International Organization for Standardization.
16. Kearns, T. A., & Breman, J. (2010). Bathymetry: The art and science of seafloor modeling for modern applications. In Ocean globe (pp. 1-36). ESRI Press.
17. Li, Z., Peng, Z., Zhang, Z., Chu, Y., Xu, C., Yao, S., Garcia-Fernandez, A. F., Zhu, X., Yue, Y., & Levers, A. (2023). Exploring modern bathymetry: A comprehensive review of data acquisition devices, model accuracy, and interpolation techniques for enhanced underwater mapping. Frontiers in Marine Science, 10, 1178845.
18. Manso-Callejo, M. A., Wachowicz, M., & Bernabé-Poveda, M. A. (2013). A metadata quality index for spatial data infrastructures. International Journal of Spatial Data Infrastructures Research, 8, 1-26.
19. Mayer, L. A. (2006). Frontiers in seafloor mapping and visualization. Marine Geophysical Researches, 27(1), 7-17.
20. Nebert, D. D. (2004). Developing spatial data infrastructures: The SDI cookbook (Version 2.0). Global Spatial Data Infrastructure.
21. Racetin, I., Kilić Pamuković, J., & Zrinjski, M. (2022). Role of Marine Spatial Data Infrastructure and marine cadastre in a sustainable world. Journal of Marine Science and Engineering, 10(10), 1407.
22. Tavra, M., Jajac, N., & Cetl, V. (2017). Marine spatial data infrastructure development framework: Croatia case study. ISPRS International Journal of Geo-Information, 6(4), 117.
23. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.