

●金更达 何嘉荪

档案信息资源集成管理中的元数据 问题及对策研究^{*}

摘 要 元数据问题是文件档案信息资源集成管理的基础问题。在集成管理过程中存在着元数据功能定位、元数据标准制定、元数据交换和对元数据的管理问题。解决这些问题的对策是:建立元数据注册系统,建立元数据仓储和充分利用有关元数据的技术协议标准。图2。参考文献8。

关键词 档案信息资源 信息资源集成管理 元数据管理 元数据仓储

分类号 G254

ABSTRACT In integrated management of archival information, there are some problems concerning metadata, such as function determination, standard drafting, metadata exchange and metadata management. The authors propose to establish a metadata registration system, establish metadata repositories and fully utilize related standards and specifications. 2 figs. 8 refs.

KEY WORDS Archival information resource. Integrated management of information resources. Metadata management. Metadata repository.

CLASS NUMBER G254

档案信息资源的集成管理已成为档案学理论界、实践部门以及广大用户普遍关注的问题。而元数据是数字信息组织和处理的基本工具。尽管在各个领域对元数据的理解和定义不同,但其重要性在信息社会中已不容置疑。无论是国际上还是国内各个行业都将元数据标准化制定作为国家和行业信息化建设的基础性工作。根据近年档案工作的实践,我们认为要实现文件档案信息资源集成管理和集成服务,首先必须解决元数据问题。

1 集成内涵

集成(integration)含综合、融合、整合成为整体和一体化的意思,其哲学内涵是指由系统的整体性及系统核心的统摄、凝聚作用而导致的使若干相关部分或因素合成为一个新的统一整体的建构及其有序化过程^[1]。文件档案信息资源集成管理,就是将集成管理理论应用于涵盖文件档案信息资源生命周期管理实践的全过程,即在管理思想上以集成理论为指导,在管理行为上以集成机制为核心,在管理视野上突破城市建设管理业务流程机构部门间的限制,并从全程管理和最优化管理的层面来对待各种文件档案信息

资源要素,提高各种管理要素的交融度,以利于优化和增强档案信息资源的真实性、完整性、有效性和有序性,最终提供以用户需求为导向的档案信息集成服务。

随着社会各部门信息化建设的逐步开展,档案信息资源集成管理已从面向机构内具体业务的纵向集成转向面向业务流程的机构部门间或跨机构的横向集成,具体表现为业务流程的集成、异构应用系统的集成、异构数据源的集成、业务领域与软件领域的集成、人机集成、界面集成等等。其中文件档案信息的集成是实现集成的基础,即集成首先需要解决的基础核心问题就是建立文件档案信息资源的形成、采集、加工、管理、长期保存、交换和共享所需要的方法、标准、工具。由于集成必须以元数据为最基本和最重要的工具,因而上述核心问题,实际上就是元数据问题,包括元数据功能定位、元数据标准制定、元数据交换和元数据管理问题。

2 元数据问题

2.1 元数据功能多元化问题

不同的应用领域对元数据的功能定位各不相同。

^{*} 本文系国家自然科学基金项目“城市建设文件档案信息集成管理与集成服务研究”(编号70273055)研究成果之一。

在图书情报领域,元数据功能侧重于描述、发现、定位等功能;在档案领域,侧重于保障真实性、完整性和有效性。但在同一领域对元数据的理解不同,也会造成功能定位的不同,例如我国电子文件归档与管理规范(GB/T 18894-2002)认为:“元数据指描述电子文件数据属性的数据,包括文件的格式、编排结构、硬件和软件环境、文件处理软件、字处理和图形工具软件、字符集等数据。”其功能侧重于对文件的描述。而文件管理国际标准(ISO15489)认为:“元数据是描述文件的背景、内容、结构及其整个管理过程的数据”^[2],其功能不仅侧重描述,也侧重管理。

在文件档案信息集成管理与服务中,元数据功能已呈现多元化发展趋势。元数据既可用于保障相关电子文件的真实性、完整性、一致性、关联性和长期有效性,又可帮助对分布式网络环境中文件档案信息资源进行有效集成管理和协助提供集成服务,更是构建文件档案信息资源组织体系必不可少的工具,同时还可以为计算机和网络系统进行自动辨析、分解、提取和分析归纳的数据。

2.2 元数据标准多样性问题

目前档案信息资源集成管理所面对的信息对象不仅是各级档案馆馆藏得到的数字化电子文件,还包括各种机构的业务系统产生的各种类型的文件档案信息。这些文件档案信息不都具有统一的结构,有些是非结构化数据(如WORD文件),有些是结构化数据(如数据库文件),有些又是半结构化数据(如电子邮件),就是说这些电子文件既具有异构性,又具有多样性,而且是分布存放在上述结构各异的信息系统平台。从我国社会信息化现状来看,这些多样化的文件档案信息不可能完全使用相同的元数据标准。因此,在文件档案集成管理与服务中,一方面要根据信息对象的类型分别建立相关的元数据标准;另一方面必须解决多样化元数据标准的应用问题。

2.3 元数据管理问题

元数据是文件档案集成管理的数据基础。通常情况下,文件档案信息元数据分散地存储于各种单位、部门的各个信息系统的关系数据库。中。这些信息系统没有担负长久保存电子文件的责任,故相应产生的元数据很容易不可预测地受到破坏,无法保证它们的长期有效。因此,有必要将文件档案信息元数据在形成之后及时集中起来统一管理,并在全

程管理和最优化管理的层面来对待各种文件档案信息资源要素,提高各种要素的交融度。这种做法的优点主要有:首先,可以集中建立专门用于保护元数据、使之安全和长期有效的屏障;其次,可以及时获取元数据信息,以确保文件档案信息资源的真实性、完整性和有效性;第三,可以有效建立统一的文件档案信息访问机制。

2.4 元数据交换共享问题

现有的集成应用范例中,集成的首要目的是实现信息共享。到目前为止,许多部门的信息系统基本属于自主开发,较少考虑到整个社会的总体需要。同时由于缺乏集成管理理论的指导、国家相关标准的滞后以及信息技术范式转换的普遍性,这些信息系统完全成为一个封闭的“竖井”,即构成了所谓的信息孤岛现象。这不仅造成文件档案信息系统以及资源建设的重复,也忽略了档案信息资源使用的不可预测性。要打通“竖井”,消除信息孤岛,实现文件档案信息在不同机构间的互通和共享,就必须在建立共同数据标准的同时,解决元数据交换共享问题。

3 对策

3.1 建立元数据注册系统

元数据注册系统(Metadata Registry, MR)是对元数据的定义信息、置标方案、转换规则、著录规则、应用指南等规范进行发布登记管理和检索的系统;它可以支持网络环境中元数据规范的发现、识别、解析、调用以及在此基础上的元数据转换、挖掘和复用^[3]。

元数据注册系统首要目标,是建立一个权威、可靠、可扩展和可持续的元数据规范及其应用系统的登记管理机制。文件档案信息集成管理将面对多样性的元数据标准。为了使各单位、各部门的信息系统内部和信息系统之间能够识别这些多样性的元数据标准,并进行相互调用、转换、挖掘和复用元数据,有效的方法是建立元数据注册系统。

目前,元数据注册系统有3种类型:(1)单一命名域元数据注册系统,负责管理一个命名域的元数据规范;(2)跨命名域元数据注册系统,管理多个命名域的元数据规范,按照统一的数据模型对不同命名域的元数据规范进行描述和管理;(3)分布式元数据注册系统,由多个元数据注册系统组成,它们可以是单一命名域元数据注册系统,也可以是跨命名域元数据

注册系统。

类型可以选择为分布式元数据注册系统(如图1所示)。

文件档案信息的管理要求实现跨文件系统、档案

系统和信息资源管理系统的集成。元数据注册系统

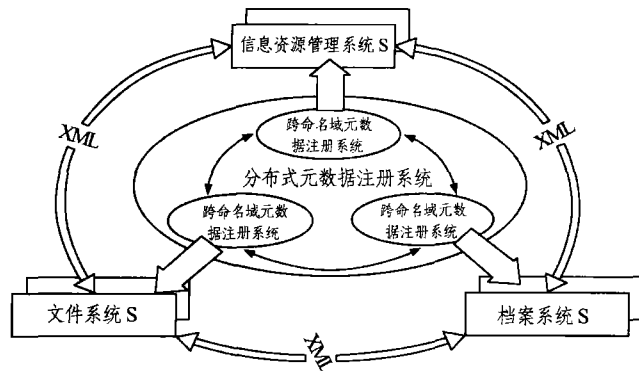


图1 分布式元数据注册系统模型

文件系统、档案系统和信息资源管理系统可以拥有自己所属范围的跨命名域元数据注册系统,元数据注册系统之间有相互调用的功能。文件系统、档案系统和信息资源管理系统的信息交换共享均采用XML技术实现。当系统接收到从其他系统输送过来的XML数据时,就启动XML解析工具,向元数据注册系统申请调用相关的元数据规范,从而起到发现、信息识别、解析、调用以及在此基础上的元数据转换、挖掘和复用等功能。

3.2 建立元数据仓储

目前,有关元数据管理的方法主要有两种:第一,

建立一个元数据访问和整个元数据生命周期管理的系统——元数据仓储,它是元数据访问和聚集的平台,用于实现元数据的集成。第二,建立一种元数据交换的途径(即元数据交换协议),不同系统中的元数据均可以通过这个途径来互访,从而把分布、异构的系统集成在一起实现元数据管理的功能^[4]。

元数据仓储可以为元数据的集中管理提供一个可操作性的框架体系。元数据仓储系统框架体系如图2所示。

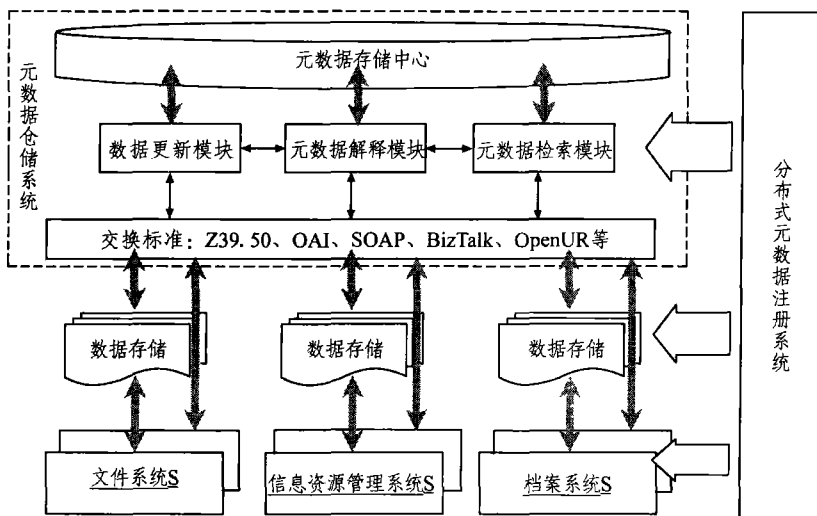


图2 元数据仓储系统框架体系

图中的元数据解析模块,为各个文件系统、档案系统和信息资源管理系统提供标准的元数据格式,以及实现对那些需要交换共享的元数据信息包的合法性验证等功能。元数据更新模块执行各个系统对元数据的更新请求,并将执行结果反馈给各请求系统。元数据检索模块执行各个文件系统、档案系统和信息资源管理系统所发出的检索命令,并将检索结果以一定的格式(如 DIP)返回给用户。

3.3 在集成管理中充分应用元数据相关技术协议标准

元数据管理、元数据交换共享的实现有赖于元数据相关技术标准、协议规范的应用。随着网络技术、计算机技术的迅速发展,元数据相关技术、协议规范得到了广泛应用。在文件档案信息集成管理和集成服务中也应该充分应用现有的元数据相关协议标准。

网络环境下,越来越多的 Web 应用通过网络能够交换数据来实现互操作,而这得益于可扩展标记语言(Extensible Markup Language, XML)的出现与发展。目前很多高层的应用协议,比如开放贸易协议、SOAP、BizTalk 等等,都是基于 XML 这个低层规范制订的。同时,XML 也是未来的数据库技术,在应用与系统之间交换数据都将被使用^[5]。XML 以及基于 XML 的标准、协议、技术将成为文件档案信息集成管理与集成服务中应用的主流技术。

架构在 Web 环境中的数字图书馆相关研究日新月异,针对 Web 环境所发展出来的分散式协议也越来越多,其中在集成管理与集成服务中最具应用价值的是 Open Archives Initiative Protocol for Metadata Harvesting(简称 OAI 协定),Simple Object Access Protocol(简称 SOAP)和 OpenURL 协议。这些基于 Web 构架的分散式协议将对实现文件档案信息元数据交换、共享发挥重要作用。

制订 OAI 协定的目的是提供一个应用程序独立、且可相互操作并且能够提供和提升 Web 上多种从事于文件内容出版发行的社群应用的框架(包括电子政务和档案信息管理)。其主要目标为:简化文件内容有效的传播,提升电子文件的存取,扩展存取数字化资源种类的范围^[6]。

SOAP,即简单对象访问协议,这是一个基于 XML 的协议。制订它的目的是在分散或分布式环境中实

现信息交换。其设计思想是利用标准化的动作,解决跨平台异构系统的沟通问题。随着计算机技术的不断发展,现代组织、企业面临的环境越来越复杂,其信息系统大多数为多平台、多系统的异构系统,迫切要求实现跨平台的、与语言无关的、也与特定接口无关的、对 Web 应用程序的可靠访问。SOAP 正是为适应这种需求发展起来的。SOAP 具有异构系统间的整合沟通能力,在实现“基于 Web 无缝集成”目标中将发挥十分重要的作用^[7]。

OpenURL 是在 Web 间传递信息的机制和应用于 Web 上超链接的一种标准表述语法,它可以根据一组已经定义好的标识(Tag),增进 Web 超链接能力。人们在因特网上获取信息时需要透过层层网卡才可取得资料。现在有了 OpenURL,则只需要手指一点就可得到,而无论信息存放在何处。事实上 OpenURL 机制在“基于 Web 的无缝集成”中发挥着越来越重要的作用^[8]。

参考文献

- 1 李宝山,刘志伟. 集成管理——高科技时代的管理创新. 北京:中国人民大学出版社,1998
- 2 ISO15489: Information and documentation——Records management
- 3 张晓林,梁娜等. 数字图书馆元数据登记系统. <http://cdls.nstl.gov.cn/mt040526/archives/docs>.
- 4 黄菊光,王申康. 元数据管理策略的比较研究. 计算机应用,2001(2)
- 5 [英]Mark Birbeck 等. XML 高级编程. 北京:机械工业出版社,2002
- 6 <http://www.openarchives.org/OAI/openarchivesprotocol.html>, The Open Archives Initiative Protocol for Metadata Harvesting
- 7 <http://www.w3.org/TR/SOAP/>, Simple Object Access Protocol (SOAP) 1.1, W3C Note 08 May 2000
- 8 <http://www.sfxit.com/openurl/openurl.html>, OpenURL syntax description.

金更达 浙江大学档案馆副研究馆员。通信地址:杭州市。邮编 310028。

何嘉荪 浙江大学信息资源管理研究所教授。通信地址同上。
(来稿时间:2005-10-24)