从书信沙龙到开放获取
——刍议学术论文形态的演化*

胡志刚，侯海燕，林歌歌
（大连理工大学科学学与科技管理研究所，大连 116024）

摘要：近年来，为适应互联网和大数据时代的新形势，学术期刊出版在形态上不断创新，给学术论文的写作、阅读、评价等带来很多新变化。学术论文文本的可读性、机读性和可交互性水平大幅提高，对于应对大数据时代信息爆炸和文献分析需求具有重要价值和意义。与此同时，开放获取运动促进新型学术期刊出版和发行模式的兴起，新的学术交流和知识传播模式被提出，简化从作者到读者的学术成果路径，推动学术成果更广、更快地传播。本文回顾学术论文从诞生到规范化、数字化、结构化、开放获取等里程碑式的过程，对于我国学术出版和数字图书馆领域找准发展方向、追赶发展潮流具有重要意义。

关键词：学术论文；开放获取；学术出版；知识传播
中图分类号：G255.75
DOI：10.3772/j.issn.1673-2286.2016.10.005

随着学术期刊数字出版和数字图书馆技术的发展，学术知识的传播和学术信息的交流渠道正变得越来越通畅，学术论文文本在写作格式、展现形式和评价方式上发生深刻变化。兴起中的开放获取运动，将数字出版和发行的模式推到一个新的高度，以PLOS、Frontier、PeerJ等开放获取期刊网站为代表，学术论文文本在美观度、机读性、交互性和多元指标的可视化展现等方面更加突出。可以说，学术论文文本已进入一个内容与形式并重、人机友好度并重的新阶段。

学术期刊和学术论文的诞生，可追溯到《学者杂志》（*JOURNAL DES SÇAVANS*）和《哲学汇刊》（*PHILOSOPHICAL TRANSACTIONS*）的创办[1-2]。从此，学术知识的传播方式从书信、沙龙和演讲的非正式形式[3]，发展到目前利用数字出版和开放获取进行知识生产和传播的新模式，经历了长达300多年的发展历程。这一历程在过去的半个世纪里，更是取得突飞猛进的进展。随着计算机和互联网技术的发展，数字出版和数字图书馆技术开始兴起并蓬勃发展[4]，学术论文文本中集成的内容越来越立体和丰富，学术知识传播的深度、广度和效率大幅提高。

本文对学术论文文档的演变历程进行回顾，梳理学术论文走向规范化、数字化和开放化的演化之路，并探讨这种演化对科学知识的传播和交流带来的影响。此外，通过展现数字出版和数字图书馆技术的新变化，探讨我国在数字出版和数字图书馆领域与国外出版商的差距。

1 学术期刊和学术论文的诞生

学术论文是记录和传播科技知识的载体，是对学术研究的描述和文本展现。早期科学家通过著作、书信、沙龙和演讲等渠道发表个人研究成果。与当今科学发展依赖高校和科研机构不同，早期科学是在沙龙和各类科学社团里发展起来的，如罗马的“山猫学会”、蒙托学

* 本研究得到国家自然科学基金项目“开放获取背景下的全文引文分析方法与应用”（编号：71503031）资助。

会、皇家科学院等。17世纪以前，欧洲科学家主要通过私人通信来交流学术问题或宣布其研究成果。例如，牛顿与胡克、哈雷等曾在通信中讨论引力的测算问题，胡克、哈雷就这一问题向牛顿提出自己的看法和建议，这对于牛顿重新完善引力的测算起到了很大作用[5]。

随着科学研究的发展加快，书信沙龙的形式越来越不能适应科学发展的需求。1665年，在科学学会的推动下，世界科技期刊《学者杂志》和《哲学汇刊》相继创刊，开启了科技期刊出版和学术论文写作的新时代。科技期刊为科技成果的展示、传播和交流提供稳定可靠的平台，逐渐成为科学家提出新发明、传播新理论的主要方式。

早期学术期刊主要用于发表自然科学领域的实验研究成果和发现，此外，还用于发布包括图书目录、著名人物讣告、民事和宗教法庭的重要文告、判决及高校的决议通告等[6]。学术论文在形式上以书信体、实验报告体为主，通常为单一作者。在一篇论文（主要是实验报告）中，可能会同时论及多个不同研究问题，论述形式不拘一格，但以时间顺序的描述为主。

2 IMRAD结构：学术论文内容的规范化

在学术论文诞生的前两个世纪里，发表的学术论文只是简单地记录做了什么和看到了什么，注重研究结果（Results）的描述和科学观点的提出。1850年以来，随着科学研究中对研究方法（Methodology）的强调，“实验和方法部分”开始在学术论文中独立成节[7]。1935年后，包含引言（Introduction）、方法（Methods）、结果（Results）和结论（Discussion）四个部分的IMRAD论文结构开始在医学领域的期刊中出现，并逐步被其他期刊和学科采用。1980年，IMRAD结构已经成为科技论文中普遍采用的一种形式[8]。

IMRAD指论文的四个主要组成部分：引言（Introduction）、材料与方法（Materials and Methods）、结果（Results）、讨论与结论（Discussion and Conclusion）。其中，“引言”部分主要描述研究问题和研究背景，并对前人的研究进行综述；“材料与方法”部分主要介绍作者采用的材料、方法和技术路线；“结果”部分主要展现研究所得结果；“讨论和结论”部分则是对研究结果的分析和解读，并对引言提出的问题进行回答。

对不同学科而言，IMRAD结构还有很多变体。比如，在数据驱动型学科，“材料与方法”被相应改成“数据与方法”（Data and Methods）；有些论文在“引言”之后还有独立的“文献综述”（Literature Review）部分，以更为系统地展现已有研究进展和前人研究成果；有些论文将“讨论”和“结论”部分分开论述，或者只有“结论”，又或者只有“讨论”。

1990年后，学术论文的规范化写作从正文过渡到摘要[9]，结构化摘要的出现始于生命科学领域。结构化摘要要求论文作者逐一列出论文背景（Background）、目标（Objective）、方法（Material and Methods）、结果（Results）和结论（Conclusions）。显然，由于摘要是读者最先阅读的部分，具有判断一篇论文是否具有阅读价值的作用，因此结构化摘要的出现大大提高读者筛选论文的效率。

高度规范化的学术论文结构，既方便学术论文的写作和编辑出版，也方便学术知识的传播、实践和检验。读者可以根据论文作者列出的材料、数据、方法和工具对结果进行重复实验，既是对论文中科学知识的消化和学习，也是对论文结论的验证。因此，在自然科学、生命科学、社会科学等领域，采用规范的学术论文结构已经成为约定俗成的学术论文写作规范。

3 PDF格式：学术论文载体的数字化

互联网技术出现之前，学者主要通过图书馆查找文献。20世纪90年代以来，得益于电子计算机和互联网技术的发展，纸质版定期发行的学术期刊出版模式日渐式微，学术期刊出版和学术论文开启数字化进程，学者逐渐习惯通过文献数据库、机构知识库乃至社交网站查找和下载电子文本的学术论文。

20世纪末，学术论文的存储和传播进入数字化时代，亟需一种通行的文档存储和分享的格式标准。由Adobe公司开发的PDF文件格式，凭借优良的设计，在与DjVu、Envoy、Common Ground Digital Paper、Farallon Replica、XPS及Adobe自身的PostScript格式的竞争中脱颖而出，成为桌面出版工作流程最受欢迎的文档格式[10]。

PDF的含义是便携式文档格式（Portable Document Format），是一种用独立于硬件、操作系统和应用程序的方式呈现文档的文件格式。PDF格式的优点是便于阅读和打印，可读性好，格式固定，不易被修改和编辑，具有很高的安全性和可靠性。PDF文件包含固定布局的平面文档完整描述，包括文本、字形、图形及其他需要

显示的信息。自PDF2.0版后，Adobe开始免费分发PDF阅读软件Adobe Reader，使得PDF迅速成为固定格式文本业界的非正式默认标准。

PDF的出现加快了学术论文的数字化进程。世界著名的科技期刊出版商们相继建立各自电子全文数据库，提供PDF格式论文全文检索和下载。例如，世界上最大的科技期刊出版集团Elsevier通过ScienceDirect提供旗下2 000多种科技期刊的全文数据检索、在线阅读和PDF下载；Springer出版集团的SpringerLink、John&Wiley集团的OnlineLibrary等也都以PDF文档全文作为学术论文的存储和下载介质。在我国，中国知网、万方数据、维普等中文期刊全文数据库中，PDF文档也是唯一提供的全文下载格式。

相较于学术论文纸质出版，数字出版方式大幅降低科技知识的生产、存储和传播的成本；数字出版模式还极大地拓展学术期刊的读者范围，使期刊发行的边际成本变得很小甚至没有，期刊的国际化问题也迎刃而解；另外，数字出版有效缩短学术期刊出版的时滞。出版时滞一直是学术期刊出版的重要障碍，不仅影响学术期刊的竞争力，也不利于学术成果的创新性保护。

站在数字图书馆领域角度，学术论文的数字化也方便了对论文文本进行检索、管理、挖掘和分析，有助于对知识的管理和利用[11]。数字图书馆领域各种新兴文献索引、文献检索、知识挖掘和知识服务技术的应用，使学者更方便地从海量文献发现内容，从专业文献中获取知识。而学术论文的数字化，是开展所有工作的基础和前提。

4 HTML/XML：学术论文数据的结构化

学术论文中包含很多结构化信息。首先，论文的元数据包含标题、作者、机构、关键词、期刊、期卷号等信息；其次，论文的正文部分，包含图、表、章节、段落、致谢等结构化信息；最后，论文的参考文献，也是学术论文不可或缺的有机组成部分。这种元数据和结构化信息，对于学术论文的高效存储、管理和解析具有重要价值。

而PDF格式文本不易标记学术论文的元数据和结构信息。虽然Adobe公司也在推动PDF文档的元数据存储功能，以提高其在互联网时代的适用性，但因其自身定位和功能的限制，在结构化信息存储方面，PDF格式并不是最合适的选择。

HTML语言格式弥补了PDF格式的这一缺陷。HTML诞生于1989年，由欧洲粒子物理研究中心设计开发，是一种超文本标记语言（Hyper Text Markup Language）。由于其丰富的扩展性和强大的表现力，得到迅速普及并开启互联网的Web时代。在数字出版和数字图书馆领域，HTML主要用于学术论文的前台显示。Elsevier、Springer、Wiley等出版商都提供HTML格式的全文阅读功能。在论文全文的HTML页面中，通过丰富的内部链接和外部链接，可清晰展示论文的元数据信息（如作者、期刊等）和正文（章节结构、图表等）信息，并且通过集成引文链接服务网站（crossref.com、dx.doi.org等），让读者可以方便地跳转到论文的参考文献页面，从而极大地提高学术论文交互性。

随着计算机可视化技术的发展，HTML语言变得越来越复杂，兼容性问题也随之出现，为文档结构化存储和数据交换带来很多问题，这就需要一种更为精简纯粹的文本语言。在此背景下，W3C成立专门工作组开发了一种新文本语言，称为XML，即可扩展标记语言（eXtensible Markup Language）。XML是一种元标记语言，可将文档分为许多元素并对这些元素加以标识。元素的类型、属性等由DTD或XML Schema进行声明和定义。DTD（Document Type Definition），即文档类型定义，用于定义文档的结构和语法，决定一个XML实体是否“有效”。

与HTML不同，XML将文本外观从数据的内容和结构中分离，变得更简洁和纯粹，因此很快成为数据交换主要的公共语言，在数字出版中广泛应用[12]。当前，很多学术期刊出版商均以XML文档作为出版工作流的媒介。沈锡宾等考察美国动物科学学会联盟基于XML的科技期刊出版工作流案例，展现了出版商如何借助相关软件将Word文档转换成XML文档，并基于XML进行HTML语言的排版和在线出版[13]。

从某种意思上说，XML有点类似于数据库，XML中的元素、元素属性及元素和元素属性间的关系，构成数据库中的表以及字段。借助简单的计算机程序，可以很容易地将XML格式的文档转化为其他格式的关联数据。

XML还与学术出版中的本体技术和语义网技术有关。本体技术是为建立统一的数字资源模型而采取的一种解决方案。在数字出版中，需要建立一个统一的数字资源模型，以表示和记录这些资源的内容结构、内容特征、表现形式、载体方式以及资源间的关系，并按不同粒度进行封装、复用和交换。本体技术作为概念化明

确的规范说明，可使人与人之间、计算机系统与计算机系统之间、人与计算机系统之间的知识处理和知识交换变得方便和顺畅。

语义网是一种数据的网络，它不仅把数据放在互联网上，而且试图将数据联系起来，并产生数据与现实事物的联系，以方便人与机器阅读、理解这些数据。当前语义网主要表达形式之一是关联数据，它采用RDF（Resource Description Framework）数据模型，利用URI（Uniform Resource Identifier）命名数据实体，发布和部署实例数据和类数据，从而可以通过HTTP协议揭示并获取数据，同时强调数据的相互关联、相互联系，以及有益于人机理解的语境信息。其中，RDF作为XML的一种衍生版本，是关联数据的基本数据模型。

5 开放获取：学术论文文本的开放化

继数字出版之后，学术出版业即将迎来一场关于开放获取的新革命。开放获取是在学术论文数字化和结构化的基础上发展起来的，数字化和结构化为开放获取提供广度和效度上的保障，使学者可以更便捷地获取所需文献资源。

自《布达佩斯开放获取计划》（*Budapest Open Access Initiative*）发布以来，越来越多的学术期刊开始尝试开放获取。近年来，随着*PLOS*（2006）、*Natural Communications*（2010）、*Scientific Report*（2011）、*PeerJ*（2012）等新型开放获取期刊的创办，人们对于开放获取运动的关注越来越多。根据DOAJ（Directory of Open Access Journals）网站的统计，截至2016年8月，开放获取期刊已经有9 188种，这个数字还在不断增长。

开放获取，指读者可以通过公共网络免费获取所需文献，开放获取的全文允许读取、下载、拷贝、分发、打印、检索，以及法律允许的其他目的[14]。根据学术出版与学术资源联盟、PLOS等机构联合发布的《期刊开放获取手册》的指导原则[15]，开放获取由6个原则构成：读者权（Reader Rights）、再使用权（Reuse Rights）、版权（Copy rights）、作者发布权（Author Posting Rights）、自动发布（Automatic Posting）以及机读性（Machine Readable）。这其中既包含读者的权利（读者的免费阅读权）和作者的权利（作者发布权），也包含第三方的权利（再使用权、自动发布、机读性）。

读者免费阅读权和作者免费发布权，打破了从作者到读者的知识流通障碍，使学术论文可以最大程度地传播，大幅提高学术交流的广度。论文作者可以借助社交媒体（如Facebook、Twitter、ResearchGate、Mendeley等），自由地分享和推荐自己的最新研究，让学术成果更快地被读者了解和讨论，从而给论文带来更多的关注和引用。Norris等研究学术论文的开放获取引用优势，发现相比非开放获取论文，开放获取论文具有更高的平均被引次数[16]。

开放获取是积极支持Altmetrics指标的先锋。Altmetrcis指标，一般译为替代计量指标，衡量学术论文在社交媒体和学术社区上的分享和转载次数等，广义上也包括引用指标之外的阅读（view）和下载（download）次数等[17]。在传统的期刊出版商对Altmetrics指标等持保守和观望态度时，开放获取期刊已普遍将Altmetrics指标作为标准配置之一。在PLOS、Frontier、PMC、BMJ Open等期刊网站的论文阅读页面上，都提供了一篇论文的Altmetrics指标信息。Altmetrics指标改变了传统基于引用次数的单一评价方式，将Web 2.0时代的新生事物囊括在对学术论文的单篇论文评价中，使学术论文在文本展现和评价体系上更加立体和多元。

此外，开放获取期刊通常还支持XML格式的论文全文下载，以方便第三方对论文的再使用和再发布。XML格式比PDF格式更便于全文解析和元数据提取[18]。XML利用标签来标记数据和定义数据类型，使得对全文的解析、元数据提取简单化[19]。一些常用的程序语言（如JAVA、Python、PHP、R等程序语言），也都支持XML文档的解析。

6 结论和启示

近代科学诞生以来，科学技术的进步与发展，一直伴随着学术期刊和学术论文的发展和演变。尤其是近50年，不仅在期刊数量上呈现爆炸式增长，在学术期刊的出版形态和发行模式方面也发生深刻的变化。学术论文文本在规范化、数字化、结构化和开放化等方面的发展，有效地提高了知识传播的效率和广度。

学术期刊和学术论文的演变是由科技发展的内在要求和外部环境共同决定的，是从量变到质变的过程。科学研究的逻辑性、严谨性和可重复性决定了学术论文必须采用一种共同遵守的规范模式进行论文写作和生产；信息化和数字化时代的到来为学术论文的传播和流通提供新的机遇和条件。随着学术期刊和论文数字

化进程的完成，由期刊出版商主导的封闭式学术论文网站开始成为学术交流中主要的甚至唯一的壁垒，在这一背景下开放获取运动应运而生。

图1展示了学术论文规范化、数字化、结构化和开放化间的内在关系。规范化和数字化构成对学术论文进行结构化存储的内容条件和形式条件。从内容方面，学术论文逐渐形成IMRAD结构写作规范，既体现知识生产过程中的完备性，又保证知识传播过程的严谨性；在形式方面，学术论文从传统的纸质出版和发行方式，转变成以PDF、HTML/XML文档为载体的数字出版和发行方式。通过对学术论文的数字化和结构化存储，打通了知识生产、传播、利用的环节和链条，为学术交流和协同创新提供全方位的知识服务平台。

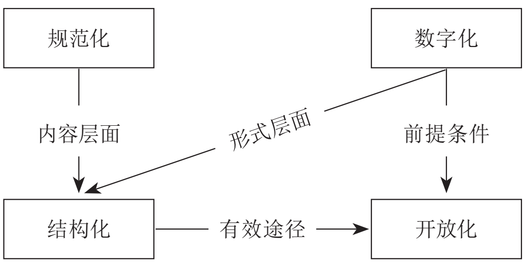

图 1 学术论文的规范化、数字化、结构化和开放化间的内在关系

论文的开放性或开放获取是未来学术出版的新模式。开放获取也是科学研究的天然属性。由于科学研究具有公有性，而学术论文最重要的作用是获得最大范围的学术承认，赢得优先发表权，所以学术论文本就应该面向读者开放获取。但是，在纸质化出版时代，由于出版和发行成本的约束，开放获取无法实现。

数字化和结构化为学术论文的开放获取提供前提条件和有效途径。数字化是实现开放获取的必要前提，保障开放获取论文可以零成本地进行分发；而结构化是进行开放获取的有效途径，只有结构化存储的数据才能最方便地被读者检索、阅读和分享。开放获取不仅进一步提高学术论文在互联网的可见度和曝光度，而且让读者更好地参与学术论文的互动中。通过集成Altmetrics计量指标，开放获取期刊网站可以记录并展现论文读者的阅读、下载、分享和引用等行为，把读者行为信息集成到学术论文的展示页面。开放获取不仅是数字出版模式上的创新，也是数字图书馆技术适应Web 2.0和大数据时代的重要举措。

应该看到，我国学术出版无论是在学术论文的文本上还是在发行模式上都远落后于国外。由于缺乏先进的期刊出版服务商的技术支持，国内期刊网站建设还比较陈旧，大部分还停留在PDF格式的数字化时代。绝大多数期刊网站乃至主要的期刊全文数据库（如中国知网、万方数据等）都不支持HTML格式的论文阅读，更难达到对各种Web 2.0元素和Altmetrics计量指标的集成。如何通过借鉴国外学术论文文本处理的先进工具和技术，提高我国学术论文的数字化水平，抢占开放获取时代学术出版的新机遇，是我国学术出版中值得关注的重要课题，也是推动数字出版和数字图书馆融合发展的重要途径。

参考文献

[1] 李武.最早的两份学术期刊[J].科技导报,2012,30(10):81.

[2] 严小涛,廖建军.期刊的诞生与存续[J].求索,2012(12):270.

[3] DAMPIERW C.科学史及其与哲学和宗教的关系[M].李珩,译.桂林:广西师范大学出版社,2009.

[4] 傅强.数字出版:新的革命[J].浙江大学学报(人文社会科学版),2008,38(4):84-89.

[5] 沃尔夫.十六、十七世纪科学、技术和哲学史[M].周昌忠,苗以顺,毛荣运,译.北京:商务印书馆, 1997.

[6] 张耀铭.学术评价存在的问题、成因及其治理[J].清华大学学报(哲学社会科学版),2015, 30(6):73-88.

[7] DAY R A, GASTEL B. How to write and publish a scientific paper[M]. 7th Rev. ed. New York: Greenwood Press,2011.

[8] SOLLACI L B, PEREIRA M G.The introduction, methods, results, and discussion(IMRAD)structure:a fifty-year survey[J].Journal of the Medical Library Association Jmla,2004,92(3):364-367.

[9] HAYNES R B,MULROW C D,HUTH E J,et al.More informative abstracts revisited[J].Annals of Internal Medicine,1990,113(1):69-76.

[10] 张立.数字出版相关概念的比较分析[J].中国出版,2006(12):11-14.

[11] 邱均平,沙勇忠,刘焕成.论数字图书馆的知识管理[J].情报资料工作,2001(5):5-8.

[12] 白杰,杨爱臣.XML结构化数字出版的特点与流程[J].出版广角,2015(5):28-31.

[13] 沈锡宾,顾恬,吕小东,等.国外一基于XML的科技期刊出版工作流个案剖析[J].中国科技期刊研究, 2011,22(4):581-583.

[14] 李春旺.网络环境下学术信息的开放存取[J].中国图书馆学报,2005(1):33-37.

[15] How Open Is It? Open Access Spectrum[EB/OL].[2016-09-01].http://

blogs.lib.ucdavis.edu/schcomm/2013/10/22/how-open-is-it-open-access-spectrum/.

[16] NORRIS M, OPPENHEIM C, ROWLAND F. The citation advantage of open-access articles[J].Journal of the American Society for Information Science&Technology, 2008, 59(12):1963-1972.

[17] PRIEM J,TARABORELLI D,GROTH P,et al.Altmetrics:a manifesto[EB/OL].(2010-10-26)[2016-09-01].http://altmetrics.org/manifesto.

[18] 李珍,田学东.PDF文件信息的抽取与分析[J].计算机应用,2003,23(12):145-147.

[19] 胡志刚,陈超美,刘则渊,等.基于XML全文数据引文分析系统的设计与实现[J].现代图书情报技术,2012(11):71-77.

作者简介

胡志刚，男，1984年生，讲师，研究方向：科学学与科学计量学、全文引文分析、科学知识图谱等。
侯海燕，女，1971年生，教授，研究方向：科学计量学、科技政策与管理，E-mail：htieshan@dlut.edu.cn。
林歌歌，女，1992年生，硕士研究生，研究方向：科学计量学。

From Letters and Salons to Open Access: Discussion on Evolution of Academic Papers

HU ZhiGang, HOU HaiYan, LIN GeGe
(Institute of Science of Science and S.T Management, Dalian University of Technology, Dalian 116024, China)

Abstract: In recent years, in order to fit in with the new era of the internet and big data, academic periodical publishing makes continuous innovation in the form and pattern, and brings a lot of new changes to academic paper writing, reading and evaluation.The machine-readability and interoperability of academic paper text is extremely helpful to deal with the information explosion of big data era, and quite useful to satisfy the literature analysis demand. Besides, open access movement, an innovatory publishing and distribution pattern of academic journals is rising recently, which puts forward a brand new mode of academic exchanges and knowledge propagation, and greatly simplified the path from the author to the reader. Due to open access movement, academic knowledge is able to spread broader and faster. In this paper, the evolution of academic papers and publishing is reviewed. Its landmarks of process, including its birth, standardization, digitalization, structuration and open access are discussed. This review could be greatly significant for China academic publishing on how to find the development direction and keep pace with the trends.

Keywords: Academic Papers; Open Access; Academic Publishing; Knowledge Propagation

（收稿日期：2016-09-02）