《现代图书情报技术》2001 年第 2 期第 9-11 页，续 15 页

元数据开发应用的标准化框架

张晓林

文摘：本文首先通过 Metadata 开发应用框架系统化地提出了 Metadata 应用开发中的相关问题及其联系，然后对各个领域 Metadata 格式、Metadata 结构、Metadata 编码语言（包括 XML）、Metadata 互操作体系（包括 RDF、元数据映射和数字对象方法）等问题及其标准进行了仔细的探讨。

关键词：元数据、元数据开发应用框架、扩展标记语言、资源描述框架、元数据映射

The Standardization Framework of Metadata

Abstract: Based on a Metadata Development and Application Framework, the paper gives a systematic presentation of related issues with the development and application of metadata. Then the paper explores in some details of the problems and standardization efforts related to metadata formats in different fields, metadata structures, metadata encoding languages (including XML), and metadata interoperability mechanisms (including RDF and Metadata mapping).

Keywords: Metadata, Metadata Development and Application Framework, XML, RDF

1． 前言

Metadata 作为“关于数据的数据”，既为各种形态的数字化信息单元和资源集合提供规范、普遍的描述基准和方法，又为分布的、由多种数字化资源有机构成的信息体系（如数字图书馆）提供整合的工具与纽带，在数字化网络化信息资源组织与利用中正发挥着日益重要的作用。本文拟从 Metadata 应用开发环境角度对 Metadata 及其应用开发的相关问题进行系统和关联的描述，并对 Metadata 格式、Metadata 结构、Metadata 编码语言、Metadata 互操作性等问题及其标准化进行分析。

2． Metadata 开发应用框架

Metadata 开发应用中涉及一系列问题。从 Metadata 生命周期的角度，这些问题包括 Metadata 的具体应用领域和应用目标、Metadata 结构、Metadata 编码语言、Metadata 制作机制、Metadata 互操作体系和检索体系、以及 Metadata 的长期保存。我们用图 1 表示按照生命周期建立的 Metadata 开发应用框架，帮助理解这些问题的结构和关系。

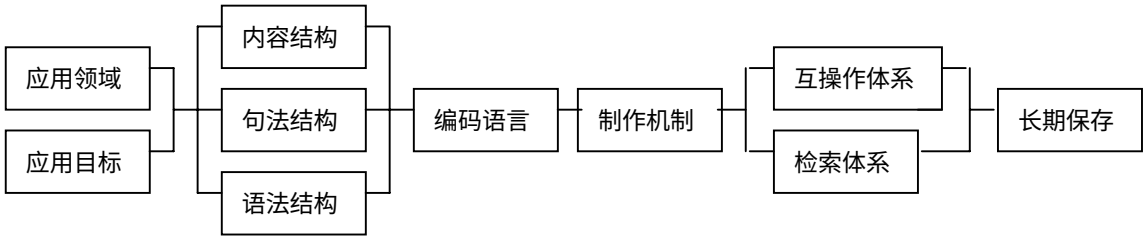

图 1 Metadata 开发应用框架

3． Metadata 应用环境

3.1 多样化的 Metadata 应用目的

在 Metadata 开发应用中，首要的问题是明确 Metadata 的应用目的。根据文献分析，可以将 Metadata 应用目的分为以下几类[1-2]：

（1）确认和检索（Discovery and Identification），主要致力于如何帮助人们检索和确认所需要的资源，数据元素往往限于作者、标题、主题、位置等简单信息，Dublin Core 是其典型代表。

（2）著录描述（Cataloging），用于对数据单元进行详细、全面的著录描述，数据元素囊括内容、载体、位置与获取方式、制作与利用方法、甚至相关数据单元方面等，数据元素数量往往较多，MARC 和 FGDC/CSDGM 是这类 Metadata 的典型代表。

（3）资源管理（Resource Administration），支持资源的存储和使用管理，数据元素除比较全面的著录描述信息外，还往往包括权利管理（Rights/Privacy Management）、电子签名（Digital Signature）、资源评鉴（Seal of Approval/Rating）、使用管理（Access Management）、支付审计（Payment and Accounting）等方面的信息。

（4）资源保护与长期保存（Preservation and Archiving），支持对资源进行长期保存，数据元素除对资源进行描述和确认外，往往包括详细的格式信息、制作信息、保护条件、转换方式（Migration Methods）、保存责任等内容。

3.2 应用于不同领域的 Metadata 格式

1990 年代以来，许多 Metadata 格式在各个不同领域出现，我们简单归纳如下：

网络资源：Dublic Core[3]、ROADS Template[4]、CDF（Channel Definition Format）[5]、Web Collections[6]、

文献资料：MARC（with 856 Field）[7]，Dublic Core

人文科学：TEI Header（Text Encoding Initiative Header）[8]

社会科学数据集：ICPSR SGML Codebook（Inter-university Consortium for Political and Social Research）[9]

博物馆与艺术作品：CIMI（Computer Interchange of Museum Information）[10]、CDWA（Categories for the Description of Works of Arts）[11]、RLG REACH Element Set[12]

视觉资料：VRA （Visual Resources Association）Core Categories for Visual Resources [13]

音乐资料：SMDL（Standard Music Description Language）[14]

政府信息：GILS（Government Information Locator Service）[15-16]

地理空间信息：FGDC/CSDGM（Federal Geographic Data Committee/Content Standards for Digital Geospatial Metadata）[17]

数字图像：MOA2 metadata[18]、CDL metadata[19]、Open Archives Format[20]、VRA Core、NISO/CLIR/RLG Technical Metadata for Images[21]

档案库与资源集合：EAD（Encoding Archival Description）[22]

技术报告：RFC 1807：A Format for Bibliographic Records[23]

FTP 文件和 FTP 文件库：IAFA Templates（Internet Anonymous FTP Archives Templates）[24]

连续图像：MPEG-7[25]

3.3 Metadata 的应用程度

由于各种原因，不同领域的 Metadata 处于不同的标准化阶段。例如，在网络资源描述方面，Dublin Core 经过多年和国际性努力，已经成为一个广为接受和应用的事实标准；在政府信息方面，由于美国政府大力推动和有关法律、标准的实行，GILS 已经成为政府信息描述标准，并在世界若干国家得到相当程度的应用，与此类似的还有地理空间信息处理的 FGDC/CSDGM；但在某些领域，由于技术的迅速发展变化，仍然存在多个方案竞争，典型的是数字图像的 Metadata，现在提出的许多标准都处于实验和完善的阶段。

不过，Metadata 开发应用经验已经明确，不可能有一个 Metadata 来满足所有领域的数据描述需要；即使在同一个领域，也可能为了不同的目的而需要不同的但可相互转换的

Metadata。

4． Metadata 结构

对于一个 Metadata 格式来说，它由多层次的结构组成：

（1）内容结构（Content Structure），对该 Metadata 的构成元素及其定义标准进行描述。例如，一个 Metadata 的构成元素可能根据其目的而包括信息内容描述性元素、技术性元素、管理性元素、结构性元素（例如与编码语言、Namespace、数据单元等的链接），Metadata 内容结构需要对所采用的元素进行准确定义和描述。但是，这些数据元素很可能是依据一定的定义标准来选取的，因此 Metadata 内容结构中需要对此进行说明，例如 MARC 记录所依据的 ISBD，EAD 所参照的 ISAD（G），ICPSR 所依据的 ICPSR Data Preparation Manual。

（2）句法结构（Syntax Structure），定义 Metadata 结构以及如何描述这种结构，例如元素的分区分段组织、元素选取使用规则、元素描述方法（例如 Dublin Core 采用 ISO/IEC 11179 标准）、元素结构描述方法（例如 MARC 记录结构、SGML 结构、XML 结构）、结构语句描述语言（例如 Extended Backus-Naur Form notation）等。在有些情况下，句法结构需要指出 Metadata 数据是否与所描述的数据对象捆绑在一起（bounded with the object）、或作为单独数据存在但以一定形式与数据对象链接，还可能描述与定义标准、DTD 结构和 Namespace 等的链接方式。

（3）语义结构（Semantic Structure），定义 Metadata 元素的具体描述方法，尤其是定义描述时所采用的标准、最佳实践（Best Practices）或自定义的描述要求（Instructions）。有些 Metadata 本身就定义了语义结构，而另外一些情况下则由具体采用单位规定语义结构，例如 Dublin Core 建议日期元素采用 ISO 8601、资源类型采用 Dublin Core Types、数据格式可采用 MIME、识别号采用 URL 或 DOI 或 ISBN[3]；又如 OhioLink 在使用 VRA Core 时要求主题元素使用 Art and Architecture Thesaurus、Thesaurus for Graphic Materials 和 Thesaurus of Geographic Names，人名元素用 Union List of Artists Names[26]。

5． Metadata 编码语言

Metadata 编码语言（Encoding Languages）指对 Metadata 元素和结构进行定义和描述的具体语法和语义规则，常称为定义描述语言（Definition Description Languages，DDL）。在 Metadata 发展初期人们常使用自定义的记录语言（例如 MARC）或数据库记录结构（如 ROADS 等），但随着 Metadata 格式的增多和互操作的要求，人们开始采用一些标准化的 DDL 来描述 Metadata。目前最为流行的当数 XML[27]和 SGML[28]。

XML 承袭 SGML 的思想和基本结构，致力于建立一个相对简单、通用、标准的文献内容与组织结构描述方法，使其独立于任何系统、设备、语言和应用。一个用 XML 标记的文献由 XML 前言（XML Prolog）和 XML 实例（XML Instance）组成，其中 XML 前言包括 XML 陈述（XML Declaration）和 XML 文献类型定义（XML Document Type Definition），而 XML 实例则是 XML 实际标记的具体文献内容。

XML 陈述详细说明文献标记使用的 XML 语言版本、字符集及是否引用外部语法规则等，XML DTD 则具体定义适用于特定类型文献的标识元素集合。在定义元素时，XML 规定了元素名称、标识符、内容模型；对每一个被定义的元素，XML 进一步定义其属性表，包括属性名、属性值表和默认值。另外，XML 还可定义有关实体（Entity）和注释（Notation）。通过这些定义，XML 可以自定义一个 Metadata 格式结构，所有嵌有 XML 解析器（XML Parser）的系统（例如标准浏览器）能利用这个定义解析 Metadata 格式，从而释读相应的 Metadata 数据。

为了进一步规范使用 XML 定义 DTD 和标记文献的行为，有关研究机构还制定了其它标准，例如：Xschema: Representation of XML DTDs as XML Documents[29]，XMI: XML

Medatada Interchange[30]，XSL: eXtensible Stylesheet Languange[31]。

6． Metadata 互操作性

我们已经看到，不同的领域（甚至同一领域）往往存在多个 Metadata 格式，当在用不同 Metadata 格式描述的资源体系之间进行检索、资源描述和资源利用时，就存在 Metadata 的互操作性问题（Interoperability）。这涉及多个不同 Metadata 格式的释读、转换和由多个 Metadata 格式描述的数字化信息资源体系之间的透明检索。

6.1 Metadata 映射

解决 Metadata 互操作问题的一种方法是进行 Metadata 格式转换，被称为 Metadata 映射（Metadata mapping、Metadata crosswalking）。目前已有大量的转换程序存在[32]，供若干流行 Metadata 格式之间的转化，例如 Dublin Core 与 USMARC、Dublin Core 与 EAD、Dublin Core 与 GILS、GILS 与 MARC、TEI Header 与 MARC、FGDC 与 MARC 等。不过，这种方法在面对多种 Metadata 格式并存的开放式环境中的应用效率明显受到限制。

6.2 标准描述方法

解决 Metadata 互操作性的另一种思路是建立一个标准的资源描述框架，用这个框架来描述所有的 Metadata 格式，那么只要一个系统能够解析这个标准描述框架，就能解读相应的 Metadata 格式。实际上，XML 和 RDF（Resource Description Framework）[33-34]从不同角度起着类似的作用。XML 通过其标准的 DTD 定义方式，允许所有能够解读 XML 语句的系统辨识用 XML-DTD 定义的 Metadata 格式，从而解决了对不同格式的释读问题。RDF 则定义了一个由资源（Resources）、属性（Properties）和声明（Statements）等三种对象组成的基本模型，其中资源和属性的关系类似于实体-关系模型，而声明则对资源与属性的关系进行具体描述。RDF 通过这个抽象的数据模型为定义和使用 Metadata 建立了一个框架，Metadata 的元素可看成 Metadata 所描述的资源的属性。进一步地，RDF 定义了标准 Schema，规定了声明资源类型、声明相关属性及其语义的机制和定义属性与其它资源间关系的方法。在进行上述声明和定义时，RDF 还规定了利用 XML Namespace 方法调用已有定义规范的机制，从而可直接在 RDF 中引用诸如 Dublin Core 或其它 Metadata 定义。在这种情况下，人们可以利用 RDF 来解读所引用的 Metadata。

6.3 数字对象方法

在数字图书馆研究中，通过建立包含 Medatada 的数字对象，人们试图从另一个角度解决 Medatada 的互操作性问题。例如在 Cornell 大学 FEDORA 项目中[35]，提出了一个由内核（Structural Kernel）和功能传播层（Disseminator Layer）组成的复合型数字对象。在内核里，可以容纳以比特流形式存在的文献内容、描述该文献的 Metadata、以及对这个文献及 Metadata 进行存取控制的有关数据；在功能传播层，有主功能传播器（PrimitiveDisseminator）支持有关解构内核数据类型和对内核数据进行读取的服务功能，还可有内容类型传播器（Content-Type Disseminators），它们内嵌 Metadata 格式转换机制。例如，在一个数字对象的内核中存有 MARC 格式的 Metadata，在功能传播层装载有请求 Dublin Core 格式及其转换服务的内容类型传播器。当数字对象使用者要求读取以 Dublin Core 表示的 Metadata 时，相应的内容类型传播器将通过网络请求存储有 Dublin Core 及其转换服务程序的数字对象，然后将被请求数字对象中的 MARC 形式 Metadata 转换为 Dublin Core 形式，在输出给用户。

6.4 分布式异构 Metadata 系统的透明检索

在现实网络信息环境中，我们往往遇到的是用不同 Metadata 描述的多个异构资源系统组成的开放型资源体系，需要有效的基于分布式系统的方法实现跨 Metadata 格式和跨系统的透明检索。这类透明检索的实施主要是基于公共检索协议，例如 Z39.50 协议[36]（用于对异构 MARC 系统和 GILS 系统的开放检索），以及 X.500、Lightweight Directory Access Protocol

（LDAP）等协议。另外，诸如 Imesh[37]、ROADS[4]等系统还对基于 Metadata 格式描述的主题信息网关交叉检索进行研究。

参考文献：

[1]张晓林。网络环境的信息组织：新问题与新方向。图书馆杂志，1998理论年刊

[2]张敏。网络信息资源组织。四川大学，硕士学位论文，1999

[3]Dublin Core. http://purl.oclc.org/metadata/dublin_core/

[4]Resource Organisation and Discovery in Subject-based services. http://www.ilrt.bris.ac.uk/roads/

[5]Channel Description Format. http://www.w3.org/pub/WWW/TR/WD-xml-961114.html

[6]Web Collections. http://www.w3.org/TR/NOTE-XMLsubmit.html

[7] USMARC. http://www.loc.gov/marc/marc.html

[8]Text Encoding Initiative Header. ftp://info.ox.ac.uk/pub/ota/TEI/doc/teij31.sgml

[9]ICPSR SGML Codebook. http://www.lib.umich.edu/codebook.html

[10]CIMI:Computer Interchange of Museum Information. http://www.cni.org/CIMI/www/framework.html

[11]CDWA:Categories for the Description of Works of Arts. http://www.ahip.getty.edu/gii/cdwa/

[12]RLG REACH Element Set for Shared Description of Museum Objects. http://www.rlg.org/reach/reach.html

[13]Visual Resources Association Core Categories for Visual Resources. http://www.oberlin.ed u/~art/vra/dsc.html

[14]Standard Music Description Language

[15]Government Information Locator Service. http://ww.usgs.gov/public/gils/

[16]赵志荣、张晓林。GILS：结构、元数据、应用。情报理论与实践，2000（待发）

[17]Federal Geographic Data Committee/Content Standards for Digital Geospatial Metadata. http://www.fgdc.gov/Metadata/ metahome.html

[18]Making of America II White Paper, Part III, Structural and Administrative Metadata. http://sunsite.berkeley.edu/MOA2

[19]California Digital Library Digital Image Collection Standards. http://www.cdlib.edu/standard/

[20]Open Archives Initiative. http://www.openarchives.org/

[21]Bearman, David. NISO/CLIR/RLG Technical Metadata for Images

[22]Encoding Archival Description. ftp://ftp.loc.gov/pub/ead/

[23]RFC 1807：A Format for Bibliographic Records. http://www.cis.ohio-state.edu/htbin/rfc/rfc1807.html

[24]Internet Anonymous FTP Archives Templates.

[25]MPEG-7 Home Page. http://www.darmstadt.gmd.de/mobile/MPEG7/index.html

[26]OhioLink. Standards for Multimedia Metadata: Art & Architecture. http://www.ohiolink.edu/

[27]Extensible Markup Languange, World Wide Web Consortium Recommendation. http://www.w3.org/TR/REC-xml/

[28]ISO 8879. Information Processing--Text and Office Systems--Standard Generalized Markup Language. http://www.sil.
org/sgml/sgml.html

[29]Xschema: Representation of XML DTDs as XML Documents. http://www.simonst1.com/xschema/

[30]XMI: XML Medatada Interchange. http://www.omg.org/cgi-bin/doc?ad/99-10-02

[31]XSL: eXtensible Stylesheet Languange. http://www.w3.org/TR/xsl/

[32]Michael Day. Metadata Mapping between Metadata Formats.
http://www.ukoln.ac.uk/metadata/interoperability/

[33]Resource Description Framework (RDF) Model and Syntax Specification.
http://www.w3.org/TR/REC-rdf-syntax

[34]Resource Description Framework (RDF) Schema Specification. http://www.w3.org/TR/PR-rdf-schema

[35]Payette, Sandra and Lagoze, Carl. Interoperability for Digital Object and Repositories: Cornell/CNRI Experiments. D-Lib Magazine, May 1999

[36]ANSI/NISO. Z39.50-1995. Information Retrieval: Application Service Definition and Protocol Specification. http://lcweb.loc.gov/z3950/agency/1995doce.html及 ftp://ftp.loc.gov/pub/z3950

[37]Imesh: International Collaboration on Internet Subject Gateways. http://www.ilrt.bris.ac.uk/discovery/imesh/

本文最初发表在《现代图书情报技术》2001 年第 2 期第 9-11 页，续 15 页。