

doi:10.3799/dqkx.2010.042

数据目录体系关键技术

刘福江¹, 彭李², 郭艳³

1. 中国地质大学信息工程学院, 湖北武汉 430074

2. 长江大学地球科学学院, 湖北荆州 434023

3. 中国地质大学计算机学院, 湖北武汉 430074

摘要: 基于目录体系的政务信息资源整合有利于信息资源的共享和交换. 针对自然资源和地理空间信息资源的特点, 分析了国家电子政务资源目录体系, 提出了面向综合决策的自然资源和地理空间信息资源整合改造的数据目录体系, 这样的整合方法为信息资源的动态有机整合建立了基础, 可以适应信息资源的动态变化, 同时通过目录编制明确了数据提供者和共享部门, 为信息资源的整合和共享建立长效机制提供了依据.

关键词: 地理信息; 数据目录; 电子政务; 自然资源.

中图分类号: TP311.13

文章编号: 1000-2383(2010)03-0375-05

收稿日期: 2010-01-15

Key Technology for Data Catalog System

LIU Fu-jiang¹, PENG Li², GUO Yan³

1. Faculty of Information Engineering, China University of Geosciences, Wuhan 430074, China

2. Faculty of Earth Sciences, Yangtze University, Jingzhou 434023, China

3. Faculty of Computer Science, China University of Geosciences, Wuhan 430074, China

Abstract: E-government information resources integration based on catalog system are beneficial to the sharing and exchange of information resources. According to the characteristics of the natural resources and geo-spatial information resources, the paper analyzed the catalog system of e-government resources, and then put forward the data catalog system of the natural resources and geo-spatial information resources integration transformation to deal with synthesis decision-making. This integrated approach establishes the basis of the organic integration of information resources, and adapts to the dynamic changes of information resources. At the same time, through cataloging, it can make data providers and share sectors clear, and provides the establishment of long-term mechanism for the integration and sharing of information resources.

Key words: geographic information; data catalog; e-government; natural resources.

0 引言

中办发[2002]17号文件确定自然资源和地理空间基础信息库作为国家电子政务4个基础信息库之一, 整合我国基础性资源环境和地理空间信息, 实现跨部门分布式的、多源、多专业、多时相、多类型和海量异构地理空间信息数据进行一体化组织和管理(自然资源和地理空间基础信息库项目办公室, 2006, 2007). 针对信息资源数量繁多、门类多样和分

布广的特点, 在自然资源和地理空间基础信息库的信息化建设中, 发现和定位信息资源成为信息资源共享和开发利用过程的基本前提. 鉴于信息资源内容繁杂和规模庞大, 社会和公众对其需求的多样性, 研究自然资源和地理空间基础信息库数据目录体系的建设就显得非常必要和紧迫.

当前, 国家正在大力开展政务信息资源目录体系的建设工作, 天津、上海、北京等城市分别开展了目录体系试点工作. 政务信息资源目录是按照信息

基金项目: 国家重点“863”项目(No. 2007AA120503)资助.

作者简介: 刘福江(1973-), 男, 博士, 副教授, 主要从事空间信息标准、GIS模型方法、GIS系统架构、3S系统集成等研究.

E-mail: felixwuhan@163.com

资源分类体系或其他方式对政务信息资源核心元数据的有序排列,相同的政务信息资源核心元数据按照不同的分类标准或者分类方式排列,在表现上形成了不同的目录树结构,但是从目录树所展现的内容上来讲,都是描述政务信息资源的核心元数据(徐枫和宦茂盛,2005)。它提供政务信息资源的发现定位服务,方便用户发现、定位和共享多种形态的政务信息资源(GB/T21063.1,2007)。

通过对国家政务信息资源目录体系进行研究(GB/T21063.1,2007;GB/T21063.2,2007;GB/T21063.3,2007;GB/T21063.4,2007;GB/T21063.6,2007),分析自然资源和地理空间基础信息库信息资源整合的现状,针对自然资源与地理信息综合决策的需要,研究数据目录体系所需的关键标准,提出适合于面向综合决策的自然资源和地理空间信息资源整合改造的数据目录体系。数据目录体系可以方便地检索和定位整合工作所需要的自然资源和地理空间信息资源,以满足综合决策的需要。数据目录是从综合决策需求出发,梳理各种数据的来源、生产、更新等。根据相关的信息资源描述规范和分类规范描述以及标识信息资源,编制面向综合决策的信息资源分类目录和共享目录,在数据目录的基础上进行数据整合改造。这样的整合方法为信息资源的动态有机整合建立了基础,可以适应信息资源的动态变化,同时通过目录编制明确了数据提供者和共享部门,为信息资源的整合和共享建立长效机制提供了依据(自然资源和地理空间基础信息库项目办公室,2006)。

1 数据目录的定义

数据目录实现按照用户需求,动态以层次化目录树的形式管理数据,按照建立动态目录树的规则及建立目录树规则的工具,用户可以根据管理的实际业务类型,扩展建立目录树的规则,提供管理多源异构数据更丰富的表现。数据目录是通过目录系统对基于服务的元数据仓库数据或非元数据仓库数据按规则进行目录组织,并利用目录规则驱动实现对数据资源以目录树的形式表现,以达到对自然资源和地理空间信息资源统一有效的组织和管理,促进自然资源和地理空间信息化健康、有序发展,为信息资源交换提供了基础。数据目录主要涉及元数据仓库、目录系统和目录规则驱动技术。

2 数据目录体系概念模型

数据目录体系的概念模型包括支撑标准、数据库和信息资源目录 3 部分,如图 1 所示。

2.1 支撑标准

支撑标准包括元数据仓库标准、资源分类标准、资源标识符标准和目录规则驱动规范,它们是建立信息资源目录体系的重要基础和保障,亦形成了数据目录标准体系,其中元数据仓库标准和资源分类标准是目录体系的核心和基础。

元数据仓库标准:规定了功能仓库与数据仓库的元数据描述,存储规范及扩展体系等。适用于用户通过查询服务的元数据来确定服务的位置和获取服务的接口,用于服务中数据仓库和功能仓库的实施,为用户发现资源、使用数据提供信息。

资源分类标准:定义了信息资源分类的原则和方法。便于建立统一的多种分类体系,为信息资源的组织、查找和管理提供多种途径。

资源标识符标准:定义了信息资源目录体系中所管理的信息资源的标识符的编码结构、管理和分配原则。在目录体系中,将按标识符编码标准赋予每项信息资源一个唯一标识性代码,保证其在跨部门资源共享和交换过程中的唯一性。

目录规则驱动规范:用于定制规则驱动实现动态目录,包括目录规则定义、目录存储、规则驱动、分组建立驱动等,适用于数据目录树的实现。

2.2 数据库

数据库包括自然资源和地理空间基础信息库、

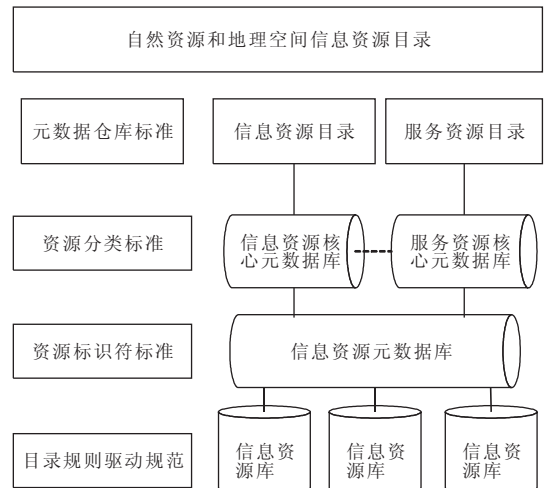


图 1 数据目录体系的概念模型

Fig. 1 Conceptual model of data catalog system

专题信息库、专题信息产品库、综合信息库、综合信息产品库、综合信息定制产品库、元数据库和核心元数据库。

自然资源和地理空间信息资源库是所有自然资源和地理空间信息资源的集合。

专题信息库由各数据分中心整合现有数据资源所建立的各种专题信息库和相应的元数据库组成,是数据主中心综合信息库的信息来源,分布在各自的数据分中心。

专题信息产品库由各数据分中心为满足电子政务和社会公众对信息产品需求利用相应的专题信息库信息所建立的各种专题信息产品库和相应的元数据库组成,分布在各自的数据分中心。各分中心的专题信息库通过交换系统汇集到主中心形成综合信息库,在综合信息库的基础上可以形式综合信息产品库和综合信息定制产品库。

分中心的元数据库是各分中心按照统一的元数据标准规范进行数据库相关联的元数据采集,提取各分中心信息库中各项信息资源的特征,形成描述信息资源本身特点的数据库。主中心的元数据库是通过交换系统将各个分中心的元数据库汇集的元数据库。

核心元数据库是在自然资源和地理空间信息资源元数据库基础上,提炼出描述自然资源和地理空间信息资源所必需的核心特征要素形成的数据库,包括数据资源核心元数据库和服务资源核心元数据库两部分。数据资源核心元数据库存放描述文本、图片、数据库等数据资源所必需的核心特征要素;服务资源核心元数据库存放描述应用系统、业务系统等服务资源所必需的核心特征要素。

2.3 信息资源目录

自然资源和地理空间信息资源目录是以自然资源和地理空间信息资源核心元数据为基础,对自然资源和地理空间信息资源进行分级、分类和格式标准化,形成的自然资源和地理空间信息资源的有序组织,包括数据资源目录和服务资源目录。

数据资源目录基于数据资源核心元数据库,对数据资源形成有序组织,便于数据资源的查询、控制和管理。

服务资源目录基于服务资源核心元数据库,对服务资源形成有序组织,便于服务资源的查询、控制和管理。

3 关键技术

3.1 SOA 技术

面向服务架构(service-oriented architecture, SOA)是一种软件系统架构模型,它将系统的每一个功能单元都定义为独立的服务并加以封装,通过良好的接口进行调用。接口采用中立的方式进行定义,屏蔽了底层复杂的基础架构,独立于实现服务的硬件平台、操作系统和编程语言。这使得异构系统中的服务可以以一种统一和通用的方式进行交互(吴信才和吴亮,2006)。

SOA 运行机制有服务提供者(service provider)、服务请求者(service requester)和服务代理者(service broker)3种角色,分别承担了不同的任务:发布(publish)、查找(find)和绑定(bind)。

3.2 元数据仓库技术

元数据仓库是实现对数据资源和功能资源这两种形式资源进行描述的集中存储(吴信才,2009)。其中,元数据仓库中描述数据资源的每条元数据包括标题、主题、摘要、全文描述、对资源的引用 URI/URL、资源类型、资源范围、资源参照、修改日期、生产厂家以及键值属性等信息;元数据仓库中描述功能资源的每条元数据包括标题、开发人员、概述、类型、资源、关联属性、功能说明以及通过扩展实现的用户增加的属性等信息。

元数据仓库管理下的元数据库是基于服务的元数据库,它描述了数据和服务的各种信息,用户通过查询服务的元数据来确定服务的位置和获取服务的接口。在分布式的网络环境中,每个节点上只有一个元数据仓库,该元数据仓库中的数据和功能资源采用唯一编码进行描述,资源标识由“服务标识+唯一标识”组成,在节点内部的元数据引用只需唯一标识。

3.3 目录系统技术

目录系统是数据组织和操作方式,它实现基于插件与类型驱动的可扩展的层次化管理多源异构数据的系统。目录系统通过基于层次数据结构存储的逻辑数据组织信息,提供一个一致、稳定的层次数据库管理器。目录系统中的一个条目对应数据中心层次数据管理结构中的一个资源节点。

3.4 目录系统驱动技术

目录系统驱动有业务规则驱动(或称 Access 分组驱动)和自定义驱动。业务规则驱动是内置驱动,

使用 AccessGDP 协议,用于访问本机中的 Access 数据库.遵循 Access 分组驱动的 XML 存储规范,可将 Access 数据库中的一个表或多个表绑定在数据中心的任意一个节点上,对应建立一颗子树,表中的字段对应子树的节点.自定义驱动:根据目录系统驱动规则,实现与目录系统具有相同基类的派生类,将自定义的数据结构驱动到目录树上管理.

数据目录对于目录树的生成有两种方式,包括单个节点配置生成目录树和通过写驱动配合规则生成目录树.其中,单个节点配置生成目录树主要是针对轻量级数据所采用的一种目录树生成方式,而对于重量级数据若采用单个节点配置的方式生成目录树将会极大地增加工作量,此时一般采用写驱动配合规则生成目录树的生成方式.这样,目录树可按不同需求进行动态分组,让用户从不同角度管理多源异构数据.

用户根据业务的需要,制定业务规则,生成目录规则 XML 文件.业务规则的形成流程见图 2 所示,分组的方式有联合分组、纵向分组、固定分组、按记录分组和项的分组.

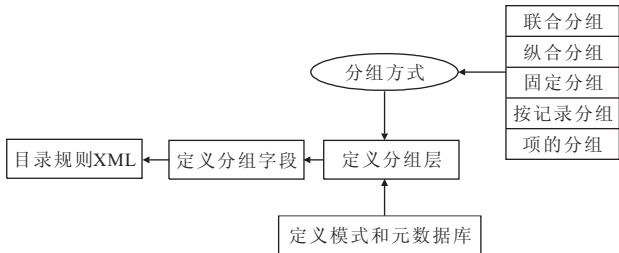


图 2 业务规则流程

Fig. 2 Process of business rules

4 数据目录形成流程

每个标准在形成数据目录所起的作用如图 3 所示,数据目录的生成包括 5 个关键步骤:

- (1)根据各专业领域特点,按照元数据仓库规范建立各专业领域的业务元数据库;
- (2)按照资源分类标准、资源标识符标准,对数据资源进行命名,生成待管理数据的编码表;
- (3)按照分组驱动建树的要求,根据各专业领域综合决策的需求,制定业务规则,生成目录规则的 XML 文件;
- (4)按照规则驱动规范,配置规则驱动文件;
- (5)生成用于展示的数据目录.

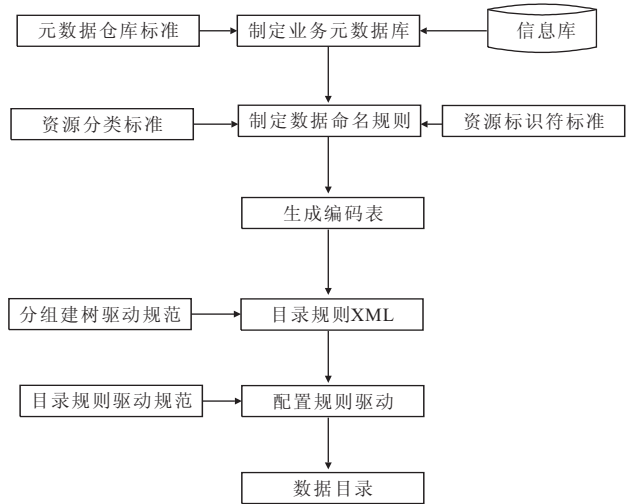


图 3 数据目录生成流程及标准

Fig. 3 Generation processes and standards of data catalog

5 结论

数据目录的表现方式非常灵活,按照信息资源粒度的不同以及信息资源类型的不同,产生不同层次和类型的信息资源目录,可以方便地根据应用需要,按行业、部门、地域、应用主题和其他使用目的变换出信息资源的各种目录,可以提供导航树,可以提供多种分类方式的浏览,可以根据用户需要对显示方式进行定制,为用户发现和定位资源提供了便利,是自然资源和地理空间基础信息库实现信息共享和交换的基础.

建立信息资源分类目录和共享目录,为实现跨部门的按需整合和共享信息资源提供依据和基础,在此基础上对信息资源进行有机地整合,可以为综合决策提供相对全面、准确、及时的信息支持.

数据目录是自然资源与地理空间信息资源开发和利用的基础设施,通过对分散在数据主中心和数据分中心的信息资源进行有效梳理,并按照统一的元数据标准、分类标准和资源标识符标准,实现对自然资源与地理空间信息资源的有序组织,整合各数据中心(数据分中心和数据主中心)的数据目录,构建自然资源和地理空间基础信息库统一的数据目录体系,形成部门间信息资源物理分散、逻辑集中的信息共享模式.

References

GB/T21063.1,2007. Government information resource catalog system-part1:framework. General Administration of

Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China, Beijing (in Chinese).

GB/T21063. 2, 2007. Government information resource catalog system-part2; technical requirement. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China, Beijing (in Chinese).

GB/T21063. 3, 2007. Government information resource catalog system-part3: core metadata. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China, Beijing (in Chinese).

GB/T21063. 4, 2007. Government information resource catalog system-part4: government information resource classification. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China, Beijing (in Chinese).

GB/T21063. 6, 2007. Government information resource catalog system-part6; technical management requirements . General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China, Beijing (in Chinese).

Natural Resource and Geo-spatial Database Project Office, 2006. Research reports of natural resource and geo-spatial database standards system. Academy of Macroeconomic Research, Beijing (in Chinese).

Natural Resource and Geo-spatial Database Project Office, 2007. Preliminary design for natural resource and geo-spatial database. Academy of Macroeconomic Research, Beijing (in Chinese).

Wu, X. C. , 2009. Datacenter integration development technology: the next generation GIS architecture and development model. *Earth Science—Journal of China University of Geosciences* , 34 (3) ; 540 — 546 (in Chinese

with English abstract).

Wu, X. C. , Wu, L. , 2006. Service-oriented distributed spatial information supporting system. *Earth Science—Journal of China University of Geosciences* , 31 (5) ; 585 — 589 (in Chinese with English abstract).

Xu, F. , Huan, M. S. , 2005. Research on the government information resource catalog system technology. *Information Technology and Standardization* , 11 ; 23 — 27 (in Chinese with English abstract).

附中文参考文献

GB/T21063. 1, 2007. 政务信息资源目录体系第1部分:总体框架. 北京:中华人民共和国国家质量监督检验检疫总局,中国国家标准管理委员会.

GB/T21063. 2, 2007. 政务信息资源目录体系第2部分:技术要求. 北京:中华人民共和国国家质量监督检验检疫总局,中国国家标准管理委员会.

GB/T21063. 3, 2007. 政务信息资源目录体系第3部分:核心元数据. 北京:中华人民共和国国家质量监督检验检疫总局,中国国家标准管理委员会.

GB/T21063. 4, 2007. 政务信息资源目录体系第4部分:政务信息资源分类. 北京:中华人民共和国国家质量监督检验检疫总局,中国国家标准管理委员会.

GB/T21063. 6, 2007. 政务信息资源目录体系第6部分:技术管理要求. 北京:中华人民共和国国家质量监督检验检疫总局,中国国家标准管理委员会.

吴信才, 2009. 数据中心集成开发技术:新一代GIS架构技术与开发模式. *地球科学——中国地质大学学报*, 34(3): 540—546.

吴信才, 吴亮, 2006. 面向服务的分布式空间信息支撑平台. *地球科学——中国地质大学学报*, 31(5): 585—589.

徐枫, 宦茂盛, 2005. 政务信息资源目录体系技术概述. *信息技术与标准化*, 11: 23—27.

自然资源和地理空间基础信息库项目办公室, 2006. 自然资源和地理空间基础信息库标准规范体系研究报告. 北京:国家发改委宏观经济院.

自然资源和地理空间基础信息库项目办公室, 2007. 自然资源和地理空间基础信息库初步设计. 北京:国家发改委宏观经济院.