



具有著录层次的元数据 编目员和元数据馆员的新挑战

Myung-Ja Han

副教授/元数据馆员
美国伊利诺伊大学香槟分校
E-mail: mhan3[at]illinois.edu

中文翻译: 林宗勇 (宁波诺丁汉大学图书馆)
Chinese Translator: LIN Zongyong
(The University of Nottingham Ningbo China)

Meeting:

80 — *Inspired moments in cataloguing* — Cataloguing

摘要

随着图书馆收藏并提供越来越多的数字资源,新的粒度单位由此在资源检索和描述中出现。数字化资源和原生数字资源使更多检索粒度层次的产生成为可能,例如,一本书或一种期刊不仅可以在整体层次检索,也可以在篇章和书页层次上进行检索。实现这些粒度层次的检索需要建立相应粒度层次的元数据。本文探讨了粒度层次检索和描述的新兴要求如何使元数据创建与编目成为一项需要高度合作的工作,并且提出了设计和创建粒度层次资源描述的元数据方案的一种方法。

背景

图书馆在系统地组织和管理馆藏资源方面有着悠久的历史,换言之,编目是为了用于提供资源检索。编目从索引卡片(3x5英寸的卡片上有单册的主要条目、附加条目、索书号和分类信息)发展到如今图书馆使用的各种不同的数据标准,如元数据标准,其中包括机读编目格式标准(MARC)和为馆藏资源创建描述性元数据的内容标准。

自 19 世纪 60 年代中期首次使用以来，MARC 成为英美编目条例（AACR）作为内容标准的图书馆描述文献的主要元数据标准。由于 MARC 具有丰富的语义组合，它一直以来被用于描述纸质图书、连续出版物、地图和微缩文件，并且毫无问题。然而，随着图书馆开始收藏越来越多其它格式的资源，特别是音频/视频资源和电子/数字资源，图书馆对其它能够更好描述非纸质资源的元数据标准的需求日渐增长，原因是有些资源通过其它系统检索比通过图书馆集成管理系统（ILS）更容易，而且用其它元数据标准能更好地描述其技术细节。随之，元数据标准，如都柏林核心元数据标准（Dublin Core）、元数据对象描述标准（MODS）、档案编码描述格式（EAD）以及可视资源学会（VRA）核心元数据标准，被用来描述数字馆藏，而图书馆也开始开发本地检索系统。根据美国高校与研究型图书馆学会（ACRL）的调查显示，大多数高校与研究型图书馆同时以 MARC 和其它元数据标准描述数字资源，并且使用 ILS 之外的系统来仓储资源（Ma 2007），这就需要具有描述具体格式信息的元素或者描述原始资源和重新格式资源之间关系的元素。比如，许多图书馆开始数字化它们的特殊馆藏和档案，尤其是图片和手稿收藏，Dublin Core、EAD 或者 MODS 被用于著录馆藏层次的查找辅助，每个数字化资源的单件层次的元数据则单独被创建。另外，由于大多数图书馆使用的 ILS 仅为 MARC 格式的记录而设计，因此图书馆一直以来都使用和开发独立的数字资源管理工具，其能在 MARC 之外的元数据标准上良好运行。

Jenn Riley（2010）编写的《眼见标准：元数据世界的可视化》囊括了 100 多个当前可用的元数据标准和内容标准。该列表不仅收录了在图书馆领域使用的元数据标准，还包括了其它领域的标准，比如，文化遗产机构，如博物馆和档案馆。它也涵盖了用于描述功能以外用途的元数据标准。决定使用何种元数据标准和内容标准的因素有很多，其中包括资源特征、元数据和数字资源的存储系统，资源应用的领域以及元数据的功能。除此之外，可用信息和编目员的学科知识也是需要考虑的因素，因为这直接影响到元数据的内容。虽然已经有很多种具体的元数据标准可供选择，但仍然有必要发展出其它元数据标准服务于更有针对性、更具粒度层次的描述，从而提供粒度层次的检索服务。

粒度层次的描述是为了粒度层次的检索服务

“实际上，任何我们提供给客户群的数字化资源都需要元数据来发现和获取（Tennant 2002, p. 32）。”元数据使我们所有的资源都可以以最优化的方式被用户所检索和获取。对创建高质量和具有一致性的元数据的积极讨论是图书馆尽力提供最佳的检索服务的表现。图书馆提供检索服务水平的高低取决于元数据所包含的信息。

直到最近，图书馆提供的默认检索层次是一本书或是一种期刊——在某些情况下是期刊的单册，换言之，“在藏有纸质书的实体图书馆中，在架册数是粒度的主要单位”（Cole, Han, and Vannoy 2012）。同样的，元数据对资源描述的默认层次，也在一本图书或一种期刊的层次。尽管如此，随着图书馆所购买的专业论文和电子资源数据库数量的增加，用户现在可以检索到资源的更多粒度层次，如章节层次和文章层次，并且期待图书馆为本馆数字化的资源提供馆藏、图书或期刊层次之上更多粒度层次的检索。为了提供粒度层次的检索服务，图书馆必须创建新的元数据标准，以容纳用户希望检索到的粒度层次的描述。

元数据粒度的研究主要集中在两个方面：元数据的元素粒度与参数粒度。Roy Tennant（2004）研究了 MARC 和文本编码计划（TEI）中姓名的元素粒度。他指出，MARC 并不具有在 100 字段区分作者姓与名的语义，TEI 在各种文献中提供粒度元素，姓名信息可与〈姓名〉或者〈作者〉元素著入 TEI 元数据的其它部分中。然而，姓和名的子元素和属性不能帮助机器以更细地方式来解析姓名部分。对比 MARC 与 TEI，MODS 中的最高层次元素〈姓名〉具有子元素〈姓名部分〉，它能用〈类型〉属性将姓与名区分开来。Dublin Core 是除 MARC 之外最常用的元数据标准，它也没有〈创作者〉元素的子元素和属性（Ma 2007）。表一显示了四种元数据标准用于描述资源中的姓名元素的不同粒度。

元数据标准	描述〈姓名〉
MARC ^[1]	100 1 _ \$a <i>Last name, First name.</i> \$d 1111-1222, \$e role.
TEI	<name type="person"> <i>First name and Last name</i> </name> or, <author> <i>Last name, First name.</i> </docAuthor>
MODS ^[2]	<name type="personal"> <namePart type="given"> <i>Last name</i> </namePart> <namePart type="family"> <i>First name</i> </namePart> <role> <roleTerm type="code" authority="marcrelator"> <i>aut</i> </roleTerm> <roleTerm type="text" authority="marcrelator"> <i>author</i> </roleTerm> </role> </name>
Dublin Core	<dc:creator> <i>Last name, First name. Date.</i> </dc:creator>

表一 用 MARC、TEI、Dublin Core 与 MODS 标准描述〈姓名〉信息

元素的扩展性是区分这些差异的方法之一。根据元数据标准，该方案使用户能扩展元素集以适应本地需求，即 Dublin Core 鼓励用户根据 2011 年三月发布的《状态报告》^[3]扩展本地元素。

如何为元数据提供信息是元数据粒度的另外一个话题，特别是所描述资源的主题。美国国会图书馆推荐在著录主题词时使用“百分之二十原则”，即只为占作品内容的至少 20% 以上的话题分配主题标目（美国国会图书馆 2008）。美国国会图书馆还提出了“三原则”和“四原则”。“三原则”规定，“如果一个普通主题包含三个以上分主题，但是作品中只涉及两到三个分主题，那就分配两到三个合适的主题，而不使用比较笼统的主题。”该原则还规定，“如果作品中涉及三个以上的分主题，那就分配一个比较笼统的主题，除非‘四原则’适用。“四原则”提出，分主题的数量不能超过四个，因为这会导致主题标目的缺失，主题

标目的作用是帮助用户发现研究和教学中的重要资源（美国国会图书馆，2008）。当粒度单位是一本书或一种期刊时，这些规则是非常适用的，描述性元数据就是为这个粒度单位而专门设计的，它几乎不可能涵盖只描述资源一小部分的主题标目。然而在数字图书馆环境下，需要有更多的著录粒度层次来提高资源的检索与发现能力。根据《CONTENTdm 元数据最佳实践》（2009），数字资源的描述，尤其针对手稿和藏书，美国国会图书馆修订了“百分之二十原则”以促进专门主题资源的发现。为了使数字资源更便于用户搜索与获取，图书馆在元数据中著入更加具体的主体标目，虽然它们只符合所描述作品的很小一部分（第 27 页）。然而，对增加元数据著录层次的需求却源于他因——用户的需求和数字人文项目的发展。

新粒度单位

随着图书馆采购越来越多的数字资源，积极建设数字馆藏以及进行大型图书数字化项目，新粒度单位随之出现。传统上以书架上的单册作为粒度单位已不再是唯一的粒度表现形式了。如今，图书馆能够提供不同粒度单位资源的检索，例如，数字馆藏中的单件资源、期刊内的文章、书本中的章节或是一张目录表，这些都成为了新的粒度单位。新的粒度单位同时也是用户对图书馆资源发现服务的期许。Yang and Hoffman（2011）在对下一代编目与检索系统的可用性测试的调查中发现，用户喜欢新的资源检索系统中提供文章与章节层次的检索以及其它服务，如分面导航服务和联合搜索功能。

随着人们对数字人文的兴趣与日俱增以及数字人文的发展，新的粒度单位与之出现。尽管大规模图书数字化项目使电子书在网络环境中以书页的方式进行阅读成为可能，但是电子书作为纸本书的替代品，人文学科的学者们希望在用于研究与课堂教学的专门图书馆藏中进行比书本层次更细的检索，比如文艺复兴时期的寓意画册以及早期剧本。他们希望数字化资源能以新的粒度单位保存。显然，这些新的粒度单位需要学者们的结构和语境分析，它们也必须具有描述性元数据的适当层次。为满足学者们的需求和用户期待，图书馆需找到建立新的元数据粒度的可持续方式，它可兼容新粒度单位的著录，因而，这需要元数据和编目馆员当新

需求出现时开发新的元数据元素或方案。其中主要在于如何与学者与用户合作，为新的结构与内容设计新的元数据标准。

设计与拓展元数据方案

著录的粒度层次需要新的“容器”——新的元素集。新元素可以加入到已经创建的元数据标准中，这些标准包含著录粒度层次所需的最常见的元素并且允许元数据的扩展，如 Dublin Core，或者还可以通过与来自不同元数据标准的语义兼容和匹配创建一个新的元数据方案，跟建立应用档案的方法类似（Heery and Patel 2000）。如果以上方法不能实现，也可以建立一个新的元数据标准来满足新语义和句法的具体要求。这两种方法中，建立新方案的方法更受欢迎，因为它能长期地保证元数据质量的一致性，并且提高元数据与更广泛用户组的互操作性和可共享性，新方案定义了元数据记录中设置的单元、属性、实体和编码方案以及它们之间的关系。另外，新方案可以自由组合使用且能更好地明确对元素中某些类型数据的限制，这样才能得到有结构性且有效的元数据记录。

大规模图书数字化项目产生了许多 16-18 世纪出版的寓意画册的数字版本，但直到现在寓意画学者仍然不能直接数字访问，除非他们亲自到访保存这些纸本资源的机构。SPINE 是为通过兼容与匹配其它元数据方案来描述寓意画册以及书中寓意画而创建的新的元数据方案。由于书中的寓意画是书的主要内容，寓意画包括三个不同的部分——标题（motto）、图片（image）和相关文字（subscriptio），寓意画学者们希望具有寓意画层次的描述，这样就可以保证寓意画层次的检索和画册层次的检索。寓意画的著录元素及其相关内容在 Stephen Rawles（2004）的研究论文《寓意画等相关电子资源的信息标目概要》中有所定义。SPINE 的可扩展标记语言方案（XML schema）于 2007 年创建，成为此方案的 1.0 版本，其中包括寓意画特定元素和借用自 TEI 的用于书本层次描述的语义（TEI Header）以及书本文本（其它 TEI 元素，如 <p> 代表段落）。^[4] SPINE 方案已经根据寓意画学者的需求和关联数据技术进行了更新，现在已是 1.2 版本（Stacker 2012）。

表二展现了 SPINE 方案是如何使用不同元数据标准的不同名称域来实现粒度层次的描述的。

书本层次描述	电子版本描述	寓意画层次描述
MODS TEI Header (*Both are transformed from MARC format bibliographic records)	SPINE of Information <copyDesc> Sub elements: <copyID> <owner> <digDesc> Attributes: countryCode, comp, scope, xml:id, globalID, pageImages	SPINE of Information <emblem> Sub elements: <motto> <transcription> <pictura> <iconclass> SKOS/RDF <skos:notation> <skos:prefLabel> <subscription> <commentatio> <keyword> ... Attributes: xmlns:xlink, xml:id, citeNo, xlink:href, globalID, xml:lang, rdf:about

表二 SPINE 方案使用的元素和元数据方案

以上已经提到过，SPINE 方案从其它已建立的方案中借用语义，也就是说，它支持书本层次的描述，不管用 MODS 还是 TEI Header，或者同时使用两者，它还在 SPINE of Information 中的元素上使用资源描述框架（RDF）和简单知识组织系统（SKOS）中的语义。因为它使用的是已经建立并被广泛使用的元数据方案，所以任何对寓意图册进行数字化的机构都能使用 SPINE 方案来做书本层次的描述，尽管寓意画层次的描述还没有实现，但这样至少能在任何寓意画册专门门户上进行书本层次的检索。

```

<biblioDesc>
<mods>
<mods:titleInfo>
<mods:title>XL [i.e. Quadraginta] emblemata miscella nova</mods:title>
</mods:titleInfo>
<mods:physicalDescription>
<mods:digitalOrigin>reformatted digital</mods:digitalOrigin>
<mods:form authority="marcform">print</mods:form>
<mods:extent>[8], xxxx p. : 41 ill. ; 20 cm.</mods:extent>
</mods:physicalDescription>
...
</mods>
<copyDesc>
<copyID>uiu2895515</copyID>
<owner countryCode="US">University of Illinois</owner>
<digDesc comp="complete" scope="all" xml:id="xliequadragintae00mure"
globalID=http://hdl.handle.net/10111/UIUCOCA:xliequadragintae00mure>
<copyID>10111/UIUCOCA:xliequadragintae00mure</copyID>
<owner countryCode="US">University of Illinois</owner>
</digDesc>
...
</copyDesc>
<emblem xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#
xmlns:skos=http://www.w3.org/2004/02/skos/core# xml:id="E000944" citeNo="I."
globalID="http://hdl.handle.net/10111/EmblemRegistry:E000944">
<motto><transcription xml:lang="de">
Alchimesterey:<normalisation>xml:lang="de">Alchemie:</normalisation></transcription>
</motto>
<pictura xml:id="E000944_P1">
<iconclass rdf:about="http://www.iconclass.org/rkd/31A247">
<skos:notation>31A247</skos:notation>
<skos:prefLabel>looking over the shoulder</skos:prefLabel>
</iconclass>
...
</picture>
</emblem>
</biblioDesc>

```

表三：SPINE 元数据记录中对《*XL [i.e. Quadraginta] Emblemata Miscella Nova.*》画册以及书中寓意画的粒度层次描述

另外，因为 MARC 对应 TEI Header 以及 MARC 对应 MODS 的映射表和转换样式表已经存在，所以用 SPINE 元数据记录进行书本层次的描述相对容易，这就

是在创建新方案时应用已建立的元数据方案的好处。表三是具有不同粒度层次的 SPINE 元数据记录描述纸本图书、数字版本和书中寓意画的样本。

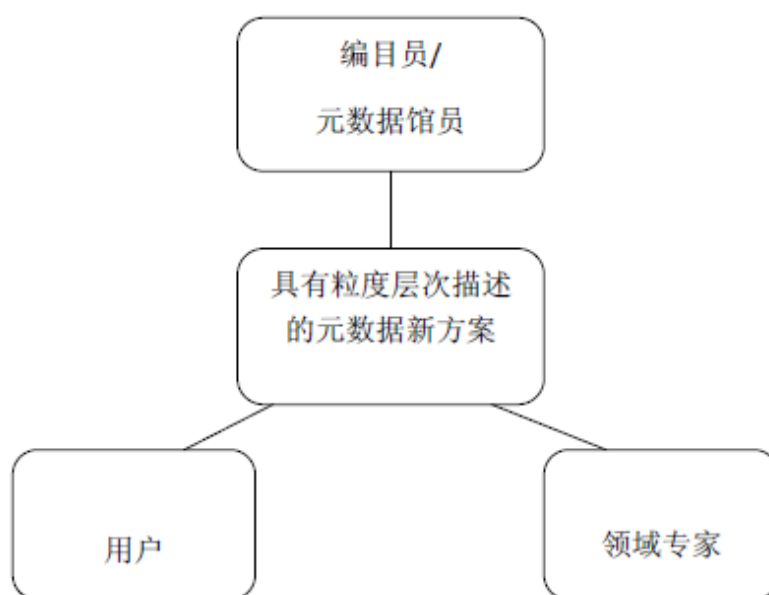
应用关联数据技术，如 RDF、SKOS 和名称域，通过使用名为 Iconclass 的受控词表，可以提供另一层面的粒度检索服务，它被用于描述寓意画和其它图片。

Iconclass 词表具有等级结构，它的词汇现有四种语言表示。使用 RDF 和 SKOS，寓意画学者能看到不同语言的 Iconclass 词汇并且查看词汇间的等级关系。RDF 和 SKOS 也是提高元数据粒度层次的工具。之前谈到过，因 Iconclass 词表可以多种语言表示，而且 RDF 包含信息，所以除了最初创建的原始语言，元数据现在还具有不同语言名称以及缩小或扩大的名称。

SPINE 方案发展的例子说明了粒度层次描述的元数据新方案可以结合 XML 来开发，这已成为编目员和元数据馆员所需掌握的必备技能（Han and Hswe 2010）。在设计元数据的数据模型方面，SPINE 的例子清楚地说明了，元数据的设计和创建工作需要图书馆员、用户（学者与学生）以及领域专家（或者学科专家）的高度配合。

同心协力设计与创建元数据

设计粒度层次描述的元数据方案需要编目员/元数据馆员、用户和领域专家的配合。对内容、结构和用户行为的分析在设计新的元数据方案时是必需的，因为它决定了粒度的层次。领域专家提供学科知识中的高质量元数据。编目员/元数据馆员可以为建立知识结构出力，用户能告知发现资源的信息使用方式以及符合他们期望的粒度层次。领域专家能设计符合具体学科知识要求的元数据。如下图所示，



图一 创建具有粒度层次描述的元数据方案

展望

元数据是设计图书馆检索服务的重要因素。基于粒度层次的元数据描述，才能确定粒度层次的检索服务，同时元数据的设计应最大程度地满足学者研究与教学的需求。随着图书馆收藏与产出越多的数字格式的资源，图书馆提供的粒度层次的检索服务因用户期待和需求的提高已发生改变。学者们希望数字化资源能以专门的方式储存和描述。根据资源的内容与结构，以及为了满足用户的需求，图书馆应开发出新的元数据方案以容纳新的粒度层次。图书馆可以通过同现有的方案兼容或匹配建立新的元数据方案，也可以用 XML 中全新的语义和句法创建新方案。然而更重要的是，新元数据方案的设计与发展和具有粒度层次描述的元数据的创建已经成为用户、领域专家和图书馆员三方需高度合作的事业，每一方都要在他们相应的专业领域里贡献自己的才能。随着图书馆在数字图书馆环境下提高对用户的服 务，编目员和元数据馆员需要掌握使用多种元数据标准，开发新的更多粒度层次描述的元数据方案用于保存数字资源，同时也要求编目员和元数据馆员不仅掌握信息技术，而且需要与学者和领域专家们合作。

参考文献

- Cole, T. W., Han, M-J, & Vannoy, J. (2012) . Descriptive Metadata, Iconclass, and Digitized Emblem Literature. *Proceedings of Joint Congress of Digital Libraries 2012*.
- CONTENTdm Metadata Best Practices. (2009) . Online. Available at http://contentdmmwg.wikispaces.com/file/view/Best_Practices_2+0.pdf
- Han, M. J., & Hswe, P. (2010) . The Evolving Role of the Metadata Librarian: Competencies Found in Job Descriptions from 2000 to 2008. *Library Resources and Technical Services*, 54/3, 129-141.
- Library of Congress. (2008) . *Subject Headings Manual* (2008 ed.) . Washington, D.C: Library of Congress, Cataloging Policy and Support Office.
- Ma, J., & Association of Research Libraries. (2007) . *Metadata*. SPEC kit. Washington, D.C: Association of Research Libraries.
- Heery, R., & Patel, M. (2000) . Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne*, Issue 25. Online. Available at <http://www.ariadne.ac.uk/issue25/app-profiles/>
- Rawles, S. (2004) *A Spine of Information Headings for Emblem-Related Electronic Resources*. In: Wade, M.R. (ed.) *Digital Collections and the Management of Knowledge: Renaissance Emblem Literature as a Case Study for the Digitization of Rare Texts and Images*. DigiCULT, Salzburg, Austria.
- Riley, J. (2010) . *Seeing Standards: A Visualization of the Metadata Landscape*. Available from <http://www.dlib.indiana.edu/~jenlrile/metadatamap/>
- Tennant, R. (2002) . The Importance of Being Granular. *Library Journal*, 127 (9) , 32.
- Tennant, R. (2004) . A Bibliographic Metadata Infrastructure for the Twenty-First Century. *Library Hi Tech*, 22 (2) , 175-181.