



中 国 船 级 社

CHINA CLASSIFICATION SOCIETY

船舶数据质量评估与安全指南

Guidelines for Safety and Assessment of Ship Data

2020

(征求意见稿)

前 言

航运业在信息化、数字化、智能化的发展中逐步积累了大量船舶数据，其已成为重要的资产，如何充分保护数据及评估船舶数据的价值，是航运业面临的重要且急迫的课题。近年，国家陆续出台数据保护相关指南及要求，如何尽快在航运业落地实施，同时，数据相关方面的研究也逐渐成为海事界研究热点，如数据完整性、数据质量相继成为 IACS 及各船级社相继推出了研究建议，对船舶数据质量评估提供相关服务，在此背景，编制船舶数据安全及质量评估相关技术要求，将是非常迫切解决的问题。

本指南研究船舶数据安全技术体系，分析船舶数据特点，制定船舶数据分类原则及分级标准，识别船舶数据安全风险，制定船舶数据营运、应用规范、隐私保护、技术标准、测评标准及要求，研究数据在采集、存储、处理、传输及共享等生命周期各环节的安全防护技术。研究船舶数据质量评估框架，研究船舶数据评估方法，确立评估指标，形成船舶数据质量评估要求。为行业规范利用数据、提升数据质量提供指导及参考。

目 录

第 1 章	通则.....	1
第 1 节	一般规定.....	1
1.1.1	一般要求.....	1
第 2 节	适用范围.....	1
1.2.1	数据范围.....	1
1.2.2	业务范围.....	1
1.2.3	数据类型范围.....	1
第 3 节	术语和定义.....	2
第 4 节	规范性引用文件.....	2
第 5 节	图纸审查.....	3
第 2 章	数据分类分级.....	4
第 1 节	一般规定.....	4
2.1.1	一般要求.....	4
2.1.2	数据分类分级原则.....	4
2.1.3	数据分类分级流程.....	4
第 2 节	数据分类.....	5
2.2.1	一般要求.....	5
2.2.2	数据分类方法.....	5
第 3 节	数据分级.....	5
2.3.1	一般要求.....	5
2.3.2	分级方法.....	5
第 3 章	数据质量评估.....	7
第 1 节	一般规定.....	7
3.1.1	一般要求.....	7
3.1.2	适用范围.....	7
3.1.3	数据类型范围.....	7
3.1.4	数据标识.....	7
第 2 节	数据质量标准.....	7
3.2.1	评估标准.....	7
3.2.2	ISO 8000.....	7
3.2.3	ISO 8000 的结构.....	8
3.2.4	ISO 8000-8.....	8
3.2.5	SQuaRE 系列标准.....	8
3.2.6	ISO/IEC 25012.....	8
第 3 节	船舶数据质量评估框架.....	9
3.3.1	数据质量总体框架.....	9
3.3.2	数据质量层级.....	9
3.3.3	数据质量评估维度.....	10
第 4 节	数据质量测量规则和指标.....	11
3.4.1	准确性.....	12
3.4.2	完整性.....	12
3.4.3	一致性.....	12

3.4.4 可信性.....	12
3.4.5 时效性.....	13
3.4.6 及时性.....	13
3.4.7 合规性.....	13
3.4.8 效率性.....	13
3.4.9 精确性.....	13
3.4.10 可追溯性.....	14
3.4.11 可理解性.....	14
3.4.12 可用性.....	14
3.4.13 可移动性.....	14
3.4.14 可恢复性.....	15
3.4.15 有效性.....	15
3.4.16 连贯性.....	15
3.4.17 合理性.....	15
3.4.18 重复性.....	15
第 5 节 船舶数据质量评估流程.....	16
3.5.1 船舶数据质量评估基本流程.....	16
3.5.2 数据质量评估需求分析.....	16
3.5.3 数据质量评估规则 and 标准.....	16
3.5.4 数据质量评估.....	17
3.5.5 数据质量评估报告.....	18
3.5.6 数据质量持续改进.....	18
第 6 节 船舶数据质量评估结果分析.....	18
3.6.1 数据质量维度测量.....	18
3.6.2 数据质量维度权重.....	18
3.6.3 数据质量指标.....	19
第 4 章 数据安全要求.....	20
第 1 节 一般规定.....	20
4.1.1 一般要求.....	20
第 2 节 数据安全框架.....	20
4.2.1 数据安全框架.....	20
第 3 节 数据采集.....	20
4.3.1 数据采集活动概念.....	20
4.3.2 安全要求.....	21
第 4 节 数据存储.....	21
4.4.1 数据存储活动概念.....	21
4.4.2 安全要求.....	21
第 5 节 数据处理.....	21
4.5.1 数据处理活动概念.....	21
4.5.2 安全要求.....	22
第 6 节 数据分发.....	22
4.6.1 数据分发活动概念.....	22
4.6.2 安全要求.....	22
第 7 节 数据共享.....	23

4.7.1 数据共享活动概念	23
4.7.2 安全要求	23
第 8 节 数据删除	23
4.8.1 数据删除活动概念	23
4.8.2 安全要求	23
附录 1 智能航行系统所应满足的元数据要求	24
附录 2 智能机舱系统所应满足的元数据要求	25
附录 3 智能能效系统所应满足的元数据要求	26
附录 4 智能船体系统所应满足的元数据要求	27
附录 5 智能货控系统所应满足的元数据要求	28
附录 6 智能集成平台所应满足的元数据要求	29
附录 7 远程控制站所应满足的元数据要求	30

第1章 通则

第1节 一般规定

1.1.1 一般要求

1.1.1.1 本指南规定了拟申请智能船舶附加标志应满足的数据质量与安全要求。

1.1.1.2 数据评估的目的找出数据错误和错误的数据元素，并测量各种数据驱动的业务流程的影响。

1.1.1.3 以数据进行风险评估的基础上，满足本指南数据安全相关要求，不单单局限于技术手段或管理措施。

第2节 适用范围

1.2.1 数据范围

1.2.1.1 本指南适用于船载系统中所采集、存储、处理、传输及共享的数据。包括完全在基于计算机系统的船上生命周期内的数据，以及与连接到船上网络的岸上系统交换的数据。

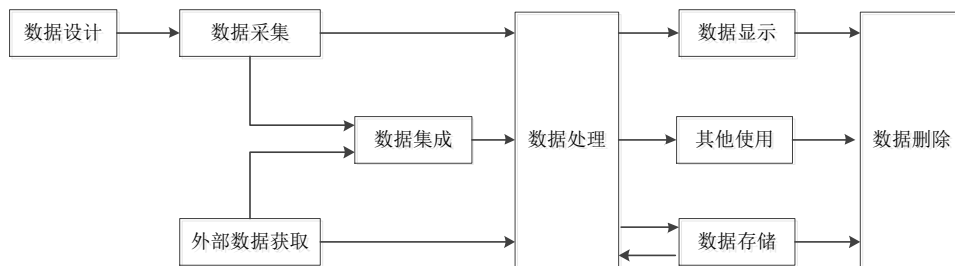


图 1.2.1.1 数据生命周期

1.2.1.2 本指南数据系指船舶及海洋结构物全生命周期产生和应用的数据，包括但不限于从设计、建造、营运及拆解等环节中生成和使用的数据，以及船舶及海洋结构物在设备接入、平台运行、系统 APP 应用等过程中生成和使用的数据。

1.2.2 业务范围

智能系统所涉及的数据见附录，其它系统可参照执行。

1.2.3 数据类型范围

可以包括但不限于：

- (1) 流数据，批数据，或静态数据；
- (2) 结构化或非结构化数据；
- (3) 大容量或小容量数据；
- (4) 孤立数据或分布式数据；
- (5) 数据价值链；
- (6) 数字孪生，仿真模型，语义模型，系统描述，或类别。

第3节 术语和定义

除另有规定外，本指南有关定义如下：

数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

数据资产 data asset

组织拥有和控制的、能够产生效益的数据资源。

数据分类 data classification

根据数据的属性、特征、来源、内容等方面，按照一定的原则和方法进行区分和归类。

数据分级 data grading

根据数据的重要性、敏感程度等方面对数据划分等级，以便于对数据采取差异化的安全措施。

数据生命周期 data life cycle

数据获取、存储、整合、分析、应用、呈现、归档和销毁等各种生存形态演变的过程。

数据质量 data quality

在指定条件下使用时，数据的特性满足明确的和隐含的要求的程度。

数据安全 data security

保持、维护数据的保密性、完整性和可用性，也可包括真实性、可核查性、抗抵赖性、可靠性等性质。

主数据 master data

满足跨部门业务协同需要的、反映核心业务实体状态属性的基础信息。

元数据 metadata

关于数据或数据元素的数据(可能包括其数据描述),以及关于数据拥有权、存取路径、访问权和数据易变性的数据。

第4节 规范性引用文件

相关文件中的条款通过本规范的引用将成为本规范的条款，凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本指南。

序号	文件编号	文件名称
1		中国船级社《钢质海船入级规范》（2018）及其修改通报
2	IACS Rec 166	Recommendation on Cyber Resilience
3	ISO/IEC 25012-2013	数据质量模型
4	ISO/IEC 25024-2013	数据质量测量
5	GB/T 37973-2019	大数据安全管理指南
6	ISO 8000-8	信息和数据质量：概念和测量

第5节 图纸审查

进行数据质量评估及安全评估时，下列图纸资料应提交审批备查：

- (1) 软件说明书；
- (2) 数据分类说明书（适用时）；
- (3) 数据分级说明书；
- (4) 系统数据设计说明书。

第2章 数据分类分级

第1节 一般规定

2.1.1 一般要求

本指南数据分类分级适用于船舶及海洋结构物、船东、船舶管理公司等开展数据分类分级工作。涉及国家秘密信息的数据，应遵守船旗国及主管机关法律法规的规定，不适用本指南。

2.1.2 数据分类分级原则

2.1.2.1 数据分类分级应满足如下原则：

- (1) 科学性，按照数据的多维特征及其相互间逻辑关联进行科学和系统地分类。
- (2) 稳定性，应以数据最稳定的特征和属性为依据制定分类和分级方案。
- (3) 实用性，数据分类要确保每个分类有数据，不设没有意义的类目，数据类目划分要符合对数据分类的普遍认识。数据分级要确保分级结果能够为数据保护提供有效信息，应提出分级安全要求。
- (4) 扩展性，数据分类和分级方案在总体上应具有概括性和包容性，能够针对组织各种类型数据开展分类和分级，并满足将来可能出现的数据的分类和分级要求。
- (5) 适用性，既要满足数据质量需求和数据安全需求，又要满足船舶实际情况。

2.1.3 数据分类分级流程

2.1.3.1 应结合自身业务特点，针对采集、存储和处理的数据，制定数据分类分级标准及要求，规范应包含但不限于以下内容：

- (1) 数据分类方法及指南；
- (2) 数据分级详细清单，包含每类数据的初始安全级别；
- (3) 数据分级保护的安全要求。

2.1.3.2 应按照图 2.1 的流程对数据进行分类分级。组织应根据数据分类分级标准及要求对数据进行分类；为分类的数据设定初始安全级别；综合分析业务、安全风险、安全措施等因素后，评估初始安全级别是否满足船舶数据安全需求，对不恰当的数据分级进行调整，并确定数据的最终安全级别。

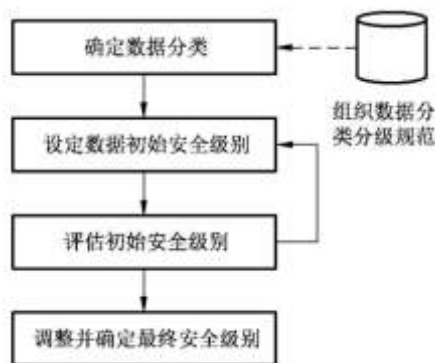


图 2.1 数据分类分级实施步骤

第2节 数据分类

2.2.1 一般要求

2.2.1.1 船东、船舶管理公司可按照组织数据资产管理形式，对船舶及海洋结构物的数据进行分类，可应结合建造模式、运营模式，分析梳理业务流程和系统设备，考虑行业要求、业务规模、数据复杂度等实际情况，对数据进行分类梳理和标识，形成数据分类清单。

2.2.2 数据分类方法

2.2.2.1 数据分类维度可以采用适用的分类方式，可以采用如下，但不局限于：

(1) 与船舶生命周期环节相关的分类方式，如研发数据（研发设计数据、开发测试数据等）、生产制造数据（控制信息、工况状态、工艺参数、系统日志等）、运维数据（货物数据、服务数据等）、管理数据（资产信息、货运信息、供应链数据、业务统计数据等）、外部数据（与其他主体共享的数据等）。

(2) 与影响船舶、人员及环境安全相关的分类方式，如操控类数据、运营类数据、分析类数据等。

(3) 与业务相关的数据分类方式，如船舶控制自身数据、船舶与外界交互数据、船舶货物数据、船舶维护保养数据等

(4) 按数据的性质分类，如定位数据（各种坐标数据、方向数据）、定性数据（事物属性的数据）、定量数据（事物数量特征的数据，如长度、面积、体积等几何量，重量、速度等物理量）和定时数据（时间特性的数据，如年、月、日、时、分、秒等）。

(5) 按表现形式分类，如数字数据（在某个区间内离散的值如符号、文字）、模拟数据（在某个区间内是连续的值，如声音、图像）。

(5) 按记录方式分类，如纸质（航行记录、海图、气象数据等）、表格、影像、磁带。

(6) 按数字化方式分类，如矢量数据、格网数据等。

2.2.2.2 船舶数据分类应随船载系统的建设与发展相适应。

第3节 数据分级

2.3.1 一般要求

2.3.1.1 应基于风险确定各类数据的数据分级规则。

2.3.1.2 数据分级则是从数据安全、隐私保护和数据使用的角度进行分级。

2.3.2 分级方法

2.3.2.1 根据船舶及海洋结构物数据遭篡改、破坏、泄露或非法利用后，数据的机密性、完整性或可用性的丧失可能对船舶、船东、船舶管理公司及利益相关者等带来的潜在影响，可将数据分为一级、二级、三级等3个级别。

2.3.2.2 机密性、完整性或可用性的丧失可能会导致灾难性影响，潜在影响符合下列条件之一的数据为三级数据：

(1) 对船舶、人员及环境造成严重的或灾难性的不利影响，可能：

- ① 在组织无法执行其一项或多项主要职能的程度和持续时间内，导致严重的船舶运营退化或损失；
- ② 对组织资产造成重大损害；
- ③ 造成重大财务损失；

- ④ 对个人造成严重或灾难性伤害，涉及生命损失或严重威胁生命的伤害。

2.3.2.3 机密性、完整性或可用性的丧失可能会对人员安全、船舶安全和/或对环境的威胁产生严重不利影响，符合下列条件之一的数据为二级数据：

- ① 在一定程度上导致船舶运营的严重恶化，直至组织可以执行其主要职能的程度和持续时间，但职能的有效性大大降低；
- ② 对组织资产造成重大损害；
- ③ 造成重大财务损失；
- ④ 对个人造成重大伤害，不涉及生命损失或严重威胁生命的伤害。

2.3.2.4 机密性、完整性或可用性的丧失可能会对人员安全、船舶安全和/或对环境的威胁产生有限的不利影响，符合下列条件之一的数据为一级数据：

- ① 在一定程度上导致船舶运营性能下降，并在一定程度上持续了组织可以执行其主要职能的时间，但功能的有效性明显降低；
- ② 对组织资产造成较小的损害；
- ③ 造成轻微的财务损失；
- ④ 对个人造成轻微伤害

表 2.3.3.1

级别	影响	系统功能	机密性	完整性	可用性
I	这些系统的故障不会导致人员安全、船舶安全和/或对环境的威胁等危险情况。	信息/管理任务的监控功能	低	中	低
II	这些系统的故障最终可能导致人身安全、船舶安全和/或对环境的威胁等危险情况。	报警和监控功能，控制功能，这些使船舶保持正常操作和居住环境的必要功能。	中	高	中
III	这些系统的故障可能会立即导致人身安全、船舶安全和/或对环境的威胁等危险情况	维持船舶推进和转向安全功能的控制功能	中	高	高

第3章 数据质量评估

第1节 一般规定

3.1.1 一般要求

3.1.1.1 测量和监控数据资产的数据质量(识别数据使用的风险)。

3.1.1.2 满足本指南质量评估结果的船舶，方可授予智能附加标志或保持智能附加标志。

3.1.2 适用范围

3.1.2.1 本指南适用于智能船舶，申请智能附加标志的系统，评估的数据范围为在智能附加标志所列的系统要求功能所涉及的数据，其它船舶及系统可参照执行。

3.1.3 数据类型范围

3.1.3.1 本指南所指数据包含但不限于：

- (1) 结构化数据（数据库、缓存、Key-Value 存储系统等等）；
- (2) 非结构化数据（文档、图片、音频、视屏等）；
- (3) 不仅包括存储的数据，还包括使用、传输、流转中的数据。

3.1.3.2 本章节进行评估的数据时间周期指在船舶年度检验、特别检验等周期，船载系统留存的数据。

3.1.4 数据标识

按照一定编码规则、语义确定数据标识。

第2节 数据质量标准

3.2.1 评估标准

3.2.1.1 本指南在评估过程中，参考以下标准：

- (1) ISO 8000 系列
- (2) SQuaRE 系列标准
- (3) 智能船舶规范（2020）
- (4) 软件可靠性及安全评估指南（2018）
- (5) 船舶网络安全要求及安全评估指南（2020）
- (6) 智能船舶集成平台检验指南（2018）

3.2.2 ISO 8000

ISO 8000 是数据质量的国际标准。ISO 断言：“创建，收集，存储，维护，传输，处理和呈现数据并以及时且经济有效的方式支持业务流程的能力，既需要了解决定数据质量的数据特征，也需要了解数据的质量。测量，管理和报告数据质量的能力。”ISO 8000 表示管理，衡量和改善特定数据类型的数

据质量的方法，这些数据类型包括但不限于主数据，交易数据和产品数据。ISO 8000 可以单独使用或与质量管理体系结合使用。

3.2.3 ISO 8000 的结构

ISO 8000 定义了数据质量的相关特征，规定了适用于这些特征的要求，并提供了改善数据质量的指南。ISO 8000 适用于数据生命周期的所有阶段。ISO 8000 由一系列部分组成，包括：

- 第 1 至 99 部分：常规数据质量
- 第 100 到 199 部分：主数据质量
- 第 200 至 299 部分：交易数据质量
- 第 300 至 399 部分：产品数据质量

3.2.4 ISO 8000-8

ISO 8000 的第 8 部分（ISO 8000-8: 2015, 2015 年 11 月）提供了计划和执行数据质量测量的基本概念。ISO 8000-8 的应用不受组织状态，信息或数据类型，硬件存储介质，软件，信息安全性和信息生命周期阶段的影响。ISO 8000-8 的主要目的是为根据以下三个类别测量数据质量提供基础：

（1）语法质量：数据符合其指定语法的程度，如元数据规定的要求。此处，元数据是指，如合法值，数据类型，参考完整性（例如数据之间的关联），业务词汇，和任何定义的业务规则。比如定位系统输出的位置信息中有表示模式的标志位，如果该标志位使用字符表示，比如使用 D 表示差分模式，则此标志位不能为非字符，且不能为规定字符之外的其他字符。

（2）语义质量：数据与数据所代表的一致程度，比如当罗经系统输出的船舶艏向为 112°，船舶的实际艏向也应为 112°，如果不是，那有一些语义错误量。比如，定位系统输出的船舶位置的纬度值，超过 90°或者为负值，也为语义错误。

（3）语用质量：数据对于特定目的的适用程度。它验证数据使用者对使用目的的适用性的看法。比如应用需要罗经系统输出的罗经频率不低于 10Hz，当实际输出频率低于 10Hz 时，认为有一些语用质量问题。

ISO 8000-8 还提供了测量数据质量的先决条件，它专门描述了要求以及与三个数据质量类别中的每一个有关的数据质量验证和确认的一组数据质量规则/维度。同时，提供了一些叙述性示例来描述违反规则的情况。

3.2.5 SQuaRE 系列标准

ISO/IEC 25012 定义的数据质量模型，应与 SQuaRE 系列国际标准结合应用，主要包括：

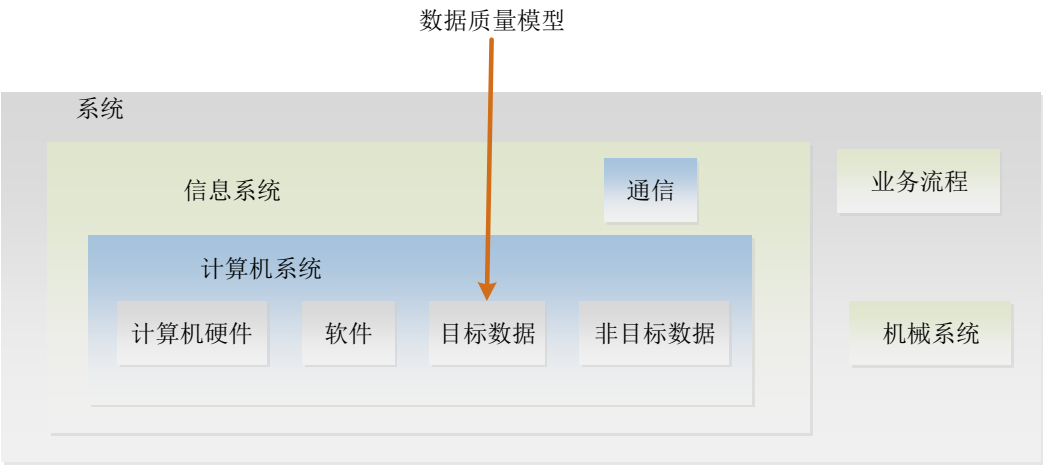
- （1）ISO/IEC 2500n – 质量管理类国际标准
- （2）ISO/IEC 2501n – 质量模型类国际标准
- （3）ISO/IEC 2502n – 质量测量类国际标准
- （4）ISO/IEC 2503n – 质量需求类国际标准
- （5）ISO/IEC 2504n – 质量评估类国际标准

3.2.6 ISO/IEC 25012

ISO/IEC 25012 是数据质量模型的国际标准。该国际标准定义了一个通用的数据质量模型，可用于以结构化方式存储数据的计算机系统。该标准把数据质量作为一个计算机系统的一部分，并对用户和系统使用的目标数据定义了质量属性。

目标数据是指组织决定通过数据质量模型来分析和验证的那些数据。非目标数据包括两类：第一类是指非持久性数据，例如由操作系统处理的数据；第二类是指包含在 IEC 25012 内，但组织选择不应用

该标准的数据。
应用数据质量模型的目标数据范围可参考下图。

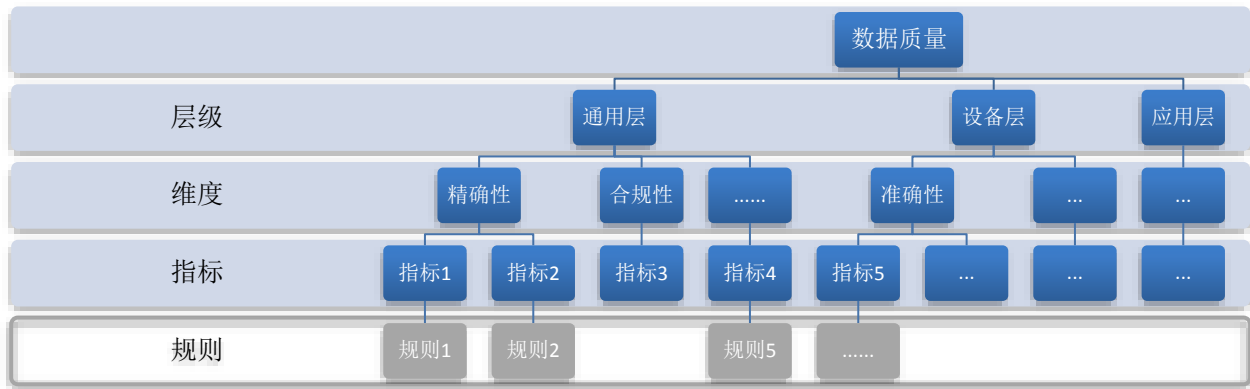


本指南涉及的质量和评估的目标数据，参考附录 1 至附录 7 中定义的智能系统的元数据。

第3节 船舶数据质量评估框架

3.3.1 数据质量总体框架

对于各种以数据为基础的应用，决策功能不同，所需性能和可靠性也不同。对于不同的应用，数据质量的需求是不同的。



3.3.2 数据质量层级

为了适应海洋和海上数据的不同级别的数据质量评估需要，建立了一个分层的数据质量评估框架，以符合 ISO 8000-8 中制定的三个数据质量类别。除层级 I 外，每一层级的数据质量应建立在上一层的基础上。

这三个层级与 ISO-8000-8 关系如表 3.3.1 所示：

表 3.3.1

层级		ISO 8000-8	
		类别	目标
I	通用层	语法	数据的结构
II	设备层	语义	数据的含义

III	应用层	语用	数据的使用
-----	-----	----	-------

三个层级评估之间的关系如图 3.3.1:

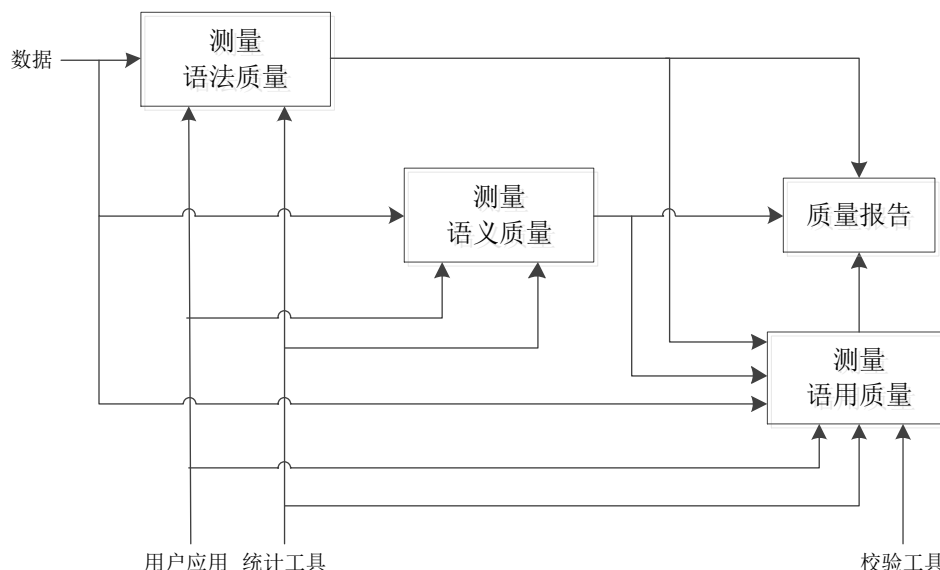


图 3.3.1

第 I 层 通用层

验证数据符合其特定语法的程度。评估目标是一致性，其中数据库/存储中特定数据元素的数据值必须符合元数据表示的预定义数据内容和结构（例如，数据层次结构，数据字段类型，数据格式，值域，精度等）。通用级别的质量检查可以应用于海洋和海上应用中的任何物联网数据。

第 II 层 设备层

验证数据与数据表示的对应程度（例如，测量如何准确反映所感测的物理值）。评估目标是数据的全面性和准确性。传感器/设备级质量检查需要了解传感器，设备和系统的工作原理，测得的物理和环境参数以及操作条件。

第 III 层 应用层

验证为特定目的而发现数据合适且值得的程度。评估目标是数据在给定用途下的可用性和有用性。可用性旨在验证数据使用者是否可以轻松访问，有效地检索和操纵数据库/存储中的数据。有用性旨在验证数据是否可以支持数据应用程序中的使用者完成任务。除了商业和技术领域的支持之外，应用程序级质量检查可能还需要 IT 和数据分析技能和知识。这可能包括确认数据可访问，并解释所采用的数据分析模型得出的不清楚的物理关系，健康状况和/或复杂的物理和操作条件。

除专业知识外，应用程序级别检查可能需要 IT 和数据分析技能。这可能包括确认数据易于访问，解释数据模型表示的物理关系和/或复杂的物理和操作条件。

3.3.3 数据质量评估维度

SQuaRE 数据质量的维度与 ISO 8000-8 的关系如下:

数据质量层级	I 通用层	II 设备层	III 应用层
--------	-------	--------	---------

<p>评估维度 (典型)</p>	<p>精确性 合规性</p>	<p>完整性 准确性 一致性 可信性 可理解性</p>	<p>时效性 及时性 效率性 可追溯性 可用性 可移动性 可恢复性</p>
----------------------	------------------------------	---	---

数据质量评估可包括如下维度：

·准确性（Accuracy）：数据具有正确表示特定使用上下文中概念或事件的预期属性的真实值的属性的程度。

·完整性（Completeness）：数据与实体在特定使用上下文中具有所有预期属性和值的满足程度。

·一致性（Consistency）：数据的属性在某种程度上是不矛盾的，并且在特定的上下文中与其他数据一致使用。它可以是有关一个实体的数据中的一个，也可以是可比实体的相似数据之间的一个或两个。

·可信性（Credibility）：数据的属性在特定上下文中被用户视为真实可信的程度。

·时效性（Currency）：数据反映其所建模的当前真实世界的程度，数据时效性度量了数据的“新鲜程度”以及在时间变化中的正确程度，可以根据数据元素刷新的频率度量数据的时效性，从而验证数据是最新的。数据时效性规则订了一个数据值在失效或需要更新之前已经历的“寿命”。

·及时性（Timeliness）：体现为数据从所需的时间点到数据真实可用的时间偏差程度。

·合规性（Compliance）：在特定使用环境中，与数据质量相关数据属性符合现行标准、惯例或法规的程度。

·效率性（Efficiency）：在特定的使用环境中，通过使用适当数量和类型的资源，数据具有可被处理并提供预期性能水平的属性的程度。

·精确性（precision）：数据在多大程度上具有精确的属性或在特定的使用环境中提供区分的属性。

·可追溯性（Traceability）：数据具有的属性对数据的访问以及在特定使用环境中对数据所做的任何更改的跟踪。

·可理解性（Understandability）：数据具有的属性，使其能够被用户读取和解释，并在特定的使用上下文中以适当的语言，符号和单位表示。

·可用性（Availability）：数据具有被授权用户和/或应用程序能够在特定上下文中检索数据的属性。

·可移动性（Portability）：数据可以在特定的使用环境中安装，替换或从一个系统迁移到另一个系统，从而保持现有质量。

·可恢复性（Recoverability）：在发生故障的情况下，数据也可以在特定的使用上下文中保持和保持指定级别的操作和质量。

·有效性（Validity）：数据满足其语法定义（格式、类型、范围等属性）的程度。

·连贯性（Integrity）：同一个数据集中，或多个数据集中，某一行中的数据与另一行的数据之间的预期关系的存在的程度。

·合理性（Plausibility）：算法使用补偿值（比如，内插或矫正数值）的程度。

·重复性（Duplication）：数据在数据集中没有重复记录的程度。根据该数据元素的标识方式，在数据集中不会记录任何内容。

第4节 数据质量测量规则和指标

数据质量验证规则用于数据质量的测量。需根据数据的使用，以及和具体数据应用关联的潜在数据

质量问题的经验和理解，制定数据质量评估规则。

3.4.1 准确性

数据具有正确表示特定使用上下文中概念或事件的预期属性的真实值的属性的程度，主要包括语法准确性和语义准确性。

语法准确性定义为一组值与在语法上认为正确的数据值的接近程度。

语义准确性定义为一组值与在语义上认为正确的数据值的接近程度。

测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为数据的记录总数

R_p 为满足定义的准确度的数据记录总数

DQ 为数据准确性的测量结果，以百分比形式记录数据质量准确性。

3.4.2 完整性

数据与实体在特定使用上下文中具有所有预期属性和值的满足程度。

通过测量没有空白值或者非控制值存在的比例，判断是不是所有必要的数据都存在。

测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为上下文特定用途范围内的数据的记录数

R_p 为特定上下文所需的数据记录数

DQ 为数据完整性的测量结果，以百分比形式记录数据质量完整性。

3.4.3 一致性

数据的属性在某种程度上是不矛盾的，并且在特定的上下文中与其他数据一致使用。它可以是有关一个实体的数据中的一个，也可以是可比实体的相似数据之间的一个或两个。

通过分析列表中显示的结构和格式模式一致性，和/或同一时间序列数据集中的采样频率分析一致性。

测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为文件中数据的记录数

R_p 为数据一致的记录数

DQ 为数据一致性的测量结果，以百分比形式记录数据质量一致性。

3.4.4 可信性

数据的属性在特定上下文中被用户视为真实可信的程度。

可信性包括真实性的概念，包括来源、属性、承诺等真实性。被一个独立且可信的组织认证过的数据可认为具备可信性。

测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为用于获取信用风险信息的数据的记录数

R_p 为内部审计校验过的数据的记录数

DQ 为数据可信性的测量结果，以百分比形式记录数据质量可信性。

3.4.5 时效性

数据反映其所建模的当前真实世界的程度，数据时效性度量了数据的“新鲜程度”以及在时间变化中的正确程度，可以根据数据元素刷新的频率度量数据的时效性，从而验证数据是最新的。数据时效性规则订了一个数据值在失效或需要更新之前已经历的“寿命”。

测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为指定字段的数据值的数据检查次数

R_p 为数据值满足时效性需求的被检查的数据检查次数

DQ 为数据时效性的测量结果，以百分比形式记录数据质量时效性。

3.4.6 及时性

体现为数据从所需的时间点到数据真实可用的时间偏差程度。

通过比较数据的期望时间点（标准时间戳）与数据的实际可用时间点（实际时间戳）之间的偏差分析数据及时性。

测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为文件中的数据记录数

R_p 为满足期望数据时间的数据记录数

DQ 为数据及时性的测量结果，以百分比形式记录数据质量及时性。

3.4.7 合规性

在特定使用环境中，与数据质量相关数据属性符合现行标准、惯例或法规的程度。

测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为文件中的数据记录数

R_p 为满足数据规范的数据记录数

DQ 为数据合规性的测量结果，以百分比形式记录数据质量合规性。

3.4.8 效率性

在特定的使用环境中，通过使用适当数量和类型的资源，数据具有可被处理并提供预期性能水平的属性的程度。

测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为文件中的总记录数

R_p 为文件中的数据记录数

DQ 为数据效率性的测量结果，以百分比形式记录数据质量效率性。

3.4.9 精确性

数据在多大程度上具有精确的属性或在特定的使用环境中提供区分的属性。

数字数据可能需要精确到几个有效数字。

分析数据准确度水平和四舍五入的测量值，获取数据的精确性。

测量规则和指标：

$$DQ = R_p / R_t$$

其中, R_t 为文件中的数据记录数

R_p 为满足期望数据精度的数据记录数

DQ 为数据精确性的测量结果, 以百分比形式记录数据质量精确性。

3.4.10 可追溯性

数据具有的属性对数据的访问以及在特定使用环境中对数据所做的任何更改的跟踪。

测量规则和指标:

$$DQ = R_p / R_t$$

其中, R_t 为文件中的测试过追溯性的数据记录数

R_p 为满足期可追属性的数据记录数

DQ 为数据可追溯性的测量结果, 以百分比形式记录数据质量可追溯性。

3.4.11 可理解性

数据具有的属性, 使其能够被用户读取和解释, 并在特定的使用上下文中以适当的语言, 符号和单位表示。

有些关于数据可理解性的信息包含在元数据中。

测量规则和指标:

$$DQ = R_p / R_t$$

其中, R_t 为文件中的数据记录数

R_p 为具有可自动链接到相关数据的元数据的记录数

DQ 为数据可理解性的测量结果, 以百分比形式记录数据质量可理解性。

3.4.12 可用性

数据具有被授权用户和/或应用程序能够在特定上下文中检索数据的属性。

一个典型的数据可用性是多个用户和/或程序的同步访问, 包括读取或更新数据。

另一个数据可用性是一定时间间隔内数据可用的能力。

测量规则和指标:

$$DQ = R_p / R_t$$

其中, R_t 为备份或恢复程序的数据项数

R_p 为备份或恢复时可用的数据项数

DQ 为数据可用性的测量结果, 以百分比形式记录数据质量可用性。

3.4.13 可移动性

数据可以在特定的使用环境中安装, 替换或从一个系统迁移到另一个系统, 从而保持现有质量。

测量规则和指标:

$$DQ = R_p / R_t$$

其中, R_t 为移动的数据数量

R_p 为移动到一个不同的计算机系统后保持当前数据质量属性的数据数量

DQ 为数据可移动性的测量结果, 以百分比形式记录数据质量可移动性。

3.4.14 可恢复性

在发生故障的情况下，数据也可以在特定的使用上下文中保持和保持指定级别的操作和质量。
可恢复性可由一些属性提供，如提交/同步点，回滚或备份恢复机制。
测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为备份/恢复程序中的数据项数

R_p 为备份/恢复操作后，成功备份/恢复的数据项数

DQ 为数据可恢复性的测量结果，以百分比形式记录数据质量可恢复性。

3.4.15 有效性

数据满足其语法定义（格式、类型、范围等属性）的程度。
测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为文件中的数据项数

R_p 为被认定为有效的数据项数

DQ 为数据有效性的测量结果，以百分比形式记录数据质量有效性。

3.4.16 连贯性

同一个数据集中，或多个数据集中，某一系列中的数据与另一系列的数据之间的预期关系的存在的程度。
测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为文件中的数据项数

R_p 为包含重要关系链接的数据项数

DQ 为数据连贯性的测量结果，以百分比形式记录数据质量连贯性。

3.4.17 合理性

算法使用补偿值（比如，内插或矫正数值）的程度。
应记录更正的方法和基本原理。
测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为文件中的数据项数

R_p 为使用补偿值的数据项数

DQ 为数据合理性的测量结果，以百分比形式记录数据质量合理性。

3.4.18 重复性

数据在数据集中没有重复记录的程度。根据该数据元素的标识方式，在数据集中不会记录任何内容。
对于时间序列数据，具有相同值但不同时间戳的数据记录不被视为重复。
测量规则和指标：

$$DQ = R_p / R_t$$

其中， R_t 为文件中的数据项数

R_p 为数据值唯一的数据项数

DQ 为数据重复性的测量结果，以百分比形式记录数据质量重复性。

第5节 船舶数据质量评估流程

3.5.1 船舶数据质量评估基本流程

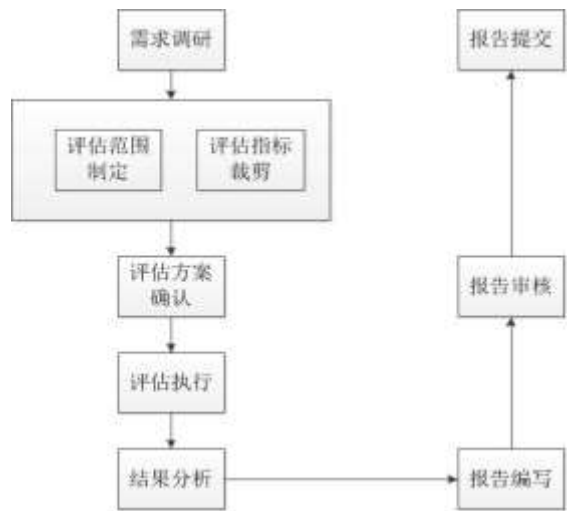


图 3.5.1 数据质量评估流程图

3.5.2 数据质量评估需求分析

数据质量需求分析通常包括数据集的识别，数据用户和主题专家的调研，初步识别数据质量问题，定义适当的数据质量规则、指标和维度，并设定质量目标等过程。主要步骤如下：

（1）选择/识别数据集

通过使用经典的统计分析方法（如数据分析或类似技术）来选择/识别用于初始评估的数据集（如待评估的数据集或类似数据应用程序生成的历史数据集的小型或全面数据集）。整理和分析数据集的过程有助于识别数据特征（如数据类型，格式，精度，范围等），发现潜在的数据质量问题和确定数据质量验证规则。

（2）初步分析和评估

① 收集数据使用用户的需求，并在数据专家的帮助下，通过对确定的数据集进行数据分析来识别出潜在的数据质量问题/异常，以及由硬件和软件引起的数据收集、传输和存储中的潜在问题。

② 收集数据相关方在技术和业务上的意见，评估已识别的数据质量问题对潜在数据使用的潜在影响（如对数据准确性和可靠性的影响以及最终对数据应用程序整体功能的影响）。

（3）定义质量需求

① 根据已识别的数据质量问题及其潜在影响来定义数据质量需求。需求定义包括数据质量验证规则，可衡量的指标和维度。

② 为每个数据质量评估级别定义阈值，加权因子和接受级别：指标，维度和整合。

3.5.3 数据质量评估规则 and 标准

（1）验证规则 and 标准

① 通过将定义的数据质量验证规则应用于一个测试数据集，而不是用于初步评估的全部数据集，来测试和验证定义的数据质量规则。

② 与用户一起查看已定义的数据质量验证规则，以确保用户理解这些规则。

（2）管理规则 and 标准

- ① 数据验证规则应以一致的格式记录，并使用清晰的自然语言描述。
- ② 应根据可衡量的数据质量指标和维度来定义数据验证规则。
- ③ 创建规则时应考虑数据的使用。已定义的数据质量验证规则需要根据实际数据集进行测试。建议在整个数据质量改进生命周期中不断完善验证规则。
- ④ 数据使用者和主题专家应该参与数据质量验证规则的制定。定义的规则应由数据使用者（例如监督数据使用的人）和主题专家确认。

3.5.4 数据质量评估

（1）评估数据质量

- ① 根据数据质量验证规则以及定义的可接受性和阈值，评估数据质量级别。
- ② 记录不符合项级别和观察到的数据质量问题。

（2）确定潜在的改进及其优先级

- ① 根据对数据使用的影响，对已知数据质量问题的重要性进行优先级排序，并评估改进方案以解决数据质量问题。
- ② 结合全面的数据分析和相关方的输入，优先考虑修复和改进工作。
- ③ 识别需要深入分析根本原因的数据质量问题以及可能的改进方案。

（3）制定数据问题处理流程

- ① 检查已确定的数据质量问题，并在数据使用者和主题专家的协助下发现问题的潜在根本原因。
- ② 解决非技术性根本原因：为数据处理人员提供适当的培训；改善数据处理流程；加强领导支持；建立明确的问责制和所有权。
- ③ 解决技术性根本原因：直接纠正有缺陷的数据；改善数据采集、传输和存储的性能；修改系统和技术流程，以防止问题再次发生；在平衡数据质量问题和纠正/改进的成本之后，持续监控并且不立即采取任何措施。

（4）解决数据质量问题

- ① 进行成本效益分析，以比较确定的潜在补救方案。应当正的投资回报率以进行改进。
- ② 咨询数据使用者和主题专家的建议，以选择最佳选项来解决问题，包括简单补救措施和根本的原因补救。
- ③ 简单的补救措施是指直接在记录中修复和更正数据（例如，数据清理/解析/格式化）。
- ④ 根本的原因补救是指针对战略变更（例如系统的修改）制定长期改进计划。它着重于修改系统以解决根本原因，首先建立机制以防止问题。预防通常比纠正更能节省成本。
- ⑤ 制定并实施补救计划，该计划旨在重新评估已修正数据集的质量水平，并确保所应用的更改不会引入其他错误，并能按预期执行。

（5）监控数据与验证规则的持续符合性

- ① 监控和趋势化数据与验证规则的持续符合性，并报告所有数据质量评估级别。
- ② 记录与制定数据质量验证规则相关的指标，维度，总体级别和数据质量得分情况。
- ③ 记录数据质量随时间变化关系。
- ④ 基于补救计划，跟踪数据质量问题的处理。
- ⑤ 设置时间间隔，以根据定义的规则定期评估质量。
- ⑥ 每次测量，设置阈值或接受标准。数据质量结果通常反应正确数据的百分比（通过验证规则）或异常的百分比（不通过验证规则），具体取决于使用的公式。如果数据符合定义的数据质量验证规则，请确认数据适合其预期的应用（例如，数据分析）。当数据不符合定义的数据质量验证规则时，及时通知和警报数据质量问题，并根据指定的补救计划，建议可能采取的措施。

3.5.5 数据质量评估报告

数据质量评估报告包括但不限于以下内容：

- (1) 评估数据集
- (2) 评估规则 and 标准
- (3) 数据质量评估得分
- (4) 数据质量可视化
- (4) 不符合项级别和数据质量问题
- (5) 数据质量问题分析 and 处理
- (6) 数据质量问题解决

3.5.6 数据质量持续改进

可以通过开始一个新的周期来实现数据质量的持续改进。新的数据质量改进周期可以根据以下情况重新开始：

- (1) 现有数据质量结果低于接受标准
- (2) 有新的待分析的数据集
- (3) 对现有数据集采用新的数据质量需求

第6节 船舶数据质量评估结果分析

3.6.1 数据质量维度测量

结合第 3.3.4 小节定义的数据质量维度范围和第 3.3.5 小节中定义的数据质量规则，测量每一种数据在每一个质量维度上的结果。

某一个质量维度上的测量结果可参考但不限于如下方法：

$$DQ_i = R_{pi} / R_{ti}$$

其中， R_{ti} 为第 i 种数据的记录总数
 R_{pi} 为第 i 种数据的通过数据质量规则的总数
 DQ_i 为第 i 种数据在该质量维度上的测量结果，以百分比形式记录数据质量。

数据质量维度的选取依赖于数据的使用需求，建议由数据使用者和评估专家共同确认。

3.6.2 数据质量维度权重

结合第 3.3.4 小节定义的数据质量维度范围和第 3.3.5 小节中定义的数据质量规则，为每一种数据的每一个质量维度制定权重。

$$\sum_{i=1}^n w_i = 1$$

其中， w_i 表示第 i 个质量维度的权重。
 n 表示某一种数据采用的数据质量维度总数。

某一个数据质量维度权重示例如下表所示：

表 3.6.2.1

数据质量维度	权重（%）	测量结果（%）
维度1	10	100
维度2	30	85

维度3	50	90
维度4	10	95
总计	100%	90

数据质量维度的权重依赖于数据的使用需求，建议由数据使用者和评估专家共同确认。

3.6.3 数据质量指标

3.6.3.1 数据质量评估需设置数据质量指标，该指标用于评估一个数据的数值质量的好坏或可接受程度，并依据此质量，可判断数据质量的一个变化趋势，并依此采取相应处理措施，持续提升数据质量。

3.6.3.2 数据质量评估指标的设立可以通过或不通过，可以是好、中和坏，也可以是一定分数范围内的具体数值。对于前两者，需要设定一个阈值，用于数据质量的评估结果划分。

3.6.3.3 数据质量评估指标可以为每一个数据质量维度设置可接受的阈值，也可以为某一个数据的总体测量结果设置一个可接受的阈值。

第4章 数据安全技术要求

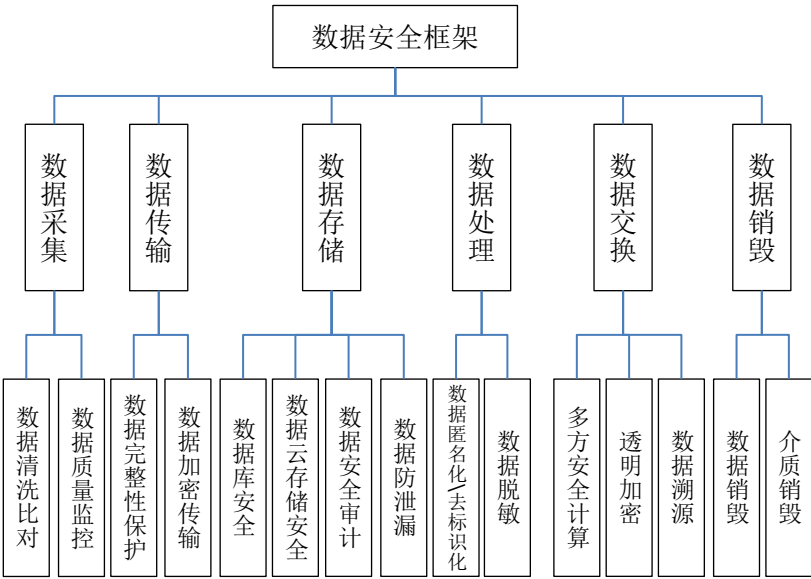
第1节 一般规定

4.1.1 一般要求

- 4.1.1.1 数据安全的总体目标是确保数据的机密性、完整性和可用性。
- 4.1.1.2 数据出境应满足船旗国及主管机关相关法律法规要求。
- 4.1.1.3 船舶数据交换可参考 ISO19645、ISO19847、ISO19848 相关技术条款。
- 4.1.1.4 船舶数据安全至少应满足《船舶网络系统要求及安全评估指南》第三章第 5 节相关要求。

第2节 数据安全框架

4.2.1 数据安全框架



第3节 数据采集

4.3.1 数据采集活动概念

4.3.1.1 数据采集的目的是获得数据，数据采集方式包括但不限于

- (1) 传感器采集数据。包括船舶各个系统在运行过程中采集和产生的数据。
- (2) 通信网络数据。通过船-船，船-岸通讯获取的数据。
- (3) 系统数据。组织内部的系统在运行过程中采集和产生的业务数据，以及各种系统、程序和服务运行产生的大量运维和日志数据等。

4.3.1.1 数据采集活动主要操作包括但不限于：发现数据源、传输数据、生成数据、缓存数据、创建元数据、数据转换、数据完整性验证等。

4.3.2 安全要求

4.3.2.1 进行数据采集活动时，安全要求应满足：

- (1) 定义采集数据的目的和用途，明确数据采集源和采集数据范围；
- (2) 遵循合规原则，确保数据采集的合法性、正当性和必要性；
- (3) 遵循最小化原则，只采集满足业务所需的最少数据；
- (4) 遵循质量保障原则，制定数据质量保障的策略、规程和要求；
- (5) 遵循确保安全原则，对采集的数据进行分类分级标识，并对不同类和级别的数据实施相应的安全管理策略和保障措施。对数据采集环境、设施和技术采取必要的安全管控措施。

第4节 数据存储

4.4.1 数据存储活动概念

4.4.1.1 数据存储指将数据静态保存在数据平台，存储的数据包括采集的数据、分析和处理的结果数据等。存储系统可以是关系数据库、非关系数据库等，应支持对不同类型和格式的数据存储，且提供多种数据访问接口，如文件系统接口、数据库接口等。直到数据被彻底删除之前，存储的数据均应由组织提供恰当的安全保护。

4.4.1.2 应充分考虑使用第三方数据存储平台保存数据的安全风险。由于知识产权、法律法规等原因，组织即使能对存储系统中的数据如个人信息或健康数据等进行有效控制，但可能不是数据的拥有者，组织仍需承担数据的存储管理责任。

4.4.1.3 数据存储活动的主要操作包括但不限于：数据编码、数据加解密、冷热数据分级存储、数据归档、数据备份、数据更新、数据访问等。

4.4.2 安全要求

4.4.2.1 组织开展数据存储活动时，应满足：

- (1) 将不同类别和级别的数据分开存储，并采取物理或逻辑隔离机制。
- (2) 遵守确保安全原则，主要考虑以下几个方面：
 - ① 存储架构安全；
 - ② 逻辑存储安全；
 - ③ 存储访问控制；
 - ④ 数据副本安全；
 - ⑤ 数据归档安全；
 - ⑥ 数据时效性管理。
- (3) 建立数据存储冗余策略和管理制度，及数据备份与恢复操作过程规范。
- (4) 船舶网络系统要求及安全评估指南第三章第5节相关要求。

第5节 数据处理

4.5.1 数据处理活动概念

4.5.1.1 数据处理活动指通过数据分析和数据可视化等技术从数据中提取信息，提炼出有用知识和

价值的系列操作。

4.5.1.2 数据处理活动的主要操作包括但不限于：数据查询、数据读取、数据索引、批处理、交互式处理、流处理、数据统计分析、数据预测分析、数据关联分析、数据可视化、生成分析报告等。

4.5.2 安全要求

4.5.2.1 组织开展数据处理活动时，应：

- (1) 依据个人信息和重要数据保护的法律法规要求，明确数据处理的目的和范围。
- (2) 建立数据处理的内部责任制度，保证分析处理和使用数据不超出声明的数据使用目的和范围。
- (3) 遵循最小授权原则，提供数据细粒度访问控制机制。
- (4) 遵循确保安全原则，主要考虑以下几个方面：
 - ① 分布式处理安全；
 - ② 数据分析安全；
 - ③ 数据加密处理；
 - ④ 数据脱敏处理；
 - ⑤ 数据溯源。
- (5) 遵循可审计原则，记录和管理数据处理活动中的操作。
- (6) 对数据处理结果进行风险评估，避免处理结果中包含可恢复的敏感数据。

第6节 数据分发

4.6.1 数据分发活动概念

4.6.1.1 数据分发活动指将原始数据、处理的数据等不同形式的数据传递给组织内部其他角色、外部实体或公众等。数据分发包括线上或线下等多种方式。

4.6.1.2 数据分发的原因包括但不限于：

- (1) 组织内部部门间的数据交换；
- (2) 为外部生成报告，例如政府部门的统计数据；
- (3) 企业间的数据交换，为客户提供使用报告等；
- (4) 数据出售给其他组织；
- (5) 业务实现需求。

4.6.1.3 数据分发涉及的主要操作包括但不限于：数据传输、数据导出、数据交换、数据交易、数据共享等。

4.6.2 安全要求

4.6.2.1 组织开展数据分发活动时，应：

- (1) 遵循责任不随数据转移原则。
- (2) 在数据分发前，对数据进行风险评估，确保数据分发后的风险可承受，并通过合同明确数据接收方的数据保护责任。
- (3) 在数据分发前，对数据的敏感性进行评估，根据评估结果对需要分发的敏感信息进行脱敏操作。
- (4) 遵循可审计原则，记录时间、分发数据、数据接收方等相关信息。
- (5) 评估数据分发中的传输安全风险，确保数据传输安全。
- (6) 提供有效的数据安全共享机制。
- (7) 建立数据发布的审核制度，明确数据发布的内容和范围对发布的数据开展定期审核。

第7节 数据共享

4.7.1 数据共享活动概念

4.7.1.1 数据共享环节涉及向第三方提供数据、对外披露数据等不同业务场景。

4.7.1.2 船舶、集成商、供应商、船东及船舶管理公司向境外组织共享数据、对外披露数据应满足船旗国相关法律法规要求。

4.7.2 安全要求

在执行数据共享安全相关管理制度规定的同时，可以建设统一数据分发平台，与数据安全域技术结合，作为数据离开数据安全域的唯一出口，进而在满足业务需求的同时，有效管理数据共享行为，防范数据遭窃取、泄漏等安全风险。统一数据分发平台需要整合所有数据共享业务场景，例如基于原始数据的处理分析结果向第三方共享、数据分析报告下载到办公终端设备等情形，对每一类数据共享场景实现差异化的线上审批流程。

第8节 数据删除

4.8.1 数据删除活动概念

4.8.1.1 数据删除活动指组织删除自有或租用的大数据平台上的数据及其副本。如果数据来自外部实时数据流，还应断开与实时数据流的链接。

4.8.1.2 数据被删除的原因包括但不限于：

- (1) 为了减少数据泄露的风险。避免数据被不适当的分发或处理。
- (2) 删除不相关或不正确的数据。数据与最初使用目的不再相关，或数据不正确。
- (3) 业务完成后的数据删除处理。数据业务完成服务目标，不再需要保存相关数据。
- (4) 满足客户的数据删除要求。法律法规要求保留的数据除外。

4.8.1.3 数据删除活动的主要操作包括但不限于：删除元数据、删除原始数据及其副本、断开与外部实时数据流的链接、删除数据的访问接口、不可恢复的数据销毁等。

4.8.2 安全要求

4.8.2.1 组织开展数据删除活动时，应：

- (1) 删除超出数据留存期限的相关数据，对留存期限有明确规定的，按相关规定执行；
- (2) 依照数据分类分级建立相应的数据删除机制，明确需要进行数据销毁的数据、方式和要求，明确销毁数据范围和流程；
- (3) 遵守可审计原则，建立数据删除策略和管理制度，记录数据删除的操作时间、操作人、操作方式、数据内容等相关信息。

附录 1 智能航行系统所应满足的元数据要求 略.

附录 2 智能机舱系统所应满足的元数据要求 略

附录 3 智能能效系统所应满足的元数据要求 略

附录 4 智能船体系统所应满足的元数据要求 略

附录 5 智能货控系统所应满足的元数据要求 略

附录 6 智能集成平台所应满足的元数据要求 略

附录 7 远程控制站所应满足的元数据要求 略