

# 富媒体学术出版资源特点 及长期保存挑战研究\*

刘惠<sup>1,2</sup> 赵子瑜<sup>3</sup> 张冬荣<sup>1,2</sup>

- (1. 中国科学院文献情报中心, 北京 100190;  
2. 中国科学院大学经济与管理学院图书情报与档案管理系, 北京 100190;  
3. 苏州大学社会学院, 苏州 215000)

**摘要:**[目的/意义] 随着信息技术的进步与应用, 学术出版模式更加多样, 学术出版物总体朝富媒体化方向发展, 此类数字资源的长期保存工作面临着诸多新的挑战。[方法/过程] 本文在对富媒体学术出版进行概念辨析和文献调研的基础上, 选取了12种代表性期刊进行个案分析, 从长期保存实践角度对其内容、形式、格式、体量等进行网络调查。[结果/结论] 本文总结出富媒体学术出版资源的六大特点, 针对这些特点, 本文讨论了该类数字资源长期保存所面临的技术、标准、服务、权益和安全五大挑战。

**关键词:** 学术出版 富媒体 数字资源 长期保存

**分类号:** G253

**DOI:** 10.31193/SSAP.J.ISSN.2096-6695.2021.04.10

## 0 引言

国际顶级学术期刊 *Nature* 从普通的科普杂志顺利转变为现代科学出版物的国际标杆, 其学术影响力的提升在很大程度上得益于出版模式的转型。首先, *Nature* 对新媒体的应用非常成熟, 其官方网站上设置了独立的新闻频道“News & Comments”和播客/视频栏目“Audio & Video”<sup>[1]</sup>; 在 Twitter、Facebook 平台上还设有与期刊同名的官方账号和新闻账号, 多角度对期刊进行媒体宣传, 与官方网站形成联动。<sup>[2]</sup> 其次, *Nature* 有成熟的增强出版实践, 它要求作者投稿时按规范提

\* 本文系中国科学院大学生创新实践训练计划项目“富媒体出版学术信息的数字资源长期保存研究”(项目编号: 20204002534)研究成果之一。

[作者简介] 刘惠, 女, 硕士研究生, 研究方向为网络信息服务, Email: liuhui@mail.las.ac.cn; 赵子瑜, 女, 本科生, 研究方向为档案资源开发利用, Email: 1482505104@qq.com; 张冬荣, 女, 硕士生导师, 研究馆员, 研究方向为网络信息服务, Email: zhangdr@mail.las.ac.cn (通讯作者)。



刘惠, 赵子瑜, 张冬荣. 富媒体学术出版资源特点及长期保存挑战研究 [J]. 文献与数据学报, 2021, 3(4): 106-118.

交增补信息, 并在论文接收后上传到网站服务器, 因此读者在浏览文章的同时可以获取相应的更丰富的内容。<sup>[3]</sup>作为新型出版模式的典型代表, *Nature* 的实践在一定程度上说明了学术出版的迅速发展使得当前学术出版资源的复杂程度不可同日而语。目前学术出版物的长期保存问题已经成为了国家战略安全问题之一, 并日益受到学界的重视。而现有的数字资源长期保存体系与政策机制未能完全适应学术出版的迅速变革, 对于新型学术出版资源给长期保存工作带来的挑战还有待学界回应。

## 1 富媒体学术出版概念辨析

随着互联网技术的飞速发展和广泛应用, 越来越多的新兴技术被运用到出版流程之中, 学术期刊不断探索新的出版模式。其中发展较早且愈发成熟的出版模式之一是增强出版。2009年以来, 国外知名出版集团如 Springer·Nature、PLoS、Elsevier、SAGE 等都已开展了学术期刊增强出版的信息系统建设和实践。<sup>[4]</sup>所谓的“增强出版”是在学术出版传统叙事方式的基础上进行成果增强的行为, 包括论文内容的扩展、呈现形式的丰富、传播渠道的扩充以及编辑流程的强化等。<sup>[5]</sup>常见的增强方式一般包括内容增强、语义增强、呈现增强和渠道增强。除了增强出版物, 当前还涌现出了许多新型学术出版物, 如数据期刊、视频期刊等。这些出版物类型不同于增强出版物, 是运用新型出版手段解构了学术成果出版的传统叙事手法, 一方面根文献不再讲述完整的科学故事, 对科研过程中涉及到的各种研究要素进行出版<sup>[6]</sup>; 另一方面根文献不再局限于文本, 而是采用音视频的形式, 以更快的传播速度、更丰富的信息量和更具实景感的阅读体验报道学术最近动态和科研最新成果<sup>[7]</sup>。此外, 出版界还出现了社交出版和新/自媒体出版等出版模式的说法。通过概念辨析可以发现, 这些学术出版模式之间并不是彼此独立、完全不同的, 而是往往在内容组织、呈现形式、分发渠道等方面呈现出相似的特点。因此总体来看, 出版模式在向媒体融合、技术融合和内容整合的方向发展, 最终导致了学术出版的“富媒体化”。富媒体 (Rich Media) 一般是指具有动画、声音、视频等多媒体形式且具有交互性的信息传播形态。<sup>[8]</sup>结合已有研究成果对富媒体出版概念的界定<sup>[9]</sup>, 本研究认为富媒体学术出版是指在学术出版物的开发、编辑、发行等流程中使用虚拟现实、增强现实、多媒体技术等多种信息技术, 实现内容呈现形式和传播手段更加多样, 用户接受体验更加丰富的出版模式。该模式可以满足用户碎片化、屏幕化、交互性、视听结合式的阅读习惯和需求, 是近年来全球学术出版数字矩阵建设的发展趋势。而富媒体化在扩大出版方影响力以及提高学术成果传播效率的同时, 也使得出版资源的情况变得更加复杂。这对长期保存工作不可避免地造成冲击, 并给参与学术出版资源永久保存甚至生命周期其他阶段的相关主体带来诸多挑战。

## 2 学术出版资源长期保存相关研究概况

国内外学界在数字资源长期保存研究领域已取得了众多成果。在知网、WOS 数据库进行检索, 初步浏览所有相关文献的题名, 发现就研究保存的对象而言, 随着数字资源的复杂化及长期

保存研究和实践不断发展,保存对象的类型越加丰富,涉及网络信息资源、社交媒体数据、数字档案、软件资源、医学数字资源、文化遗产资源、电子邮件、政府数据、开放科学数据等,说明保存实践和研究更加专门化,但其中以“数字资源”这一概念宽泛的保存对象出现次数最多,研究的问题涉及数字资源长期保存的系统工具<sup>[10]</sup>、标准规范<sup>[11]</sup>、技术方法<sup>[12]</sup>、保存格式<sup>[13]</sup>、政策法规<sup>[14]</sup>等。进一步浏览题名中出现“数字资源”的长期保存相关文献,发现大部分文献虽然以数字资源为保存对象进行讨论,但潜在认知主要针对的是或者默认包含了国内外学术出版资源。如中国科学院文献情报中心长期保存工作团队产出的系列研究成果<sup>[15-16]</sup>基于的是NDPP保存体系建设实践。而NDPP的保存资源主体就是学术期刊,截至2020年,NDPP已保存了国际重要的68种文献数据库,保存期刊达11300多种,并保存了国内万方和维普数据库收录期刊14000多种。此外,专门针对学术出版物尤其是学术期刊这一类型数字资源的研究成果不仅发文时间都较早,数量较少,而且研究主题具有较为明显的集中性和同质化。如HONEY S L<sup>[17]</sup>介绍了电子学术出版物归档保存的三种方法。赵毅<sup>[18]</sup>探讨了电子期刊长期保存方案所涉及的基本问题,比较分析了LOCKSS、CLOCKSS、Portico以及e-Depot四个全球著名保存方案的特征。这一现象的主要原因是学术出版资源是国内外最早实现长期保存实践和研究的,因此研究体系和实践成果较为成熟。虽然近年来长期保存领域的研究对象不断拓展,学界对长期保存过程中的技术和非技术问题不断提出新的解决方案,但没有学者将保存对象再次聚焦到已然发生巨大变化的学术出版领域。在数字生态愈加复杂的当下,学术出版实践不断创新,既有的学术出版资源长期保存策略应按当前富媒体化的特点进行调整和改进。

### 3 富媒体学术出版资源的特点分析

#### 3.1 调研方案和结果

学术期刊是学术出版最重要的渠道和形式,且限于研究精力和论文篇幅,本研究最终选取了12本国内外影响力较大的代表性学术期刊作为调研对象(见表1)进行网络调查。案例选取标准是既要包含从传统出版订阅模式转型的期刊如*Nature*、*Plos one*等,也要包括年轻的新型期刊如数据期刊*Data in brief*和视频期刊*VJEP*,并尽量覆盖多个学科和多种增强出版方式。由于研究目的是把握富媒体学术出版物的特点,并在此基础上分析给长期保存工作带来的诸多挑战,因此本文结合长期保存的实践要求,从出版物内容、形式、格式、大小、增强方式等几个维度对样本期刊进行网络调查,详细调查结果见表2。

表1 调研对象基本情况

期刊	学科	性质	出版机构	是否OA
<i>Data in brief</i>	全学科	数据期刊	Elsevier	√
<i>Biodiversity Data Journal</i>	生物学	数据期刊	Pensoft Publishers	√
<i>HardwareX</i>	全学科	要素期刊	Elsevier	√

续表

期刊	学科	性质	出版机构	是否 OA
<i>Plos one</i>	生物学	传统期刊	Public Library of Science	√
<i>elife</i>	生物学、医学	传统期刊	eLife Sciences Publications	√
<i>Nature</i>	自然科学	传统期刊	Springer Nature	√
<i>The Innovation</i>	自然科学	传统期刊	Elsevier	√
<i>Video Journal of Education and Pedagogy (VJEP)</i>	教育学	视频期刊	Springer	√
《中华心血管病杂志(网络版)》(CVJC)	医学	视频期刊	中华医学会杂志社	√
<i>Journal of Visualized Surgery (JoVS)</i>	胸外科、食管外科、心外科等外科领域	视频期刊	AME	√
<i>Journal of Visualized Experiments (JOVE)</i>	生物学、环境科学、物理学、医学和心理学等	视频期刊	MyJoVE Corporation	√
<i>The Journal of Minimally Invasive Gynecology (JMIG)</i>	医学	视频期刊	Elsevier	×

注: 是否 OA 选项中, √表示该期刊完全 OA 或者是混合 OA。

表 2 调研结果基本情况

出版对象	内容	形式	文件格式	大小 (范围)	增强方式
<i>Data in brief</i>	传统论文、辅助理解论文的研究数据、软件、代码、模型、算法、数学工具等	图表(可交互)、模型、图片、表格、组件、链接	PDF、ZIP、JPG(文本中嵌入的表格是文本形式, 不可以直接下载)	466MB [ 2021.4 (第 35 卷) ]	内容增强 语义增强 呈现增强
<i>Biodiversity Data Journal</i>	传统论文、增补文件(支撑论文假设或结果的数据文件、图表等)	图表、数据文件、文本、视频、音频、3D 和其他交互式模型	XML、EXCEL、HTML、PDF、PPT、CSV、TXT	131MB (2021.2.03-2.25)	内容增强 语义增强 呈现增强 渠道增强
<i>HardwareX</i>	实验视频、研究数据及相应文件	图表、数据、视频、链接	XML、XLSX、ZIP、STP、DOC、PNG、JPG、PDF、MP4	190MB [ 2021.4 (第 9 卷) ]	内容增强 语义增强 呈现增强
<i>Plos one</i>	传统论文、辅助理解论文内容的研究数据和图表等	文本、图表、数据文件、PPT	XML、TIF、PNG、JPG、PDF、PPT、XLSX、DOC	78MB (2021.2.08-2.17)	内容增强 语义增强 渠道增强
<i>elife</i>	传统论文, 以及文中所提到的程序代码、使用到的数据、实验效果图像等	文本、链接、图片	PDF、HTML、ZIP、JPG	46.2MB (2021 年 1-2 期)	内容增强 语义增强 渠道增强
<i>Nature</i>	传统论文、增补材料(数据、代码、3D 模型、拓展图表、数据、同行评议情况等)、播客、视频等	链接、图片、表格、文本、视频、音频	HTML、ZIP、WEBP、CSV、PDF	1.27G (589 卷-7840 号)	内容增强 语义增强 渠道增强 呈现增强

续表

出版对象	内容	形式	文件格式	大小 (范围)	增强方式
<i>The Innovation</i>	传统论文、图形摘要等	文本、图片	JPG、PDF	260MB (2021年第1期)	呈现增强 语义增强
<i>VJEP</i>	视频内容主要是教育研究的演示或示范、学者采访、行为观察等	音视频、图表、文字、代码	PDF、MP4、JPG、EXCEL、HTML	12MB+ [2020.9 (第5卷)]	语义增强 呈现增强 渠道增强
<i>CVJC</i>	视频论文 (具有一般期刊论文的结构)、增补材料	视频、音频、动画、图表	JPG、MP4、PDF	49.7MB+ (2020卷)	语义增强 呈现增强 渠道增强
<i>JoVS</i>	可视化手术, 包括教学视频和外科手术示意图	视频、图片	PDF、JPG、HTML、MP4	16MB+ [2021.1 (第7卷)]	语义增强 呈现增强 渠道增强
<i>JOVE</i>	实验或方法类视频、传统论文 (可以制作成视频)	文本、视频	PDF、MP4、HTML	—	呈现增强 语义增强 渠道增强
<i>JMIG</i>	视频论文、书面手稿 (补充支持或证明研究发现)	视频、文字、图像、图表、统计数据	MOV、MPEG-4、MPEG-2、WMV、AVI、EPS、PDF、TIF、JPEG	—	呈现增强 语义增强 渠道增强

注: JOVE 和 JMIG 分别因为是混合 OA 和非 OA 期刊, 无法按一定范围统计出版物大小。

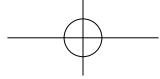
## 3.2 特点解析

### 3.2.1 出版物内容更加复杂化

如表 2 所示, 当前学术期刊的内容特点发生显著变化。一方面, 不再局限于用讲述科学故事的传统形式呈现科研成果, 而是转向对科研环节中的各类要素进行出版, 如 *Data in brief* 出版了研究数据、软件等; 另一方面, 语义增强和内容增强是普遍的增强出版方式, 因此学术出版物的范畴由单一的期刊论文扩展至具有富媒体特征的信息包, 表意更加丰富和准确, 且具有动态性、超文本性、异构化等特点。OCLC 2014 年的报告《不断发展的学术记录》<sup>[19]</sup> 中指出, 与以前的书面印刷甚至早期数字形式相比, 当下构成学术记录的内容变得更加动态, 而且“界限”更少。数字内容日益复杂成为长期保存需要积极应对的问题之一。长期保存的对象不再是论文本身, 需要重新思考保存对象和范围。

### 3.2.2 出版物格式更加多样化

调查样本中的学术出版物均具有多样化的文件类型, 如文本 / 文档、图像、电子表格 / 数据库、应用程序、演示文稿、音视频等, 每一种文件类型又包含了多种文件格式, 如表 3 所示。实际上, 不同的文件格式受到时间因素的影响在有用性方面存在差异, 这就需要考虑为各种格式的数字资源选择合适的保存工具和保存策略。<sup>[20]</sup> 保存策略一般应满足: (1) 能够识别和解析最适合可持续保存的格式; (2) 该文件格式必须能在较长时间内可以被解释和重用, 即较长一段时间后, 必须能重新创建和渲染多媒体文档的文本、图形和声音组件, 以产生与创建最初用户感知和使用体验相同的效果。<sup>[21]</sup>



刘惠, 赵子瑜, 张冬荣. 富媒体学术出版资源特点及长期保存挑战研究 [J]. 文献与数据学报, 2021, 3(4): 106-118.

表 3 调查样本中出现的文件格式汇总

文件类型	文件格式说明	文件格式拓展名
应用程序	用于对启动应用程序的可执行文件进行编码的文件格式	TAR、ZIP
图像	用于编码静态图像的文件格式	TIFF、PNG、JPEG、JPG、WEBP、STP
演示文稿	用于编码演示文稿的文件格式	PPT
电子表格 / 数据库	用于编码表格数据的文件格式, 例如电子表格和数据库	XLS、CSV
文本 / 文档	用于编码文本信息的文件格式	PDF、HTML、RTF、TXT、ASC、DOC
视频	用于编码动态图像的文件格式	AVI、MPEG、MPEG-2、MOV、MP4、FLV、WMV

### 3.2.3 出版物体量迅速增长

当前学术出版物中音视频、代码等类型内容的占比不断增加, 从而导致数字资源体量不断增长。<sup>[22]</sup> 本文根据这 12 种期刊各自的出版频率选择了一期 / 卷的内容进行浏览、下载, 统计其出版物的体量大小。如图 1 所示, 调研期刊的文本体量远小于全部出版内容体量, 而且富媒体化越成熟、增强出版方式越多的期刊其差距越明显, 如 *Nature* (周刊) 一期出版物的内容总量就超出 1G。此外, *CVJC*、*VJEP*、*JoVS* 这几种视频期刊不允许下载其视频, 在统计时无法将音视频的体量计入其中, 如 *CVJC* 2020 年的全部出版内容除了 13.8MB 的文本和图片, 还有 355 分钟的视频。因此可以推断若加上音视频的体量, 全部出版内容的存储体量一定远大于当前调研的结果。

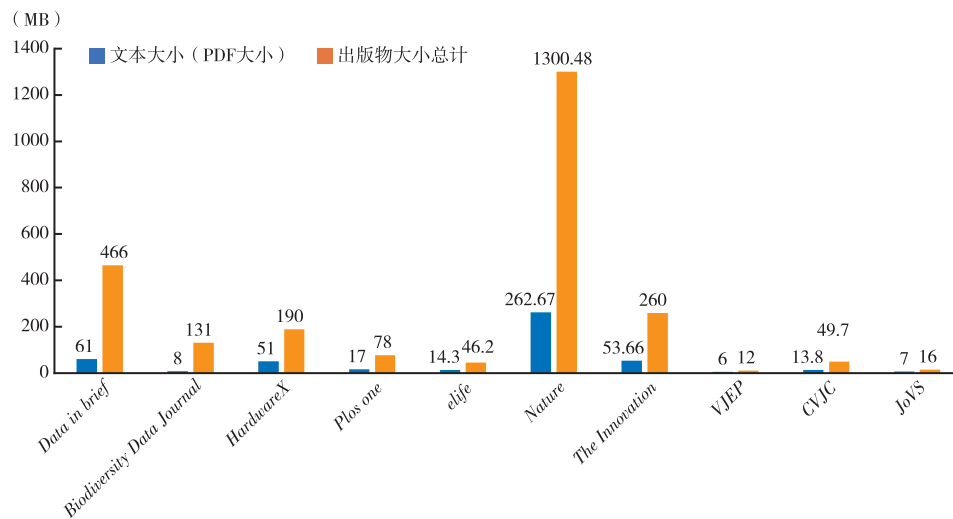
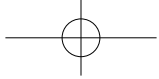


图 1 出版物存储体量比较条形图

### 3.2.4 出版物技术依赖度不断提升

数字资源长期保存对象既包括信息本身, 还包括能够发挥保存对象性能的存储介质。出版物内容复杂度大大提升, 导致其文件类型越来越多样化之外, 对硬件技术的要求越来越高, 尤



其是数据集、网络记录、视听记录和3D绘图等资源类型对存储介质的依赖程度更大。同时存储技术更新换代的速度快，依赖特定硬件创建和存储的出版资源一旦缺乏提供商的支持，可能会遭受永久性损失。即使采用迁移或仿真技术，也需要为工作人员和硬件提供一定的适应时间。<sup>[23]</sup>由此可见，当前学术出版物的长期保存工作需要解决存储介质快速更新带来的一系列衍生问题，否则将影响数字资源的完整性和可用性。<sup>[24]</sup>

### 3.2.5 开放出版资源占比越来越高

本次调研的12种期刊中，10本都是开放获取（open access, OA）期刊或者提供了OA选项，在一定程度上说明OA学术资源总量在学术出版物总体中的占比越来越大。但目前国内外长期保存的优先级往往针对图书馆付费访问的内容<sup>[19]</sup>，缺乏对OA内容长期保存的关注。出版商尤其是小型出版社的出版物中有较多的OA期刊，由于不作为商业资源被公众订购，无法被纳入长期保存体系。可喜的是，近两年来OA资源的保存问题在我国得到了越来越多的关注，如2020年Springer Nature的570多种OA期刊顺利被纳入国家数字科技文献资源长期保存体系并建立了长期保存数据的持续更新机制。<sup>[25]</sup>

### 3.2.6 出版资源分布日益分散化

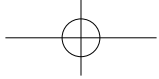
当前原生数字学术出版物分布日益分散，其原因主要有三点：一是日益增多的OA学术资源具有开放性、分散性、重复性的特征；二是论文包括的增补材料可能具有多个版本副本，存储在多个位置，使用多个文件名和多种文件格式；<sup>[26]</sup>三是渠道增强出版方式也导致资源更加分散。学术期刊逐渐形成自己的融媒体阵地，建设多平台的社交媒体账号来宣传自己的刊物，传播学术会议信息，与用户进行话题互动，从而在社交媒体上生产了大量非正式出版但与刊物密切相关的学术信息。

## 4 富媒体学术出版资源长期保存面临的挑战

### 4.1 技术挑战

数字资源长期保存通常重点保存的是运行系统的软硬件环境，而目前常见的保存技术有拷贝、仿真、迁移、冗余备份等。富媒体出版物快速更新的软硬件环境给现有的长期保存方法带来了许多技术挑战。如由于技术快速过时，就需要持续迁移。迁移基本上是复制数据或从一种技术转换为另一种技术，目的是保持数字信息的可用性，并维持用户面对技术快速更新时检索、浏览和使用资源的能力。迁移技术需要使用专门的格式转换工具对大量异构数据进行处理，因此持续迁移不仅成本高，且数据格式转码等步骤不可避免地会导致部分信息更改或丢失，从而不断降低出版物内容的完整性和真实性。<sup>[27]</sup>仿真技术方法也有自己的缺点，如缺乏一套确切的规范能证明当前模拟的系统与原始情况一致，当出版物的相关属性发生更改时就会容易发生内容不可读取的问题。但由于富媒体学术出版资源具有较强的交互功能，使用仿真技术对超文本、动画等交互式文档实现保存后，系统对用户输入的信息能做出与原始效果相同的反应。因此为了维护该类资源保存后的用户体验，未来需要进一步完善仿真方法，着力解决难度大、高耗时的问题。<sup>[28]</sup>

除了上述提到的几种典型的技术方法，学界还在积极探索将新技术应用到长期保存中，以



刘惠, 赵子瑜, 张冬荣. 富媒体学术出版资源特点及长期保存挑战研究 [J]. 文献与数据学报, 2021, 3(4): 106-118.

应对富媒体学术出版资源体量迅速增长、内容愈加复杂等特点。如 EL-FAKD<sup>[29]</sup> 提出应该更充分地利用人工智能来为搭建功能更强大且成本更低的长期保存系统做贡献。如受生物进化启发全局优化算法以及基于生态系统内集体交互行为进行智能运算的群体智能两种 CI 技术可以为设计系统工具提供理论指导, 实现一种以保存对象为中心、本质上自组织的保存方法, 即数字对象将按算法程序自动选择保存最有价值且寿命最长的内容。WOODALL<sup>[30]</sup> 则提出面对软件快速更新的挑战, 使用区块链技术对数字资源进行保存能更好地维护资源的完整性、真实性和可靠性。总之, 面对学术出版资源的富媒体特征, 目前已有的保存方法还不能全面满足其保存需求, 新技术应用也还在理论探讨和实践尝试中。因此对保存机构来说, 应在具体的工作实践当中综合考虑所处环境、保存目标、数据类型等因素, 选择多种合适的技术与策略制定组合保存方案。

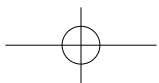
#### 4.2 标准挑战

目前我国保存机构多元化, 凭借自身职能优势参与数字资源长期保存的研究机构、高校、图书馆、档案馆等在各自选择保存资源时制定了其所在领域的政策标准。面对快速发展的学术出版实践, 原有的标准规范未能体现富媒体学术出版资源内容复杂、边界模糊、资源分布愈加分散等特点。因此各个保存机构的行动框架存在差别, 保存模式各成一派, 多领域协同无法实现, 从而阻碍了机构间数字资源的互操作和共享。而且我国国家层面的纲领性文件仍旧空缺, 数字资源保存活动缺少统一的组织和协调。国家层面统一的长期保存标准规范的缺失不利于对我国数字资源进行整体规划, 更限制了多机构主体间的合作。<sup>[31]</sup> 因此有必要针对学术出版资源富媒体特征, 制定实行统一的标准规范, 实现不论在基础架构层集成异构数据, 在传输利用层跨平台处理, 还是在顶层服务层实现主体协作的统一指导。制定统一的长期保存标准规范体系时, 应包括不同媒体类型的数字资源加工与组织标准、服务管理标准和技术方案标准等。<sup>[32]</sup> 同时鼓励多机构协同参与制定相关标准规范, 并根据自身实际情况提出针对性建设方案。

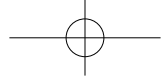
此外, 长期保存标准规范中的元数据标准一直以来是一个较为独立且十分重要的问题。美国学者 PRISCILLA KAPLAN 提出了一个数字保存金字塔模型, 描述了数字资源长期保存流程中各个节点的实现目标, 包括可获得性、可识别性、完整性、持久生存能力、可呈现能力、真实性和可理解性。<sup>[33]</sup> 而元数据则是贯穿整个长期保存流程生命周期的实现这些目标的重要信息, 其质量影响了长期保存工作的整体质量。元数据标准是元数据质量的重要保障。因此应借鉴 PREMIS 通用保存元数据框架标准, 制定符合我国本土特色的通用元数据标准, 指导从业者实践。且因为元数据标准描述的是数字信息资源的保存特征, 因此还需要各保存机构在具体应用实施标准前, 结合自身学术出版资源的类型特点和格式多样性评估元数据标准的适用范围和可操作性, 调整元数据框架, 控制元数据粒度。并将存储在元数据中的信息与最新的技术趋势进行比较, 以及及时掌握文件的过时情况, 为进行数据迁移和搭建合适的虚拟仿真环境做准备。<sup>[34]</sup>

#### 4.3 权益挑战

版权持有者和信息使用者之间的权利平衡被认为是“不确定的不对称分布”, 互联网使用的普遍性加剧了这一不平衡关系的脆弱性,<sup>[28]</sup> 而富媒体学术出版资源所呈现出的复杂特点则使权







益问题的解决变得更加困难，尤其是法律法规问题。问题表现如下：

(1) 富媒体学术出版物的资源内容和相关软硬件的知识产权可能属于不同的个人或组织，获取所需的合理保存权利将是一项漫长而复杂的任务。且我国相关法律中对各保存主体的复制例外权和责任界定并不明晰。随着以富媒体学术出版资源为代表的复合类型数字资源长期保存成为研究与实践的新增长点，为了划定合理的数字保存法律边界，我国应以利益平衡的立法原则修订和完善法律，寻求更为合理的版权作品合理使用空间。一方面，可以借鉴法国《法定呈缴本法》，将多类型数字信息资源如软件、数据库、多媒体文件以及公开网络资源纳入出版物法定呈缴范围。另一方面，在《著作权法》《关于制作数字化制品的著作权规定》等法律法规中对保存主体的保存权、复制权和使用权等作出明确规定，设置适当的技术措施限制与例外范围，增加应对技术措施滥用的内容。<sup>[35]</sup>

(2) 长期保存是一项需要全球协作的工作。而法律并不总是与技术进步保持同步，并且国家和地区间的差异较大。多数情况下，各个国家的法律对数字资源复制、存储、访问、修改等行为施加了不同的限制，设置了不同的权限范围。这就给国家间的合作带来不小的挑战，容易引发法律纠纷。因此在与国外学术出版者或版权持有者进行前期商谈及后期合作时，保存主体一方面应周期性整理自身数字资源涉及国家或地区的相关法律文件、版权政策、自存储政策等，分析其对长期保存的法律限制和豁免权要求；另一方面可以寻求与法律顾问的密切合作，法律顾问可以在数字保存过程的所有阶段提供有价值的建议和指导，包括协助保存机构起草保存项目的版权协议，就数字化作品如何向公众公开或与其他机构共享做出风险评估决策等。<sup>[36]</sup>

(3) 富媒体学术出版物牵涉的利益相关者更加多样化，包括图书馆、档案馆、出版商、学术协会、政府机构和其他内容创作者等。而在确保长期访问以及提供未来使用方面，这些利益相关者分散的、独特的和互补的作用和责任分别是什么还尚不明确。<sup>[37]</sup>权责不清不利于进行有效的战略指挥和统筹规划，从而阻碍构建我国数字资源长期保存与安全治理体系。可以依照在数字资源长期保存中的作用和角色将所有利益相关者划分为生产者、使用者和保存者三类主体，厘清保存者与其他主体角色之间的法律关系，明确各主体职责、做好权利分工。除了在理论上和立法过程中厘清法律关系，避免权责混乱，还应不断对在实践中遇到的涉及权利与义务关系的问题进行分析和总结，在各主体的权利与义务之间寻求平衡点，从而调动各方的积极性。<sup>[38]</sup>

#### 4.4 服务挑战

从长远来看，在学术出版数字化推进过程中，学术出版物的富媒体化给长期保存的服务环节带来了更加艰巨的挑战：(1) 为了保证出版内容集合的完整性，需要捕获和维护的上下文和关系信息更多，这导致发现、访问和渲染的需求日益复杂。如果保存不当，就可能导致存储的数字内容难以被用户理解，用户也难以在无边的数字海洋中准确找到所需资源。(2) 学术出版资源对技术的依赖性使得用户使用体验受技术影响较大，并进一步影响了服务方式和服务场所。例如在提供对保存材料的访问权限时，是提供与原始用户体验无异的服务，还是随着当前技术的发展而增强用户原始体验？此外，由于设备更新换代快，若干年后用户可能无法在线访问某些数字作品，只能在保存机构现场进行检索使用。这时值得思考的问题就变成了保存机构如何提供更方便、优质的服务。



刘惠, 赵子瑜, 张冬荣. 富媒体学术出版资源特点及长期保存挑战研究[J]. 文献与数据学报, 2021, 3(4): 106-118.

因此长期保存工作必须适当考虑未来目标用户可能的需求和期望, 创新服务机制: (1) 优化检索发现服务。借鉴日本国立国会图书馆的做法<sup>[39]</sup>, 通过深度挖掘并利用大数据技术提供完善的资源发现服务和可视化分析服务, 使海量、异构的原生数字记录保存之后, 用户能够快速找到所需信息, 并得到一定的洞察, 从而提高数据集的重用率。还可以借鉴哈佛大学图书馆, 加工富媒体学术出版物数据, 开发能够突出资源特色的服务产品和功能, 提供多种资源使用方式, 从而为用户提供更直观、维度更丰富的使用体验。(2) 实时跟进技术发展。保存机构应及时跟进软硬件更新, 改进技术方案的同时不断优化系统服务功能, 同时也应注意用户使用体验是否受到影响。还可以结合不同类型服务对象的需求及时调整服务策略, 必要时设计线下服务方式和提供服务场所。此外在行使服务权利时, 保存机构应注意对提供的服务加以限制。避免侵犯版权持有者的合法权益和经济利益, 注意平衡版权持有者的个人权益与公共利益间的关系, 兼顾商业利益, 维护信息服务市场的正常运转。<sup>[40]</sup>

#### 4.5 安全挑战

学术出版资源富媒体化不仅增加了保存技术难度, 还增加了长期保存过程中的风险因素, 如存贮媒体的不稳定性增大, 导致其有效寿命缩短; 标准不统一, 导致数字资源在新旧平台间转换时易被改变; 随着文件格式技术的更新换代, 当前的保存格式不可用风险增加等。因此在数字生态环境中, 长期保存工作的灾害规划和风险防范至关重要。而目前我国针对长期保存风险管理的研究和实践不足, 缺乏较为全面的安全治理战略规划, 因此是时候将注意力转移到风险管理建设上来了。风险评估是长期保存工作风险管理的必要环节, 是主动识别和处理与保存格式或保存介质相关风险的过程。美国、加拿大、英国、澳大利亚、瑞士等国家都在建立和推进风险评估流程。如美国的做法是确定电子记录保存优先级后, 根据计划进行数字保存活动, 并通过日常扫描技术环境和监视通信设备防范文件格式、存储介质变得过时和不可持续的风险。通过借鉴上述国家成熟的实践经验, 我国可以根据本土实际情况制定风险管理计划, 设计风险评估流程, 并为此开发具体功能, 持续扩大高效、可靠的长期保存规模。风险评估流程应设置以下几点: (1) 总结影响文件格式可持续性的若干因素, 并根据这些因素对文件格式的可持续性进行量化风险评估, 遴选出适合长期保存的文件格式。各保存机构可根据自身资源的内容类型, 形成文件格式风险评估指标, 如开放性、自描述性、稳定性、外部依赖性等。<sup>[41]</sup> (2) 控制元数据质量, 借鉴“全面质量控制理论”, 对保存元数据质量开展全面评估。质量评估可以贯穿元数据的整个管理周期, 确保符合保存主体特定保存需求, 降低保存后资源不可复用、再现和共享的风险。<sup>[42]</sup> (3) 制定保存成功的评估原则, 如完整性、可访问性、真实性和体验可用性等。确保长期保存管理流程的合理性和可信性。在细化评估标准时, 既要从保存对象的组织状态和数据质量出发, 也要关注用户的使用体验。<sup>[43]</sup> (4) 保存机构应系统分析保存过程中可能会出现的一切风险因素, 对风险因素类型和风险等级进行科学划分, 并与其他主体角色即生产者 and 使用者建立更密切的合作伙伴关系, 共同考虑已知和潜在的组织和技术风险。必要时可以设置一份风险防范的措施清单, 包括潜在、中度和远程风险识别, 灾难管理团队的通信渠道加密, 防火墙等的安全性检查等, 并应定期重新修订以防范新发意外风险。



## 5 结语

本研究通过网络调研和个案分析,总结出富媒体学术出版物的诸多特点,以及这些特点给长期保存工作带来的严峻考验,并在讨论各方面挑战时对相关主体未来如何改进技术方法、完善规范标准、维护权益平衡等提出了一点建议。此外,本研究仍有可以深入拓展的研究空间,例如随着新媒体技术的发展和移动智能设备的普及,社交媒体将成为学术出版愈加重要的发展舞台。这一渠道增强带来了社交媒体上学术信息的丰富和增值,进而启发我们思考这些非正式学术出版的学术信息是否具有保存价值。而学界对该问题的认识和讨论尚不充足。未来长期保存研究领域应引起对社交媒体学术信息保存相关问题的重视,本研究团队也会继续关注该问题并开展下一步研究工作。

### 【参考文献】

- [1] 张新玲,谢永生. 国外顶级学术期刊《Nature》新媒体应用研究[J]. 中国传媒科技, 2017, (4): 75-76.
- [2] 徐婷婷,柴玥. 基于6种中外生科类期刊对比的新媒体化传播研究[J]. 中国科技期刊研究, 2020, 31(2): 159-165.
- [3] 姚长青,田瑞强. 新科学研究范式下的学术期刊出版趋势研究[J]. 科技与出版, 2018, (5): 31-36.
- [4] 杨晖. 基于UGC的“网站+微信”学术期刊增强出版模式——以《电子技术应用》为例[J]. 科技传播, 2020, 12(8): 35-36.
- [5] 李静,亢小玉. 科技期刊增强出版的融合模式分析与推进策略研究[J]. 编辑学报, 2020, 32(6): 615-618.
- [6] 刘凤红,张恬. 开放科学背景下新兴学术论文出版类型——研究要素出版[J]. 中国科技期刊研究, 2017, 28(2): 138-144.
- [7] 刘冰,史红,常青云. 医学视频期刊出版探索及视频科技期刊发展面临的相关问题[J]. 编辑学报, 2020, 32(5): 549-554.
- [8] 武小菲. 泛阅读时代富媒体在数字出版中的应用[J]. 出版发行研究, 2014, (8): 54-57.
- [9] 王珍珍. 我国富媒体出版研究[D]. 河南大学, 2016.
- [10] 孙艳,马炳厚,王栋. LOCKSS还是Portico——谈图书馆电子资源的长期保存[J]. 图书馆建设, 2010, (1): 47-49.
- [11] JIN P H. A Study of Metadata for Digital Preservation Based on PREMIS 3.0[J]. Journal of Korean Society of Archives and Records Management, 2020, 20(4): 185-203.
- [12] 王伟. 数字资源长期保存的技术研究[J]. 情报科学, 2012, 30(11): 1751-1754.
- [13] YUVARAJ M, ASHRAF T, KUMAR N. Digital File Formats for Digital Preservation[M]. Hershey, IGI Global, 2017.
- [14] ABBUD GRACIO J C, FADEL B. Copyright laws and their relation with digital preservation[J]. Ibersid-Revista De Sistemas De Informacion Y Documentacion, 2010, (4): 159-164.
- [15] 张晓林,吴振新,付鸿鹄,等. 国家科技数字资源长期保存体系建设与发展[J]. 数字图书馆论坛, 2020(7): 12-19.
- [16] 张晓林,吴振新,赵艳,等. 国家数字科技文献资源长期保存体系的战略与实践[J]. 图书馆杂志, 2017, 36(12): 14-19.
- [17] HONEY S L. Preservation of electronic scholarly publishing: An analysis of three approaches[J]. Portal-



刘惠, 赵子瑜, 张冬荣. 富媒体学术出版资源特点及长期保存挑战研究 [ J ]. 文献与数据学报, 2021, 3 ( 4 ) : 106-118.

Libraries and the Academy, 2005, 5(1): 59-75.

[ 18 ] 赵毅. 电子期刊长期保存方案的比较研究 [ D ]. 郑州大学, 2017.

[ 19 ] WITTENBERG K, GLASSER S, KIRCHHOFF A, et al. Challenges and opportunities in the evolving digital preservation landscape: reflections from Portico [ J ]. Insights, 2018, (31): 28.

[ 20 ] TRACZYK T. Requirements for Digital Preservation [ M ]//TRACZYK T, OGRYCZAK W, PAŁKA P, et al. Digital Preservation: Putting It to Work. Cham: Springer, 2017: 3-13.

[ 21 ] BORGHOFF U M, RÖDIG P, SCHMITZ L, et al. Long-term preservation of digital documents [ M ]. Cham, Springer, 2006.

[ 22 ] BHAT WASIM A. Long-term preservation of big data: prospects of current storage technologies in digital libraries [ J ]. Library Hi Tech, 2018, 36(3): 539-555.

[ 23 ] TRIPATHI S. Digital preservation: some underlying issues for long-term preservation [ J ]. Library Hi Tech News, 2018, 35(2): 8-12.

[ 24 ] SO J-E, HAN H-J, YANG D. A Comparative Analysis of Long-Term Preservation Policies in Foreign Electronic Records: NARA, LAC, TNA, NAA, and SFA [ J ]. Journal of Korean Society of Archives Records Management, 2018, 18(4): 125-148.

[ 25 ] IPRES2021. OA 期刊长期保存合作进展 [ EB/OL ]. (2021-01-27) [ 2021-03-01 ]. <https://mp.weixin.qq.com/s/dVCKd5DdXYAh6ED9m6lBpA>.

[ 26 ] HOUGHTON B. Preservation challenges in the digital age [ J ]. D-lib magazine, 2016, 22(7/8): 1-6.

[ 27 ] UHEREK A, MAIER S, BORGHOFF U M. Requirements for Long-Term Preservation of Digital Videos and First Experiments with an XMT-Based Approach. proceedings of the International Conference on Computer Aided Systems Theory, December 203-210, 2015 [ C ]. Cham:Springer,2015.

[ 28 ] TLOU M. MASENYA, PATRICK NGULUBE. Digital preservation practices in academic libraries in South Africa in the wake of the digital revolution [ J ]. SA Journal of Information Management, 2019, 21(1): 1-9.

[ 29 ] EL-FAKDI A, DE LA ROSA J L. Analysis of Nature-Inspired Algorithms for Long-Term Digital Preservation [ J ]. 2021, 9(18): 2279.

[ 30 ] WOODALL A, RINGEL S. Blockchain archival discourse: Trust and the imaginaries of digital preservation [ J ]. 2020, 22(12): 2200-2217.

[ 31 ] 耿志杰, 陈佳慧. 数字资源多领域协同长期保存的实践与优化 [ J ]. 图书馆学研究, 2021, (14): 58-67.

[ 32 ] 韩子静, 孟琼. 数字资源领域标准规范体系建设比较研究 [ J ]. 图书馆学研究, 2020, (4): 60-67.

[ 33 ] NIU Y, LI H. Progress report on the role of digital resource preservation and utilization for libraries in China [ J ]. Publishing Research Quarterly, 2017, 33(4): 483-492.

[ 34 ] 杨永清, 黄毕惠. 文献资源组织中元数据集成建设问题研究 [ J ]. 数字图书馆论坛, 2020, (9): 24-31.

[ 35 ] 柴会明. 图书馆数字资源长期保存的技术措施限制与例外研究——基于著作权保护“技术路径”的考量 [ J ]. 图书馆杂志, 2021, 40(1): 48-56.

[ 36 ] JUNE M. BESEK. Digital Preservation and Copyright [ EB/OL ]. (2008-09-01) [ 2021-04-01 ]. [https://www.wipo.int/wipo\\_magazine/en/2008/05/article\\_0010.html](https://www.wipo.int/wipo_magazine/en/2008/05/article_0010.html).

[ 37 ] SOTO A. Digital preservation of scientific e-journals: Colombian case study [ D ]. Oslo: Oslo and Akershus University, 2015.

[ 38 ] 王爱霞, 许乐, 赵霞. 数字信息资源长期保存中保存者与其他相关主体间的法律关系探析 [ J ]. 图书馆理论与实践, 2021, (3): 65-69.

[ 39 ] 杨云鹏. 日本国立国会图书馆互联网资源存档研究与启示 [ J ]. 数字图书馆论坛, 2021, (1): 24-31.



[40] 한희정, 오효정, 양동민. 전자기록물의 장기보존을 위한 보존포맷 선정 방안에 관한 연구[J]. Journal of Korean Society of Archives Records Management, 2020, 20(1): 69-87.

[41] 王扬扬, 张照余. 电子文件格式的可持续性风险评估[J]. 浙江档案, 2021, (1): 36-37.

[42] 朱泽, 李玉海, 王常珏, 等. 重构之路, 我国数字资源保障体系的发展与未来——“2021年全国数字资源保障体系重构学术研讨会”评述[J]. 数字图书馆论坛, 2021, (6): 30-35.

[43] ABRAMS S. Tacit attitudinal principles for evaluating digital preservation success [J]. Archival Science, 2021, 21(3): 295-315.

## Research on the Characteristics and Digital Preservation Challenges of Rich Media Academic Publishing Resources

LIU Hui<sup>1,2</sup> ZHAO Ziyu<sup>3</sup> ZHANG Dongrong<sup>1,2</sup>

(1. National Science Library, Chinese Academy of Sciences, Beijing 100190, China;

2. Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China;

3. School of Society, Soochow University, Suzhou 215000, China)

**Abstract:** [ Purpose/significance ] With the advancement and application of information technology, academic publishing models become more diverse, and academic publications are generally developing towards rich media. The long-term preservation of such digital resources is facing many new challenges.

[ Method/process ] Based on the concept analysis and literature research on rich media academic publishing, this article selects 12 representative journals for case analysis, and analyzes its content, form, format, volume, etc. from the perspective of long-term preservation practices Internet research. [ Result/conclusion ] This article summarizes the six major characteristics of rich media academic publishing resources. In response to these characteristics, this article discusses the five major challenges of technology, standards, services, rights, and security.

**Keywords:** Academic publishing; Rich media; Digital resources; Long-term preservation

( 本文责编: 魏 进 )