长期保存存档信息包的研究与构建

付鸿鹄　吴振新　王玉菊

摘　要　在对长期保存存档系统模型(OAIS)定义的存档信息包(AIP)模型及可用的相关标准JATS标签集、PREMIS、METS等详细分析的基础上,对现有标准进行集成,构建了一个比较理想的AIP应用模型。结合该应用模型,对长期保存中AIP构建实例进行分析和说明,总结在保存实践活动中构建AIP应考虑的几个主要问题,包括技术标准的成熟性和可扩展性、AIP本身的可扩展性和兼容性、AIP主要功用的影响、格式迁移策略等。图7。参考文献18。

关键词　数字资源　长期保存　存档信息包　元数据　标准规范

Research and Construction on AIP in Digital Preservation

Fu Honghu　Wu Zhenxin　Wang Yuju

Abstract: In this paper, the definition of OAIS Archival Information Package (AIP) model and related standards such as JATS, PREMIS, METS have been analyzed. An ideal AIP application model has been formed through the integration of existing standards. Based on the application model, the AIP construction in digital preservation has been discussed as an example. Several key issues that should be considered in AIP constructing are summarized, including the maturity and scalability of standards, scalability and compatibility of AIP, the main function of AIP, and format migration policy. 7 figs. 18 refs.

Keywords: Digital Resources; Digital Preservation; Archival Information Package; Metadata; Standard Specification

1　引言

数字资源长期保存的目标是对数字化信息进行长期保存以保证在未来长期可用,同时确保被保存的信息对于目标用户而言是独立可理解的。随着计算机技术飞速发展,存储介质、操作系统、软件平台等数字信息所严重依赖的技术环境都在发生着快速变化。除了技术因素,长期保存目标用户的知识库(Knowledge Base of the Designated Community,即长期保存服务目标用户群的知识背景)也在不断的变化。如何确保存档的信息在技术环境和用户知识发生变化后还能够被正确呈现、理解和使用,是长期保存必须解决的问题。长期保存系统必须要保存比被保存的信息对象本身的内容多得多的信息,才有可能保证被保存对象的长期可用性,这也是长期保存系统区别于其他信息系统的重要特征。

成功保存信息对象的关键是保存系统能够明确识别和理解数据对象以及相关联的必要信息,并对其进行有效的组织和维护,这些给数字信息的保存带来了重大挑战。作为长期保存的基本存档单元,存档数据包(Archival Information Package,AIP)的组织和构建一直都是保存领域一个非常重要的研究主题,也是保存实践首先要解决的一个关键问题。

本文基于长期保存存档系统模型OAIS[1],结合相关技术标准和应用实例,对AIP的组成和构建进行详细深入的分析和探讨。

2　OAIS中AIP的相关定义及描述

在OAIS参考模型中定义了三种信息包,提交信息包(Submission Information Package,SIP)、存档信息包(Archival Information Package,AIP)和分发信息包(Dissemination Information Package,DIP)。存档信息包(AIP)是长期保存系统中用于存储保存的信息包。

信息包是OAIS中的核心概念,作为功能模

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

型中各模块之间传递的基本对象，它是一个概念化的容器，包含了两种类型的信息：一种是存档对象本身，即内容信息；另一种是长期保存需要的相关信息，即保存描述信息（Preservation Description Information，PDI）。OAIS 要求 PDI 与内容信息封装在一起保存，如图 1 所示。AIP 作为 OAIS 的一类信息包，其构造遵循上述要求，即包括内容信息和保存描述信息。

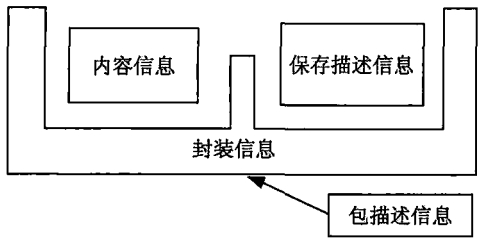

图 1　信息包的概念和关系[1]

AIP 的内容信息是长期保存的基本目标，本身是数字化的信息，由比特（bits）组成。这些比特只有与呈现信息（Representation Information）相结合的时候，才能被转换为有意义的信息，也只有经过呈现的信息才能够为长期保存的目标用户所理解和使用。即，“数据经其呈现信息的转译生成信息”。因此，AIP 的内容信息由数据对象（物理对象或数字对象）和呈现信息（包括结构信息、语义信息等）组成。

AIP 的保存描述信息是长期保存内容信息所需的相关信息，OAIS 将这些信息分为以下 5 种类型：参考信息（Reference）、起源信息（Provence）、环境信息（Context）、不变性信息（Fixity）以及访问权限信息（Access Rights）。

为了更清楚地理解 OAIS 模型对 AIP 的内容和结构的详细定义和描述，本文将相关信息汇集在一起如图 2 所示。

包描述信息包含从数据对象内容信息提取的内容描述和从 PDI 提取的保存描述信息，从而提供检索辅助，如描述元数据等信息；数据对象的呈现信息是使数字对象可被理解的信息，如数字对象的格式、正确展示所需要的软件及其版本等技术元数据；指引信息用于数据对象的引用，是数字对象的唯一标识符；起源信息记录数字对象创建、变更等相关事件的信息，是重要的保存元数据信息内容；环境信息包括数据对象创建的原因、与环境的关系及与其他数据对象的关系等，也称为上下文信息。

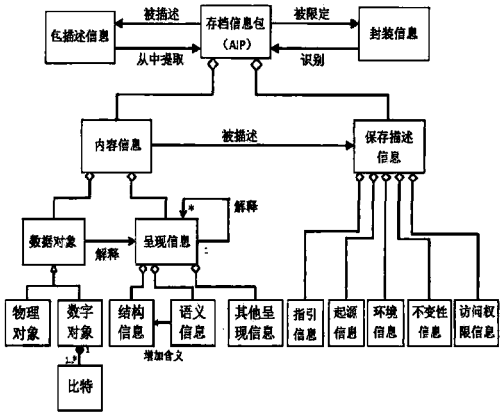

图 2　AIP 详细结构与内容[1]

AIP 独立于存储介质存在，通常是保存迁移中的主要对象。良好的数据结构和遵从标准的元数据格式将极大地减少数据对象的管理、分发和迁移的工作量，对构建可信赖的长期保存系统具有重要意义。

3　保存领域与 AIP 构建相关的标准规范

针对 AIP 的组成和构建，长期保存领域已经逐步形成了一些相关的标准规范，并陆续被众多系统和项目采用。本部分将对几个主要的标准进行比较详细的介绍和分析。

3.1　期刊文章标签套集 JATS

JATS[2]（Journal Article Tag Suite，期刊文章标签套集）由美国国家生物技术信息中心（NCBI）开发，其目的是提供一种通用的期刊数据存档和交换格式。JATS 根据应用场景的差异，针对存档交换、期刊出版、文章创作定义并描述了三种文章（Article）数据模型，形成了期刊存档和交换（Journal Archiving and Interchange Tag Set）、期刊出版（Journal Publishing Tag Set）和文章创作（Article Authoring Tag Set）三个标签集。JATS 的前身是 NCBI 的 NLM 存档和交换 DTD[3]，2012 年作为 NISO 的标准发布（Z39.96－2012）。此外，NCBI

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

在JATS的基础上扩展实现BITS[4](Book Interchange Tag Suite,图书交换标签集)以提供用于出版商和存档机构交换图书内容的通用格式。

在AIP的构建中,主要采用其中的期刊存档和交换标签集,该标签集的目的是不依赖于数据的原始提交形式保存期刊中的学术内容。使用该标签集可以捕获现有材料中的结构和语义成分,而无需文本格式或特定序列的建模[5]。该标签集定义了描述期刊文章的内容和元数据的元素和属性。

JATS存档和交换标签集格式文档的根为<Article>(对应期刊中的论文、信件、社论、书评等期刊的顶级组件),<Article>下包含一个或多个部分(如果有多个部分,则必须按规定的顺序出现),如图3。

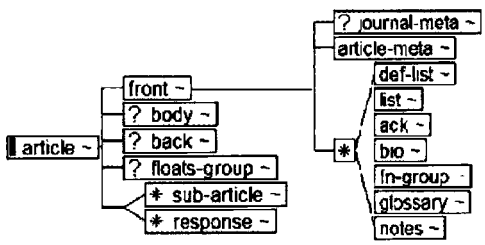

图3 Article标签格式

<front>为必备部分,其内容包括文章的元数据,如标题、所在期刊、出版日期和所属卷期、版权声明等。期刊级元数据使用标签<journal-meta>,包括期刊的题名、ISSN、贡献者、出版者等相关描述元数据字段;文章级元数据和卷期级元数据使用标签<article-meta>,包括文章题名、摘要、作者、机构、出版日期、所属卷期等相关描述元数据字段。

其他标签皆为可选部分。<body>是文章的主体(Body of the article),包括文章主要的图文内容,通常由段落和章节构成,可能包含图形、表格、工具条等。对于不保存全文的存档库可不包含此部分。

<back>包含正文的附属成分,如词汇表、附录、参考文献列表等。<floats-group>为方便发布者处理,用于存放一些附属材料的单独的容器元素,如表格、图片、文本框等。最后可能有一个或多个<Response>或<Sub-article>。

<front>部分包含了数据对象的元数据信息,如果不需要保存或交换数据对象的具体内容(全文),JATS文档可以只包含<front>。

3.2 保存元数据标准PREMIS

PREMIS[6](Preservation Metadata:Implementation Strategies,保存元数据:实施战略)项目由OCLC和RLG发起,建立在大量的专家意见的基础上,已经成为保存元数据国际上的事实标准,被越来越多的现有或者正在构建的长期保存系统所接纳和采用。2005年发布1.0版本,国内已有文章进行详细介绍[7];在2008年发布的2.0版本中做了许多改进,调整了数据模型,特别是对权利声明实体进行了比较大的扩展。当前为2.2版本。

PREMIS基于OAIS参考模型,从实施的角度定义自己的数据模型,可以看作是对OAIS概念模型到可执行语义单元的翻译框架。PREMIS在所有保存管理元数据中定义了一个子集(如图4),涉及数字对象描述元数据、保存元数据、业务规则、具体格式的技术元数据等多方面内容[8]。这些数据可分别对应AIP中的保存描述信息、呈现信息的不同部分。

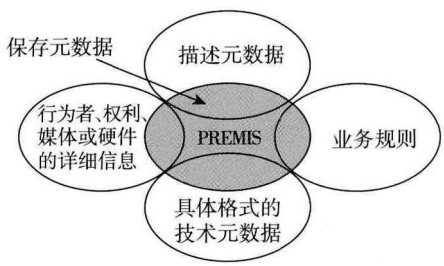

图4 PREMIS与长期保存元数据间关系

PREMIS数据模型中定义的5种实体类型,包括知识实体(Intellectual Entities)、数字对象(Objects)、权利声明(Rights Statements)、事件(Events)、行为者(Agents)。

知识实体是一种概念上的实体,也可被称为"书目实体"(bibliographic entities),为了管理和描述可以作为一个单一知识单元考虑的一组内容,如一本书、地图、图片、数据库等。PREMIS没有定义相关的元数据,因为已有许多可供选择的

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

描述元数据标准,如 DC、MODS 等。

对象实体是实际管理和保存的数字对象。PREMIS 的主要内容用于描述数字对象。描述信息包括:唯一标识符、校验信息、对象大小、格式、原始名称、创建信息、限制信息、重要属性、环境、存储地和介质、数字签名、与其他数字对象的关系等。

事件实体是影响数字对象的活动的相关信息。准确可靠的时间记录是维护对象的数字起源的关键,这对数字对象的真实性非常重要。关于事件需要记录的信息包括:唯一标识符、事件类型、事件发生时间、详细描述、事件结果编码、事件结果的详细描述、涉及的行为者及角色、涉及的数字对象及其角色。每个保存系统需要决定哪些事件要作为数字对象的历史永久保存,PREMIS 建议对数字对象造成改变的操作都要记录。

行为者实体包括参与事件或权利声明的人、机构或软件程序。由于已经有一些外部标准可用于记录更详细的信息,PREMIS 仅定义少量的语义单元来标识行为者,包括:唯一标识符、行为者名称、行为者类型。事件或权利声明中引用行为者应同时标识其角色。每个行为者都可以有多个角色。

权利声明实体包括一个或多个与对象实体/代理人实体相关的权利或权限声明。

PREMIS 数据字典中定义了语义单元而不是元数据元素。语义单元是信息或知识的片段,元数据元素定义了用元数据记录、框架或数据库表达信息的方法。

PREMIS 定义 3 种类型的语义单元:语义单元(Semantic Unit),代表信息或知识片;容器(Container)是一种特殊的语义单元,本身不包含数据值,用于将一组有相互关系的语义单元组织在一起,提供了数据字典中的层次结构,其中包含多个子单元;扩展容器(Extension Container),是一种特殊的容器,其中不包含任何子单元,其目的是提供一个记录非 PREMIS 的元数据的位置。通过这种方式,PREMIS 可以包含外部元数据,从而扩展了 PREMIS 的覆盖范围。当前版本没有定义知识实体的语义单元,但随着应用需求的发展,在 PREMIS 3.0 中知识实体将成为对象的另一个层次,并定义相关的语义单元。

PREMIS 不指定元数据在系统中应该怎样表示,只是定义系统应该知道和能够导出给其他系统的内容。PREMIS 是考虑和组织保存元数据时的一个通用数据模型,可以作为保存系统中核心元数据的清单指导本地实施,也可以作为保存系统之间进行信息包交换的标准。但 PREMIS 并不是一个可直接使用的解决方案,需要在系统中实例化成元数据元素[9]。用于数据交换时 PREMIS 推荐采用 XML 表示,并提供了一个直接与数据字典对应的简单的 XML schema,以描述对象、事件、行为者和权利声明。

3.3 元数据封装规范 METS

METS[10](Metadata Encoding and Transmission Standard)由美国数字图书馆联盟 DLF(Digital Library Federation)2001 年开发,由美国国会图书馆的网络发展和 MARC 标准办公室负责维护,采用 W3C 的 XML Schema 语言表达,是一个针对数字对象进行封装的描述性、管理型和结构性元数据标准。METS 是一个 XML 框架,可将数字对象所有关联的元数据进行存储,为多种类型的元数据提供一个整合容器。

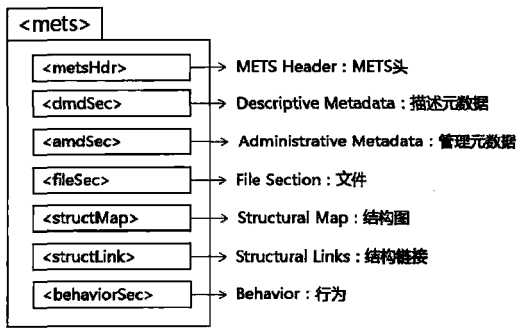

图5 METS 文件结构

如图5所示,METS 文档由七个主要部分组成:

< metsHdr > METS 文件头,包含了描述 METS 文档自身的元数据,比如创建者、创建时间、上次修改时间、编辑者、文档状态等。

< dmdSec > 为描述元数据。METS 文档可以包含一个或多个 < dmdSec > 元素,可以指向该

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

METS 文档外部的描述元数据(外部描述元数据,<mdRef>元素,例如,引用 OPAC 中的一条 MARC 记录),也可以在该 METS 文档内部嵌入描述元数据(内部描述元数据,<mdWrap>元素),或二者兼有。

<amdSec>为管理元数据,包含文件是如何创建和保存的、关于知识产权和该数字对象原始资源的元数据、构成该数字对象的文件的起源信息(比如,文件的主体和派生关系,以及迁移和转换信息)。与描述型数据一样,管理元数据既可以位于 METS 文档外部,也可以被编码在文档内部。METS 规定了管理元数据的四种主要格式:技术元数据(<techMD>),关于文件的创建、格式和使用特征的信息;知识产权元数据(<rightsMD>),版权和许可信息;来源元数据(<sourceMD>),该数字对象用于模拟其来源文档的描述型元数据和管理型元数据;数字起源元数据(<digiprovMD>),关于作品的原始数字化形态与作为数字对象的当前形态之间的关系信息,包括文件之间的来源/目标关系、主体/派生关系和迁移/转换关系。

<fileSec>为文件部分,包含一个或多个文件组(<fileGrp>)。数字对象由若干内容文件<file>组成,这些文件要全部列在<fileSec>中。<file>元素可以被<fileGrp>元素划分成组,以便按照对象版本加以细分。

<structMap>用于描述数字对象的结构图,是 METS 文档的核心。它描述了数字对象的层次结构,并将该结构中元素与相应的内容文件和元数据关联起来,允许用户通过结构图导航。利用一系列嵌套的 元素体现这种层次结构,用 METS 指针元素<mptr>和文件指针元素<fptr>进行指向。

<structLink>用于记录结构图层次节点之间存在的超链接,特别适于用 METS 存档 web 站点。

<behaviorSec>行为部分,用于关联可执行代码与 METS 对象内容。其中的每一个行为(behavior)包含了接口定义元素(interface definition element)和机制元素(mechanism element)。前者明确了一组行为的抽象定义,并用特定的行为(behavior sector)表示;后者则标识了在接口定义中已被抽象定义的若干行为所对应的可执行代码模块。

METS 为信息对象及其各种元数据的封装提供了一个标准框架,支持灵活的扩展,已经为多数长期保存项目和系统采用。

4 构建基于标准的 AIP 应用模型

OAIS 定义了存档信息包(AIP)的概念模型,在保存实践中,需要根据具体的应用需求定义各自的应用模型。从开展长期保存之初,各项目就开始了 AIP 的研究和实践,已有论文对早期的 AIP 应用模型进行了总结和分析[11]。随着保存领域主流标准的出现和趋于成熟,大多数项目在实际应用中或多或少都采用了上述标准,并根据实际需求和不同的系统目标形成了各具特色的 AIP 模型。

本文通过对几个现有标准的集成,结合 OAIS 定义的 AIP 概念模型,形成了一个比较理想的 AIP 应用模型,如图 6。

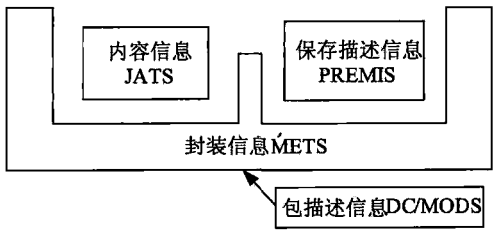

图 6 基于相关标准的 AIP 应用模型

METS 文档在其设计用途中指出,它可以担当 OAIS 参考模型中的 SIP、AIP 或 DIP 的角色[12],即可用于这三类信息包的封装。METS 提供一个灵活的元数据封装框架,通过 METS 框架可实现对数字对象及其所有关联元数据的封装。在实际应用中,多数项目都采用 METS 进行 AIP 的封装或采用可转换成 METS 的封装格式。

JATS 标签集(及 BITS)为期刊(图书)类数字对象的存档和交换提供了结构化的描述,作为通用的 XML 格式,包括元数据和全文内容的描述标签,提供了对 AIP 中内容信息的规范化描述。JATS 格式在世界范围内被用于标记数千种期刊,

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

PubMed Center 和 HighWire 等将其用于期刊在线出版[13]。采用 JATS 标签创建可存档的 XML 格式文档,能够对期刊(或图书)相关元数据及全文内容进行有效地保存。如 Portico[14] 将接收的 XML 文档转换 JATS 格式的可存档 XML 进行存档。

PREMIS 为保存元数据提供了参考和数据交换标准,在 AIP 的构建中,可以灵活地使用其中的语义单元,许多长期保存项目以不同的方式使用 PREMIS,如:Archivematica、DAITSS、HathiTrust、Rosetta 等[15]。

大多数项目在实际应用中或多或少都符合上述标准应用模型,特别是组合使用 METS 和 PREMIS。如 DAITSS 项目[16]采用 METS 进行数据包封装并支持几乎所有 PREMIS 数据对象,其内容信息采用 MODS 描述;HathiTrust[17] 定义了基于 METS 和 PREMIS 的数据规范,对内容信息采用 MARC 描述,并在 METS 结构外进行了封装。

Portico 项目是符合本应用模型的典型代表。Portico 项目在其 2.0 版本系统中出于简化处理、提高保存效率等方面的考虑使用了自定义的数据模型,该模型可以与 PREMIS 兼容,并对数据模型到 PREMIS 和 METS 进行了详细映射,存档数据内容采用 JATS 格式,并从中提取描述元数据。

5 长期保存 AIP 构建实例分析

中科院文献情报中心长期保存项目(DPS)基于 Fedora Repository[18] 建设,其目标是对中科院订购的数据资源进行长期保存,在突发事件发生的情况下提供应急公共访问。目前主要存档内容为电子期刊、电子书和实验室手册等。Fedora 是一个开源的数字仓储软件,在许多长期保存项目中使用。其定义的存档信息包(AIP)由作为保存对象的数字对象的信息内容和一个描述文件构成。其中信息内容可以包括多个数字对象文件,Fedora 在文件系统中存储的对象为数字对象文件及与之相关的描述文件。

Fedora 定义了一个可扩展的存档内容模型 FOXML(Fedora Object XML),由永久标识符(PID)、对象属性(Object Properites)及一系列数据流(Datastream)构成。其中,对象属性是系统定义的一组管理和跟踪数据对象所必需的描述性属性;Datastream 是 Fedora 数字对象的具体内容项,所有的数字内容,包括内容信息和保存描述信息,都可以通过 Datastream 保存。模型中定义了几个系统保留 DataStream。此外,用户可根据保存需要,通过增加不同的 DataStream 定义来实现对内容模型的扩展。

DPS 对 FOXML 数字对象模型进行了扩展,定义了自己的 AIP 内容模型,如图 7 所示。DPS 基于 FOXML 的 AIP 内容模型中包括 Fedora 保留的 DataStream、DPS 自定义的 DataStream 及 DPS 自定义的可选 DataStream 三个部分,其中附加文档及其技术元数据可多次重复出现。

DPS 的 AIP 封装中,保存元数据和技术元数据采用 PREMIS 定义的相关语义单元,描述元数据(NSLDescMeta)目前为自定义格式,正在探索向 JATS 的迁移,其中电子书保存采用 BITS 的元数据标签格式作为描述元数据。

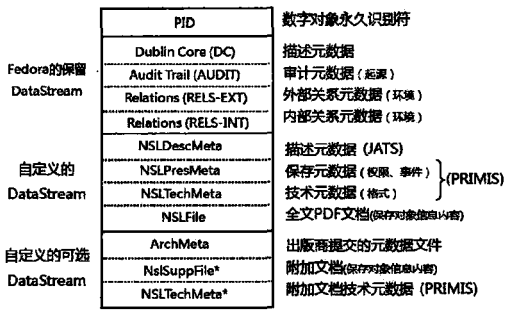

图7 基于 FOXML 的 AIP 内容模型

Fedora 以 FOXML 格式进行数据封装,同时也支持 METS 格式的封装数据的摄入和导出。Fedora 数字对象在采用 METS 格式封装时,主要使用 <amdSec> 和 <fileSec> 两部分,根据 Datastream 的属性 CONTROL_GROUP 值,分别映射到 <amdSec> 和 <fileSec>。一般来说,元数据等 FOXML 内部保存的 XML 格式 DataStream(CONTROL_GROUP = X)映射成 <amdSec>;外部引用或内部管理的数据文档文件 DataStream(CONTROL_GROUP = M/E/R)映射到 <fileSec> 中的 <fileGrp>,其中的每一个保存版本映射到一个

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

<file>。DPS 内容模型映射到 METS 结构形式如下：

```
<METS:mets>
<METS:metsHdr RECORDSTATUS="A"/>
<METS:amdSec ID="NSLDescMeta" STATUS="A"></METS:amdSec>
<METS:amdSec ID="DC" STATUS="A"></METS:amdSec>
<METS:amdSec ID="RELS-EXT" STATUS="A"></METS:amdSec>
<METS:amdSec ID="RELS-INT" STATUS="A"></METS:amdSec>
<METS:amdSec ID="NSLPresMeta STATUS="A"></METS:amdSec>
<METS:amdSec ID="NSLTechMeta" STATUS="A"></METS:amdSec>
<METS:amdSec ID="Audit" STATUS="A"></METS:amdSec>
<METS:fileSec>
<METS:fileGrp ID="DATASTREAMS">
<METS:fileGrp ID="NSLFile">
<METS:file ID="NSLFile.0"></METS:file>
  </METS:fileGrp>
  <METS:fileGrp ID="ArchMeta">
<METS:file ID="ArchMeta.0"></METS:file>
  </METS:fileGrp>
  <METS:fileGrp ID="NslSuppFile">
<METS:file ID="NslSuppFile.0"></METS:file>
  </METS:fileGrp>
</METS:fileGrp>
</METS:fileSec>
</METS:mets>
```

DPS 早期保存的全文数据主要为 PDF 文档，目前，越来越多的出版商在提交数据的时候提交 XML 格式全文，为了保证数字资源的保存忠于原始提交数据，DPS 正在探索采用 JATS 存档 XML 格式的结构化全文数据的具体实施方案。

6 保存实践中构建 AIP 的关键问题

在 AIP 构建中尽量遵循相关标准是长期保存可持续发展的重要保障。在遵循标准的前提下，各具体的保存实践还需要考虑保存系统的灵活性、处理效率、可靠性等多方面的问题，需要根据具体应用需求设计具体的 AIP 构建方案，以更好地实现长期保存的最终目标。其中需要考虑的问题主要包括以下几个方面：

(1)技术标准的成熟性和可扩展性。从长远发展角度，AIP 的构建要坚持遵循相关标准，在保障系统效率和有效性的前提下，尽量采用现有成熟的技术标准。由于相关标准随着技术和应用的变化也会发生变化，在标准的使用中要充分考虑其未来发展及向前兼容性，以保证标准发生变化时数据仍能标准化保存和解析，并在必要时能够高效率地完成迁移。

(2)AIP 本身的可扩展性和兼容性。保存的数字对象的类型和复杂度是确定元数据标准和封装结构最主要的影响因素。实践中应根据数字对象的特点确定其在保存系统中的存储方式，设计数字对象的内容及结构模型，选择特定的描述元数据标准和封装结构。此外，随着时间的推移和保存系统自身的发展，可能需要处理更多类型的数字对象，构建 AIP 时需充分考虑未来系统扩展的需求，保证数字对象结构的灵活性和可扩展性，以适应未来的需求变化。

(3)AIP 主要功用的影响。系统的建设目标对 AIP 的构建也有重要影响。对于主要用于存档的系统(如 DAITSS 系统)，可以对信息内容进行压缩打包，从而简化存储、拷贝过程，节约存储空间；对于需要在特定情况下提供公共访问的服务系统(如 DPS 系统)，对信息内容进行压缩打包将不利于用户访问，需要考虑将信息内容直接存储。

(4)格式迁移策略。随着技术环境的变化，保存系统中的数据内容可能会出现格式过时问题，为有效地进行数据对象的格式翻新，AIP 中必须保存完整的呈现信息。如果系统存储对象包括压缩包，则压缩包中数据对象的相关技术元数据也需要提取、保存。由于数据对象的复杂性和多变性，AIP 中相关元数据字段或结构的设计必须具有良好的可扩展性，并确保数字对象与其呈现信息的对应关系。

实际应用中，面对复杂的数据对象和应用需求，单一的 AIP 结构可能无法满足所有的保存需求。需要对 AIP 进行类型、功能等方面的区分，在统一的应用模型指导下构建多种实例化模型，保证 AIP 结构上的统一性和应用上的灵活性。

7 结语

信息内容的不变性、对技术环境的依赖性与不断发展变化的技术环境、应用环境之间的矛盾是长期保存必须解决的基本问题。存档信息包 AIP 的构建与封装是数字资源长期保存系统处理流程中的重要内容，是数字资源得以保存并保证

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

未来可理解可使用的基础。构建遵循相关保存标准规范的AIP,对于在未来长期的保存活动中保证系统的可信赖性、数据对象的可迁移性、以及与其他系统的互操作和数据交换具有非常重要的意义。

JATS提供了丰富的元数据标签,特别是期刊存档和交换标签集,充分考虑出版和存档双方的应用需求,具有很高的参考价值和实际应用价值,可用于数据对象描述元数据和结构化全文的标识和保存、生成可存档的XML文档。PREMIS作为事实上的保存元数据标准,为保存系统的元数据提供了详尽的参考,用于保存事件、行为、保存权限等管理元数据和技术元数据,可以灵活地嵌入AIP的封装结构中。METS提供了一个灵活的封装容器,将不同标准的各类元数据灵活地组织在一起,在长期保存相关项目中获得了广泛的应用。这三个主流标准的出现和发展极大地便利了长期保存实践中AIP的构建。本文基于这三个标准构建AIP应用模型,希望能够对未来工作提供依据,为同行提供有益参考。

参考文献

1 OAIS[EB/OL]. [2014-05-13]. http://public.ccsds.org/publications/archive/650x 0m2.pdf.

2 Journal Article Tag Suite[EB/OL]. [2014-05-13]. http://jats.nlm.nih.gov/.

3 NLM DTD Suite[EB/OL]. [2014-05-13]. http://dtd.nlm.nih.gov/.

4 Book Interchange Tag Suite[EB/OL]. [2014-05-13]. http://jats.nlm.nih.gov/extensions/bits/.

5 Journal Archiving and Interchange Tag Set[EB/OL]. [2014-05-13]. http://jats.nlm.nih.gov/archiving/.

6 PREMIS[EB/OL]. [2014-05-13]. http://www.loc.gov/standards/premis/.

7 高蒿,张智雄. PREMIS保存元数据体系分析[J]. 现代图书情报技术,2006(4).

8 Understanding PREMIS[EB/OL]. [2014-05-13]. http://www.loc.gov/standards/premis/understanding-premis.pdf.

9 Metadata for Preservation: A Digital Object's Best Friend[EB/OL]. [2014-05-13]. http://www.niso.org/news/events/2013/webinars/preservation/.

10 METS[EB/OL]. [2014-05-13]. http://www.loc.gov/standards/mets/.

11 张智雄,等. 数字保存系统中的信息模型研究[J]. 中国图书馆学报,2006(5).

12 METADATA ENCODING AND TRANSMISSION STANDARD: PRIMER AND REFERENCE MANUAL[EB/OL]. [2014-05-13]. http://www.loc.gov/standards/mets/METSPrimerRevised.pdf.

13 NISO Publishes Journal Article Tag Suite (JATS) Standard[EB/OL]. [2014-05-13]. http://www.niso.org/news/pr/view?item_key=d92a2bc93b43db6831e68914e134c731d83cbdd1.

14 Portico[EB/OL]. [2014-05-13]. http://www.portico.org/digital-preservation/.

15 PREMIS Implementation Registry [EB/OL]. [2014-05-13]. http://www.loc.gov/standards/premis/registry/premis-fulllist.php.

16 DAITSS, an OAIS-based preservation repository[EB/OL]. [2014-05-13]. http://daitss.fcla.edu/sites/daitss.fcla.edu/files/DAITSS%20in%20ACM%20rev_0.pdf.

17 Digital Object Specifications (METS and PREMIS)[EB/OL]. [2014-05-23]. http://www.hathitrust.org/digital_object_specifications.

18 Fedora[EB/OL]. [2014-05-13]. http://www.fedora-commons.org/.

(付鸿鹤 馆员 中国科学院文献情报中心,吴振新 研究馆员 中国科学院文献情报中心,王玉菊 馆员 中国科学院文献情报中心)

收稿日期:2014-05-13

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net