Systematic Compilation and Application of Historical Marine Scientific Investigation Data for the South China Sea and Its Affiliated Islands and Reefs (1)

Authors: Xu Chao, Li Sha, Yuan Li, Xu Chao

Date: 2023-08-17T00:00:00+00:00

Abstract

The South China Sea constitutes a core national interest of China, characterized by its vastness and depth, abundant resources, complex natural conditions, and strategic geographical location. Since the 1950s, China has organized over ten large-scale comprehensive marine scientific expeditions to the South China Sea and its affiliated islands and reefs, obtaining a substantial volume of important original data. However, early materials lacked electronic processing means, with inconsistent standards for recording and archiving expedition data, and some original materials scattered among individual researchers. Due to personnel turnover, death, and other factors, historical data and materials have been lost or are on the verge of being lost, creating an urgent need for rescue and systematic compilation.

The project "Systematic Compilation of Historical Data from Marine Scientific Expeditions to the South China Sea and Its Affiliated Islands and Reefs" plans to collect, organize, and compile historical data and materials from over 400 large- and small-scale South China Sea scientific expedition projects conducted since the 1950s. The project will first achieve electronic conversion of expedition cruise data, then construct a comprehensive database and geographic information system, and finally integrate 5-6 specialized databases (hydrometeorology, biological ecology, chemistry, fisheries, geology, islands and reefs, etc.), 4-6 map libraries, 1-2 sets of visualization products, 3-5 research reports, and 6-8 monograph series. The rescue and systematic standardized compilation of historical data are of significant importance for understanding the resource and environmental conditions of the South China Sea and its islands, promoting data openness and sharing, enhancing data application value, and maximizing the benefits of national scientific and technological investment.

This series of articles is divided into three parts: Article 1 specifies the concrete content and implementation procedures for project data submission and standardized compilation, proposing corresponding practical strategies for technical challenges; Article 2 addresses the scientific management of compiled data, briefly discussing research and practice in South China Sea marine data governance technologies; Article 3 focuses on the sharing services of compiled data, briefly discussing the construction practices and innovative applications of "Digital South China Sea". Through this series, we aim to provide a relatively comprehensive discussion of the research process and practical experience in developing the comprehensive database for compiling historical South China Sea expedition data, namely "Digital South China Sea", thereby offering technical and application references for relevant data compilation projects under the National Special Program for Science and Technology Basic Resources Survey.

Full Text

Preamble

The South China Sea represents a core national interest for China—vast, deep, resource-rich, and strategically vital, yet characterized by complex natural conditions. Since the 1950s, China has organized more than ten large-scale comprehensive marine scientific surveys of the South China Sea and its affiliated islands and reefs, yielding substantial quantities of important raw data. However, early records lack electronic formats, survey documentation standards vary across expeditions, and archival standards remain inconsistent. Moreover, some original data remain scattered among individual researchers, and with personnel changes and deaths over time, much of this historical data has been lost or faces imminent loss, creating an urgent need for rescue and systematic reorganization.

The "Reorganization of Historical Data from Scientific Investigations of the South China Sea and Its Affiliated Islands and Reefs" project aims to collect, rescue, sort, and reorganize historical data from over 400 scientific investigations in the South China Sea since the 1950s. The initiative will first digitize scientific expedition data, then construct a comprehensive database and geographic information system, ultimately integrating 5–6 specialized databases (hydrometeorology, biological ecology, chemistry, fisheries, geology, islands and reefs), 4–6 image libraries, 1–2 visualization products, 3–5 research reports, and 6–8 monographs. This rescue and systematic standardization of historical data is crucial for understanding the resource and environmental conditions of the South China Sea and its islands, promoting open data sharing, enhancing data application value, and maximizing the benefits of national scientific investment.

This series comprises three articles. The first article defines the specific content and implementation procedures for data submission and standardized reorganization, proposing practical strategies for technical challenges. The second article addresses the scientific management of reorganized data, discussing research and practice in South China Sea marine data governance. The third

article focuses on sharing services for reorganized data, examining the construction practices and innovative applications of the "Digital South China Sea." Through this series, we aim to comprehensively document the research process and practical experience of the comprehensive database for reorganizing historical South China Sea survey data—the "Digital South China Sea"—providing technical and application references for related data compilation projects under the National Special Program for Basic Science and Technology Resource Surveys.

Keywords: South China Sea; scientific investigation; historical data; systematic reorganization; technology application

Since the 1950s, China has conducted numerous large-scale survey projects in the South China Sea and its affiliated islands and reefs. These include the 1958–1960 National Comprehensive Marine Survey, which established 36 survey sections, 237 surface observation stations, and 57 continuous observation stations in the South China Sea—though surveys near Taiwan Province in the East China Sea region and large areas of the South China Sea could not be completed due to limitations at the time [**?**, **?**]. The 1973–1978 Comprehensive Survey of the Central South China Sea covered eleven expeditions totaling over 50,000 nautical miles, reaching every reef of the Xisha Islands, repeatedly traversing the Zhongsha Islands, landing on Huangyan Island multiple times, and passing through the northern part of the Nansha Islands. Survey items encompassed marine geology, seafloor geomorphology, marine sedimentation, marine meteorology, marine hydrology, seawater chemistry, marine biology, and island geomorphology [**?**]. The 1979–1982 Comprehensive Survey of the Northeastern South China Sea conducted 14 comprehensive and thematic expeditions investigating various natural phenomena and their patterns, including circulation, water masses, temperature-salinity-density structures, meteorological and wave conditions, sediment characteristics and distribution, gravity and magnetic field features, distribution and variation of chemical elements, pollution status and countermeasures, and biological species and abundance distribution [**?**]. The 1984–2009 Comprehensive Survey of the Nansha Islands and Adjacent Waters, conducted as a national special project across the Seventh through Eleventh Five-Year Plans, yielded large quantities of measured data and samples in marine geology, geomorphology, geography, surveying, geophysics, geochemistry, biology, ecology, chemistry, physics, hydrology, and meteorology. These efforts produced massive amounts of firsthand data and important original records, with over 60 publications on scientific investigation results from the Nansha Islands and South China Sea since the 1980s. These achievements hold significant scientific importance and have provided reference materials for China's diplomacy, national defense, and production construction, generating substantial social and economic benefits [**?**].

Implementing the current project according to unified reorganization technical specifications to construct a comprehensive database will enable more thorough understanding of the resource and environmental conditions of the South China

Sea and its islands. This work not only demonstrates China's historical management and development of the South China Sea but also provides fundamental scientific data support for major national strategic needs in resource development, military operation support, diplomatic negotiations, and rights protection in the new era.

1 Main Content of Data Reorganization

The "Reorganization of Historical Data from Scientific Investigations of the South China Sea and Its Affiliated Islands and Reefs" is led by the South China Sea Institute of Oceanology, Chinese Academy of Sciences, in collaboration with nine institutions: the Institute of Oceanology (Chinese Academy of Sciences), the Third Institute of Oceanography (State Oceanic Administration), Xiamen University, the South China Sea Fisheries Research Institute (Chinese Academy of Fishery Sciences), the Qingdao Institute of Marine Geology (China Geological Survey), the Institute of Geographic Sciences and Natural Resources Research (Chinese Academy of Sciences), Nanjing University, China University of Petroleum (East China), and Tsinghua University. The project systematically reorganizes historical data from past South China Sea scientific investigations, covering firsthand measured data from numerous marine research fields including marine hydrology, meteorology, chemistry, biology, ecology, geology, geophysics, fisheries, and island reef scientific drilling.

The main tasks include: (1) collecting, rescuing, sorting, and electronically entering historical data from large-scale marine scientific surveys, including locating various archival materials and ship logs, gathering scattered data, digitizing paper records, and interviewing retired veteran scientists and crew members to reconstruct historical "memory recovery," ensuring the reliability, accuracy, and authenticity of scientific expedition events and data; (2) performing quality control, standardization, and normalized reorganization of historical data to construct electronic databases; and (3) producing datasets, atlases, and systematic summary reports related to observations and analyses of South China Sea territory, marine physics, biodiversity, ecological environment, fishery resources, marine chemistry, sedimentary geology, and marine geophysics, while establishing a comprehensive database and sharing platform, publishing scientific monographs, producing audio-visual materials from South China Sea surveys, and developing geographic information systems.

2 Overall Implementation Steps

The project implementation can be divided into three major phases: comprehensive collection and sorting, standardized reorganization and analysis, and scientific management and sharing [Figure 1: see original paper].

2.1 Comprehensive Collection and Sorting of Historical Data

The project's primary task is to assess historical South China Sea survey data domestically and internationally. Seven research groups implement this by determining the spatial-temporal distribution and specific content for data collection according to their objectives, systematically sorting historical data to achieve effective aggregation of dispersed materials.

By 2017, Research Group 1 had completed an inventory of existing South China Sea survey documents and audio-visual materials at the South China Sea Institute of Oceanology, collecting 101 CDs, 909 photographs, 17 videotapes, 8 audiotapes, 71 diskettes, and 598 documents from South China Sea surveys conducted between 1959 and 1990. By 2021, all research groups had completed collection, sorting, and scanning of paper documents, as well as preliminary submission of data entities for items listed in the "Electronic and Reorganization Project Inventory," with summaries provided in Table 1 .

2.2 Standardized Reorganization and Analysis

2.2.1 Data Aggregation Following data submission management methods and standard specifications for basic work data, six research groups conduct classified collection of data according to their specialties: physical oceanography and meteorology, marine chemistry, marine biology and ecology, marine fisheries, marine geological environment, and island reef geological samples. This approach achieves effective aggregation of dispersed data materials.

2.2.2 Standardized Reorganization According to project reorganization technical procedures and quality control specifications, the six research groups perform data inspection, classification, and standardization on original materials, completing normalized reorganization of data elements to establish six specialized databases: physical oceanography and meteorology, marine chemistry, marine biology and ecology, marine fisheries, marine geological environment, and island reef geological samples. These are integrated into one comprehensive database for South China Sea survey historical data reorganization, with construction undertaken by the South China Sea Marine Data Center of the South China Sea Institute of Oceanology (hereinafter referred to as the South China Sea Marine Data Center).

The project systematically sorts all electronic South China Sea scientific investigation historical materials submitted by the six groups (including project name, survey time, survey area, survey content, undertaking unit, archiving unit, and funding) and conducts spatial-temporal distribution visualization to construct one South China Sea Survey Geographic Information System. Interviews with retired veteran scientists enable historical "memory recovery" to ensure reliability, accuracy, and authenticity of expedition events and archival data, producing one set of audio-visual materials from South China Sea surveys.

2.2.3 Analysis and Publications The project uniformly plans publication of one series of South China Sea survey monographs. For hydrometeorological historical data reorganization, data undergo proofreading, quality control, and standardization to produce format-standardized data products. Historical and modern products are compared and analyzed to produce the *South China Sea Physical Oceanography Atlas* and *South China Sea Marine Meteorology Atlas*. Marine chemistry data reorganization involves proofreading, quality control, and standardization using unified mapping tools and legends to produce planar, sectional, vertical, or time-series distribution maps of chemical elements, forming the *South China Sea Marine Chemical Elements Atlas*. Marine biology and ecology data reorganization includes proofreading, quality control, and standardization, updating and revising taxonomic information for marine species, compiling a South China Sea biodiversity species catalog, mapping population distributions, and writing monographs such as *South China Sea Plankton* and *South China Sea Macrobenthos*. Fisheries resources survey data reorganization produces distribution maps of South China Sea fishery resources, conducts trend analysis, and writes the monograph *South China Sea Fishery Resources*. Geological environment data reorganization performs data mining to explore correlations and release new value, compiling a comprehensive South China Sea geological atlas and writing monographs including *Geology and Engineering Environment of South China Sea Coral Reefs*. Island reef geological sample data reorganization, based on drilling and survey data since the 1950s, compiles catalogs and identification atlases of reef-building organisms and their diversity from Miocene core samples, produces data atlases of island reef geological samples, and writes four monographs including *Eolian Biogenic Reefs of the Xisha Islands*.

2.3 Scientific Management and Sharing Services

The project implements a "co-construction and sharing" data work mechanism [**?**] to visualize and scientifically manage historical South China Sea survey data and reorganized materials. By establishing internal connections among diverse, heterogeneous data according to marine disciplinary backgrounds and knowledge structures, the project builds the most comprehensive, longest time-spanning, and widest spatial-ranging South China Sea marine scientific investigation comprehensive database. Strengthening research on data integration technology and information-knowledge fusion applications develops capabilities for massive information integration analysis and services oriented toward national strategic needs. The project constructs marine spatial information systems and thematic service systems, develops a new "Digital South China Sea" data management and sharing platform, and provides one-stop sharing services for reorganized South China Sea survey historical data to the entire Chinese Academy of Sciences and nationwide. This significantly enhances data integration, sharing service support, and application effectiveness, providing crucial foundational data support for South China Sea marine scientific research, government decision-making, and building China into a maritime power.

3 Difficulties and Strategies

Since the 1950s, over 50 years of South China Sea scientific investigation and research have produced massive quantities of measured data and samples across marine geology, geomorphology, geography, surveying, geophysics, geochemistry, biology, ecology, chemistry, physics, hydrology, and meteorology, accumulating vast amounts of firsthand data and important original records. Particularly with evolving situations in the South China Sea, historical survey data from areas like Huangyan Island and the Wan'an Tan Basin are especially precious. However, valuable historical data have long remained scattered across various expeditions, projects, or institutions, with some original materials held by individual scientists. Without electronic data technology at the time, these materials were never effectively integrated or standardized, lacking completeness. Due to the long time span, personnel changes, and deaths, some data have been lost or face imminent loss. Archival transfers, relocations, and expired retention periods have made some materials unlocatable. Personal recording habits and handwriting have rendered some paper archives illegible. Most historical data exist in non-digital formats—paper, diskettes, magnetic tapes, optical discs—with some severely damaged and unreadable. Additionally, inconsistent archival standards and specifications have resulted in complex and diverse data. All these factors have complicated historical materials and increased the difficulty of systematic reorganization.

To address these challenges, the data center has focused on implementing collaborative mechanisms, constructing resource systems, and developing technical procedures, using information technology to support project implementation.

3.1 Establishing a Collaborative Data Resource Construction Mechanism

Following the "co-construction and sharing" data work mechanism, the project established a collaborative data resource construction mechanism encompassing inter-institutional, project-task, and inter-scientist coordination to better collect and integrate historical data from diverse sources. Oriented toward reorganization objectives and the innovation and national development demands for South China Sea data, the mechanism determines specific content for reorganizing data across different disciplines, including processing existing data to enhance application value and timely supplementation and updates. It establishes data quality evaluation and control methods, develops detailed reorganization technical implementation rules, and standardizes formats to ensure uniform description scales for data information. The mechanism also coordinates division of labor and coverage across projects and expeditions among research groups and facilitates coordination between data producers and managers.

3.2 Constructing a Data Resource System for South China Sea Survey Historical Data Reorganization

Observational elements from South China Sea and affiliated islands and reefs scientific investigations are diverse, obtained through different methods, with variations in precision, format, expression, and data structure, involving different temporal scales, spatial reference systems, and coordinate systems. These characteristics present features of multi-source, multi-state, and multi-type data [**?**]. The data resource system is constructed according to requirements for authenticity, completeness, standardization, non-repetitiveness, and reliability. Based on preliminary data collection and sorting, combined with submitted reorganized materials, the data center has built a data resource system for South China Sea survey historical data reorganization, as shown in Table 2 .

3.3 Constructing a Technical Regulations System for South China Sea Survey Historical Data Reorganization

The standardized reorganization and analysis implementation process for South China Sea and affiliated islands and reefs marine scientific investigation historical data primarily includes digitization, standardization, quality control, and metadata extraction. The project has developed eight regulatory documents: *Technical Regulations for the Systematic Reorganization of Historical Data from Scientific Investigations of the South China Sea and Its Affiliated Islands and Reefs—Part 1: Marine Hydrology and Meteorology*; *Part 2: Marine Chemistry*; *Part 3: Marine Biological Ecology*; *Part 4: Marine Fishery Resources Survey*; *Part 5: Island Reef Geological Environment*; *Part 6: Island Reef Geological Samples*; *Part 7: Scientific Data/Atlas Metadata Standards*; and *Part 8: Scientific Data/Atlas Documentation Writing Specifications*. These documents standardize technical details for implementing systematic reorganization of historical investigation data across disciplines.

3.3.1 Historical Data Digitization This step primarily achieves paper material electronic conversion. Scanned materials must match the "Electronic and Reorganization Project Inventory" in project documents with assigned serial numbers. Each project forms one folder, with multiple paper documents processed and merged into single PDF or DOC files. Scanned pages must be flat and clear.

3.3.2 Historical Data Standardization In this step, six research groups extract data according to the "Electronic and Reorganization Project Inventory," processing materials by category to form reorganized format data text files. For GIS compatibility, CSV format is used for data storage; other specialized data retain professional formats while providing data reading software. One project may extract one or multiple digital data entities. Data entity files are named following the convention: "Submitting Unit-Data Source (Project Name/Observation Method)-Data Name (Instrument Type)-Start

Date (YYYYMMDD)-End Date (YYYYMMDD).File Extension."

For dataset standardization, the project developed data reorganization technical regulations and standard data templates to process over 50 years of materials from different sources into unified datasets with consistent formats, encoding, units, and references. Special emphasis is placed on including station latitude-longitude values, observation times, observation elements, and dimensions in observational data reorganization.

3.3.3 Quality Control of Reorganized Data In this step, six research groups determine quality control methods and procedures by discipline and category, performing quality control on reorganized data. Comprehensive quality inspections identify and resolve potential issues, flag errors, and produce technical reports, data analysis reports, or data correction papers to ensure reasonable quality control and guarantee that reorganized data can be directly applied to scientific research.

3.3.4 Metadata Extraction Core metadata cataloging for reorganized datasets follows the core metadata standards for basic science and technology work data and the metadata cataloging specifications of the National Earth System Science Data Center. Emphasis is placed on temporal span, spatial coverage, acquisition scale, spatial reference and coordinates, spatiotemporal expression methods, scale, and spatiotemporal resolution. This ultimately forms "six-in-one" shared release data comprising metadata, data entities, data documentation, data thumbnails, data samples, and data classification [**?**].

Since China launched the Special Program for Basic Science and Technology Work in 1999, hundreds of projects have been supported, accumulating substantial foundational scientific data, rescuing and reorganizing precious historical scientific materials, and developing industry standards and reference materials [**?**]. However, most historical data from completed basic work projects remain scattered among institutions without effective integration or standardized reorganization, with some data facing loss [**?**]. The State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, conducted data submission for 133 special program projects initiated between 2006–2012 through the key project "Integration and Standardized Reorganization of Basic Science and Technology Work Data" (2013FY110900). By July 2017, 72 projects had completed data submission, accounting for 54.14% of total projects [**?**]. Compared with this project, the "Reorganization of Historical Data from Scientific Investigations of the South China Sea and Its Affiliated Islands and Reefs" faces greater implementation difficulty, with historical data spanning over 50 years from 1950–2015, involving more than 400 projects, and distributed across over ten government agencies, research institutes, and universities, requiring material collection, coordination, overcoming difficulties, systematic reorganization, and comprehensive integration.

Historical South China Sea survey data serve as important strategic resources and new production factors supporting China's maritime power strategy, with non-renewable characteristics that necessitate urgent systematic rescue and reorganization. Scientific data management aims to enable open sharing, which essentially realizes the opening and common use of scientific data resources. Opening is the prerequisite; common use is the purpose. Only through common use can scientific data resources maximize their potential scientific, social, and economic value through extensive social services, and only through common use can repeated national investment for the same purposes be reduced to achieve maximum investment benefits [**?**].

The "Reorganization of Historical Data from Scientific Investigations of the South China Sea and Its Affiliated Islands and Reefs" has defined specific reorganization content, with all scanned materials currently in the integration phase. According to the overall workflow deployment, the project is being implemented intensively, with some reorganized datasets already meeting submission requirements and entering quality control and analysis stages. The project has proposed corresponding strategies for reorganization difficulties, conducting beneficial exploration and practice in collaborative mechanisms, resource systems, and technical regulations. Construction of the comprehensive database and its technical applications will be discussed in subsequent articles, with plans to openly share long time-series, multi-disciplinary, multi-dimensional South China Sea marine data resources by 2023.

References

Xu Hongru. (2008-11-01). An epoch-making national marine census—Commemorating the 50th anniversary of the National Comprehensive Marine Survey [EB/OL]. https://blog.sciencenet.cn/blog-74364-45008.html

Xu Du. 2010. 1958–1960: National Comprehensive Marine Survey [J]. *Marine Sciences* (04), 109-110. (in Chinese).

South China Sea Institute of Oceanology, Chinese Academy of Sciences (Ed.). 1982. *Comprehensive Survey Research Report of the South China Sea (I)* [R]. Beijing: Science Press: 1.

Liu Zhaoshu. 1985. New progress in comprehensive survey and research of the northeastern South China Sea [J]. *Marine Science Bulletin* (03), 55-58. (in Chinese).

Zhao Huanting, Wang Lirong, Yuan Jiayi. 2017. Overview of scientific investigations and research on the South China Sea islands—Commemorating the 70th anniversary of the recovery of the South China Sea Islands (2) [J]. *Tropical Geography* (05), 649-658. ZHAO Huanting, WANG Lirong, YUAN Jiayi, 2017. Scientific Investigations and Research on the South China Sea Islands: The 70th Anniversary of Recovery of the South China Sea Islands (2) [J]. *Tropical Geography*, 2017, 37(5): 649-658 (in Chinese with English abstract).

Xu Chao, Li Sha, Chen Rongyu, He Yunkai, Chen Shaoyong. 2016. Data management and sharing of scientific investigation on South China Sea marine sections [J]. *Marine Information*, 2016(1): 19-28 (in Chinese).

Bai Yan, Yang Yaping. 2020. Practice and enlightenment of data submission for Special Program of Basic Science and Technology Work [J]. *China Science & Technology Resources Review*, 2020, 52(04): 70-79.

Yang Yaping, Jiang Hou, Sun Jiulin. 2020. Practice of scientific data sharing: A case study of the national earth system science data center [J]. *Journal of Geoinformation Science*, 22(6): 1358-1369. Yang YP, Jiang H, Sun J L, 2020. Practice of scientific data sharing: A case study of the national earth system science data center [J]. *Journal of Geoinformation Science*, 22(6): 1358-1369 (in Chinese with English abstract).

Zhu Yunqiang, Sun Kai, Yang Yaping, Wang Xiaoxuan, Yue Xiafang, Zhu Huazhong, Song Jia, He Mingyue, Chen Youqi, Peng Keyin, Fu Lei, Yang Yanchen, Yang Mei, Zou Jinqiu, Chen Yan, Li Jinbin, Liu Yan, Lei Lei, Yang Jie, Li Weirong, Zhang Jinqu. 2017. Data resources collection and reorganization for National Special Program on Basic Works for Science and Technology of China [J]. *China Science & Technology Resources Review*, 49(05): 12-20. ZHU Yunqiang, SUN Kai, YANG Yaping, WANG Xiaoxuan, YUE Xiafang, ZHU Huazhong, SONG Jia, HE Mingyue, CHEN Youqi, PENG Keyin, FU Lei, YANG Yanchen, YANG Mei, ZOU Jinqiu, CHEN Yan, LI Jinbin, LIU Yan, LEI Lei, YANG Jie, LI Weirong, ZHANG Jinqu, 2017. Data Resources Collection and Reorganization for National Special Program on Basic Works for Science and Technology of China [J]. *CHINA SCIENCE & TECHNOLOGY RESOURCES REVIEW*, 49(05): 12-20 (in Chinese with English abstract).

Huang Dingcheng. 2003. Theoretical foundation and mechanism of the scientific data-sharing [J]. *China Basic Science*, 2003(2): 22-27. Huang Dingcheng, 2003. Theoretical Foundation and Mechanism of the Scientific Data-sharing [J]. *Chian Basic Science*, 2003(2): 22-27 (in Chinese with English abstract).

Note: Figure translations are in progress. See original paper for figures.