DOI:10.13530/j.cnki.jlis.172006

面向电子文件保存的统一元数据模型的构建*

刘越男　杨建梁

摘　要　元数据是电子文件管理的基本工具,在维护电子文件真实、完整、可用、可理解等方面的重要作用已经得到了广泛关注。目前,国际主流的与电子文件管理相关的元数据模型中,没有任何一个模型的设计初衷是面向电子文件保存的。与此同时,电子文件保存单位元数据管理的需求又在呼唤着统一模型的出现。本文以保护既有应用、支持持续管理、坚守专业原则、采用模块化设计思路等为原则,从业务逻辑、实体及其级次、实体关系等方面对 ISO 23081、PREMIS、PRONOM 模型予以分析、对比与整合,面向电子文件保存构建元数据模型,主要包括文件、技术环境、责任主体、业务、法规标准五个实体。该模型可以与现有主流模型建立映射,为支持文件保存机构设计元数据应用纲要提供统一的概念基础和体系框架。图 8。参考文献 29。

关键词　电子文件　数字保存　元数据模型　互操作

分类号　G270.7

Construction of a Unified Metadata Model for Electronic Records Preservation

LIU Yuenan & YANG Jianliang

ABSTRACT

Metadata is an essential tool for electronic records management and preservation; it has received high attention in maintaining the authenticity, integrity, usability of electronic records and enabling the understanding of information objects. Among the related metadata standards and guidelines, metadata models, which play the role of 'meta-standards', can provide a basic framework for both the authorities and the organizations to develop concrete schemes. The most widely used metadata models in records and archives management field include recordkeeping metadata conceptual model of ISO 23081, preservation metadata model of PREMIS and technical environment information model of PRONOM. But none of them initially serves the purpose of the long-term preservation of electronic records, no matter which metadata model is applied in records preservation organizations, some kind of refinement, expansion or adjustment is always needed.

This paper aims at building a unified metadata model for electronic records preservation through analyzing, comparing and integrating the metadata models of ISO 23081, PREMIS and PRONOM. As an

* 本文系 2013 年教育部新世纪优秀人才支持计划"电子文件元数据方案实施策略和方法研究"(编号:NCET-13-0574)的研究成果之一。(This article is an outcome of the project "Implementation Strategies and Methods of Metadata Schemes of Electronic Records" (No. NCET-13-0574) supported by Program for New Century Excellent Talents in University of Ministry of Education in 2013.)

通信作者:刘越男,Email:liuyuenan@ruc.edu.cn,ORCID:0000-0002-5216-2111(Correspondence should be addressed to LIU Yuenan, Email:liuyuenan@ruc.edu.cn, ORCID:0000-0002-5216-2111)

inclusive solution of the existing standards, this unified model provides a common conceptual framework for electronic records preservers to formulate metadata application profile and support metadata reuse between electronic records preservation system and electronic records management system.

The construction of unified metadata model starts from the actual demands of electronic records preservation by the way of modular design. By comparison the authors find that the records entity, people (agents) entity, business entity and mandates entity of ISO 23081 metadata model can be successively mapped to the object entity, agent entity, event entity and rights statement entity of PREMIS metadata model; furthermore, the records entity and the object entity can complement each other, so do the business entity and the event entity of these two models. The metadata model of PROMOM enriches the description of the technical environment of electronic records, which has great significance to digital information representation. Compared with PREMIS and PRONOM, ISO 23081 can better support the professional principles including the sustainable management of electronic records and their metadata throughout the records life circle, multi-level description of records and the integration of records and business. Therefore the authors lay the basis of architecture and terminology of the unified metadata model more on ISO 23081 while taking the advantages of the other two. The model is comprised of five entities including records entity, technological environment entity, agent entity, business entity and mandate entity. The records entity has multiple levels including archives collection, fond, series, file and item in managerial dimension, whereas in technical dimension the levels cover representation, computer file and bit stream. The technological environment entity refers to the systems and technologies used in creating, managing, preserving and rendering electronic records. The agent entity refers to the individuals, organizations or automated equipment taking part in creating, managing and utilizing electronic records. The business entity refers to the business of creating, managing and preserving electronic records. The mandates entity refers to the rules that normalize business execution and records access. The relationships between entities are set as attributes of the entities rather than an independent entity.

The unified metadata model for electronic records preservation constructed in this paper, in comparison to ISO 23081, enriches the content of records entity by integrating records preservation business with records creation business and recordkeeping business, and promotes the ability to meet long-term preservation requirements by adding a new technological environment entity. Compared with PREMIS, it turns the object entity into records entity and event entity into business entity, which better reflects the characteristic of records management; moreover, it makes the technological environment an independent entity from object entity. As for PRONOM, the technical components entity can be mapped to the technological environment entity of the unified metadata model, whereas the identifiers entity, intellectual property rights entity, documentation entity and actors entity can be mapped to the attributes of the technological environment entity of the unified metadata model. Some of the findings in this paper need further practical verification. 8 figs. 29 refs.

KEY WORDS

Electronic records.　Digital preservation.　Metadata model.　Interoperability.

0 研究背景

元数据是电子文件管理的基本工具,其在维护电子文件真实、完整、可用、可理解等方面的重要作用已经得到了广泛关注。制定合适的元数据方案,对各类组织机构的电子文件加以管理,成为理论和实践领域关心的核心问题之一。20世纪90年代以来,各级各类标准化组织、档案主管部门先后制定了多部电子文件元数据相关标准和指南,以指导文件形成单位和文件保存机构等实际部门制定具体的元数据方案,开展元数据管理。在这些标准规范中,规定元数据模型的标准可以视为元标准,元标准不仅可以指导实际部门制定元数据方案,而且可为各个国家、地区制定元数据标准或指南提供依据。

1 文献综述

目前已经颁布的与电子文件管理相关的元数据模型主要有:ISO 23081《信息与文献管理文件元数据》提出的文件管理元数据(Recordkeeping Metadata)概念模型(以下简称ISO 23081模型)①,PREMIS(Preservation Metadata:Implementation Strategies)规定的保存元数据(Preservation Metadata)数据模型(以下简称PREMIS模型)[1],ISO 14721:2003《空间数据与信息移交系统开放档案信息系统(OAIS)参考模型》中的信息模型(以下简称OAIS信息模型)[2],以及英国国家档案馆电子文件格式管理项目PRONOM制定的电子文件技术环境信息模型(以下简称PRONOM模型)[3]等。其中,ISO 23081模型主要面向文件形成单位的文件管理工作②,描述的是文件管理元数据及其相关关系。PREMIS模型、OAIS模型和PRONOM模型均面向档案的长期保存机构,前者描述的是数字信息长期保存所需元数据及其相互关系,中者从数据包的角度逐层揭示了描述信息对象的各类元数据及其相互关系,后者反映的是电子文件技术环境各类要素及其相互关系。

在研究层面,程妍妍在翻译的基础上,全面介绍ISO 23081模型的构成[4]。刘越男基于ISO 23081模型,按照模块化设计的思路,面向ERMS的实施,提出从实体、实体级次到具体元数据属性的元数据方案设计流程,肯定了ISO 23081模型的实用价值[5]。张正强通过详细比较OAIS信息模型的表征信息层、保存描述信息层与ISO 23081元数据模型的实体、属性,认为ISO 23081的顶层框架有电子文件管理理论的支持,更适合作为电子文件管理元数据顶层框架设计的标准[6]。钱毅认为,综合档案馆主导的电子文件中心,应参考ISO 23081模型设计元数据方案;由于这样的电子文件中心除了文件服务、归档管理之外,还承担电子文件长期保存的职能,所以还应根据PREMIS考虑保存元数据的需求;该研究暗示了ISO 23081和PREMIS结合的可能性[7]。程妍妍指出,ISO 23081模型描述的是电子文件凭证性相关的元数据,OAIS信息模型描述的是和长期保存相关的技术元数据,两者可

① ISO 23081共包括三个系列标准:ISO 23081-1:2006《信息与文献文件管理流程文件元数据第1部分:原则》,阐述了文件管理元数据的意义和作用、角色与职责、类型及其关系等原则性问题,该标准已被采纳为国家标准GB/T 26163.1-2010;ISO 23081-2:2009《信息与文献管理文件元数据第2部分:概念与实施》,着重阐述了元数据概念模型,以及元数据方案的制定与实施;ISO 23081-3:2011《信息与文献管理文件元数据第3部分:自我评估方法》,为文件形成机构评估其元数据方案符合文件管理需求的程度提供检查清单。本文阐述的ISO元数据模型及其详细解释主要见于ISO 23081-2:2009。

② 指从形成文件的业务系统中捕获文件、加以维护并在保管到期后加以处置的过程,这样的管理过程由专业化的电子文件管理系统[Electronic Records Management System,ERMS]支撑,该系统中管理的元数据即为文件管理元数据。

以结合起来共同解决电子文件长久凭证问题，但并未就如何结合加以阐释[8]。

在实践层面，ISO 23081 模型已被欧盟、澳大利亚、中国、美国佛蒙特州等多个国家和地区的档案主管部门采纳，并基于此制定了本土化的电子文件管理元数据标准，如欧盟先后于 2008 年、2010 年推出的 ERMS 标准《文件系统通用要求》MoReq2[9]、MoReq2010[10] 中所包括的元数据方案，澳大利亚国家档案馆制定的《澳大利亚政府机关文件管理元数据标准》(2015)[11]，我国国家档案局制定的档案行业标准《文书类电子文件元数据方案》(DA/T 46-2009)，美国佛蒙特州档案与文件局、信息与创新局联合发布的《公共机构文件管理元数据指南》(2008)[12]等。PREMIS 模型是在 OAIS 信息模型的基础上开发而成的，并对之进行延伸和扩展，可以将 PREMIS 模型看成是 OAIS 信息模型的具体化[13]。PREMIS 的正式发布使保存元数据实现了从理论标准到实践操作的升级[14]。美国档案与文件署的电子文件档案馆(ERA)[15]、加拿大国家图书档案馆的可信数字仓储(LAC-TDR)[16]等较有影响力的项目都应用了该标准。PRONOM 模型则主要应用在英国国家档案馆的技术登记系统中，该系统以及据此开发的文件格式识别工具 DROID 也被 ERA[17]、LAC-TDR[17]、加拿大温哥华市数字档案馆系统[18]、澳大利亚新南威尔士州数字档案馆[19]等多个项目应用。

然而，上述元数据模型的设计初衷均非直接面向电子文件长期保存。文件保存机构应用 PREMIS 时，需要根据档案文件的特点对之进行改造；若将 ISO 23081 从电子文件管理阶段延伸应用至电子文件长期保存阶段，也必须对 ISO 23081 模型进行扩展和调整；PRONOM 则只描述了电子文件的技术信息，未涉及其他元数据信息。钱毅、程妍妍等学者提出：为了更好地支持电子文件保存，需要将 ISO 23081、OAIS 和 PREMIS 进行结合。而目前尚缺乏具体的解决方案，这样的解决方案正是本文研究的核心。

本文的具体研究问题包括：①是否必要并可能在 ISO 23081、PREMIS、PRONOM 等已被广泛应用的元数据模型的基础上，通过衔接和整合，构建面向电子文件保存的统一元数据模型(本文也称之为“电子文件保存元数据模型”)；②如何综合上述三个模型的成果；③统一的电子文件保存元数据模型的框架、构成如何，是否可以和上述三个模型进行映射，从而既保护现有应用又指导未来发展。

本研究并非要在 ISO 23081、PREMIS、PRONOM 等现有标准之外建立新标准，也不面向具体的应用，而是有效衔接、整合这些在文件档案管理领域得到实际应用的元数据标准，建立具有包容性的解决方案。即通过对上述三大元数据模型的衔接和整合，构建面向电子文件保存的统一元数据模型，为文件保存机构(如档案馆)制定元数据应用纲要(Application Profile)提供统一的概念基础和体系框架，支持数字档案馆系统(也称数字保存系统、电子文件长期保存系统等)和 ERMS 之间的元数据复用，满足电子文件长期保存的实际需求。

2　研究思路和研究方法

本文的基本研究思路是：首先分析构建统一电子文件保存元数据模型的必要性和可行性；在获得肯定性答案之后，根据研究目的，归纳构建统一电子文件保存元数据模型的基本原则；鉴于新模型需要包容文件生命周期不同阶段的元数据模型，而 ISO 23081 模型与 PREMIS 模型在文件管理领域和数字保存领域各领风骚，因此这两大模型将是电子文件保存元数据模型最重要的基石，笔者分别从业务逻辑、实体、实体级次等方面比较两个模型，进而归纳统一模型的关键要素；在此基础上，综合 PRONOM 模型的特色，构建出统一的电子文件保存元数据模型。本文主要采用比较、归纳等定性分析方法，最后将有关研究发现集成。

3 构建电子文件保存元数据模型的必要性和可行性

3.1 必要性

(1)电子文件保存机构制定元数据应用纲要的需要

电子文件保存机构在制定本单位的元数据应用纲要时,可以从现有的一个或多个元数据元素集中选择自己所需的元数据元素,构成元数据方案。然而,文件保存机构按照什么样的思路选择元数据元素,如何判断需要哪些特定元数据元素,元数据模型的作用在这个时候就能显现出来,全面、系统的模型可为文件保存元数据方案提供自上而下的设计思路。而现实世界中电子文件保存元数据模型是缺位的,单独使用上述三大模型中的任何一个,都不足以支撑电子文件保存元数据的管理。

ISO 23081 的主要目的在于通过标准化的元数据框架支持文件管理与业务的集成,并维护电子文件的真实、完整、可用。虽然该标准指出数字档案馆系统的主要元数据来自 ERMS,标准化的元数据有助于提高系统互操作性。但是,它也承认,保存元数据和保存策略主要是由图情和档案管理领域负责。如果要以该模型作为文件保存元数据框架,势必要进行一定的拓展。以"业务"实体为例,业务实体主要包括形成文件的业务和文件管理业务两个方面,未涉及长期保存业务。我国档案行业标准 DA/T46 既面向文件形成单位又面向档案保存机构(档案馆),变相地将 ISO 23081 作为覆盖电子文件生命周期全程的框架模型。在"业务行为"元数据的"值域"中,以示例的方式列举了"接收、迁移"等长期保存行为,这样的规定已经超出了 ISO 23081 模型的范围。

PREMIS 模型由图情界开发,主要目的在于维护数字信息的生存能力(Viability)、可呈现性(Renderability)、可理解性(Understandability)等。一方面,由于其并非单独针对电子文件(档案)资源,故其设置的实体不能完全反映文件、档案的特征,术语也与文件、档案管理领域有一定的差异;另一方面,PREMIS 强调数字保存机构在应用该标准时需要同时整合其他既有元数据标准,它甚至没有覆盖数字信息长期保存所需要的所有元数据[20]。档案部门在应用的时候,需要补充其他元数据项目,比如 ERA 项目只用到部分 PREMIS 元数据;美国佛罗里达州数字档案馆在应用 PREMIS 时,需要对已经保存的信息进行整理并对应到 PREMIS 的元数据模型中,这是 PREMIS 应用的障碍之一[21]。

PRONOM 模型只描述电子文件技术环境信息,主要用以支持一类长期保存活动——电子文件格式管理。尽管其范围有限,却也引发了我们对于"技术环境"这个元数据实体的关注,思考其作为元数据模型中独立实体的可能性。事实上无论 ISO 23081 模型还是 PREMIS 模型,都没有将"技术环境"作为单独的实体。

调查显示,截至 2012 年底,我国不足一半的省级、副省级档案馆在接收电子文件的同时依据相关标准接收元数据,近八成的档案馆在电子文件接收进馆后不再形成元数据,无法支持电子文件持续、动态的管理[22],这与缺乏顶层设计框架或多或少有一定的关系。随着元数据知识的普及,文件保存机构对电子文件保存元数据模型的需求也越来越强烈。倘若将上述几个国际主流模型综合起来,加以衔接和整合,则有可能建立一个可以满足长期保存过程中电子文件持续管理需求的元数据模型,从而为实践部门提供参考。如图 1 所示。

(2)数字档案馆系统复用元数据的需要

电子文件的形成、管理和保存是个持续的过程,保存并非起始于保存单位对电子文件的接收(保存阶段),而是发端于电子文件的形成阶段,保存与形成、管理相集成,贯穿于电子文件生命周期的全过程[23]。电子文件形成、管理的质量直接影响电子文件保存的质量。电子文件保存元数据中的很大一部分,来自其形成和管理系统。维护电子文件的长久真实、完整、可

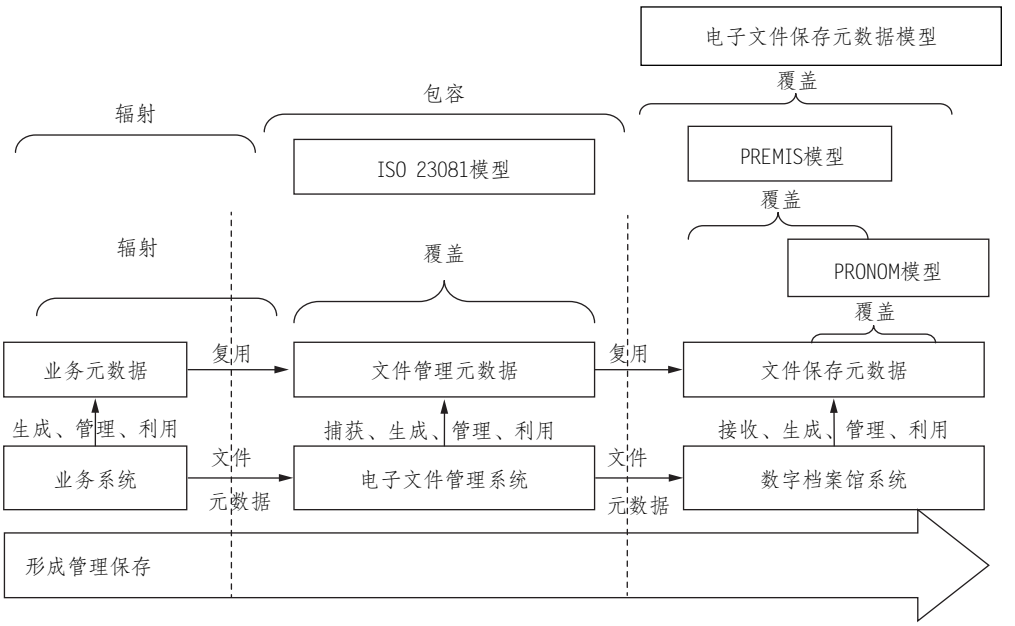

图 1　电子文件生命周期中的系统、元数据及元数据模型

用和安全，离不开其生命周期中元数据的持续积累和恰当应用。高效的电子文件保存，要求元数据在相应系统中准确生成，在不同系统之间无缝传递，即元数据在其源头一次性捕获，在后续管理系统中重复使用。有研究指出，跨环境（包括跨业务、跨系统、跨空间等）的元数据复用对于文件管理和支撑业务流程有着至关重要的意义[24]。

澳大利亚莫纳什大学在 2006 年发起了一项智能文件管理元数据研究项目（Clever Recordkeeping Metadata Research Project，CRKM），该项目以解决业务系统、ERMS 以及数字档案馆系统之间的元数据互操作问题为己任，以便在这些系统之间最大程度地复用元数据，而不是重新创造元数据。CRKM 提出了“元数据中介（Metadata Broker）”的概念，在面向服务系统（SOA）环境中，将元数据中介以中间件的形式嵌入到系统中，对不同系统的元数据方案进行登记，并在不同的元数据方案之间进行自动化的翻译与转换，可以实现业务系统、ERMS 以及数字档案馆系统元数据的互操作[25]。从长远来看，这种方式比传统接口方式更能降低系统互操作的成本。

数字档案馆系统需要复用来自电子文件管理系统、业务系统的元数据，对电子文件生命周期不同阶段元数据模型（ISO 23081 和 PREMIS 等）的整合提出了要求。我们认为，如果数字档案馆系统采用的元数据模型能够包容 ERMS 的元数据模型，并对业务系统元数据的规范化生成提出要求，元数据实体类型、属性实现无缝衔接，则可以简化元数据中介的设计和应用，进一步降低元数据复用的成本。

3.2　可行性

（1）既有模型相互补充且得到广泛应用

尽管设计初衷不同，但是 ISO 23081、PREMIS 和 PRONOM 三个元数据模型并非彼此对立，而是相互补充，服务于电子文件不同生命阶段中的不同管理目的。ISO 23081 服务于 ERMS 中元数据的捕获和管理，并反向辐射形成电子文件的业务系统，对业务系统元数据的生成提出要求；PREMIS 服务于数字档案馆系统中

信息的长期可展示和可理解,从而保证数字信息的生存能力;PRONOM 则通过对电子文件技术环境的细致描述,为使用合适方式保证其长期可展示性提供支持。管理目的互补性降低了在不同模型之间选择的难度,从而将模型构建的任务转化为模型整合。

此外,上述三个元数据模型已在电子文件管理与长期保存实践中得到较为广泛的应用,属于国际主流的元数据模型,在此基础上构建的统一模型,与现有元数据模型有着较好的互操作性,易于实践应用。

(2)基于实体描述电子文件元数据及其相互关系

本文重点研究的三大模型,都采用"实体"这个基本单元来描述电子文件元数据及其相互关系。ISO 23081 认为实体是指"任何现在、过去或是未来可能存在的具体或抽象事物,包括这些事物间的相互关系"。我们可以将实体理解为元数据描述的对象类型。各模型描述方式的一致性在操作层面保证了模型整合的可行性。

ISO 23081 模型由文件(Records)、责任主体(Agent)、业务(Business)、法规标准(Mandate)、关系(Relation)五大实体组成,如图 2 所示。其中文件是组织机构业务活动的记录;责任主体即从事业务活动的人员或者组织机构;业务包括形成文件的业务和文件管理业务,两者是相互集成的;法规标准指约束业务执行和记录方式的规则;关系即上述各实体之间的关系,这是一类特殊的实体。

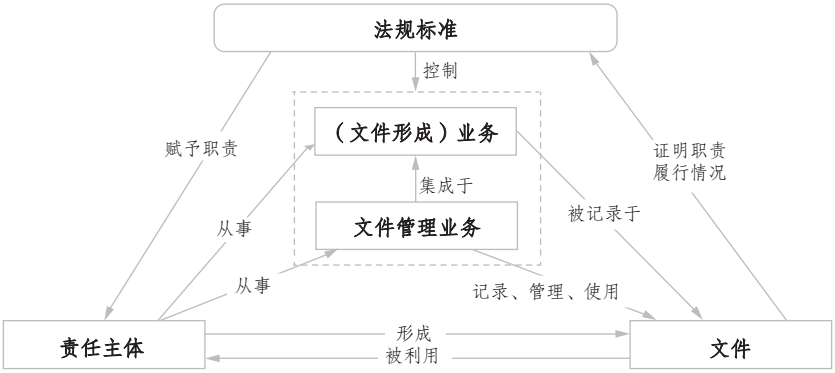

图 2 ISO 23081 文件管理元数据概念模型的主要实体及其关系

PREMIS3.0 版数据字典所建立的数据模型,明确了四类实体:对象(Object)、事件(Event)、责任主体(Agent)和权利声明(Rights Statement)。对象是指被保存的数字信息的离散单元,包括智能实体、表征、文件和比特流等层次,其中环境(Environment)是指在某种程度上支持数字对象的软硬件技术,在数字保存系统中,环境被作为智能实体加以描述,并作为表征、文件或比特流被捕获并保存。事件是指至少涉及或影响数字保存系统中一个对象或一个责任主体的行动。责任主体是指在对象生命周期中与事件或权利相关的个人、组织或软件程序。权利声明是指对对象、责任主体一项或多项的权限声明。实体之间的关系如图 3 所示。

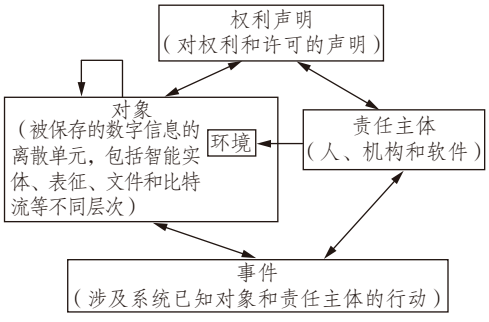

图 3 PREMIS 3.0 版本数据字典中的数据模型

PRONOM 信息模型包含技术组件(包含文件格式、软件、存储介质、硬件组件四个子实体)、标识符、行为者、知识产权和文档五大实体。其中,技术组件实体描述电子文件生成、读取与管理所需的技术环境,标识符指可描述其他实体的外部标识符,行为者指对另一实体执行某种已定义动作的个人或组织机构,知识产权指对另一实体适用的知识产权信息,文档指与另一实体相关的文档。如果某实体有从自身出发并指回自身的箭头,说明此实体与同类型实体之间存在着可描述关系;如果某实体有从自身出发并指向其他实体的箭头,说明此实体与所指向的实体之间存在着可描述关系[26]。从图 4 可以看出,技术组件是其最核心的实体,其他实体围绕着技术组件而非文件本身展开。因此,笔者在综合分析的时候,主要针对技术组件实体,而忽略其他实体。

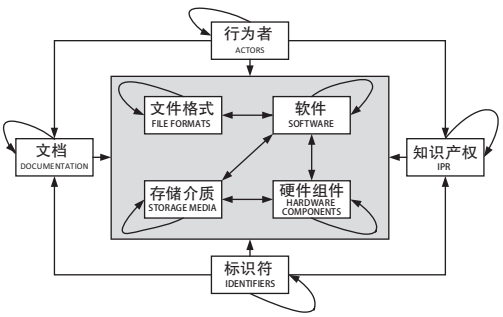

图 4　PRONOM 信息模型

4　构建电子文件保存元数据模型的基本原则

(1)保护既有应用

本文构建统一电子文件元数据模型的首要原则就是保护模型的既有应用,即不推翻各国、地方、机构对相关模型的现有应用,不打破重来,而是在已有模型的基础上,进行有序整合和适度完善。一方面为现有应用之间的元数据转换、映射、复用提供统一的概念基础;另一方面为未来业务系统、ERMS、数字档案馆系统之间的元数据转换、映射、复用降低成本,提高效率。

(2)支持持续管理

数字档案馆系统对元数据的复用具有很强的特殊性:复用元数据的系统之间是顺序衔接的关系,而不是平行关系,数字档案馆系统需要依次从 ERMS 和业务系统中接收电子文件元数据,将其作为保存元数据的一部分,元数据复用在很大程度上表现为元数据的继承。这就要求保存元数据模型能够包含文件管理元数据模型,即本文所构建的统一元数据模型要包含 ISO 23081 模型,以支持电子文件生命周期中元数据的持续管理。

(3)坚守专业原则

面向电子文件保存构建统一的元数据模型,应始终坚持文件、档案管理的专业原则。笔者认为,在元数据领域,至少应该坚持两大原则:一是多级著录原则;二是文件与业务的集成原则。多级著录是《国际标准——档案著录规则(总则)》[ISAD(G)][27]确立的基本原则,它明确档案著录必须由总到分进行,即依次对全宗、类目、案卷、单份文件进行描述,且以等级结构形式对不同级别的著录结果加以联接。多级著录是全宗原则在著录领域的具体要求,是维系文件之间有机关联的手段。集成是电子文件管理的基本原则之一,其中文件与业务的集成是核心[28]。国家标准《信息与文献 文件管理第 1 部分:通则》(GB/T 26162.1-2010,采标自国际标准 ISO 15489-1:2001)也肯定了这一原则。

ISO 23081 模型中,文件、责任主体、法规标准、业务四个实体都具有多个层次。其中文件实体涉及档案集合、全宗、系列、案卷、业务流、文件等层级,责任主体实体包括机构、部门、工作组、个人等层级,业务实体包括联合职能、职能、活动、事务等层级,法规标准实体包括法律法规、政策、业务规则等层级,可以有效支持档案多级著录的原则,并扩展了多级著录的范畴[29]。

作为 ISO 15489 的配套标准,ISO 23081 亦坚持文件管理和业务的集成,突出体现在将业

务实体视为形成文件、管理文件两类业务的集成,从概念模型的层面就明确了 ERMS 应该捕获来自业务系统、描述文件形成业务以及来自 ERMS 自身、描述文件管理业务的元数据。正是因为对业务的全面描述,元数据能够完整地说明文件的来龙去脉,从而维护电子文件的核心价值——真实性(Authenticity),这也是电子文件元数据所具有的一大独特作用。

基于上述原因,笔者在构建统一元数据模型时,更多地着眼于 ISO 23081 的框架。

(4)遵循模块化设计思路

上文指出,现有模型都采用实体的方式来描述元数据及其关系。元数据模型所支持的管理需求,来自电子文件形成、管理和保存的业务流程,而这样的业务流程呈现出复杂的时空结构,很难直接应用于模型设计。将复杂的电子文件管理需求抽象简化为概念实体及实体间的关系,则可以自上而下地构建元数据模型,符合统一元数据模型顶层设计的定位;可以实现元数据的模块化设计与应用,各实体的元数据各司其职,减少冗余元数据项;基于实体的元数据管理目标明确,元数据框架的层次也更为清晰,逻辑严谨。此外,模块化设计还有助于实际单位灵活扩展元数据,方便利用(见图 5)。

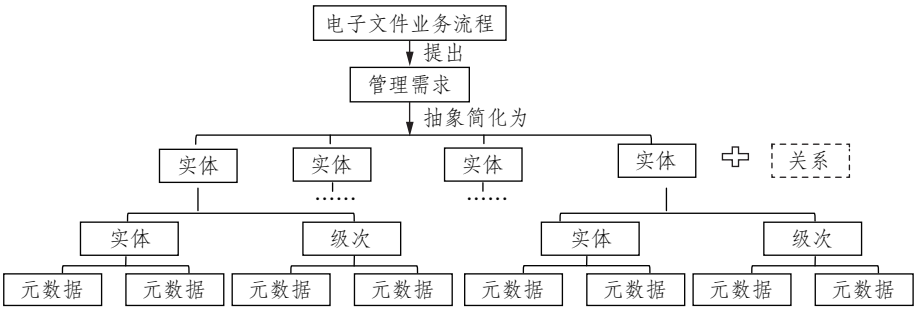

图 5 基于实体的模块化设计思路

5 既有模型的比较与整合

鉴于 PRONOM 模型只描述了电子文件的技术环境,不涉及电子文件(数字信息)及其管理的其他内容,覆盖面与 ISO 23081 模型和 PREMIS 模型不对等,因此本部分主要比较后面两个模型。仅在 5.2 部分对技术环境元数据的讨论中,参照了 PRONOM 模型。

5.1 业务逻辑

虽然概念实体及其关系不能直接反映业务流程,但却可以反映出业务逻辑。经分析,我们认为,ISO 23081 模型的业务逻辑是:在法规标准的约束下,责任主体在文件形成业务中形成文件,在文件管理业务中管理文件,这两类业务相互集成。PREMIS 模型的业务逻辑是:经权利声明授权或许可,责任主体(在技术环境中)管理和利用①数字对象。

可以发现,这两个模型都采用了“规则—主体—行为—对象”的逻辑表达方式。在电子文件管理领域,两者可以整合为:在法规标准(包括权利声明)的约束下,责任主体基于技术环境,在文件形成业务中形成文件,在文件管理业务中管理文件,在文件保存业务中保存文件,在整个业务过程中利用文件。

5.2 实体及其级次

(1)管理对象实体

在 ISO 23081 模型中,管理对象是文件实

① 此处“管理和利用”是对“事件”的释义。

体，包括档案集合、全宗、系列、案卷、业务流、文件等层级。而 PREMIS 模型中，管理对象是对象实体，包括智能实体（Intellectual Entity）、表征（Representation）、文件（File，指计算机文件）和比特流（Bit Stream）四个层级。其中，智能实体是指需要数字保存的一个智力作品或艺术创造，比如一本书、地图、照片、数据库等；表征是指完整呈现一个智能实体所需的一系列文件以及表达其结构的元数据，比如构成一篇期刊论文的一个 SGML 文件和两个图像文件；文件指为计算机系统理解的数字信息；比特流是指文件内部一段连续的或不连续的数据。如图 6 所示，智能实体、表征、文件和比特流可以是单个对象，也可以是聚合对象。

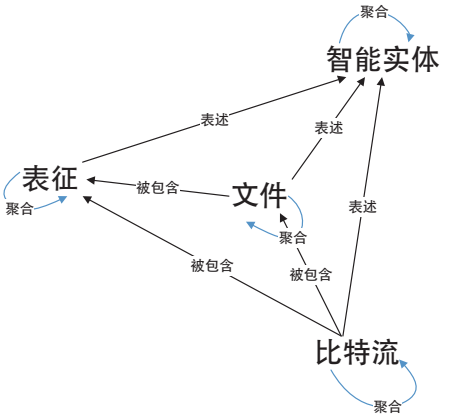

图 6　PREMIS 管理对象层次关系

ISO 23081 模型中文件实体的层级是文件管理单元的层级，揭示的是文件与文件之间的业务联系。而 PREMIS 模型中对象实体的层次是技术单元的层次，揭示的是保管对象之间的技术联系。从智能实体到比特流，是从技术层面上对保管单元自上而下的分解。技术环境也被作为一类智能实体，更鲜明地体现出“对象”实体的技术特征。我们认为，只有智能实体这个层次可以和 ISO 23081 的文件实体直接映射，单个智能实体和单份文件映射，智能实体的聚合和文件集合映射；表征、（计算机）文件和比特流皆可以看成是文件实体的技术构成。这样的映射关系也提醒文件、档案管理人员关注电子文件在技术层面的构成要素，数字环境下需要细化管理的颗粒度，在表征、（计算机）文件和比特流等层次开展必要的元数据管理工作。就元数据模型而言，文件实体的层次可以进一步丰富。

（2）责任主体实体

在 ISO 23081 模型中，责任主体实体是在开展业务过程中形成、管理和利用文件的责任人员或机构。在自动化环境下，自动生成和利用数据的设备也可能是一种特殊类型的责任主体。在 PREMIS 模型中，责任主体是指在对象生命周期中与事件或权利相关的个人、组织或软件程序。可以看出，在责任实体的概念上，ISO 23081 与 PREMIS 有较强的一致性，都包含个人和组织机构，也指出了自动化环境下设备、软件程序作为责任主体的可能性。出于文件管理的需要，ISO 23081 模型对组织机构责任者层次划分更为详细，包含机构、部门、工作组等。

（3）业务实体

ISO 23081 模型的业务实体包括形成文件的业务和文件管理业务，两者相互集成。PREMIS 的事件实体指至少涉及或影响数字保存系统中一个对象或一个责任主体的行动。在概念上 PREMIS 的事件实体比 ISO 23081 更加宽泛，这意味着 ISO 2308 的业务实体与 PREMIS 的事件实体相互映射，可以加以融合。鉴于电子文件元数据持续形成的特点，电子文件元数据的业务实体在文件形成业务、文件管理业务之外，还应集成文件保存业务，对应到 PREMIS 即是文件形成事件、文件管理事件和文件保存事件。

（4）权利声明实体与法规标准实体

ISO 23081 法规标准实体是约束业务执行和记录方式的规则，包括法律法规、政策、业务规则等层级。而 PREMIS 的权利声明实体是指对对象、责任主体的一项或多项权限的声明。PREMIS 的权利声明实体可以和 ISO 23081 的法规标准实体建立间接的映射关系，这是因为对于电子文件长期保存来说，关于文件秘密程度、知识产权保护、隐私权保护、访问权限等方面的文

件利用法规标准是最重要的业务规则之一,如果将法规标准限定在利用类法规标准的范围内,两者是一致的。但是从电子文件全程管理来说,法规标准并非仅仅包括利用类法规标准,还包括文件生成、管理的规则,如保管期限规定等,因此ISO 23081的法规标准实体更具包容性。

(5)对技术环境元数据的处理

数字信息的形成、管理、展示和利用都离不开技术环境的支持。ISO 23081模型并不直接包含技术环境,在分析各实体的通用属性元数据时,将技术环境元数据作为"利用"属性元数据的一个构成提出来,除了技术环境之外,"利用"属性元数据还包括权利、存取、受众、语言、完整性、成文方式等。可见,技术环境在ISO 23081中并没有得到额外的重视,这或许与该模型主要服务于文件管理工作有关。相比而言,PREMIS模型更加强调数字信息的软硬件技术环境,PREMIS 3.0数据词典中明确指出,和数字对象显示、呈现相关的软硬件技术环境可以被当作智能实体加以管理。笔者以为,在技术层面,技术环境固然可以被当作智能实体,但是这样处理的结果就是混淆了数字对象及其技术环境的区别。

PRONOM模型指出,电子文件技术环境不仅包括通常意义上的软件、硬件,还包括格式、软件、存储介质、硬件组件等多个要素,每个要素都可能影响到电子文件的长期可用性。在构建元数据模型的时候,不应忽略"技术环境"这个概念实体。

5.3 实体关系

ISO 23081模型中,关系实体表达其他各实体之间的关系,是一类特殊的实体。虽然PREMIS模型中没有直接将关系作为一个实体,但是在PREMIS 3.0的数据字典中,详细解释了两个数字对象之间可能的关系类型,具体包括包含和被包含的结构关系、文件格式转换过程中出现的派生关系;在实体关系方面,PREMIS认为责任主体与权利声明、管理对象、事件,权利声明与管理对象、责任主体会存在关系,并在各个实体上集成了关系元数据。从实施的角度来讲,由于关系本身错综复杂,不宜单独作为一个实体,笔者更倾向于PREMIS的处理。当然,ISO 23081也指出,可以将"关系"作为实体的一个通用属性加以实施。

6 面向电子文件保存的统一元数据模型

通过上文分析,可以发现:ISO 23081模型和PREMIS模型在业务逻辑、实体及其级次、实体关系方面既有共通之处,亦各有所长。其管理对象实体、责任主体实体、业务实体、权利声明和法律法规实体均可以相互映射,在管理对象实体、业务实体等方面还可以互相补充和融合。PRONOM模型则在"技术环境"实体方面给我们以思考和启发。

鉴于ISO 23081模型更能体现文件管理的特点,我们在构建统一元数据模型的时候,在整体框架结构和术语层面,将其作为主要参考。与此同时,笔者认为,技术环境是数字环境下文件的生存环境,故主张将技术环境作为单独实体,与文件、责任主体、业务、法规标准四个实体一起共同构成统一的元数据模型,如图7所示。其中文件实体在管理层面包括档案集合、全宗、类目、案卷、文件等层级,在技术层面包括表征、(计算机)文件、比特流等层级;技术环境实体是指形成、管理、展现电子文件信息的系统和技术;责任主体实体是指形成、管理和利用文件的个人、组织机构或自动化设备;业务主体包括形成、管理和保存文件的业务;法规标准主体是指约束业务执行和记录方式的规则,包括文件利用的规则。在实体关系的处理上,我们将"关系"当作实体的属性,而不单独设为实体。

面向电子文件保存的统一元数据模型,是衔接、整合ISO 23081、PREMIS、PRONOM等模型的桥梁,应具备与现有模型的兼容性和互操作性,能够保证数字档案馆系统最大程度地复用业务系统和ERMS中的元数据,也支持数字

档案馆系统直接复用外来数据库，如链接 PRONOM 数据库中的技术信息。图 8 揭示了统一电子文件保存元数据模型的实体与既有元数据模型的实体(子实体)的映射关系。

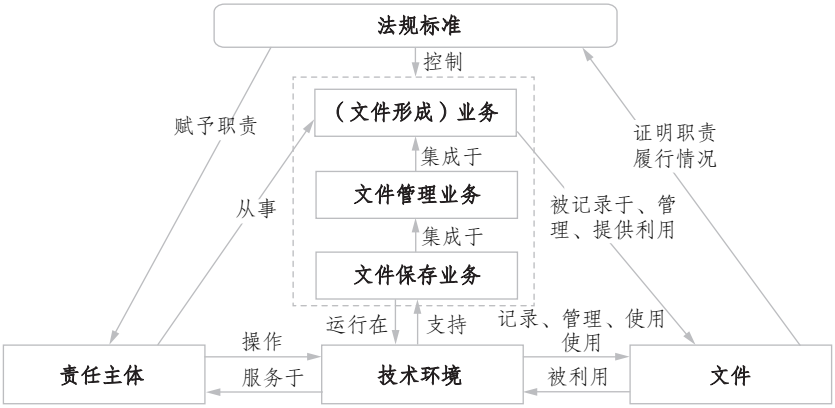

图 7　面向电子文件保存的统一元数据模型

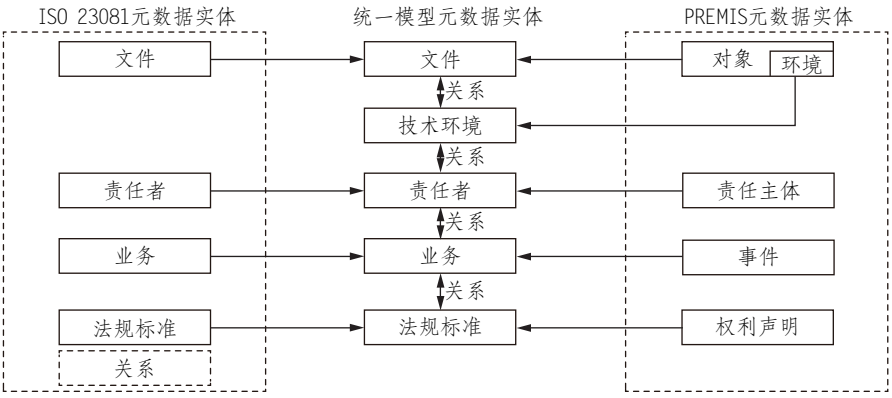

图 8　统一模型元数据实体与其他模型的映射关系

7　小结

本文从电子文件全程管理的需求出发，面向电子文件保存，基于 ISO 23081 模型、PREMIS 模型、PRONOM 模型中实体的相互映射、融合和互补，尝试构建统一的元数据模型，包括文件、技术环境、责任者、业务和法规标准五个实体。相比 ISO 23081 模型，统一模型丰富了业务实体的内涵，增加了技术环境实体，统一模型增加了对长期保存需求的应对能力；相比 PREMIS 模型，将“对象”“事件”实体，改造为更符合文件管理要求的“文件”“业务”实体，并将“技术环境”从“对象”实体中抽离出来；相比 PRONOM 模型，后者的“技术组件”实体可以直接和统一模型中的“技术环境”实体映射，原模型中“标识符”“行为者”“知识产权”和“文档”则可映射为统一模型中“技术环境”实体的属性。

在后续的研究中，可以持续对所提出的统一模型进行实践验证和改进，并基于该模型提出实施指南。不同类型的行业、不同类型的组织机构，也可以结合其特征，建立更具针对性的保存元数据方案和元数据应用纲要。

参考文献

[1] PREMIS Editorial Committee. PREMIS data dictionary for preservation metadata version 3.0[EB/OL]. [2016-12-07]. http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf.

[2] CCSDS. Reference model for an open archival information system (OAIS) [S/OL]. [2016-12-07]. http://public.ccsds.org/publications/archive/650x0m2.pdf.

[3] The National Archives. PRONOM 4 information model [EB/OL]. (2005-01-04)[2016-12-07]. http://www.nationalarchives.gov.uk/aboutapps/fileformat/pdf/pronom_4_info_model.pdf.

[4] 国际标准 ISO 23081元数据模型分析[J]. 程妍妍,编译.现代图书情报技术,2008(9):31-35.(ISO 23081 Metadata model analyses[J]. Cheng Yanyan,trans.New Technology of Library and Information Service,2008(9):31-35.)

[5] 刘越男,梁凯,顾伟. 电子文件管理系统实施过程中元数据方案的设计[J]. 档案学研究,2012(2):56-64.(Liu Yuenan,Liang Kai,Gu Wei. Metadata schema design in the implementation of electronic records management system[J]. Archives Science Study,2012(2):56-64.)

[6] 张正强. 论电子文件管理元数据顶层框架设计的标准化[J]. 中国图书馆学报,2009,35(2):12-19.(Zhang Zhengqiang. On the standardization of top-level framwork design for records mamagement metadata[J]. Journal of Library Science in China,2009,35(2):12-19.)

[7] 钱毅. 论电子文件中心元数据方案的管理策略[J]. 档案学通讯,2012(6):76-79.(Qian Yi. Management strategy for metadata schema of electronic records center[J]. Archives Science Bulletin,2012(6):76-79.)

[8] 程妍妍. 电子文件管理元数据模型研究[J]. 浙江档案,2008(6):38-41.(Cheng Yanyan, Research on metadata model for electronic records management[J]. Zhejiang Archives,2008(6):38-41.)

[9] Serco consulting,appendix 9 to the MoReq2 specification:metadata model (version 1.04) [EB/OL]. (2008-09-08) [2011-06-29]. http://www.dlmforum.eu.

[10] DLM forum foundation,MoReq2010 modular requirements for records systems,Volume 1 Core services &plug-in modules Version 1.0 [EB/OL]. [2011-06-29]. http://www.dlmforum.eu.

[11] National Archives of Australia. Australian government recordkeeping metadata standard (AGRkMS,Version 2.2) [EB/OL]. [2016-07-08]. http://www.naa.gov.au/Images/AGRkMS-Version-2.2-June-2015_tcm16-47131.pdf.

[12] Vermont State archives & records administration, Vermont department of information &innovation [EB/OL]. (2008-10-01) [2016-12-08]. https://www.sec.state.vt.us/media/67296/MetadataGuideline2008.pdf.

[13] 高嵩,张智雄. PREMIS 保存元数据体系分析[J]. 现代图书情报技术,2006,1(4):19-23.(Gao Song,Zhang Zhixiong. A study on PREMIS preservation metadata framework[J]. New Technology of Library and Information Service,2006,1(4):19-23.)

[14] 刘建华,张智雄. 保存元数据的发展趋势研究[J]. 图书馆杂志,2016(6):10-16.(Liu Jianhua,Zhang Zhixiong.Study on the trend of metadata preservation[J]. Library Journal,2016(6):10-16.)

[15] NARA. ERA status and accomplishments [EB/OL]. (2016-08-30) [2016-12-07]. http://www.archives.gov/era/about/status-accomplishments.html.

[16] Armstrong P. Library and archives Canada:towards a trusted digital repository [EB/OL]. (2008-01-07)

[2016-12-08]. http://archive.ifla.org/IV/ifla74/papers/084-Armstrong-en.pdf.

[17] NARA. Contributes to improved digital records preservation and access system[EB/OL]. (2016-08-15)[2016-12-07]. https://www.archives.gov/press/press-releases/2011/nr11-31.html.

[18] Dingwall G. City of Vancouver Archives digital preservation program[R]. UBC, Vancouver, 2015.

[19] The State Records Authority of New South Wales. Digital archives migration methodology[EB/OL]. [2016-12-07]. https://www.records.nsw.gov.au/recordkeeping/advice/digital-archives-migration-methodology.

[20] Caplan P. Understanding PREMIS [EB/OL]. (2009-02-01)[2016-12-07]. http://loc.gov/standards/premis/understanding-premis.pdf.

[21] Donaldson D R, Conway P. Implementing PREMIS: a case study of the Florida Digital Archive[J]. Library Hi Tech, 2010, 28(2): 273-289.

[22] 刘越男，祁天娇. 我国省级、副省级档案馆电子文件接收及管理情况的追踪调查[J]. 档案学通讯，2014(6)：10-15.（Liu Yuenan, Qi Tianjiao. Continuous investigation on the ingest and management of electronic records in provincial and vice-provincial archives[J]. Archives Science Bulletin, 2014(6): 10-15.）

[23] 谢丽. 数字文件管理：数字保存不可或缺的基础[R]. 第四届中国电子文件管理论坛. 北京：中国人民大学，2013.（Xie S. Digital records management: indispensable foundation of digital preservation[R]. the 4th China Electronic Records Management Forum. Beijing: Renmin University of China, 2013.）

[24] Evans J, Mckemmish S, Bhoday K. Create once, use many times: the clever use of recordkeeping metadata for multiple archival purposes[J]. Archival Science, 2005, 5(1): 17-42.

[25] Monash University. Final report of the CRKM project [EB/OL]. [2016-12-08]. http://www.infotech.monash.edu.au/research/groups/rcrg/crkm/docs/rpt-final.doc.

[26] 韩若画.英国电子文件格式管理项目 PRONOM 研究[M]//刘越男，马林青.2010—2015电子文件发展及前沿报告.北京：电子工业出版社，2016：232-254.（Han Ruohua. The electronic records format management project: PRONOM of United Kingdom[M]//Liu Yuenan, Ma Linqing. Report on electronic records management development and advancement (2010-2015). Beijing: Publishing House of Electronics Industry, 2016: 232-254.）

[27] ICA.ISAD(G): General international standard archival description [S/OL]. [2016-12-08]. http://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf.

[28] 冯惠玲. 政府电子文件管理[M]. 北京：中国人民大学出版社，2004：35-36.（Feng Huiling, Governmental electronic records management[M]. Beijing: China Renmin University Press, 2004: 35-36.）

[29] 刘越男. ISO 23081带来的启示与困惑[J]. 北京档案，2008(7)：26-29.（Liu Yuenan. Inspiration and confusion of ISO 23081[J]. Beijing Archives, 2008(7): 26-29.）

刘越男　中国人民大学信息资源管理学院教授，数据工程与知识工程教育部重点实验室研究员，博士生导师，副院长。北京 100872。

杨建梁　中国人民大学信息资源管理学院博士研究生，数据工程与知识工程教育部重点实验室研究助理。北京 100872。

（收稿日期：2016-12-16；修回日期：2017-02-11）