

学习对象元数据的分层描述

沈中南, 史元春

(清华大学计算机科学与技术系 北京 100084)

szn00@mails.tsinghua.edu.cn

[摘要]

随着学习对象的增加, 对学习对象的管理、查找、评价和共享需要通过元数据来实现。因此, 学习对象元数据的标准化成为一个十分迫切的问题。LOM (Learning Object Metadata) 规范和“教育资源建设技术规范”就是在这种条件下产生的。本文对 LOM 规范和“教育资源建设技术规范”作了简要的介绍。重点阐述了由这两个规范所产生的元数据的分层描述, 并对每一层都进行了详细的分析。同时指出了这种分层结构产生的原因。

[关键字]

学习对象元数据, 互操作性, 扩展, 教育资源

[abstract]

With the rapid growth of learning objects, the information or metadata plays an important role in managing, searching, evaluating and sharing learning objects. So the standardization of learning object metadata becomes an impending problem. LOM (Learning Object Metadata) Specification and “Educational Resources Construction Technology Specification” are produced under such situation. In this paper, we give a brief introduction to these two specifications and focus on the multilevel descriptions of metadata for learning objects. The paper analyzes every layer in detail and illustrates the cause for the layered structure.

[keywords]

Learning Object Metadata, Interoperability, Extension, Educational resources

1. 简介

学习对象元数据是关于学习对象的信息, 这儿学习对象是指用于教学和培训的任何学习资源。随着网络的发展和教学需求的发展, 学习资源在飞速地增加, 但随之带来的问题是难以共享学习资源和发现有效的学习资源。学习对象元数据的作用就在于为学习者或教育者对学习对象的查找、评估、获取和使用等提供支持, 同时也支持学习对象的共享和互换。它通过描述学习对象的属性, 使得用户能够获得学习对象的相关信息而不用直接操作学习对象。基于这些属性, 用户可以对学习对象进行分类管理、浏览查找和共享互换等等。

为了达到互操作的目的, 学习对象元数据的格式应该是一致的, 也就是他们应该有相同的语义模型和绑定方法。较早且影响较大的学习资源的元数据规范是 Dublin Core[6], 它规定了元数据属性的一个最小集合, 用于描述类似文档一类的学习对象。现有的很多关于元数据的规范或多或少的受到 Dublin Core 的影响[7], 有些直接是对 Dublin Core 的扩充, 例如 GEM[8]项目 (Gateway to Educational Materials Project)。GEM 扩展了 Dublin Core 的数据元素, 用于支持对文档以外的学习资源的描述, 如学习计划, 课程等。最近几年, IMS (Instructional Management System) 也致力于开发学习资源的元数据规范, 于 2000 年 6 月形成了规范的最终稿[9]。

1996, 一个由学术、政府和企业代表组成的组织成立, 称为教育技术标准化委员会 (Learning Technology Standards Committee), 简称 LTSC, 并受到 IEEE (Institute of Electrical and Electronic Engineers) 的资助, 标号为 IEEE P1484[4]。委员会中 1484.12 小组负责学习

对象元数据规范的制定工作。该小组对现有的规范/标准进行了总结，形成了学习对象元数据规范草案（Draft Standard for Learning Object Metadata），简称 LOM[2]。LOM 正受到越来越广泛的重视，有望成为下一代学习资源元数据的标准。

我国从 1999 年开始着手远程教育技术的标准化工作，并与 2000 年成立了现代远程教育技术标准化委员会（Distance Learning Technology Standards Committee），简称 DLTSC[5]。DLTSC 在 LOM 的基础上，进行了一系列的本地化工作，包括数据元素属性的确定，取值范围的修改等等，形成了我国的学习对象元数据规范草案（也简称为 LOM）[1]。同时 DLTSC 还对 LOM 进行了进一步的扩展和细化，起草了“教育资源建设技术规范草案”[3]。本文以下部分将对 LOM（在不引起混淆的情况下，均指我国的学习对象元数据规范草案）和“教育资源建设技术规范草案”作简要介绍，重点阐述它们对学习对象元数据分层描述的支持。

2. 学习对象元数据（LOM）

2.1 LOM 概述

LOM 规范为学习对象的属性描述提供了一个语义模型，用于支持信息的交换。但它并不涉及任何的实现问题，包括编码绑定、API 绑定、用户界面以及学习对象如何使用等等。这样做的目的在于：

- 语义模型与绑定的分离：保证了语义模型的独立性，而同时又没有对绑定问题（和实现有关）做出任何的限制。一个语义模型可以对应多种绑定方法，如 XML 编码绑定、HTTP 元绑定等等，用户可以根据自身的需求任意选取。
- 语义的互操作性：通过建立一个通用的数据模型，保证了元数据的不同绑定有较好的语义互操作性。同时，不同绑定之间转换也会变得简单明了（这儿需要注意的是：DLTSC 也将对绑定问题制定出相应的规范，遵循绑定规范的实现将会获得最好的数据互操作性）。

LOM 规范所建立的数据模型是学习对象属性的一个最小集合，它只对学习对象的一些公用属性进行描述。而且对于某些数据元素，LOM 规范直接给出了推荐的取值集合（称为词汇表），以获得更好的数据互换性。LOM 规范的主要特点如下：

- 学习对象的属性被组合成 9 个类别，每个子类相互比较独立，并且有明显的语义含义。在具体实现时可以有选择的支持不同的子类。
- LOM 规范支持多种语言，这种对多语言的支持体现在两个方面：学习对象可以使用多种语言和学习对象元数据本身可以用多种语言来描述。但不可否认的是，这大大增加了 LOM 规范的复杂度，也增加了实现的复杂性。
- LOM 规范对扩展的支持，这也体现在两个方面：词汇表的扩展和结构上的扩展（通过第 9 个子类实现）。扩展问题将在以下部分详细论述。

2.2 LOM 的结构

如上所述，LOM 的数据元素被分成 9 个类别[1][2]，分别是通用类、生存期类、元元数据类、技术类、教育类、权利类、关系类、评注类和分类类。LOM 的结构可以看作是树状的，只有叶节点才直接有值，如图 1 所示。

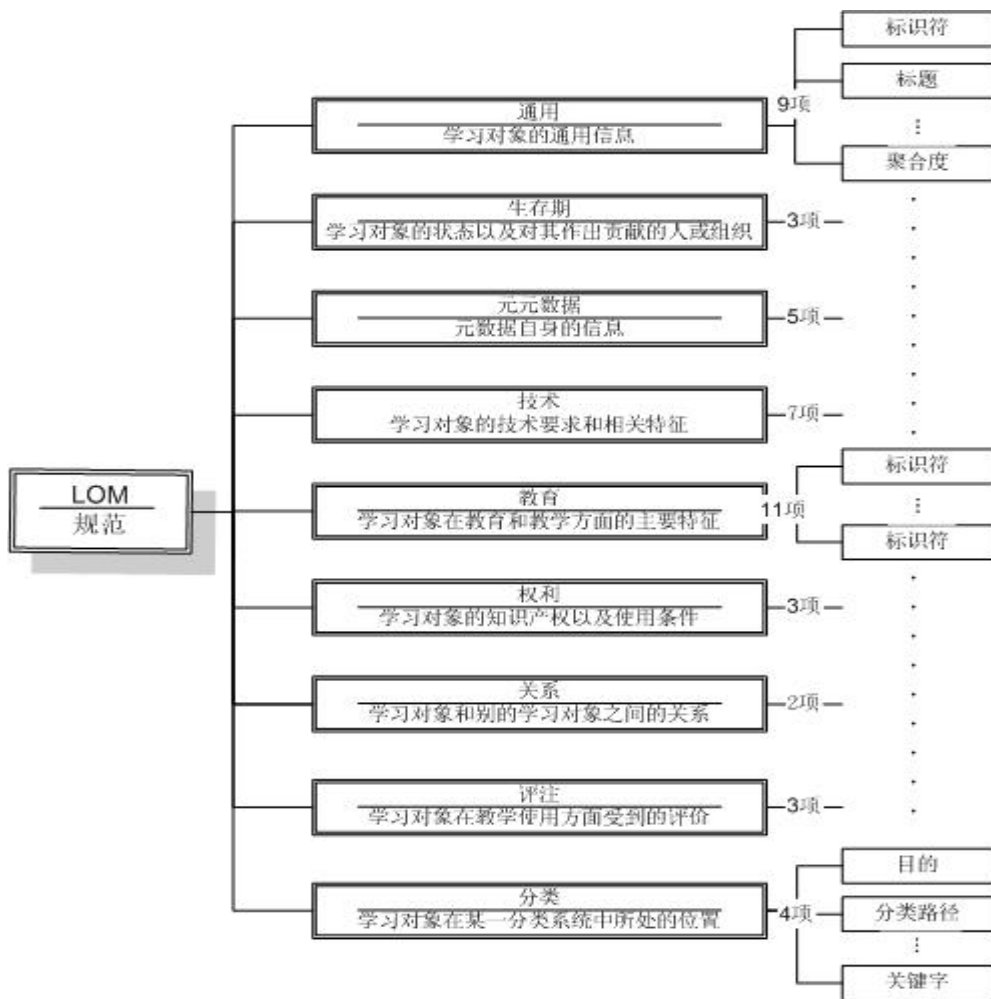


图1 LOM数据模型的结构

3. 教育资源建设技术规范

“教育资源建设技术规范” [3]包括四部分的内容，分别为：教育资源开发的技术要求；教育资源的属性标注；教育资源的评价标准；教育资源管理系统的体系结构和基本功能。在该规范中，我们所关心的是第二部分，它把教学资源大体上分为 8 类，对于每一类资源，在 LOM 规范的基础上添加了一些该类资源特有的扩展属性，如图 2 所示。

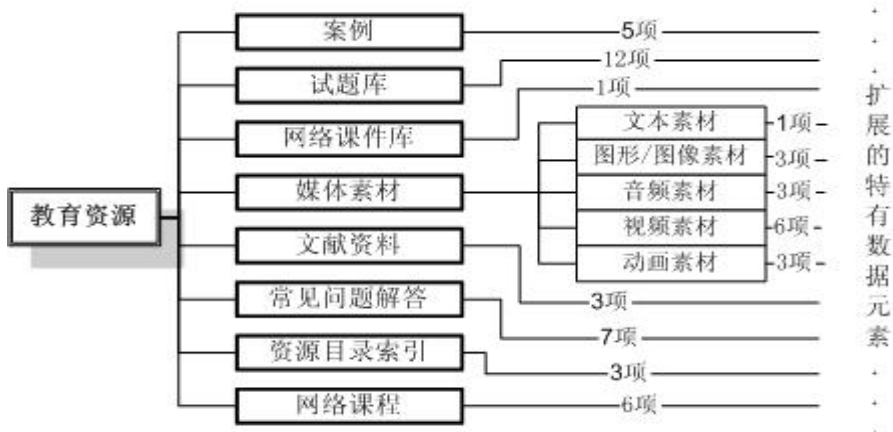


图2 基于资源的分类及扩展属性

4. 元数据的分层描述

结合 LOM 规范和“教育资源建设技术规范”，可以看到，对于一个学习资源的属性可以分层进行描述。在 LOM 规范中，所有的数据元素被分为两类：必需数据元素和可选数据元素。所有的必需数据元素组成了学习资源属性的核心集。所有的可选数据元素组成了学习资源的可选集。“教育资源建设技术规范”中特定资源的扩展数据元素组成了学习资源的特定扩展集。除此之外用户扩展的数据元素组成了学习资源的任意扩展集。这四个集合组成了学习对象元数据的四层描述，如图 3 所示。

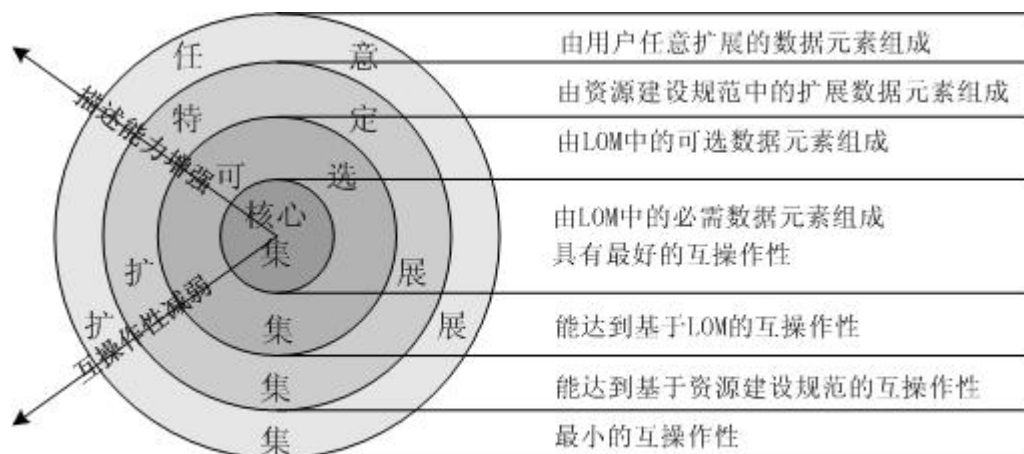


图3 学习资源元数据的四层描述

学习资源元数据四层描述的产生从本质上来讲是元数据的灵活性与互操作性制约平衡的结果。严格遵循标准的元数据实例显然会有最好的互操作性，因为它能被其它任何遵循标准的应用所接受。然而，标准所规定的元数据对用户来说是远远不够的。每个开发商都希望自己的产品有特色，而且某些特定的资源需要特定的元数据来描述，这就需要在标准的基础上对元数据进行扩展，以增强灵活性和描述能力。这些扩展的数据元素显然不具有通用性，很可能被别的应用所拒绝，因此互操作性就很差了。由图 3 可以看到：随着扩展数据元素的增加，随之而来的是描述性的增强和互操作性的减弱。

4.1 核心集

所谓核心集，是指由 LOM 规范中的必需数据元素（共 17 项）组成的具有强制性的集合。核心集中的数据元素具有最大的通用性，一般来说对于任意一个学习对象都是必要的。对于任何一个和 LOM 规范一致的元数据实例或应用都必须支持核心集中的所有数据元素。核心集是学习资源元数据实现互操作的基础。对于元数据实例，核心集中的数据元素是能被任何和 LOM 规范一致的应用所接受的，同样对于任何一个应用，都能识别和 LOM 规范一致的元数据实例核心集中的数据元素。

4.2 可选集

所谓可选集，是指由 LOM 规范中的可选数据元素组成的集合。对于集合中的任意一个数据元素，对于元数据实例或应用来说都是可选的。可选集可以看作是对核心集的扩展。用户在确定核心集以外的数据元素时要尽可能地选用可选集中的数据元素，因为可选集为数据互操作提供了一个可能被采用的数据元素的集合。

学习对象的核心集和可选集构成了 LOM 规范的一致性问题[1][10]。和规范严格一致的

元数据实例或应用必须包含或支持核心集中的所有数据元素，可以包含或支持可选集中的数据元素，不能包含或支持核心集和可选集以外的扩展数据元素。和规范一致的元数据实例或应用必须包含或支持核心集中的所有数据元素，可以包含或支持可选集中的数据元素，也可以包含或支持核心集和可选集以外的扩展数据元素。一致性和互操作性是有密切关系的，一致性程度的增强伴随着互操作性的提高，但却是以牺牲扩展性为代价的。开发商在进行一致性声明时需要进行取舍权衡。

这儿需要着重指出的是 LOM 规范本身提供了一定的扩展机制。这种扩展机制为学习对象的开发商和使用者提供了极大的方便。

● 词汇表的扩充

如前所述，LOM 规范给出了某些数据元素的取值集合，但如果用户并不想采用词汇表中的值，则可以对词汇表进行扩充。数据类型为词汇表的数据元素的值是一个二元组（来源，值）（在以下的说明中，我们省略对语言的描述）。如果用户采用 LOM 推荐的值，则二元组的形式为（“LOMv1.0”，“XXX”）。如果用户使用自己的词汇表，则二元组的形式为（URI，“XXX”），其中 URI 为用户扩展的词汇表的标识符。具体例子见图 4。

● “分类”子类的使用

LOM 的第 9 个类别为数据元素的扩展提供了一个有效的途径。通过它，用户可以任意指定分类系统（无论是正式的，还是用户自定义的），并给出学习对象在该分类系统的位置这实际上是为用户提供一种扩展数据元素的机制。例如，用户可以通过它把学习对象归入某一学科领域（见图 5）。一个好的分类系统对于实现学习对象的管理、浏览和查找是十分有利的。但是要形成一个统一的分类系统却是一件十分困难的事[12]。

5.2 学习资源的类型

9分类

值1 { (en, LOMv1.0), (zh, 练习) }

9.1目的 { (en, LOMv1.0), (zh, 学科) }

值2 { (en, USER_DEFINED), (zh, 案例) }

9.2分类路径

其中，USER_DEFINED定义为 {媒体素材，网络课件，案例，常见问题解答，……}

9.2.1 来源 (zh, 国家标准)

9.2.2 分类单元 { (zh, 物理学) / (zh, 声学) /……}

图4 词汇表扩展的例子

图5 “分类”子类使用的一个例子

4.3 特定扩展集

特定扩展集中的数据元素是由“教育资源建设技术规范”中特定资源的扩展属性组成的。我们可以把它看作是某一领域中的一个“协议”。对于那些遵守“协议”的用户或开发商，他们的元数据实例或应用系统是能够达到互操作的。而对于那些不遵守此“协议”的用户或开发商，他们的扩展数据元素是不具有互换性的。

在确定特定扩展集时需要遵守以下原则：

- 尽可能地使用 LOM 中的数据元素。只有当学习对象的某个属性不能用 LOM 中的数据元素来表示时，才考虑通过添加扩展数据元素来实现。
- 特定扩展集中的数据元素不能损害 LOM 中数据元素的语义。
- 特定扩展集的数据元素在特定领域中要有一定的通用性。

4.4 任意扩展集

任意扩展集是用户或开发商根据自身的需要所确定的扩展数据元素的集合。确定特定扩展集的前两条原则同样适用于确定任意扩展集。对于那些能用特定扩展集描述的属性，也尽量采用特定扩展集中的数据元素，而不要随意地添加新的扩展数据元素。很明显，任意扩展集的互操作性很差，只有同一开发商的产品才能达到互操作。当然，开发商或用户之间可以

通过某种“双边协定”来提高数据互换的能力。

5. 小结

学习对象元数据在学习对象的管理、浏览、查找和共享等方面起着重要的作用。随着学习对象的增加，确定学习对象元数据的统一规范显得尤为重要。LOM 规范和“教育资源建设技术规范”的目的就在于给出学习对象的标准化属性描述，尽可能地达到资源共享。结合这两个规范，形成了学习对象元数据的四层描述结构：核心集，可选集，特定扩展集和任意扩展集，这种分层的现象是数据互操作性和描述灵活性之间相互平衡的结果。也就是说通过分层结构，提供互操作性的多个层次，尽可能地满足不同应用的需要。用户或开发商可以根据自身的需求有选择地采用。

6. 参考文献

- [1] “DLTS-3.1 学习对象元数据规范”，DLTSC，2001 年 12 月。
- [2] “Draft Standard for Learning Object Metadata”，IEEE P1484.12/D6.1，18 April 2001，http://ltsc.ieee.org/doc/wg12/LOM_WD6-1_1.pdf。
- [3] “教育资源建设技术规范（征求意见稿）”，DLTSC，2001 年 12 月。
- [4] Learning Technology Standards Committee，IEEE P1484，online: <http://ltsc.ieee.org>。
- [5] Distance Learning Technology Standards Committee，online: <http://www.dlts.moe.edu.cn/cltsc>。
- [6] “Dublin Core Metadata Element Set, Version 1.0: Reference Description”，<http://www.dublincore.org/documents/>。
- [7] L. Jay Wantz, Michael Miller, “Educational metadata in use”, Computer Networks and ISDN Systems, Volume: 30, Issue: 1-7, April, 1998, pp. 701-703。
- [8] Gateway to Educational Materials (GEM), <http://www.geminfo.org/>。
- [9] Thor Anderson, Tom Wason, “IMS Learning Resource Meta-data Information Model”, final specification, version1.1, <http://www.imsproject.org/metadata/mdinfov1p1.html>。
- [10] “DLTS-4.1 数据扩展技术的指导规范”，DLTSC，2001年12月。
- [11] Mayumi Okamoto, Yasuhiro Okui, “Investigation of Learning Object Metadata and Application to a Search Engine for K-12 Schools in Japan”, Advanced Learning Technologies, 2001. Proceedings。
- [12] Daniel D. Suthers, “Evaluating the Learning Object Metadata for K-12 Educational Resources”, Advanced Learning Technologies, 2001. Proceedings。