中文 Altmetrics 数据整合分析平台的建立研究*

□陈铭　叶继元

摘要　替代计量学(Altmetrics)的发展极大地促进了在线科学交流和开放科学发展的进程，已普遍被研究机构和研究人员看作是一种基于社交网络数据进行科研交流、传播和评价的新型计量学。为了在 Web2.0 环境下，保证科学评价和计量研究的准确性和影响力，最理想化的方式是按照一致性的标准建设能融合各社交网络 Altmetrics 数据的整合分析平台。通过网络调研法和深入访谈法，借鉴了国外已有针对 Altmetrics 数据整合分析平台的有益经验和需要避免的问题，分析了在国内建立中文 Altmetrics 数据整合分析平台由图书馆界来牵头组织的必要性和可行性。提出由图书馆界来整合 Altmetrics 数据的要点，包括：拓展原始数据源、使用并大力普及 DOI、制定统一的数据标准和科学设计指标、构建 Altmetrics 数据采集加工的整体方案、广泛开展合作和保护用户的隐私数据。

关键词　替代计量学　数据整合分析平台　图书馆信息资源建设　开放数据
分类号　G250
DOI　10.16603/j.issn1002－1027.2022.04.013

1　Altmetrics 的发展背景及社交网络平台数据整合的必要性

1.1　Altmetrics 的发展背景

随着网络技术的快速发展和普及，人类社会的交流方式发生了巨大变化。特别是大量社交媒体平台出现后，由于其使用便捷且成本低廉，成为数以万计的普通公众进行信息、思想和知识等交流的主要方式，这其中也包括了科研工作者们对于学术知识和思想的传播交流。这不但促进了学术信息更广泛的交流，也实现了学术研究成果的多元化评价。Altmetrics 正是在这样的背景下悄然诞生，促成了计量学学科结合社交媒体在线交流特点进行了 Web2.0 的创新和革命，也成为图书情报领域的研究热点，产生了巨大的影响。科研机构、科研人员以及出版商都纷纷通过社交网络如推特(Twitter)、小木虫、脸书(Facebook)等分享交流科研成果，Altmetrics 已普遍被研究机构和研究人员看作是一种基于社交网络数据进行科研交流、传播和评价的新型计量学，旨在通过其建立的快速、全面和新型的科研评价体系补充仅仅依靠传统引文指标或同行评议的科研评价体系。

1.2　社交网络平台及其数据呈现的问题

源于科学在线交流环境中诞生的 Altmetrics 新型计量学的运行基础是各类型学术成果(如论文、图书、数据集、程序、视频等)在各种社交网络平台中发生交互而不断产生并逐步积累的网络数据，即学术成果的 Altmetrics 指标测量的是其在主流新闻媒体、社交媒体或在线社区被下载、提及、分享、点赞和评论次数等的关注度和影响力。所以通过 Altmetrics 获得可以进行指标分析的数据受到以下两方面的影响：一是研究成果所出现的各种社交网络平台，这是 Altmetrics 原始数据源的产生地，具体来说可细分为以下六类：①社交媒体网站，国外有谷歌加(Google＋)、Facebook、Twitter 等，国内有微信、微博、知乎等；②在线学术网站，国外有 Scopus、Web of Science、EBSCO ，国内有中国知网、万方数据库等期刊综合网站等；③新闻媒体网站，国外有科学新闻(Science News)、时代新闻(Time News)，国内有

* 国家社会科学基金重大课题“新时代我国文献信息资源保障体系重构研究”(编号：19ZDA346)的研究成果之一。
通讯作者：陈铭，ORCID：0000-0001-5061-6，邮箱：chenming@nju.edu.cn。

科学网等；④文献管理平台，国外有 CiteULike、Mendeley 等，国内有道客巴巴、百度文库、豆丁网等；⑤学术社区网站，国外有 F1000 等，国内有丁香园、小木虫等；⑥百科平台，国外有维基百科（Wikipedia）等，国内有 MBA 智库百科等。二是统计各种社交网络平台上的研究成果及其“活跃”情况，如在小木虫上对某一研究成果的评论数或在 Mendeley 上对某一出版物的保存量[1]。

目前对于 Altmetrics 的一种主要研究方法是通过收集社交网络平台上的数据进行实证评价分析得出相应结论。因此，实时快速收集数据并保证这些数据的准确性、一致性、全面性和有效性就显得非常重要。但是由于不同的社交网络平台数量众多，各平台的使用率和普及率差别很大，在这些社交网络平台上进行交流传播产生的大量网络数据呈现出庞大、多态、异构、不稳定和繁杂的特征，且各平台的数据和指标都存在一定的差异，导致数据去重和整合的难度很大。

1.3 整合 Altmetrics 数据的必要性

鉴于以上情况，很多科研人员虽然已广泛使用各社交网络上的数据来进行基于 Altmetrics 的评价研究，但是在做研究时只能选取不同的有代表性的社交网络平台来采集、处理和汇总平台数据，难以形成统一的标准，这必将严重影响评价结果的准确性和全面性，也会限制 Altmetrics 的长期可持续发展。因此为了在 Web2.0 环境下，保证科学评价和计量研究的准确性和影响力，最理想化的方式是构建有一致性和通用意义的 Altmetrics 社交网络标准化数据框架和评价体系，并按照一致性的标准建设能融合各社交网络平台上的 Altmetrics 数据的整合分析平台。目前尚未有一个能把各种零散的社交网络数据整合起来的中文平台出现，这在一定程度上会影响这种创新评价方式的发展和开放数据的有效使用。因此建立中文 Altmetrics 数据整合分析平台是必须且紧迫的。

2 国外 Altmetrics 数据整合分析平台的经验与问题

2.1 国外 Altmetrics 数据整合分析平台的经验

目前国外已有针对 Altmetrics 数据的整合分析平台，其本身不产生互动数据，而是汇聚并整合了多个不同社交网络平台的原始数据源。国外 Altmetrics 数据整合分析平台主要包括 Altmetric.com、PLoS ALM、PlumX、Kudos、ImpactStory 和 Webometrics Analyst 等，是由不同的出版商或服务商在不同时间开发的。通过对上述平台网站的调研，总结了一些可以借鉴的经验。

（1）数据来源和成果类型较丰富。国外 Altmetrics 整合分析平台的数据来源广泛，大部分来自于社交媒体网站、在线学术网站、文献管理平台、学术社区网站、新闻媒体网站和百科平台等，Altmetric.com 和 PLoS ALM 覆盖的数据源最为全面，囊括了上述六种社交网络平台二十种左右的原始数据源。数据来源越广泛多样，评价数据就越能准确科学地反映被评价对象的影响力。评价的成果类型也是多种多样的，从学术论文到博客、数据集、软件、程序代码以及图片等都囊括其中，其中 PlumX 评价的成果类型最多，高达 27 种。

（2）通过 DOI 等标识符来识别学术成果。数字对象唯一标识符（Digital Object Identifier，DOI）是国外最常用的一种标识符技术，Altmetrics 数据的准确性主要取决于文献的 DOI，DOI 是否可用在很大程度上决定了 Altmetrics 数据的质量表现[2]。学术文献还有一些其他的标识符，如 PMID，ArXiv ID 和 SlideShare 的统一资源定位符（Uniform Resource Locator，URL）等[3]。国外 Altmetrics 数据整合分析平台通过学术成果的 DOI 等统一标识符来实时追踪各社交网络平台上的 Altmetrics 数据，保证了在网络环境下对学术文献对象的准确识别，有效地避免了重复。

（3）建立了系统的指标体系。国外 Altmetrics 数据整合分析平台都建立了系统的指标体系和一致的评分系统，然后以报告或者评分的形式对某一项科研成果的社会影响力进行评价。因此可以提供相对系统和标准化的可用数据，为科研人员的研究和科研机构的评价提供数据维度的方便和实时的支持，科研人员可以在其平台工具上一站式查询 Altmetrics 数据，省去了在多个社交平台寻找数据的麻烦。

由于易用性和开放性的特征，国外 Altmetrics 数据整合分析平台受到了学界的欢迎，为精准的科学评价和合理的科技政策的制定提供了全面参考，对 Altmetrics 的发展也起到了重要作用。

2.2 国外 Altmetrics 数据整合分析平台存在的问题

从理论上来说，不同 Altmetrics 数据整合分析

平台应该提供一致的 Altmetrics 数据，但是由于 Altmetrics 数据本身的多源性和复杂性，现在国外也还没有一家 Altmetrics 数据整合分析平台能够涵盖所有社交网络平台的数据，并且各数据整合分析平台的数据源存在着以下不一致的问题。

(1)数据来源不一致。目前各个国外 Altmetrics 整合分析平台的数据来源不一样，所收集数据的策略不一样，其在发展过程中根据自身目标和愿景形成了独具特色的数据源[4]，比如 Altmetric.com 对博客文章的收集最多，而 PlumX 更多采集来自新闻媒体的数据[5]。这些平台还制定了各不相同的数据提取清洗政策，这对数据的使用产生了很大影响。

(2)指标聚合方式不一致。各数据整合分析平台所提供的 Altmetrics 指标有些直接来自某个社交网络平台应用程序编程接口(Application Programming Interface，API)提供的一个字段，有些是多个字段的组合形成的一个新的指标[6]，以满足评价的需求。这些指标只有名称，并没有说明是如何构建的。所以到目前为止这些数据整合分析平台都还不能够提供准确、全面和一致的 Altmetrics 指标。这些都会影响评价的准确性。

(3)数据更新速度不一致。不同的数据整合分析平台对不同来源数据的更新频率不一致。比如 Altmetric.com 平台中声明对 Twitter、Scopus、Wikipedia 的数据是实时更新的，而 Facebook、YouTube、Mendeley 等是每天更新。实时更新的具体含义以及如何实现、每日更新的方式和具体时间等都是不透明的，这也是导致 Altmetrics 数据整合分析平台数据质量的问题之一。

2.3 对国外 Altmetrics 数据整合分析平台的借鉴

国外常用的 Altmetrics 数据整合分析平台都以寻找更多样和全面的社交网络数据源为主要目标，比如 Altmetric.com 整合分析平台还收录了新浪微博数据，未来各平台将会收集到更全面的社交网络平台数据。而且不同的数据整合分析平台之间是具有互补关系的，不同平台之间的数据如果可以相互融合，那么所能提供的数据就会更加准确和全面，可以认为这是 Altmetrics 数据整合分析平台的发展趋势，也是建立中文 Altmetrics 数据整合分析平台的目标。

在中文环境下，首先要考察学术成果受到哪些中文社交网络平台的关注；其次要尽可能寻找多样化和全面性的学术成果网络社交数据源，借鉴国外 Altmetrics 数据整合分析平台的经验；再次能提供的 Altmetrics 指标需要有详细统一的标准，要能避免如上所述国外各 Altmetrics 数据整合分析平台现存的问题。最后这个平台的数据不能单靠高成本和低效率的人工方式收集，需要设计专业和智能的数据收集工具，快速准确地通过 API 接口从相关平台获取网络数据，并对其进行清洗、格式转换和特征提取等工作。而这些工作如果由图书馆这样的社会公共服务机构进行领导和组织，将能取得比较好的效果。笔者深入访谈了 10 位图书馆学界和业界的专家，对图书馆界是否适合承担建立中文 Altmetrics 数据整合分析平台领导组织的角色、图书馆界应在数据整合分析平台中提供什么样的服务向专家进行了详细的咨询访谈(访谈提纲见附录)。根据对专家意见的整理和综合，笔者认为在建立中文 Altmetrics 数据整合分析平台时，由图书馆界来组织实施是非常必要且可行的。

3 图书馆界整合中文 Altmetrics 数据的必要性和可行性

3.1 图书馆界整合中文 Altmetrics 数据的必要性

3.1.1 图书馆作为社会公共服务机构的必然使命

数据作为获取知识和开展知识服务的重要性和价值已经引起全球的重视，它被认为是世界上最有价值的资源并且改变了竞争的本质[7]。“开放数据”在维基百科中的定义是：“不受任何知识产权和管理机制的限制，是经过挑选与许可的数据，可以免费开放给公众，任何人都可以自由使用[8]。”开放数据包括开放的馆藏数据、科研数据、政府数据、商业数据和一些用户产生的数据等。公众对开放数据需求的不断增加促进了开放数据运动的长足发展，随着越来越多数据的开放，开放数据的种类不断增长，开放数据的良好环境逐步形成。

Altmetrics 社交网络数据属于用户产生的数据，但由于均被托管给了第三方，所以在授权的情况下也可以认为转变为一种商业数据，因此它也属于开放数据。根据开放的理念和思维以及开放数据的内涵，Altmetrics 社交网络数据应该免费开放给公众让其自由使用。但是大部分社交网络平台是营利性质的，数据还未能无条件地提供给大众使用。把受限制的数据无条件地向任何人开放，也是用户拥

有“数据开放权”的最高目标。虽然在开放数据环境下如何对开放数据进行知识产权保护尚没有明确的法律条文，但在对 Altmetrics 数据进行整合管理的过程中，图书馆界也可以借鉴国外如德国国家图书馆、英国大英图书馆、学术出版与学术资源联盟等制定的对于开放数据在馆内应用及许可协议，明确使用者与数据提供体系之间的关系，促成用户与原始数据之间的对接与再利用[9]。

国外 Altmetrics 的数据整合分析平台开发商大多是出版商或服务商，比如 Altmetric.com、PlumX、Kudos、ImpactStory 和 Webometrics Analyst 都是由服务商提供的，PLoS ALM 是由出版商提供的[10]。由于具有商业性质，所以其提供的数据服务大多不是免费的，比如 PlumX 和 ImpactStory 都是收费的，Altmetric.com 也是部分收费的，这将极大制约 Altmetrics 的推广和应用。因此国内整合中文 Altmetrics 数据来开发建立整合分析平台不适合由商业盈利机构来主导。并且图书馆作为社会公共服务机构有义务有责任承担开放数据管理和领导组织的角色，负责对社交网络开放数据进行遴选、采集、描述、组织、分析与评价，提高数据资源的可用性和价值性，给用户免费提供社交网络数据的保存、检索、分析挖掘等服务，并根据数据连续使用的视角提供数据关联、标识和发布等服务[11]。此外，图书馆相比于商业机构也更有利于对这些开放数据进行长期保存。

3.1.2 开放数据环境下的责任推动

数据已成为体现图书馆服务水平和核心竞争力的重要因素，是涉及图书馆服务模式创新、提高个性化服务能力和增强服务透明度的战略资源。数据也是图书馆发现用户需求、进行服务决策和评估服务有效性的直接依据，是图书馆用户服务“数据权”和“知情权”的有效载体[12]。图书馆很早就开始进行科学数据的管理和政府开放数据的整合管理，如美国卡内基图书馆整合了农业、教育、建筑、卫生等多类别的政府开放数据于平台上供人们随时下载分享[13]。国外的知名高校如哈佛大学、斯坦福大学、剑桥大学等都有针对本校科研数据整合管理的平台[14]。我国虽然起步较晚，但近几年也越来越重视对开放数据的整合管理，国内的“双一流”大学如北京大学、武汉大学、复旦大学等也都建立了科学数据的共享平台[15]。2014 年国内 9 家高校图书馆还在复旦大学的牵头下共同发起成立了“中国高校图书馆研究数据管理推进工作组”，并建立了能够实现科研数据存储、发布、交换、共享与在线分析等功能的复旦社会科学数据平台[16]。Altmetrics 数据属于开放数据，是一种公共资源，这种类型数据的开放和加工整合能为图书馆服务质量的提高提供强有力的数据资源支持。因此，图书馆的 Altmetrics 社交网络数据的整合开放是否可用、安全和具有公信力，将会对图书馆服务模式的有效性以及用户权益的保障产生很大影响。图书馆界在这样的新契机下要充分发挥自身价值和功能，以用户为中心，以开放的姿态利用复杂多样的社交网络数据满足用户多元化的需求，跟上时代发展的潮流，使图书馆与社会的关系更加紧密，从而提高图书馆的地位，实现图书馆的积极转型，推动图书馆事业的跨越式发展。

3.1.3 图书馆实现创新信息资源建设的途径

Altmetrics 的发展为图书馆实现创新的信息资源建设、开展数据相关的服务提供了新的机遇。图书馆的信息资源建设对象除了有纸质资源和数据库资源外，数据也已成为图书馆信息资源建设中的重要组成部分。社交网络数据是重要的数据资源，对于科研人员开展评价活动具有重要作用。国家图书馆已收藏了新浪微博上的 2000 亿条博文，美国国会图书馆也已收录了千亿条 Twitter 上的推文[17]。图书馆必须把社交网络数据纳入到资源建设的范畴，图书馆要改变传统的资源建设思想和建设模式，整合各社交平台上大量的 Altmetrics 数据，提高社交数据资源的利用率，进而提高图书馆的服务水平和服务效率。这有利于推动整个社会开放共享的形成，也让公民更了解图书馆的价值所在。

3.2 图书馆界整合中文 Altmetrics 数据的可行性

3.2.1 图书馆具有丰富的信息和数据服务经验

随着社会的快速发展，数据时代的到来带给图书馆巨大的变革，图书馆不再只是存储纸质文献的场所，图书馆本身的纸质馆藏资源就很丰富，并在经过多年的数字图书馆建设后，已拥有大量各种类型的数字化资源，是最适合承担大数据时代数据开放与整合的实践者。而且图书馆作为信息和知识的保存和传播机构，具有面向公众开展信息资源服务的丰富经验。这些经验完全可以移植到数据服务方面。这也是图书馆界相比于商业机构更适合整合中文 Altmetrics 数据的优势之一。图书馆界组织协调

开发的、可以供用户免费使用的项目将会有更多的利用率。各图书馆也不需要额外斥巨资向商业机构购买这样的整合平台。国外图书馆在数据服务方面已开展了较多实践[18]，其以开放数据为原则推动图书馆把现有的数字资源转变为开放数据资源，并将社会各方丰富的公开数据资源纳入馆藏资源体系，如政府数据、气象数据、人文数据、科学数据和商业数据等，加强开放数据资源之间的融合与关联，打破时空对于公众获取数字资源和服务的限制，为用户提供特色数据服务，满足公众对于数据资源的个性化需求。

3.2.2 科技发展使图书馆具有数据加工分析的能力

图书馆从发展初期开始就持续追踪读者使用馆藏资源的情况，从基础简单的读者调查到书籍期刊资源的使用追踪，再到用计算机来进行图书馆借阅统计分析，直至到出现电子资源后进行电子资源使用情况的复杂分析，因此图书馆具有分析整合用户数据的经验和能力。随着科技的发展进步，图书馆的信息化设备已经十分完备，具有先进的计算机设备和高速的网络设施，以及成熟的数据采集加工能力，与图书馆相关的用户使用资源的数据分析模型也越来越多样化，并且图书馆采集、存储和整合数据的成本也逐渐降低。因此由图书馆来进行整合各中文社交网络平台上的 Altmetrics 数据具有充分的可行性。

3.3 COUNTER 和 DRAA 的启示及图书馆界的角色和定位

3.3.1 COUNTER 和 DRAA 成功运作的启示

网络电子资源在线利用统计(Counting Online Usage of Net-worked Electronic Resources，COUNTER)是规范电子资源使用统计报告数据处理、审核和提交的国际化标准，其目的是为在线信息服务商和用户提供可靠的、一致的、兼容的使用统计标准和方案。2019 年，COUNTER 发布了第五版《COUNTER 电子资源使用统计实施规范》(以下简称 COUNTER R5)。COUNTER R5 采用了新的报告体系，新的元素和属性，以及新的报告格式和获取方式[19]。

高校图书馆数字资源采购联盟(Digital Resource Acquisition Alliance of Chinese Academic Libraries，DRAA)是由中国部分高等学校图书馆共同发起成立的，目的是合作开展引进数字资源的采购工作，规范集团采购行为，通过联盟的努力为成员馆引进数字学术资源谋求最优价格和最佳服务。DRAA 于 2013 年初开始支持通过标准化的电子资源使用统计获取协议(The Standardized Usage Statistics Harvesting Initiative，SUSHI)自动获取 COUNTER 格式报告。2015 年 9 月正式成立 DRAA 使用统计工作组，目的一是促进数据库商配合收集数据库的使用数据；二是建立收集使用数据的长效机制，并对数据质量进行检查；三是深入研究 COUNTER 规范，实现对使用数据统计的进一步应用。DRAA 使用统计模块分为数据获取、数据处理和报告展示三个层次来进行使用数据统计的管理[20]。

COUNTER 项目对于图书馆开展整合 Altmetrics 数据具有很大的启发。COUNTER 最重要的组成部分是图书馆联盟，它还整合了电子资源从生产到利用环节的出版商、资源提供商和行业组织等。新的 COUNTER R5 报告体系除了依旧保持了详细的统计术语和严格的报告撰写标准外，还拓展了数据类型，比如数据集、音频、视频和图片等，并更精准和更新定义了计量类型、访问类型和访问方式等。DRAA 通过引入了全新的 SUSHI 协议，直接自动收割 COUNTER 的电子资源使用统计报告，可以使图书馆高效地获取更准确的电子资源使用数据，而不需要登录多个数据库商的网站下载 COUNTER 使用数据，加强了图书馆和数据库商对于资源使用评估的互动，为各图书馆制定科学有效的采购策略提供了方便[21]。因此由图书馆联盟来进行多方参与和协调发展电子资源使用数据的统计，制定统计的格式、内容和术语，并控制数据质量方面是有成功典范的，而且自动化收割和整合管理使用数据还极大地方便了图书馆自身使用这些数据来更好地开展服务。这也更证明了数据时代由图书馆界整合中文 Altmetrics 数据的可行性。

3.3.2 图书馆界在中文 Altmetrics 数据整合中的角色及可提供的服务

图书馆界在中文 Altmetrics 数据整合中应借鉴国际上 COUNTER 项目和我国 DRAA 统计工作组项目的成功实施经验，充分发挥其在数据整合过程中协调组织的作用，确立其协调者、组织者、建设者和服务提供者的角色。Altmetrics 数据整合项目可以由中国图书馆学会(以下简称中图学会)或教育部高等学校图书情报工作指导委员会(以下简称高校

图工委)来牵头组织，联合主要的省级公共图书馆、高校图书馆或图书馆联盟，以及部分社交网络平台的企业代表等共同开展此项工作。经费可以由各参与的图书馆缴纳的会费并结合中图学会或高校图工委下拨的部分研究经费组成，并争取以项目方式获得一些基金的资助。对于每一家参加数据整合项目的图书馆要明确分工职责，分别负责统计标准的制定、Altmetrics 数据的采集、Altmetrics 数据的加工处理、Altmetrics 数据的保存和发布、与社交网络企业沟通等。

最终应能建立起 Altmetrics 数据整合分析平台，平台需要让用户自由查询所需要文献的 Altmetrics 全面数据，还要能定期提供多样化的统计报告，包括按时间、内容、类型划分等。对于参与 Altmetrics 数据整合项目的图书馆，可以优先使用 Altmetrics 数据，该图书馆可以在第一时间把 Altmetrics 数据整合分析平台推送到该馆的主页，让用户根据需要进行浏览和获取。对于未参与的图书馆，由该馆与项目负责人进行联系，取得 Altmetrics 数据整合分析平台的使用授权后，为该馆用户提供服务。Altmetrics 数据整合分析平台工作流程和服务要点的架构见图 1。

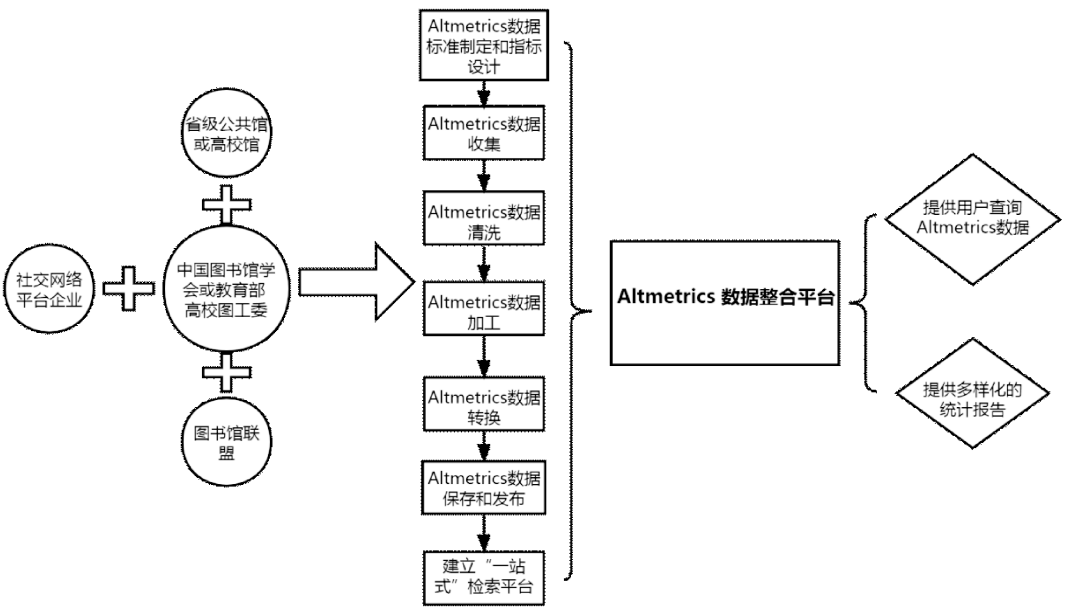

图 1　Altmetrics 数据整合分析平台工作流程和服务要点

4　图书馆界构建 Altmetrics 数据整合分析平台的要点

4.1　拓展原始数据源

Altmetrics 强调对多类型学术成果的认可和评价，不仅需要格式多样，还需要内容多样。所以在格式上学术成果不仅是传统论文的文本格式，还可以是图片、视频等。内容上除了学术论文外，程序片段、数据算法和科学数据集等新型的学术成果也是重要的需要被关注的类型[3]。比如关于科学数据集，可以利用数据引证的方式来评估科学数据集的价值，但由于过于局限于规范的学术论文，无法捕捉所有科学数据集被广泛应用的情况和多元化的价值[22]。因此用 Altmetrics 的指标来测量科学数据集在社交网络平台的提及、下载等情况，可反映其被分享和应用的情况[23]。但由于不同学科的科研人员对于社交网络平台的使用偏好具有较大差别，当前 Altmetrics 的指标涉及的研究成果以适用于科学、技术、工程与数学(Science，Technology，Engineering，Mathematics，STEM)领域的学术论文偏多，而人文艺术社会科学领域的很多研究成果(如唱曲、画作、雕刻)很难在社交网络平台上通过定量数据来衡量其影响力[24]。

研究者往往希望 Altmetrics 数据整合分析平台能够全面评价成果的多种影响力，这就需要通过不同类型的多样化的数据源来支撑，才能实现精准和全面的统计和评价。因此图书馆界在进行中文 Alt-

metrics 数据整合时要尽量拓展原始数据源，寻找和收集格式多样化（文本、图片及视频等）和内容多样化（学术论文、程序片段、数据算法、科学数据集等）的学术成果。

4.2 使用并大力普及 DOI

国外的 Altmetrics 数据整合分析平台在统计学术成果的 Altmetrics 数据时，一般是基于学术成果的唯一标识符来获取论及这篇学术成果的数据。发达国家数字对象唯一标识符（Digital Object Unique Identifier，DOI）的普及率非常高，而大多数发展中国家还没有普及 DOI。DOI 系统是由国际 DOI 基金会（International DOI Foundation，IDF）进行全球分布式管理，2007 年 3 月，IDF 正式授权中国科技信息研究所和万方数据成立中文 DOI 注册机构[25]。我国中文社交网络平台的中文学术成果很少有标注来源出处的唯一标识，甚至有一些连 URL 都没有，所以无法通过 DOI 来追踪中文学术成果的社交网络痕迹。因此图书馆界在整合 Altmetrics 数据时需要将不同标识符的相同目标文献采用一致的方法进行聚合，从而确定指标的一致性，保证数据的质量。图书馆界可以呼吁和敦促相关出版机构和期刊编辑机构尽快加入 DOI 系统，并提供相关技术支持和服务，促进 DOI 的普及使用。

4.3 制定统一的数据标准和科学设计指标

Altmetrics 的评价和研究离不开高质量的 Altmetrics 数据。Altmetrics 数据质量问题主要发生在社交网络平台、数据整合分析平台和用户三个层面，如前所述，社交网络平台数据的差异、数据的不稳定性、数据的不一致性和数据的覆盖率不同会影响 Altmetrics 数据的质量，数据整合分析平台中数据的来源不一致、聚合方式不一致以及数据更新速度不一致也会影响 Altmetrics 数据的质量。所以 Altmetrics 数据的准确性、一致性、动态性和持续性成为图书馆界在进行数据整合分析时最需要关注的问题。

图书馆界作为 Altmetrics 数据整合分析的主导者，应当重视数据质量的问题。首先通过与整合分析平台的开发者进行深入沟通和协调，制定统一的 Altmetrics 的数据标准，合理遴选各社交网络平台的 Altmetrics 数据，保证整合分析平台采集并记录到的数据与来源社交网络平台真实的原始数据相一致，保证各来源社交网络平台的数据相统一，保证各数据更新的频率相一致。其次科学地设计整合分析平台中的数据指标。由于各网络社交平台功能相近，许多 Altmetrics 指标界线模糊，在评价时会存在含义重合或者相似的情况，因此有必要区分不同类型的指标，把同类型的指标进行整合。比如出版商 PLoS 以及服务商 ImpactStory 将 Altmetrics 指标分为访问、引用、讨论、推荐和保存五类，PlumX 将指标分为使用、获取、提及、社交媒体以及引用五类。国内有学者将 Altmetrics 计量指标分为传播、获取、利用三个层次[26]。因此在聚合过程中图书馆界需要选取更具代表性、覆盖范围更大以及使用频率较高的指标作为评价指标。最后还需要注意一些保证 Altmetrics 数据质量的关键问题，比如数据整合分析平台的性能问题，这样才能更好地促进 Altmetrics 平台的开发与应用，提升图书馆界对于 Altmetrics 数据管理的有效性。

4.4 构建 Altmetrics 数据采集加工的整体方案

图书馆界对于 Altmetrics 数据的使用需要重视数据从采集、处理到转换等各个阶段的数据质量，并且在每个阶段能够采取合适的方法和策略避免容易产生问题的因素。

在数据收集阶段，图书馆界要根据数据规划要求，多渠道筹措资金提供数据整合分析平台的建设资金，设计开放数据服务机制，完成数据收集前的准备工作。然后可以通过元数据收割协议，对各社交网络平台进行元数据收割。在对 Altmetrics 数据进行收集时要注意实现登录接口、入口的设置。

在数据处理阶段，首先要进行数据清洗。在此过程中图书馆员将社交网络平台上的初始数据通过 API 进行提取，然后根据规划的需求，利用合适的数据清洗工具以可靠性、真实性和唯一性为原则核实数据的来源，剔除不一致、重复、不准确的数据，修正不精准的数据，保证 Altmetrics 数据质量。其次进行数据加工。图书馆界在数据加工前要创建元数据框架，基于已有元数据，确定字段结构，统一标准，制定元数据规范和关联数据应用；实现元数据的录入、排序、补充和存储，便于数据资源的关联、分析和应用等。图书馆界应创新开放数据格式，使其朝着资源描述框架（Resource Description Framework，RDF）格式转变，重视 API 标准化；还应实现开放数据管理的通用设计，方便各专业背景的用户使用，让开放数据转变为更方便使用的简单数据。

在数据转换阶段，Altmetrics 数据通过清洗、加

工之后成为结构化的干净数据，但是还要通过数据转换才能利用 API 接口对外开放。图书馆员可以和技术人员协作，利用 RDF 格式转变工具把数据转换存储到 RDF 存储库中。然后再利用统一或者分类的开放方式对外开放[27]。

在实现数据从采集、处理到转换等阶段后，图书馆界应基于在文献资源管理方面的经验，进行数据平台的数据存储、分类、组织、检索、管理等，使用户可利用图书馆的“一站式”检索平台一键快速搜索到自己所需的数据资源，并完成基于数据共享框架下的智能判断和决策。图书馆界应对开放全程实时监管，当发现错误或不精确的数据时，应及时对其进行修改。

4.5 广泛开展合作

图书馆界对于 Altmetrics 数据的开放、整合、管理应广泛听取用户和社会的意见，了解其需求，坚持公开、透明、可扩展和合作的原则，合理选择数据开放的对象、内容和方式，不断增强 Altmetrics 数据整合分析平台的可用性和价值密度，最终实现数据的开放性增值。各图书馆之间应加强合作，对 Altmetrics 数据进行采集整合不是某一家图书馆的事，而是整个图书馆界的责任，需要各图书馆之间协作来完成。图书馆界需协调各方开展对数据标准、采集整合方法工具和策略的研究。

图书馆界还需加强与社交网络平台企业的联系和合作，作为 Altmetrics 数据的来源，要想获得准确可靠的数据不能仅靠 API 自动收集，特别是一些企业没有开放给大众的数据，可以由图书馆界与这些企业进行沟通协商，从利益相关者的角度出发，提出解决方案以便获得这些数据。总之要加强图书馆界与社会各部门的联系，促使图书馆作为公共文化服务机构能更好地完成 Altmetrics 数据整合分析的工作，开展创新性的由数据驱动的公共服务，从而也可以扩大图书馆的社会影响力。

4.6 保护用户的隐私数据

Altmetrics 数据中包含着社交网络平台用户大量的阅读内容、参与内容、社会关系和地理位置等个体特征和行为数据。虽然社交网络数据权属于数据收集的企业一方，但是这些包含大量个人数据的隐私也是需要被保护的。在采集利用 Altmetrics 数据过程中如果图书馆界对用户 Altmetrics 数据资源进行无限制地完全开放，一些用户不想公开的隐私和个人信息就会被泄露[28]。图书馆界应加强 Altmetrics 数据开放过程中用户的隐私保护，确保数据提供服务具有较高的安全性并能保障用户的名誉权。首先，图书馆界应保证用户拥有对自身社交行为数据采集、使用和共享的知情权与决定权，让用户依据保护的需求决定开放的内容、程度和方式。其次，图书馆界在 Altmetrics 数据开放过程中，应采用对用户隐私信息匿名，或转为采集用户群体特征的数据，力争在能保持数据价值、可用性和开放性的前提下保护用户的隐私安全。最后，图书馆界还应根据发展变化中的用户隐私保护需求，不断更新完善相关的行业规范和政策法规，确保用户隐私保护可及时被评估和界定[29]。因为图书馆界对于用户隐私数据的判定标准也是关乎用户隐私保护有效性和可控性的重要因素。

5 结语

Altmetrics 的长期可持续发展能够保证科学评价和计量研究的准确性和影响力，能够给科研人员和公众提供更好的开放数据服务，因此建立中文 Altmetrics 数据整合分析平台势在必行。作为社会公共服务机构的图书馆界最适合承担牵头组织的角色，前有 COUNTER 项目和 DRAA 成功运用的经验借鉴，后有图书馆界丰富的信息和数据服务的经验，图书馆界可以充分发挥组织协调作用，把握整合分析的要点，避免国外 Altmetrics 数据整合分析平台存在的问题，协调各方力量做好 Altmetrics 数据整合分析平台。

参考文献

1 Alexand R A J，Hoffmann C P，Kunne S，et al. Altmetrics for large，multidisciplinary research groups：comparison of current tools[J].Bibliometrie-Praxis und Forschung，2014，3 (1)：1－19.
2 Haustein S. Grand challenges in Altmetrics：heterogeneity，data quality and dependencies[J]. Scientometrics，2016，108(1)：413－423.
3 吴胜男，赵蓉英.Altmetrics 应用工具的发展现状及趋势之分析[J].图书情报知识，2016(1)：84－93.
4 Ortega J. Altmetrics data providers：a meta-analysis review of the coverage of metrics and publication[J]. Profesional De La Informacion，2020，29(1)：1－23.
5 余厚强，尹梓涵.不同替代计量数据库数据政策与数据数值的比较研究[J].情报杂志，2021，40(5)：111－117.
6 Ding J D，Guo J，Liu X J，et al. Research on the relationship between citation and Altmetrics of Open Access papers from different geographical regions[C]. Proceedings of the International

Conference on Scientometrics and Informetrics. Italy：Sapienza Univ Rome，2019.
7 新浪财经.《经济学人》：数据经济时代要求革新互联网巨头监管方式[EB/OL].[2021－06－06].http://news.jstv.com/a/20170515/14948336099.shtml.
8 Wikipedia.Open data[EB/OL].[2021－03－09].https//en.wikipedia.org/wiki/Open_data.
9 杨敏，夏翠娟，徐华博.开放数据许可协议及其在图书馆领域的应用[J].图书馆论坛，2016，36(6)：91－98，141.
10 金贞燕，侯景丽，孙华丽.Altmetrics 数据整合分析工具的现状特点及相关问题研究[J].情报理论与实践，2019，42(4)：89－95，70.
11 张连分.大学图书馆开展数据管理服务的实践和成效评析[J].图书馆建设，2018(10)：45－51，58.
12 容春琳.公共图书馆应用大数据的策略研究[J].图书馆建设，2013(7)：91－95.
13 赵宁，黄铁娜，曹洋.图书馆融合政府开放数据服务模式探索[J].新世纪图书馆，2020(12)：62－65.
14 李正超.国外高校图书馆科学数据服务调研及启示[J].图书馆学研究，2018(19)：79－84.
15 张群，张以淳，彭奇志.嵌入“双一流”建设的高校图书馆科学数据服务研究[J].图书馆工作与研究，2018(11)：15－19，31.
16 刘敏.“双一流”高校图书馆科学数据服务现状及优化策略[J].图书馆工作与研究，2020(11)：15－24.
17 快科技.价值连城：2000 亿条微博被国家图书馆保存[EB/OL].[2021－08－10].https://baijiahao.baidu.com/s?id＝1631287155333511535&wfr＝spider&for＝pc.
18 黄如花，王春迎，周力虹.国外公共图书馆开放数据服务实践分析及对我国的启示[J].图书情报工作，2018，62(13)：139－144.
19 侯景丽.COUNTER R5 的新特性及对图书馆的影响[J].图书馆杂志，2018，37(12)：46－55.
20 杨巍，叶仁杰，吴元业，等.COUNTER Release 5 的新特征及其应用研究[J].大学图书馆学报，2020，38(1)：18－25，41.
21 陈大庆，叶兰，丁培.电子资源使用统计收割标准 SUSHI 的实施与应用研究[J].中国图书馆学报，2018，44(2)：46－60.
22 李龙飞，余厚强，尹梓涵，等.替代计量学视角下科学数据集价值的定量测度研究[J].情报理论与实践，2020，43(9)：47－52，71.
23 Bornmann L. Do Altmetrics point to the broader impact of research? an overview of benefits and disadvantages of Altmetrics[J].Journal of Informetrics，2014，8(4)：895－903.
24 刘丽敏，王晴.国外 Altmetrics 理论研究与实践进展[J].情报理论与实践，2017，40(3)：132－137.
25 徐行，王爽.图书情报领域如何推动 DOI 在中国的发展[J].图书馆学研究，2011(11)：60－62.
26 余厚强，邱均平.替代计量指标分层与聚合的理论研究[J].图书馆杂志，2014(10)：13－19.
27 凌霄娥.数字人文下的图书馆开放数据服务机制分析[J].图书馆，2021(5)：68－73.
28 马晓亭，尚庆生.大数据时代图书馆开放数据服务平台与开放数据服务模式研究[J].图书馆理论与实践，2015(5)：72－75，102.
29 李佳佳.信息管理的新视角：开放数据[J].情报理论与实践，2010，33(7)：35－39.

作者单位：南京大学信息管理学院，江苏南京，210023
收稿日期：2021 年 9 月 22 日
修回日期：2022 年 3 月 13 日
（责任编辑：关志英）

附录：专家访谈提纲

尊敬的_____专家：

您好！我们的国家社科基金重大课题（“新时代我国文献信息资源保障体系重构研究”项目编号：19ZDA346）正在研究关于构建中文 Altmetrics 数据整合分析平台的内容，您是图书馆学界的资深教授，特想向您请教一些问题，了解您对于构建平台的想法。以下是希望能当面访问您时的问题提纲：

（1）您觉得国外建立的 Altmetrics 数据整合分析平台（如 Altmetric.com、PLoS ALM、PlumX、Kudos）有哪些优势和劣势？

（2）您觉得是否需要建立中文 Altmetrics 数据整合分析平台？如果需要的话，您觉得可以由谁来组织建立？

（3）您觉得由图书馆界来建立中文 Altmetrics 数据整合分析平台是否必要和可行？

（4）如果由图书馆界来进行牵头组织建立中文 Altmetrics 数据整合分析平台，图书馆界在其中应该起到什么样的作用，可以提供什么样的服务？

期待与您的面谈，感谢您的支持！

国家社科基金课题组成员

Research on Establishing Chinese Altmetrics Data Integration and Analysis Platform

Chen Ming　Ye Jiyuan

Abstract：The development of Altmetrics has greatly promoted the development of online scientific communication and open science，and has been widely regarded by various research institutions and researchers as a new metric for scientific communication，dissemination and evaluation based on social network data. In order to ensure the accuracy and influence of scientific evaluation and measurement research in the Web2.0 environment，the most ideal way is to build data integration analysis platforms that can integrate Altmetrics data on various social platforms according to consistent standards. Network research method and in-depth interview method have been used. There are some foreign integrated analysis platforms for Altmetrics data，which can provide some useful reference experience，but there are also some problems to avoid. It is necessary and feasible to establish Altmetrics data integration and analysis platform by the library Community in China. Key points for libraries to integrate Altmetrics data are：expanding original data sources，using and popularizing DOI，developing unified data standards and scientifically designing indicators，constructing an overall approach to Altmetrics data collection and processing，collaborating extensively，and protecting users' private data.

Keywords：Altmetrics；Data Integration Analysis Platform；Construction of Library Information Resources；Open Data

（接第 99 页）

Review of Fine-Grained Sentiment Analysis Based on Text

Tan Cuiping

Abstract：Through literature research，this paper expounds the influence and promotion of fine-grained sentiment analysis from the perspective of different granularity levels，summarizes the latest tasks and technical methods of fine-grained sentiment analysis，and finally forecasts future trend of this research field.The relevant research results can provide the reference for subsequent research.

Keywords：Fine-Grained；ABSA；Textual Analysis