

# arXiv 可持续发展计划的运营与管理机制剖析\*

■ 白林林<sup>1,2</sup> 祝忠明<sup>1</sup>

<sup>1</sup> 中国科学院兰州文献情报中心 兰州 730000 <sup>2</sup> 中国科学院大学 北京 100049

**摘要:** [目的/意义]对 arXiv 实施的可持续发展计划、运营与管理模式进行剖析,以期对我国 2016 年成立的“中国科学院科技论文预发布平台(ChinaXiv)”在管理运营方式、服务、发展等方面提供借鉴。[方法/过程]对 arXiv 实施的可持续发展计划中采用的管理与经营模式、执行路径图进行详细剖析。[结果/结论]arXiv 可持续发展计划强调从外部获取资金资助,成员费用是根据下载文章数量计算的机构排名,筹资方式透明化。在科学咨询委员会和成员咨询委员会代表成员机构利益和对用户的需求调研情况下,来源于实际的系统发展、升级等方面的意见值得借鉴。且 arXiv 开始关注与其他机构知识库的互操作性,这种思想也值得我国在运营 ChinaXiv 时考虑。

**关键词:** arXiv 可持续发展计划 管理模式 运营模式 路径图

**分类号:** G254

**DOI:** 10.13266/j.issn.0252-3116.2017.11.007

## 1 引言

arXiv<sup>[1]</sup>(以前的 xxx.lanl.gov)是一个预印本论文存储、检索、发布和交流的开放知识库平台和服务器,覆盖领域包括物理、数学、计算机科学、非线性科学、定量生物学、定量金融学和统计学。最初是由量子物理学家保罗·亨利·金斯帕(P. H. Ginsparg)于 1991 年在美国洛斯阿拉莫斯国家实验室成立,自 2001 年开始,转而由美国康奈尔大学图书馆运营。2010 年 1 月,为了确保 arXiv 的可持续发展,康乃尔大学图书馆开始了历时 3 年的 arXiv 长期可持续发展(arXiv sustainability initiative)模式探索,以促进 arXiv 由康奈尔大学图书馆向多机构协同管理模式的转变。该可持续发展计划从 2013 年 1 月份开始执行,为期 5 年。截至 2017 年 2 月,arXiv 可持续发展计划已执行 4 年,共有来自 25 个国家的 206 个成员机构参与<sup>[2]</sup>,存储的数据量在 2016 年底达到了 1 219 219 篇<sup>[3]</sup>。在西班牙国家研究委员会(CSIC)网络计量学实验室(Cybermetrics Lab)在 2016 年 7 月、2017 年 1 月相继公布的第 19 版、20 版全球知识库(rankings web of repositories)<sup>[4]</sup>排名中,arXiv 都居于第一,可见其影响力之大,当然这与其

管理运营方式、提供的服务密不可分。arXiv 自成立以来,国外学者围绕其中的论文使用情况进行了大量研究,表面上看 arXiv 中论文确实具有较高的引用率,究其原因在于 arXiv 存储的论文先于期刊出版<sup>[5]</sup>。P. M. Davis 等人将 1997 年-2005 年发表在 4 种数学期刊的 2 765 篇论文作为数据源,得出在 arXiv 仓储论文的篇均被引频次为 1.1 次,比非 arXiv 存储的论文高出 35%<sup>[6]</sup>。国内学者的研究主要集中在对 arXiv 2013 年可持续发展计划实施之前运营机制的介绍及对国内预印本系统的启示、与国内预印本系统的比较和存储论文的定量分析方面,如 2009 年刘银娣从业务、财务和社会影响 3 个方面详细分析了 arXiv 的运营情况<sup>[7]</sup>;王欣等对 2008 年-2011 年发表于 Frontiers of Physics 上的 161 篇论文进行数据分析,建议国内物理学期刊出版者加强网站建设,充分利用 arXiv 平台提升期刊引用水平<sup>[8]</sup>。

本研究将对 arXiv 于 2013 年开始实施的可持续发展计划中的管理与经营模式、实施路径图进行详细分析,以期对我国 2016 年成立的“中国科学院科技论文预发布平台(ChinaXiv)<sup>[9]</sup>”在管理运营、服务和发展等方面提供借鉴。

\* 本文系中国科学院文献情报能力建设专项“中国科学院知识产权存储管理中心建设”项目(项目编号:Y6ZG421001)研究成果之一。

作者简介:白林林(ORCID:0000-0003-2265-7399),博士,E-mail:bailinlin@mail.las.ac.cn;祝忠明(ORCID:0000-0002-2365-3050),信息学部主任,研究员。

收稿日期:2017-03-28 修回日期:2017-05-12 本文起止页码:55-62 本文责任编辑:王传清

## 2 arXiv 可持续发展计划的管理和经营模式

### 2.1 arXiv 可持续发展计划的提出

2010 年,为了 arXiv 的可持续发展,康奈尔大学将 arXiv 管理模式从康奈尔单独管理转变为多家机构协同管理,旨在减少 arXiv 的财务负担和对单一机构的依赖。为了实现这个目标,康奈尔大学制定了一个临时机构自愿资助模式,邀请来自全世界 200 个 arXiv 高频下载机构的图书馆和研究实验室以会员的身份对 arXiv 进行资助。arXiv 的管理不仅仅是运营成本的管理,而且要在基于用户需求和不断发展的学术交流模式基础上提高学术资源的价值。

在 2010 年-2012 年为期 3 年的计划过程中,为了将 arXiv 定位为协同管理、社区支持的开放获取资源,康奈尔大学广泛寻找相关利益者的投入。计划中还包一场与几个出版社和协会代表讨论协同管理的可行性和可取性的会议,会议讨论了科研资料的交叉链接、互操作性和生命周期、附件材料支持<sup>[10]</sup>。

可持续发展计划过程旨在研究如何使收入模式多样化、如何确保 arXiv 满足一系列管理原则和如何提供透明和可靠的社区支持服务。最后在西蒙基金会(Simons Foundation)资助下,2010 年-2012 年 3 年计划的主要成果为 arXiv 可持续发展计划的成员协同管理模式的诞生。

### 2.2 arXiv 可持续发展计划组织管理模式

arXiv 由康奈尔大学图书馆在 arXiv 科学咨询委员会(Scientific Advisory Board, SAB)和成员咨询委员会(Member Advisory Board, MAB)的指导下,以及许多学科审核者的协助下进行维护和操作(见图 1)。其中,康奈尔大学图书馆主要分为领导小组和运营小组来对 arXiv 进行管理决策和技术维护。成员咨询委员会和科学咨询委员会是并行运营的,相辅相成,分别从发展和运营方面对康奈尔大学图书馆就 arXiv 的管理提供建议<sup>[11]</sup>。

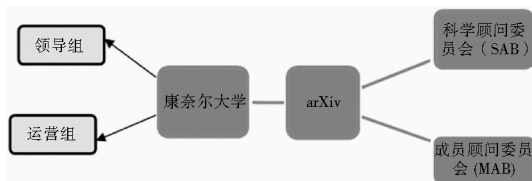


图 1 arXiv 的组织结构

其中康奈尔大学图书馆主要承担的工作是:在 arXiv 科学咨询委员会和成员咨询委员会的指导下,承担 arXiv 运营和管理的全面责任,包括最终经济责任;对提交资料的审核过程和用户支持进行管理,包括提

交过程和政策的制定和实施;arXiv 的技术基础设施运营,包括系统维护、标准实施以及用于吸收新功能和服务的计划的制定;协调和促进科学咨询委员会和成员咨询委员会的会议和审议活动;提供数据接口访问以支持 arXiv 数据的开放重用;承担归档的责任,以确保 arXiv 镜像站点长期访问和维护;通过互操作共享工具和软件与相关项目活动建立和保持合作,以提高科学界的服务;保持透明和开放的交流战略,以吸引 arXiv 的主要利益相关者;如果 arXiv 不再承担目前的服务时,负责制定“继任计划”;为了确保透明度和问责制,arXiv 为成员机构提供年度经营和财务报告。

科学咨询委员会由 arXiv 所涉及领域的科学家和研究人员组成,并从中选举一名主席,与 arXiv 的用户支持主任一起协调科学咨询委员会的活动,其中有两名是来自于成员咨询委员会的无选举权的当然委员(ex officio members)。科学咨询委员会提供有关 arXiv 知识监管的建议和指导,特别关注 arXiv 审核系统的政策和操作;审核 arXiv 中的存缴标准,有必要时建议对这些标准进行修订;提出 arXiv 覆盖的新主题或新学科领域,并对这些领域进行审核;对成员咨询委员会提出的 arXiv 发展计划,提供科学家反馈意见;就发展计划提出建议,特别是关于改进支持提交和审核处理的系统的建议。

成员咨询委员会从 arXiv 的成员中选出,成立于 2013 年。arXiv 的成员咨询委员会的代表主要是为图书馆、研究机构、实验室和对 arXiv 服务提供资金支持的基金会预留的,其中两名是来自于科学咨询委员会的无选举权的当然委员,由康奈尔大学图书馆的 arXiv 项目主任担任主席。成员咨询委员会代表的是参与成员机构的利益,就 arXiv 知识库管理和发展、标准实施、互操作性、发展重点、经营计划、外展和宣传方面的问题向康奈尔大学图书馆提供建议;对康奈尔大学图书馆提出的 arXiv 预算方案进行审议。

### 2.3 arXiv 可持续发展计划经营模式

arXiv 由康奈尔大学图书馆和全球机构成员集体提供资金支持,可能还有基金会赠款。主要包括 4 方面<sup>[2]</sup>:

(1) 康奈尔大学图书馆承诺每年提供 7.5 万美元的现金补贴,用于支持 arXiv 的运营成本和所有间接成本的实物捐助,目前占有运营成本的 37%。

(2) 每个成员机构承诺支持 arXiv 的 5 年资助。成员费分为 4 个层级,根据下载的文章数量计算机构排名,每年计算一次,从 1 500 美元到 3 000 美元不等

(见表 1), 主要集中在排名前 250 名的机构上, 因为其占了下载量的 75% (2016 年以前是前 200 名)。通过联盟协议提供支付的成员机构可以给予折扣。

表 1 成员费分级

| 层级  | 下载排名    | 年度成员费(美元) |
|-----|---------|-----------|
| 1 级 | 1-50    | 3 000     |
| 2 级 | 51-100  | 2 500     |
| 3 级 | 101-150 | 2 000     |
| 4 级 | 151+    | 1 500     |

(3) 西蒙基金会 2013 年-2015 年每年捐赠 5 万美元, 2016 年增加至 10 万美元, 用于支持康奈尔大学图书馆对 arXiv 的管理。

(4) 对于成员支付产生的税收, 西蒙基金会将匹配 30 万美金的补助。

在经费的利用方面, 对于需要额外资金的, 董事会建议必须在服务升级或开发项目实施之前确定此类资金的来源。增值服务的任何费用将被设定在足以收回提供服务的成本和促进业务运营成本的水平。需要以透明和非歧视性条款提供增值服务, 避免独家合作; 成员机构成功提供初始资金后, 如果成员的财政支持低于康奈尔认为足以保持 arXiv 的质量水平, 将会启动使用两年储备金, 以便康奈尔获得足够的资金支持; 如果不这样做, 可能导致需要制定一个退出战略, 包括将 arXiv 转型成另外一个机构。

2017 年启动的 Next Generation arXiv<sup>[12]</sup> 大改造计划成功获得了斯隆基金会 (Alfred P. Sloan Foundation) 44.5 万美金的捐款<sup>[13]</sup>, 用于启动 arXiv-NG 计划。除此之外, 支持 arXiv-NG 第一阶段的另一个资金来源是艾伦人工智能研究所 (Allen Institute for Artificial Intelligence, AI2), 提供了 20 万美元<sup>[14]</sup>, 用于支持自动质量控制和改进搜索和检索体验工具的开发。

为了鼓励全球更多的图书馆和科研机构支持 arXiv 知识库, 2017 年 1 月 26 日康奈尔大学公布 arXiv 的管理者和 Google scholar 的共同创始人 A. Acharya 合作制定了一个新的资助模型<sup>[15]</sup>: 个人赞助商可以为无力承担费用的机构提供成员费用。为了宣传这个资助模型, A. Acharya 赞助了印度 3 家领先的物理研究机构 (哈里什-钱德拉研究所<sup>[16]</sup>、物理研究实验室<sup>[17]</sup>、塔塔基础研究院<sup>[18]</sup>) 的成员费用。

### 3 arXiv 可持续发展计划的执行

下文从技术、用户支持和审核、管理新的合作伙伴和特别计划 5 方面, 对 arXiv 可持续发展计划在过去 4 年的执行情况进行总结。

#### 3.1 技术

arXiv 在过去 4 年的技术发展中, 比较注重审核系统、数据质量控制、数据保存等技术方面的改善 (见表 2)。在审核系统的技术改善方面, 根据收集到的用户需求改进和简化审核流程, 使得审核者可以更加直接高效地与 arXiv 系统和管理者进行交互, 更快地完成内容质量审核。在数据质量控制技术方面, 集成 P. H. Ginsparg 开发的自动重叠检测应用程序, 将新提交的内容与现有语料库进行比较; 添加了 ORCID 作者标识符支持, 以实现与其他知识库更好的互操作性。在数据保存技术方面, 将所有服务器迁移到虚拟机基础架构来增加存储容量; 扩展 arXiv ID 范围为 yymm.nnnnn, 使得提交数量扩展至 1 万/月。除此之外, arXiv 系统支持将研究数据作为辅助文件进行存缴和向提交系统添加自动分类检查。

#### 3.2 用户支持和审核

在用户支持和审核方面, 截至 2016 年底, 来自全球的 150 多名 arXiv 学科专家是 arXiv 成功的关键原因, 他们根据 arXiv 确定的审核流程、审核政策和公认的学术交流标准, 对提交的内容是否符合主题、是否符合科学界兴趣、适当的时候对提交的内容进行重新分类、删除、文本重叠度注释等一系列内容审核手段来进行严格的质量把关, 同时 arXiv 也制定了一些指标来对审核者进行绩效考核。除此之外, arXiv 将与用户和审核者的通信管理转移到基于 Web 的问题跟踪系统 (请求跟踪器), 取代目前基于邮件的 arXiv 管理流程, 通过问题跟踪系统可以回答用户问题、解决技术问题、与审核者进行交互、对审核申诉问题予以回应和解决。相比基于邮件的管理流程, 问题跟踪系统不仅可以更好地进行跟踪, 而且可以建立报告和问题分类工具, 用以更好地管理 arXiv。见表 3。

#### 3.3 管理

在 arXiv 管理方面, 2013 年-2016 年每年会举行科学咨询委员会和成员咨询委员会会议, 用以讨论技术发展重点、财务状况、审核工具和政策以及筹资战略等; 并对这种新的协同管理模式持续进行测试与完善, 用以明确康奈尔大学图书馆、科学咨询委员会和成员咨询委员会各自的责任与权利。可看出, arXiv 可持续发展计划特别注重科学咨询委员会和成员咨询委员会章程的制定、人员的安排和相关会议的筹办; 除此之外, 筹资也是 arXiv 可持续发展计划管理关注的重点, 尝试从其他机构或者联邦政府获得筹资机会, 从 2015 年开始了网上捐款试验, 2016 年通过网上捐款获得 3.2 万美元的收入。

表 2 arXiv 技术路径

| 年份   | 技术改进   |
|------|--|
| 2013 | <ul style="list-style-type: none"> <li>将所有服务器迁移到虚拟机基础架构来增加存储容量</li> <li>采用请求跟踪器,取代目前基于邮件的 arXiv 管理流程</li> <li>从收集用户需求开始改进支持审核过程的工具和界面</li> <li>研究数据支持:审查了 2013 年 3 月结束的数据保护试点的使用情况,将继续使用现有的辅助文件机制支持适度大小的数据集和其他材料,并将从试点获取的数据作为辅助文件,以支持其长期性</li> </ul> |
| 2014 | <ul style="list-style-type: none"> <li>开始了改进支持审核过程的工具和界面的工作</li> <li>扩展 arXiv ID 范围为 yymm.nnnnn,使得提交数量扩展至 1 万/月</li> <li>向提交系统添加自动分类检查</li> <li>向科学咨询委员会提供 arXiv 代码的摘要文档</li> <li>集成 P. H. Ginsparg 开发的自动重叠检测应用程序,将新提交的内容与现有语料库进行比较</li> </ul>       |
| 2015 | <ul style="list-style-type: none"> <li>根据科学咨询委员会和审核员的意见,持续改进工具和界面,使审核员可以更加直接高效地与 arXiv 系统和管理员进行互动</li> <li>启动了对 TeX 系统进行升级、重组的过程</li> <li>添加了 ORCID 作者标识符支持,以实现与实施权威控制的其他知识库更好的互操作性,并且作为向成员组织提供机构统计数据的途径</li> </ul>                                  |
| 2016 | <ul style="list-style-type: none"> <li>继续对 TeX 系统进行升级、重组</li> <li>简化审核者和管理员工作流程、互动过程</li> </ul>  |

表 3 arXiv 用户支持和审核路径

| 年份   | 用户支持和审核  |
|------|--|
| 2013 | <ul style="list-style-type: none"> <li>与科学咨询委员会合作,正式确定主题咨询委员会的职责</li> <li>为科学咨询委员会和成员咨询委员会举行年度会议,讨论技术发展重点、财务状况、审核工具和政策以及筹资战略</li> <li>继续评估和完善新的管理模式的运营</li> </ul>  |
| 2016 | <ul style="list-style-type: none"> <li>通过网上捐款,arXiv 收入了 3.2 万美元</li> <li>举行科学咨询委员会和成员咨询委员会年度会议,讨论技术发展重点、财务状况、审核工具和政策以及筹资战略</li> <li>继续评估和完善新的管理模式的运营</li> <li>C. Myers 离职后,继续评估 arXiv 科学主任的岗位责任,2017 年重新填补这一位置</li> <li>在人员配置方面,为了加强 arXiv 的日常监督,在 2016 年,将 arXiv 团队模式转变为有两个全职经理,一个是 IT 经理,另一个是用户支持运营经理</li> </ul> |

### 3.4 新的合作伙伴和交流

arXiv 在过去的 4 年中,一直在扩大其成员(见图 2 和表 4)。截至 2017 年 2 月,已有 25 个国家 206 个成员加入到 arXiv,按表 1 所示的下载量进行排名,排在 1 级、2 级、3 级、4 级的成员各有 42、45、36、73 个。1 级成员中有 9 个成员(如欧洲核子研究组织 CERN<sup>[19]</sup>、德国研究中心亥姆霍兹协会( DESY)<sup>[20]</sup>、哈佛大学、苏黎世理工学院、普林斯顿大学等)承诺承担更多费用以支持 arXiv-NG<sup>[21]</sup>。除此之外,5 个白金成员,包括加州理工学院、伦敦大学帝国学院、马克斯·普朗克数字图书馆、麻省理工学院、剑桥大学将在提供基本成员费用的基础上,还会承担更多费用以支持 arXiv-NG;5 个非成员机构,包括美国数学学会(AMS)、奥地利科学基金会、丹麦电子研究图书馆( DEFF)、欧洲研究委员会、Google Deep Mind 也为 arXiv 运营提供资助。在分布的 25 个国家中,以美国所占比例较高,达到 40.8%,包括 84 个机构;其次是德国、英国、日本、加拿大,分别有 32、14、14、8 个机构,我国只有中国科学院高能物理研究所和国家科学图书馆(文献情报中心)、清华大学两家机构,且 2016 年下载排行均在 4 级,下载量较低。

arXiv 支持以联盟的形式成为会员,比如德国除了马克斯·普朗克数字图书馆和欧洲南方天文台(ESO)之外的 30 个机构全部以 arXiv-DH 和 HGF 联盟的形式参加,英国的 14 个机构全部通过英国联合信息委员会(JISC)联盟参加,日本的 14 个机构全部通过 NII Japan Consortia 联盟参加,法国的 6 个国家全部通过法国开放科学交流中心(CCSD)<sup>[22]</sup>参加,而美国的 84 个成员中有 15 个成员是通过十大学术联盟(BTAA)<sup>[23]</sup>、15 个是以美国西部图书馆联盟成员的身份(GWLA)<sup>[24]</sup>参加,加州大学的 10 个分校通过加州数字图书馆(CDL)<sup>[25]</sup>参加,通过这种联盟协议参加的成员所缴的费用相比其他独立的成员费用有所降低。

arXiv 还继续与一些对康奈尔可持续发展计划工作感兴趣的出版商、协会进行对话,探讨不断发展的公开政策和强制性开放获取政策;同时也收到来自图书馆、资助机构提出的关于 arXiv 在支持新型强制性开放获取政策方面的问题和功能需求。arXiv 还计划使用 ORCID、资助 ID、机构 ID 来整合标准化元数据,继续探讨存缴和链接与论文相关的科研数据来提升 arXiv 在学术交流生态圈中的作用。

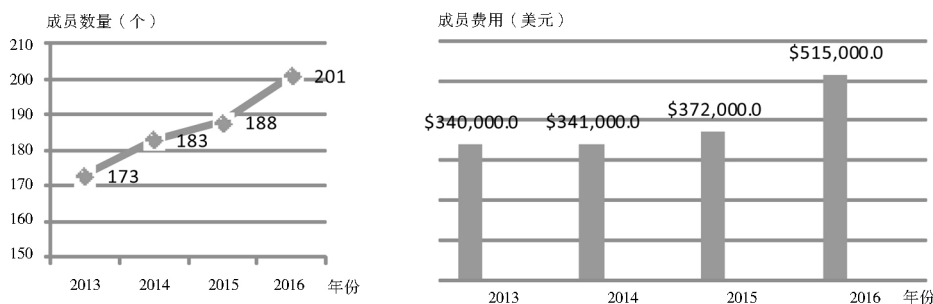


图2 arXiv 成员

表4 arXiv 新的合作伙伴与交流

| 年份   | 新的合作伙伴   |
|------|--|
| 2013 | <ul style="list-style-type: none"> <li>来自 22 个国家的 173 个成员提供了 34 万美金的成员费用</li> <li>继续与一些对康奈尔可持续发展计划工作感兴趣的出版商/社团进行对话,对新型的开放获取权限和合作想法进行讨论</li> <li>制定新伙伴关系参与 arXiv 的准则</li> <li>加强与用户(科学家)的沟通</li> <li>寻求外部资金来推进 arXiv</li> </ul> |
| 2014 | <ul style="list-style-type: none"> <li>来自 24 个国家的 183 个成员提供了 34.1 万美金的成员费用</li> <li>继续与几个出版商和社会对话,以评估在文章发布后(即在制作过程中)将文章的作者版本存入文档的想法</li> <li>让成员咨询委员会成员参与对其本国机构的 arXiv 用户体验的非正式调查,以扩大对当前新兴的使用模式的了解</li> </ul>                  |
| 2015 | <ul style="list-style-type: none"> <li>来自 23 个国家的 188 个成员提供了 37.2 万美金的成员费用</li> <li>调研了 arXiv 与其他机构和学科知识库互操作性需求</li> <li>arXiv 在学术交流生态学中的作用</li> </ul>   |
| 2016 | <ul style="list-style-type: none"> <li>来自 25 个国家的 201 个成员提供了 51.5 万美金的成员费用</li> <li>arXiv 在学术交流生态学中的作用,与其他系统实现一站式存缴</li> </ul>   |

目前,除了康奈尔大学图书馆的网站,arXiv 还在中国、德国、印度、西班牙、洛斯阿拉莫斯 5 个国家有镜像站点<sup>[26]</sup>。在欧洲的高能物理信息检索系统(stanfordphysics information retrieval system,INSPIRE)<sup>[27]</sup>、美国航空航天局天体物理数据系统(astrophysics data system,ADS)<sup>[28]</sup>和中国的中国科学院文献情报中心<sup>[29]</sup> 3 个机构的数据库中也有 arXiv 的检索界面。

### 3.5 arXiv-NG

从用户的角度看,arXiv 是一个成功的知识库系统,服务于世界各地的许多科学家。但是 arXiv 面临着巨大的压力,科学咨询委员会和成员咨询委员会在 2015 年会议上得出的结论是:arXiv 目前的重点除了维护经营模式外,还需要努力继续通过赞助和合作筹集资金。arXiv 需要首先制定一个“引人注目的愿景”,以便能够说服筹资目标,这个愿景能够超越目前可持续发展计划。

2016 年 8 月 arXiv 迎来 25 周年之际,由科学家和图书馆员组成的顾问团队开始考虑将平台改造得更为现代化,经过 2016 年几个月的规划,康奈尔大学图书馆公布 arXiv 将启动为期 3 年的 Next Generation arXiv (下一代 arXiv),简称 arXiv-NG,主要目的是对 arXiv 系

统基础设施进行大改造<sup>[9]</sup>。改造计划主要参考 2016 年 6 月该网站公布的针对大量用户调查的结果:大部分用户不希望网站大改,但提出了一些调整建议,比如升级搜索功能、网站实现对参考文献的链接、更多格式的引用导出功能等;喜欢网站现在内置的质量控制体系,包括文本检查、重复率检查、将论文分类到正确的主题领域以及驳回没有科学价值的研究论文等。康奈尔大学的物理学家 P. H. Ginsparg 说:“调查的结果表明,人们或多或少地希望网站坚持基本的传播任务,不希望因过度延伸或商业化而转型<sup>[18]</sup>。”

通过多阶段设计和开发,改造计划所需资金预算大概为 250-300 万美元。在 2016 年 12 月已完成了资金筹集,成功获得了斯隆基金会(Alfred P. Sloan Foundation)44.5 万美金的捐款,用于启动 arXiv-NG 计划。除此之外,支持 arXiv-NG 第一阶段的另一个资金来源是艾伦人工智能研究所(Allen Institute for Artificial Intelligence,AI2),提供了 20 万美元<sup>[14]</sup>,用于支持康奈尔大学图书馆与康奈尔大学计算和信息科学系之间的合作,允许 arXiv 聘请一位研究与创新研究员与 arXiv 团队合作,为 arXiv 设计和集成一系列现代化的、以研究为导向的功能,最终目标是从研究中产生的工具

集成到 arXiv 系统中,以提高用户和审核者的体验。2017 年为这项计划的第一阶段,为期 18 个月,主要是创建和实施工作计划和人员安排。

在未来开发 arXiv-NG 过程的三年中,arXiv 将继续依赖现有系统(classic arXiv),并致力于继续强大其服务。arXiv-NG 改造计划具体包括:①互操作性和公共访问权限支持:添加有关资助信息、文章状态和旧内容迁移的元数据字段;支持 arXiv 和机构知识库之间的互操作性;在数据知识库中添加数据集的链接;创建更好地与计算机科学会议进行集成的工具和设施,为数据分配 DOI;将 arXiv 内容吸收到康奈尔大学机构知识库。②界面现代化:现代化用户提交、搜索界面、添加分面分类,包括作者标识符;更换和改进警报系统;标记撤回的文章。③软件调整和改进:调整提交系统、加速传统代码库的改进。

## 4 arXiv 对中国科学院科技论文预发布平台(ChinaXiv)管理运营方面的启示

### 4.1 建立长期规划和有效的管理运营模式

arXiv 可持续发展计划协同管理模式的提出经过了一系列长期规划与短期规划,并建立了有效的组织管理体系。arXiv 的可持续发展计划从计划到诞生经历了 3 年的时间,全面实施已过了 4 年,接下来将继续花 3 年的时间实现对 arXiv 系统基础设施进行大改造。在组织管理方面,arXiv 由康奈尔大学图书馆在 arXiv 科学咨询委员会和成员咨询委员会的指导下,以及许多学科审核者的协助下进行维护和操作,同时实现与相关出版商、协会进行对话,探讨不断发展的公开政策和强制性开放获取政策各自分工明确,并定期进行管理模式的评估与完善,这种有效的协同组织管理模式确保了 arXiv 的有效管理。在运营方面,通过协同组织管理模式,确保了资金来源的多元化,以比较创新的依据 arXiv 的下载量排名的筹资方式从外部获取资金资助,筹资方式透明化,同时还扩展新的资金赞助模式(网上捐款试验、个人赞助无力承担费用的机构),这种合理的资金来源模式确保了 arXiv 持续与稳定发展。

ChinaXiv 经过前期的规划,自 2016 年中旬开始实施。在组织管理方面,由中国科学院文献情报中心信息系统部单独管理。资金来源仅仅依靠政府拨款,缺乏更为广泛和合理的资金来源模式,很难维持其持续和稳定发展。因此,ChinaXiv 可适当借鉴 arXiv 预印本

系统的一些做法。由于目前 ChinaXiv 系统处于初步探索阶段,国内权威学者仍然更倾向于将其成果通过传统核心期刊发布,因此目前要做的是,加强宣传力度,通过不断呼吁权威专家投稿以加强存储质量和学术影响力,带动其他学者加入到这个活动中来,确立系统的地位后再逐步扩大范围。在经营方面,除了借鉴 arXiv 按论文下载量对机构收费之外,还可以按下载率单独对作者收费、提供个性化服务、实行会员制管理、网上捐款试验等措施,来获得系统运行、管理经费。

### 4.2 注重质量审核控制和简化流程

arXiv 的成功一方面得益于物理学家对预印本这种交流形式的认可;另一方面与其运行中严格的质量审核控制和系统功能的不断完善是分不开的。在质量控制方面,经历了从一开始的通过对作者的注册地址中需要有个“.edu”后缀作为合法科研单位的控制、认证(endorsement)制度审核机制的引进到将新提交的内容与现有语料库进行比较的自动重叠检测应用程序。同时,全球的 150 多名 arXiv 学科专家是 arXiv 成功的关键原因,他们根据 arXiv 确定的审核流程、审核政策和公认的学术交流标准,对提交的内容是否符合主题、是否符合科学界兴趣、是否符合分类等一系列内容审核手段来进行严格的质量把关,同时 arXiv 也制定了一些指标来对审核者进行绩效考核。在保证基本质量的同时,根据收集到的用户需求改进和简化审核流程,使得审核者可以更加直接高效地与 arXiv 系统和管理者进行交互,更快地完成内容质量审核。

ChinaXiv 作为一个可靠、规范的自然科学领域的中国科研论文开放仓储库,在质量控制方面,要求用户使用邮箱注册,提交的论文进入审核阶段,通过开放评论、同行评议等即时交互,促进交流与合作,在这方面 ChinaXiv 起步做的很好。后续 ChinaXiv 还可借鉴 arXiv 系统经验,制定合理的审核流程、审核政策等,定期修订,对审核者资格制定合理的考核指标。

### 4.3 数字资源长期保存技术及时更新

在数据保存技术方面,arXiv 将所有服务器迁移到虚拟机基础架构来增加存储容量;扩展 arXiv ID 范围为 yymm.nnnnn,使得提交数量扩展至 1 万/月。除此之外,arXiv 系统支持将研究数据作为辅助文件进行存缴。arXiv 对存缴的每一条记录和修改记录都必须保持永久的记录,不允许撤销。长期保存既是保证作者作品存储安全的重要措施之一,也是保证资源能够持

续稳定地为读者所用的重要手段,因此,ChinaXiv 应借鉴 arXiv 在资源长期保存上的做法,对存缴的记录进行长期保存与记录,对不同版本之间的记录进行控制。

#### 4.4 根据实际调研对系统业务和服务进行完善

arXiv 在技术改进、系统升级或新增业务等方面,通过实际用户调研或者举行会议,根据用户调查结果或者科学咨询委员会和成员咨询委员会、学科专家审核者的意见来开展,如 arXiv-NG 改造计划主要参考 2016 年 6 月所做的用户调查结果。

ChinaXiv 刚刚起步,在保持自身已有的特色业务“我的定制推送”“最新活动”等基础上,在以后进行系统升级、新增业务等方面可以借鉴 arXiv 的成功经验,根据用户需求或者专家调研意见来进行,达到事半功倍的效果。

#### 4.5 实现系统的互操作与融合

实现 arXiv 系统与机构知识库的内容共享是 arXiv 自 2015 年起特别关注的项目:开始研究是否可将学者存缴到机构知识库的论文副本自动推送至 arXiv;同时 arXiv 开展新的项目通过 SWORD 接口更好地整合计算机科学会议论文;通过使用诸如 ORCID、Grant-ID 或机构 ID 之类的 ID 来整合标准化的元数据,存缴和链接与论文相关的研究数据。可以说,实现系统之间、信息之间的互操作与深度融合,为用户提供更好地服务是 arXiv 的更高目标。

ChinaXiv 在互操作方面,通过 OAI-PMH 开放接口获取网站期刊和文章的 DC 元数据,同时通过 ORCID 对接机构知识库、iAuthor 中国科学家在线、arXiv.org 平台等,在互操作方面目前发展水平可与 arXiv 相媲美。ChinaXiv 收录的文献大多数以物理学、天文学、能源科学等自然科学较多,这类科学一般以国家或者机构项目资助形式进行,所以在以后的发展中,可以借鉴 arXiv 系统通过 Grant-ID 进行数据与元数据融合。

## 5 结语

本研究对 arXiv 可持续发展战略中的管理与经营模式、执行路径进行了详细分析,以期为我国 2016 年启用的“中国科学院科技论文预发布平台(ChinaXiv)”在管理运营方面提供借鉴。arXiv 可持续发展计划强调建立长期规划和有效的管理运营模式,协同成员管理模式分工明确,运营模式以比较创新和透明的方式进行,确保了合理的资金来源,严格质量审核标准、流

程与政策,有效的数字保存机制,根据实际全面开展系统发展、升级。且 arXiv 开始关注与其他机构知识库的互操作性和期刊系统上传论文的同期存缴。ChinaXiv 虽然处于萌芽阶段,但在质量审核和互操作方面的做法还是比较成功的,在资金来源方面、数字保存方面、系统业务和服务方面可以借鉴 arXiv 的成功做法。目前最主要的任务是直视国内权威学者仍然更倾向于将其成果通过传统核心期刊发布这一事实,加强宣传力度,通过不断呼吁权威专家投稿以加强存储质量和学术影响力,吸引更多学者加入到这个活动中来,确立系统的地位后再逐步扩大范围。

#### 参考文献:

- [1] ArXiv Sustainability Initiative [EB/OL]. [2016-12-05]. <https://confluence.cornell.edu/display/culpublic/arXiv+Sustainability+Initiative>.
- [2] ArXiv Member Institutions (2017) [EB/OL]. [2017-01-15]. <https://confluence.cornell.edu/pages/viewpage.action?pageId=340900096>.
- [3] ArXiv submission rate statistics [EB/OL]. [2017-01-15]. [https://arxiv.org/help/stats/2016\\_by\\_area/index](https://arxiv.org/help/stats/2016_by_area/index).
- [4] World Ranking Web of Repositories [EB/OL]. [2017-01-15]. <http://repositories.webometrics.info/en/world>.
- [5] 刘银华,王瑰玮. 基于 arXiv 平台的仓储论文实证分析[J]. 现代情报, 2013, 33(2): 155-158.
- [6] DAVIS P M, FROMERTH M J. Does the arXiv lead to higher citation and reduced publisher downloads for mathematics articles[J]. *Scientometrics*, 2007, 71(2): 203-215.
- [7] 刘银娣. 电子印本仓储——arXiv 运营情况研究[J]. 出版科学, 2009, 17(3): 78-82.
- [8] 王欣,董洪光. 国内物理学期刊 arXiv 自存档论文的引用优势研究——以 *Frontiers of Physics* 期刊为例[J]. 图书情报工作, 2011, 55(11): 144-147.
- [9] 中国科技论文预发布平台[EB/OL]. [2017-01-15]. <http://www.chinaxiv.org/user/mySubmit.htm>.
- [10] ArXiv @ 20: Collaboration Agenda for Research and Innovation [EB/OL]. [2017-01-15]. <https://confluence.cornell.edu/pages/viewpageattachments.action?pageId=340895275&metadataLink=true>.
- [11] ArXiv Governance Model [EB/OL]. [2017-01-15]. <https://confluence.cornell.edu/display/arxivpub/arXiv+Governance+Model>.
- [12] Next Generation arXiv [EB/OL]. [2017-03-18]. <https://confluence.cornell.edu/display/arxivpub/Next+Generation+arXiv>.
- [13] Alfred P. Sloan Foundation awards grant for arXiv upgrade [EB/OL]. [2017-03-18]. <https://www.library.cornell.edu/a->

- bout/news/archive/alfred-p-sloan-foundation-awards-grant-arxiv-upgrade.
- [14] ArXiv and Allen Institute for Artificial Intelligence (AI2) collaborate on semantic scholar pilot for computer science papers [EB/OL]. [2017 - 03 - 18]. <https://confluence.cornell.edu/display/arxivpub/2016/12/12/arXiv+and+Allen+Institute+for+Artificial+Intelligence+%28AI2%29+collaborate+on+Semantic+Scholar+pilot+for+Computer+Science+papers>.
- [15] New sponsorship model broadens arXiv membership [EB/OL]. [2017 - 03 - 18]. <https://www.library.cornell.edu/about/news/archive/new-sponsorship-model-broadens-arxiv-membership>.
- [16] Harish-Chandra Research Institute [EB/OL]. [2017 - 03 - 15]. <http://www.hri.res.in/>.
- [17] Physical Research Laboratory [EB/OL]. [2017 - 03 - 15]. <https://www.prl.res.in/prl-eng/#main-content>.
- [18] Tata Institute of Fundamental Research ,Mumbai [EB/OL]. [2017 - 03 - 15]. <http://www.tifr.res.in/>.
- [19] CERN [EB/OL]. [2017 - 03 - 15]. <http://home.cern/about>.
- [20] Helmholtz Association of German Research Centres [EB/OL]. [2017 - 03 - 15]. <https://www.helmholtz.de/en/>.
- [21] ArXiv@25 Review Strategy [EB/OL]. [2017 - 03 - 15]. <https://confluence.cornell.edu/display/arxivpub/arXiv@25+Review+Strategy>.
- [22] The Center for Direct Scientific Communication (CCSD) [EB/OL]. [2017 - 03 - 15]. <https://www.ccsd.cnrs.fr/en/>.
- [23] Big Ten Academic Alliance [EB/OL]. [2017 - 03 - 15]. <http://www.btaa.org/home>.
- [24] Greater Western Library Alliance [EB/OL]. [2017 - 03 - 15]. <http://www.gwla.org/>.
- [25] California Digital Library [EB/OL]. [2017 - 03 - 15]. <http://www.cdlib.org/>.
- [26] ArXiv mirror sites [EB/OL]. [2017 - 03 - 15]. <http://lanl.arxiv.org/help/mirrors>.
- [27] INSPIRE-HEP [EB/OL]. [2017 - 03 - 15]. <http://inspirehep.net/info/general/project/index>.
- [28] SAO/NASA ADS [EB/OL]. [2017 - 02 - 26]. <http://adsabs.harvard.edu/>.
- [29] ArXiv Search Interface [EB/OL]. [2017 - 02 - 27]. <http://arxivsi.las.ac.cn/home.htm>.

作者贡献说明:

白林林: 负责论文的数据获取、提纲与撰写;

祝忠明: 负责论文的修订。

Dissection on ArXiv Sustainability Initiative Governance and Business Model

Bai Linlin<sup>1 2</sup> Zhu Zhongming<sup>1</sup>

<sup>1</sup> Lanzhou Library , Chinese Academy of Sciences , Lanzhou 730000

<sup>2</sup> University of Chinese Academy of Sciences , Beijing 100049

**Abstract:** [Purpose/significance] Dissection on arXiv Sustainability Initiative governance and business model hopes that it can provide helpful experiences to ChinaXiv lauched in 2016 in the aspects of governance and business model , services and development and so on. [Method/process] This paper dissects the governance , business model and roadmaps adopted by arXiv's Sustainability Initiative in detail. [Result/conclusion] The arXiv Sustainability Initiative emphasizes funding from external sources , membership fees are based on an institutional ranking calculated according to the number of articles downloaded and financing method is transparent. Under the circumstances that Scientific Advisory Board and Member Advisory Board represent the interests of member institutions and investigation of the user needs , the suggestions of system development , upgrades and other aspects from the actual are well worth learning and drawing upon. And the idea that arXiv begins to pay attention to the interoperability with other institutional repositories is also worthy of our considering in the operation of ChinaXiv.

**Keywords:** arXiv Sustainability Initiative governance model business model roadmaps