

# 国际海洋观测计划数据规范化存储与管理的述评

许艳苹

(厦门大学 近海海洋环境科学国家重点实验室, 福建 厦门 361102)

**摘要:** 基于世界大洋环流实验(World Ocean Circulation Experiment, WOCE)、全球联合海洋通量研究(Joint Global Ocean Flux Study, JGOFS)、全球海洋船基水文调查计划(Global Ocean Ship-based Hydrographic Investigations Program, GO-SHIP)和痕量元素及同位素海洋生物地球化学循环国际研究(An International Study of the Marine Biogeochemical Cycles of Trace Elements and Isotopes, GEOTRACES)四个国际海洋观测计划的科研数据规范化存储与共享管理经验, 分析多学科综合计划如何让获得的海量观测数据进行长期、有效的存储, 具有可传承性, 探讨了建立数据中心和数据仓储对于数据的规范化管理和推进数据共享的重要性。总结出一套数据管理方案: 在项目启动时, 建立项目/数据管理办公室, 推行科研数据标准化管理。针对项目研究主题形成完整的实体参数目录, 统一数据名称、单位、分类及内容, 制定元数据模板等, 数据管理办公室要能存储和共享数据, 并确保数据安全, 在项目成员内部就数据质量的保证措施、数据共享与引用等管理规定达成共识。

**关键词:** 国际海洋观测计划; 数据规范化存储; 数据中心; 数据仓储; 数据共享与引用

**中图分类号:** P73, G203 **文献标识码:** A **文章编号:** 1000-3096(2021)07-0127-10

**DOI:** 10.11759/hyxx20201019002

国际海洋观测计划, 始于 20 世纪 60 年代, 约每 10 年开展一次。1956—1959 年的国际地球物理年, 是海洋观测史上第一次大规模的国际联合行动。1971—1980 年的国际海洋考察十年, 经历 1980—1985 年间的中型计划, 发展为 1985—1990 年的全球变化。物理海洋学观测计划, 经历中大洋动力学实验, 发展为世界大洋环流实验(WOCE)计划。生物地球化学海洋学观测计划, 经历海洋地球化学断面研究(Geochemical Ocean Sections, GEOSECS)计划, 发展为全球联合海洋通量研究(JGOFS)计划<sup>[1]</sup>。在 20 世纪 60—70 年代, 船测的科学数据和资料主要通过出版研究图集的形式公开, 如 GEOSECS 计划。随着全球变化计划的蓬勃发展, 获取了海量高分辨率的观测数据, 完善而有效的数据管理方案尤为重要, 如何让这些数据具有长久的使用性和可传承性, 国际计划如何管理数据, 哪些经验值得借鉴?

本研究选取世界大洋环流实验(World Ocean Circulation Experiment, WOCE)、全球联合海洋通量研究(Joint Global Ocean Flux Study, JGOFS)、全球海洋船基水文调查计划(Global Ocean Ship-based Hydrographic Investigations Program, GO-SHIP)和痕量元素及同位素海洋生物地球化学循环国际研究(An Interna-

tional Study of the Marine Biogeochemical Cycles of Trace Elements and Isotopes, GEOTRACES)四个海洋学科领域的国际计划, 包含物理海洋、海洋化学、海洋生物与生态、海洋气象、光学等不同的学科方向, 对其数据管理流程和管理政策进行分析和总结, 借鉴国际海洋观测计划数据管理规范化理念和共享经验, 探讨建立适合我国的数据管理方案。

## 1 国际海洋观测计划简介

### 1.1 世界大洋环流实验(WOCE)简介

WOCE(1988—1998 年)是世界气候研究计划的主要组成部分, 是规模最大的国际海洋学合作计划之一。该计划为全球海洋观测提供了前所未有的全球海洋观测数据, 数据量大, 覆盖海域广, 包含 350 多种示踪剂分布, >18 500 个站位的高精度温盐深仪(Conductivity-

收稿日期: 2020-10-19, 修回日期: 2021-01-25

基金项目: 国家自然科学基金项目(41942022); 近海海洋环境科学国家重点实验室(厦门大学)技术人员开放基金

[Foundation: National Science Foundation of China, No. 41942022; Open fund for technician of State Key Laboratory of Marine Environmental Science (Xiamen University)]

作者简介: 许艳苹(1980—), 女(汉), 山东荣成人, 工程师, 硕士, 主要从事海洋科学的科研数据共享与管理、海洋科学数据库系统研究, 电话: 0592-2182833, E-mail: ypxu@xmu.edu.cn

Temperature-Depth, CTD)数据, >17 400 个站位的瓶采数据(包含水文、营养盐和示踪剂), 540 个航次的船载声学多普勒流速剖面仪(Acoustic Doppler Current Profile, ADCP)流速数据等。这些数据以电子图集(Electronic Atlas of WOCE Data, eWOCE)的形式发布<sup>[2]</sup>, eWOCE 是 WOCE 计划数据集发布的一部分, 收录在 DVD2<sup>[3]</sup>中。

WOCE 中国计划, 原国家海洋局依托“向阳红 5 号”于 1991 年 11—12 月在西太平洋实施首次 WOCE 中国海上多学科综合考察, 历时 45 天, 共完成 10 条断面的现场作业。其中有 4 条完整的 WOCE 断面(P27-P30, 即 PR21-PR24), 两条为 WOCE 断面的一部分(P8 和 P9, 即 PR1 和 PR3)<sup>[4]</sup>。首航共完成 CTD 测站 82 个, 抛弃式温深仪(XBT)147 个, 收集各种水化学样品 6 100 多个, 进行了全剖面观测, 这些资料全部汇入国际 WOCE 资料集。

## 1.2 全球联合海洋通量研究(JOGFS)简介

JGOFS(1989—2000 年)是一个包含水文、生物、化学、光学等多学科的国际研究计划, 研究区域包含太平洋、大西洋、印度洋、南大洋等海盆, 来自美国、德国、英国、法国等 26 个国家的科学家参与其中。中国是最早参与国际 JGOFS 计划活动的国家之一, 在 1989 年 2 月 JOGFS 中国委员会在青岛成立, 制定了 JOGFS 在中国实施的基本要点, 即黄河和长江与具有宽广陆架的中国边缘海间的通量, 为全球陆海间的通量研究提供重要的数据支撑。1991 年国家自然科学基金委员会启动重点项目“东海陆架边缘海洋通量研究”(1991—1994 年), 开启了全球陆架边缘海洋通量研究的序幕<sup>[5]</sup>。中国于 1990—1999 年在台湾海峡、东海、西菲律宾海等海域共实施 80 多个航次的研究工作, 均纳入了国际 JOGFS 计划<sup>[6]</sup>。

## 1.3 全球海洋船基水文调查计划(GO-SHIP)简介

GO-SHIP 由国际海洋碳协调项目(The International Ocean Carbon Coordination Project, IOCCP)与气候和海洋项目(Climates and Ocean: Variability, Predictability and Change, CLIVAR), 于 2007 年联合成立<sup>[7]</sup>。旨在进行海岸到海岸、海岸到冰区的持续重复的水文学断面调查计划, 于 2012—2023 年计划 55 条核心断面(包含 WOCE 计划的重复断面), 截止 2020 年 1 月已完成 80%的断面调查, 仍有 11%断面还没有列入调查计划。GO-SHIP 计划汇集物理海洋学、碳循环、海洋生物地球化学、生态系统等观测, 是全球气候观测

系统/全球海洋观测系统(Global Ocean/Climate Observing System)的一部分。

## 1.4 痕量元素及同位素海洋生物地球化学循环国际研究(GEOTRACES)简介

GEOTRACES 于 2010 年由国际海洋研究委员会启动, 是国际海洋学界的一项全球尺度的重大研究计划。来自美国、英国、德国、法国等 30 多个国家的科学家参与其中, 航迹遍布太平洋、大西洋、印度洋和极地海洋, 中国是首批成员国之一。目前已实施 130 多个航次, 其中仅 2017 年发布的数据产品包含了来自 25 个国家的 326 名科学家于 2007—2014 年实施的 39 个航次的数据集, 总计 1 810 个站位, 458 种参数, 590 张断面图、130 多个 3D 动画场景<sup>[8]</sup>。下一期数据产品预期于 2021 年发布。

其中有关痕量元素及其同位素的海洋生物地球化学及生态气候效应的研究, 受洁净采样与分析技术的限制, 被视为各国海洋科学研究实力的体现。第一个 GEOTRACES 中国断面调查航次—GP09 断面调查航次, 于 2019 年 4—6 月由厦门大学牵头依托厦门大学“嘉庚号”科考船执行。

## 2 国际计划的数据管理规范

WOCE、JGOFS、GO-SHIP 和 GEOTRACES 计划都制定了数据标准, 实施科学数据规范化。规范化的科研数据包含元数据和实体数据, 元数据是对样品的采集、测试、质控步骤及每个步骤的操作人、参照标准等的详细描述。元数据如同实体数据的说明书, 在数据使用过程中对任何一个步骤产生疑问、遇到问题都可以溯源, 对数据的历史版本、版本的更新内容及解释也可以溯源。实体数据是数据集实体, 通常包含采样日期及时间、站位名称、纬度、经度、水深、采样深度、布放 cast 号(即所在站位科研仪器下水顺序)、采样瓶号和具体参数名称, 需要对数据集的每项内容如参数名称进行解释, 对参数单位等进行统一规定。

### 2.1 元数据

WOCE、JGOFS、GO-SHIP 和 GEOTRACES 计划对元数据都有规定的模板内容, 包含航次时间、调查船、首席科学家、调查海域、参数采集的流程、分析步骤、质控措施、参照标准等。具体而言, WOCE 的元数据文档, 包含航次时间、调查船、调查断面/航次名称、首席科学家、航次实施的国家、调查区域经纬度范围、调查参数采样和分析信息、数据结

果的不确定性等。GO-SHIP 作为 WOCE 历史断面的重复调查,元数据文档大部分与 WOCE 是相同的,增加了数据/数据集发表情况。

JGOFS 的元数据模板,又称采样和分析方法的说明文档,是 ASCII 文本文件,包含采样流程、分析步骤、设备校验、数据处理技术等,采样方法可以引用参考文献。

GEOTRACES 的元数据模板,是 Excel 文件,除了共同信息外,增加“质控标签”,不同机构和项目通常采用不同的质控标签标准。同样都是 GEOTRACES 的数据产品(GEOTRACES Intermediate Data Product, IDP),不同 IDP 版本,标准也是不同的。如 IDP2014 和 IDP2017,采用的是国际海洋学数据与信息交流(International Oceanographic Data and Information Exchange, IODE)的质控标签;而即将发布的 IDP2021,采用的 SeaDataNet(Pan-European Infrastructure for Ocean & Marine Data Management)的质控标签。质控标签是对数据进行质控后的数据质量评估,如 IODE 的质控标签有 5 个,分别为:1 代表数据质量好(good quality),2 代表数据质量不确定(not evaluated, not available or unknown quality),3 代表数据质量存疑(questionable/suspect quality),4 代表数据质量不好(bad quality),9 代表数据丢失(missing data)。随着分析检测方法和技术的提高,数据说明文件逐渐被规范化,对数据的说明越来越详细,质控标签的个数也不断增多,如 SeaDataNet 增加了低于仪器检测限(value below detection)、内插值(interpolated value,即从项目的其他数据中衍生得出的值)等。

## 2.2 实体数据

为了保持整个项目参数的一致性,需要对项目的实体数据参数名称、单位等进行统一规定,即数据库框架,也称数据目录。已有的数据管理经验表明,随着研究的深入,开展的参数不断增加,参数列表需要实时更新,每个项目的参数列表都记录着更新的版本日期。

### 2.2.1 建立数据目录

WOCE 的参数分类包括海气通量、水文、海洋表层温度/盐度、声学多普勒流速剖面仪(ADCP)、海流计(Current Meters)、漂浮浮标(Drifters)、次表层浮标(Subsurface Floats)、剖面浮标(Profiling Floats)、海平面/海表高度数据(Sea Level Data)、海洋气象等。

JGOFS 的参数分类包括物理海洋、海洋化学、痕量元素、有机化合物、生物学-生物丰度/生物量、生物学-生产力/吸收/呼吸、同位素和光学等。

GO-SHIP 的参数分类包含 CTD 和基于 CTD 采集的参数、碳数据、船载声学多普勒流速剖面仪数据、投放式声学多普勒流速剖面仪数据、海洋气象数据和走航数据。

GEOTRACES 的参数分类借助参数树探索工具(Parameter Tree Exploration Tool)和参数搜索工具(Parameter Search Tool)拓展,两者可以相互转换。分为 8 大方向:气溶胶、生物、溶解态痕量金属、水文和生物地球化学、木质素、颗粒态痕量金属、极地、沉降。在每个方向中再按照采样系统/采样方式分为若干类,如痕量洁净 CTD、常规 CTD、原位大体积泵、拖鱼、自动收集的雨水、分粒级的颗粒物、船舶的走航采样等,共计 11 000 多种参数。GEOTRACES 的参数树按字母排序,如同词汇表,便于查找的同时,也方便后续参数的添加,是很好的参数树框架范例。

### 2.2.2 参数名称和单位

四大计划都对参数的名称、缩写进行定义;在参数目录下,可检索每项参数的名称、解释和单位。对参数列表中没有包含的参数,也都提供了命名方式的参考依据。

### 2.2.3 文件格式及内容

WOCE 的实体数据,综合网页公开发布的 DVD 和数据中心网站,数据格式包含 NetCDF、ASCII 和 Excel。

JGOFS 的数据管理系统,可读取任何格式的数据。对提交的实体数据的内容有以下 3 点要求:(1) 记录航次的事件编号,(2) 从 CTD 中采集的样品,记录 cast 号和采样瓶号,(3) 数据记录不能为空白,须标明理由,如坏的数据标记为“nd”,低于检测限标记为“zero”。

GO-SHIP 的实体数据,综合不同的数据中心,数据格式包含 NetCDF、Csv、Excel 等。

GEOTRACES 的实体数据,在公开发布的 IDP 数据页面包含 4 种格式:ODV、ASCII、Excel 和 NetCDF。

## 3 国际计划的数据管理政策

### 3.1 数据质量保证政策

1980 年末,时值 WOCE 计划启动伊始,国际海

洋学界尚没有一个数据系统能满足 WOCE 计划的数据管理。为此,该计划建立了一系列的数据整合中心(Data Assembly Center, DAC)和特殊分析中心(Special Analysis Centers),这些中心分布在各实验室和研究院校,每个数据中心只处理一个数据流,需要具备一定的专业知识来确保数据质量<sup>[9]</sup>。

GO-SHIP 计划参照 WOCE 计划的数据管理方式,依托不同的数据中心存储不同参数。修订了 1994 年 WOCE 水文项目手册,于 2010 年初发表。新手册规定了每个参数从产生到质控的细节,还规定了使用认证的参考物质/标准物质(Certified Reference Materials, CRMs),航次及测定期间尽可能频繁地使用,以确保不同航次相同断面的数据具有可比性。

JOGFS 计划,各个国家相继建立 JOGFS 项目数据管理办公室,负责数据质量的把关。

GEOTRACES 计划设有国际计划的数据中心—GEOTRACES 国际数据中心(GEOTRACES International Data Assembly Centre, GDAC),依托英国国家海洋数据中心(British Oceanographic Data Centre, BODC)。数据中心建立后,由专家组成数据管理委员会,制定了详细的数据质量控制流程。开展不同实验室之间的国际互校;对同一海域不同国家执行航次的校验,即在同一地理位置设置互校站位;提供铁等痕量元素的标准样品等。

## 3.2 数据保护期和公开发布

每个大计划都产生了大量高质量、高分辨率的多学科参数,科学家们普遍的愿望是全面公开船测资料,为了平衡整个大计划对数据的集成需求及科学家们的数据知识产权之间的矛盾,各大计划相继制定了数据共享与引用政策,规定了数据的公开时间、公开方式等。

### 3.2.1 数据保护期

WOCE 计划规定数据产生 2 年后公开,科学家可以在这段时间分析和发表数据,同时确保大计划对全球数据的收集。

美国 JOGFS 项目的大部分数据是在线公开,如过程研究项目中的北大西洋藻华实验(North Atlantic Bloom Experiment, NABE)、阿拉伯海过程研究(Arabian Sea Process Study)、赤道太平洋过程研究(Equatorial Pacific Process Study),BATS 和 HOTS 时间序列站数据,CO<sub>2</sub> 调查数据,遥感数据等<sup>[10]</sup>。只有

过程研究中的南极环境与南大洋过程研究(Antarctic Environment and Southern Ocean Process Study, AESOPS)项目对数据设置为期 2 年的数据保护期。法国 JOGFS 项目对数据设有 2—4 年的保护期。日本 JOGFS 项目对数据设有 3 年的保护期。

GO-SHIP 计划要求各项参数依托数据中心公开,依据航次的类型和参数的分类,规定航次结束后 6 周公开初步数据,6 个月公开质控数据,2 年内公开所有数据。

国际 GEOTRACES 计划要求所有参与国家在数据产生两年后公开。其中美国 GEOTRACES 项目要求 2 年内(自样品收集时间计算)公开所有数据集、元数据文件、衍生的数据产品如模型结果等。加拿大 GEOTRACES 项目要求在项目结束 3 年后,公开所有数据。

### 3.2.2 数据公开发布方式

四大计划数据都公开发布数据产品,包含数据集和图集,发布方式略有不同。

WOCE 计划数据的公开发布方式是 DVD(WOCE Data DVD<sup>[11]</sup>)和数据中心网站。GO-SHIP 计划数据的公开方式有两部分:(1)项目网站公开航次调查计划及完成情况,(2)数据中心网站公开数据。作为 WOCE 断面的重复计划,部分航次数据同时公开在 WOCE 项目的数据网站。

JOGFS 计划的数据产品“国际 JOGFS 数据集,第 1 卷:离散数据集(1989—2000)”,于 2003 年 5 月召开的 JOGFS 开放科学会议,首次以 DVD 的形式公开发布<sup>[12]</sup>。为了实现长期存储和在线访问,2003 年秋季由世界海洋环境科学数据中心(World Data Centre for Marine Environmental Sciences, WDC-MARE, Germany)负责整合国际 JOGFS 数据集,出版了“国际 JOGFS 数据集,第 2 卷:集成数据集”及对应的数据报告;近 40 000 个数据记录可以在 PANGAEA 数据库里实现在线检索<sup>[12]</sup>。JOGFS 数据产品的发布经历了每个国家单独发布其数据产品,而后集成发布,期刊公开发表和网站发布相结合。

GEOTRACES 计划在实施的第四年公开发布第一个数据产品,即 IDP2014<sup>[13]</sup>,包含两部分:(1)实测数据集,(2)电子图集(eGEOTRACES),eGEOTRACES 基于实测数据集,展示方式包含断面图和 3 维动画场景。3 年后又发布了第二个数据产品 IDP2017<sup>[8,14]</sup>,数据量增加了一倍,数据质量和展示形式上也有所

创新。目前正在准备第三个数据产品 IDP2021。IDP 的特点是以航次为连接,将来自不同国家科学家的数据进行集成,以期刊的形式公开发表,结合网站发布。

### 3.3 数据共享与引用

四个大计划都制定了数据共享政策,以保护知识产权。对还没有公开/发表的数据,未经提供数据的科学家同意不得传递和使用。已公开公布的数据,采用科学引用的方式使用。如 WOCE 的数据引用公开的 DVD 和 eWOCE。JGOFS 计划制定了引用政策,国际计划汇编来自各个国家的单个数据集,引用格式:责任科学家姓名+数据集名称+CD-ROM 名称,每个国家的格式略有不同<sup>[15]</sup>。GO-SHIP 的数据使用,需致谢“全球海洋船基水文调查计划(GO-SHIP)”,并引用气候和海洋项目与碳水文数据办公室发布的数据集等。GEOTRACES 的数据和图集引用公开发表的 IDP<sup>[8, 13]</sup>和 eGEOTRACES。

## 4 国际计划数据管理依托的平台

### 4.1 WOCE

WOCE 计划设立国际项目办公室,最初位于英国伍姆利的海洋科学研究所执事实验室(Institute of Oceanographic Sciences Deacon Laboratory, Wormley, U.K),1995 年夏季搬到英国南安普敦大学南安普敦海洋学中心。项目办公室的主要任务是获取、评估和传播科学家们参与 WOCE 领域的活动,数据分析和建模方案等。WOCE 的数据集存储在美国国家海洋数据中心(US National Oceanographic Data Centre, NODC)。

### 4.2 JGOFS

JGOFS 计划设立 JGOFS 国际项目办公室,位于挪威卑尔根大学。没有国际计划的数据中心,数据管理依托各参与国家建立的项目数据管理办公室。

JGOFS 美国建立 JGOFS 美国数据管理办公室,位于伍兹霍尔海洋学研究所(Woods Hole Oceanographic Institution, WHOI),负责 JGOFS 美国的数据收集、存储与管理。JGOFS 德国数据中心设在德国基尔海洋科学研究所(Institut für Meereskunde, Kiel, Germany),有数据系统网站,可以检索航次信息、参数及负责责任科学家、数据可用性。数据申请和使用时,需要邮件联系责任科学家得到许可。JGOFS 英国数据中心设在 BODC,负责数据的

存储、共享与发布,其中得到授权的用户可以在在线系统中提取数据。JGOFS 法国数据中心设在法国自由城海洋观测站(Villefranche Oceanographic Observatory, France),负责 JGOFS 法国数据的收集和存储,并维护数据库网站。JGOFS 日本设有数据管理办公室,数据存储在日本海洋数据中心(Japan Oceanographic Data Center, JODC);数据中心的网站由名古屋大学维护。JGOFS 加拿大数据中心设有专门的数据管理网站,与加拿大的国家海洋数据中心—加拿大海洋环境数据服务(Marine Environment Data Service, MEDS)相连接,存储 JGOFS 加拿大项目的数据和相关事件文件集;通过 MEDS 的离线数据申请系统申请数据。JGOFS 印度数据存储在南印度的国家海洋数据中心(Indian National Oceanographic Data Centre, NODC),负责数据的收集和存储<sup>[10]</sup>。

### 4.3 GO-SHIP

GO-SHIP 计划,依托不同的数据中心对数据进行收集、存储与发布<sup>[8]</sup>:(1) 气候和海洋项目与碳水文数据办公室(CLIVAR and Carbon Hydrographic Data Office, CCHDO),负责 CTD 和瓶采参数;(2) 海洋碳数据系统(Ocean Carbon Data System, OCADS, former CDIAC),负责碳数据;(3) 夏威夷联合存储船载 ADCP(Hawaii Joint Archive for Shipboard ADCP)和全球海洋表面航行数据项目(The Global Ocean Surface Underway Data Project, GOSUD),负责船载声学多普勒流速剖面仪数据;(4) 夏威夷大学海流组(Hawaii Joint Archive for Shipboard ADCP),负责投放式声学多普勒流速剖面仪数据;(5) 上层海洋气象数据整合中心(Surface Marine Meteorological Data Assembly Center, COAPS, FSU),负责海洋气象数据;(6) 全球海表走航数据项目(The Global Ocean Surface Underway Data Project, GOSUD),负责走航数据。

### 4.4 GEOTRACES

GEOTRACES 计划设有国际计划数据中心 GDAC,由英国国家海洋数据中心管理,由海洋研究科学委员会、美国国家科学基金会和英国自然环境研究中心共同资助。

GEOTRACES 美国项目数据中心设在生物化学海洋学数据管理办公室(Biological and Chemical Oceanography Data Management Office, BCO-DMO),

负责数据的收集、存储与发布。GEOTRACES 加拿大项目数据中心设在加拿大的国家极地数据中心—极地数据目录(Polar Data Catalogue), 由加拿大国际极地年(The Canadian International Polar Year, IPY)计划提名, 加拿大卓越中心网(ArcticNet)和加拿大冰冻圈信息网(CCIN)联合开发, 项目办公室设在滑铁卢大学和魁北克市拉瓦尔大学。GEOTRACES 德国项目数据中心设在德国的国家极地数据中心—PANGAEA, 负责数据的收集、存储与发布。GEOTRACES 瑞典项目数据中心设在瑞典气象水文研究所(SMHI), 负责数据的收集、存储与发布。GEOTRACES 日本项目数据中心设在日本海洋数据中心 JODC, 负责收集和管理所有海洋学数据, 并负责数据的质量控制。GEOTRACES 法国项目数据中心设在法国自由城海洋观测站, 负责数据的收集、存储与发布。GEOTRACES 荷兰项目数据中心设在荷兰皇家海洋研究所(Royal Netherlands Institute for Sea Research, NIOZ), 负责数据的收集、存储与发布。GEOTRACES 中国项目数据中心设在厦门大学, 负责数据的收集、存储与发布。

## 5 国际海洋观测计划数据管理的对比

综合 WOCE、JGOFS、GO-SHIP 和 GEOTRACES 四个国际海洋观测计划, 项目的执行期都超过 10 年, 研究区域遍布全球大洋, 多个国家的科学家共同参与, 均依托国家数据中心/大型数据中心, 依次为美国国家海洋数据中心(NODC)、地球与环境科学数据发布(PANGAEA)、气候和海洋项目与碳水文数据办公室(CCHDO)、GEOTRACES 国际数据中心(GDAC)等。在数据管理方面都制定了数据规范化存储、数据质量控制措施、数据保护期, 在数据的公开发布方式、共享与引用、数据的下载使用等方面略有不同。

四大计划的不同之处, 对比如下:

**实体数据内容:** 四个计划只有 JGOFS 计划要求提供 cast 号和采水瓶号, 同一站位不同 cast 号, 可能一个采自白天, 一个采自夜晚; 这两个参数决定了各参数是否是采自同一水团, 对过程研究非常重要。

**数据质量保证:** WOCE 和 JGOFS 计划的数据质量保证依靠各数据中心和项目数据管理办公室, 没有统一的质控措施和标准。GO-SHIP 计划在 WOCE 计划的基础上, 修改制定了新的操作手册, 统一了

从采样到质控的流程, 规定了样品测定过程中使用标准物质进行质量控制。GEOTRACES 计划建立国际数据中心, 成立数据管理委员会, 规定对于不同国家/研究团队实施的不同航次需设定互校站位, 各实验室参与国际比对等措施, 对各参与国家的数据质量进行统一把关。

**数据产品发布:** GEOTRACES 计划在执行期间, 每隔 3—4 年更新发布一次数据产品, 采用数据图集和实体数据相结合, 期刊发布和网站发布相结合的方式, 能够更快地促进学科间的交流及数据的共享。而 WOCE 和 JGOFS 计划都是项目结束后再组织发布数据产品。GO-SHIP 计划仍在执行期, 还没有发布整合的数据产品, 依托数据中心发布数据集。

**数据申请与共享:** WOCE、GO-SHIP 和 JGOFS 计划, 打开数据中心网址选择需要的断面/航次/参数, 可以直接下载实体数据<sup>[16, 17, 12]</sup>; GEOTRACES 计划, 打开数据中心网址先注册账号选择数据使用用途, 可以下载不同格式的实体数据和图集<sup>[14]</sup>。

## 6 建立数据中心/数据仓储对于数据规范化管理的重要性

国际计划的数据管理经历了 WOCE 时代从刚开始没有数据中心可用, 到建立各种组合和处理数据中心管理数据; JGOFS 计划依托各国家建立的项目数据管理办公室管理数据; GO-SHIP 依托不同的数据中心管理不同类型的数据; 到 GEOTRACES 计划建立国际计划数据中心并依托各国家的国家数据中心。数据管理方面从数据规范化内容、数据质量控制方案、数据发布与引用方式、共享政策等都在不断发展完善。

其中数据的发布和引用方式, 从 WOCE 和 JGOFS 计划的 CD-ROM, GO-SHIP 计划的航次数据集, 到 GEOTRACES 计划的 IDP, 目的都是让数据能独立于文章单独发表/发布。数字对象标识符(digital object identifiers, doi)的应用, 解决了数据要与文章捆绑发表的困局, 确保数据知识产权的同时能够最大程度地使用数据, 是全球大数据时代海洋科学发展的重大机遇。建立规范的数据仓储中心, 必定能推进数据共享与使用。

国际计划数据共享与管理的成功经验, 得益于他们所依托的科学数据中心。我们选取了其中四家海洋学数据中心, 分析其功能与特点, 汇总在表 1。

表 1 四个科学数据中心的功能与特点汇总

Tab. 1 Summary of functions and characteristics from four scientific data centers

数据中心	所属机构	创建时间	学科	功能与特点
BODC	英国国家海洋数据中心	1988	海洋环境学	数据库设计具有四大功能: 关系数据库管理系统、国家海洋学数据库、项目数据库、网络数据库; 也是数据仓储库
BCO-DMO	伍兹霍尔海洋研究所	2006	海洋学	只接收美国国家科学基金会海洋科学部的生物和化学海洋学部门以及极地项目部的南极生物和生态系统项目提供数据管理服务
PANGAEA	德国不莱梅大学、AWI 研究所	1995	地球与环境科学	支持全球的科学家使用、存储和发布数据。数据仓储库, 可以使用数字对象标识符 doi 来标识, 共享, 发布和引用每个数据集
JODC	日本海洋数据中心	1965	海洋学	自 1995 年, 海洋数据服务系统实现互联网在线, 用户可以使用多个关键字在线检索数据, 并下载。不支持其他来源的数据存储

### 6.1 英国国家海洋数据中心(BODC)

BODC 成立于 1988 年, 起初负责北海项目的数据管理, 如今已发展成为具有成熟的样本框架数据库, 负责管理大型多学科研究项目的数据, 如 JGOFS 英国、GEOTRACES 英国和国际计划的数据管理。除了项目数据库管理, BODC 的数据库设计还兼具其他三大功能: 关系数据库管理系统、国家海洋学数据库和网络数据库<sup>[18]</sup>。

BODC 还是一个巨大的数据存储中心, 接收全球海洋环境领域数据的存储。存储在 BODC 的数据, 会配备 doi。doi 是用于永久和稳定地标识(通常是数字)对象的代码, 提供了一种用于检索有关对象的元数据的标准机制, 且通常提供一种访问数据对象本身的方法。

对数据知识产权的保护, 除发表文章之外, 将其存储在数据仓储/数据中心以获取 doi, 俨然已经成为保护其知识产权的最好办法。

### 6.2 美国生物化学海洋学数据管理办公室(BCO-DMO)

BCO-DMO 于 2006 年由前 JGOFS 美国项目办公室和前全球海洋生态系统动力学(GLOBal Ocean ECosystems Dynamics, GLOBEC)美国项目办公室合并成立。位于 WHOI, 目前也负责 GEOTRACES 美国的数据管理, 由美国国家科学基金会资助<sup>[19]</sup>。

BCO-DMO 目前只接收美国国家科学基金会海洋科学部的生物和化学海洋学部门以及极地项目部的南极生物和生态系统项目, 提供数据存储和数据管理服务。其他来源的数据存储需要付费。

### 6.3 地球与环境科学数据发布(PANGAEA)

PANGAEA 成立于 1995 年, 是地球与环境科学数

据发布信息系统, 支持全球的科学家使用、存储和发布数据。可以使用 doi 来标识、共享、发布和引用每个数据集<sup>[20]</sup>。由阿尔弗雷德·韦格纳研究所, 亥姆霍兹极地和海洋研究中心(the Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, AWI)和不来梅大学海洋环境科学中心(the Center for Marine Environmental Sciences, University of Bremen, MARUM)主办。

### 6.4 日本国家海洋数据中心(JODC)

JODC 成立于 1965 年, 是日本的海洋综合数据库, 收集和管理日本政府机构、大学和其他海洋研究机构观测到的海洋数据<sup>[21]</sup>。负责管理日本的国际项目数据, 如 JGOFS, GEOTRACES, PICES(North Pacific Marine Science Organization)等。

自 1995 年, 海洋数据服务系统实现互联网在线使用功能, 用户可以使用多个关键字在线检索数据, 并下载。不支持其他来源的数据存储。

## 7 开展综合性项目科学数据管理工作的启示

基于以上对比分析, 在开展综合性项目数据管理工作时, 为了让数据具备长久的使用性和可传承性, 平衡项目对数据的集成需求和科学家对数据的共享需求。建议项目启动时, 建立项目/数据管理办公室, 制定科研数据标准化规范。针对项目研究内容和目标, 制定出一套完整的项目的实体参数目录, 统一数据名称、单位、分类及内容, 制定元数据模板, 根据数据系统功能设定可接收的数据格式。数据管理办公室要具备存储、共享与发布数据的功能。在项目成员内部就数据质量的保证措施、数据共享和使用等管理规定达成共识。

## 7.1 建立实体参数表, 明确责任分工

针对项目的关键科学问题、主要研究目标和预期目标等, 结合责任科学家的研究方向和项目的课题设置, 制定出实施参数与责任科学家一一对应的项目计划实施参数汇总表, 依照航次执行情况实时更新。

## 7.2 推行数据标准化

所谓数据标准化, 即制定数据规范化标准, 包含元数据和实体数据, 规定其具体内容, 同时预留新增学科数据的空间。

### 7.2.1 元数据

需要建立元数据模板, 包含所有与科学数据产生相关的基本信息, 做到每个步骤细节可溯源, 如航次基本信息, 参数的采样、测样、分析流程及质控措施等相关流程和操作人等。

### 7.2.2 实体数据

(1) 参数名称和单位, 同一个参数在不同的研究机构和不同的科学家的笔下, 常常有不同的名称。参数名称需要标准化, 第一步是汇总同一参数尽可能多的名称, 第二步要求参数名称归一化。或者制定出适合项目参数的命名规则, 并统一参数的单位。

(2) 文件模板和格式, 根据数据库设计的功能, 规定实体数据的内容和提交格式。

(3) 数据质量的保证方案和质控标签

数据质量的保证方案, 尽可能详细具体。比如规定各参数从采样、分析测定到质量控制每个步骤的操作规范, 使用统一的标准物质, 设置共同的互校站位(每个海域设置一个站位或者一条断面), 在国内开展各实验室之间的比对, 参与国际互校等工作方案。

质控标签, 可以创立项目/数据中心的质控标签, 也可以直接采用国际上已有的质控标签。

(4) 参数分类, 根据计划参数的特点, 建立项目/数据中心的参数分类体系, 创建“参数字典表”, 便于在数据库中检索查找; 预留新增数据的空间。

## 7.3 规定数据提交及公开时间

数据提交时间要视项目研究特点, 依据数据生产周期, 可以设置不同的时间节点提交数据。比如在船上能够完成测定的参数, 可以设置 1—6 个月提交原始数据、质控数据和相关报告。测定周期比较长的数据如同位素等可以设置 1—2 年提交。国际惯例

一般要求在航次结束后两年/数据产生后两年完成提交并公开, 公开时间同样需要视项目研究特点, 在项目成员内部达成共识。

## 7.4 制定数据共享和引用政策

数据共享管理政策, 尽可能详细具体。比如: 数据未公开前, (1) 在期刊论文中使用未发表数据, 投稿前需要联系数据提供者征得同意, 并商讨数据的具体使用方式。(2) 不得将申请到的数据提供给第三方等。

综合性项目的特点强调多学科交叉, 如何既能满足学科交叉对数据共享交流的需求, 又能保证科学家的知识产权, 除了项目成员共同遵守学术道德规范, 还需出台法律/政策规范进行保障和限制<sup>[22]</sup>。

## 7.5 建立/依托数据共享平台

数据要集成, 各学科要进行充分的共享交流, 促进项目的成果产出, 必须建立/依托数据共享平台, 同时配备共享平台的开发小组和共享管理的业务小组, 共同合作运营。同时为数据按时提交如期公开提供技术支撑与保障。数据共享平台, 可以是分布式国家级数据中心群<sup>[23]</sup>, 可以是与国际接轨的数据仓储中心群, 数据集提交到数据仓储/数据中心获得 doi, 使用该数据集, 必须引用 doi。

## 8 结语

本文介绍了 WOCE、JGOFS、GO-SHIP 和 GETRACES 四个国际海洋观测计划对数据管理的一整套流程。数据在管理方式上都制定了数据规范化、数据质量控制措施、数据保护期、数据公开发布、共享与引用政策。

全球海洋观测计划的数据管理经历了近 30 年的发展, WOCE 计划由刚开始没有数据中心可用, 到建立各种组合和处理数据中心管理数据; JGOFS 计划依托各国家建立的项目数据管理办公室管理数据; GO-SHIP 依托不同数据中心存储和管理不同类型的数据; GEOTRACES 计划建立国际计划数据中心并依托各国家的国家数据中心, 数据管理从数据规范化内容、数据发布与引用方式、共享政策都在不断发展完善。在此基础上对比了美国 BCO-DMO、英国 BODC、德国 PANGAEA 和日本 JODC 四个国家的国家海洋数据中心的的功能与特点, 探讨了建立数据中心和数据仓储对推行数据规范化存储与管理推进数据共享的重要性。



借鉴国际海洋观测计划数据管理形式和各国的数据共享经验,探讨建立一套适合我国的数据管理方案。项目启动时,依托数据中心,建立项目/数据管理办公室,制定科研数据标准化规范。针对项目研究内容和目标,制定出一套完整的项目的实体参数目录,统一数据名称、单位、分类及内容,制定元数据模板,根据数据系统功能设定可接收的数据格式。数据管理办公室要具备存储、共享与发布数据的功能。在项目成员内部就数据质量的保证措施、数据共享和引用等管理规定达成共识。本研究可以为我国的海洋观测计划制定数据管理方案提供参考和借鉴。

致谢:感谢厦门大学戴民汉教授的指导与建议,感谢厦门大学蔡毅华教授、杨进宇助理教授、李骁麟教授、郭香会副教授、林宏阳副教授和集美大学何碧烟教授在论文修改过程中提出的建议,感谢两位评审专家对论文提出的评审建议。

参考文献:

- [1] JENNINGS F D. The Role of NSF in “Big” Ocean Science: 1950 to 1980[C]//50 Years of Ocean Discovery: National Science Foundation 1950-2000. National Academy of Sciences. Washington, D.C: National Academy Press, 2000: 141-148.
- [2] eWOCE-Electronic Atlas of WOCE Data[DB/OL]. [2019-12-11]/[2020-09-06] <https://www.ewoce.org/>.
- [3] WOCE Data Products Committee. WOCE Global Data Version 3.0. WOCE International Project Office, WOCE Report No. 180/02[R], Southampton, UK, 2002.
- [4] 范元炳, 蒲书箴. 我国海洋科学领域的全球变化研究进展[J]. 地球科学进展, 1998, 13(1): 62-71.  
FAN Yuanbing, PU Shuzhen. Research progress in oceanographic sciences of China relevant to global change[J]. Advances in Earth Science, 1998, 13(1): 62-71.
- [5] 戴民汉, 张经, 杨桂朋, 等. 海洋化学. 《中国海洋学学科史》(苏纪兰主编, 王颖副主编)[M]. 北京: 中国科学技术出版社, 2015: 183-218.  
DAI Minhan, ZHANG Jing, YANG Guipeng, et al. Chemical oceanography, In: History of oceanography in China (J. Su and Y. Wang, Editors), China Science and Technology Press, 2015: 183-218.
- [6] Joint Global Ocean Flux Study (JGOFS)[EB/OL]. (2003-12)/[2020-09-06]. <http://ijgofs.whoi.edu/>.
- [7] The Global Ocean Ship-based Hydrographic Investigations Program (GO-SHIP)[EB/OL]. (2020-12-23)/[2021-01-20]. <https://www.go-ship.org/>.
- [8] SCHLITZER R, ANDERSON R F, DODAS E M, et al. The GEOTRACES Intermediate Data Product 2017[J]. Chemical Geology, 2018, 493: 210-223.
- [9] WOCE International Project Office. “WOCE observations 1990-1998; a summary of the WOCE global data resource”. WOCE Report No. 179/02[R], Southampton, UK, 2003.
- [10] LOWRY R K, BALIÑO B M, eds. JGOFS Data Management and Synthesis Workshop. JGOFS Report No. 29/99[R]. JGOFS International Project Office, Bergen, Norway, 1999.
- [11] World Ocean Circulation Experiment Global Data Resource 3.0[DB/OL]. (2000)/[2020-09-06]. <https://www.nodc.noaa.gov/woce/>.
- [12] JGOFS International Collection DVD[DB/CD]. (2003-05)/[2020-09-06]. <http://ijgofs.whoi.edu/>.
- [13] MAWJI E, SCHLITZER R, DODAS E M, et al. The GEOTRACES Intermediate Data Product 2014[J]. Marine Chemistry, 2015, 177: 1-8.
- [14] GEOTRACES Intermediate Data Product 2017 (Version 2)[DB/OL]. (2018-02-14)/[2020-09-06] <https://www.bodc.ac.uk/geotraces/data/>.
- [15] CONKRIGHT M, AVRIL B, CHANDLER C, et al. JGOFS International Data Collection. V1: Discrete Datasets[R], 2003.
- [16] WOCE Program Data[EB/OL]. (2020-08-06)/[2021-01-20]. <https://www.ncei.noaa.gov/access/ocean-carbon-data-system/oceans/CDIACmap.html>.
- [17] CLIVAR and Carbon Hydrographic Data Office(CCHDO)[EB/OL]. (2020-12-01)/[2021-01-20] <http://cchdo.ucsd.edu/>.
- [18] British Oceanographic Data Centre (BODC)[EB/OL]. (2020-09)/[2020-09-06] <https://www.bodc.ac.uk>.
- [19] Biological and Chemical Oceanography Data Management Office (BCO-DMO)[EB/OL]. (2020-09)/[2020-09-06]. <https://www.bco-dmo.org/>.
- [20] PANGAEA[EB/OL]. (2020-09)/[2020-09-06]. <https://pangaea.de/>.
- [21] Japan Oceanographic Data Centre (JODC)[EB/OL]. (2020-09)/[2020-09-06]. <https://www.jodc.go.jp/jodcweb/aboutJODC.html>.
- [22] 路鹏, 苗良田, 莫纪宏, 等. 科学数据共享领域的政策规范和法律规范[J]. 国际地震动态, 2008, 4: 35-41.  
LU Peng, MIAO Liangtian, MO Jihong, et al. The norm of policy and law in scientific data sharing[J]. Recent Developments in World Seismology, 2008, 4: 35-41.
- [23] 刘闯, 王正兴. 美国全球变化数据共享的经历对我国数据共享决策的启示[J]. 地球科学进展, 2002, 17(1): 151-157.  
LIU Chuang, WANG Zhengxing. The experience of U.S global change data and information sharing and its indication to China[J]. Advances in Earth Science, 2002, 17(1): 151-157.

# Review of standardized data archiving and management of international ocean observation programs

XU Yan-ping

(State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361102, China)

**Received:** Oct. 19, 2020

**Key words:** international ocean observation programs; standardized data archiving; data center; data depository; data sharing and citation

**Abstract:** Based on the standardized data archiving and management practices of four international ocean observation programs (i.e., World Ocean Circulation Experiment (WOCE), Joint Global Ocean Flux Study (JGOFS), Global Ocean Ship-based Hydrographic Investigations Program (GO-SHIP), and An International Study of the Marine Biogeochemical Cycles of Trace Elements and Isotopes (GEOTRACES)), long-term scientific data availability, data archiving effectiveness, and data reuse for multi-disciplinary research programs were assessed. Results showed that a data center/data depository is essential to standardize data management and promote data sharing. Hence, this paper summarizes a standard protocol of data management. A data management office (DMO) or program office (PO) is initially set up to conduct data standardization and frame a set of parameter dictionaries, including names, units, classifications, and metadata templates. The DMO/PO should be able to ensure data archiving, safety, and distribution. The principal investigators/committee members also need to reach an agreement on data policy regarding data quality control, data sharing, and citation.

(本文编辑: 康亦兼)