

试论科技期刊数字化数据保存

● 陈 庄

摘要: 期刊数字化数据存在着较严重的文件质量问题, 同时存在保存机制缺陷。针对数字化文件来源与制成对其文件质量的影响, 期刊编辑应发挥职业特点, 在编辑加工过程中提高数字化文件质量, 并选择适当的文件格式, 采用异地迁移保存方法, 维护科技期刊原始数字化数据长期保存安全性和再利用有效性。

关键词: 科技期刊 数字化数据 数据保存 科技期刊编辑

作为记录科学技术文献的载体, 科技期刊在社会发展中具存史价值。^[1] 科技期刊存史功能发挥的基础是期刊文献保存。科技期刊保存包括印本保存与数字化数据保存, 这两类保存都是必不可少的, 网络出版时代数字化数据保存尤其重要。但长期以来我国书刊的计算机排版文件资源流失严重, 给我国的数字出版建设带来重大损失。^[2] 科技期刊的数字化数据由哪些机构负责保存, 以及应当把哪些数字化数据当作最适宜保存对象, 这是期刊数据保存必须考虑的问题。期刊数据保存关系到期刊数字化建设的成效, 期刊数字化文件的格式、内容质量又与电子文稿的编辑加工过程息息相关。因此, 期刊编辑在工作中应当发挥职业特性, 做好科技期刊的数字化数据保存, 为数字信息安全提供保障。

一、科技期刊数字化数据保存现状与问题

欧洲的 DELOS、NEDLIB、ERPANET、PRESTO、SEPIA、MMP、EVA, 澳大利亚的 RANDORA, 美国的 Minevra 等项目致力解决数字资源的长期存储, 使得数字长期保存在技术及标准方面取得长足进展。^[3] 为做好数字出版学术期刊的保存, 我国国家数字图书馆于 2003 年加入了 LOCKSS 项目, 但受版权许可、缺乏社会支持等影响, 数字期刊的保存在我国

仍然是一项长期而艰巨的任务。^[4]

科技期刊数字化文件包含了插图、表格、版式文件、内容文本文件等, 这些反映期刊信息内容的数字资源, 成为将来数字化信息资源整合、进一步出版传播的基础。科技期刊数字化文件保存对象应当包含这些内容在内的所有信息。就现有的网络发行学术期刊来说, 我国的学术期刊数字化文件普遍存在质量问题, 突出表现在版式差、有效文本化率低等方面^[5], 因此不是最佳期刊原始数字化文件对象选择。我国科技期刊数据库的出版流程中, 先由期刊出版单位提供数字化原始文件, 再经数据库加工汇编上网。大多数期刊同时在多家数据库出版发行, 各数据库加工制成的上网文档质量参差不齐, 且各自标识数据库标徽。少数期刊委托数据库独家收录, 所出版文件也未达到原版式质量要求。

虽然目前期刊出版单位的编辑出版电脑化程度各不相同, 但总的来说, 其期刊数字化数据可分为成品文件与原稿文件。成品文件包括用于网络出版的期刊网络版文件、用于印刷的排版文件; 原稿文件主要为作者投寄的电子原稿, 包括各种格式的电子文本, 以及测试输出、影像、扫描等文件。经考查, 国内科技期刊数字化文件的保存状况如表 1 所示。

以上各类文件中, 期刊网络版文件主要是由各

表1 科技期刊数字化数据的保存情况

文件	用途	来源	保存单位	安全信任度	保存期限
网络版文件	网络出版	数据库	数据库	高	合作期
排版文件	印刷	印刷厂、期刊单位	印刷厂 期刊单位	低 中	无 不确定
原稿文件	编辑、排版	作者	作者 期刊单位	低 中	不确定 不确定

合作的大型数据库制作和保存，保存安全系数高，但保存期限仅限于合作期之内。合作期之后，数据库方不保证数据的继续保存及移交。期刊出版单位在往后的过刊网络出版中，可能面临数字化资料再制作。缺乏排版文件的众多期刊数字化再制作工程耗费巨大。这种社会资源浪费，在多年前国内学术期刊初次加入数据库出版时已经出现过，究其原因，是期刊出版单位不能提供必需的原始数字化数据，特别是排版文件。当时期刊印刷单位只负责交付印刷本，不提供排版数据，也不负责保存数据，造成了网络出版早期几乎全部期刊原始数字化数据遗失的局面。

二、科技期刊数据保存类型和选择

科技期刊数据信息可分为许多种类，以出版成形状态可分为原稿文件、排版文件和原版式网络版文件，同属于科技期刊原始数字化数据；以文件格式可分为方正书版、方正飞腾、Word、pagemaker、Adobe indesign 等所生成的文件；以内容形式可分为：文本文字、插图、表格、数据、公式等；以及文档背景记录：文件创建时间、修改时间等。排版文件主要用于印刷和输出 PDF 格式网络版文件，是期刊出版内容和版式的最直接来源，因此，各类期刊数字化文件中以排版文件最为重要。笔者调查了 370 家依托大型数据库发行网络版的科技期刊的排版数据情况，结果表明我国科技期刊排版以方正书版为主，占 82.2%（表 2）。其中 87 家（23.5%）期刊

编辑部不存有期刊排版文件，这说明接近 1/4 的科技期刊出版单位还未存有期刊排版数据；另外 31 家期刊单位（8.4%）只存 PDF 文件，其中超过一半不能完全文本复制，这部分不适宜作为网络出版文件。

文件长期保存必须考虑在再使用中体现保存价值。期刊排版使用比例最多的方正书版，其排版文件转制网络版（如 PDF）存在版式精确还原与文本化不能同时兼得隐患^[6]。现存方正书版排版文件能用于再次印刷，但作为网络出版的储备资源，因其在直接应用制作网络文件中存在明显不足，成为科技期刊网络出版发展中的瓶颈。在我国，万方和维普数据库极少直接使用方正排版文件来转制网络版文件，则说明了这个问题。

Word 成功用于网络环境下的科技期刊出版^[5]，则为科技期刊排版文件选择与数据保存方式提供了一条新途径。虽然 Word 作为期刊出版排版之用目前为数不多（表 2），但以 Word 为代表的办公软件具备广泛使用基础，在网络出版更是可以高效、准确地完成格式转换对接，将来以它作为储备文件格式，可满足实用、安全、长期保质的要求。

期刊出版流程中，文件格式转换一般循“→排版文件→网络版文件”方向，如操作方法得当，到达流程末端时文件信息能保质保真，但往往其间会出现信息损耗，比如原稿中图像文件分辨率下降。部分期刊出版单位只保存 PDF 格式文件，这会因字体、符号、图像分辨率等问题不适合再次印刷。^[7]在期刊出版流程中，PDF 格式文件属于“下游”文件，从保质的角度考虑，还应注重“上游”文件收

表2 科技期刊排版文件格式与期刊数量

文件格式	方正书版 (.s2、.s72、.s92、 .ps2、.mps、.s10 等)	方正飞腾 (.fit)	Word (.doc、.docx)	其他 (.p65、.cdr、 .tex、.indd)
数量	304	37	17	12
比例	82.2%	10%	4.6%	3.2%

藏。“中游”中排版文件包括文字信息和排版信息,便于网络出版进行各种信息处理和再加工;“上游”文件就是原稿文件。“排版文件+原稿文件”是文件“复原”的基础,对期刊数字化文件保存来说,排版文件与 Word 原稿文件同等重要。

三、期刊编辑应在原始数字化 数据保存中发挥重要作用

网络出版各环节中,因作者只拥有初始论文数据,而数据库商又局限于合同义务与保存期限制约,并且不收藏原始数据,所以两者都不是最理想的期刊原始数据保存者。期刊出版单位拥有最完整的期刊出版原始资料,包括电子原稿、排版文件等,这些都是再次出版的基础资料,具备期刊原始数据保存运作的最理想物质条件。期刊出版单位掌握了“上游”、“中游”文件资源。“上游”文件来源于作者,集中于期刊出版单位。“中游”文件中的修改稿是编辑与作者共同编辑加工而得,排版文件则是编辑、作者与印刷排版部门共同编校的结果。这些原始数据是期刊出版单位独有的,也是现有文件中最值得保存的。表 2 表明,为数不少的期刊单位对原始数据的保存意识不强。当前,期刊出版单位人员编制很少,机构内很难配有专人用于数据维护。在信息安全与知识产权问题未得到较好解决之前,借用单个商业机构进行数据维护管理并不是成熟做法。数据库收藏数据并不是原始数据,再说数据库无法保证合作期限之后的数据保存。原始数据一旦在缺少有效备份的情况下遗失就几乎无法再获取。可见,科技期刊数字化数据保存工作必须及早在编辑队伍中开展。期刊编辑应当在科技期刊发挥传承科技文献的社会功能中担负起不但仅为期刊生产者,还应当成为期刊文件管理者的责任,增强文件保存意识。

期刊出版单位应当采取必要的措施,规避数据损失和失效风险,保障期刊数据的长期安全。数字化信息长期保存的措施是定期更新数据介质以及数

据异地迁移保存^[8]。数据迁移是数据保存的有效方法:将数字信息迁移到稳定性高的媒介上,从对软件依赖强的格式转换到对软件依赖程度低的格式上^[3]。根据数字化数据安全保存原则,期刊编辑可通过当地硬盘和电子邮箱存放相结合的方式,为数据保存多个复本,提高数据保护安全系数。

1. 提高备份数据存放介质数量

一般单期期刊的数字资源包括原稿、修改稿、排版文件等,其大小在几十至两百兆(M)之间,近两年主流硬盘(包括移动硬盘)数百吉(G),已能存放成百上千期期刊资料。期刊出版单位一般拥有多台电脑,期刊编辑可利用这些已有资源,在多台电脑磁盘中存放期刊备份文件,还可复制到移动硬盘中,异地保存,以增强数据备份安全性。为提高磁盘使用寿命,选用品牌硬盘,另外还要加强硬盘维护:防病毒,防灰尘,避免置于过热过冷环境(温度以 15~25℃为宜)^[9]。高校主办期刊还可复制文件到学校档案馆,进一步提高数据保存效果。

2. 借用电子邮件服务邮箱空间异地存放

使用电子邮件网络储存,可以降低保存成本,实现数据转移^[10]。国内各大型电子邮箱相继扩容,例如网易邮箱容量无限,能发送 2G 或更大的附件,服务器均托管在国家级顶级机房,配备国际顶级杀毒引擎,保证保存与通信安全。在专门费用和设施配置未到位情况下,选择多个电子邮件邮箱寄存资料,是目前各期刊出版单位数据异地安全保存的最佳途径。以《广东海洋大学学报》为例,每期各类数据平均总计 145M,同时发送包括网易在内的 4 个本刊注册邮箱保存,快捷简便。

3. 收集原始数据务必齐全、高质

保证资源的完整性及长期有效性,拥有高质量和丰富的原始数据对数据库和期刊单位来说是至关重要的。数据齐全是未来数据重现的最基本保证。理论上,上下游文件可互相转换,使用上游文件可生成下游文件,但会耗费人力物力;有时缺失上游文件会影响到下游文件的正常使用,比如,在缺少

谈文化产业竞争力视域下 青年期刊的发展

●王玉玲 陈训明

摘要：当前，非时政类报刊出版单位文化体制改革正进入新阶段，从事业向企业的转向，背后的核心是产业化发展观，关键是培育核心竞争力。笔者立足文化产业发展规律，提出辨证、思危、超前看待青年期刊的发展，并在战略、定位、团队、机制等四个方面夯实推进。

关键词：文化产业 青年期刊 发展策略

青年期刊是我国传媒业的独特形态。伴随着 20 世纪 80 年代的“新闻改革”，90 年代的“媒体改革”，新世纪的“文化体制改革”的时代浪潮，历经着“从文化工业到文化产业”与“从文化事业到文化产业”的业态洗礼，青年期刊走过一条起步辉煌、

中程震荡、重塑自信的发展道路。面向新世纪，惟有以创新求奋进，顺应科技进步的快速变迁与受众需求的细分变化，从文化产业发展规律去探索实践，着力打造内生的核心竞争力，才能重新走出属于自己的独特路径并创造二次辉煌。

图片文件情况下，方正小样文件（fbd）不能正常使用。委托专业公司制作排版文件的期刊单位，争取将每期排版资料悉数收回本部，与其他原始文件一同保存。期刊数据的收集不但要齐全，还需保证数据应用质量——永久有效性。数据主要应用于未来数字化出版，文本化是数字化出版的基础。目前各大型科技期刊数据库现行版本距离原版式要求还有一定距离，文本化覆盖率还有待提高，并且所出版文件标识有数据库所有权徽标，因版权所限，不宜用作未来新版权的网络再出版。鉴于今后越来越多的海外发行需求，科技期刊应着眼于未来世界出版业的发展方向，建议各期刊出版单位尝试排版文件格式更改与保存，排版文件制作符合世界出版和阅读使用主流要求，考虑使用 Word、Fontpage 等通用排版软件，保证期刊数字化数据长期有效性。

参考文献

[1] 中国科学技术期刊编辑学会. 科学技术期刊编辑教程 [M]. 2 版. 北京：人民军医出版社，2007：24~25.

[2] 张海游. 国外电子期刊的存取及长期保存现状研究 [J]. 科技情报开发与经济，2010，20（2）：35~37.

[3] 周伟. 数字资源长期保存问题研究 [D]. 长春：吉林大学，2008.

[4] 安艳杰. 数字学术期刊的长期保存研究 [J]. 现代情报，2006（6）：129~130.

[5] 陈庄，禚胜修. 实现从电子稿到期刊印刷版和网络版的编排一体化 [J]. 科技与出版，2009，21（5）：48~50.

[6] 陈庄. 科技期刊网络版的版式研究 [J]. 中国出版，2008（8）：52~54.

[7] 孟玫. 什么是适合印刷的 PDF 文件（上） [J]. 印刷杂志，2007（7）：45~47.

[8] 刘家真，许杰. 影响我国数字信息长期保存的问题与解决方案 [J]. 档案学研究，2008（4）：50~55.

[9] 杨丽萍. 电子图书档案载体的科学保护 [J]. 档案管理，2005（5）：73.

[10] 孟立军，闫斌. 用电子邮箱实现网络存储 [J]. 陶瓷研究与职业教育，2005，3（4）：39~40.

（本文作者单位：《广东海洋大学学报》编辑部）