

论电子文件管理元数据顶层框架设计的标准化 *

张正强

摘 要 研究电子文件管理元数据顶层框架设计标准化既有现实意义,又有理论意义。通过对电子文件管理元数据顶层框架设计标准化可供选择的两大基础性标准,即 OAIS 参考模型的元数据顶层框架和国际标准 ISO/23081 元数据的顶层框架的比较分析,认为在电子文件管理元数据标准的制定上,应该选择国际标准 ISO/23081 的电子文件管理元数据顶层框架。图7。参考文献4。

关键词 电子文件管理元数据 顶层框架设计 OAIS ISO/23081 元数据框架

分类号 G254

ABSTRACT The author thinks that there is practical and theoretical significance for the study of the standardization of top-level framework design for records management metadata. Between the two available basic standards, i. e. OAIS and ISO 23081, he proposes to use ISO 23081. 7 figs. 4 refs.

KEY WORDS Records management metadata. Top-level framework design. OAIS. ISO 23081. Metadata framework.

CLASS NUMBER G254

1 研究电子文件管理元数据顶层框架设计标准化的意义

1.1 现实意义

电子文件管理元数据已在我国电子文件领域引起了高度重视,目前正起草的电子文件管理元数据的标准,在文件领域有电子文件管理元数据的国家行业标准和电子文件管理元数据的区域标准,在电子政务领域有电子公文管理元数据的国家标准。此外,还有专门行业如科技部、核工业等的电子文件管理元数据行业标准也正在制定之中。

电子文件管理元数据标准的起草与制定在我国文件领域引起如此广泛的重视,凸显了电子文件管理元数据在我国电子文件管理中的地位与作用,说明我国电子文件管理正朝着更深的方向发展,这是十分值得关注的。电子文件管理

元数据是电子文件管理的“命脉”,是电子文件管理中保证电子文件凭证价值的不可缺少的核心部分,这一点已在国际文件领域形成共识。所以,电子文件管理元数据标准必然备受关注,在该背景下,研究电子文件管理元数据顶层框架设计的标准化就十分重要的现实意义。

1.2 理论意义

电子文件管理元数据顶层框架设计的标准化,从目前来看,可供选择的基础性标准主要有两类:一类是长久保存元数据框架标准,即按照《ISO 14721:2003 空间数据和信息传输系统 - 开放档案信息系统 - 参考模型》(以下简称 OAIS 参考模型)的元数据框架来进行设计;一类是文件元数据框架标准,即按《ISO 23081 - 1:2006 信息与文件 - 文件管理过程 - 文件元数据第一部分:原则》和《ISO 23081 - 2:2007 信息与文件 - 文件管理过程 - 文件元数据第二部分:

* 本文系国家社科基金重点资助项目“基于 XML 电子文件管理元数据标准研究”(项目批准编号:04ATQ002)的系列研究成果之一。

概念与实施问题》的元数据框架来进行设计。当然,电子文件管理元数据顶层框架设计也有按其他框架标准设计的,但不占主流。

在这样的情况下,就存在一个如何实现电子文件管理元数据顶层框架设计的标准化的问题,或者说如何选择电子文件管理元数据顶层框架的基础标准问题,因为,对这个问题解决的科学与否直接关系到电子文件管理标准化程度的高低,也直接关系到电子文件管理元数据能否充分、全面、有效及可持续地保证电子文件的凭证价值。所以,对于该问题的研究,是建立电子文件管理元数据标准的基础理论研究,具有重大理论意义,可以直接指导电子文件管理元

数据顶层框架设计标准化的实践。

2 电子文件管理元数据标准顶层框架设计选择的基础性标准——OAIS 参考模型中的元数据顶层框架

2.1 OAIS 参考模型中的元数据顶层框架模型

OAIS 参考模型是一个各领域通用的长久保存数字信息以便于检索利用的基础性标准。在该标准中,其给出了一个针对数字对象元数据的顶层框架模型,该顶层框架主要体现在 OAIS 参考模型的信息模型中,如图 1 所示。

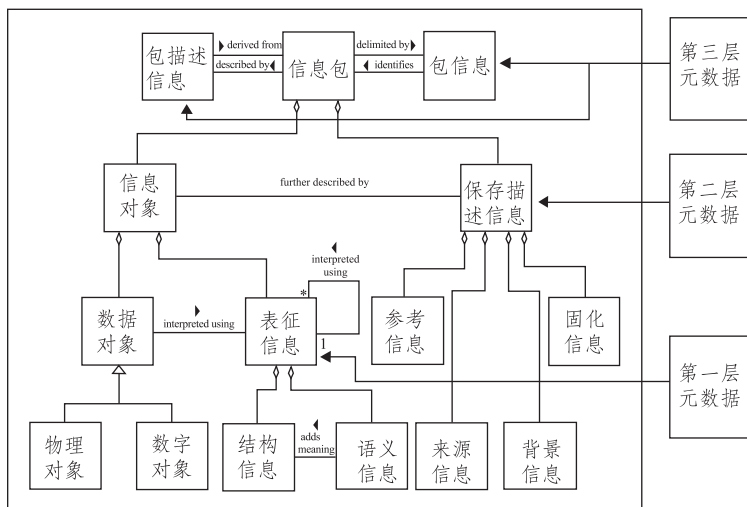


图 1 OAIS 参考模型的元数据框架模型示例

从图 1 可以看出,该元数据模型的基本框架实际上是一个三层架构元数据框架:

第一层元数据框架是关于数据对象的,即图 1 中的表征信息部分。其作用就是说明数据对象的特征,以便于显示和理解数据对象。在数据对象为数字对象时,表征信息的作用主要是说明数字对象的结构和语义,从而使由“0”和“1”所构成的数字对象转化为人—机可读的有意义的数据。

第二层元数据框架是关于信息对象的,即图 1 中的保存描述信息部分。其作用就是从信息长久保存的角度说明信息对象的特征,以确

保信息对象可被确认和理解信息对象被创建时所处的环境。保存描述信息包括四类:参考信息、来源信息、背景信息和固化信息。

第三层元数据框架是关于信息包的,即图 1 中的包信息与包描述信息部分。其作用是从总体上说明、描述信息包的特征,以便用户确认、定位与检索信息包。

在以上三层元数据框架中,第一层元数据框架的性质是属于信息对象在物理层面上的技术元数据框架;第二层元数据框架的性质是属于信息对象在逻辑层面上的长久保存元数据框架;第三层元数据框架,主要是从第二层元数据

中抽取出来的、用于检索的元数据框架。可见,在这三层元数据框架中,第一层与第二层是关键。所以,在 OAIS 参考模型的信息模型中,表征信息与保存描述信息实际上就构成该标准的长久保存元数据的顶层框架。

2.2 OAIS 参考模型中的元数据框架的具体内容

在 OAIS 参考模型的信息模型中,在信息对象物理层面上的第一层技术元数据框架,即表征信息,是由三个部分构成:

(1)结构信息:主要是通过将数字对象的比特流组织成特定的数据类型、数据类型组和更高层的意义来解释比特流,它包括数据格式的说明,以及对存取数据所需的硬件与软件环境进行的描述。结构信息主要是面向机读的信息。

(2)语义信息:就是在由结构信息对数据结构定义的基础上提供更多的意义,如结构信息将比特流定义为 ASCII 文本字符,而语义信息就是将这文本语言定义为汉语。所以,语义信息主要是面向直读(人读)的信息。

(3)其他表征信息:就是理解结构信息和语义信息所需的其他表征信息。如果把结构信息和语义信息看作是第一层表征信息,那么其他表征信息就是理解、说明第一层表征信息所需的第二层、第三层表征信息。

在信息对象逻辑层面上的第二层长久保存元数据框架,即保存描述信息,其框架是由四个部分构成:

(1)参考信息:主要是为信息对象提供能唯一确认该信息对象的标识符,如系统中产生的内部标识符,或一个 ISBN 的标识符。

(2)背景信息:主要记录信息对象与其环境的关系,如关于信息对象被创建的原因,以及信息对象与其他信息对象的关系,这些关系可以是内容主题上的相关关系或等级关系。

(3)来源信息:主要是记录信息对象的历史,说明信息对象的创建,以及创建后所经历的各种变化,保管情况的变化。

(4)固化信息:主要是提供数据完整性校验,或提供用于保证特定的信息对象不被未授

权改动的校验码。

在 OAIS 参考模型中,由表征信息元数据框架和保存描述信息元数据框架所构成的长久保存元数据的顶层框架,为国际范围内各领域的信息对象的长久保存提供了一个标准化的架构。所以,从整个国际范围内来看,OAIS 参考模型的长久保存元数据框架作为基础性框架标准,其产生的标准化的作用与影响是巨大的,意义也是深远的。

3 电子文件管理元数据标准顶层框架设计选择的基础性标准——国际标准 ISO/23081 元数据顶层框架

3.1 国际标准 ISO/23081 第一部分文件实体类元数据的顶层框架

国际标准《ISO 23081 - 1 信息与文件 - 文件管理过程 - 文件元数据第一部分:原则》(以下简称国际标准 ISO/23081 第一部分)于 2006 年正式颁布,该标准是直接根据文件管理的需求标准——国际标准《ISO 15489 信息与文件 - 文件管理》起草制定的。国际标准 ISO/23081 第一部分提出了一个被称之为“法规三元组”的文件管理元数据顶层框架,这个框架实际上是一个关于文件实体类的元数据顶层框架,如图 2 所示。

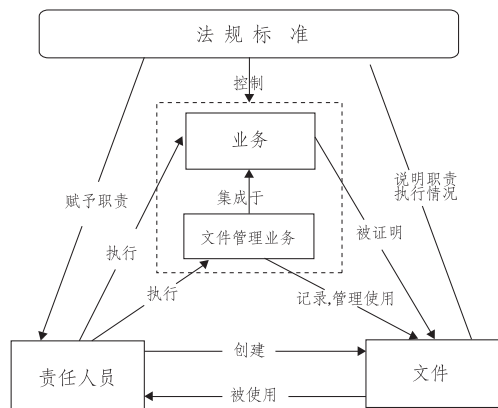


图2 国际标准 ISO/23081 第一部分的元数据法规三元组框架示例

从图2看,法规三元组的具体内容包括:

- (1)文件实体类元数据:主要是关于文件自身的元数据;
- (2)责任人员实体类元数据:主要是关于责任人员的元数据;
- (3)业务实体类元数据:主要是关于业务工作或业务过程、文件管理过程的元数据;
- (4)法规标准实体类元数据:主要是关于业务规章制度、政策以及法规的元数据。

由于法规三元组框架没有显示动态的文件管理元数据框架部分,因此,国际标准 ISO/23081 第一部分又进一步规定了上述法规三元组框架中的每一个实体类元数据,在文件捕获之前或之后同样适用,每一实体类的元数据都由下列两类元数据部分动态地构成:

(1)与文件一起捕获的元数据:将元数据固化于自身的业务背景之中,使管理过程能够开始进行(如在文件捕获点上的元数据)。

(2)不断创建与捕获的元数据(如过程元数据):这类元数据已不是由文件创建组织机构所生成的,但是任何一个负责自始至终地管理文件的组织机构都必须确保这类元数据的创建与捕获。

由此可见,国际标准 ISO/23081 第一部分的元数据顶层框架是紧紧围绕文件的实体而提出的,对文件适用具有鲜明的针对性;又由于它是根据国际标准《ISO 15489 信息与文件—文件管理》的文件需求标准起草制定的,其框架实现文件管理需求的功能十分清晰,具有明确的目的性。

3.2 国际标准 ISO/23081 第一部分文件实体类元数据顶层框架的理论基础

国际标准 ISO/23081 第一部分元数据顶层框架——法规三元组框架的提出有其理论基础,该理论基础就是文件连续体理论,可以说法规三元组框架就是文件连续体理论模型的具体化。

如果将法规三元组框架与文件连续体理论模型对应起来,就可以更清晰地看出其相互之间的内在联系,如图3所示。

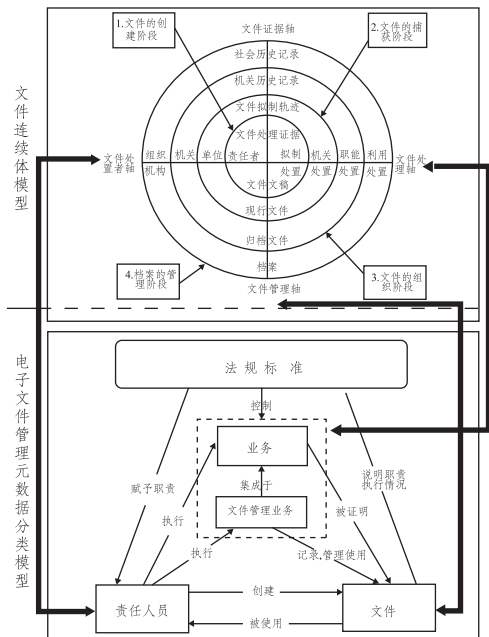


图3 法规三元组框架与文件连续体理论模型的对应

从图3可以清楚地看出法规三元组中的“文件、业务、责任人员”分别对应于文件连续体理论模型中的“文件管理轴、文件处置轴、文件处置者轴”,其渊源关系十分清晰。

3.3 国际标准 ISO/23081 第二部分的元数据顶层框架

3.3.1 国际标准 ISO/23081 第二部分文件实体类元数据框架——法规四元组框架的扩展

国际标准化组织在继2006年颁布国际标准 ISO/23081 第一部分之后,又于2007年7月30日颁布了国际标准《ISO 23081-2 信息与文件—文件管理过程—文件元数据第二部分:概念与实施问题》(以下简称国际标准 ISO/23081 第二部分)。国际标准 ISO/23081 第二部分在第一部分的基础上将文件管理元数据法规三元组框架进行了扩展,提出了法规四元组的框架,如图4所示。

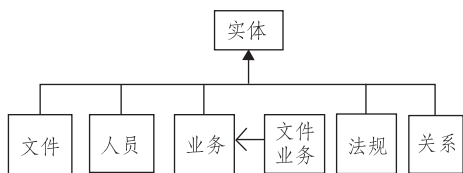


图4 元数据法规四元组框架示例

从图4可以看出,法规四元组框架与法规三元组框架相比,在原先四个实体类元数据的基础上又增加了关系实体类元数据。这个关系实体类,实际上在法规三元组框架中就存在,但由于它是隐性实体,所以在第一部分的框架中没有揭示出来,因此,法规三元组框架只是明确了四类显性实体。而法规四元组顶层框架却揭示了隐性实体类元数据——关系实体类元数据,所以,在元数据框架上,法规四元组框架是对法规三元组框架的扩展,比法规三元组框架更为完善。

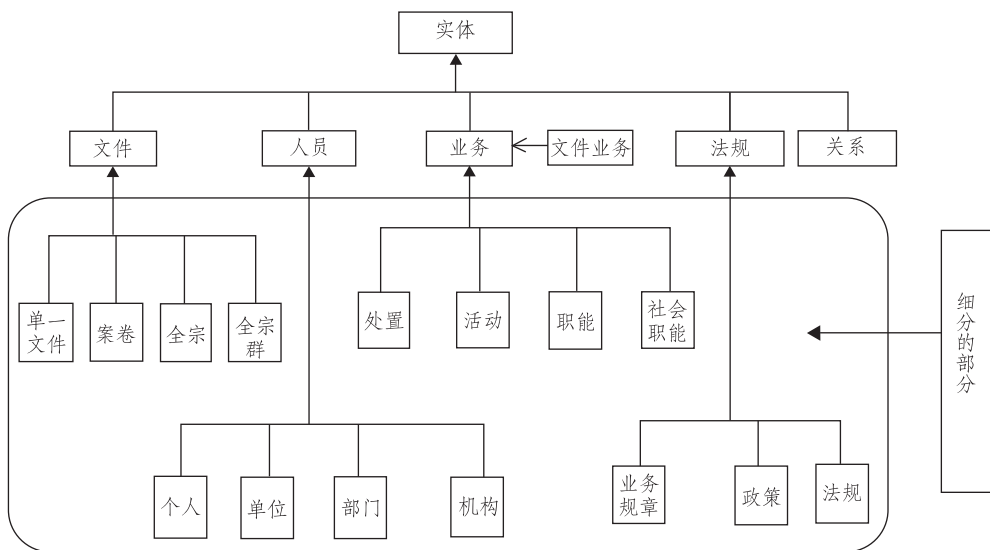


图5 国际标准 ISO/23081 第二部分的元数据法规四元组完整框架示例

其实图5所示的四个方面细化的内容在国际标准 ISO/23081 第一部分元数据法规三元组框架中也包含有,但是并没有被揭示出来,而在第二部分的法规四元组完整框架中得到了明确的揭示与表达。这无疑又是对第一部分元数据

3.3.2 国际标准 ISO/23081 第二部分文件实体类元数据框架——法规四元组框架的细化

国际标准 ISO/23081 第二部分元数据法规四元组框架不仅对第一部分进行了扩展,而且对其进行了细化,主要体现在四个方面:

(1) 对文件实体类元数据进行了细化,分出了单一文件、案卷、全宗和全宗群;

(2) 对责任人员实体类元数据进行了细化,分出了人员、单位、部门和机构;

(3) 对业务实体类元数据进行了细化,分出了处置、活动、职能和社会职能;

(4) 对法规标准实体类元数据进行了细化,分出了业务规章、政策和法规。

由于国际标准 ISO/23081 第二部分元数据法规四元组框架对以上四个方面的细化,从而形成了更为完整的文件实体类元数据框架,如图5所示。

法规三元组框架的完善和具体化,因此从这个意义上而言,国际标准 ISO/23081 第二部分元数据法规四元组完整框架又可被称为“法规四元组级次框架”,因为该框架把文件实体类元数据的等级级次表现得清清楚楚。

3.3.3 国际标准 ISO/23081 第二部分的属性类元数据框架

由于法规四元组级次框架仅是关于文件实体类的元数据的顶层框架,还没有具体表现出文件实体类元数据的属性元数据,因此,国际标准 ISO/23081 第二部分又提出了文件实体类元数据的属性类元数据的框架,从而构建了系统的文件元数据顶层框架体系。国际标准 ISO/23081 第二部分所提出的属性类元数据共包括六组:

- (1) 标识组:主要是确认文件元数据实体的元数据,包括实体类别、实体级次和实体注册号等。
- (2) 描述组:主要包括所需的用于确定被检索的实体是否可用的元数据,如题名、分类、提要、存贮位置、所有权及外部标识符等。
- (3) 使用组:主要包括有助于实体长久使用的元数据,如技术环境、权限、利用、文件使用对象、语种、完整性及文件种类等。
- (4) 计划事件组:主要是用于管理实体的元数据,由具有联系的、并具有顺序的元数据构

成,如事件时间、事件类型、事件描述、事件关系和触发事件等。

(5) 事件历史组:主要是关于文件的过去和曾作用于文件实体及其元数据的其他管理事件的元数据。对每一个事件,都由事件时间、事件类型、事件描述、事件关系以及处理的责任人员的元数据来说明。这类元数据是顺序记录每一个专门事件。元数据元素包括事件编号、时间、类型、事件描述和关系。

(6) 关系组:主要是指向关系实体并说明实体与其他实体之间关系的元数据,包括关系标识符、相关实体的标识符、关系类型和关系时间。

由于国际标准 ISO/23081 第二部分既提出了文件实体类元数据的框架,又提出了属性类元数据的框架,所以形成了一个很系统的元数据顶层框架体系(如图6所示),比起第一部分来,其框架更系统、更具体,在元数据顶层框架的标准化程度上更进了一步。

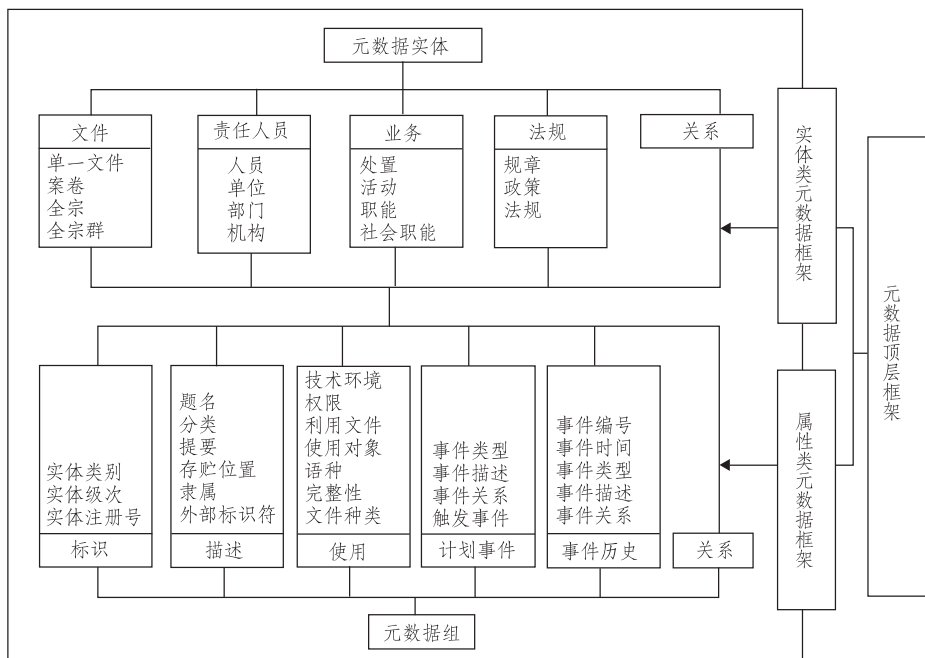


图6 国际标准 ISO/23081 第二部分元数据顶层框架体系示例

4 电子文件管理元数据标准顶层框架设计选择的基础性标准的比较

4.1 OAIS 参考模型的元数据顶层框架与国际标准 ISO/23081 第二部分元数据顶层框架的比较

如前所述,在 OAIS 参考模型的元数据顶层框架中,主要是两层元数据框架,其中第一层元数据框架是对数字对象进行描述的特征信息,或者说是信息对象在物理层面上的技术元数据,主要是提供与数字对象显示相关的技术手段和用户所用的操作设备方面的元数据,从而保证数字对象能够长久地得到存取。这一层元数据,实际上对应于国际标准 ISO/23081 第二部分的属性类元数据的使用组中的“技术环境”元数据,因为技术环境元数据是专门在文件实体的物理层面处理“格式信息、解码信息以及各种所需的技术”方面的元数据,也就是说技术环境元数据包括了使用文件实体所必需的相关技术环境的元数据。所以,正好与 OAIS 参考模型的元数据顶层框架中的第一层元数据框架相对应。

OAIS 参考模型的元数据顶层框架中的第二层元数据框架是信息对象的保存描述信息,或者说是信息对象在逻辑层面上的长久保存元数据,主要是关于信息对象长久保存所必需的信息,共包括四类元数据:参考信息、来源信息、背景信息和固化信息。

(1)参考信息。实际对应于国际标准 ISO/23081 第二部分的属性类元数据中的“标识组、描述组”元数据,因为参考信息是关于信息对象的标识符的元数据,其范围包括了“对象标识符、题名、版本号”等,而国际标准 ISO/23081 第二部分的属性类元数据中的“标识组”也是提供标识类的元数据,“描述组”则主要是提供如“题名、外部标识符”等元数据。

(2)来源信息。实际对应于国际标准 ISO/23081 第二部分的属性类元数据中的“事件历史组”元数据,因为它们都是描述对象内容历史的元数据类。

(3)背景信息。实际对应于国际标准 ISO/23081 第二部分的属性类元数据中的“关系组”元数据,因为这两类元数据都主要是处理关系的元数据。

(4)固化信息。这类信息主要是关于信息对象完整性的元数据,因此,其正好与国际标准 ISO/23081 第二部分属性类元数据的使用组中的“完整性”元数据相对应。

4.2 国际标准 ISO/23081 第二部分元数据顶层框架与 OAIS 参考模型的元数据顶层框架的比较

在国际标准 ISO/23081 第二部分元数据顶层框架中,第一层元数据框架是文件实体类元数据,其在 OAIS 参考模型的元数据顶层框架中没有相对应的部分。这一点不难理解,因为在 OAIS 参考模型中的元数据顶层框架是一个通用框架,它不可能对各个领域中的数字对象列出实体。不同的专业领域对数字对象有不同的实体划分,要列出一个满足各领域的通用的实体类元数据框架也不太现实。而在国际标准 ISO/23081 第二部分元数据顶层框架中,却列出了文件实体类元数据,这是因为在文件档案专业领域内对作为文件的实体划分有一致的理论认识与实践共识,所以也就可以列出对实体具有一致认识与共识的实体类元数据框架。

国际标准 ISO/23081 第二部分元数据顶层框架中的第二层元数据框架是属性类元数据,其在 OAIS 参考模型的元数据顶层框架中有相对应的部分:

(1)标识组元数据,对应 OAIS 参考模型元数据顶层框架的第二层元数据中的“参考信息”;

(2)描述组元数据,对应 OAIS 参考模型元数据顶层框架的第二层元数据中的“参考信息”;

(3)使用组元数据,对应 OAIS 参考模型元数据顶层框架的第一层元数据中的“表征信息”和第二层元数据中的“固化信息”;

(4)事件历史组元数据,对应 OAIS 参考模型元数据顶层框架的第二层元数据中的“来源

信息”;

(5)关系组元数据,对应 OAIS 参考模型元数据顶层框架的第二层元数据中的“背景信息”;

(6)计划事件组元数据,在 OAIS 参考模型

的元数据顶层框架中没有对应的部分。

国际标准 ISO/23081 第二部分元数据顶层框架与 OAIS 参考模型的元数据顶层框架相对应的模型如图 7 所示。

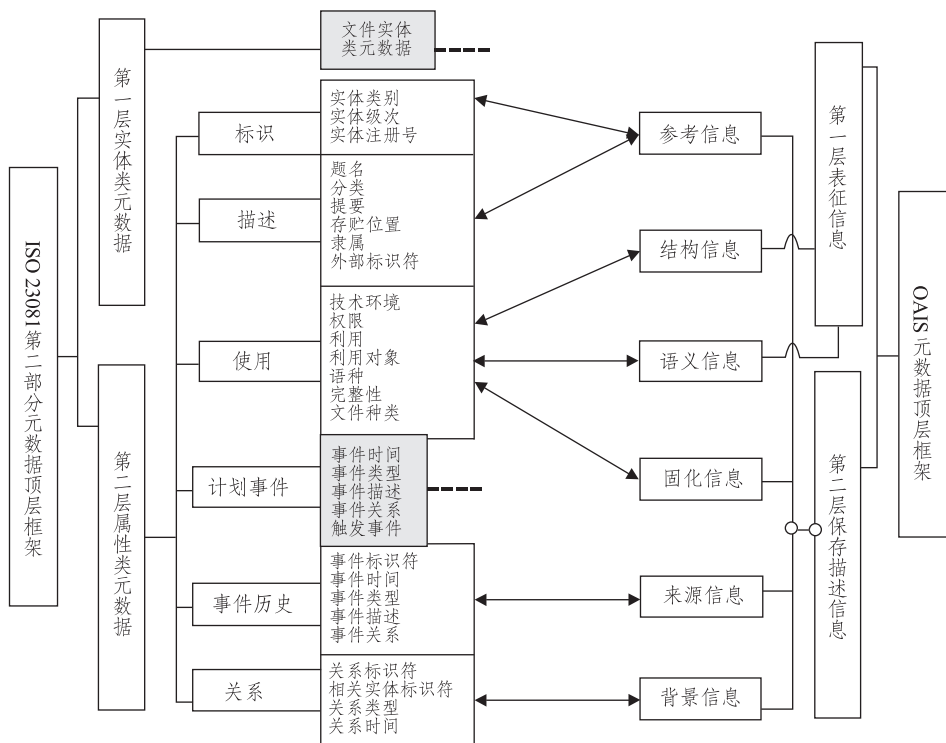


图 7 国际标准 ISO/23081 第二部分元数据顶层框架与 OAIS 元数据顶层框架的对应模型示例

从图 7 可以清楚看出:①国际标准 ISO/23081 第二部分元数据的顶层框架中的“文件实体类元数据、计划事件组元数据”是 OAIS 参考模型的元数据顶层框架中所没有的。②国际标准 ISO/23081 第二部分元数据的顶层框架中第二层属性类元数据包括了 OAIS 参考模型的元数据顶层框架所表述的各类元数据,或者说,OAIS 参考模型的元数据顶层框架所表述的各类元数据实际上就相当于国际标准 ISO/23081 第二部分元数据的顶层框架中第二层属性类元数据。③国际标准 ISO/23081 第二部分元数据的顶层框架比 OAIS 参考模型的元数据顶层框架更系统,因为在国际标准 ISO/23081 第二部分的

属性类元数据中,不仅规定了组的类别,而且还规定了每一组类别中的元素,从而形成了层、组(相当于类)、元素的三级架构;而 OAIS 参考模型的元数据顶层框架只有层和类,是二级架构。

5 结论

综上所述,从电子文件管理元数据标准的制定角度而言,在选择电子文件管理元数据标准顶层框架的基础性标准时,应该首先选择国际标准 ISO/23081 第二部分元数据的顶层框架,因为基于上述分析可以清楚地得出以下结论:

(1) 国际标准 ISO/23081 第二部分元数据

的顶层框架有电子文件管理理论的支持,因此是专为制定电子文件管理元数据标准而“量身定做”的国际标准。

(2) 国际标准 ISO/23081 第二部分元数据的顶层框架既具有实体类元数据,又具有属性类元数据,且属性类元数据又是三级的系统架构,因此从标准化角度看,采用国际标准 ISO/23081 第二部分元数据的顶层框架可以达到很高的标准化程度。

(3) 国际标准 ISO/23081 第二部分元数据的顶层框架是国际标准 ISO/23081 第一部分元数据原则的具体化,又是支持国际最权威的文件标准——ISO 15489 的元数据顶层框架,因此采用国际标准 ISO/23081 第二部分元数据的顶层框架具有明确的国际范围内普遍认可的电子文件管理的需求牵引。

(4) 国际标准 ISO/23081 第二部分元数据的顶层框架是一个宏观框架,从应用上讲,具有结合各国电子文件管理实践制定各国电子文件管理元数据标准的空间,因此在文件档案领域具有广泛应用性。

如果从目前国际标准化组织第 46 技术委员会—信息与文献工作技术委员会/第 11 分会—档案/文件管理分委员会的电子文件管理元数据工作组正在着手制定的有关电子文件管理元数据评估的国际标准和如何制定电子文件管理元数据标准的

核检表来看,遵循国际标准 ISO/23081 第二部分元数据的顶层框架可以得到国际电子文件管理元数据系列标准的支持,从而实现对电子文件管理元数据标准“需求→原则→实施→评估”整个过程的标准化,进而可以实现对电子文件管理元数据标准的整个生命周期的标准化管理。

参考文献:

- [1] 国际标准化组织. ISO 14721:2003 空间数据和信息传输系统—开放档案信息系统——参考模型[S],2003:32—60.
- [2] 国际标准化组织. ISO 23081—1 信息与文件—文件管理过程—文件元数据第一部分:原则[S],2006:11—13.
- [3] 国际标准化组织. ISO 23081—2 信息与文件—文件管理过程—文件元数据第二部分:概念与实施问题[S],2007:7—21.
- [4] Upward, Frank. Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond—a personal reflection[J]. Records Management Journal, Nov. 2001.

张正强 南京政治学院上海分院信息管理系教授,博士生导师。通讯地址:上海四平路 2575 号。邮编 200433。

(收稿日期:2008-08-07;修回日期:2008-10-08)