

海洋环境调查资料分类与编码研究

刘志杰¹, 孔敏¹, 舒雨婷¹, 崔晓健¹, 杨扬¹, 吴春芳²

(1. 国家海洋信息中心 天津 300171; 2. 自然资源部第二海洋研究所 杭州 310012)

摘要: 为确保海洋环境调查资料在汇集、处理、管理和服务全过程中语义表达的一致性, 需采用统一分类和编码技术, 设计资料分类编码体系, 以满足海洋环境调查资料的汇集、管理与服务需求。在对海洋调查资料分类现状研究的基础上, 系统分析了分类影响因素, 提出了分类原则和方法, 初步构建了基于学科、数据类型和调查方式为主的海洋调查资料分类体系, 并对其进行规范化编码。目前该分类体系与编码已应用到海洋环境资料和信息管理业务化工作中。

关键词: 海洋环境资料; 分类; 编码; 管理; 海洋信息活动

中图分类号: P731. 11 文献标志码: A 文章编号: 1005—9857(2020)01—0034—05

Classification and Coding of Oceanographic Survey Data

LIU Zhijie¹, KONG Min¹, SHU Yuting¹, CUI Xiaojian¹, YANG Yang¹, WU Chunfang²

(1. National Marine Data and Information Service, MNR, Tianjin 300171, China;
2. Second Institute of Oceanography, MNR, Hangzhou 310012, China)

Abstract: In order to ensure the consistency of semantic expression of marine environmental survey data in the whole process of collection, processing, management and service, it is necessary to adopt unified classification and coding technology and design data classification coding system to meet the requirements of collection, management and service of marine environmental survey data. On the basis of research on the classification status of marine survey data, this paper systematically analyzed the classification influencing factors, proposed the classification principles and methods, constructed a five-level classification system based on disciplines, data types and survey methods, and carried out standardized coding for it. The classification system and coding had been applied to the operational management of marine environment data management.

Key words: Marine environmental data, Classification, Coding, Management, Marine information activities

0 引言

海洋调查资料分类编码是从不同角度对资料进行抽象和分类, 以便对资料进行更好的概括, 是

海洋调查资料管理的一项基础性工作, 关系到海洋调查资料汇集、处理、管理和共享各个方面。中华人民共和国成立以来, 我国海洋事业进入大发展时

收稿日期: 2019-07-02; 修订日期: 2019-12-19
基金项目: 全球变化与海气相互作用专项(GASI).
作者简介: 刘志杰, 高级工程师, 博士, 研究方向为海洋地质资料管理与处理技术

期,以专项调查、综合科考、资源勘探和专题研究为目的,先后在我国近海及深海大洋区域开展了一系列调查活动,取得了丰富的海洋调查资料和信息。随着科技进步,海洋调查技术日新月异,水下滑翔机、海底地震仪(OBS)、自制无人深潜器(AUV)、缆控水下机器人(ROV)、深海载人深潜器(HOV)等被广泛应用于现代海洋调查活动中,数据类型、种类不断丰富,数据量呈指数级增长。如何对海洋资料进行合理分类和高效管理,一直是海洋资料管理者所面临的一个重要问题。针对不同应用目的,往往会出现多角度多级别的分类标准^[1-3]。这些标准规范在一定程度上为海洋调查资料分类提供了依据,但侧重需求不同,往往自成体系,难以满足海洋调查资料综合管理需要。目前尚未形成面向海洋调查资料管理的科学分类编码体系,从而影响和制约了海洋调查资料的管理和使用效率。本研究在系统分析现有分类现状的基础上,结合已有数据资源,探讨形成了基于学科、数据类型和调查方式为主的数据分类编码体系,以便对海洋数据资源进行合理规划、管理和利用。

1 分类与编码研究现状

海洋调查资料的分类研究实际上贯穿于海洋信息活动的全过程,由于信息活动过程或应用目的不同,海洋调查资料的分类也呈现多样化。比如在海洋调查资料采集、处理阶段,往往以学科为体系,进行各学科资料种类和类型的划分;在共享服务阶段,则会根据服务目的不同,采用不同组织分类,总体上更注重资料查询检索实效^[4]。随着海洋信息化进程的加快,我国在海洋信息分类标准化方面开展了深入研究,并形成相应标准规范。2005年原国家海洋局颁布了海洋行业标准《海洋信息分类与代码》^[1],该标准是一个庞杂的分类体系,涉及海洋资源信息、海洋经济统计信息、海洋环境数据信息、海洋基础地理信息、海洋情报文献信息和海洋法规信息,且下延到每一数据类型的数据项,其目的主要为海洋数据库系统建设提供有效支撑。其中海洋环境数据既包含船载调查数据又有业务化观测数据,从数据库建设的角度以学科和数据类型为主要分类依据,对数据要素项进行了统一编码,以保证

数据库中存储和交换的一致性。该标准在历史时期数据库建设过程中发挥了重要作用,但难以适用于前期数据采集与处理。我国国家标准《海洋调查规范第7部分:海洋调查资料交换》^[2]作为海洋调查标准体系的一部分,主要针对海洋调查资料交换内容和记录格式进行了规定,以各学科标准数据集为分类对象,涉及海洋水文、海洋气象、海洋声光、海水化学、海洋生物、海洋地质地球物理六大类。该分类层级更具有概括性,部分调查数据如海洋遥感、沉积化学等学科数据分类尚未囊括其中。2008年,依托国家科技基础条件平台建设项目,为规范、统一海洋信息数据共享发布,国家海洋信息中心编制了《海洋信息分类和分级共享标准》,该标准主要强调数据共享的分类和用户分级,分类体系更为宽泛。此外,2016年颁布的国家标准《海洋要素图式图例及符号》^[5]和2017年颁布的海洋行业标准《极地考察要素分类代码和图式图例》^[3]中所涉及的分分类仍是建立在学科基础上,分别对海洋和极地专题调查要素进行分类研究,主要规定了不同要素图式图例及符号,用于海洋和极地考察专题要素图集及地图产品的制作,侧重于要素的空间化表达和管理。

综上,为满足不同业务应用需求,国内已形成不同的海洋调查资料分类体系,但主要针对某一个专题或一项调查制定^[6]。随着海洋信息化发展和调查技术手段的进步,其分类层次和内容也需要不断更新。目前,尚未建立基于文件的面向海洋调查资料汇集、处理和管理的分类体系与编码,不利于海洋信息资源的高效汇集和统筹管理。为此,在《海洋调查标准体系》^[7]框架指导下,建立海洋调查资料分类与编码,是实现跨部门、跨领域、多源、多时相、多尺度的海洋调查资料整合与管理的前提,将有利于我国海洋调查资料科学规划、管理和后续集成应用。

2 分类与编码原则

海洋环境调查资料分类是一项复杂的工作,涉及学科领域多,数据类型复杂且具有一定的综合性和交叉性。不同学科资料特点显著不同,有的资料种类少,调查手段复杂,如海洋水文资料;有的资料种类繁多,调查手段相对单一,如海洋生物和海洋底质。因此,在分类时要充分考虑不同学科资料的

特点,厘清数据间逻辑关系,并遵循用户使用习惯,结合学科未来发展需求。另外,为便于海洋资料汇集、处理和管理,保持分类的连续性和稳定性,其分类也不宜过细,总体以文件为分类最小基本单元,遵循基本原则如下。

(1)科学性原则:海洋环境调查资料的分类要从实际需求出发,结合数据资源的具体特点,着眼于实用,分类合理,具有可操作性,能够满足数据管理需求。

(2)系统性原则:综合考虑资料主题的一致性,按其内在联系对每一类都进行系统化排列,确保类目唯一、结构层次清晰,既体现数据资源总体架构,又能减少冗余。

(3)规范性原则:资料分类与编码简洁、规范,资料类别采用通用名称,确保语义无二义性。

(4)兼容性原则:充分考虑已有分类基础,尽可能引用相关标准或与之充分衔接。最大限度保留已被人们所熟悉和惯用的框架,便于用户掌握和运用。

(5)可扩充性原则:基于编码对象的属性特点,保证分类对象增加或细化时,已有分类体系能够延续和拓展,确保分类编码稳定,可修订和完善。

3 分类体系构建

任何资料都具有多维属性,海洋调查资料也不例外。资料分类有两个要素:一是分类对象;二是分类依据。资料分类依据即分类粒度问题是分类首要考虑因素,影响到资料分类层级,又与分类目的紧密联系。海洋调查资料门类的划分总体以《海洋调查标准体系》^[7]为依据,从调查内容考虑,分为海洋环境、海洋资源和海洋经济三大部分。本研究即以海洋环境调查资料作为门类分类对象,探讨该门类下资料层级关系。学科、调查方式、资料类型、调查仪器、介质以及数据加工程度和自身属性等都是影响海洋环境调查资料分类的影响因素。从资料汇集和管理角度,重点需要考虑的分类影响因子为学科、资料加工程度、资料类型和调查方式等,再根据分类因子之间的逻辑关系,组成关联有序的资料分类体系。

海洋环境调查通常以项目为主线开展,调查任务常作为资料管理单元,涉及现场采集、室内测试分析、处理解释和研究等多个环节。海洋环境资料

包括资料种类和资料类型两个维度。资料种类的划分适用于所有学科,是根据数据加工程度进行的划分,通常包含任务文档类、原始数据集、整编数据集、标准数据集、报告专著类、图件图集类、图像摄像类、软件类和其他类(表 1),其中原始数据集、整编数据集和标准数据集是调查数据的主体部分,资料类型的划分主要基于数据集,将海洋环境数据细分为大类、中类、小类和子类的嵌套体系。

表 1 海洋环境调查资料种类及编码

资料种类	代码	英文名称	说明
任务文档类	1DO	Documents	任务合同、任务实施方案、航次报告、航次计划等
原始数据集	2RD	Raw dataset	现场原位测试及调查仪器自记观测数据的集合,含参数和格式说明
整编数据集	3CD	Complicating dataset	根据整编技术规范要求进行整理的分析测试及后处理数据的集合,含数据处理与质量评价报告
标准数据集	4SD	Standard dataset	经代码和记录格式转换、质量控制、排重合并等处理后形成的标准文件的集合
报告专著类	5RA	Report & article	研究报告、专著、发表论文等
图件图集类	6AL	Atlas	编制的图件或图集(矢量文件)
图像摄像类	7IC	Image & camera	由专业设备获取的摄像视像及图片等
软件类	8SW	Software	配套软件及说明
其他	9OT	others	除上述范围内数据

海洋环境调查资料大类划分并不等同于学科分类,二者有联系又不完全一一对应,但资料分类一定是建立在学科分类基础上^[8]。海洋环境调查资料学科大类以 2007 年国家颁布执行的《海洋调查规范》系列国家标准^[9]为主要依据,并参考《海洋调查

4 海洋环境资料编码

资料编码是建立在分类基础上,将资料赋予具有一定规律的,易于计算机和人识别处理的符号^[10]。资料种类和类型由于考虑维度不同,应分别进行编码设计。调查资料种类编码共由 3 位编码组成,包括 1 位数字和 2 位资料种类英文缩写码(表 1),与资料类型码无冲突,也可组合使用。资料类型编码基于学科体系建立,主要由门类码、大类码、中类码和小类码组成的有序无含义码,其中门类码用 1 位阿拉伯数字表示,大类、中类和小类码分别由 2 位阿拉伯数字组成(图 1)。个别数据类型分类层级较多,在小类基础上可再分为子类,子类码由 1 位阿拉伯数字表示,与小类码之间用“—”连接(图 1,表 2)。根据资料分类编码的兼容性原则,门类码继承《海洋信息分类与代码》中约定海洋环境门类编码“3”。海洋环境资料分类编码长度由分类粒度决定,编码长度 5~9 位。如某一类资料只分到中类,下无嵌套小类,则编码只到中类即可,长度为 5 位;如数据编码需要分到小类,则编码长度为 7 位;如分到子类,则编码长度为 9 位。以海洋底质沉积物常量元素数据为例,其编码为 3070105-1,从左至右,“3”代表门类码——海洋环境,“07”代表大类码——海洋底质,“01”代表中类码——沉积物,“05”代表小类码——沉积物化学,最后一位“1”则代表子类码——常量元素。为保证编码的灵活性,编码不采用补位方式,编码长度由资料类型决定。

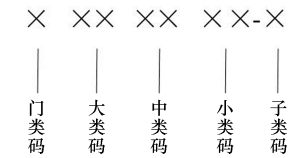


图 1 海洋环境资料编码结构

5 结语

海洋环境资料分类与编码是一项复杂的基础性

工作,其目的是为了更有效地组织和管理海洋环境资料。在海洋资料分类研究基础上,结合已掌握数据资源实际情况,清晰地梳理出海洋环境资料脉络体系,建立海洋环境资料分类与编码,有效避免了资料在使用过程中表达的二义性,也能更好地满足未来调查多学科综合发展的需求,便于自上而下部署和组织数据集的生产和集成^[4],有效提升海洋调查环境资料汇集、处理和管理服务水平。该分类体系是一个动态开放的体系,会随着调查技术水平的提高,资料类型的丰富,不断成熟和完善。

参考文献

[1] 国家海洋局.HY/T 075—2005 海洋信息分类与代码[S].北京:中国标准出版社,2005.

[2] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会.GB/T 12763.7—2007 海洋调查规范第 7 部分:海洋调查资料交换[S].北京:中国标准出版社,2008.

[3] 国家海洋局.HY/T 221—2017,极地考察要素分类代码和图式图例[S].北京:中国标准出版社,2017.

[4] 王卷乐,林海,冉盈盈,等.面向数据共享的地球系统科学数据分类探讨[J].地球科学进展,2014,29(2):265—274.

[5] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会.GB/T 32067—2015 海洋要素图式图例及符号[S].北京:中国标准出版社,2016.

[6] 白降丽,彭道黎,杨馥宁.森林资源信息分类及编码体系研究[J].浙江农林大学学报,2007,24(3):326—330.

[7] 国家海洋局.HY/T 244—2018 海洋调查标准体系[S].北京:中国标准出版社,2019.

[8] 廖顺宝,蒋林.地球系统科学数据分类体系研究.[J].地理科学进展,2005,24(6):93—98.

[9] 《海洋调查规范》系列国家标准[J].中国标准化,2011(5):22—23.

[10] 耿庆斋,张行南,朱星明.基于多维组合的水利科学数据分类体系及其编码结构[J].河海大学学报:自然科学版,2009,37(3):346—350.