



项目编号	INFO-115-C01
文档编号	TR-REC-013

中国科学院数据应用环境建设与服务

元数据参考模型

（征求意见稿）

中国科学院数据应用环境建设与服务 项目组

2009 年 9 月

目 录

1	范围.....	1
2	术语与定义.....	1
2.1	资源.....	1
2.2	数据集.....	1
2.3	数据实体.....	1
2.4	数据模型.....	2
2.5	元数据.....	2
2.6	元数据元素.....	2
2.7	元数据实例.....	2
2.8	元数据复合元素.....	2
2.9	元数据子集.....	2
2.10	编码体系.....	2
2.11	词汇编码体系.....	3
2.12	句法编码体系.....	3
2.13	数据类型.....	3
2.14	值域.....	3
2.15	元数据注册系统.....	3
3	元数据参考模型.....	3
4	调研分析.....	4
4.1	需求调研.....	4
4.2	现有方案调研.....	5
4.3	需求分析.....	5
4.4	确保兼容性.....	6
5	方案制定.....	6
5.1	方案制定.....	6
5.1.1	方案制定.....	6
5.1.2	意见征求和修订完善.....	6
5.1.3	制定应用规范.....	7
5.2	结构化定义.....	7
5.3	文档规范.....	9
5.3.1	正文要素.....	10
5.3.2	附录要素.....	11
5.4	元数据注册.....	12
6	系统建设.....	12
6.1	设计开发.....	12
6.2	试用、评估和完善.....	12
6.3	系统发布.....	12
7	应用服务.....	12
7.1	扩展方案.....	13
7.1.1	元数据的定制.....	13
7.1.2	元数据的扩展.....	13

7.1.3	元数据应用方案的设计方法.....	14
7.2	元数据映射规则.....	18
7.2.1	元数据映射.....	18
7.2.2	元数据映射内容.....	18
7.2.3	元数据映射方法.....	19
7.2.4	映射编写规定.....	20
8	一致性测试规则.....	20
8.1	标签检查.....	20
8.2	完整性测试.....	21
8.3	最大出现次数测试.....	21
8.4	标识符测试.....	21
8.5	数据类型测试.....	21
8.6	值域测试.....	22
8.7	模式测试.....	22
8.8	用户自定义的扩展元数据测试规范.....	22
8.8.1	排他性测试.....	22
8.8.2	定义测试.....	22
8.8.3	元数据应用方案测试.....	22
9	评估完善.....	23
9.1	应用评估.....	24
9.2	方案完善.....	24

元数据参考模型

1 范围

本规范面向数据应用环境的数据资源整合，规范化元数据研制和应用的流程模型，主要包括调研分析、方案制定、系统建设和应用服务等方面，特别是规范化了元数据的格式、语义、语法、注册、一致性测试和评估完善等，以实现元数据方案的规范化，以及基于不同元数据方案的数据之间的互相访问、内容交换和整合集成。

本规范适用于中国科学院数据应用环境建设与服务项目中各类元数据的研制和应用。

2 术语与定义

2.1 资源

具有身份标识的任何对象。

例如，数据集、服务、人员或机构等。

[GB/T 20533--2006，定义 3.27]

2.2 数据集

可以识别的数据集合。

[ISO 19115，定义 4.2]

本规范中，数据集指中国科学院数据应用环境建设与服务项目的主题数据库、参考数据库、专题数据库、专业数据库及其各级子库。

注：通过对诸如覆盖范围或者属性类型的限制，数据集在物理上可以是更大数据集中的一个比较小的数据组。

2.3 数据实体

包含数据内容的一个逻辑的或物理的存储单元。

例如，文本文件、关系数据库数据表、电子表格、GIS 图层等。

注：一个数据集可能由一个或多个数据实体组成。

[GB/T 20533--2006，定义 3.6]

2.4 数据模型

以反映信息结构的某种方式对数据组织描述。

[GB/T 18391.1, 定义 3.23]

注：数据模型指明数据的属性、结构和内在关系。

2.5 元数据

关于数据的数据。本规范中，具体指描述数据及其环境的数据。

2.6 元数据元素

元数据的基本单元。

[GB/T 19710--2005, 定义 4.6]

2.7 元数据实例

与某一元数据模式相符的某一具体数据资源的元数据。

[GB/T 20533--2006, 定义 3.8]

2.8 元数据复合元素

一组说明数据相同特性的元数据元素。

[ISO 19115, 定义 4.7]

注 1：可以包含一个或多个元数据复合元素。

2.9 元数据子集

元数据的子集合，由相关的元数据复合元素和元素组成。

[ISO 19115, 定义 4.8]

2.10 编码体系

提供语境信息或句法分析规则，以帮助对元数据元素的值进行解释的受控词表或者格式规范。

[GB/T 20533--2006, 定义 3.14]

注：编码体系包括词汇编码体系和句法编码体系。

2.11 词汇编码体系

作为元数据元素值域的受控词表。

例如，汉语主题词表、世界各国和地区名称代码表等。

[GB/T 20533--2006，定义 3.15]

2.12 句法编码体系

元数据元素的值在格式上遵从的正式表示法。

例如，“2000-01-01”作为日期的标准表示。

[GB/T 20533--2006，定义 3.16]

2.13 数据类型

有效值域和允许对该值域内的值进行的合法操作的规定。

例如，整型、实型、布尔型、字符串、日期以及本标准中自定义的数据类型等。

注 1：数据类型包括基本预定义类型和用户定义的类型。

注 2：数据类型用术语标识，例如整型。

[GB/T 20533--2006，定义 3.18]

2.14 值域

允许值的集合。

[GB/T 18391.1，定义 3.75]

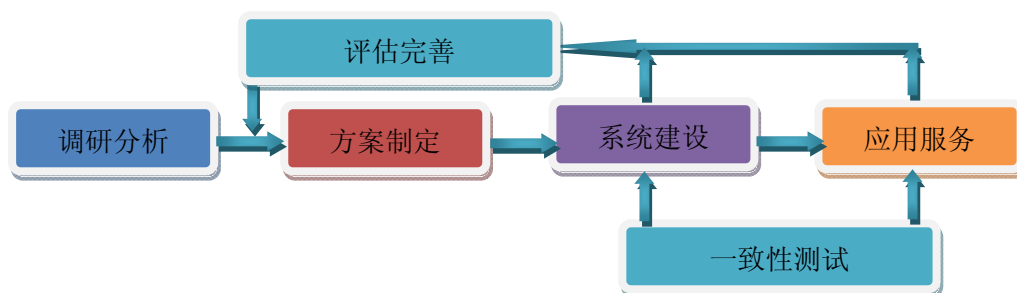
2.15 元数据注册系统

注册元数据的信息系统。

[GB/T 20533--2006，定义 3.12]

3 元数据参考模型

元数据参考模型面向数据应用环境的数据资源整合，立足于元数据研制和应用等全流程的规范化，从调研分析、方案制定、系统建设和应用服务四个关键环节，规范化了元数据的结构、格式、语义、语法、注册等内容，实现了据向数据应用环境中元数据工作的规范化，以及不同元数据方案对数据互相访问、内容交换和整合集成的支持。



元数据参考模型

调研分析是元数据工作的起始点，直接关系到元数据成果的实用性，所以在该阶段主要完成各方需求的掌握和客观分析，为方案制定奠定基础。

方案制定则基于调研分析的结果完成方案的研制，并通过结构化定义、规范化文档实现元数据方案的规范表达，且针对数据交换和集成的需求将元数据方案注册到开放的系统中。

系统建设主要实现元数据管理系统设计开发、实用评估和完善，并最终发布后服务。

应用服务主要包括元数据服务、扩展方案和元数据映射规范，以支持灵活多样的元数据应用需求。

一致性测试则是保证系统建设和应用服务在采用元数据方案方面的一致性所应实现的功能要求，是元数据方案得以贯彻执行的保障。基于元数据方案实现的系统建设和应用服务有可能为非方案研制方承担，所以系统建设和应用服务等环节的执行过程中应通过本一致性测试，以确保符合本规范的要求。评估完善则基于系统建设和应用服务，特别是应用服务的效果，对元数据方案进行评估和完善，提升元数据方案的效用。

4 调研分析

调研分析是元数据工作的起始点，直接关系到元数据成果的可用性和实用性，所以在该阶段主要完成各方需求的掌握和客观分析，为方案制定奠定基础。

调研分析的主要内容是全面掌握需求，以及现有业界成果，并在深度分析二者的基础上给出相应元数据工作的建议方案。

4.1 需求调研

元数据研制人员与内容专家、领域用户进行深入沟通，通过与有关人员的接触、访谈，了解该工作的背景、目的、需求、目标等信息，掌握各方对数据资源元数据的需求，主要

包括：

- 应用背景；
- 是否存在现有标准规范及其基本现状信息；
- 元数据的目的、范围与属性，如描述人、事、时、地、物，或词汇控制等属性；
- 元数据的功能定位，如资源的描述、检索、索引或管理等功能；
- 元数据预计或预期的进展安排；
- 有关人员联系信息。

4.2 现有方案调研

了解元数据需求后，应调研和分析同领域内当前元数据方案的建设情况，并对现有标准及其应用情况进行详细分析，评估其对当前需求的满足程度。在标准分析时，应从适用对象（资源类型、学科）、功能等方面综合考虑，归纳总结出其对当前需求的满足程度与不足。而应用分析则重点在于发现标准规范在用户应用时的发现、总结和反馈，辅助充分了解标准对当前需求的满足情况。同时，在研究中还应了解本学科领域其元数据的应用趋势和议题，以作为未来实施与发展的参考。

此步骤分析的成果，将有助于了解与其它同质或相似标准规范间的差异，取长补短藉以完善需求、目标。

4.3 需求分析

融合领域现有标准规范及其应用的分析研究成果，在元数据制定工作展开前应更深入的对需求进行分析，适当扩大沟通对象范围，与领域专家、用户和系统设计人员进行深入的交互，充分了解各方面的需求，为元数据制定奠定良好基础。

深入分析阶段经过与更大范围有关人员的交互，将获得如下信息：

1. 元数据元素需求，包括元素名称、定义及著录规范；
2. 元数据元素属性，包括数据型态、必填性、多值性等；
3. 元数据元素关系；
4. 元数据元素代码表；
5. 元数据著录范例；
6. 功能需求；
7. 数据查询与呈现需求，包括不同层次的检索（如关键词查询、进阶查询）与呈现（如简要显示、详细显示）；
8. 元数据管理需求，如权限控制。

4.4 确保兼容性

利用之前的研究发现，根据需求所发展的元数据内容应当尽可能的与现有的元数据方案在内容上兼容，支持基于此元数据方案的元数据能够在更大范围与其他元数据实现数据交换和互操作。

5 方案制定

方案制定则基于调研分析的结果完成方案的研制，并通过结构化定义、规范化文档实现元数据方案的规范表达，且针对数据交换和集成的需求将元数据方案注册到开放的系统中。

5.1 方案制定

5.1.1 方案制定

根据“调研分析”的成果，实现元数据方案制定所涉及的全部内容，至少应包括以下内容：

1. 标准制定的背景、规范性内容引用、术语等；
2. 元数据方案完整正文，主要包括：
 - a) 复合元素、元素及其关系等内容的详细描述和图例；
 - b) 元数据复合元素、元素代码表；
 - c) 标准规范字典表；
3. 标准规范之 XML Schema 或 UML 描述；
4. 元数据与参考规范之间的映射关系；
5. 标准规范应用方案制定的约定和方法。

5.1.2 意见征求和修订完善

元数据方案制定完成后，应当再次与领域专家、用户和系统设计人员进行深入的交流，征求上述人员对该标准规范的意见，并努力就内容及其细节等达成共识。

当存在需要对标准进行必须完善的意见时，应根据共识意见把对标准规范进行相应的修订完善，必要时可对标准进行再次的意见征求和修订完善。

5.1.3 制定应用规范

为了推动标准在更大范围内的应用，提高标准的应用质量，特别是控制和保证元数据记录的质量等，应当制定系列应用指导性规范，帮助用户理解和使用标准有关的内容，包括元数据元素定义、著录原则、著录范例，以及元数据服务过程中计划供应用标准的检索元素表与参考数据等。

应用规范应同时兼顾标准管理系统的设计与实施，在元数据管理系统的设计和开发过程中为系统分析人员和设计人员等提供重要参考。

5.2 结构化定义

元数据在定义中应当包含以下几个方面的内容：

(1) 适用范围说明。说明元数据所适用的特定研究对象和适用范围。

(2) 词汇及其约束集合。词汇集合是表达信息对象属性的词汇（即数据元素）及其语义定义的合集，对元数据中的每一个数据复合元素、元素，都应对其名称、定义、数据类型等信息做出明确规定；而约束规则集合则定义元素（复合元素）、结构约束和句法特征的规则，诸如元素（复合元素）的可选性、可重复性、数据元素之间的关系等。常常词汇和约束集合交叉定义，通过语义定义实现对复合元素、元素的概念及其约束的声明。标准之复合元素、元素的定义应遵循如下规则：

- 各元数据应最大可能采用推荐的元数据项并在语义上保持严格一致；
- 对推荐的元素不能描述的特性可以增加元素但新增加元素不能与已有元素有任何语义上的重复；
- 为了更准确地描述对象，允许向下再设若干层数据元素（子元素），数据元素间的语义是不重叠的，合起来不能超过复合元素定义的内涵；
- 数据元素不可再分。

对于元数据的元素（复合元素）定义方法，推荐采用与 Dublin Core 大体一致的方法，即采用 ISO/IEC 11179-3 标准，通过中文名称、英文名称、标识、定义、数据类型、值域、可选性、最大出现次数、注解九个属性来描述元数据。

● 中文名称

元数据对象的规范化中文名称，一般指全称。

● 英文名称

元数据的英文名称，一般用英文全称。

● 标识

标识用于本标准规范体系内唯一识别元数据复合元素或元素，其确定规则如下：

- 在本标准范围内必须唯一。

- 对存在国际或行业领域惯用英文缩写的元数据复合元素和元数据元素，采取该英文缩写为其标识。
- 对于根据英文名称或其他认识自定义的标识，在保持唯一性的前提下统一取每个单词前三个字母作为其标识，当如此取词不能保证唯一性时应延展取词位数，通常仅增加一位，如此仍不能保证唯一性时如前继续延长取词，直至保证唯一性为止。
- 对于元数据复合元素的标识的写法是，所有组成词汇的缩写为无缝连写，并且每个词汇缩写的首字母大写；
- 对于元数据元素的标识的写法是，所有组成词汇的缩写为无缝连写，首词汇全部采用小写字母，其余每个词汇缩写的首字母大写。

● 定义

描述元数据对象的基本含义与内容。

● 数据类型

元数据的有效值域和允许对该值域内的值进行有效操作的规定。

例如整型、实型、布尔型、字符串、日期等。

● 值域

说明元数据元素取值范围。

● 可选性

说明元数据复合元素或元素是否必须选取的属性，包括必选（M）、可选（O）和条件必选（C）三种情况。特别当该元数据为条件必选时，应注明其约束条件。

必选（M）

表明该复合元素或元数据元素必须选择。

可选（O）

根据实际应用可以选择也可以不选的复合元素或元数据元素。

可选复合元素可以有必选元素；但这些元素只当可选复合元素被选用时才成为必选的。如果一个可选复合元素未被选用，则该复合元素所包含的元素（包括必选元素）也不选用。

条件必选（C）

说明该选择复合元素或元数据元素的条件，当该条件满足时，至少一个元数据复合元素或元数据元素必选。“条件必选”用于以下三种可能性之一：

——表示在 2 或 2 个以上复合元素或元数据元素中进行选择。至少存在一个复合元素或元数据元素必选。

——当已经选用另一个复合元素或元数据元素时，此复合元素或元数据元素为必选。

——当另一个元数据元素已经选择了一个特定值时，此元数据元素为必选。

● 最大出现次数

说明复合元素或元数据元素可以具有的最大实例数目。只出现一次的用“1”表示，重复出现的用“N”表示。不为1的固定出现次数用相应的数字表示，如“2”、“3”、“4”等。

● 注解

对元数据的含义的进一步解释。

元数据方案的适用范围、词汇与约束规则集合是必备内容，对于一个元数据来说，在词汇和约束规则集合定义之前需要根据实际情况可能需要说明以下内容：

- 规范化引用：该标准所引用性包含、涉及的其他标准的内容。
- 概念和术语：标准中涉及的有关概念、术语的释义说明。
- 约束规则或约定：标准内容所使用的规则或者约定，如唯一标识符的命名规则等。

规范化引用等为选择使用内容，当存在应用的必要时应该在标准中定义上述内容，当然对于上述没有涉及的内容需要在标准词汇及其约束前定义的内容，也应该根据情况予以说明或定义。

此外，还应该根据实际需要，在标准中为用户提供编码体系、编码语言、句法描述语言等方面的内容。

(a) 编码体系 (Encoding Scheme)。编码体系是对数据元素值进行规约的受控词表或者规则，也是多数元数据都具有的重要组成部分。建议为用户提供推荐编码方案。

(b) 句法描述语言。建议为用户推荐以某种句法描述语言（例如，XML Schema、UML等）对元数据的描述。

(c) 编码语言。包含对特定编码语言（例如，XML、HTML、Oracle 关系数据库语言等）的绑定规则 (binding rules)，建议为用户推荐使用的编码语言。

为了使所定义的标准规范具有更灵活的适应性，元数据一般地都包括相应的扩展和制定元数据应用方案 (Application Profile) 的规则。本参考模型针对数据应用环境元数据扩展和制定元数据应用方案的需要提供了普适的系列规则，所以在定义具体标准时不需再单独提供上述规则，只需参照本参考模型对应内容即可。如若对现有规则有所补充，可以仅增加补充的内容；但若有修改，则应明确所修改的内容。

标准内容的字典表是上述内容之外的一项选用内容呈现方式，用表格的方式将前面定义的内容复合元素、元素及其约束、关系以表格的形式呈现给用户。

5.3 文档规范

元数据方案最终应当以文档的形式呈现给用户，故在参照GB/T 1.1—2000《标准化工作导则 第1部分：标准的结构和编写规则》的基础上确定了标准文档书写规范用于指导、规范化元数据文档的工作。

元数据应包含的内容要素如表 1 所示。

表 1 内容要素

要素类型	要素编排
概述	封面 目次 前言 引言
正文	名称 1 范围 2 规范性引用文件 3 术语和定义 4 符号和缩略语 5 要求 6 元数据模式 7 元数据 8…定义与元数据有关的其它章。
附录	附录A 元数据数据字典 附录B…定义元数据有关的其它附录
参考资料	元数据制定过程中的参考资料

注：黑体表示元数据必备要素，正体表示元数据可选要素

5.3.1 正文要素

- **范围**
介绍该元数据所定义的主要内容，标准适用范围及对象。
- **规范性引用文件**
说明该元数据所引用的文件。
- **术语和定义**
定义该元数据中所采用的术语和定义。
- **符号和缩略语**

说明该元数据对元数据的描述方法及相关的约定。

● 要求

说明参照本标准制定元数据应用方案时应与本标准保持一致。

● 元数据模式

概要说明元数据的内容参考模型，并逐一简要介绍元数据的各个组成复合元素。

● 数据标准

以摘要表示的形式对组成元数据的根元素、以及复合元素、数据元素进行描述。摘要表示应遵从本参考模型关于标准复合元素、元素等定义的规定，并按照根元素—复合元素—数据元素的层次结构组织行文条款。

对于元数据的元素定义方法，推荐采用 ISO/IEC 11179-3 标准，通过中文名称、英文名称、标识、定义、数据类型、值域、可选性、最大出现次数、注解九个属性来描述。

● 其他

与元数据有关的其他正文内容要素，如扩展元素信息等。

5.3.2 附录要素

● 元数据数据字典

元数据数据字典作为标准的规范性附录，通常表现为表格形式。数据字典有助于清晰地定义每一个元素、复合元素、代码表，方便读取元数据定义和描述信息。表 2 是元数据数据字典描述示例：

表 2 元数据的字典形式示例

序号	中文名称	英文名称	标识符	定义	数据类型	值域	约束/条件	最大出现次数	注释
1	元数据	Metadata	metadata	定义生态科学数据的元数据的根元素。	类	第 2-11 行	M	1	
2	角色名称：标识信息	Role name: Identification Information	identInfo	元数据描述的数据资源（数据集或数据集系列）的基本信息。	关联	标识（A.1.2）	M	1	
3	角色名称：数据质量信息	Role name: Data Quality Information	dqInfo	数据资源的质量评价信息。	关联	数据质量（A.1.3）	O	N	

注：数据字典对每个元数据元素的中文名称、英文名称、标识符、定义、数据类型、值域、约束/条件、最大出现次数、注释等九个属性进行了完整说明。

在数据字典中，对于复合元素和数据元素的约束/条件属性，用“M”表示必选，“C”表示条件必选，“O”表示可选；对于复合元素和数据元素的最大出现次数，只出现一次用“1”表示，不固定次数的重复出现用“N”表示，允许不为 1 的固定出现次数，并用相应的数字（即“2”、“3”、……）表示。

5.4 元数据注册

修订完善后的元数据方案应当通过有效途径、网站注册发布，以促进元数据方案在更大范围内被传播、接受认可和使用，以及支持基于不同元数据的元数据之间的交换、互操作和共享。

元数据注册需要根据注册系统的要求，把标准的内容“注入”到系统中，具体的注册方法和手段需参考相应元数据注册系统的要求和指导。

6 系统建设

系统建设主要实现元数据管理系统设计开发、实用评估和完善，并最终发布后服务。

6.1 设计开发

元数据制定完毕后，应发展相应的元数据管理系统。系统开发任务在于开发与标准规范相一致的元数据系统与工具。开发前及期间，为保证系统设计和实现的针对性和有效性，主题计划、元数据及系统发展者之间的参与成员，应持续交流与讨论意见。

6.2 试用、评估和完善

系统雏型完成后，系统设计开发人员应邀请元数据有关工作人员、用户等相关人参与系统的试用和测试，并要求测试人员和试用者完成相应的试用评估情况说明，供系统人员修正完善系统参考。

6.3 系统发布

系统在参考测试和试用意见的基础上完善后，经有关系统建设领导认可后可以对用户发布，同时还应为用户提供全方位的技术支持，协助用户解决安装过程中遇到的各种问题，以及指导用户试用系统完成元数据内容的著录、管理和网络发布。

7 应用服务

应用服务主要包括元数据服务、扩展方案和元数据映射规范，以支持灵活多样的元数据应用需求。

元数据服务主要通过方案的外向普及，或者管理系统的应用推进，实现基于元数据方案的元数据服务。元数据扩展方案和数据映射则是在原方案基础上的高级应用，二者各有所重，内容自成特点。

7.1 扩展方案

数据资源学科广泛、类型复杂的特点，决定了不可能只制定一个元数据就可以描述所有的科学数据集，而是需要建立多个元数据。通常由于资源和需求的多样性导致某个元数据对于应用需求“过剩”或（及）“不足”，因此需要根据应用对该标准定义的元数据及其关联结构进行定制裁剪，以及（或者）补充元数据。

应用方案包括元数据的定制和扩展两个方面的内容。

7.1.1 元数据的定制

元数据的定制是根据应用需求，从原标准定义的元数据复合元素和元素中去除掉不需要的复合元素和元素。在裁剪之前，必须认真分析应用需求，并仔细地查阅原标准中现有的复合元素和元素。通过将应用需求与原标准中的复合元素和元素进行比照，保留应用需求所需要的复合元素和元素，而舍弃不需要的复合元素和元素。

在进行元数据定制裁剪时，可以灵活地采用从上向下的遍历策略或者从下向上的遍历策略选择需要的复合元素和元素，并删除不需要的复合元素和元素。

元数据裁剪的原则是如果一个复合元素被删除，则意味着其所包含的复合元素和元素皆被删除。反之，如果一个复合元素或被保留，则其所在的元数据复合元素也应保留。

7.1.2 元数据的扩展

元数据的开展是现有标准不能满足应用需求时，在现有标准的基础上适当增加内容对象实现需求的方法。在扩展之前应认真分析应用需求，并仔细地查阅原标准中现有的复合元素和元素。通过将应用需求与原标准中的复合元素和元素进行比照，保留应用需求所需要的复合元素和元素，并在适当的位置增加适合的内容对象。

7.1.2.1 扩展原则

- 扩展的元数据元素不应改变原标准中现有元数据元素的名称、定义或数据类型属性；

- 增加的元数据元素应按照原标准所确定的等级结构进行合理地组织。努力使增加的元数据元素组织进现有的复合元素中,无法为增加的元数据元素找到在逻辑上恰合或者自然恰合的上层复合元素时,可以定义新的复合元素;
- 增加的复合元素可以包含增加的和现有的复合元素/元素,作为其组成部分;
- 允许以代码表替代值域为自由文本的现有元数据元素的值域;
- 允许增加现有代码表中值的数量,扩充后的代码表应与扩充前的代码表在逻辑上保持一致;
- 允许对选择保留的元数据元素的值域进行缩小(例如,在原标准中规定元数据元素的值域中有 7 个值,在定制后可以规定它的值域只包含其中的 4 个值);
- 允许对保留的复合元素/元素的可选性和最大出现次数施以更严格的限制(例如,在原标准中定义为可选的描述符,在扩展后可以是必选的;在原标准中定义为可无限次重复出现的元数据复合元素/元素,在扩展后可以是只能出现 1 次);
- 不得扩展原标准所不允许的任何内容。

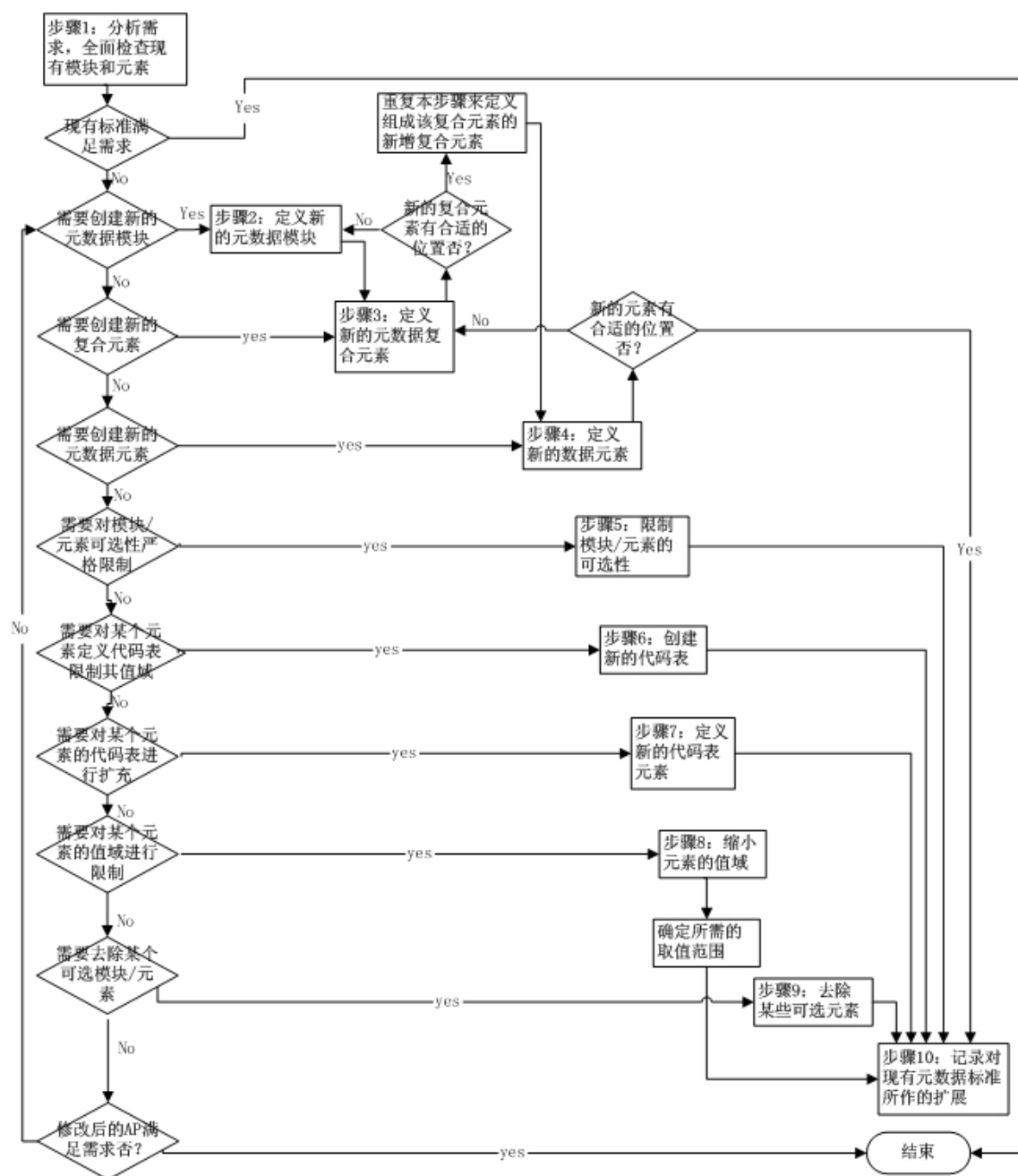
7.1.2.2 扩展类型

允许下列扩展类型:

- 1) 增加新的元数据元素;
- 2) 增加新的元数据复合元素;
- 3) 建立新的代码表,代替值域为“自由文本”的现有元数据元素的值域;
- 4) 创建新的代码表元素(对值域为代码表的元数据元素的值域进行扩充);
- 5) 对现有元数据复合元素/元素施加更严格的可选性限制;
- 6) 对现有元数据复合元素/元素加更严格的最大出现次数限制;
- 7) 缩小现有元数据元素的值域。

7.1.3 元数据应用方案的设计方法

可以遵循以下这个包括 10 个步骤的方法来制定元数据应用方案。下图展示了元数据应用方案的设计流程。



元数据应用方案流程

步骤 1：分析元数据需求，并全面检查原标准中的元数据复合元素和元素

首先要分析元数据要描述的资源对象是什么，描述的粒度有多大，明确元数据描述的对象数据集所属的学科专业、资源类型以及其他资源特征。结合分析结果，仔细检查原标准现有元数据复合元素和元素。这种检查不仅应该覆盖元数据复合元素和元素的名称，而且覆盖其定义、数据类型、可选性、值域、最大出现次数以及注释等，以确保对原标准所做的扩展都是符合扩展规则的。

方法：

- 1) 如果现有元数据已经满足应用的需求，自然无需对现有元数据进行扩展；
- 2) 如果标准现有复合元素不能满足需求，需要创建新的元数据复合元素，则进行步骤

2;

3) 如果需要创建新的元数据复合元素来满足需求, 则进行步骤 3;

4) 如果需要创建新的元数据数据元素来满足需求, 则进行步骤 4;

5) 如果某个现有元数据复合元素或元素可以满足需求, 但是需要对其可选性作更严格的限制, 则进行步骤 5;

6) 如果某个现有元素可以满足需求, 但是其值域过于宽泛, 能够用一个代码表来限制其值域, 则进行步骤 6;

7) 如果某个现有元素可以满足需求, 但是需要将作为其值域的代码表进行扩充, 则进行步骤 7;

8) 如果某个现有元素可以满足需求, 但是应用方案中所需值域只是核心元数据中所规定值域的一个子集, 则进行步骤 8;

9) 如果确定核心元数据中的某个可选元素确实不为应用方案所需要, 即需要去除某个可选元素, 则进行步骤 9。

步骤 2: 定义新的元数据子集

如果现有子集不能从总体上满足需求, 而且确定无法通过扩展其中任一子集来满足, 确实需要创建一个新的元数据子集, 可以定义一个新的元数据子集。

为了使数据应用环境系统元数据都保持一致的风格, 应依据 5.2 节元数据定义规范中规定的元素属性对新增子集进行定义, 既需要依次定义该新建子集的中文名称、英文名称、标识、定义、数据类型 (对于子集来说, 其数据类型为“复合类型”)、值域、可选性、最大出现次数、注释等方面的属性。

方法:

1) 通过步骤 3 来定义组成该的复合元素;

2) 通过步骤 4 来定义组成该复合元素的数据元素;

3) 进行步骤 10。

步骤 3: 定义新的元数据复合元素 (步骤 3)

如果现有元数据复合元素不能满足需求, 而且确定无法通过扩展其中任一复合元素来满足需求, 这种情况下, 可以定义一个新的元数据子复合元素来满足应用方案的特定需求。

为了使数据应用环境系统元数据都保持一致的风格, 应依据 4.3 节元数据定义规范中规定的元素属性对新增复合元素进行定义, 即需要依次定义该新建复合元素的中文名称、英文名称、标识、定义、数据类型 (其数据类型为“复合类型”)、值域、可选性、最大出现次数、注释等方面的属性。

方法:

1) 确定该复合元素的位置, 即确定该复合元素属于哪个复合元素。如果无法在现有结构中为该复合元素找到合适的位置, 则进行步骤 2;

2) 确定该复合元素所包含的元数据元素;

- 3) 通过步骤 3 来定义组成该复合元素的新增复合元素;
- 4) 通过步骤 4 来定义组成该复合元素的新增数据元素;
- 5) 进行步骤 10。

步骤 4: 定义新的数据元素

如果现有数据元素不能满足需求, 而且确定无法通过扩展其中任一数据元素来满足需求, 这种情况下, 可以定义一个新的数据元素来满足应用方案的特定需求。

为了使数据应用环境系统元数据都保持一致的风格, 应依据 4.3 节元数据定义规范中规定的元素属性对新增数据元素进行定义, 即需要依次定义该新建数据元素的中文名称、英文名称、标识、定义、数据类型、值域、可选性、最大出现次数、注释等九个方面的属性。

方法:

- 1) 确定该数据元素的位置, 即确定该数据元素属于哪个复合元素。如果无法在现有结构中为该数据元素找到合适的位置, 则进行步骤 3;
- 2) 进行步骤 10。

步骤 5: 限制复合元素/元素的可选性

某个现有复合元素/元素可以满足需求, 但是应用方案需要对其可选性进行更严格的限制, 即将其可选性由可选改为必选。若确定进行此改动, 则进行步骤 10。

步骤 6: 创建新的代码表

某个现有元素可以满足需求, 但是, 需要对其值域进行限制, 而现有的代码表都不能满足需求, 需要定义一个新的代码表来满足应用方案的特定需求。

如果新定义的代码表以某一国家或国际标准为依据, 应注明该依据; 如果新定义的代码表无据可查, 则应在代码表中设置一栏, 用来说明每个代码元素的含义。进行步骤 10。

步骤 7: 定义新的代码表元素

某个现有元素可以满足需求, 但是需要将作为其值域的代码表进行扩充, 这时应该参考现有代码表元素定义新的代码表元素。

扩充后的代码表必须与扩充前的代码表在逻辑上保持一致, 必须是原代码表在逻辑上的一个扩展, 否则, 则可能是该元素的值域不适宜扩展, 应该返回步骤 1。

为了准确无误地将新定义的代码表元素反映到应用方案的文档中, 进行步骤 10。

步骤 8: 缩小元素的值域

某个现有元素及其值域可以满足需求, 但是应用方案中所需该元素的值域只是本标准中所规定值域的一个子集。

方法:

- 1) 确定所需要的取值范围;
- 2) 进行步骤 10。

步骤 9：去除某些可选元素

在确信核心元数据中的某个可选元素确实不为应用方案所针对的数据资源所需要的情
况下，可以不在应用方案中保留将该可选元素，并进行步骤 10。

步骤 10：记录对现有元数据所做的扩展

一旦对现有元数据做了以上任何形式的扩展，都应该及时将此扩展清晰地记录下来。而
且，对现有元数据所做的所有扩展都必须在派生自原标准文档的应用方案文档中以一定格式
写明，并具体反映到应用方案的“复合元素概述”、“复合元素详细介绍”等其它相关部分，
以及 XML Schema（如果采用 XML 作为编码语言）等表现形式中。

7.2 元数据映射规则

本规则适用于数据应用环境各建设单位研究和制定本参考模型中定义的元数据与参考
模型外元数据的映射。

目前各元数据方案在适用领域、适用范围和复杂程度上各不相同，这给元数据集之间的
互操作带来了问题。因此，作为本参考模型之外的其他元数据，我们需要研究它们和参考模
型内元数据之间的相互映射关系，从而便于本参考模型内的元数据与参考模型外元数据之
间的互操作。

由于不同的元数据在广度、深度、侧重点和覆盖范围等方面都存在不同之处，所表达
的概念也难以完全对等，例如可能源标准中的一个或多个元素在目标标准中找不到相应的元
素；有可能某元素在源标准中为可重复元素，具有多个值，但在目标标准中却不可重复；还
有可能元素语义取值范围存在差异，如源标准中可用自由文本描述，而目标标准要求必须使
用规范代码表，因此实际映射时需要根据具体情况做出合适的处理方式。本建议不涉及具体
的元数据映射规则，只对元数据映射的概念、内容和方法作出阐述，对元数据映射的行文格
式作出规定，具体的映射细节将在具体的元数据中作为附录给出。

7.2.1 元数据映射

元数据映射（Metadata Mapping）是实现元数据互操作的一种手段，其具体含义是指分
析元数据中的元素，将不同标准中相同或相似含义的元素联系起来建立对应关系。元数据之
间元素映射形成的集合称为元数据映射表（Metadata Crosswalk），也可称为元数据映射字典。

7.2.2 元数据映射内容

元数据映射可分为语义映射和结构映射两个方面。语义映射主要针对不同的元数据（如
DC、EAD、TEI 等），分析标准中的元素含义，建立元数据映射表，近似的实现数据资源的

“跨库”揭示。结构映射主要解决不同元数据之间的对应关系，更多地表现为一种“格式转换”，例如将 RDF 转换成 XML Schema 形式，以此来提供异构系统的互操作。下面对语义映射的主要内容进行说明。

1) 元素映射

由于不同元数据的制定是针对不同的学科领域、不同的国家、地区，所覆盖的深度、广度有所不同，因此元素之间的映射关系比较复杂。主要有以下几种：

- 一对一关系：这是最简单的一种映射，即源元数据中的某元素与目标元数据中的某元素语义相同，并与其一一对应。这种一一对应关系有时候是等值对应，有时目标元数据中没有与源元数据完全对应的元素，这时可用相似含义的元素替代。但由于元数据的复杂性，这种一一对应关系在元数据映射中并不多见。
- 一对多关系：如果源元数据中的元素概念比较粗略和宽泛，而目标元数据中的元素定义更加细化和明确，那么源元数据中某个元素表达的语义可能包括目标元数据若干元素表达的语义。这样源元数据中的这个元素根据其语义可以分别对应到目标元数据中的若干元素。
- 多对一关系：与一对多关系相反，当源元数据中的元素概念更具体和细化，而目标元数据中的元素定义比较笼统的时候，源元数据中的不同元素可能映射到目标元数据的同一元素中。
- 其他：除了以上几种关系外，在具体建立元数据映射表时还要考虑源元数据和目标元数据中没有映射关系的元素，元数据中元素的层次结构在映射时的处理，以及允许多个复合元素对象的元数据与单个复合元素对象标准之间的映射关系。

2) 值域映射

在进行元数据映射时，不仅要考虑到元素之间的对应关系，还要考虑对应元素间值域的映射。有时源元数据中某元素的取值是自由文本，而与之对应的目标元数据中的元素取值需要采用受控词；或者源元数据和目标元数据对应的元素采用不同的受控词表。这些元素值的转换规则都要在元数据映射表中详细说明。

7.2.3 元数据映射方法

元数据映射可以采取动态和静态的方式。动态方式根据元数据映射表建立元数据转换中间件，将相应的查询请求中的有关内容转换成资源站点支持的元数据模型或可以识别的元数据格式，返回时再转换成本系统支持的元数据格式，以支持本系统的查询结果处理。静态转换类似于目前的搜索引擎，将资源站点的数字对象抓取到本地，按照本地的元数据模型建立索引，提供服务。

7.2.4 映射编写规定

由于元数据映射只能是一种近似的映射，所以各单位在制定元数据时，只能根据实际情况和项目外标准建立一种相对合理的映射方案。本部分只对行文中映射部分需要包含的内容作出最小约定。

与项目外标准的映射可以作为可选性附录置于元数据文末，具体内容应该包括以下两部分：

- 1) 元素映射表：以表格的形式描述源标准和目标标准间的元素映射关系，该表格至少包含三个字段：源标准元素，目标标准元素，说明。其中说明字段用于对具体元素映射时的处理方式、元素值域的映射方式作出解释。
- 2) 实施建议：对于所作出的元数据语义映射给出实施建议，是能够以此进行格式转换还是只能达到标准间部分的语义映射。

8 一致性测试规则

一致性测试适用于本参考模型中定义的所有元数据以及由上述标准发展而来的任何元数据应用方案，用于保持数据应用环境元数据参考模型中所定义的元数据及其扩展应用方案的一致性。数据应用环境项目内定义的任何元数据或应用方案应当按照接受测试标准或应用方案的相应部分提供元数据，用户扩展的元数据应当遵循本参考模型中所定义的元数据扩展规范确定和提供元数据，并通过一致性测试。

本一致性测试规则作为仅适用于元数据与其扩展应用方案的一致性测试，而不适用于：

- 测试该标准或应用方案的元数据子集、复合元素和元素是否存在逻辑上的不一致；
- 测试该标准或应用方案的元数据子集、复合元素和元素的设置是否存在专业、学科和表述上的错误或不妥之处。

一致性测试则是保证系统建设和应用服务在采用元数据方案方面的一致性所应实现的功能要求，是元数据方案得以贯彻执行的保障。

8.1 标签检查

1) 测试目的：保证元数据方案应用中所采用元数据子集、复合元素和元素的标签与方案完全一致。

2) 测试方法：检查元数据方案应用中每个元数据子集、复合元素或元素的标签，将之与所采用方案相应内容的对应元数据子集、复合元素或元素的标签进行对比。每个元数据子

集、复合元素或元素的标签与方案完全一致。

8.2 完整性测试

1) 测试目的：保证接受测试的元数据应用方案中来自被扩展标准中的所有必选或者条件必选的元数据子集、复合元素和元素的一致性。

注：很多规定为必选的元素包含在任选或条件必选的元数据复合元素中。只有选用包含它们的元数据复合元素时，这些元素才成为必选的。

2) 测试方法：将被扩展的元数据与接受测试的元数据应用方案进行对比，查看被扩展标准中相应部分所定义为必选的元数据子集、复合元素和元素是否在元数据应用方案中出现。当符合被扩展标准设定的有关条件时，还要对比检查相应部分定义为条件必选的元数据子集、复合元素和元素是否在元数据应用方案中出现。

以下测试适用于各种约束/条件——必选、条件必选和可选——的元数据子集、复合元素和元素。

8.3 最大出现次数测试

1) 测试目的：保证元数据应用方案中来自被扩展标准的每个元数据子集、复合元素和元素出现次数不超过被扩展标准规定的次数。

2) 测试方法：检查元数据应用方案中每个元数据子集、复合元素或元素的最大出现次数，将之与被扩展标准中相应部分定义的对应该元数据子集、复合元素或元素的“最大出现次数”属性进行对比。每个元数据子集、复合元素或元素的最大出现次数不应超过被扩展标准中规定的对应该元数据子集、复合元素或元素的最大出现次数。

8.4 标识符测试

1) 测试目的：保证接受测试的元数据应用方案中来自被扩展标准的每个元数据元素使用规定的标识符。

2) 测试方法：检查元数据应用方案中每个元数据元素的标识符，确定它是否在被扩展标准中定义。

8.5 数据类型测试

1) 测试目的：保证接受测试的元数据应用方案中来自被扩展标准的每个元数据元素使

用符合规定的数据类型。

2) 测试方法: 检查元数据应用方案中每个元数据元素的数据类型, 确定其数据类型是否符合被扩展标准中的规定。

8.6 值域测试

1) 测试目的: 保证接受测试的元数据应用方案中来自被扩展标准的每个元数据元素的取值范围在规定的域内。

2) 测试方法: 检查元数据应用方案中每个元数据元素的取值范围, 确定它们是否落在被扩展标准规定的值域内。

8.7 模式测试

1) 测试目的: 保证接受测试的元数据应用方案遵循被扩展标准定义的模式。

2) 测试方法: 检查元数据应用方案中来自被扩展标准的每个元数据元素, 确定它们是否包含在被扩展标准定义的相应元数据复合元素中。

8.8 用户自定义的扩展元数据测试规范

8.8.1 排他性测试

1) 测试目的: 保证每个用户自定义的元数据子集、复合元素和元素是唯一的, 且尚未在被扩展标准中定义过。

2) 测试方法: 检查元数据应用方案中每个用户自定义的元数据子集、复合元素和元素, 确定其是唯一的, 且此前未曾在被扩展标准中使用。

8.8.2 定义测试

1) 测试目的: 保证用户自定义的每个元数据子集、复合元素和元素按照被扩展标准的规定进行了定义。

2) 测试方法: 检查元数据应用方案中每个用户自定义的元数据子集、复合元素和元素, 保证被扩展标准中要求的属性都已经被定义。

8.8.3 元数据应用方案测试

1) 测试目的: 检查接受测试的元数据应用方案是否遵循被扩展标准确定的规则。

2) 测试方法: 按照 2 和 3 的规定进行测试。

8.8.3.1 元数据应用方案测试报告

测试方对元数据应用方案进行一致性测试后，给出最终测试报告，应包含以下内容：

8.8.3.2 测试方

执行测试的经过授权的机构名称。

8.8.3.3 测试员

代表测试方对元数据应用方案进行测试的测试人员的姓名。

8.8.3.4 被测试的元数据应用方案

被测试的元数据应用方案及其被扩展标准的名称、制定方和版本号。

8.8.3.5 日期和时间

测试所进行的日期和时间。

8.8.3.6 最终结果

被测试的元数据应用方案是与被扩展元数据严格一致、一致或不一致的元数据应用方案。

8.8.3.7 建议

综合被测试的元数据应用方案对必选、可选、扩展的元数据子集、复合元素和元素的处理情况，给出建议以提高该元数据应用方案与其被扩展元数据的互操作性。

9 评估完善

评估完善则基于系统建设和应用服务，特别是应用服务的效果，对元数据方案进行评估和完善，提升元数据方案的效用。

9.1 应用评估

依据项目计划的需求，检查元数据整体实施程序和效益，主要应评估以下项目：

- 元数据记录的质量：包括完整性、准确性、数据粒度(granularity)和数据可用性(serviceability)，如从决策和程序二方面，评估元数据检索质量。
- 采用标准对于检索的效益：从其中关键数据项查询内容，评估其检索效率与精确度。
- 元数据管理工具的使用性：如评估元数据工具的发展效益，及后续设计的重点。

9.2 方案完善

元数据管理系统的建设和应用、应用服务都是元数据方案适用能力的检验，所以应该在评估元数据方案适用于多方需求的基础上，针对方案的不合理或不适用之内容进行修订完善，以便更好的支持各类应用对元数据方案的需求。