

研究数据管理工具发展研究

马建玲 曹月珍

【摘要】研究数据管理服务是图书馆在 e-research 环境下将要面对的一项新型服务，研究数据管理工具是辅助开展研究数据服务的重要保障。参考研究数据生命周期模型，根据数据生命周期各阶段研究数据管理需求的不同，提出一种研究数据管理工具分类方法，将研究数据管理工具划分为数据处理类工具、数据分发与出版类工具、数据分析类工具和集成管理工具，并对各类典型数据管理工具进行分析比较和推荐。

【关键词】研究数据 管理工具 工具分类 数据生命周期

Abstract: Research data management service is a new type of service that libraries will be faced on e-research environment. Research data management tools are the important guaranty for providing research data management service. Based on data management life cycle model and the different needs of data management, the paper presents a classification method for research data management tools, they can be divided into the tools of data processing, the tools of data distribution and data publishing, the tools of data analysis and the tools of integrating data management. The paper also analyzes and recommends some typical data management tools.

Key words: research data management tool tool classification data life cycle

1 引言

随着数字科研环境的发展，科学数据迅速激增和积累，图灵奖获得者 Jim Gray 提出了一个新的经验定律：网络环境下每 18 个月产生的数据量等于有史以来数据量之和^[1]。美国高校和研究图书馆协会（ACRL）2012 年发布了专题报告“学术图书馆和研究数据服务”^[2]，报告中指出当科学的合作性、数据驱动和可计算化不断加强，研究人员将面对大量的数据管理需要。研究数据在整个学术研究生命周期中的重要性不断显现，英国联合信息系统委员会 JISC 的报告“管理和共享数据（Managing and Sharing Data）”^[3]中指出，研究数据是非常有价值的资源，通常需要更多的时间和金钱来生产和创造，很多数据的价值可能会超过原始的研究。但对于研究数据的管理与服务目前还处于研究和发展阶段，相比较学术论文产出与出版的系统化规范化流程与管理，研究数据的管理与出版还亟待完善。研究数据管理与服务具有很强的复杂性和综合性，包括管理政策、管理机制及技术工具等多个方面。研究数据管理工具是辅助开展数据管理的重要基础和保障，目前针对研究数据生命周期各个环节对数据管理的需求已开发了一系列的工具软件和系统。本文将系统地研究数据管理工具进行梳理和研究，提出一种基于数据生命周期的研究数据管理工具分类方法，并对典型数据管理工具进行分析比较，希望对图书馆开展研究数据管理服务提供借鉴和参考。

2 研究数据管理生命周期模型

研究数据管理与数据生命周期紧密相关，研究数据管理工具的开发与应用需要基于数据生命不同阶段的需求，目前已出现了多种数据生命周期模型和框架，分别从不同角度描述数据从产生、收集、描述、存储、发现、分析到再利用的整个生命周期，此处主要列举两种应用比较广泛的数据生命周期模型。

2.1 DataONE 数据生命周期模型

地球数据观测网（Data Observation Network for Earth，简称 DataONE）为了向用户提供高效数据管理服务构建了一个数据生命周期模型（见图 1^[4]）。数据生命周期分 8 个组件，分别是：计划（Plan）：包括数据的描述以及如何管理数据和在数据的生命期中如何获取数据；收集（Collect）：包括通过手工、传感器及其他设备观测的

数字化数据; 保证 (Assure): 通过检查和检测确认数据的质量; 描述 (Describe): 使用适当的元数据标准精确完整地描述数据; 保存 (Preserve): 将数据提交到一个适当的长期保存系统 (如数据中心); 发现 (Discover): 潜在有用的数据包括与数据相关的信息能够被获取; 集成 (Integrate): 将不同来源的数据组合成一套均质的数据以便能够真正进行分析; 分析 (Analyze): 分析数据。

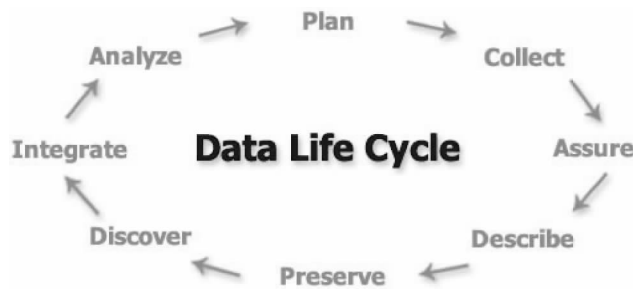


图1 DataONE 数据生命周期模型

2.2 DDI 组合数据生命模型

数据文档计划 (Data Documentation Initiative, 简称 DDI) 构建了组合数据生命模型 DDI 3.0 (如图 2^[5] 所示), DDI 3.0 模型包含了一个组合的研究数据生命周期, 特别适合于社会科学领域的的数据。模型主要根据科学研究生命周期中数据流确定了数据的处理流程, 包括研究概念确定、数据收集、数据处理、数据分发、数据存档、数据发现及数据分析。

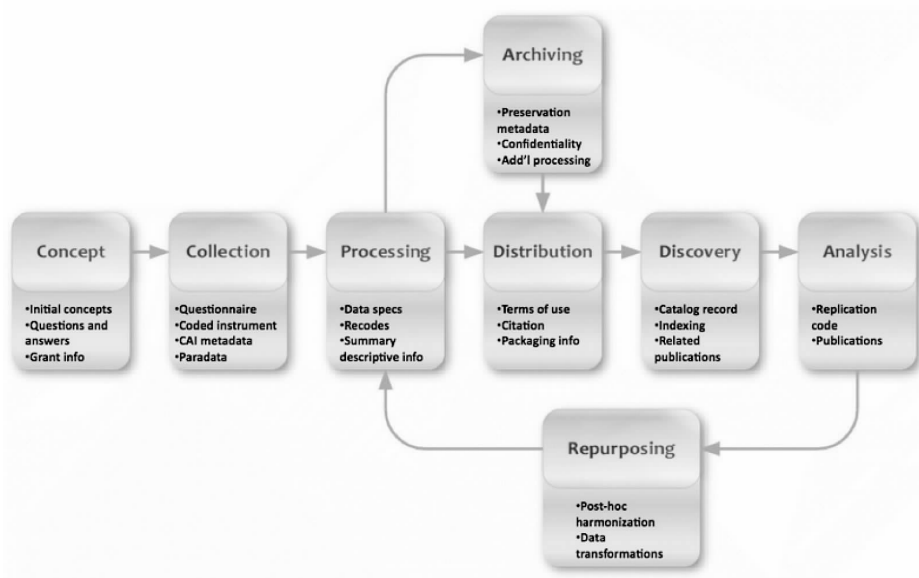


图2 DDI 3.0 组合生命周期模型

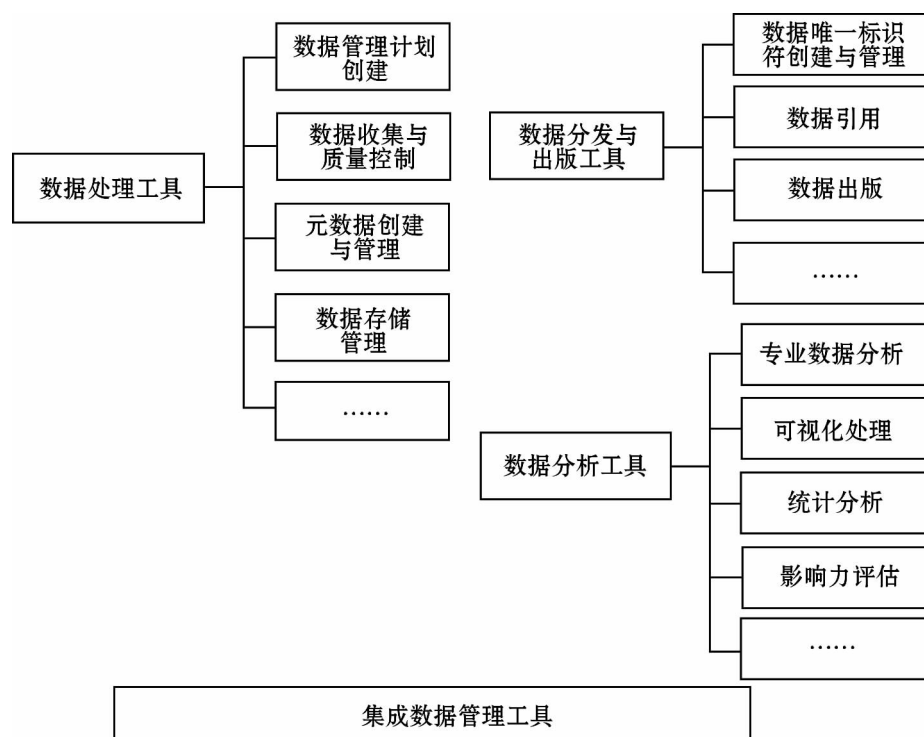
3 研究数据管理工具分类

数据管理贯穿整个科研过程, 数据生命周期各阶段有不同的数据管理需求, 结合上述两个数据生命周期模型中数据处理流程和步骤, 本文把研究数据管理工具划分为数据处理类工具, 数据分发与出版类工具和数据分析类工具, 如图 3 所示。其中数据处理类工具包括辅助进行数据收集、数据质量控制与保证、数据描述和数据存储的各类工具; 数据分发与出版类工具主要包括辅助创建唯一标识符、提供数据发布、引用服务和数据出版的工具; 数据分析工具包括影响力评估、可视化处理、数据统计分析工具等。此外, 在研究数据管理工具中还有一类工具属于集成数据管理工具, 即可以满足数据生命周期多个环节需求的综合类工具。

3.1 数据处理工具

3.1.1 数据管理计划创建工具 (Data Management Plan Tool)

美国基金组织如美国自然科学基金会 (NSF)^[6]、美国国家医学研究院 (NIH)^[7] 等都要求研究人员在提交项



目申请时提交研究数据管理计划（DMP），英国研究理事会（RCUK）^[8] 下属的7个专业理事会也都强制要求制定数据管理计划。在此背景下，为帮助科研机构 and 研究人员快速创建符合基金组织要求的数据管理计划（DMP），美国、英国先后开发了数据管理计划创建工具。目前应用最广泛的工具有：（1）加利福尼亚数字图书馆等开发的辅助创建和撰写 DMP 的工具 DMPTool，它的主要功能是帮助研究者选择某一特定的基金组织，创建即时可用的数据管理计划，DMPTool 主要提供给美国高校和科研机构，用户需要注册后使用，目前在美国已有 75 所大学注册使用。（2）英国数字化策管中心（Digital Curation Centra，简称 DCC）开发的帮助研究者和研究支持人员创建数据管理计划（DMPs）的工具 DMPonline^[10]，它可以帮助研究者明确如何创建、管理和共享数据，帮助研究者组织数据以便可以在需要时发现和理解它，通过更多的引用和更强的影响促进数据的可见性，用户需要用个人的 Email 或已认证的机构信息注册登录后使用系统。（3）综合地球科学数据应用社区（IEDA）开发的数据管理计划创建工具 IEDA Data Management Plan Tool v.2^[11]，其设计目的是以简单的方式创建 DMP，并且可以产生一个格式友好的 PDF 文档，清晰显示相关的信息。当 IEDA 聚焦有关海洋、陆地和极地环境的固体地球科学数据时，DMP 还设计了一种通用的形式，能够同时向 NSF 其他分部提交 DMP。

3.1.2 数据收集与数据保证（Collect and Assure）

数据收集因学科领域差异、项目目标不同，数据收集的内容和方法差异很大，因此一般没有通用的数据收集工具，但有些机构或项目提供了一些标准的数据收集检测方法。如 DataONE 项目对于数据收集制定了一套最佳实践的方法^[12]，主要内容包括：在数据收集时考虑创建一个使用模板，确保所有与数据相关的内容都会被完整收集，特别是有多个收集者同时收集数据时这点更加重要，模板中要描述数据文件的内容，包括定义每个参数、数据格式、使用单元、缺失值代码、提供公共参数格式示例等，数据描述文件应该作为一个“readme.txt”文件与数据文件共同保存，元数据文件应该采用通用的元数据标准。DataONE 也建议采用一致的数据组织方式，包括电子表格方式和数据库方式。此外，DataONE 对于编码标准、文件名命名、裸数据的收集和保存等都有详细的建议。

数据保证是在数据收集、登记和分析过程中确保数据质量和质量控制的过程。在数据收集过程中描述的任何条件都可能影响数据的质量，如收集数据过程中对缺失值的估算误差、人工输入错误、数据格式等问题。

3.1.3 元数据创建与管理工具（Metadata Creation）

元数据是关于数据的描述信息，包括如何收集数据、谁来收集以及收集的目的、描述项目和数据的基本元素等，这些信息是未来数据仍然可读与可重用的关键保证。元数据描述更多地是提供数据描述元数据模板、指南、

不同学科领域元数据标准。各领域也有很多专门的元数据创建与管理工具,例如: EPA Metadata Editor (EME)^[13]是一款简单的地理空间科学元数据编辑器,可以帮助用户创建和编辑符合美国环保署(EPA)和联邦地理科学数据委员会(FGDC)元数据要求的记录。Metavist^[14]是用于联邦地理数据委员会(FGDC)空间元数据标准的元数据编辑器,它的主要目标是提高对于属性和实体的编辑能力,提高对更多元数据元素的支持,提高对其他生态学元数据标准的支持能力等。Morpho Data Management Software^[15]是生物复杂性知识网络(KNB)创建的数据管理软件,它可以帮助生态学家创建元数据。英国数字化策管中心(DCC)开发了一系列开源的创建和操作元数据的工具^[16],包括 Curator's workbench, ICA - AtoM, Nesstar, PERMIS in METS Toolbox 等。

3.1.4 数据存储工具与系统

数据的保存与组织管理是数据管理中的重要内容,数据的数字化保存可以提高研究的影响力,保证数据的安全和长期可读与可用,同时也是很多基金组织的要求。用于研究数据存储与长期保存的工具和系统比较多,一般可以分为以下几种类型:

(1) 公共数据仓储库

当前各国基金组织、学术期刊、研究机构普遍采用的研究数据管理政策是鼓励研究人员将数据存储到所属学科的权威数据仓储库或公认的综合数据仓储库,如存储基因芯片数据到 Array Express 或 GEO,存储基因序列到 GenBank、EMBL 或 DDBJ,存储生态学数据到 Dryad 中。各学科领域的专门数据仓储库比较多, Nature^[17]、PLoS ONE^[18] 等期刊都给出了详细的推荐仓储库列表。下面重点推荐几个目前应用广泛的通用数据仓储系统,以及用于数据仓储库发现的数据仓储库目录和数据仓储库登记系统。

①Dryad^[19]是一个生命科学领域的仓储资源库,可以存储电子表格、文件、数据等各类出版数据,除了某些期刊有临时时间限制要求的数据外,全部存储在 Dryad 的数据允许下载和重用, Nature、Science、PLoS ONE 等期刊都将 Dryad 作为重要的数据仓储库推荐给作者存储数据。

②Figshare^[20]是一个开放获取的数据仓储系统,允许研究者以可引用、可检索和可共享的模式出版他们所有的数据,允许用户上传任何格式的文件包括图像、数据集、多媒体、论文、展板、演示文稿和文件等,提供长期保存,允许科学家开放的检索和共享数据。

③Databib^[21]是一个帮助研究人员标识和定位 Internet 网上研究数据仓储库的工具,是一个联机研究数据仓储库目录,它是由普渡大学图书馆联合宾州州立大学创建的,目前已经登记记录了 600 多个仓储库。

④re3data.org^[22]是由德国研究基金会创建的一个全球研究数据仓储注册系统,提供研究数据仓储长期保存并向研究者、基金组织、出版者和学术机构提供检索服务,目前已注册了 610 多个研究数据仓储。

(2) 数据存储管理系统创建工具

越来越多的大学和研究机构认识到研究数据的价值和重要性,再加上各基金组织也都提出了明确的数据管理要求,很多机构开始计划或着手建设机构数据仓储库,实现对研究数据的保存和管理,并帮助研究人员管理和组织数据。对研究数据的存储管理包括对个人数据、项目数据、研究室数据、机构数据等不同层次数据的存储和管理,针对不同的需求也开发了一系列研究数据管理系统创建管理工具,如 CKAN、UC3 Merritt、DataStage、DataBank、ScholarSphere 等,下面重点推荐 3 个工具系统。

①CKAN^[23]是由非营利组织开放知识基金会(Open Knowledge Foundation)开发的一个功能强大的开源数据门户平台软件系统,提供一套完整的即装即用的数据存储管理与发现的软件解决方案,实现数据的出版、共享、发现和利用。CKAN 目前主要用于政府机构和全球用户创建政府或社区数据门户,已经有超过 40 个数据仓储或门户系统使用 CKAN,如英国政府开放数据门户 Data.gov.uk。

②UC3Merritt^[24]是由加州大学数字图书馆策管中心(UC3)开发的新型的仓储服务系统,它可以向 UC 的组织和机构提供管理、存储和共享有价值的数字化资源。Merritt 能够提供对于数字资产的长期保存、与其他研究者共享研究服务,并且满足基金项目要求的数据共享和保存要求。

③ScholarSphere^[25]是由宾州州立大学开发的仓储服务系统,宾州州立大学科研机构可以基于网络共享他们的研究和研究产出,教师和学生能够使用 ScholarSphere 收集和存贮研究产出并且创建一个可读和可引用的记录,资源类型包括论文、演示文稿、出版物、数据集等。研究者也能够利用此服务完成基金组织共享和管理研究数据的要求。

(3) 扩展机构知识库(IR)系统功能,增强对于研究数据的管理与服务。

高等院校和研究机构建设机构知识库(简称 IR)已成为一种普遍趋势,据 OpenDOAR 统计全球已建立的 IR

已达到 2 545 个, 其中使用最多的开源 IR 软件为 Dspace^[27]。Dspace 软件本身具有对多种研究产出的存储支持, 包括研究数据, 但支持功能有限, 无法有效实现对研究数据的灵活描述、存储、组织和管理。因此, 有机构或项目开始研究扩展 IR 功能, 增强研究数据管理与服务功能的开发和扩展。如斯坦福大学图书馆开发的 Stanford Digital Repository (SDR)^[28], 能够提供对论文、开放获取资源、研究数据、网络资源、空间数据等多种数据的长期存储与发现服务。基于开源 IR 系统 Dspace 开发的中国科学院机构知识库网格^[29] 系统也开始扩展 IR 的功能和服务, 重点围绕科研项目过程, 对于科研过程中可能产生的各类过程文档和数据实现存缴、描述和长期保存, 并逐步支持对非文本资源的挖掘、识别和利用。

3.2 数据分发与出版工具

数据出版是将研究数据作为与学术论文、专利等同样重要的研究数据进行发布和出版, 它将有助于数据的发现与集成, 作者对科学数据资源进行引用, 出版商能进行链接, 同时也有助于对研究数据资源进行产出分析等。

3.2.1 数据唯一标识符创建和管理工具

为研究数据资源赋予数字对象唯一标识 (DOI) 是进行科学数据出版的基础, 只有对每一个数据对象进行唯一识别, 才能有效地对其产权信息、质量信息进行认证, 有助于形成科学数据领域的评价体系。国际上已经出现了提供数据唯一标识服务的系统以及创建和管理数据唯一标识符的工具。

(1) DataCite^[30]: 由德国国家科技图书馆牵头建设的 DataCite 系统是一个非营利组织, 目前已有 16 个国家 28 个机构加入成为 DataCite 会员, 它的目标是数据应该如其他学术资源如论文与图书一样被方便引用, 数据引用能够帮助我们方便地使用和验证数据, 能够跟踪数据的影响力, 可以构建一个承认和奖励数据生产者的学术框架。DataCite 通过与研究者、研究机构和出版者合作使研究数据可见和可检索。

(2) EZID^[31]: EZID 是由加州大学数字图书馆开发的创建、管理数据唯一永久标识符工具, 它的主要功能有为学术资源如文本、数据、术语等创建标识符, 以多种格式存储标识符的可引用元数据, 更新当前的 URL 保证引用永久可用, 支持不同的唯一标识符, 包括 ARKs 和 DataCite 的 DOIs, 能够通过 EZID 程序界面自动操作和生成唯一标识符。

3.2.2 数据出版系统和工具

(1) Dryad 是一个生命科学领域仓储库, 同时也是一个数据出版系统, 它有效地把期刊、出版者、科学家和其他利益相关者连接在一起。Dryad 系统的核心是“论文与数据的集成提交 (Integration of Manuscript and Data Submission)”, 允许期刊出版者方便地将论文稿件和数据提交到 Dryad 中^[32], 实现论文与数据的集成出版。

(2) 数据期刊出版系统 (data journal): 数据期刊是一种新型的学术出版物, 它一般以网络期刊形式出版, 它并不是只出版数据, 而是以出版 data papers 为核心, 所谓 data paper 指描述一个数据集, 给出数据收集、处理过程、软件工具、文件格式等的细节描述^[33], 它是一类以描述一个或一组数据集为首要目标的出版物, 有些数据期刊只出版短 data papers, 如 Geoscience data journal, 出版物可以不包括对于假设和结论的分析和背景, data papers 直接链接到存储在数据中心的数据集。但也有部分数据期刊不仅只出版数据, 如数据期刊的代表之一生物多样性数据期刊 (Biodiversity Data Journal, BDJ)^[34] 创建了一种基于同行评议、开放获取、综合的联机出版平台, 可以实现多样化的数据出版, 如支持期刊站点提供附加数据文件下载, 数据可以存储在特殊指定的数据仓储中 (如 Dryad, Pangaea), 数据可以以标记和机器可读形式出版, 提供扩展的多媒体和语义增强服务等^[35]。

(3) PANGAEA 是一个开放获取的地球和环境科学领域数据仓储库, 同时也是一个地球系统科学数据出版系统^[36]。数据内容可以通过一个搜索引擎、一个数据仓储和 Web 服务检索获取, 它基于 DOI 标识符提供对于数据集的同行评议和引用, 同时与多个学术出版商如 Elsevier, Springer, AGU, Thompson Reuters 等合作。PANGAEA 还提供数据与学术论文的同步出版。

3.3 数据分析处理工具

3.3.1 专业数据分析处理工具

不同学科领域产生数据的环境不同, 产出数据类型、范围等也有很大区别, 如生物学领域需要对有机体数据进行分析, 物理学、计算机科学可能需要对大量的传感器数据进行处理和分析, 化学家则需要处理大量的实验数据, 因此, 在各个学科领域开发了大量的专业数据分析处理工具, 此类工具数量较多、专业性很强。如生物学领域的 Mesquite、Predictive Ecosystem Analyzer (PEcAn)、PyDSTool、BioSens (UCSB) 等软件工具, 化学领域的 Chemometrics、SLIMS (Small Laboratory Information Management System) 等, 地球科学领域的 ArcGIS

Desktop、ESRI ArcMap 等。

3.3.2 数据可视化处理工具

数据可视化是一门对数据进行视觉化表达的应用学科,数据可视化的目的是为了有效传达信息,通过对海量的复杂数据进行分析,以直观的视觉手段揭示数据间的复杂数据和隐性关系。数据可视化也是研究数据分析的一项重要重要手段,相应地开发了一系列研究数据可视化分析处理工具。如 OMERO 是一款用于管理、分析和可视化处理生物图像数据的工具, Arcgis 可以实现对地理空间数据的可视化表达, VMD 主要用于实现对大蛋白数据的可视化表达和显现, ChemWindow 是一个化学领域的画图工具, CyBy(2) 是一个化学和生物科学领域基于框架的数据管理工具,用于存储和可视化结构化数据以及其他信息, GeoMapApp 是一个数据探索和可视化工具,已经成为海洋地球科学数据系统的一部分。此外还有很多专业软件工具包含可视化功能和模块,如 GeoSetter、Data Science Toolkit、CrystalMaker 等。

3.3.3 数据统计分析工具

通过观测、实验、测量和调查得到的数据是一种以数量形式显示的事实,数据统计分析的目的是把隐没在大量杂乱无章数据中的信息、知识萃取和提炼出来,找出研究对象的内在规律,因此,数据统计是数据分析的重要内容,特别是在社会科学领域,数据统计分析更是数据分析的最重要手段。数据统计工具已发展的比较成熟,出现了一批应用比较广泛的工具,如 SPSS、SAS、Statistica、BMDP、S-PLUS 等。美国统计协会(American Statistical Association)^[37]是全球最大的统计组织,它在网站上列举了大量的数据统计工具、项目、软件包等,包括用于一般数据统计、空间统计、可视化、数据挖掘、实验设计、数学计算、数据转换等。

3.3.4 影响力评估工具

传统文献计量学已经有一批成熟的工具,如 Web of Science、TDA、Thomson Innovation、CiteSpace 等,通过文献引用、合作等数据分析评估作者、机构、国家、研究团队等的学术影响力。网络计量学是文献计量学在网络环境下的最新发展,且逐渐形成和发展了一些适应网络计量分析的工具和方法,如利用搜索引擎、Web 爬虫、Citation Analysis 软件包分析比较链接数、URL 引用数等;此外,近几年随着网络的发展,也出现了一批基于社会网络的学术影响评论工具或系统,甚至产生了一个新的研究领域——补充型指标计量学(Altmetrics),工具包括 Twitter、Citeulike、Mendeley、ORCID、Zotero^[38]。这些工具可能不直接用于影响力计量计算,但是它们能够帮助发现影响力和成功的线索或证据。2012 年 Thomson Reuters 开发了 Data Citation Index 数据库,这是目前第一个数据索引工具,它精选高质量的研究数据、数据集、数据仓储建立索引库,允许研究和发现科学数据并且链接到对应的出版文献,可以通过引用提高研究数据的影响力。

3.4 集成数据管理工具

随着基金组织、科研机构、学术期刊对研究数据管理的要求不断明确和深化,研究人员和科研机构的数据管理任务不断加重,为解决这些问题,美国、英国等先后启动了一批研究数据管理研究框架项目,相应地开发出一批集成研究数据管理工具和系统。所谓集成数据管理工具是指能够帮助完成数据生命周期多个环节需求的综合管理工具,可以利用一个工具全面完成多项数据管理任务。这类工具功能完备,特别是一些机构开发的工具非常符合机构数据管理的要求和特点,取得了较好的应用效果。具体如下:

(1) 英国 JISC 的管理研究数据计划 2011-13 (Managing Research Data Programme 2011-13)^[39]中包含多个数据管理计划框架项目,每个项目产生了一批数据管理工具或系统。如由牛津大学主持的 ADMIRAL and DataFlow 项目,它主要目标是提供组织、收集和标注生物学数据集的环境和辅助工具,项目中开发的工具有 DataStage 和 DataBank。其他的数据管理框架项目还有 FISHnet、I2S2、IDMB、MaDAM、SUDAMIH 等,每个项目也都开发了相应的各类数据管理工具。澳大利亚教育部资助莫纳什大学的 ARCHE R 项目^[40]开发了一套用于研究数据管理的 e-Research 工具,它提供一个框架辅助研究人员收集、管理、存储、合作和出版科学数据。

(2) DataUp 是由加利福尼亚大学数字图书馆开发的一个开源工具,它的主要功能是帮助研究人员建立数据文档、管理和存储他们的数据^[41],是能够嵌入科学家 workflow 并且可以集成在 Excel 中。

(3) DataConservancy^[42]是由约翰霍普金斯大学 Sheridan 图书馆牵头的项目。DataConservancy 开发了一套专门用于研究数据组织与管理的工具软件包,可以完整实现数据收集和保存、数据发现与共享。工具的主要特点在于提供基于空间、时间和术语的知识化的数据组织,可以实现跨学科的数据集成,提供多种互操作的 API,是基于数据对象的仓储库,已经实现了与 NSDIC, ArXiv 等系统的集成。

(4) Dataverse Network^[43]是哈佛大学开发的一个开源应用系统,主要用于出版、共享、引用、发现和分析研究数据,研究人员可以通过系统把数据提供给其他用户使用,也可以备份自己的工作,研究人员、作者、出版者、数据分发人员和机构都可以获得相应的数据保证服务。它也是一个用于长期保存和存档实践的数据仓储库。通过稳定的数据引用支持研究数据共享,也能够支持基于数据重现科学研究。

4 结语

研究数据在学术产出中被视为“二等公民”^[44],这不是说研究数据不重要,而是表明研究数据的管理和出版亟需规范和加强,在 e-Research 环境下产生的研究数据几乎全部是数字化的数据,对于这类数据的管理必须依靠各类数据管理工具。本文系统梳理了对应研究数据生命周期各环节需求的研究数据管理工具的类型并且推荐了一批典型工具,可以说研究数据管理工具是一个非常大的集合,也是一个快速发展的领域。研究数据来自不同学科领域,数据管理的需求差异很大。图书馆要开展研究数据管理服务,数据管理工具是必不可少的,需要学科馆员根据不同学科的特点,进一步发现和精选研究数据管理工具,一方面向科研人员推荐工具,另一方面也在研究数据服务中筛选和应用工具;需要信息技术人员根据研究人员、研究机构实际需求开展研究数据管理工具和系统的研究与开发,逐步提升图书馆研究数据管理与服务的能力,真正提供有效和高质量的研究数据管理服务。

注释

- [1] Jim Gray. What Next? A Dozen Information - Technology Research Goals [EB/OL]. [2013-12-10]. http://research.microsoft.com/pubs/68660/ms_tr_99_50_turingtalk.pdf.
- [2] Carol Enopir, Ben Birch, Suzie Allard [EB/OL]. [2013-12-12]. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf.
- [3] Veerle Van den Eynden. Managing and Sharing Data [EB/OL]. [2013-12-12]. <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>.
- [4] Data Life Cycle [EB/OL]. [2013-12-13]. <http://www.dataone.org/best-practices>.
- [5] DDI Lifecycle [EB/OL]. [2013-12-13]. <http://www.ddialliance.org/what>.
- [6] NSF Data Management Plan Requirements [EB/OL]. [2013-12-20]. <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.
- [7] NIH Data Sharing Policy [EB/OL]. [2013-12-20]. http://grants.nih.gov/grants/policy/data_sharing/.
- [8] RCUK Common Principles on Data Policy [EB/OL]. [2013-12-20]. <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>.
- [9] DMPTool [EB/OL]. [2013-12-22]. <https://dmp.cdlib.org/>.
- [10] DMPOnline [EB/OL]. [2013-12-22]. <https://dmponline.dcc.ac.uk/>.
- [11] Data Management Plan Tool v. 2 [EB/OL]. [2013-12-22]. <http://www.iedadata.org/compliance/plan>.
- [12] Primer on Data Management [EB/OL]. [2013-12-23]. http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf.
- [13] The EPA Metadata Editor (EME) [EB/OL]. [2013-12-23]. <https://edg.epa.gov/EME/>.
- [14] Metavist Version 2 [EB/OL]. [2013-12-23]. <http://metavist2.codeplex.com/>.
- [15] Morpho Data Management Software [EB/OL]. [2013-12-23]. <http://knb.ecoinformatics.org/morphoportal.jsp>.
- [16] Creating and Manipulating Metadata [EB/OL]. [2013-12-25]. <http://www.dcc.ac.uk/resources/external/category/creating-and-manipulating-metadata>.
- [17] Recommended Data Repositories [EB/OL]. [2013-12-25]. <http://www.nature.com/scientificdata/for-authors/data-deposition-policies/>.
- [18] PLOS Editorial and Publishing Policies [EB/OL]. [2013-12-26]. <http://www.plosone.org/static/policies.action>.
- [19] Dryad [EB/OL]. [2013-12-27]. <http://datadryad.org/>.
- [20] Figshare [EB/OL]. [2013-12-27]. <http://figshare.com/>.
- [21] Databib [EB/OL]. [2013-12-27]. <http://databib.org/>.
- [22] Re3data [EB/OL]. [2013-12-27]. <http://www.re3data.org/>.
- [23] CKAN [EB/OL]. [2013-12-28]. <http://ckan.org/about/>.
- [24] UC3Merritt [EB/OL]. [2013-12-28]. <https://merritt.cdlib.org/>.
- [25] What is Scholar Sphere [EB/OL]. [2013-12-28]. <https://scholarsphere.psu.edu/>.
- [26] Databank [EB/OL]. [2013-12-29]. <http://www.dataflow.ox.ac.uk/index.php/about/about-databank>.
- [27] OpenDOAR [EB/OL]. [2013-12-30]. <http://www.opendoar.org/find.php?format=charts>.
- [28] Stanford Digital Repository (SDR) [EB/OL]. [2013-12-30]. <http://lib.stanford.edu/sdr>.

- [29] 中国科学院机构知识库服务网络 [EB/OL]. [2013-12-30]. <http://www.irgrid.ac.cn/>.
- [30] DataCite [EB/OL]. [2013-12-30]. <http://www.datacite.org/>.
- [31] EZID [EB/OL]. [2013-12-30]. <http://n2t.net/ezid>.
- [32] Dryad: Scalable and Sustainable Infrastructure for the Publication of Data [EB/OL]. [2013-12-31]. http://wiki.datadryad.org/wg/dryad/images/e/ec/Dryad_abil1_short.pdf.
- [33] Geoscience Data Journal [EB/OL]. [2013-12-31]. <http://onlineibrary.wiley.com/journal/10.1002/> (ISSN) 2049-6060.
- [34] Biodiversity Data Journal [EB/OL]. [2013-12-31]. <http://biodiversitydatajournal.com/>.
- [35] Lyubomir Pernv, etc. Making Small Data Big! The Biodiversity Data Journal [EB/OL]. [2013-12-31]. http://vsmith.info/files/2012Biodiversity_Data_Journal.pdf.
- [36] PANGAEA? Data Publisher for Earth & Environmental Science [EB/OL]. [2013-12-31]. <http://www.pangaea.de/>.
- [37] American Statistical Association [EB/OL]. [2014-01-02]. <http://www.amstat.org/>.
- [38] Euan Adie. New Perspectives on Assessment How Altmetrics Measure Scholarly Impace [EB/OL]. [2014-01-02]. <http://www.slideshare.net/BaltimoreNISO/niso-webinar-new-perspectives-on-assessment-how-altmetrics-measure-scholarly-impact>.
- [39] Managing Research Data Programme 2011-13 [EB/OL]. [2014-01-03]. <http://www.jisc.ac.uk/whatwe-do/programmes/diresearchmanagement/managingresearchdata.aspx>.
- [40] eResearch Toolset Overview [EB/OL]. [2014-01-21]. <http://www.ijdc.net/index.php/ijdc/article/view/99/74><http://archer.edu.au/>.
- [41] DataUP [EB/OL]. [2014-01-04]. <http://dataup.cdlib.org/>.
- [42] DataConservancy [EB/OL]. [2014-01-05]. <http://dataconservancy.org/>.
- [43] Dataverse Network [EB/OL]. [2014-01-05]. <http://thedata.org/>.
- [44] University of California Curation Center California Digital Library. The Data Management Ecosystem [EB/OL]. [2014-01-05]. <http://www.slideshare.net/jakkb1/jak-data-mgmtecosystem>.

马建玲 中国科学院国家科学图书馆兰州分馆, 研究馆员。

曹月珍 中国科学院国家科学图书馆兰州分馆, 研究生。

(上接第 39 页)

学体系的研究不仅有助于解决各个地区和各个服务机构自行开发编目标准所带来的一些问题: 如著录规则差异较大, 标引语言不统一, 标准体系滞后等问题; 还有助于建立专业、权威、公正和具有自我建设能力的 RAD 编目标准第三方认证主体, 较好地解决文献服务机构长期存在的编目标准差异化的问题; 也有助于建立健全相关的制度及法规问题, 规范行业行为, 使国内与国际接轨。该方法体系的具体实施过程牵涉本体 (Ontology)、企业架构 EA 和企业架构 EA 成熟度模型等相关技术, 将这些技术熟练且能融会贯通地应用, 促使编目标准认证能更好地服务于编目工作。虽然该编目标准认证体系在实施的过程中存在一定的困难且有相当的难度, 但是通过对 RDA 编目标准认证的细致分析研究, 可以为日后更新我国编目规则做好充分准备, 也可加速我国编目行业的健康发展。

注释

- [1] [5] 张秀兰. RAD 与我国编目工作的改革 [J]. 情报资料工作, 2013 (1): 86-89.
- [2] [3] [4] 杨莉萍. 美国国会图书馆 RDA 服务实践所感 [J]. 现代情报, 2011 (11): 75-79.
- [6] 国家图书馆外文采编部. “RDA 理论与实践” 培训班暨 “RDA 在中国的实施和挑战” 研讨会 [EB/OL]. [2012-10-03]. http://www.lsc.org.cn/CN/News/2012-07/EnableSite_ReadNews1923460761342281600.html.
- [7] 王春英. 浅议我国 NGO 管理中的第三方评估机制 [J]. 学会, 2009 (6): 23-26.
- [8] 肖花, 刘春年. Ontology 及其在企业架构中的应用 [J]. 图书馆学研究, 2012 (15): 65-68.
- [9] 业务架构和微软 IT 规划方法 [EB/OL]. [2013-12-02]. <http://www.uml.org.cn/qiyexinxi/200907314.asp>.
- [10] 张睿, 王泐. 利用企业架构理论规避信息孤岛问题探析 [J]. 现代商贸工业, 2012 (4): 240-241.
- [11] 李屹, 王学颖. 企业架构 EA 成熟度模型比较研究 (II) [J]. 沈阳师范大学学报 (自然科学版), 2010 (1): 105-108.
- [12] [13] [14] [15] 杨周南, 朱建国, 刘锋. XBRL 分类标准认证的理论基础和方法学体系研究 [J]. 会计研究, 2010 (11): 10-15.

肖花 江西农业大学图书馆, 助理馆员, 硕士研究生。

刘春年 南昌大学信息管理系, 教授。