● 专论：健康数据资源长期保存 ●

人口健康科学数据长期保存系统建设路径探析*

杨晨柳 方 安 胡佳慧 姚宽达 王 茜

（中国医学科学院/北京协和医学院医学信息研究所 北京 100020）

〔摘要〕 立足人口健康科学数据长期保存需求，分析 5 个国外高校科学数据长期保存系统建设实践及特点，详细阐述其保存策略和技术措施，总结其在科学数据长期保存过程中的关注重点和问题，为人口健康科学数据长期保存建设路径提供参考和借鉴。

〔关键词〕 人口健康科学数据；长期保存；数据管理；可信存储

〔中图分类号〕R－058 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2021.09.003

Exploration of Population Health Data Long－term Preservation System Construction *YANG Chenliu, FANG An, HU Jiahui, YAO Kuanda, WANG Qian, Institute of Medical Information, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100020, China*

〔**Abstract**〕 In view of the long－term preservation needs of population health data, the paper analyzes the construction practice and characteristics of long－term science data preservation systems in five overseas universities, expounds the preservation strategies and technical measures in detail, and summarizes the key points and issues in the process of long－term preservation of science data, so as to provide references for population health data long－term preservation construction.

〔**Keywords**〕 population health data; long－term preservation; data management; trusted archiving

1 引言

科学数据作为数据驱动和数据密集时代的基础战略资源，因其对科技创新和经济发展的重要意义而受到高度重视[1]。在国务院办公厅、国家互联网信息办公室、全国人大常委会相继发布的《科学数据管理办法》[2]《数据安全管理办法（征求意见稿）》[3]《中华人民共和国数据安全法》中[4]，要求科研院所、高等院校和企业等法人单位建立科学数据保存制度，配备数据存储、管理、服务和安全等必要设施，实行更加严格的管理制度（实施数据

〔修回日期〕 2021－01－15

〔作者简介〕 杨晨柳，助理研究员，发表论文 8 篇，参编论著 1 部；通讯作者：王茜，馆员，发表论文 20 余篇。

〔基金项目〕 国家人口与健康科学数据共享服务平台后补助项目“人口与健康科学数据长期保存”（项目编号：NCMIKD07N－201905）；中国医学科学院医学与健康科技创新工程服务“一带一路”战略先导科研专项“卫生信息服务研究”（项目编号：2017－I2MB&R－10）；中国医学科学院医学与健康科技创新工程协同创新团队项目“中文临床医学术语系统构建研究”（项目编号：2017－I2M－3－014）。

技术防护、开展数据安全风险评估等），保障科学数据完整性和安全性。人口健康科学数据涉及生物医学、基础医学、临床医学、公共卫生、中医药学、药学、人口与生殖健康等多个学科领域，具备较高的科研和利用价值，有效支撑着医学创新发展、临床诊疗和疾病预防等工作，一旦数据丢失，将会带来无法弥补的损失，亟需对其进行长期保存。因此，本文围绕人口健康科学数据有效保存和管理的目标，调研国外建设程度较好的科学数据长期保存系统，梳理其在解决科学数据长期可获取、可复用、可解释等问题方面所实施的保存策略和技术措施，探析人口健康科学数据长期保存系统建设路径，以期为人口健康科学数据长期保存实践提供指引和参考。

2 研究对象及方法

2.1 调研对象

本文从代表性和实用性角度选择5个建设较早且较为完善的国外高校科学数据长期保存系统，包括美国俄勒冈州立大学学术保存系统（Scholars Archive，SA），美国明尼苏达大学数据保存系统（Data Repository for University of Minnesota，DRUM），澳大利亚迪肯大学数据商店（Deakin University Data Store，DUDS）和在线研究数据系统（Deakin Research Online，DRO），美国斯坦福大学数字保存系统（Stanford Digital Repository，SDR）以及澳大利亚国立大学数据档案系统（Australian Data Archive，ADA）。整体来看上述系统多为高校图书馆建设并维护，面向教职工和在校服务生提供数据保存管理和共享发布平台服务，为包含研究、学术、历史记录等内容的科学数据提供10年以上的长期保存服务以便于对其后续利用。

2.2 研究方法

基于已选取的5个科学数据长期保存系统，从保存策略、技术措施两个方面，数据遴选、接收限制、保存管理、评估审计、开放获取、数据分类、格式推荐、命名规范、版本控制、数据备份10个维度，比较分析上述系统建设的具体做法和实践特点，为人口健康科学数据建设路径提供借鉴。

3 结果分析

3.1 保存策略

3.1.1 数据遴选 并非所有数字资源都需要长期保存，因此根据一定标准或优先级对数字资源进行遴选能够为长期保存活动提供完整、准确、可靠的数据资源。目前已有部分高校创建数据遴选规则，针对保存需求进行保存数据质量筛选。如DRUM重点对关键、高价值的研究数据进行长期保存，推荐数据为最终或已发布状态[5]；SDR优先保存非专有、未压缩、普遍使用、遵循开放标准的科研数据资源，便于实现数据长期可获取和可使用[6]。值得注意的是，在进行保存数据资源遴选时，应重视被保存科学数据资源的去隐私问题，强调被保存数据资源不能包含诸如人名、地址、电话号码等涉及隐私、机密的敏感数据项，或受法律保护的信息。

3.1.2 接收限制 长期保存数据接收方案主要基于院校数据保存要求制定。除澳大利亚迪肯大学外其他高校均制定比较详细的数字资源接收方案且重点强调数据传输要求和描述信息两方面。数据传输部分，高校普遍支持基于网页的数据接收模式并限制上传数据文件大小，对于无法在线上传的较大文件则根据需求协商存储计划以实现数据存储；描述信息部分，要求数据所有者或提交者提供描述数据性质的说明文档，包含但不限于数据内容、范围、格式、保存和元数据等信息，用于数据的保存、重用和发现。如ADA提供多种科学数据接收方式，包括FTP、AARNet、CloudStor等工具安全在线传输数据、利用电子邮件传输较小数据文件、将数据文件复制到DVD和USB驱动器等产品上通过密码保护或加密方式进行邮寄等[7]。

3.1.3 保存管理 保存管理是长期保存过程中一系列必要管理活动的统称，其确保在必要时间内持续访问科学数据，实现保存数据长期可查找、可访问、可互操作和可重用。目前高校主要通过参考并引用已有相对成熟的数据管理规范实现长期保存的

数据管理。如DRUM参照数据管理计划（Data Management Plans，DMP）[8]，制定元数据创建、数据质量控制、操作流程监督、数据格式转换、资源共享协议、数据引用规则等涉及数据全生命周期的管理规范，实现数据下载和浏览频率的记录、统计和分析功能；澳大利亚迪肯大学参考《澳大利亚负责任研究的行为守则》制定数据保存策略[9]，以满足其科研数据长期保存管理需求。

3.1.4 评估审计 审计是长期保存最重要的功能之一，主要对实时故障进行监控和检测，处理和降低数据丢失风险，特别是访问和使用较少的数字资源，确保长期保存系统稳定性、可靠性和可用性。除SA未明确提及数据审计方案外其他高校已普遍完成数据长期保存审计体系建设。如DUDS&DRO制定数据机密性、隐私权以及知识产权风险控制策略，降低科学数据丢失、被盗、损坏、腐蚀的可能性[10]；SDR强调为数据提供高度安全、可管理的存储环境，保护存储文件数位和字节免受损坏。为解决数字资源保存可信赖问题，欧洲委员会讨论通过欧洲数字存储库审计和核证基本框架，有效地创建分层认证方法，如基于自我评估和同行审查的数据批准印章（CoreTrustSeal）、要求全面的自我评估信息和文件——可信赖的数字档案标准（DIN 3164）以及将自我评估与外部审计相结合的审核和认证值得信赖的数字存储库（ISO 16363：2012）。目前SA[11]和ADA[12]已通过Core Trust Seal认证要求成为国际认可的可信赖长期保存平台。

3.1.5 开放共享 组织按照统一管理策略向外部有选择性地提供其所掌握的数据内容，实现数据跨组织、跨行业流转和交换的行为。所调研高校均为其在校师生提供科学数据开放共享服务。为保证数据创建者权益，部分院校制定6个月到3年不等的数据发布限制期。实际应用中除DRUM要求数据必须支持公开访问外，其他系统则通过设置公开、部分公开、非公开等数据访问权限实行对科研数据的分级分类访问。其中SDR基于图书馆目录Search-Works共享科学数据支持搜索、浏览数据相关内容最新进展；ADA则要求用户填写科学数据存储表格和许可表格，为数据提供足够描述信息，有效界定数据访问条件。

3.2 技术措施

3.2.1 数据分类 数据格式管理的前提。通常情况下长期保存系统支持所有格式数据存储，但在实际操作中高校会对数据进行初步分类以便进行针对性保存和管理，维护数字格式可持续性。如OSU基于数据生成方式将其分为观测、实验、衍生或编译、仿真、引用或规范数据5类；UM将能够实现格式标准化的科学数据分为数据集、地理空间数据、动态影像数据、声音、统计信息、图片、表格、文本、网页、容器等；SDR依据数据通用标准以及单一学科特殊要求将数据分为书籍、手稿、地图、照片、口述历史、音乐、视频、数据、网站、学位论文、期刊文章、3D对象等[13]；ADA将其保存资源分为定量和定性数据，方便进行存储及管理。

3.2.2 格式推荐 数据格式描述数字对象的文件或位流的特定结构并指明能够处理该结构的相关应用程序，保证数据文件长期可读取。格式管理作为数字对象管理的核心内容，受到长期保存机构广泛关注。整体来看所调研高校普遍要求被保存数据文件格式能够满足公开获取、普遍可用的特征，管理方法分为3类：一是参考已被广泛接收且认可的数据文件格式指南，解决长期保存数据格式过时问题。如SDR参考美国国会图书馆发布的《推荐格式声明》[14]将数据文件保存为开放格式，涉及专有格式文件时在目录中创建一个readme.txt文件，记录生成该文件的软件名称、版本等信息，帮助用户识别并再次打开文件。二是基于长期保存需求提供不同类型科学数据的保存格式建议。如DUDS&DRO参考格式指南ANDS Guide[15]的同时，使用澳大利亚迪肯大学软件目录推荐的应用程序和工具；SA将科学数据按照适合长期保存要求的格式进行存储并提供对应说明文档；ADA不仅针对数据类型推荐存储格式，还提供数据格式转换服务以处理数据格式过时问题，保证数据长期可用。三是引入数据管理专家监督与审核机制，配合格式管理技术支持。如DRUM基于数据管理专家与数字化数据拥有人员沟通

协商机制，确保所存储数据采用能够支撑长期访问、发现和重用的格式和结构，同时创建虚拟计算环境以模拟运行过时数据格式所需原始系统或软件。

3.2.3 命名规范 通过赋予目标对象及文件统一的命名规则实现对特定文件和数据的索引和定位，能够支持系统对数据进行存取、更新、替换和查找等操作。从规范角度看所调研院校均已创建文件命名规范并要求遵循一致性和描述性原则。如OUS强调文件名称需包含足够多的描述信息、提供文件背景信息等，便于从文件版本、目录结构、文件命名、文件结构、备份方面对保存数据内容进行管理。从实施现状看长期保存系统普遍支持对数据库中数据便捷引用的永久链接和唯一标识符方案，此外DRUM等系统还基于Google Scholar、Web of Science对搜索引擎进行优化，确保其所有数据集都包含完整索引信息[16]。

3.2.4 版本控制 由于保存需求不同，各机构提出的长期保存版本控制方案及实施程度存在差异。DUDS&DRO参考英国莱斯特大学的版本控制图[17]设计版本控制功能并制定明确规范以确保文件版本追踪，如将日期、时间作为文件名的一部分，保留文件主副本内容；SDR通过跟踪数据对象变更行为提供版本回溯功能，基于数据最新版本，依据需求访问之前任一版本。

3.2.5 数据备份 保证科学数据长期保存可持续性的必要手段，通过对数据文件进行冗余存储避免数据损失，满足其长期访问需求。传统档案化存储策略已基本普及，该方法主要通过安全备份、数据校验、灾难恢复等保障手段为数据文件提供安全保存环境。近年来随着虚拟化和云服务的快速发展，数据存储服务开始逐步由线下转变为线上。作为在线存储服务的典范，DUDS&DRO通过提供不同类型网络硬盘实现数据文件存储、共享、访问、备份等管理功能[18]。为防止存储介质退化，DUDS&DRO利用工具在计划时间内执行自动备份策略，将数据转移到网络驱动器上，同时异地多副本存储以便于随时进行数据恢复。区别于DUDS&DRO多类型网络硬盘，SDR将存储内容保存在旋转磁盘上，而多个副本存储在LTO磁带上，以保证可随时提供在线访问副本。

3.3 小结

分析对比后形成高校科学数据长期保存系统建设情况概览，见表1。所调研高校不同程度制定了保存策略方案并基于自身数据保存需求各有侧重，其中接收限制、保存管理、评估审计建设较为完善。具体实施层面，各系统普遍强调对于科学数据类型和文件格式的规范管理，以应对数据创建和保存技术逐步更新淘汰带来的数据无法正常使用和读取问题，同时提出制定有效的数据备份机制并选择长期可用的存储介质方案，如网络硬盘、云存储等，保证数据安全性和稳定性。

表1 高校长期保存系统整体建设情况概览

系统名称	保存策略					技术措施				
	数据遴选	接收限制	保存管理	评估审计	开放获取	数据分类	格式推荐	命名规范	版本控制	数据备份
SA	×	√	√	×	√（C）	√	√	√	√	-
DRUM	√	√	√	√	√	√	√	√	√	√
DUDS&DRO	×	×	√	√	√（C）	×	√	√	√	√
SDR	√	√	√	√	√（C）	√	√	√	√	√
ADA	√	√	√	√	√	√	√	√	√	-

注：√表示已建设，×表示未建设，√（C）表示有条件的建设，-表示未明确表述。

4 启示与借鉴

4.1 概述

通过对上述国外高校长期保存系统建设实践的分析对比，可以发现科学数据长期保存主要涉及元数据、文件命名、可信赖存储、开放获取、格式管理、可持续保存共6个方面，通过参考和借鉴国外高校建设经验，人口健康科学数据长期保存从上述内容出发，开展方案设计与系统建设。

4.2 制定元数据接收规范

元数据是数据的数据，主要用于记录数字资源结构以及生成、保存、检索、应用的全过程，方便对数据进行有效组织和统一管理，确保数据可以被查找和识别。同时元数据被用作数据一致性、真实性、完整性的判断依据。为保证数据可理解性，接收数据时，人口健康科学数据长期保存系统要求科学数据创建者/所有者提供描述数据性质的说明文档，帮助完成对数据真实性、完整性、可靠性的评估和审核，分析和解释保存数据详情。依托人口健康科学数据元数据，基于数据资源有效的描述、组织和管理信息以及数据结构、环境信息、变动历史等重要内容，能够实现对科学数据全生命周期管理的溯源，保障科学数据可靠性。

4.3 重视数据文件命名规范

文件命名是人口健康科学数据组织管理的重要内容，将接收的每个文件都基于已制定的命名规则生成唯一标识符，用来避免与其他文件产生混淆，便于实现数据存取、更新、替换、查找等各项操作。遵循《科技资源标识符》[19]（CSTR）（GB/T32843－2016）方案，人口健康科学数据对象标识符内容包含资源代号（4位CSTR字母代码）、注册机构（5位CSTR注册机构代码）、资源类型（2位CSTR资源类型代码）和内部标识符（6位日期+6位流水号）4部分内容，此外允许内部标识符扩展“字母+数字”形式的文件编号信息。基于人口健康科学数据长期可获取需求，清晰明确的命名规范将更有利于对数据文件的查找和识别，方便系统对存储数据管理和维护，推动人口健康科学数据长期存储、管理、应用和共享。

4.4 配置可信赖存储环境

提供安全的数据保存环境是长期保存活动开展的基础，可信赖长期保存系统强调数据长期可发现、可访问、可交互、可复用，要求从基础设施、保存系统、数据对象、人员管理等方面加强安全管理，避免对数据安全造成威胁，增强保存系统本身及内部数据可信度。人口健康科学数据长期保存系统基于ISO16363可信赖仓储和审计标准，从组织基础设置、数字对象管理、基础设施和安全风险管理3方面开展数据长期存档和审计环境建设，实现限制用户访问内容、避免数据被篡改或泄露、提供安全数据访问环境的目的，有助于实现可信赖的数据安全存储。

4.5 重视数据所有者权益

确认并保护数据所有者权益是长期保存活动可持续进行的重要前提。数据管理层面，保存方通过发布时间限制手段进一步完善人口健康科学数据接收、摄入、保存、管理等内容，保护数据创建者及所有者权益，保证数据集完整性；用户管理层面，针对数据特征及所有者授权访问的用户及数据范围，通常采用用户认证方式对数据访问权限进行控制，将设定好的密钥分配给不同用户，针对密钥链进行集中管理和维护。同时积极制定人口健康科学数据所有者权益保护方案，结合数据实际保存和共享情况，考虑国家政策及行业认可与接受的数据管理标准，更好地实现数据合理合法的保存和获取。

4.6 开展数据文件格式维护

随着科技发展进步，数据文件本身格式类型、数据生成读取依托的软件以及支撑软件的操作系统均存在过时甚至被淘汰风险。人口健康科学数据长期保存系统引入PRONOM格式登记系统[20]、DROID格式识别工具以及PUID唯一标识符，其中采用格式工具DROID识别并验证所提交的科学数据

文件的格式，识别的格式被记录为 PRONOM 认证格式，未标识的格式被记录为未经认证格式，两类来源格式均被统一存入人口健康科学数据长期保存系统的本地格式库，与对应数据文件相关联。同时基于 PRONOM 系统定期发布的格式更新文档，为人口健康科学数据提供长期保存格式风险提醒和迁移推荐服务，保证数据长期可用性。

4.7 使用可持续保存技术

软件方面，开源技术和软件能够更为灵活地实现长期保存系统所需功能，避免引发软件版权纠纷问题。人口健康科学数据长期保存系统采用 Fedora[21] 作为管理和共享科学数据的开源仓储库，其最新版本 Fedora 6 引入基于增量版本控制的 OCFL 文件布局标准，用于优化长期保存文件层次结构及共享的方法，实现独立于应用程序的数字对象存储方法，保证科学数据存储、迭代，从保存、维护、迁移、备份等方面促进对数字文件的访问及管理。硬件方面，为提升数据保存和恢复能力，使用磁盘、磁带等稳定性较高的存储介质进行数据存档及异地备份，定期检查备份数据质量，避免单点故障，满足多种保存需求下的数据迁移活动，灵活提供数据访问服务，提升数据存储和共享能力。

5 结语

数据驱动时代下，对于具有较高科研和应用价值的人口健康科学数据而言，长期保存是一项重要而又艰巨的任务。本文面向人口健康科学数据长期保存实践需求，从保存策略、技术措施两方面，数据遴选、接收限制、保存管理、评估审计、开放获取、数据分类、格式推荐、命名规范、版本控制、数据备份 10 个维度，总结梳理 5 所国外高校科学数据长期保存规划设计与建设经验，在此基础上，从制定元数据规范、创建数据文件命名规范、配置可信赖存储环境、关注数据开放获取权益、开展数据文件格式维护、使用可持续保存技术 6 个方面开展系统设计和建设工作，支撑人口健康科学数据长期保存和管理。

参考文献

1 胡佳慧，杨晨柳，方安，等. 医学数字资源长期保存实践与分析［J］. 中国数字医学，2019，14（12）：95－98.

2 国务院办公厅. 国务院办公厅关于印发科学数据管理办法的通知［EB/OL］.［2021－07－20］. http://www.gov.cn/zhengce/content/2018－04/02/content_5279272.htm.

3 国家互联网信息办公室. 数据安全管理办法（征求意见稿）［EB/OL］.［2021－07－20］. http://www.gov.cn/xinwen/2019/05/28/content_5395524.htm.

4 全国人民代表大会. 中华人民共和国数据安全法［EB/OL］.［2021－06－10］. http://www.npc.gov.cn/npc/c30834/202106/7c9af12f51334a73b56d7938f99a788a.shtml.

5 Libraries Digital Conservancy. DRUM Polices and Terms of Use［EB/OL］.［2020－10－19］. https://conservancy.umn.edu/pages/drum/policies/.

6 Stanford Libraries. SDR Overview［EB/OL］.［2020－10－19］. https://library.stanford.edu/research/stanford－digital－repository/sdr－overview.

7 Australian Data Archive. ADA Data Deposit: study description form［EB/OL］.［2020－10－19］. https://ada.edu.au/wp－content/uploads/2018/03/ADA_Deposit_Form_v1－0.pdf.

8 University of Minnesota Libraries. Data Management Plans（DMPs）［EB/OL］.［2020－10－19］. https://www.lib.umn.edu/datamanagement/dmp.

9 Deankin University Library. Why Should I Use University Services to Stone My Data?［EB/OL］.［2020－10－19］. https://www.deakin.edu.au/library/research/manage－data/store/why－use－university－data－storage.

10 Deankin University Library. Collections［EB/OL］.［2020－10－21］. https://www.deakin.edu.au/library/collections.

11 Oregon State University Libraries. User Guide［EB/OL］.［2020－10－21］. https://guides.library.oregonstate.edu/Scholars－Archive/Datasets.

12 Australian National University. The Australian Data Archive［EB/OL］.［2020－10－21］. https://ada.edu.au/.

13 Stanford Libraries. Stanford Digital Repository［EB/OL］.［2020－10－21］. https://library.stanford.edu/research/stanford－digital－repository.

（下转第 51 页）

‘位于’‘舌苔’)，(‘腻’‘位于’‘舌苔’)，(‘淡’‘位于’‘舌质’)，(‘色红’‘位于’‘口唇’)，(‘3～5 天一行’‘位于’‘大便’)，(‘沉’‘位于’‘脉’)，(‘细’‘位于’，‘脉’)，(‘弱’‘位于’‘脉’)，(‘弦’‘位于’‘脉左’)，(‘稍’‘描述’‘弦’)，(‘多’‘位于’‘汗’)]。重组三元组从而获取实际的症状抽取结果为：“手肿”“手胀”“手无力”“盗汗”“舌苔薄”“舌苔腻”“舌苔黄”“舌质淡”“口唇色红”“大便 3～5 天一行”“脉沉”“脉细”“脉弱”“脉左稍弦”。

6 结语

本文提出一种基于三元组抽取策略的高血压医疗实体提取模型，有效解决传统 NER 无法解决的中医实体识别中出现的实体离散问题。实验发现基于大型中医临床语料库训练出的针对中医特定场景的 BERT_ TCM，与常规中文语料库训练出的 BERT - base - chinese 相比，在中医高血压病历关系抽取任务中具有更好的性能。与仅进行单一关系抽取的 BiGRU 模型相比，联合抽取模型显著提高了各项性能指标。可能的原因是联合抽取同时提取实体和关系，避免实体识别任务中 CRF 层语义信息丢失。CASREL 模型性能比联合抽取模型更加优越，在使用相同预模型 BERT_ TCM 且未添加对抗情况下，CASREL 模型的 *F*1 值远超联合抽取模型。此外引入对抗训练技术能够有效提升模型鲁棒性。

参考文献

1 Zheng S, Wang F, Bao H, et al. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme [C]. Vancouver: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017.

2 Giannis B, Johannes D, Thomas D, et al. Joint Entity Recognition and Relation Extraction as a Multi - head Selection Problem [J]. Expert Systems with Application, 2018 (114): 34 - 45.

3 Wei Z, Su J, Wang Y, et al. A Novel Hierarchical Binary Tagging Framework for Relational Triple Extraction [EB/OL]. [2020 - 06 - 22]. https://arxiv.org/abs/1909.03227v2.

4 Qin C, Martens J, Gowal S, et al. Adversarial Robustness through Local Linearization [C]. Vancouver: Proceedings of the 33rd Annual Conference on Neural Information Processing Systems, 2019.

5 Devlin J, Chang M W, Lee K, et al. BERT: Pre - training of Deep Bidirectional Transformers for Language Understanding [C]. Minneapolis: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019.

6 Yi R, Hu W. Pre - trained BERT - GRU Model for Relation Extraction [C]. Beijing: Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition, 2019.

(上接第 16 页)

14 Library of Congress. Recommended Formats Statement 2020 - 2021[EB/OL].[2020-10-19]. https://www.loc.gov/preservation/resources/rfs/RFS%202020 - 2021.pdf.

15 ANDS Guide. File Formats [EB/OL]. [2020 - 10 - 19]. https://www.ands.org.au/_ _data/assets/pdf_file/0003/731775/File - Formats.pdf.

16 Libraries Digital Conservancy. About the Data Repository [EB/OL].[2020 -10 -21]. https://conservancy.umn.edu/pages/drum/.

17 University of Leicester. Good Practice and Guidance - document Version Control Chart (Draft) [EB/OL]. [2020 - 10 - 19]. https://www2.le.ac.uk/services/research - data/old - 2019 - 12 - 11/documents/UoL_ VersionControlChart_ d0 - 1.pdf.

18 Deankin University Library. Where Should I Store My Digital Data?[EB/OL]. [2020 - 10 - 21]. https://www.deakin.edu.au/library/research/manage - data/store/where - should - i - store - my - digital - data.

19 中华人民共和国国家质量监督检验检疫总局，中国国家标准化管理委员会. GB/T32843 - 2016 科技资源标识 [EB/OL]. [2021 - 07 - 20]. https://wenku.baidu.com/view/77931fbd7dd5360cba1aa8114431b90d6c8589ae.html.

20 钱毅. 基于长期保存视角的电子档案格式管理研究 [J]. 档案学通讯，2016，4 (6)：52 - 57.

21 DuraSpace Organization. Fedora Commons Repository Developer Documentation [EB/OL]. [2021 - 07 - 20]. http://docs.fcrepo.org/.