

# 中国海水经济 (养殖) 动物种质资源 元数据库的概念、内容与体系<sup>\*</sup>

李 军 尤 锋

(中国科学院海洋研究所, 青岛 266071)

随着中国海水养殖业的大力发展, 其种质资源保护和种质保存的研究日趋广泛, 由此获得的数据越来越多, 人们对这些数据的共享要求也越来越迫切。种质资源元数据库建立的目的在于将大量分散的资料在新的层次上系统地加以收集、整理和补充, 使动物种质信息管理规范化, 提高整体管理水平, 并逐步实现网络化和资源共享, 使海水经济动物种质资源和种质库的研究成果得到充分有效地利用, 发挥其科学价值和社会价值。所以, 为有效地进行海水经济动物遗传多样性保护, 为向海水养殖业和科学研究提供优良种质信息, 为政府部门提供决策信息, 促进中国养殖业的可持续发展, 建立海洋经济动物种质资源信息系统元数据库是很有必要和迫切的。

目前已建立起来的生物相关的各类元数据库有: 中国农业信息元数据库系统 (符海芳等, 2002)、中国极地科学数据库系统 (张侠等, 2002)、林业资源数据信息网络系统 (易浩若, 2003) 和中国生物多样性元数据库 (徐海根, 1999), 其中农作物和陆上生物等的种质元数据库均已建立, 并正在发挥着重要作用, 而海洋方面的共享数据则只包括海洋资源数据库、海洋环境数据库和海洋经济统计信息数据库, 海洋经济 (主要是养殖) 动物如鱼、虾、贝类等种质资源和种质保存的数据库尚没有建立, 其种质资源信息系统元数据库则更无人涉足。作者将就元数据的概念、元数据库的组成和建立以及海洋经济动物种质资源元数据库建立的策略进行介绍, 以期为建立中国海水经济动物种质信息数据库提供有益的参考依据。

## 一、元数据的概念

元数据实际上并不是个新的概念, 只是由于近年来各专业领域信息化和数字化的发展, 产生了海量的数据, 为了更有效地管理、维护、查找和利用这些数据, 元数据技术的开发和利用才被人们广泛重视起来。

<sup>\*</sup> 通讯作者: 李军, 男, 博士生导师, 研究员, E-mail: junli@ms.qdio.ac.cn

收稿日期: 2004 年 7 月 16 日。

## 1. 元数据的概念

一般来讲, 元数据 (metadata) 是用于提供某种资源的有关信息的结构数据 (structured data), 说明数据的内容、质量、条件、状况和其他相关特性的背景信息 (EEA 1993, UNEP 1992, NMIQ 1981)。元数据有时也称为描述数据 (data about other data) 和诠释数据或中介数据, 描述数据的内容和结构。其使用目的在于: 识别资源, 评价资源, 追踪资源在使用过程中的变化, 实现简单高效地管理大量网络化数据, 实现信息资源的有效发现、查找、一体化组织和对使用资源的有效管理。例如, 检索数据时需要知道数据的主要内容, 因此元数据中一般包含对数据的描述信息和质量信息, 质量信息是为了保证数据的可比性和可靠性。另外, 元数据还应包括数据的访问信息、数据的可获得性信息、数据的访问限制等等内容。在数据共享中, 有些数据是有条件共享的, 它是《生物多样性公约》资料交换机制 (CHM) 的主要数据形式, 为用户提供信息源的信息, 并提供指向信息源的链路。因此, 它的作用并不是监管, 而是交换与共享。

## 2. 元数据的特点

①是用于描述信息资源的高度结构化数据; ②可以管理和组织信息; ③可以挖掘信息资源; ④帮助人们准确地查找所需要的信息, 并通过比较和对比不同来源数据的相同元数据元素, 获得自己所需要的信息内容 (赵改善等, 2002)。

# 二、元数据库及其组成

元数据的实体包括机构 (指得到认可的、储存和维护有关信息的组织)、数据集 (指保存在一个机构中的数据及伴随文件的集合, 数据的集合则指的是涉及一个具体主题或地理区域一个或一系列数据成员) 和成员 (被看作相当于一个数据文件、书面报告和地图, 是最低层次的数据实体, 包括不同的类型, 如报告、地图或数据库等) 3 个等级序列。

## 1. 元数据库的组成

一个为保存和处理元数据而设计和建立的数据库, 就叫元数据库 (metadata base)。元数据库的组成包括以下几个方面 (徐海根, 1999; 徐海根等, 1999; 纪力强, 1999; Pinborg 1992)。

### (1) 关于部门和机构的元数据项

主要信息有: 机构信息、机构概述及机构职能的信息 (如部门和机构的名称、代码、英文名称、网站、主管部门、类型、主要职能、联系人及其身份、电子邮件和通信地址等)。

### (2) 关于数据集的元数据项

数据集的名称、数据集描述 (如编号、学科、专业、关键词、说明、项目、数据范围)、采集、数据的用途、采集者及其信息、质量控制、建立时间、出版信息等等。

(3) 关于数据集中数据的元数据项

数据集名称、数据库名称、保存地点、建库单位、建库人、所用软件、记录数、数据库结构、数据名称、存储名称、类型、说明、存储长度、单位、精密度、准确度、数据的上下限、获得数据的手段、时间或环境等。

(4) 专著信息

作者、书名、版次、出版社、出版地、出版年代、字数(万字)、主要内容、主要特点、关键词、收藏地点等。

(5) 图件信息

图件名称、主持单位/专家、成图年代、出版年代、出版社、出版地、比例尺、覆盖区域、主要特点、关键词、收藏地点等。

2. 中国动物元数据库简介

目前中国动物物种编目数据库, 主要收录分布在中国境内的动物物种的基本信息, 已经收集了分布在中国动物物种哺乳类、鸟类、爬行类、两栖类、鱼类、蜘蛛、蟹类及部分昆虫共 14 000种及亚种的基本信息。记录内容包括每个物种和亚种的分类代码, 分类阶元, 拉丁学名, 中文名称, 定名人, 定名年代, 原始文献, 模式产地, 同物异名, 经纬度, 海拔, 地理分布等。在中国濒危和保护动物数据库中记录了中国濒危动物物种基本信息, 有包括 660余种(及亚种)濒危动物的物种代码, 拉丁学名, 中文名, 俗名, 商品名, 英文名, 保护级别和颁布年代, 濒危程度, 受威胁因素, 保护措施, 彩色图片等。

三、元数据库的建立和标准

1. 元数据库的建立

元数据库的建立应当由国家主管部门或经国家主管部门授权的机构实施, 并定期更新数据。这个机构作为元数据库的数据监管人, 具有一般数据监管人的权利, 并履行数据监管人的义务。如果是一个以上的机构建立元数据库, 应当预先商定各自的工作范围, 统一建库的方法和工具, 统一用户界面, 协调工作进度, 避免出现空白领域或重复性劳动。各数据集的数据监管人有义务在数据集完成时将元数据信息交给元数据的数据监管人, 并保证元数据的准确性和完整性。在每次数据集数据更新时, 要将更新的元数据信息交给元数据的数据监管人。每个数据集完成时, 都要建立相应的数据集数据项的元数据库, 与数据集共同保存。数据监管人应保证用户可以方便地得到元数据库。作为数据集的数据监管人, 必须制度化地保证将元数据库的建立和维护作为整个数据集建立和维护工作的一个重要组成部分(WCMC 1996)。

2. 元数据制定标准

制定元数据的标准, 统一元数据的定义和内容, 是当前信息时代数据管理和使用的需求。只有这样, 才能方便不同数据库数据进行交换或共享甚至管理。所以, 通过建立

不同等级的与数据库实体链接的元数据库, 可以规范数据管理, 使数据具有继承性, 达到管理数据、说明数据、了解数据和获得数据的目的, 从而使信息共享成为可能。元数据标准还可以使信息共享的数据库能建立统一的、规范化的元数据库, 保证用户在理解水平上的一致, 有利于数据交换和相互操作; 便于开发适用于不同数据库和多种操作系统的元数据操作工具, 简化元数据的输入、编辑与维护, 方便数据查询与检索, 维护数据的版权 (蒋景瞳等, 1999)。元数据不仅为用户使用数据提供方便, 也使数据拥有者的利益得到保护。因此, 元数据的标准应该尽量覆盖面广、时间序列和空间序列并存、属性数据和空间数据并存、电子数据和书面数据并存等 (徐海根, 包浩生, 2002)。同时, 还要注意数据的简洁性、科学性和适用性。

中国元数据的数据应用研究工作是近年来才在有关领域开展起来的。目前, 比较典型的元数据标准方案有: 国家基础地理信息中心的《中国可持续发展共享元数据标准》; 国家信息中心的《国土资源环境与地区经济信息系统 (NREDIS) 信息共享元数据内容标准草案》; 中国科学院计算机网络信息中心的《科学数据库元数据标准 (讨论稿)》; 中国生态系统研究网络 (CERN) 的《CERN元数据标准 (修订稿)》; 中国科学院动物研究所的《中国生物多样性信息系统数据管理标准之一——数据集的元数据标准 (试运行)》 (赵改善等, 2002); 中国 21 世纪议程管理中心, 1999) 5 种, 分别制定了一些元数据标准, 但并不是正式颁布的国家标准。目前还没有国家元数据标准正式颁布。

## 四、海洋经济动物种质资源元数据库建立的策略及其应用

在海洋资源信息方面, 实际上已经开始应用元数据技术了。如国家海洋局在网上已经公布的共享元数据信息, 林绍花等 (2003) 对海洋学科数据库群建设具体内容涉及标准与规范的研究、制定, 海洋资料资源大调查, 海洋资料元数据库系统建设, 海洋资料数据库群建设, 中国近海海洋基本场产品的研制, 海洋资料的共享服务网络系统建设, 海洋潮汐历史资料的水尺零点考证研究以及海洋环境资料质量控制自动化系统等进行了报道。中国海洋资源数据库集由国家海洋信息中心研制开发, 含数据库 16 个: 中国海岛资源数据库、中国沿海渔业资源数据库、沿海盐场资源数据库、沿海旅游资源数据库、海洋石油天然气资源数据库、沿海主要港口资源数据库、沿海港址资源及环境条件数据库、国内海洋综合经济数据库、海洋水产业数据库、海洋盐业数据库、海洋运输业数据库、海洋石油与天然气业数据库、滨海旅游业数据库、海洋环境资料数据库、海洋潮汐预报数据库、海洋法规数据库。其中海洋水产业数据库包括了历年海洋捕捞、海水养殖的生产情况和海洋水产业从业人员等 195 条信息; 开发目的是为了了解海洋水产业的发展状况及趋势, 调整其在海洋经济方面所占的比重, 为合理开发利用海洋水产资源提供信息服务。但这些都集中在海洋自然资源的数据库建设, 而有关海水经济动物种质资源的元数据库尚无人涉及。

建立海洋经济动物种质资源元数据库是非常必要的, 这是海洋经济动物养殖健康发展和国家可持续发展战略决策的需要、水产研究和水产科技发展的需要、抢救水产科学数据资源的需要。其建立原则和策略包括: ①与国际标准接轨; ②与中国可持续发展信

息共享元数据标准尽量保持一致;③吸取我国现有的一些农业、地理和生物多样性元数据库的合理部分;④简洁明了,重点突出,灵活性强,方便用户补充、更新、查询、使用,并针对海洋经济动物种质资源的特殊性进行调整(易浩若,2003)。

## 1. 种质资源数据库的建立

种质资源数据库包括以下内容:

(1)制定种质资源科学数据的采集、加工、建库和共享的标准规范应根据建立数据库和提供共享服务的需要,在国家标准和行业标准的指导下制定水产业科学数据采集、加工、建库和共享的标准规范,包括数据内容、分类与代码、组织结构、数据格式、采集方法、加工规程、处理平台、数据库结构、元数据、分发服务等。

(2)对数据采集进行框架设计。一个优化的数据采集框架设计,将保证用有限数量的数据来真实地反映客观事物的全貌。

(3)制订元数据标准,包括元数据内容和组织结构等;开发元数据的生成软件。

(4)建成各类海水经济动物科学数据的元数据库并与各类数据库同步更新。

(5)构建数据加工与集成平台。以数据共享为目标,构建一致的数据加工平台和数据集成平台。该数据加工与集成平台具有海量数据的处理和存储能力,具有较强的数据导入能力和较广泛的格式兼容性。

(6)数据的采集和加工。采集、加工、建库和共享的海水经济动物科学数据包括水产科学实验数据(海水养殖鱼、贝、虾类培育数据;水产品加工数据;水产病害数据;遗传育种数据),水产种质资源数据(中国和引进海水养殖鱼、贝、虾类种质数据;种质保存和培育数据;原种场数据),海水养殖生态环境本底监测数据(海水经济动物养殖海域近岸和离岸海区普查与监测数据;海边和海上大型工程如码头、石油平台等数据),经济动物产卵场的保护数据,海水水产信息技术基础数据(海水经济动物基础地理数据;种质资源遥感监测评估数据;水产业专题数据;水产业应用软件数据),水产业文献资料及科技信息(水产业法律法规、种质标准与文献资源数据;科技信息)。

## 2. 元数据技术的未来发展

(1)元数据技术的应用前景。采用元数据技术,将以更规范化、标准化的方式全面地管理企业的数据资产,保证数据质量,提高查找数据和应用数据的效率,充分发挥数据资产的增值作用,降低新品种和品系开发利用的风险和成本;采用元数据技术,将使应用互联网进行的数据资产电子商务活动效率更高,效果更好,方便客户对不同数据源的了解和对比,开发潜在的数据市场,降低交易成本和交易时间;采用元数据技术,还可以针对水产养殖企业分布范围广、各种数据和信息量大的特点,充分利用计算机网络建立分布式的企业信息管理系统,为企业管理和决策提供全面、丰富、可靠的信息支持,提高决策的效率和正确性。数据集成和应用集成,采用元数据技术,可以在分布式环境下实现同一项目内部不同数据源、不同项目数据源之间,以及企业数据银行与项目数据库之间实现有机的数据集成和应用集成,提高项目管理水平和工作效率,方便多学科学者之间的协同工作;在数据管理的基础上,采用元数据技术可以更有利于实现企业

级的数据挖掘和知识发现, 从而作为企业知识管理系统中有效的知识获取途径, 并在此基础上实现企业的技术创新和管理创新, 提高企业的决策水平和竞争实力。

(2) 元数据已经成为数据管理、数据共享和数据使用的一种重要工具。随着元数据技术的完善, 元数据的研究和应用对于海水种质资源的管理、持久安全保存和开发有极其重要的价值, 为发掘有用的新种质提供了科学依据, 为政府部门提供决策信息。同时对于完善种质资源的保护措施, 建立核心种质, 研究品种的生态类型, 制定各种的生产区划, 加强与国外的物种种质资源交流也有重要参考价值, 从而提高海水养殖业的经济效益, 这也将是中国可持续发展网上的一个重要组成部分。

## 参 考 文 献

- 符海芳, 牛振国, 崔伟宏. 2002. 中国农业信息元数据内容体系研究. 地理学与国土研究, **18**(4): 6~9
- 纪力强. 1999. 中国生物多样性信息系统数据字典和元数据标准. 中国 21 世纪议程管理中心编. 中国地理信息元数据标准研究. 北京: 科学出版社, 101~114
- 蒋景瞳, 刘若梅, 贾云鹏. 1999. 国际元数据标准的发展研究现状. 中国 21 世纪议程管理中心. 中国地理信息元数据标准研究. 北京: 科学出版社, 1~1
- 林绍花, 陈奎英. 2003. 地球科学数据库系统 (WDC-D)——海洋学科数据库群建设项目. 海洋信息技术, **1**: 1~3
- 徐海根, 包浩生. 2002. 中国生物多样性核心元数据标准的探讨. 中国环境科学, **20**(2): 106~110
- 徐海根, 薛达元, 吴小敏. 1999. 生物多样性信息元数据库的开发. 农村生态环境, **15**(2): 16~21
- 徐海根. 1999. 中国生物多样性元数据库的研究与开发. 兰州大学学报 (自然科学版), **35**(4): 103~108
- 易浩若. 2003. 加强基础数据采集促进林业科学数据共享. 林业科技管理, **3**: 20~22
- 张 侠, 程少华, 朱建钢. 2002. 中国极地科学数据库系统的研究与建立 1. 数据资源的组织以及元数据与数据集的关系研究. 极地研究, **14**(1): 62~72
- 赵改善, 曹邦功. 2002. 元数据: 勘探开发数据管理的一种新工具. 石油物探, **41**(2): 236~242
- 中国 21 世纪议程管理中心. 1999. 中国地理信息元数据标准研究. 北京: 科学出版社
- EEA 1993. Analysis and Suggestions for a metadata system and service for the EEA. European Environment Agency
- NMD. 1981. Handbook on Characterization of Environmental Data. Nordix Council of Ministers Environmental Data Group
- Pinto J. U. 1992. Catalogue of Data Sources (DS) for the Environment
- UNEP. 1992. The Field Metadata (MDB) Entity Attribute Definitions. UNEP-GRID database. UNEP-GRID Geneva
- WCMC. 1996. Information Management Guidelines. The World Conservation Monitoring Center. London

# The Concept, Content and System of Genetic Resources Metadatabase on China Marine Economically Valuable (Cultivated) Animals

LI Jun YOU Feng

(Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071)

## ABSTRACT

Study on the protection and conservation of marine biological genetic resources has gained much widerly with the development of marine aquaculture in China. The more and more data

have been collected but how to share these data demands attention to. Sharing of nature resources and biodiversity information useful for aquaculture and related research is the key to their sustainable utilization. Metadata refers to information on the location, source, content, or other specifics in relation to actual data. Metadatabase is a database designed and developed for metadata conservation and management, which provides services for the management, query and use of actual data. This system has powerful information query functions, such as query by institute catalogue, query by thematic information category, query by name, and query by author. It is rich in information. There are some metadatabases such as China Information Network for Sustainable Development and other links on the Internet. But there is no mechanism for marine economically valuable (cultivated) animal genetic resources information exchanging and sharing in China at present. It is therefore necessary and urgent to set up an information query channel with metadata to help people know where to find required information on marine cultivated species and how to use them. This paper clarifies the definition, role and meaning of metadata, and outlined the concept, key elements and standards of metadatabase. The general status of biodiversity metadatabase in our country is also reviewed. We discussed the contents and strategies to set up the China Marine Economically Valuable (Cultivated) animal genetic resources metadatabase in the present paper, including standing criterion institute on collecting, processing, metadatabase constituting and sharing of information, sharing of scientific data on genetic resources, and data collection framework designing. The conceptual basis on how to organize and construct this kind of database is presented.