

海洋地质档案资料元数据著录

戴勤奋,王圣洁,张兆代,王 红,孙佃奇

(中国地质调查局青岛海洋地质研究所,青岛 266071)

摘 要:综合参考国内外档案著录与档案元数据相关标准,以当前海洋地质档案资料内容特征及整理体系为基础,以未来大范围信息资源服务与共享为导向,提出了适用于原始、实物、成果三大类海洋地质档案资料的多实体、多层次著录元数据模型,并给出了元数据内容、格式和样式的落实方案,期望推动传统档案著录向元数据置标著录迁移并转型。

关键词:海洋地质;档案资料;元数据;档案著录

中图分类号: P712

文献标识码: A

DOI:10.16028/j.1009-2722.2016.11009

1 引言

传统档案著录定义为“在编制档案目录时,按照一定规则,对档案内容和形式特征进行分析、选择和记录的过程”^[1,2],著录遵循标准《档案著录规则》(DA/T 18-1999)^[3],该标准适用于传统载体档案(如纸质档案)手工整理过程中的后控著录及编目。随着计算机与互联网的大众化及数字信息资源的主流化,传统著录规则的局限性在应用过程中逐渐显现^[4],于是元数据作为电子档案的著录与管理工具进入档案领域^[5,6],以适应档案资料数字化、信息管理集成化、信息服务网络化建设的需要。

元数据称为“关于数据的数据”,是用于资源(能满足需求的可标识对象)特征描述的一种信息载体。国际标准化组织的标准族 ISO 23081 将元数据定义为:描述记录资料(records)背景、内容、结构及其管理过程的数据^[7-9]。其中“records”的释义在国内档案界颇有争议^[10],参考美国档案工

作者协会(SAA)“档案与记录术语汇编”中的相关词条^[11],本文将“records”定义为:某一项目、任务或活动在规程控制下运行所形成的具有再利用价值的过程文件集合。“海洋地质档案资料”囊括海洋地质调查及研究过程中形成的原始记录及其派生记录,符合“records”的概念范围界定,因此“海洋地质档案资料元数据”等同采用 ISO 23081 的元数据定义。

依据 ISO 23081 的元数据定义,档案资料元数据包含两大部分内容信息,首先是描述归档资料背景、内容、结构信息的捕获节点元数据;然后是描述档案管理过程的捕获后元数据。与传统著录相比,元数据著录信息由档案资料内容向其形成背景扩展,背景信息将有助于追根溯源,确定档案资料的可靠性与可用性;其次,元数据著录将传统的编目阶段著录拉伸至档案资料生命期全程动态著录,由此保障档案资料不受时空限制的可理解性。

在当前的地质档案管理体系中,海洋地质资料按原始地质资料、实物地质资料和成果地质资料分类整理,海洋地质资料的档案著录工作按照行业标准《地质资料档案著录细则》(DA/T 23-2000)实施^[12]。《地质资料档案著录细则》在《档案著录规则》基础上定制,描述实体为档案本身,著录层级为案卷级和文件级,着重机构内部的档

收稿日期:2016-01-22

基金项目:中国地质调查局项目“海洋地质资料服务产品开发示范”(中地调研[2013]131)

作者简介:戴勤奋(1963—),女,研究员,主要从事海洋地质数据库设计工作。E-mail:dqinfen@cgs.cn

案管理与服务需求。如果将档案管理拓展至整个行业,如地调局系统的档案委托保管;服务范围扩展至互联网,如公益性的地质资料公共服务,那么档案资料的全过程多实体著录、以及由总到分的多级著录就十分必要。分层分块组织方式既有利于信息总览、回溯,实现档案资料不受人员变动及地点限制的可理解性,也有利于信息分拆隔离,实现不同目的、不同网络环境的元数据差异化著录和表达。为此,本文提出了适用于原始、实物、成果三大类海洋地质档案资料的多实体、多层级著录元数据模型,并给出了元数据内容、格式与样式的落实方案。

2 信息资源分析

随着档案资料的电子化与数字化,档案工作的重点正从档案资料整理与保管转向信息资源的

检索与利用,元数据作为在计算机中构建、完整反映现实资源特征的虚拟资源,在资源运转过程中起着媒介与桥梁作用。

海洋地质档案资料来源于海洋地质调查与研究项目的实施过程,包括:项目外业实测调查、样品实验分析、数据处理解译等工作过程直接形成的归档资料;项目阶段性成果资料和结题成果资料;项目采集的实物样品资料;项目购置仪器设备携带的技术资料;项目购置或编制的软件资料;以及工作过程中收集、交换、购买的技术资料。表 1 列出了面向档案管理的海洋地质档案资料类别(依据 DA/T 41-2008 第 4.3 条分类)^[13];同时,表 1 及图 1 也给出了各类档案资料相对应的海洋地质信息资源类别(按表现形式分类,与资源载体及格式无关)及其分级,所列资源便是元数据的描述对象,元数据的介入将减少现实环境对原态档案资料的操作,实现资源运转效率的最大化。

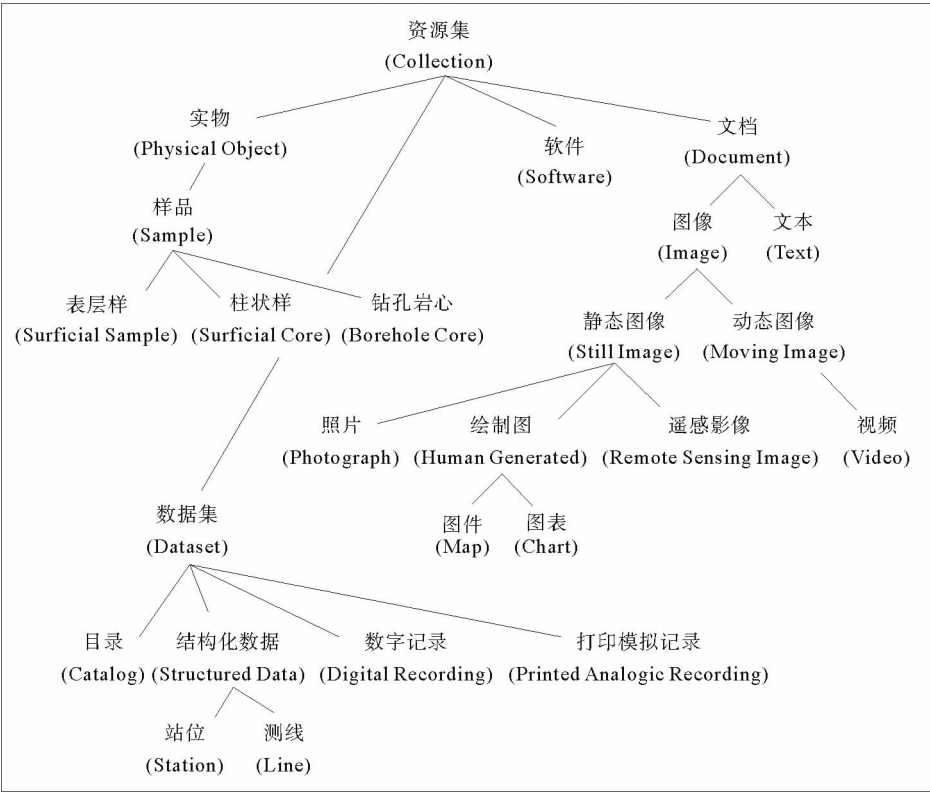
表 1 海洋地质档案资料及其信息资源类别

Table 1 Archival records and information resource types for marine geology

档案资料类别	信息资源类别	说 明
成果报告及图件	目录、文本、图件	项目年度报告、结题报告正文及其附图、附表;成果图件或系列成果图件及其说明书;成果评审文据
实物地质资料	目录、样品、文本、照片	底质取样采集的表层样、柱状样;钻孔取样采集的岩心、岩屑等。实物样品的数字化形式是照片及描述性文字
测绘资料	文本、图件、数字记录、结构化数据、站位、测线	导航定位的仪器原始记录、班报、质量自检记录及工作报告;定位获取的观测点位、取样站位、钻孔孔位、走航调查测线位置数据及位置图。站位及测线要素可作为地理信息检索入口
野外地质观察	文本、照片、视频	海岸带地质调查(如岸滩地质踏勘、岸滩剖面测量)获取的实测滩面资料、实地影像资料及调查报告
勘探工程及现场试验	文本、图件、结构化数据、数字记录、打印模拟记录	油气勘探钻探过程中的取心、录井、测井现场记录,现场试验记录,施工报告、录井报告、测井解释报告,完井地质报告等
采样测试鉴定	文本、照片、结构化数据	外业水体(悬浮体、海水)采样记录,底质(表层样、柱状样)采样记录,浅钻孔取心回次记录及班报,样品现场描述及测试记录,航次报告,工作报告,外业验收文据。室内岩心剖分、照相、岩性描述记录,样品送样记录,水样、沉积样、岩样测试分析及鉴定报告(如粒度分析、矿物鉴定、微体古生物鉴定、化学分析、测年、古地磁测量、物理力学性质测试等)
试油、试采、采油	文本、图件、结构化数据	油气勘探与开发过程的试油、试采、采油原始记录、班报、日志、地质日报、地质月报,试油、试采、采油总结报告等
仪器记录及动态资料	文本、数字记录、打印模拟记录	定点或走航测量(如单波束测深、多波束测深、同步验潮,侧扫声呐、浅地层剖面、重力、磁力、地震、海底热流、海流、海水浊度及温盐深等测量)仪器原始记录、打印模拟记录、班报、质量自检记录、数据现场处理报告、航次报告、工作报告,外业验收文据
遥感影像	遥感影像	从专业机构有偿或无偿获取的正量遥感或航空遥感影像资料

续表 1

档案资料类别	信息资源类别	说 明
中间性综合资料	文本、图件、结构化数据、数字记录、打印模拟记录	数据后处理结果(如地震叠加、偏移成果剖面,以及水深、重磁异常、速度谱、热流、海流、温盐深等数据);资料解释推断结果(如浅地层剖面、地震剖面的地震反射层、断层、地层层序解释结果);数据后处理报告,资料解释成果报告及附图
仪器设备资料	文本、软件	仪器设备携带的技术资料及专用软件
软件资料	文本、软件	项目购置或编制的软件及说明文档
管理性文件	文本	与项目管理相关的各种文件,如立项申请、项目预算、任务书、合同书、协议书、委托书、设计书,设计更改和批准文件等
外部收集资料	文档、数据集	收集、交换或购买的数据资料(如浪潮流长期观测资料、地磁日变资料、密度、磁化率、热导率等物性资料)、相关单位成果资料、以及公开出版的地质文献资料、图集、标准规范等



3 元数据模型

3.1 元数据实体

元数据是档案资料管理及资源检索利用过程中实体对象的代表,因此实体决定了元数据所代

表的角类型。参考 ISO 23081 的多实体模型,图 2 给出了海洋地质档案资料的元数据实体信息集合模型。

海洋地质档案资料元数据实体由“档案”、“业务”及“责任者”三实体聚合(一种松散的整体与部分关系),其中档案实体描述档案资料的标识、内容结构、来源背景及分发利用信息;业务实体描述

档案的管理活动信息;责任者实体描述相关责任单位、责任人及其联系信息。三实体中的“档案”为必备实体,如果“档案”或“业务”实体引用“责任者”实体,则“责任者”实体也必备,此时责任者实体的变动将会对其使用方产生影响。

描述“档案”实体的元数据信息包括:①标识信息:提供档案资料的唯一识别信息;②内容信

息;描述档案资料内容范围、覆盖区域及档案整理与载体信息;③来源信息:提供档案资料来源及其依托项目信息;④分发利用信息:提供资源开放控制、获取方式与数据格式信息;⑤相关资料信息:提供相关档案资料与出版物信息;⑥附注信息:提供需要解释与补充的信息或具特殊意义的备注信息。

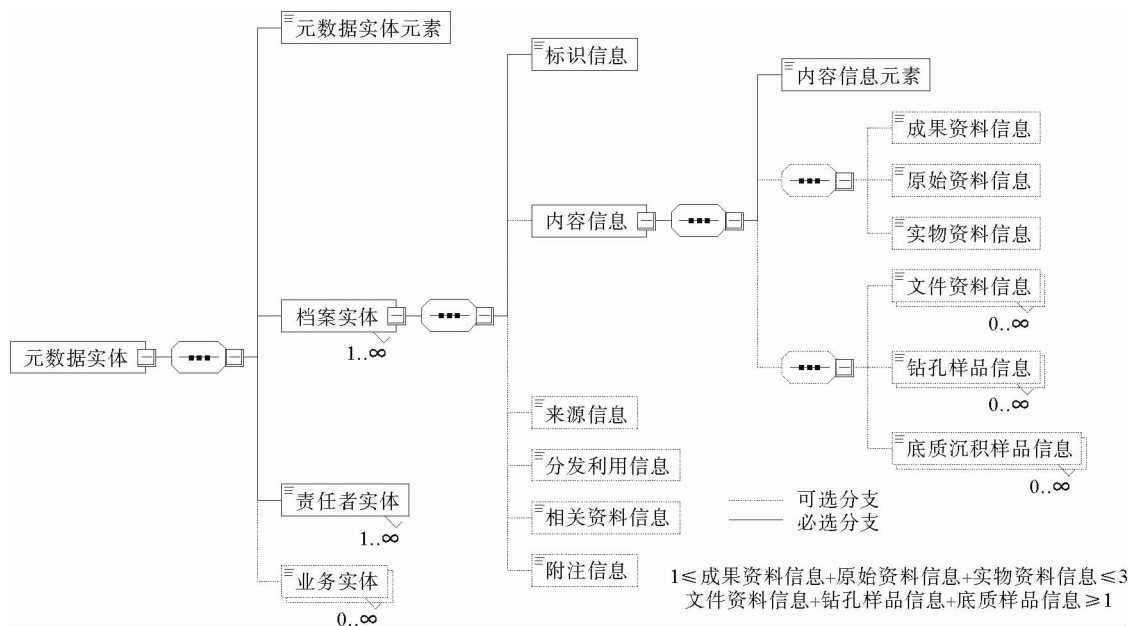


图2 海洋地质档案资料元数据实体信息集合

Fig. 2 Metadata entities and information sets for marine geological archival records

3.2 著录层级

遵循档案的来源原则和尊重全宗原则,参考《国际标准—档案著录通则》(ISDA(G),国际档案理事会(ICA)颁布)的多级著录模型^[14,15],同时依据海洋地质档案资料的分类、分级整理方式,将档案实体的元数据著录层级由高至低分为:全宗级、项目级、类别级、案卷级和文件级。实际中可按传统方式只著录案卷与文件,或以项目为单位整体著录,或原始资料、成果资料、实物资料分类著录,或选择相应层级后由总到分按序著录,高层级中已著录的内容不必在低层级中重复描述。

4 元数据内容、格式与样式

4.1 元数据内容定制

元数据著录并非传统著录的另一种形式,引

入元数据的目的是期望实现不受地点限制的信息资源检索、管理与共享,该目标要求元数据必须能支持互操作性,即异构平台间的元数据注册、交换与共享。解决互操作问题的途径之一是元数据内容及其格式的标准化,即行业共同遵循统一的内容描述标准、使用统一的数据组织框架,达到信息的互联互通。

元数据内容定制在现行的档案著录内容基础上扩充,因必须满足现行档案著录标准及规范的规定与要求;考虑到未来大范围共享的可能性,也兼顾了与国际主流档案著录标准的兼容性;因海洋地质资料与地理位置密切相关,也参考了国内、国际的地理信息元数据标准。最终,依据定制内容著录的元数据能支撑基于主题、地理位置及时间信息的网络资源检索,并能支持基于目的需求的资源可用性判断,也能提供资源的访问与获取途径信息,详见图3~5。

元数据内容定制技术策略包括:①以元数据在实际中可能的角色类型为切入点,分实体、分层、分块组织档案资料元数据的内容信息;②在现行档案著录内容基础上扩充,传统著录元素全部

保留,只是予以重新组织;③尽可能简单化,避免复杂的层次嵌套;④以数据类型为容器封装可复用描述结构;⑤最小化无输入控制的自由文本类元素,最大化代码表类型的元素,因便于输入控制

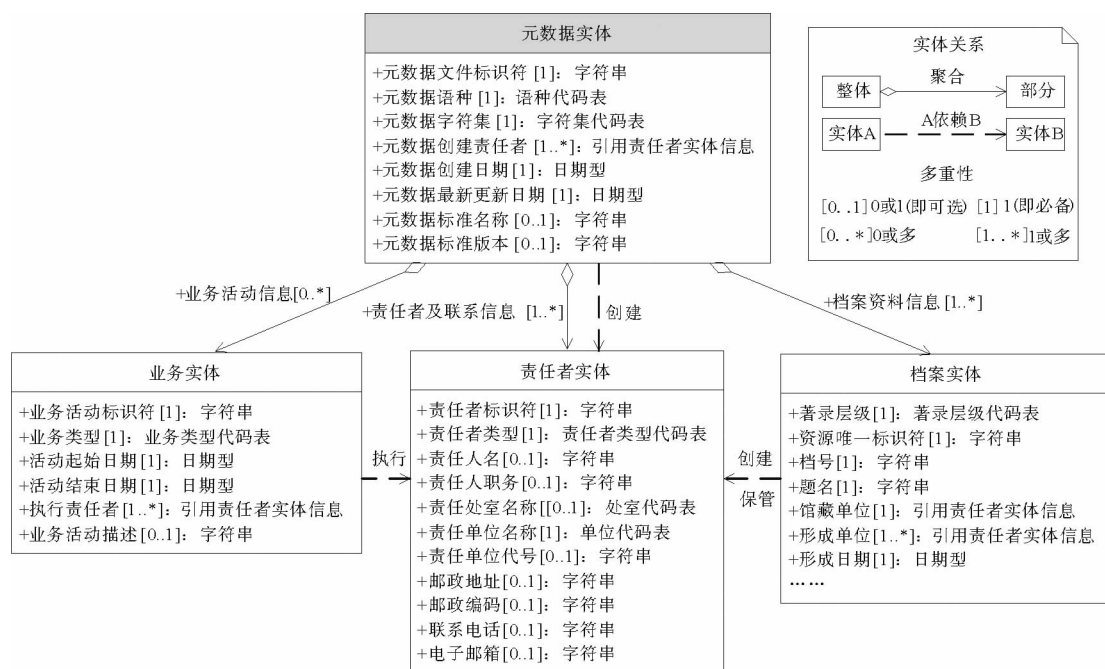


图 3 元数据实体信息

Fig. 3 Metadata entity information

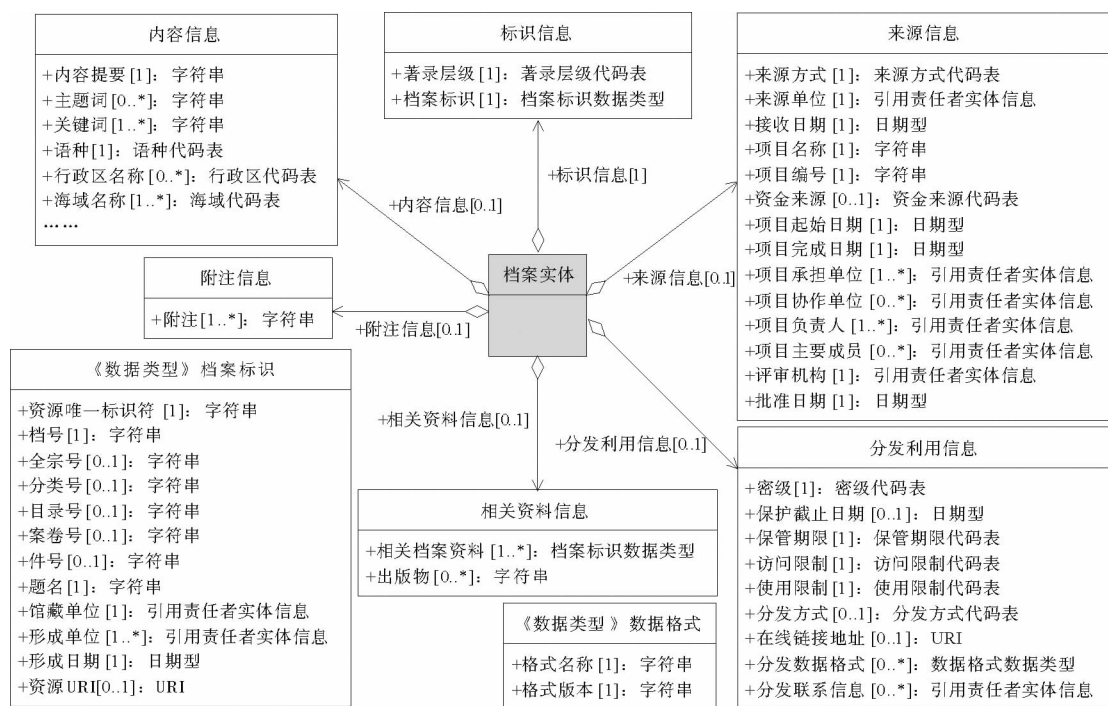


图 4 档案实体元数据信息

Fig. 4 Archival records entity metadata information



图 5 档案实体内容信息

Fig. 5 Content information for archival records entity

且可扩充。

4.2 元数据格式定义

元数据属于结构化或半结构化的信息集合, XML(Extensible Markup Language, 可扩展标记语言)格式文件是组织元数据内容信息的最佳选择。XML 将不同类别的信息保存在各置标节点, 有利于表达元数据信息之间的层次与嵌套关系; 而且节点信息抽取、节点后期扩展也灵活方便; 同时, XML 文件也便于网络传输、交互以及跨系统、跨平台共享。

元数据 XML 文件的结构通过 XML Schema (XML 模式) 定义, 定义创建的 XML Schema 文件(元数据 XML 模式)可在数据交换中心、数据互操作平台或数据库系统中注册登记, 经注册的 XML Schema 描述框架能自动校验所提交元数据的合法性及有效性, 保证元数据 XML 文件符合 XML 模式的格式要求。

元数据 XML 模式涉及元数据结构框架的搭建, 以及各置标节点标签名、数据类型、可选性、多重性及其值域的定义, 其中标签名的命名规则是需要讨论的问题。

XML 标签是机器识别和数据交换的中间体, 即计算机利用该标签检索及提取置标节点的信息, 标签名的变动将对已保存的 XML 元数据文件、以及外部应用工具产生影响; 因此, 要求 XML 标签遵循行业共同的词汇表、同时能便于机器快速识别且保持长期稳定。目前, 国内尚无统一的行业词汇表, 而且相关标准中的标签命名规则差异较大, 如《基于 XML 的电子文件封装规范》(DA/T 48—2009), 直接采用元数据元素的中文名称作为标签名^[16]; 中国地质调查局的《地质资料目录数据标准规范》采用中文名的拼音首字母组合作为标签名^[17]。直接采用与表达内容完全相同的元素名称作为标签名, 虽然方便, 但有不稳定隐患, 因为外部的变化将对内部的 Schema 产

生影响;此外,如果不同标准对于同义词给出不同的中文表达,例如“内容提要”、“提要”、“摘要”、“内容描述”,那么对计算机而言,它们属于不同的标签,元数据的互操作性将因此受到影响。国外元数据标准基本上均采用英文缩略码组合作为标签名,只是组合形式有差异,例如,国际标准化组织的《地理信息元数据》标准采用驼峰式组合标签(如:<archDesc>)^[18]、美国档案工作者协会的《档案置标著录》(Encoded Archival Description,简称EAD)标准则采用全小写组合标签(如:<archdesc>)。EAD的标签库由SAA档案置标著录技术分会维护更新,最新版本为2015年8月颁布的EAD3^[18]。考虑未来国际范围数据交换的可能性,海洋地质档案资料元数据适合采用EAD的标签命名方式,标签中文名称留待样式单中映射。

元数据XML模式定义策略:①尽最大可能保证XML Schema在一定程度上的稳定性,使其变动不致引起外部应用的大幅度调整;②保持代码表类型定义的可扩展性,且代码值尽可能采用无意义的字符串(如:01、02…)占位,具体值域留待样式单补充;③不确定因素移交样式单处理。

4.3 元数据表达样式定制

元数据内容及格式必须通过相应标准控制,但元数据的显示内容及表达方式却可按需求随意定制。XML具有内容与样式分离的特点,元数据显示内容可按需从相关置标节点提取,提取内容的显示方式可通过XML样式单(XSLT stylesheet)定制,最终利用样式单(xslt文件)将存储信息的XML源树转换为显示信息的结果树(如HTML5)。不同的样式单可以使相同的数据呈现不同的外观,同一XML文件采用多种样式表达,将有利于满足不同类型用户群的实际需求。

元数据表达样式定制策略:①按实际需求定制样式单,或按规范中规定的报表格式定制样式单;②鼓励职能部门定制专用的样式单;③创建样式库以便统一管理样式单。

5 结语

以当前海洋地质档案资料内容特征及整理体

系为基础,以未来大范围信息资源服务与共享为导向,本文给出了海洋地质档案资料的元数据模型设计方案及实现策略,期望推动传统著录向元数据置标著录迁移并转型。未来,创建实用的著录工具或在线网络应用工具,将是推动元数据著录真正落实的关键。著录工具拟隐藏元数据内容与结构的设计细节,保持统一的页面或界面风格及操作方式,使著录页面或界面视图既易于理解也易于操作使用;其次,应能自动生成XML格式文件,并实现元数据文件合法性及有效性的自动校验;最终按用户所需样式显示元数据,通过元数据,使档案资料成为满足用户需求的信息资源。

参考文献:

- [1] DA/T 1-2000 档案工作基本术语[S].
- [2] GB/T 3792.1-2009 文献著录 第1部分:总则[S].
- [3] DA/T 18-1999 档案著录规则[S].
- [4] 赵芳. 对我国《档案著录规则》再修订的思考[J]. 档案管理, 2009(5):37-38.
- [5] 毛海帆. 电子档案元数据方案设计与应用初探[J]. 档案学研究, 2010(1):74-78.
- [6] 钱毅. 论电子文件中心元数据方案的管理策略[J]. 档案学通讯, 2012(6):76-79.
- [7] ISO 23081-1:2006 Information and documentation—Records management processes—Metadata for records—Part 1:Principles [S].
- [8] GB/T 26163.1-2010/ISO 23081-1:2006 信息与文献 文件管理过程 文件元数据 第1部分:原则[S].
- [9] ISO 23081-2:2009 Information and documentation—Managing metadata for records—Part 2: Conceptual and implementation issues [S].
- [10] 王英玮,熊朗宇. 论文件、记录和档案的术语含义及其生命周期[J]. 档案学通讯, 2015(6):4-7.
- [11] Richard Rearce-Moses. A Glossary of Archival and Records Terminology [M/OL]. Chicago: The Society of American Archivists, <http://www.archivists.org>, 2005.
- [12] DA/T 23-2000 地质资料档案著录细则[S].
- [13] DA/T 41-2008 原始地质资料立卷归档规则[S].
- [14] International Council on Archives. ISAD(G): General International Standard Archival Description—Second Edition [S]. 2000.
- [15] 张正强. 论《国际标准—档案著录规则(总则)》的主要著录规则、特点及启示[J]. 档案学研究, 2013(3):75-81.
- [16] DA/T 48-2009 基于XML的电子文件封装规范[S].
- [17] 中国地质调查局发展研究中心. 全国地质资料目录服务中心系统——地质资料目录数据标准规范[S]. <http://www.ngac.cn/mlzx/>, 2012-12-01.

- [18] ISO 19115-1:2014 Geographic Information-Metadata-Part 1: Fundamentals [S].
- [19] Technical Subcommittee for Encoded Archival Description of

the Society of American Archivists. Encoded Archival Description Tag Library, Version EAD3[S]. <http://www.archivists.org/>, 2015-08-28.

METADATA DESCRIPTION FOR MARINE GEOLOGICAL ARCHIVAL RECORDS

DAI Qinfen, WANG Shengjie, ZHANG Zhaodai, WANG Hong, SUN Dianqi

(Qingdao Institute of Marine Geology, CGS, Qingdao 266071, China)

Abstract: In order to move and transform the archival description to encoded metadata description, this paper puts forward a metadata model for the multi-entity and multi-level archival description of marine geological raw records, physical samples and final reports. Schemas for metadata content, format and style sheet implementation are also presented. Domestic and international standards for both archival description and archival metadata are comprehensively referenced, based upon the present content characteristics as well as arrangement systems of marine geological records. Marine geological information resources are expected to be served widely and shared well in the future.

Key words: marine geology; archival records; metadata; archival description