

气象元数据标准与信息发布技术研究*

王国复¹⁾ 徐 枫²⁾ 吴增祥¹⁾

¹⁾(国家气象中心,北京 100081) ²⁾(国家信息中心,北京 100053)

摘 要

元数据在数据交换、网络信息资源发现和共享系统建设中具有很重要的作用。文章简论了元数据和元数据标准的最新研究和应用概况,并重点介绍了WMO核心元数据标准,以及我国气象数据集元数据格式标准制定的原则、依据和内容。还论述了元数据系统的构建和发布流程;最后结合气象科学数据共享项目建设规划和分布式共享平台的建设实践,详细阐述了我国气象科学数据共享服务网元数据发布系统的设计方案和实现方式。

关键词:元数据 发布 共享 网关 元数据服务器

引 言

随着信息技术和信息技术应用的不断发展,信息资源的规模在不断扩大,并呈现出多样性,不同机构、不同应用系统之间交换、共享数据的需求日益突出。如何发现、获取分布于网络中的各类信息资源已成为公认的需要首先解决的一个问题。“元数据”则为这一问题的解决提供了一条有效的途径。

迄今,对元数据的认识是各种各样的,没有一个统一的定义。一般认为元数据是“关于数据的数据”或“关于信息的信息”^[1]。元数据的英文词是 Metadata。“Meta”来源于希腊文,有“alongside, with, after, next 等”的意思^[2],所以也有规范、有序的意思。元数据并不是随着信息技术的发展而产生的,传统意义上的图书卡片、磁盘标签、歌曲目录都可称为元数据。而在网络时代,元数据的提出最早是从地理信息系统开始的。对于元数据定义的研究,可能以下两个论述更具有指导意义:(1)元数据是关于资源的结构化信息,或称“把无序信息变成有序的、关于数据的结构化数据”;(2)元数据支持信息系统运行,即“是关于网络资源的机器能够识别的信息”^[3]。

通过元数据,一方面能够对信息资源进行详细、深入的了解,包括信息资源的格式、质量、处理方法和获取方法等方面细节;另一方面借助它能实现网络共享,使得信息资源的用户(或用户应用系统)可以迅速地发现与其需求匹配的信息资源,进而通过网络或其他途径取得它并加以利用,从而促进信息资源的共享。

元数据是对信息资源的规范化描述,它是按照一定标准(即元数据标准),从信息资源中抽取出相应的特征,组成的一个特征元素集合。这种规范化描述必须准确和完备地说明信息资源的各项特征。元数据标准的制定是元数据应用的前提。

* 本文由科技部国家科技基础性工作专项资金项目(2002DEA30041)资助。

2003-04-07 收到,2004-04-22 收到再改稿。

对元数据标准的研究起始于 20 世纪 90 年代中期,而且主要集中在数据域。目前,正在应用的元数据标准有:美国联邦地理数据委员会(FGDC)元数据标准^[3],美国航空航天局(NASA)的目录交换格式(DIF)标准,美国国际地球科学信息网络中心(CIESIN)元数据标准、都柏林(Dublin)核心元数据标准、国际标准化组织 ISO/TC211 元数据标准草案^[4]等等。这些标准可简单划归为两类:一类是针对互联网上的信息系统的,典型的是 Dublin 核心元数据标准;另一类是针对行业的,如 FGDC 元数据标准等。

1 气象元数据标准

1.1 元数据标准制定的一般原则

元数据包括三层结构:元数据子集、元数据实体和元数据元素^[1]。元数据元素是元数据最基本的信息单元,元数据实体是同类元数据元素的集合,元数据子集是相互关联的元数据实体和元素的集合。在同一个子集中,实体可以有简单实体和复合实体两种。简单实体只包含元素,复合实体既包含简单实体又包含元素,同时复合实体与简单实体及构成这两种实体的元素之间具有继承关系。图 1 是以 WDC-D 气象元数据来说明的元数据三层结构。

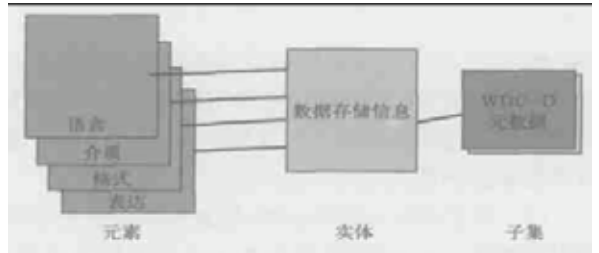


图 1 元数据层次结构

元数据标准的制定须遵从一定的原则。标准要支持元数据在行业或其他领域的应用,以提供数据的基本状况为目的。标准要提供一个实体与元素集,并定义元素的性质:必选、一定条件下必选以及可选等。须特别指出的是元数据标准定义的对象是数据,而非定义与数据相关的计算机系统、传输手段和信息表现方式等。一个元数据标准除要详细列出其实体与元素集以外,还要说明元数据标准适用范围、参考标准、术语与定义、元数据分级、元数据性质、标识、格式和类型,以及元数据扩展原则等。

1.2 世界气象组织(WMO)核心元数据标准草案

因 WMO 各部门对其数据集归档有不同的要求,所以制定适用于各部门和各 WMO 项目的综合元数据标准是非常困难的。为此,WMO 分别于 2001 年 11 月、2002 年 5 月和 2003 年 12 月经过三次专家组会议讨论通过了一个核心元数据标准草案^[5]。WMO 各执行计划或项目可依据该标准指导分析本部门和项目的元数据要求,制定一个部门或项目内部使用的标准^[5],并可以根据其具体情况扩展其元素、关键词和代码表等。WMO 核心元数据标准草案共 27 项(见表 1)。

1.3 我国气象数据集元数据格式标准的制定原则

我国气象数据集元数据格式标准的制定主要依据以下参考原则^[1]:

(1) 参考气象科学数据的特点:由于气象观(探)测记录种类、气象要素的多样性,以及观测记录的连续性、时间性、空间性、地域性等特点,在描述数据集实体和属性时,应明确表述气象科学数据特点的相关内容和项目,如数据类型、气象要素名称、记录时间、观

表 1 WMO 核心元数据标准草案

必选项	Title(数据集名称)	Temporal extent(时间跨度)
	Identifier(数据集标识)	Metadata language(元数据语言)
	Reference Date(数据集制作日期)	Metadata point of contact(元数据责任人)
	Responsible party(数据集责任人)	Metadata character set(元数据字符集)
	Abstract(摘要)	Metadata date stamp(元数据制作日期)
	Key words(关键词)	Access rights or restrictions(限制密级)
	Geographic Extent(地理范围)	Distribution mechanisms(分发系统)
	Topic Category(数据集类别)	Processing level(处理程度/质量状况)
一定条件下必选项(C)	Dataset language(数据集语言)	Character set(数据集语言字符集)
可选项	Spatial resolution(空间分辨率)	Metadata file identifier(元数据标识)
	Spatial representation type(空间显示类型)	Metadata standard name(元数据标准)
	Reference system(参考系)	Metadata standard version(版本)
	Vertical extent(垂直跨度)	Validity time(元数据制作时间)
	On-line resource(在线资源)	

(探)测次数,以及记录数据的空间和地理范围等信息。

(2) 参考气象科学数据共享工作的需求:气象元数据标准的制定要特别考虑气象科学数据分布式共享系统平台建设的需要,如分布式数据管理、数据搜索、导航、用户认证、数据检索服务等。

(3) 参考相关元数据标准:气象科学数据是地球科学数据的重要组成部分之一,其元数据的内容和格式,应以国内外相关的元数据标准为指导和参考依据,以便与国内外标准接轨。当前可参照的标准有:WMO 核心元数据标准草案、ISO/TC211 元数据标准草案、WMO 气候数据集目录款目格式、FGDC 元数据标准、中国可持续发展信息共享元数据标准草案等。

1.4 我国气象数据集元数据格式标准(草案)的内容

基于上述原则,制定了气象数据集元数据格式标准(草案),它适用于气象科学数据共享系统建设和数据集的编目、归档、发布和交换等。描述对象是“气象科学数据集”。本标准草案规定了气象数据集元数据的核心元素或编目信息,即一级元数据的最基本、最主要的实体和元素的性质、内容、标识、结构及有关细则,用于了解气象数据集总体。它可以被扩展到二级,即数据集的更详细信息,又可分为若干子集,分别说明数据集某一方面的信息。该标准草案由 71 个元数据元素组成,包括 19 个必选项(M),19 个一定条件下必选项(C)和 33 个可选项(O)。

2 元数据管理系统与元数据发布

作为对信息资源的描述,元数据只有完整地包含其描述对象的各种特征信息,并且其内容和组织方式需要遵循一定的规范,人们才能借助元数据正确地了解其所描述的对象进而促进信息资源或产品的共享或交换,或者通过元数据实现流程有效控制。元数据的完整性和规范性需要通过元数据的有效管理加以保证,这要求元数据管理系统必须能够适应元数据的应用目的和特点,在具备一般信息管理系统的共同功能之外,还应着重解

决以下几个方面的问题:

(1) 充分支持元数据内容的标准:元数据的目的是在不同人或系统之间共享有关信息资源、产品的说明信息,从而间接地实现对信息资源或产品的管理和交换。达到这一目标的前提是不同的人或系统能够一致地理解元数据中的各项信息,因而要求元数据在其使用范围内必须在内容组织和语意上遵循一定的规范。相应地,元数据管理系统必须对用户所采用的元数据内容标准给以有效地支持。根据需要元数据内容标准中有时需要定义描述元素之间的约束关系,例如描述元素之间的互斥、互为前提、甚至元素值之间相互限制。元数据管理系统必须能够正确地处理这些逻辑关系,严格按照标准规范对元数据的各种处理,以便正确地规约元数据的采集和维护工作。此外,不同领域的元数据内容标准必然有所不同,同一领域的标准也会随着应用需求的改变而有所变化。元数据管理系统必须具备足够的标准适应能力,以使用户能够及时根据需求的变化进行必要的调整。

(2) 高效的元数据网络检索:元数据的主要作用是借助计算机网络交换信息,使人们能够通过元数据及时、准确地了解他们所需的产品(数字化或非数字化)。因此,元数据管理系统必须提供元数据的网络查询检索功能。元数据的网络查询既不同于关系型数据检索,也不同于一般网络搜索引擎常用的全文检索。元数据是非结构化的,关系型数据的索引机制不能很好的适应元数据的不稳定结构。另一方面,元数据在信息组织上又存在数据域(描述元素)的划分,采用全文检索机制则不利于通过数据域的区分来减小查询命中范围。因此,元数据管理系统的需要采用与元数据特点相适应的新的检索机制,以求提高元数据查询的整体效率。

(3) 标准的网络搜索协议:不同部门之间在元数据共享方面的合作要求各部门的元数据管理系统之间必须能够互联,并实现元数据的网络交换。为实现这一目标,元数据管理系统的网络查询服务必须遵循一种通用的协议实现对元数据的网络搜索的提取。目前,在网络信息搜索和提取方面最重要的协议是 Z39.50 协议,该协议由 ISO 建立,用于规范网络信息搜索和提取过程中的各种请求与响应,并对服务器和客户机的处理进行规范。

2.1 元数据管理系统结构

元数据管理系统结构主要由元数据网关、元数据服务器和元数据库组成^[6](图 2)。其中,各分节点安装元数据节点服务器,用于提供该节点数据中心元数据信息的发布,并按照统一的元数据标准建设元数据库;主节点部署安装元数据服务系统网关软件,用于连接各分节点元数据服务器,实现元数据和数据的全网发布。

元数据网关是支持元数据服务的中心枢纽。其功能一般要包括:

(1) 服务器代理功能:可以有效的避免远程客户对元数据库直接存取,屏蔽了非法入侵,保证用户数据安全。

(2) 服务器注册管理功能:对于加入到元数据共享系统的服务器,需要对其服务器名称、地址等进行注册登记,使其连接到元数据共享系统中。

(3) 网络客户管理功能:提供用户注册、数据库访问权限管理等网络客户管理功能,便于网络客户权限的控制。

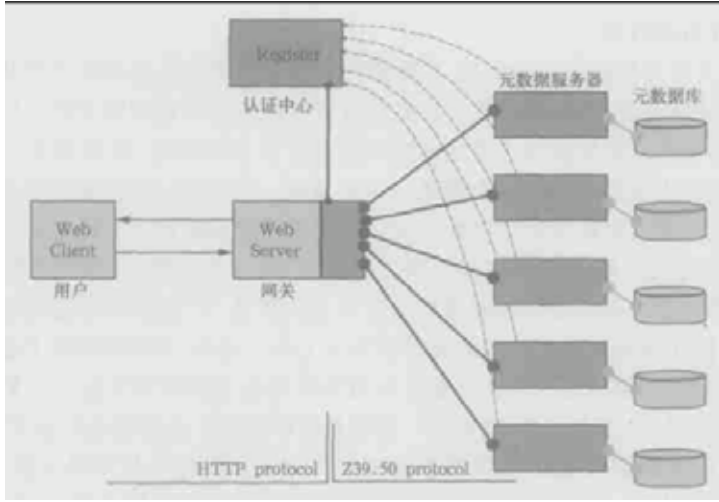


图 2 元数据系统结构

元数据服务器用于发布元数据。各元数据服务器一方面通过申请注册，把本节点元数据信息纳入到整个系统中，另一方面又接受 Web 服务器对本节点的元数据和数据搜索指令。这样用户通过该系统就可以透明访问任一节点上的元数据和数据信息。

元数据库是元数据发布系统的核心内容，元数据的采集可以利用元数据编辑器手工方式采集，也可以进行自动采集，但都要按照统一的元数据标准进行处理。

2.2 元数据发布

元数据发布过程分 4 个步骤(图 3)：

- (1) 用户向元数据网关发送元数据查询指令；
- (2) 元数据网关将用户的指令发布到各节点的元数据服务器；
- (3) 各节点元数据服务器搜索本地的元数据库，并将结果返回到元数据网关；
- (4) 元数据网关将查询到的元数据记录进行综合后，返回到用户手中。

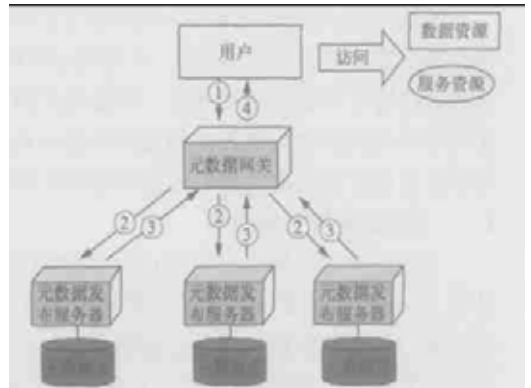


图 3 元数据发布过程

最后，用户对检索到的元数据进行评估，以决定是否访问该信息资源，经认证后获取数据服务。

3 中国气象科学数据共享服务网元数据统一发布

3.1 中国气象科学数据共享服务网分布式共享平台总体结构

中国气象科学数据共享服务网是由国家级和区域省级有机组成的，覆盖全国，分布式的网络化科学数据共享服务系统^[7](图 4)。它由 1 个主节点和若干个分节点组成，主节点部署于中国气象局，分节点部署于中国气象局气象信息中心、中国气象科学研究院、国

家卫星气象中心和省级各气象中心。国家级与区域省级分节点在不同层次上维护、管理数据资源以及相应的元数据信息：国家级分节点存储、管理国家基础气象数据、国际交换数据(WDC-D系统)、卫星气象数据和气象科研数据；区域省级分节点存储、管理省级及专题气象数据(例如暴雨、干旱、沙尘暴等数据)，各分节点维护其数据和相应的元数据。

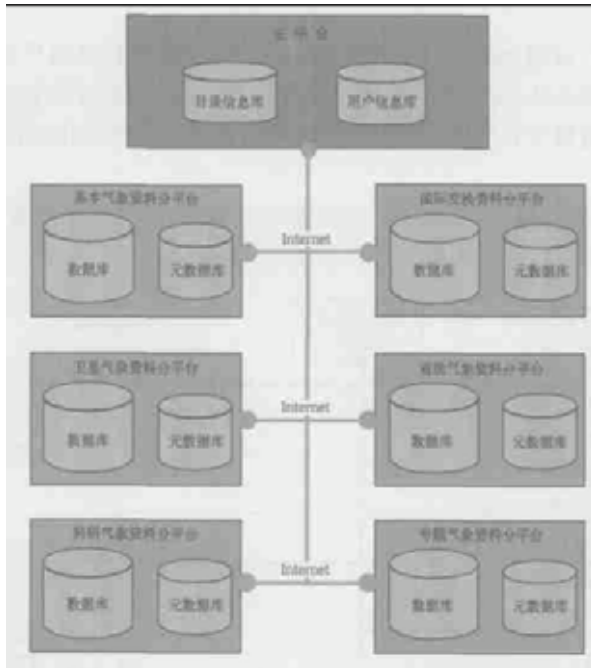


图4 中国气象科学数据共享服务网分布式平台总体结构

主节点及各分节点建立网站系统。用户通过访问网站系统以获取分布在不同数据中心的、基于 Web 的数据服务，如：元数据信息检索、共享数据集的浏览、访问和下载等。

系统具有系统认证服务功能，可以提供系统用户的统一认证或独立认证。

3.2 发布流程

信息发布包括两个过程：数据搜索和数据访问。数据搜索是用户通过访问共享服务网站来发现数据资源的过程。对共享数据资源的搜索是通过节点(包括主节点及各节点)网站系统所提供的元数据搜索引擎来实现的。

用户首先访问主节点门户网站，并提交元数据检索请求，Web 服务器处理用户发送的 Web 页请求，在应用服务器上执行元数据服务等应用程序，并把用户请求构造成 Z39.50 协议消息发送至元数据网关。网关将根据协议消息中所指定的节点检索范围，通过元数据网关发送到特定分节点元数据服务器。节点服务器访问本地的元数据库(一般以 XML 格式存储)，执行 XML 查询命令，生成检索结果返回给应用服务器。其中，元数据网关起到非常重要的作用，它不但要发送用户检索条件到各元数据节点服务器，而且还要汇总、合并来自被检索节点服务器的检索结果。最后，应用服务器将检索结果返回浏览器。元数据搜索不需要认证。

用户得到检索结果(描述共享数据资源的元数据)后，可以对结果所描述的共享数据

资源的内容、质量、格式等进行评估,以确定是否需要获取该数据集,若是,就进入数据访问阶段。

数据访问是指根据元数据中共享资源定位信息(URL)对共享资源进行访问、下载、在线操作等的过程。对共享数据资源的访问是通过节点网站系统所提供的数据检索应用系统实现的。

用户通过 URL 访问分节点网站系统,该节点 Web 服务器对用户数据访问请求进行处理,或执行标准的 SQL 语句检索共享数据库(URL 提供的是数据检索参数输入页),或通过 HTTP、FTP 直接下载(URL 直接提供访问路径)。数据访问时用户还要进行注册和认证。

从纵向结构来说数据访问分四层:浏览器客户端、Web 与应用服务器、网关与元数据节点服务器和数据资源(数据库或文件目录结构),见图 5。

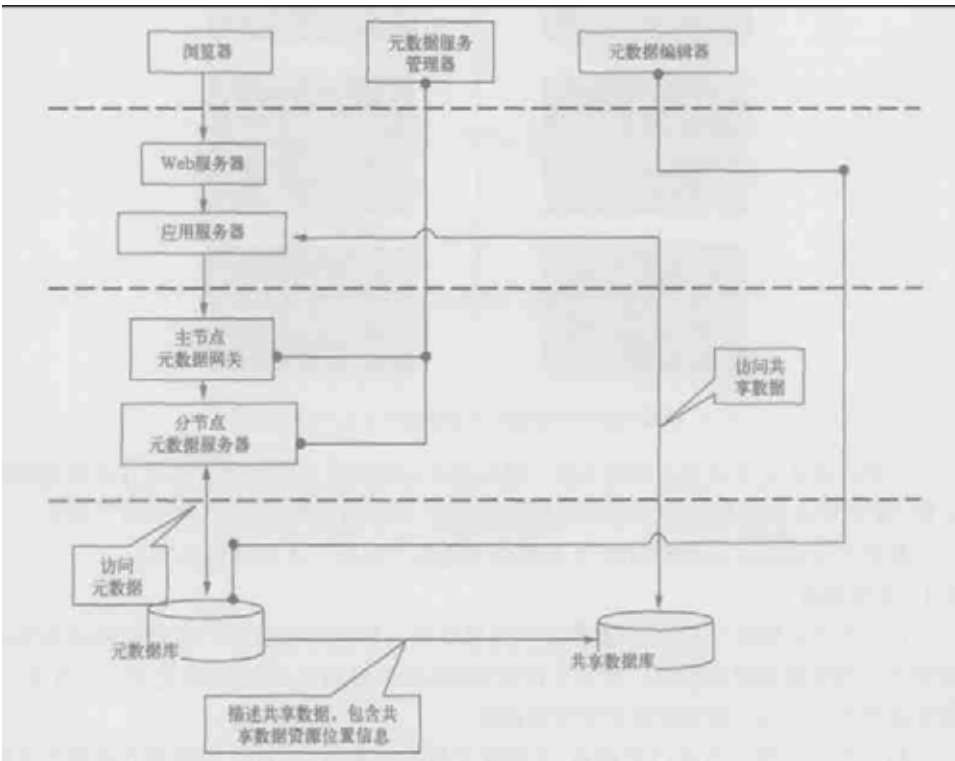


图 5 流程图

4 结 语

元数据对气象科学数据共享工作具有非常重要的意义,它甚至对整个信息技术的发展也具有不可估量的作用。通过几年的努力,气象数据集元数据格式标准已广泛应用于气象数据管理及业务系统建设中。基于元数据系统建成的中国气象科学数据共享服务网

由 1 个主节点和 8 个分节点组成,其中 3 个分节点分别部署于国家级业务和科研中心(中国气象局气象信息中心、国家卫星气象中心和中国气象科学研究院),5 个分节点分别部署于湖北、甘肃、黑龙江、江苏省气象局和西藏自治区气象局。系统实现了基于统一元数据标准的信息发布和用户一点登录全网数据透明访问,用户可以通过访问本系统获取分布在不同节点的、基于 Web 的数据检索与下载服务。

参 考 文 献

- 1 吴增祥,臧海佳,王国复.气象科学数据与气象数据集元数据.科学数据管理与共享.北京:中国科学技术出版社,2002.186~189.
- 2 蒋景瞳,刘若海,贾云鹏.国际元数据标准的发展和研究现状.中国地理信息元数据标准研究.北京:科学出版社,1999.110~113.
- 3 Federal Geographic Data Committee. Content Standard for Digital Metadata Workbook. 2000.
- 4 The International Organization for Standardization (ISO). Geographic Information Metadata (ISO TC 211/DIS 19115). 2001.
- 5 WMO Commission for Basic Systems Expert Team on Integrated Data Management Final Report. 2002.
- 6 徐枫.元数据技术及其在科学数据共享中的应用.科学数据共享管理研究.北京:中国科学技术出版社,2002.178~196.
- 7 李集明,汤绪,王国复,等.气象科学数据共享试点实践.科学数据共享管理研究.北京:中国科学技术出版社,2002.213~235.

THE RESEARCH OF METEOROLOGICAL METADATA STANDARD AND INFORMATION ISSUING

Wang Guofu¹⁾ Xu Feng²⁾ Wu Zengxiang¹⁾

¹⁾ (National Meteorological Center, Beijing 100081)

²⁾ (State Information Center, Beijing 100053)

Abstract

Metadata is making an important effect on the data exchange, information search in Web and sharing system design. The up-to-date development of metadata and metadata standards is introduced concisely. The establish principle, gist and content of our metadata standard are emphasized. The architecture and flow of metadata system are explored. At last, based on the layout and developing practice of "Meteorological Data Sharing Project", the blue print and implement of metadata issuing system for the China Meteorological Data Service Network are expounded in detail.

Key words: Metadata Issuing Sharing Gateway Metadata server