

●吴 鹏 强韶华 苏新宁

政府信息资源元数据描述框架研究^{*}

摘 要 XML作为主流的元数据描述语言,可以在电子政务系统中,为政府信息资源提供统一的描述框架。框架主要由模式层、应用层、环境层三部分组成。模式层由政府信息资源元数据描述模式框架描述,基于XML的元数据描述应用框架支撑了元数据的具体应用,而XML应用框架生命周期和XML元数据技术的协同工作解决了元数据应用的可操作性和可持续性。图1。参考文献9。

关键词 电子政务 政府信息资源 元数据 描述框架

分类号 G254

ABSTRACT XML is a mainstream metadata description language and can be used in e-government systems to provide an unified description framework, which consists of model layer, application layer and environment layer. The model layer is described with the framework for the description of metadata for government information resources, the XML-based metadata description application framework supports practical metadata applications, and the lifespan of metadata application framework and XML metadata technology work together to solve the problems of operability and continuity in metadata applications. 1 fig. 9 refs.

KEY WORDS e-Government. Government information resource. Metadata. Description framework.

CLASS NUMBER G254

根据CNNIC的统计,截至2004年6月,我国政府gov.cn下的域名总数约1.4万个,政府网站约2.4万个。然而伴随着这些令人备受鼓舞的统计数据一起增长的,却是用户对政府网站信息资源建设状况的普遍不满。究其原因,当然涉及政策、体制、技术、资金、人才等方面,但有一个关键性的因素无法忽略,即缺少对现有信息资源进行标准化描述的有效机制,直接导致人们查找、利用信息不便。构建新型的标准化政府信息资源描述机制成为电子政务建设的当务之急。

元数据作为“关于数据的数据”,可以处理各种数字化和非数字化资源,特别是能较好地解决网络信息资源描述、发现、控制和管理问题。元数据可以有效描述政府信息资源,实现政府信息资源的高效发现和交流。

1 政府信息资源元数据发展现状

在电子政务建设中,统一的元数据描述框架可以进行真正意义上的政府信息资源整合,面向公众提供统一的政府信息查询,提高检索的精确率,促进信息资源的共享和利用,完善政府信息资源管理体制。

目前有MARC、DC、RDF等几种元数据格式。政府信息资源元数据描述有两种类型,分别是“国家信息定位服务”(GILS, Government Information Locator Service)和“都柏林核心元数据标准”(Dublin Core Metadata)。

依照GILS发展指引服务的国家较多,包括美国、加拿大、日本、俄罗斯等。GILS是一个辨识描述政府信息资源、提供获得该资源方式的系统,它是由美国政府在政府信息化的长期实践中建立的。而澳大利亚、加拿大、新西兰等国则尝试以都柏林核心为建置系统的标准。“都柏林核心元数据首创计划”(DC-MI, Dublin Core Metadata Initiative)应用委员会2001年9月在DC基础上推出专门针对政府信息资源管理的应用扩展集DC-Government。

GILS、DC等都是相关国际组织制定的适应广泛需求的标准体系,而在具体应用实现时,会受到各国政府的基础建设、资源状况、社会需求甚至是政治、文化、经济等方面的影响。如何结合本土需求更好地应用相关元数据标准,是各国政府共同关注的问题。许多国家的政府元数据本土化制定工作已经取得相当的成果,制定了符合本国需求的元数据规范。

为了有效辨识和描述行政部门的信息,并向用户提供获得信息方式的支持,电子政务建设领先的国家纷纷制定基于元数据政府信息资源描述框架。美国政府在1980年制定了“联邦信息指引系统”(FILS, Federal Information Inventory/Locator Systems),在1994年制定了“国家信息定位服务”。1998年澳大利亚政府基于都柏林核心元数据标准制定了澳

^{*} 本文是国家自然科学基金资助项目(项目编号70373028),同时受南京理工大学经济管理学院青年教师科研扶持基金项目资助。

大利亚政府定位服务 (AGLS, Australian Government Locator Service)。2000年5月,加拿大政府制定了“通用信息描述标准”(CLF, Common Look and Feel, www.cio-dpi.gc.ca/clfupe),这是一种基于都柏林核心元数据标准。它规定了加拿大政府网站的信息和服务的描述格式、数字加密方式、元数据格式、主机域名配置等等。2001年新西兰政府制定了新西兰政府定位服务元数据计划(NZGLS, New Zealand Government Locator Service Metadata Schema),用来指导全国的政府信息资源描述,该计划是新西兰官方的关于政府信息(网上和非网上信息)分类、发布和发现的项目,目的在于使政府机构和公众之间的交流更有效。英国政府则在2001年5月2日制定了英国电子政务元数据标准(UK e-GMS, e-Government Metadata Standard)。该标准陈述了英国政府关于建立和实现官方信息系统元数据的政策,概述了元数据在政府中的作用和应用方针,列出了发展、管理元数据结构和标准的原则。

虽然 GILS 较之 DC-Government 在专业元素设置、应用实施成熟度方面占优势,但从长远发展来看,后者将占据未来政府信息描述领域的主流地位。这主要是因为“都柏林核心元数据标准”已经是最主要的资源描述格式,采用该标准的扩展应用有利于资源共建共享,另外,随着 DC-Government 的进一步完善,它对政府信息资源的描述能力将更趋专业,与 GILS 的差距必将不断缩小。

电子政务已成为我国国家信息化建设的重点工程。一方面,由于政务公开的实施,政府信息资源得到开发,政府信息量大增,而且新增的政府信息资源数字化程度提高,原有的文本信息也进一步转化成数字信息而存在;另一方面,随着政府信息化进一步推进,我国各级政府不仅通过网络向公众提供信息服务,并且基于政府信息资源提供在线交易。虽然政府信息资源开发迅速,但利用、管理还比较混乱,政府部门间数

据难以共享,通过网络检索政府信息资源、获得政府服务还比较困难。主要原因是我国目前还没有一个统一的面向政府信息资源描述的元数据描述框架。建立一个全国范围内统一的政府元数据框架已成为发展电子政务的当务之急。

近年来,我国学者开始关注这一领域的研究。中山大学的曹树金教授对比分析了一些外国政府信息资源的元数据描述标准,并分析了中国政府的元数据应用现状,指出在中国适合采用 DC-Government 作为元数据核心集。北京大学的肖珑教授从数字图书馆的角度提出中文元数据描述框架。南京大学苏新宁教授就政府辅助决策中信息资源的元数据描述框架展开研究。笔者认为在此基础上,可以对电子政务流程中涉及的信息资源展开进一步研究,构建一个具有通用性和兼容性的元数据描述框架。

2 政府信息资源元数据描述框架

在电子政务系统中,政府信息资源支撑着政府程序和业务流程操作。为了流程化管理政府和其协作者之间的信息交换,人们需要在总体上了解政府和顾客、管理者和协作者之间交互的格式、内容。政府信息资源元数据描述框架是应用于电子政务系统中的,因此该标准需要采用一种语言格式加以定义,XML 作为主流的元数据描述语言,可以提供统一的描述框架。XML 可以根据政府需要,根据对象的属性以及对象的关系绘制,建立面向电子政务系统物理数据源和数据流程的元数据模型。政府信息资源元数据描述框架根据中国电子政务建设的需要,形成一个基于 XML 语法的元数据语义描述。该框架主要由模式、应用、环境三部分组成(图1),其中模式层由政府信息资源元数据描述模式框架描述,基于 XML 的元数据描述应用框架支撑了元数据的具体应用,而 XML 应用框架生命周期和 XML 元数据技术的协同工作解决了元数据应用的可操作性和可持续性。

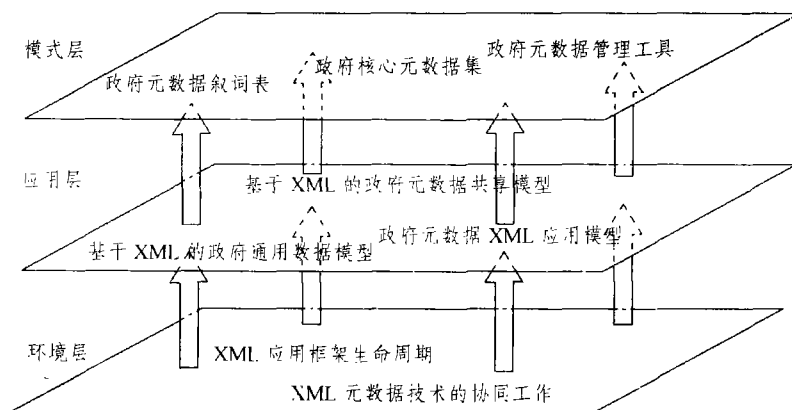


图1 政府信息资源元数据描述框架

2.1 政府信息资源元数据描述模式框架

电子政务信息资源元数据描述模式框架制定了关于政

府信息(网上和非网上信息)分类、发布和发现的规则,它的目的在于为政府和其服务对象之间建立无缝隙的信息交流

和共享平台。该模式框架包括核心元数据集、元数据叙词表、元数据管理工具。核心元数据集是政府机构用来描述信息和服务的描述结构,包括以下五类:一是为检索者提供概要性信息;二是提供所描述资源的细节信息的元素;三是描述数据存在理由的元素;四是描述资源间关系的元素;五是提供书目控制信息的元素。元数据叙词表是描述信息和服务所用的词汇,由国家图书馆等图书文献机构管理,政府信息资源描述模式框架的功能和主题元素需要使用功能词表(FONZ)和主题词表(SONZ),这两个词表中的词汇结合起来共同描述政府信息、服务以及其他资源。元数据管理工具是使政府机构获得、管理和应用描述的工具,元数据管理工具是有关政府服务和资源信息的电子目录,由政府机构建立并进行内容维护。

2.2 基于XML的元数据描述应用框架

元数据描述框架得以顺利实施的关键是政府建立统一的基于XML技术的应用框架,该框架可以概述元数据在政府中的作用和应用方针,列出展示、实现、管理元数据结构的原理。该框架包括下列内容:一是基于XML的政府元数据共享模型,它描述元素和限制属性的元数据共享模式及其使用指南,从而为不同种类的信息建立分类标准,以识别异构信息资源,为政府信息资源建立主题索引,按照服务内容进行合理组织,以服务不同类型和层次的顾客。二是基于XML的政府通用数据模型,它定义政府业务的信息安全、信息隐私、信息访问和记录管理,规定了政府网站信息和服务的描述格式、数字加密方式、主机域名配置等。三是政府元数据XML应用模型,它在实施过程涉及技术标准和应用标准,技术标准主要用于对数据的加工和处理,包括基于XML的数据描述模式、统一建模语言(UML)的应用、资源描述框架(RDF)等;应用标准包括基于XML的表格标准、中间件等相关标准。

2.3 XML应用框架生命周期研究

电子政务环境中,政府的元数据描述框架应该是独立于技术之外的。除此之外,元数据描述框架应该有稳定生命周期,支持信息资源的管理、发现。目前,W3C的XML标准获得业界的广泛支持,W3C的XML标准目前是作为各国电子政务的主流标准,但是其主要功能是被设计用来为XML文档提供精确的建模规则,而不是支持在兼容环境中开放的、可扩展的元数据应用模式,这样W3C的XML标准并不非常适合于描述元数据。因此在开发中国电子政务元数据标准时就需一个通用的XML描述机制,而不局限于W3C的XML标准的特定规则。

2.4 XML元数据技术的协同工作

互操作性是指一个系统或产品在用户无须作出特别努力的情况下,与其他系统或产品协同工作的能力。网络分布式环境中,互操作性作为整合最大范围的可获取信息服务的一种手段越来越重要。

XML元数据技术领域有一系列的标准,这些标准之间

通常是互补而不是竞争的关系,但是这些标准来自不同的应用行业,并且都面向互联网发展应用技术,例如来自信息科技领域的Dublin Core,来自人工智能领域的DAML/OIL,电子出版领域的ISO 13250主题地图,元数据领域的ISO 11179标准,以及来自W3C组织的标准。尽管目前电子政务元数据需要标准的XML描述,但是更为重要的是政府组织需要从不同的行业标准中选择技术解决方案。这一点尤其重要,因为政府的具体部门和某个具体行业是紧密相连的。

为了使资源在最大范围内得到共享和检索,需要一个XML元数据技术的协同工作模式,其中包括:XML框架能够为使用XML表示元数据提供工具;XML框架能够提供元数据的消息接口,以兼容不同的元数据标准;XML框架对于XML元数据技术的选择能够提供帮助和实例。

“政府信息资源元数据描述框架”是电子政务信息资源管理的前沿研究领域,在我国该项研究尚处于起步阶段。该研究的开展将会对我国电子政务环境下政府信息资源描述的研究产生积极影响,同时对各级政府电子政务的健康发展有一定指导作用。

参考文献

- 1 赵培云. 理性思考电子政府信息资源建设. 现代情报, 2003(10)
- 2 陈莹芳. 美国政府信息指引服务的发展成因及诠释资料. 大学图书馆学报(台湾), 1999(1)
- 3 曹树金等. 论政府信息资源的元数据标准. 情报学报, 2004, 23(6)
- 4 肖珑等. 中文元数据标准框架及其应用. 大学图书馆学报, 2001, 19(5)
- 5 DC-government working group meeting. Shanghai, China, 11 october 2004
- 6 Office of the e-Envoy. e-Government Metadata Standard 1.0, 2002
- 7 New Zealand E-Government Unit. New Zealand Government Locator Service (NZGLS) Metadata Schema Compliance Study, 2001
- 8 Ann Wrightson. Representing Metadata in XML: the UK e-Government Metadata Standard
- 9 Alan Bundy. Australian and New Zealand Information Literacy Framework. Australian and New Zealand Institute for Information Literacy, 2004

吴 鹏 南京理工大学信息管理学系博士, 讲师。通信地址: 南京。邮编 210094。

强韶华 南京大学商学院博士研究生, 讲师。通信地址: 南京。邮编 210093。

苏新宁 南京大学信息管理学系教授, 博士生导师。通信地址: 南京。邮编 210093。 (来稿时间: 2006-05-31)