智慧海洋多平台数据管理规范研究

李兆钦[1,2],刘增宏[1,2],许建平[1,2],孙朝辉[1,2],卢少磊[1,2]

(1. 自然资源部第二海洋研究所 杭州 310012;2. 卫星海洋环境动力学国家重点实验室 杭州 310012)

摘要:我国未来智慧海洋观测系统将集成多种观测平台,获取全球范围、长时间和准实时的海洋数据。为更加科学而高效地管理数据,文章结合我国智慧海洋发展趋势,分析存在的问题,借鉴相对先进的数据管理规范,为建设智慧海洋国际共享应用平台提供参考。研究结果表明:与观测平台、传感器和通信技术的快速发展相比,我国海洋数据管理的发展十分缓慢,缺乏数据质量控制和存储等规范,严重制约海洋观测数据的有效管理和共享应用,不利于海洋大数据的融合和信息挖掘技术的发展;为各类观测平台制定科学而合理的数据管理规范,对于智慧海洋建设至关重要,亟须尽快组织专业人员成立工作组,落实和保障该项工作的有序开展,从而提升我国在海洋大数据信息处理方面的能力。

关键词:智慧海洋;海洋大数据;数据管理规范;数据共享;海洋观测

中图分类号:P717 **文献标志码**:A **文章编号**:1005-9857(2020)04-0078-06

The Multi-platform Data Management Specification for 'Smart Ocean'

LI Zhaoqin[1,2],LIU Zenghong[1,2],XU Jianping[1,2],SUN Chaohui[1,2],LU Shaolei[1,2]

(1. Second Institute of Oceanography,MNR,Hangzhou 310012,China;
2. State Key Laboratory of Satellite Ocean Environment Dynamics,Hangzhou 310012,China)

Abstract: In the future, China's smart ocean observation system will integrate multiple observation platforms,combining with advanced communication and internet technologies,which makes it possible to get large-scale and long-term oceanic information in near real-time from regional to global scale.In order to manage data more scientifically and efficiently,this paper analyzed the existing problems in combination with the development trend of smart ocean in China, and drew lessons from the relatively advanced data management standards,so as to provide important reference for the construction of the international shared application platform of smart ocean.The research results showed that,the progress of ocean data management was relatively slow compared with the rapid development of observation technology and lack of standardized data quality control and storage,which had greatly hindered the effective management and sharing

收稿日期:2019-09-09;**修订日期**:2020-04-04

基金项目:中国海洋发展研究会与中国海洋发展研究中心联合资助项目"打造智慧海洋国际共享应用平台可行性研究".

作者简介:李兆钦,研究实习员,硕士,研究方向为物理海洋学数据分析

通信作者:刘增宏,高级工程师,研究方向为物理海洋学调查

of observational data and was unfavorable for the integration of ocean big data and the development of information mining technology. It was therefore the drafting of scientific and reasonable data management specifications for various types of observing platforms scientific was essential for the construction of the China′s smart ocean system. This had required great attentions from related departments. It is necessary to establish working groups as soon as possible to conduct related works, thus will enhance the ability to process ocean big data.

Key words: Smart ocean, Ocean big data, Data management specification, Data sharing, Ocean observation

0　引言

随着海洋观测技术和通信技术的快速发展，海洋仪器设备不断更新换代，未来海洋观测将步入智慧海洋的物联网时代，观测方式不再局限于船只走航，而是包含各种固定和移动平台的组网观测，在空间上呈点、线、面分布，观测层次从海表至海底，时间尺度覆盖小时至年甚至年代，结合互联网和计算机技术，实现从海洋信息采集、传输和处理到数据分析、产品制作和信息服务的智能化。当前海洋多平台的组网观测已不难实现，真正亟须解决的关键技术问题是统一化和规范化的数据管理以及多平台的信息融合和挖掘，其中针对多观测平台的数据管理规范是基础和须首先解决的问题[1]。

各种海洋观测平台都有独特的工作模式和数据格式，只有制定科学合理的数据管理规范，才能高效利用这些观测数据，做到有据可依，进而为海洋立体监测和认知海洋提供更多的有效信息。多年来我国已制定部分与海洋数据相关的标准和规范，但很多标准不一致，尚未建立统一的海洋数据规范和标准体系，导致目前海洋数据的质量难以保证、利用率较低以及兼容性和可比性较差[2]。为各种观测平台制定数据管理规范往往须投入大量人力和物力，对于相关人员来说，数据质量控制方法的设计须具备专业知识，数据存储方式和格式以及数据处理流程的制定须同时具备计算机知识和海洋知识。本研究借鉴较成熟的观测平台数据管理规范，结合我国智慧海洋的发展趋势，对多平台数据管理规范进行研究，以期为建设智慧海洋国际共享应用平台提供参考。

1　智慧海洋

美国 IBM 公司于 2008 年率先提出“智慧地球”的概念，其核心是把传感器嵌入各种物理基础设施，连接形成物联网并与互联网进行整合，使人类能以更加精细和动态的方式管理生产和生活，实现全球智慧状态[3]。智慧海洋是基于海洋综合立体感知、互联网实时信息传输、大数据、云计算和知识挖掘等高新技术，以海洋综合感知网、海洋信息通信网和海洋大数据云平台等信息基础设施为主体，搭建海洋信息智能化应用服务群，并建立贯穿各环节的标准质量、运维服务、技术装备和信息安全体系[4]。智慧海洋能力建设包括感知网、通信网、大数据平台和应用群，具备的功能包括智能化信息采集、信息传输、信息处理和信息服务。已建成的美国大洋观测计划(IOO)、美国综合海洋观测系统(IOOS)、加拿大东北太平洋时间序列水下观测网(NEPTUNE)和欧洲 EMSO 观测网等单一或综合观测系统均可视为智慧海洋的初级产品[5-8]。

海洋强国建设离不开智慧海洋建设。智慧海洋建设事关重大战略，事关国家利益，事关长远建设[9]。未来智慧海洋观测系统将集成基于“空天地海”的各种观测平台，自动获取海洋内部各种物理和生物地球化学信息，并通过现代通信网络传输到大数据平台。大数据平台是重要的中间环节，发挥数据接收以及信息解译、处理和分发的作用[10]。由于工作方式和携带传感器不同，每种观测平台具有不同的采样方式和观测数据，其信息处理流程、数据质量控制方法和数据存储格式也不同，须制定相应的数据管理规范，才能合理和有效地利用这些数据，以开展后续的大数据分析和挖掘。

2 数据管理规范

2.1 信息处理流程

每种观测平台都有独特的工作模式和采样方式，并可能使用不同的通信手段传输观测数据，信息处理流程存在差异，因此须详细制定尽可能完善的信息处理流程。在制定信息处理流程的过程中，应有包括平台使用、系统设计、程序编写、质量控制、数据存储和数据使用等在内的相关人员共同参与，商讨流程的总体框架和每个模块实现的功能。其中，总负责人负责流程总体框架的设计和制定，各模块负责人负责各模块的流程设计和制定，数据用户提出数据使用需求，各模块之间还须进行协调。信息处理流程通常包含数据（编码）接收、信息解译、质量控制、格式写入以及数据分发和共享 5 个模块。

中国 Argo 实时资料中心制定的 Argo 剖面浮标信息处理流程如图 1 所示。

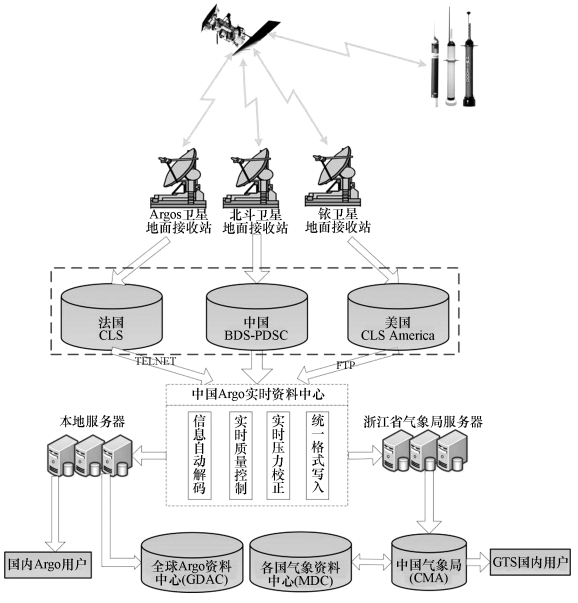

图 1 中国 Argo 实时资料中心的浮标信息处理流程

该信息处理流程包括 3 个部分：①针对 Argos 卫星、铱卫星和北斗卫星 3 种卫星通信系统的浮标数据接收模块；②具有信息自动解码、实时质量控制、实时压力校正和统一格式写入功能的信息处理模块；③负责数据分发的模块。按照该流程，中国 Argo 实时资料中心基于 LINUX 操作系统建立可批量接收、处理和分发各型浮标和传感器观测资料的系统并投入业务运行，这是支撑中国 Argo 实时资料中心运行的核心技术。该系统内的各模块由大量基于 MATLAB 和 Python 语言的代码以及 LINUX SHELL 脚本组成，最终通过 LINUX Crontab 进行任务调度，实现系统程序的定时自动运行[11]。

未来智慧海洋共享应用平台的信息处理流程更加复杂，可参考中国 Argo 实时资料中心的信息处理流程进行规划。

2.2 数据接收

使用不同通信系统传输数据（编码），其数据接收方式不尽相同：①Argos 卫星通信的数据包通常采用 32 字节的十六进制编码，用户可使用 Telnet 协议或 Email 接收数据[12]；②目前使用最广泛的铱卫星通信的数据包通常采用二进制编码，其中 RUDICS 方式支持大数据包传输，SBD 方式支持最大 300 字节的数据包传输，数据接收方式分别为 FTP 协议和 Email[13-14]；③国产北斗卫星通信支持约 100 字节的二进制或十六进制数据包传输，用户须安装北斗终端接收数据[15]；④在近岸无线信号覆盖区域内，4G/5G 无线通信模块支持至少 20 Mb/s 的数据传输速度，用户可使用 TCP/IP 和 M2M 协议接收数据[16]。

由此可见，使用不同通信方式的观测平台须使用不同的通信协议并从不同服务器或数据中心接收观测数据。在制定各种观测平台的数据接收规范时，应综合考虑卫星（无线）通信的速率和用户对数据时效性的要求，明确数据的最低接收频次。同时，针对编程使用的协议提出建议，避免数据包丢失。

2.3 信息解译

通过卫星（无线）传输的观测数据通常以二进制或十六进制编码，因此须严格按照设备研制厂家提供的用户手册解码。即使是同一型号的设备，因设备版本和携带传感器的不同，其数据编码格式也有差别，如美国 Teledyne Webb 公司生产的 APEX 型浮标从 2000 年至今有 10 余种编码格式。使用 Argos 卫星和铱卫星通信传输的浮标数据编码格式

也完全不同，如不严格按照用户手册的编码格式说明进行解码，将无法获得正确的浮标观测数据和技术信息。因此，制定智慧海洋观测平台的信息解译规范对于获取正确的观测数据十分重要[17]。

信息解译规范的制定过程通常包含解码器的编写、测试、试运行和业务运行。解码器编写规范应规定使用的语言和操作系统、解译信息类型和程序结构规范化等内容，应由具备专业计算机知识和海洋知识的人员共同完成；解码器测试规范应规定测试的过程、内容和结果分析以及程序改进等，同样应由具备专业计算机知识和海洋知识的人员共同完成；解码器试运行规范应包含试运行的方式和结果分析以及程序改进等，应由该模块负责人和解码器程序编写人员共同完成；解码器顺利通过试运行方可投入业务运行，投入业务运行的解码器应正式编入信息解译手册，说明该解码器对应观测平台的种类、型号、版本和携带传感器等，并逐一说明程序的输出变量。

信息解译的重要环节是将传感器的输出值或观测平台通过卫星（无线）通信的输出值计算成真正有效的海洋要素值。很多生物地球化学传感器的输出值并不代表观测要素值，须通过给定的方法和实验室标定系数进行计算后，才能获得需要的物理量。例如：使用荧光法测量海水溶解氧浓度的 Aanderaa 溶解氧传感器，其输出值通常为红蓝光的 2 个相位值，须通过相应的方法（如 Stern-Volmer 方程）反演得到真正的海水溶解氧浓度[18]。因此，应在规范中明确规定各种传感器输出值的要素计算方法，并说明计算程序。

信息解译后，须将用户需要的观测数据、设备技术信息、时间和定位等写入文件。这就须针对每种观测平台制定相应的数据存储格式，主要包括存储方式（二进制或可读的文本文件）、文件命名规则、存储时间段、存储格式（包括表头信息、定位、时间和观测数据等）、变量命名规则、数值有效数字、缺省值、采用的单位和数据排列方式。此外，应规范每种观测平台在计算机中的目录组织结构，目录的命名应遵循一定的规则。

2.4　数据质量控制

数据质量控制是在信息获得正确解译后确保数据可靠性的重要步骤，不同观测平台和传感器的数据质量控制方法虽各不相同，但可相互借鉴。质量控制分为实时和延时 2 种模式：实时质量控制完全由计算机自动完成，可保证数据无明显异常值；延时质量控制须由具备专业知识背景的人员进行人工判断，通过赋予质量控制标记符来确定数据质量。无论是实时质量控制还是延时质量控制，所有数据都应保留原始观测值，经延时质量控制的数据还须包含校正值。除观测数据外，定位和时间信息也须进行质量控制，尤其是对于移动观测平台，当其无动力时应满足通过定位和时间信息计算的平台速度不超过海域最大流速，当其有动力时还应考虑平台的前进速度。因此，制定数据质量控制方法是复杂的工作，应由具备专业海洋知识和了解平台运行情况的技术人员共同完成。

在每种观测平台制定数据质量控制方法前，应组建工作小组，负责该平台数据质量控制方法的制定和更新工作。由数据质量控制方法总负责人进行各小组间的协调，目的是保证将各类观测平台具有共性的内容进行统一，如确定质量控制标记符和选取阈值。在制定数据质量控制方法时可借鉴国际较成熟的方法，如水下滑翔机和 Argo 具有相似的工作特点和可携带的传感器[19]，水下滑翔机的观测数据质量控制方法即可借鉴国际 Argo 计划制定的方法。每种观测平台的数据质量控制方法应独立成册，并由工作小组负责更新，对于目前尚无完善数据质量控制方法的传感器观测数据，应随时跟踪国内外最新研究进展，并有计划地更新。数据质量控制手册中应规定质量控制的各项测试步骤、赋予的质控标记以及方法的出处和原理等内容。当数据质量控制方法确定后，可设计相应的程序，以实现观测数据的质量控制，并对每个观测值赋予质量标记符号。每个测试步骤的程序均应进行测试，从而获得手册规定的测试效果。

2.5　数据存储

科研数据的共享要求数据可被科研人员和公众发现、访问、互操作和重复使用，因此数据产出后管理的目的并不仅是数据存储，更重要的是数据流转和利用，这就要求共享数据能以通用或可转换的

格式保存和管理，通用而高效的数据存储方式可方便用户使用数据，并促进数据的推广使用。智慧海洋共享应用平台的建设也须制定通用而高效的数据存储规范，既可节省人力和物力投入，还能提高平台开发应用服务的效率，降低用户的时间成本。在海洋与大气领域，目前国际较通用的数据存储方式有 NetCDF、JSON 和 XML 等，其中：NetCDF 格式应用最为广泛，适合存储网格化和散点数据；JSON 和 XML 为结构性标记语言，易于机器解析和生成，通常用于存储元数据[20-21]。

各类观测平台的数据类型通常分为元数据、观测数据、技术信息和漂移轨迹（移动平台）等。其中，元数据文件存储设备的序列号、固件版本、通信和定位系统、传输平台号、携带的传感器、布放时间、布放位置、布放使用的船只、负责人以及所属项目等信息，观测数据文件存储平台通过卫星（无线）通信传回的输出值、计算后的要素值、时间和定位信息以及质量控制标记符号等内容，技术信息文件存储平台、传感器和电池等的技术参数，漂移轨迹文件存储移动平台在海上漂移时的定位和时间信息以及在水下漂移时测量的各类要素值。

在制定数据存储规范时，须首先制定有关变量的命名和类型、文件的命名和存储方式、各种平台和传感器的型号和生产厂家以及测量要素的命名等规则，使各观测平台的变量名统一，如温度变量名可统一使用 TEMP。数据文件名应体现平台的类型、型号、序列号、投放时间和循环序号等信息，如“glider_seawing1000J003_20190701_001.nc”即代表序号为 1000J003 的“海翼”号水下滑翔机在 2019 年 7 月 1 日投放后观测的第 1 条剖面。对于各类观测要素，须统一确定其数据类型、详细名称、单位、极值范围、分辨率、保留小数位数和缺省值等属性。

各工作小组可根据各观测平台的工作特点和观测要素等信息，制定相应的数据存储规范或格式说明，确定需存储的数据和信息以及每个变量的维数和长度等。例如：某个锚碇浮标在 10 m、50 m、100 m 和 200 m 水深层各安装 1 个温度传感器且观测频率一致，那么其温度变量可定义为 TEMP(N_LEVELS,N_TIME)，其中 N_LEVELS＝4，N_TIME＝UNLIMITED，即代表该浮标的观测层数为 4 层，观测值数量（时间序列）不限定，可随时间的推移不断增加。又如：某个 Argo 剖面浮标在某个循环内观测的盐度可定义为 PSAL(N_PROF,N_LEVELS)，其中 N_PROF＝1，N_LEVELS＝75，即代表该浮标观测 1 条盐度剖面，层数为 75 层（有些浮标携带的 CTD 传感器同时采用散点和连续采样模式，在 1 个循环内可观测 1 条以上温、盐度剖面）。因此，在制定数据存储规范时应综合考虑观测平台的工作模式、携带的传感器和采样模式等情况，提高文件的兼容性，避免多次修订规范。

需要指出的是，对于参与国际共享的数据，如已有数据存储的相应国际规范（如 Argo 计划），应直接采用这些国际规范，从而保证数据存储格式与国际大科学计划无缝衔接。

2.6 数据分发共享

科研数据的开放共享具有重要意义：一方面，可基于前人研究成果，有效提高学术成果质量；另一方面，可避免低效和重复工作，加快创新，并提高科研过程的透明度[22]。通过参与国际大科学计划与其他国家共享和交换观测数据，是未来智慧海洋建设必不可少的环节，也是人类共同应对全球气候变化的必然趋势。因此，我国亟须以国家需求为牵引，深度参与国际合作，科学制定智慧海洋数据分发共享的管理办法和条例，明确可参与国际共享的数据类型和数据共享的分级制度等。例如：对于 Argo 观测，在大洋海域获取的数据资料，可无条件与其他 Argo 计划成员国共享和交换；在我国管辖海域获取的数据资料，可实行有限共享策略和分级共享制度。对于数据分级共享制度的制定，相关部门应牵头成立专家组，并根据专家组的意见做出科学决策。

3 结语

智慧海洋在“空天地海”观测系统的大趋势下应运而生，其中具有国际影响力的智慧海洋共享应用平台是智慧海洋建设工程的重要组成部分。制定针对多观测平台的数据管理规范是重要基础，既能确保观测数据的高效接收、处理、存储和推广应

用，又能有效保证数据质量。制定高效的数据管理规范是非常复杂的工程，须投入大量人力和物力，并须由具备专业计算机知识和海洋知识的人员以及从事观测平台设计和使用的人员紧密配合，加强顶层设计和明确分工。科学而有效的数据管理规范可大大节约智慧海洋共享应用平台业务运行所需的人力和物力资源以及计算成本，更重要的是可提升我国在海洋大数据信息处理方面的能力。因此，该项基础性工作亟须引起相关部门的高度重视，尽快组织专业人员成立工作组，落实和保障该项工作的有序开展。

本研究涉及的数据类型主要包括温度、盐度和压力等物理海洋环境要素以及溶解氧和叶绿素等部分生物地球化学环境要素，但不包括海洋地质、地磁和声学等海洋物理和海洋地质数据。

参考文献

[1] 曹丽娜.海洋大数据管理与应用技术研究[D].舟山：浙江海洋大学，2019.

[2] 杨锦坤，董明媚，武双全.推进我国海洋数据深入共享服务的总体考虑[J].海洋开发与管理，2015，32(3)：68－72.

[3] 徐德明.智慧中国地理空间智能体系研究报告[M].北京：社会科学文献出版社，2014.

[4] 姜晓轶，潘德炉.谈谈我国智慧海洋发展的建议[J].海洋信息，2018(1)：1－6.

[5] COWLES T，DELANEY J，ORCUTT J，et al. The ocean observatories initiative：sustained ocean observing across a range of spatial scales[J]. Marine Technology Society Journal，2010，44(6)：54－64.

[6] OCEAN US. An integrated and sustained ocean observing system(IOOS) for the United States：design and implementation[Z]. 2002.

[7] BARNES C R，BEST M M R，ZIELINSKI A. The NEPTUNE Canada regional cabled ocean observatory[J]. Sea Technology，2008，49(7)：10－14.

[8] FAVALI P，BERANZOLI L. EMSO：European multidisciplinary seafloor observatory[J]. Nuclear Instruments and Methods in Physics Research Section A：Accelerators，Spectrometers，Detectors and Associated Equipment，2009，602(1)：21－27.

[9] 潘德炉.大数据是智慧海洋建设的灵魂[EB/OL].(2018－06－20)[2019－09－01].https://www.sohu.com/a/236889688_335896.

[10] 黄冬梅，赵丹枫，魏立斐，等.大数据背景下海洋数据管理的挑战与对策[J].计算机科学，2016，43(6)：17－23.

[11] 刘增宏，吴晓芬，许建平，等.中国 Argo 海洋观测十五年[J].地球科学进展，2016，31(5)：445－460.

[12] 张少永，林玉池，熊焰.Argos 卫星发射平台研究与 Argos 通讯系统应用[J].海洋技术，2005(1)：25－28.

[13] Iridium Satellite LLC. Iridium short burst data service developers guide，Release 2.0[Z]. 2007.

[14] Iridium Communications. Iridium 9602 SBD transceiver developer's guide(2014)[Z]. 2014.

[15] 张素伟，沈锐.北斗卫星系统在 HM 2000 型剖面浮标中的应用[A].许建平.西太平洋 Argo 实时海洋调查研究文集[C].北京：海洋出版社，2017：388－398.

[16] 王胡成，徐晖，程志密，等.5G 网络技术研究现状和发展趋势[J].电信科学，2015，31(9)：149－155.

[17] 中国 Argo 实时资料中心.严重影响 Argo 资料质量的关键技术：剖面浮标解码器[Z].2017.

[18] 涂梦迪，殷高方，赵南京，等.基于时域荧光寿命的水体溶解氧浓度检测方法[J].光学学报，2018，38(10)：39－44.

[19] WONG A，KEELEY R，CARVAL T. Argo quality control manual for CTD and trajectory data Version 3.2.01 February 2019[Z]. 2019.

[20] LEDOUX H，ARROYO O K，KUMAR K，et al. CityJSON：a compact and easy-to-use encoding of the CityGML data model. Open geospatial data，softw. stand[Z]. 2009.

[21] SCHMIDT A，KERSTEN M，WINDHOUWER M，et al. Efficient relational storage and retrieval of XML documents[A]. International workshop on the world wide web and databases[C]. Berlin，Heidelberg：Springer，2000：137－150.

[22] GUEDJ D，RAMJOUÉ C. European commission policy on open-access to scientific publications and research data in Horizon 2020[J]. Biomedical Data Journal，2015，1(1)：10.11610/bmdj.01102.