

中国科学院水生生物研究所科学数据中心: 助推大数据时代水生生物学研究新发展

缪荣丽 魏朝军 王宝强 彭新奎 陈凯 熊凡 方成池 甘小妮 曾宏辉

SCIENCE DATA CENTER OF INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES: PROMOTING THE NEW DEVELOPMENT OF HYDROBIOLOGY RESEARCH IN THE ERA OF BIG DATA

MIAO Rong-Li, WEI Chao-Jun, WANG Bao-Qiang, PENG Xin-Kui, CHEN Kai, XIONG Fan, FANG Cheng-Chi, GAN Xiao-Ni, ZENG Hong-Hui

在线阅读 View online: <https://doi.org/10.7541/2023.2023.0050>

您可能感兴趣的其他文章

Articles you may be interested in

水生所与三峡和葛洲坝水利工程中的鱼类保护

CONTRIBUTIONS OF THE INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES TO FISH PROTECTION IN THE CONSTRUCTION OF THREE GORGES DAM AND GEZHOUBA DAM

水生生物学报. 2020, 44(5): 1040–1044 <https://doi.org/10.7541/2020.120>

天鹅洲故道水生态研究进展

PROGRESS IN STUDIES ON WATER ECOLOGY IN TIAN' E ZHOU OXBOW

水生生物学报. 2017, 41(4): 935–946 <https://doi.org/10.7541/2017.117>

微藻种质资源库——藻类科学研究和产业发展的重要平台

MICROALGAL CULTURE COLLECTION: FUNDAMENTAL PLATFORM FOR ALGAL RESEARCH AND INDUSTRIAL DEVELOPMENT

水生生物学报. 2020, 44(5): 1020–1027 <https://doi.org/10.7541/2020.118>

菱湖鱼病工作站: 现代科学改造中国传统养鱼业的序曲

LINGHU FISH DISEASE RESEARCH STATION: THE OVERTURE OF MODERN SCIENTIFIC TRANSFORMATION OF CHINESE TRADITIONAL FISH FARMING INDUSTRY

水生生物学报. 2020, 44(5): 1028–1039 <https://doi.org/10.7541/2020.119>

基于多源数据的“安全公正空间”在大通湖流域环境管理中的应用

APPLICATION OF “SAFE AND JUST OPERATING SPACE” BASED ON MULTI-SOURCE DATA IN ENVIRONMENTAL MANAGEMENT OF DATONG LAKE CATCHMENT

水生生物学报. 2021, 45(1): 197–205 <https://doi.org/10.7541/2021.2020.064>

从公共数据库中利用数据挖掘的方法开发15条黄金鲈EST-SSR标记

DEVELOPMENT OF FIFTEEN EST-SSR MARKERS IN YELLOW PERCH *PERCA FLAVESCENS* BY DATA MINING FROM PUBLIC EST SEQUENCES

水生生物学报. 2018, 42(6): 1186–1193 <https://doi.org/10.7541/2018.145>



关注微信公众号，获得更多资讯信息

中国科学院水生生物研究所科学数据中心： 助推大数据时代水生生物学研究新发展

缪荣丽 魏朝军 王宝强 彭新奎 陈凯 熊凡 方城池 甘小妮 曾宏辉

(中国科学院水生生物研究所, 武汉 430072)

摘要: 中国科学院水生生物研究所(以下简称“水生所”)是研究内陆水体生命过程、生态环境保护与生物资源利用的综合性学术机构。自20世纪50年代开始,水生所科研人员率先有计划地针对全国主要水体开展了实地科学考察,产生了大量的水生生物原始生态学数据;随着显微拍照和测序技术的进步,调查还产生了各种浮游生物显微照片和大量的遗传数据。这些数据为水生态监测服务系统和数据库的构建提供了数据支持,有助于推进水生生物学研究的发展。此外,面向国家在水环境保护、渔业可持续发展方面的重大战略需求,中国科学院水生生物研究所科学数据中心(以下简称水生所科学数据中心)(<http://sdb.ihb.ac.cn/>)聚焦科学数据标准规范、安全可控和开放共享,对上述科学数据进行收集保存、汇交整合、分级管理和分析挖掘,以期为生态环境保护与生物资源保护利用等领域的基础性、战略性和前瞻性的科学研究和技术创新服务。

关键词: 大数据; 水生生物; 水环境保护; 渔业可持续发展; 科技创新

中图分类号: Q178.2 **文献标识码:** A **文章编号:** 1000-3207(2023)10-1719-06



水生生物科学数据是指在水生态环境保护、水生生物多样性和资源利用研究中产生的原始、基础性数据,这些数据支撑了国内水生生物多样性、水体评价、水环境修复等方面的重大科学研究和发现,为揭示国内水环境变化的影响因素和响应规律,提高水华、鱼病灾害预测和预警能力提供了数据支持。

为充分发挥水生生物科学数据的系统分析利用和共享价值,中国科学院水生生物研究所(以下简称“水生所”)于2019年6月成立了水生生物数据分析管理平台,并设立水生生物大数据中心,该数据中心于2021年12月23日通过中科院的首批认定,成为中科院所级科学中心。数据中心建立了水生生物数据规范管理和层级共享机制,汇集整合中科院及国内本学科领域主要科学数据资源,汇交和保存本学科领域国家财政支持的科技项目产生的科学数据,形成了规范、系统、具有时间尺度和空间覆盖范围的水生生物特色数据集共计58.7 TB。

在当今大数据时代背景下,科学数据作为科研

成果的信息载体,是国家科技创新和社会经济发展的基础和保障,是促进科技创新和科技进步的重要战略资源。党的十八大以来,习近平总书记围绕生态文明建设作出了一系列重要论断,指出要“提升生态系统多样性、稳定性、持续性”。随着中国可持续发展战略的实施以及人民生活水平质量的提高,人们对水生态环境的保护越来越重视。水生生物作为水生态系统的主体,是研究水体演变的关键依据,也是水体生态系统健康修复与水生态文明建设的重要内容,是水生态环境保护的重要对象。

1 水生所特有的“黄金资源”——全国水生生物调查数据

水生所科学数据中心拥有自20世纪50年代至今,全国33个省级行政区(21个省、5个自治区、4个直辖市)范围内(图1),包括江、河、湖、库及湿地等各类水体的水生生物调查数据。这些数据是国内淡水生态环境领域历时最长、调查范围最广、调查类群最全面的科学数据,是了解内陆水体

收稿日期: 2023-02-16; 修订日期: 2023-03-24

基金项目: 国家自然科学基金(32270489)资助 [Supported by the National Natural Science Foundation of China (32270489)]

作者简介: 缪荣丽(1985—),女,实验师;主要从事藻类多样性与基因调控研究。E-mail: mronli@ihb.ac.cn

通信作者: 曾宏辉,高级工程师; E-mail: zhh@ihb.ac.cn

生命过程、生物资源利用和生态环境保护方面的宝贵资源。

自20世纪50年代开始,水生所的科技工作者就在无锡太湖^[1]、武汉东湖^[2]、丹江口水库^[3]、三峡水库^[4]、黑龙江和松花江流域^[5]、长江流域和淮河流域^[6,7]等地进行了大量的水生生物和渔业资源调查研究,调查轨迹遍布全国大部分水域,并在此基础上制定了我国首部《湖泊调查基本知识》规程^[7],推动了国家对水生生物科学研究的重视,引领了水生生物学的快速发展。新中国成立以来,青藏高原研究成为了我国国家战略层面的科技任务。20世纪60年代,水生所的科技工作者积极响应国家科技战略要求,实现了对世界第一高峰——珠穆朗玛峰(珠峰)地区水生生物的首次调查^[8-10];20世纪70年代以来,随着第一次青藏高原综合科学考察拉开序幕,我们在号称“世界屋脊”的青藏高原开展了多尺度、大范围的科学考察研究^[11-13]。对该地区的鱼类、藻类、水生无脊椎动物等水生生物进行了专题性研究,并整理、总结了相关科学数据和一手科研资料,形成了一批以《珠穆朗玛峰地区科学考察报告》^[14]《青藏高原科学考察丛书》^[12,15]《青藏高原横断山区科学考察丛书》^[16]和《中国西藏硅藻》^[17]为代表的宝贵成果。

除此以外,水生所在武汉东湖^[18]、三峡水库、

丹江口库区^[19]、洪泽湖及赤水河^[20]等地设有长期的野外科学观测站/科研基地,这些科研基地为水生所研究人员在东湖(1956年至今)^[2,21,22]、长江流域、淮河流域(1956年至今)^[7,23]、三峡库区(1958年至今)^[4,24]及丹江口库区(1958年至今)^[3]等地区进行长期连续的生态监测与调查提供了基础设施和科研保障。科学观测站的建立也直接推动和加速了水生所对水环境指标、水生生物的时空分布格局及水生生物的食物网等基础资源数据的获取。

以上科学数据是水生所特有的“黄金资源”,曾为青藏高原隆起的生物地理学研究、我国淡水渔业发展、三峡工程、南水北调等工程的方案实施、各地饮用水源地探测管理、流域管理规划、珍稀水生生物物种保育、水体富营养化管理、水华发生预警与研究等做出了重要贡献。

2 数据资源整合助力水生生物科学新发展

随着科技的快速发展,大数据时代的数据爆炸为水生生物科学的发展带来了机遇和挑战。由数据驱动的科技创新改变了各学科当前的研究范式。

水生生物作为水生态系统的主体,是研究水体演变的最重要依据,也是维护水生态系统健康的关键,没有水生生物的水是没有生命的“一潭死水”。“水”里有什么?“水”好不好?如何保护“水”?针对上

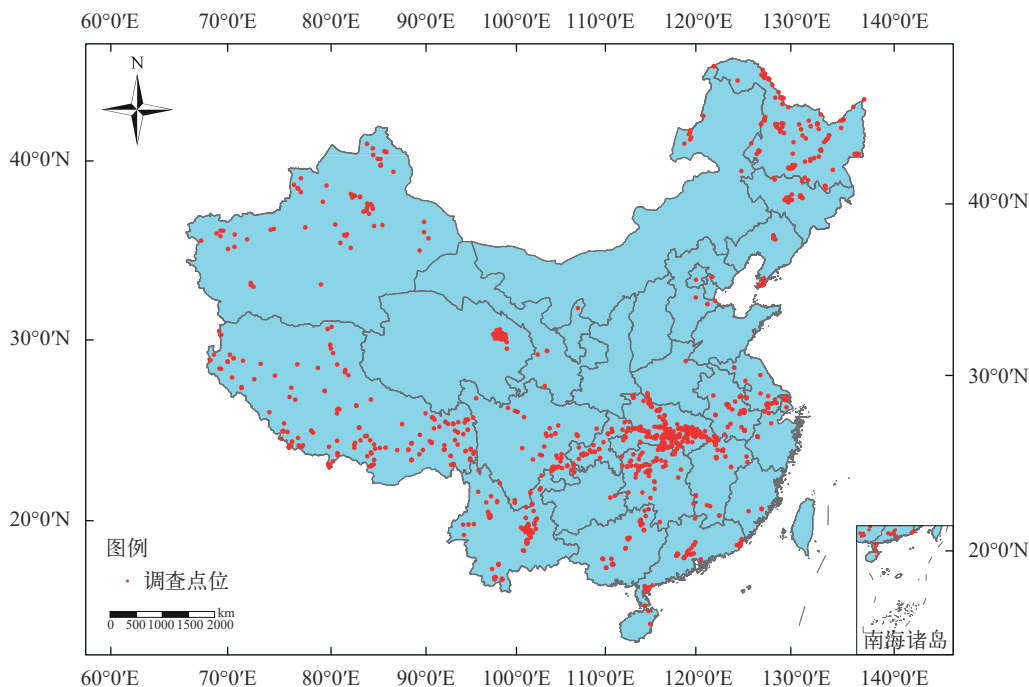


图1 水生生物历史调查数据分布图

Fig. 1 The distribution of historical data of hydrobiological survey

根据中华人民共和国自然资源部监管的标准地图编辑[审图号为GS(2016)1569]

Edited based on a standard map supervised by the Ministry of Natural Resources of the People's Republic of China [No. GS (2016) 1569]

各类预警的能力, 为政府决策和监测环境管理提供技术支持, 为公众提供全方位的环境状况信息。结合eDNA技术^[31, 32]提供物种遗传资源信息库服务, 入库信息涵盖物种DNA序列、图像、采集地点和日期、采集和鉴定人等, 以期达到调查数据的高效管理、历史追溯、分析共享及决策预警的作用。

2.3 构建水生生物图像智能识别系统

水生生物物种鉴定中心现存海量物种图像数据, 专业鉴定人员可以通过人工对物种特征点进行标注, 生成物种形态学特征点训练集, 通过该训练集开发以卷积神经网络算法为核心的机器学习模型, 继而进行图像识别, 实现对水生生物的人工智能鉴定。基于图像识别技术再结合多通路浮游生物自动进样技术、多景深显微拍摄与图片处理技术等, 构建水生生物图像智能识别系统(图2C)。并将《SL733-2016内陆水域浮游植物监测技术规程》《HJ1216-2021 水质 浮游植物的测定 0.1 mL计数框-显微镜计数法》《SCT 9402-2010 淡水浮游生物调查技术规范》和《水和废水监测分析方法》(第四版)第五篇《水和废水的生物监测方法》等技术指南应用到智能识别自动计数、分析输出等计算模块中。通过与专业技术人员的鉴定结果进行比对, 制定浮游生物智能识别检测技术标准, 并进行技术试点和推广, 提高水生生物检测速度和准确度, 减少人工专业技术差异带来的误差。

3 展望

水生所科学数据中心还将整合学科优势, 制定水生态监测相关的行业标准, 如浮游生物智能识别监测技术标准, 及eDNA采样、保存、运输、提取、分析等全过程的统一标准等, 保证水生生物检测和数据分析报告的稳定性和准确性。与此同时, 数据中心还将坚持务实发展, 在确保数据安全的前提下, 加强数据平台建设, 构建集动力数值模型建设、多源遥感数据重构、水华预警技术、水质评价技术及物种遗传资源信息库服务为一体的水生生物科学大数据中心。

参考文献:

- [1] Guo Y. Investigation of Wuli Lake [J]. *Chinese Science Bulletin*, 1951, 2(10): 1103-1106. [郭扬. 五里湖调查工作报告 [J]. *科学通报*, 1951, 2(10): 1103-1106.]
- [2] Rao Q Z, Zhang Z S. Ecological changes of phytoplankton in Lake Donghu, Wuhan, during 1956-1975 and the eutrophication problem [J]. *Acta Hydrobiologica Sinica*, 1980, 7(1): 1-17. [饶钦止, 章宗涉. 武汉东湖浮游植物的演变(1956-1975年)和富营养化问题 [J]. *水生生物学报*, 1980, 7(1): 1-17.]
- [3] Borutsky E. B, Wu H W, Pai K T, et al. Hydrobiological survey of the region of the projected dam reservoir of Tankiangkou, with propositions for fisheries management [J]. *Acta Hydrobiologica Sinica*, 1959(1): 33-56. [E. B. 波鲁茨基, 伍献文, 白国栋, 等. 丹江口水库库区水生生物调查和渔业利用的意见 [J]. *水生生物学集刊*, 1959(1): 33-56.]
- [4] Borutsky E. B, Wang Q L, Chen S Z, et al. Hydrobiological survey of the Three Gorges Reservoir Area of Yangtze River, with planting propositions for fisheries management [J]. *Acta Hydrobiologica Sinica*, 1959(1): 1-32. [E. B. 波鲁茨基, 王乾麟, 陈受忠, 等. 长江三峡水库库区水生生物调查和渔业利用的规划意见 [J]. *水生生物学集刊*, 1959(1): 1-32.]
- [5] Villigin E. B. Brief opinions on fishery in Heilongjiang Province [J]. *Acta Hydrobiologica Sinica*, 1959(2): 119-127. [E. B. 维利金. 对黑龙江省渔业的简短意见 [J]. *水生生物学集刊*, 1959(2): 119-127.]
- [6] Chu S L. On fishes of Ichang, with notes on their distribution in the Yangtze River [J]. *Acta Hydrobiologica Sinica*, 1955(2): 81-95. [褚新洛. 宜昌的鱼类及其在长江上下游的分布 [J]. *水生生物学集刊*, 1955(2): 81-95.]
- [7] Rao Q Z. The Basics of Lake Survey [M]. Beijing: Science Press, 1956: 1-367. [饶钦止. 湖泊调查基本知识 [M]. 北京: 科学出版社, 1956: 1-367.]
- [8] Rao Q Z. Some fresh-water algae from Southern Tibet [J]. *Oceanologia et Limnologia Sinica*, 1964, 6(2): 169-192. [饶钦止. 西藏南部地区的藻类 [J]. *海洋与湖沼*, 1964, 6(2): 169-192.]
- [9] Rao Q Z, Zhu H Z, Li R Y. Summary of algae in Qomolangma area of South Tibet [J]. *Chinese Science Bulletin*, 1973, 18(1): 30-32. [饶钦止, 朱惠忠, 李尧英. 我国西藏南部珠穆朗玛峰地区藻类概要 [J]. *科学通报*, 1973, 18(1): 30-32.]
- [10] Zhang C L, Yue Z H, Huang H J. Notes on fishes of Southern Tibet, China, with proposal of one new genus, Tetrostichodon [J]. *Acta Zoologica Sinica*, 1964, 16(2): 111-121. [张春霖, 岳佐和, 黄宏金. 西藏南部的鱼类 [J]. *动物学报*, 1964, 16(2): 111-121.]
- [11] Jiang X Z, Shen Y F, Gong X J. Scientific Investigation Series on the Qinghai-Tibet Plateau-Aquatic invertebrates of Tibet [M]. Beijing: Science Press, 1983: 1-491. [蒋燮治, 沈毓芬, 龚循矩. 青藏高原科学考察丛书——西藏水生无脊椎动物 [M]. 北京: 科学出版社, 1983: 1-491.]
- [12] Li R Y. Thermal cyanophytes of the Xizang Plateau and Hengduan Mountain Region [J]. *Acta Hydrobiologica Sinica*, 1985, 9(3): 264-279. [李尧英. 青藏高原及横断山区的温泉蓝藻 [J]. *水生生物学集刊*, 1985, 9(3): 264-279.]
- [13] Wu X W, He M C, Chu X L. On the fishes of Sisoridae from the region of Xizang [J]. *Oceanologia et Limnologia Sinica*, 1981, 12(1): 74-79. [伍献文, 何名巨, 褚新洛.

- 西藏地区的鲢科鱼类 [J]. 海洋与湖沼, 1981, 12(1): 74-79.]
- [14] Tibet Scientific expedition team, Chinese Academy of Sciences. Report of scientific investigation in Qomolangma Region 1966-1968 Biology and alpine Physiology [M]. Beijing: Science Press, 1974: 75-172. [中国科学院西藏科学考察队. 珠穆朗玛峰地区科学考察报告 1966—1968 生物与高山生理 [M]. 北京: 科学出版社, 1974: 75-172.]
- [15] Li R Y, Wei Y X, Shi Z X, *et al.* Scientific Investigation Series on the Qinghai-Tibet Plateau-Algae of Tibet [M]. Beijing: Science Press, 1992: 1-509. [李尧英, 魏印心, 施之新, 等. 青藏高原科学考察丛书——《西藏藻类》[M]. 北京: 科学出版社, 1992: 1-509.]
- [16] Chen Y Y. Scientific Investigation Series in Hengduan Mountain Area of Qinghai-Tibet Plateau-Fish in Hengduan Mountain Region [M]. Beijing: Science Press, 1988: 1-364. [陈宜瑜. 青藏高原横断山区科学考察丛书——横断山区鱼类 [M]. 北京: 科学出版社, 1988: 1-364.]
- [17] Zhu H Z, Chen J Y. Diatoms in Tibet, China [M]. Beijing: Science Press, 2000: 1-353. [朱蕙忠, 陈嘉佑. 中国西藏硅藻 [M]. 北京: 科学出版社, 2000: 1-353.]
- [18] Huang X F. An important base of limnology research in China-East Lake Ecosystem Experimental Station of Chinese Academy of Sciences [J]. *Lake Science*, 1994, 6(1): 96-96. [黄祥飞. 我国湖沼学研究的重要基地——中国科学院东湖湖泊生态系统试验站 [J]. 湖泊科学, 1994, 6(1): 96-96.]
- [19] Zhang H. Research base of Institute of Hydrobiology, Chinese Academy of Sciences in Danjiangkou Reservoir Area settled in Danjiangkou [N]. Fishery Guide to be Rich, 2009(11): 9-9. [张华. 中科院水生所丹江口库区科研基地落户丹江口 [N]. 渔业致富指南, 2009(11): 9-9.]
- [20] Huang Z L. Free flowing Chishui River-a river with unique characteristics and protection value in the upper reaches of the Yangtze River [N]. China Three Gorges Construction, 2008(2): 10-19. [黄真理. 自由流淌的赤水河——长江上游一条独具特色和保护价值的河流 [N]. 中国三峡建设(科技版), 2008(2): 10-19.]
- [21] Liu J k, Xie P. Unraveling the enigma of the disappearance of water bloom from the East Lake (Lake Donghu) of Wuhan [J]. *Resources and Environment in the Yangtze Basin*, 1999, 8(3): 312-319. [刘建康, 谢平. 揭开武汉东湖蓝藻水华消失之谜 [J]. 长江流域资源与环境, 1999, 8(3): 312-319.]
- [22] Lu M, Xie P. Studies on the structure of crustacean zooplankton in Houhu Region of Lake Donghu, Wuhan [J]. *Acta Hydrobiologica Sinica*, 2002, 26(2): 123-129. [鲁敏, 谢平. 武汉东湖后湖区浮游甲壳动物群落结构的研究 [J]. *水生生物学报*, 2002, 26(2): 123-129.]
- [23] Chen P X, Deng Z L, Yu Z T, *et al.* Fishes of Yangtze River [Z]. Hubei Science Conference Award, 1978. [陈佩薰, 邓中彝, 余志堂, 等. 长江鱼类 [Z]. 湖北省科学大会奖, 1978.]
- [24] Liu H Z, Li M Z, Chang T. Contributions of the Institute of Hydrobiology, Chinese Academy of Sciences to fish protection in the construction of Three Gorges Dam and Gezhouba Dam [J]. *Acta Hydrobiologica Sinica*, 2020, 44(5): 1040-1044. [刘焕章, 黎明政, 常涛. 水生所与三峡和葛洲坝水利工程中的鱼类保护 [J]. 水生生物学报, 2020, 44(5): 1040-1044.]
- [25] Fan G, Song Y, Yang L, *et al.* Initial data release and announcement of the 10000 Fish Genomes Project (Fish10K) [J]. *Gigascience*, 2020, 9(8): giaa080.
- [26] Miao W, Song L, Ba S, *et al.* Protist 10000 genomes project [J]. *Innovation*, 2020, 1(3): 100058.
- [27] Yang L D, Xu Z T, Zeng H H, *et al.* FishDB: an integrated functional genomics database for fishes [J]. *BMC Genomics*, 2020, 21(1): 801.
- [28] Yang W T, Zhang J, Yan G X, *et al.* *Tetrahymena* Functional Genome Database Updated in 2016 on Resources Construction of Life Cycle Transcriptome, Meiosis Transcriptome and Phosphoproteome Data [J]. *Genomics and Applied Biology*, 2016, 35(6): 1443-1450. [杨文涛, 张晶, 闫冠雄, 等. 四膜虫功能基因组数据库增量更新2016: 生活史和减数分裂转录组及磷酸化蛋白组资源建设 [J]. 基因组学与应用生物学, 2016, 35(6): 1443-1450.]
- [29] Xiong J, Lu X Y, Lu Y M, *et al.* *Tetrahymena* Gene Expression Database (TGED): a resource of microarray data and co-expression analyses for *Tetrahymena* [J]. *Science China Life Sciences*, 2011, 54(1): 65-67.
- [30] Chen K, Fang C C, Wu Z G, *et al.* AeDNA: aquatic environmental DNA database [J]. *Acta Hydrobiologica Sinica*, 2022, 46(11): 1741-1747. [陈凯, 方成池, 吴志刚, 等. AeDNA: 水生生物eDNA数据库 [J]. *水生生物学报*, 2022, 46(11): 1741-1747.]
- [31] Xiong F, Shu L, Zheng H H, *et al.* Methodology for fish biodiversity monitoring with environmental DNA metabarcoding: the primers, databases and bioinformatic pipelines [J]. *Water Biology and Security*, 2022, 1(1): 1-10.
- [32] Ruppert K M, Kline R J, Rahman M S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA [J]. *Global Ecology and Conservation*, 2019(17): e00547.

SCIENCE DATA CENTER OF INSTITUTE OF HYDROBIOLOGY, CHINESE ACADEMY OF SCIENCES: PROMOTING THE NEW DEVELOPMENT OF HYDROBIOLOGY RESEARCH IN THE ERA OF BIG DATA

MIAO Rong-Li, WEI Chao-Jun, WANG Bao-Qiang, PENG Xin-Kui, CHEN Kai, XIONG Fan, FANG Cheng-Chi, GAN Xiao-Ni and ZENG Hong-Hui

(Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China)

Abstract: The Institute of Hydrobiology (IHB) of the Chinese Academy of Sciences (CAS) is a comprehensive academic research institution engaged in the study of the life process, ecological environment protection and biological resources utilization of inland water. Since the 1950s, researchers in IHB have been at the forefront of conducting hydrobiological surveys in main waters of China, and vast amounts of original survey data have been generated. With the advancement of photomicrography and sequencing technology, a variety of plankton micrographs and a wealth of genetic data have been produced, which provide data support for the construction of monitoring service system and database of hydro-ecology, and promote the development of the hydrobiology research. To maximize the value of this data, the Scientific Data Center of IHB, CAS (<http://sdb.ihb.ac.cn/>) addresses the major strategic needs of the state in water environment protection and the sustainable development of fisheries, focusing on scientific data standards, security and control, open sharing. To serve the ecological environment protection and biological resources protection and utilization of basic, strategic and forward-looking scientific research and technology innovation, the Scientific Data Center of IHB collects, saves, integrates, manages, analyzes and digs the aquatic biological data.

Key words: Big data; Hydrobiology; Water environment protection; Fishery sustainable development; Scientific and technological innovation

