

2012 年第 5 期（总第 15 期）

长期保存跟踪扫描

主办单位：中国科学院国家科学图书馆

2012 年 5 月

为传播科学知识，促进业界交流，
特编译《长期保存跟踪扫描》，仅供个人
学习、研究使用。

目 录

【信息扫描】	3
网络研讨会系列开放注册：在个人数字知识库中管理和保存音频与视频	3
美国国家人文基金会保存访问部开设网站	4
简单而经济的云端保存和归档的特点和定价	5
2012 年数字音像保存国际会议——“播放，暂停和向前推进”	6
【动态追踪】	6
Chronicle 项目：二十世纪六七十年代 BBC 的北爱尔兰电视新闻	6
云预报：JISC Curation 在云研讨会上的报告	7
Chronopolis 赢得“值得信赖的数字仓储库”称号	9
DCC 工具和服务目录	10
数字资源长期保存管理工作组成立十周年纪	11
NDSA 发布其网络存档调查结果	12
手机短信存档	15
2012 年 ALA 保存周	16
《技术与工具》	18
新的开源视频数字化工具	18
知识库与内容管理系统的区别	19
数字保存中的 XML Schema	20
《专题报道》	25
个人数字存档专题报道	25
【资料推荐】	36
《内存块：如何保持历史资料的“鲜活”》	36
NDIIPP 的报告、白皮书和文章	36

【信息扫描】**网络研讨会系列开放注册：在个人数字知识库中管理和保存音频与视频**

2012 第二届热点问题系列网络研讨会——DuraSpace 社区研讨会系列正在进行。

该研讨会题为：在个人数字知识库中管理和保存音频与视频。就像知识库管理者们说的那样，管理、保存丰富的数字媒体音频和视频文件并创建对其的访问给管理者、内容创造者和用户都带来了一系列复杂的挑战。系列策划人 Karen Cariani 和来自 WGBH、美国印第安纳大学、摇滚名人堂 (The Rock and Roll Hall of Fame)、AVPS 和斯坦福大学的专家们利用三个网络研讨会向与会者介绍在个人机构知识库中保存音频和视频数字媒体的新策略、最佳实践和学术应用。这三个网络研讨会分别是：

- **概述：保存音频和视频数字媒体。**

Karen Cariani 和其他 WGBH 成员对数字音频和视频的保存做了一个概述，跟与会者分享了 workflow、挑战、最佳实践、具体问题和不同视频格式的复杂性等内容。已于 2012 年 5 月 16 日结束。

- **创建对音频和视频数字媒体的访问：视频项目和摇滚名人堂的变更 (Variations on Video project and the Rock and Roll Hall)。**

该研讨会将于 2012 年 6 月 6 日召开，Karen Cariani 和来自“视频项目和摇滚名人堂的变更”的代表将向与会者介绍在机构知识库中保存音频和视频数字媒体的新策略、最佳实践和学术应用。他们将与大家分享他们在访问、视频问题和保存保管方面的经验，并讨论与视频访问和知识库相关的新的开源初创计划。注册地址：

<http://events.r20.constantcontact.com/register/event?oeidk=a07e5rc9i4ib0f7902a&llr=5iy95gcab>

- **用元数据描述音频和视频数字媒体。**

该研讨会将于 6 月 13 日召开，Karen Cariani 和来自 AVPS 与斯坦福大学的代表们将对元数据和格式进行概述，包括视频元数据的复杂性、可用格式和这些复杂性所带来的挑战。

注册地址：

<http://events.r20.constantcontact.com/register/event?oeidk=a07e5rc9jzq407a7c12&llr=5iy95gcab>

编译自：

<http://duraspace.org/webinar-series-registration-open-managing-and-preserving-audio-and-video-your-digital-repository>

(么媛媛编译, 李红培 吴振新校对)

美国国家人文基金会保存访问部开设网站

美国国家人文基金会保存访问部宣布推出其完全重新设计的NEH网站(<http://www.neh.gov>)。除了能更容易申请和管理补助金以外,新网站还展示了特色项目、NEH人员新闻、可访问的NEH资助的在线内容。

保存访问部的网址是:<http://www.neh.gov/divisions/preservation>。该网站突出显示了所有该部门资助项目的新闻和活动:“人文馆藏和参考资料”,“对较小机构的保存援助赠款”,“可持续文化遗产馆藏”,“保存和访问的研究与发展”、“教育和培训”,“记录濒危语言”,和“国家数字报纸计划”,以及“记载美国”(网址:<http://www.chroniclingamerica.org>)。该网站由美国国家人文基金会和美国国会图书馆共同主办。

网站上有四个获奖项目的专题报道——“弗吉尼亚在线百科全书项目”,“弗里克美术收藏馆的数字图片档案馆项目”,“南方信使数字存档项目”,“沃尔特斯艺术博物馆的中世纪手稿数字化项目”。有“记载美国”中记录的NEH旧址——旧邮政总局大楼的趣味历史故事,还有最新更新的即将推出的资助项目的截止日期(保存援助赠款和研究与发展)以及计划人员提供的未来巡回会议活动的一些细节。

保存和访问页面上不仅有近期新闻和事件的目录,还将通过研究最新的工具和最佳实践、数字资源长期保存领域的进展、复杂保存问题的创新可持续发展解决方案、新的培训方案、保存音像和原生数字馆藏,以及从小型历史学会和博物馆到研究型大学档案馆和图书馆的创造性合作等,探讨保存社区的趋势和发展。

该网站上的内容将定期更新,可以通过Twitter(@NEH_PresAccess)来跟踪其最新的功能、项目、资金公告和新闻通知。

编译自:

<http://duraspace.org/national-endowment-humanities-division-preservation-and-access-launches-web-site>

(么媛媛编译, 李红培 吴振新校对)

简单而经济的云端保存和归档的特点和定价

DuraSpace推出了新的DuraCloud (<http://DuraCloud.org>)服务方案,提供三个层次的经济型云端保存和归档服务。新的服务方案的收费标准与商业云存储提供商相比很有竞争优势,并且不需要额外的转换或者增加可变成本。

三种服务方案提供了多种保存措施和企业应用案例,欲定制DuraCloud托管服务的用户可以从中选择适合的一种。

- DuraCloud 基本保存方案非常适合那些想在一个云数据中心中将原始内容保存为两个冗余副本的机构。
- 如果想要有更多的保护, DuraCloud 保存叠加方案尤其适合那些想在两个云数据中心将原始内容保存为四个冗余副本的机构。
- DuraCloud 企业级方案将提供一套完整的 DuraCloud 特色服务,是那些需要为部门、研究小组、跨机构项目或个体设置多个子账户的机构的理想选择。

为了确保对不可替代内容的长期保存, DuraCloud 每周自动对内容进行健康检测、报告、文件修复等。更多的信息,包括各个方案的完整特点介绍,参见: <http://duracloud.org/pricing>。

DuraCloud 特点一览

对于学术机构、博物馆、政府部门、专业组织的研究成果和数字图书馆资料的归档保存来说, DuraCloud 是唯一的将内容存档于多个云提供商的托管式软件服务,所有的文档、图像和视频资料都是可以随时访问和获取的。DuraCloud 运行于 Amazon AWS 和 RackSpace 上,不久还将在美国最大的学术云存储系统——SDSC 云上运行。DuraCloud 为用户提供了以下服务:

- 内容备份并存储在多个供应商处
- 为内容进行自动的健康检查,所以你的文件永远不会毁坏或丢失
- 为选择保存叠加方案的客户提供对已损坏文件的自动修复
- 一套完整的报告帮你更有效地管理你的内容
- 在线共享并发送到任何与互联网连接的设备中

编译自: <http://duraspace.org/new-duracloud-digital-archiving-and-preservation-services>

(李红培编译, 齐燕 吴振新校对)

2012年数字音像保存国际会议——“播放，暂停和向前推进”

PrestoCentre 基金会与南加州大学 Shoah 基金会研究所及数字知识库计划联合举办 2012 年数字音像保存国际会议。5 月 21 日至 23 日，来自世界各地的 250 多位与会者齐聚坐落在洛杉矶的南加州大学 (USC) 罗纳德教导大厅，共同讨论和探析数字音像资料保存的发展现状。大会召集档案保管员、制作公司、电影制片人、电视节目制作人、首席技术官、科学家、供应商、战略家、投资者、政策制定者等各界人士，来共同研究如何应对音像资料保存所面临的最紧迫的问题。所有与会者协助相关领导机构制定出了音像资料保存的未来研究和政策议程，并从中受益。

“本年度的会议主题是‘播放、暂停和向前推进’，会议着眼于面向越来越多的视听内容的保存和共享的现状、未来前景和战略，” PrestoCentre 基金会主席 Jan Müller 说到，“数字存档是链接我们的过去、现在和未来的桥梁，本次讨论的议题涵盖但不限于相关技术、标准、用户期望、存档成本，以及日益复杂的组织、技术和融资模式之间的交互等。”

“所有这一领域的工作人员都有责任继续开拓创新数字存档的技术和工艺，为保存人类遗产做出应有的贡献，” USC Shoah 基金会研究所首席技术官兼 USC 图书馆学院副院长 Sam Gustma 说到，“我们受到了当前最大的公共视频数据库之一的委托对其音像资料进行长期保存，我们必须努力保证所存储的档案文件，当然也要包括世界各地的其它所有档案资料，即使科技剧烈演变，也能保持其可获取、可理解、持久生存、完整、真实等性能，这是非常重要的。”

编译自：<http://www.prestocentre.org/events/screeningfuture/2012>

(齐燕编译，李红培 吴振新校对)

【动态追踪】

Chronicle 项目：二十世纪六七十年代 BBC 的北爱尔兰电视新闻

有着合作伙伴关系的 BBC、英国大学电影和电视局 (BUFVC: British Universities Film and Video Council) 与 JISC 于 2012 年 4 月 19 日共同宣布推出 Chronicle 项目，使整个学术界可以在线获取二十世纪六七十年代的 BBC 北爱尔兰电视新闻。

BBC 的视听档案包括从二十世纪二十年代 BBC 创立以来收集的大量资料，但其中大部分仍是不可访问的，这些资料以电影或录像带的形式保存，用以满足 BBC 节目制作人的需

求。Chronicle 项目对 BBC 北爱尔兰库存中的新闻和时事节目进行数字化, 将这些档案的一部分提供给英国的高等教育和继续教育领域使用。这样就从政治、历史和文化角度提供了一个丰富的情景体验, 教师、学生和研究人员就有机会研究北爱尔兰历史的一个重要时期(1963年至1976年)的事件。

BBC 和 JISC 进行了数年的非正式合作, 现在这两个组织开始了新的合作伙伴关系, 签署了一份理解备忘录 (MOU: Memorandum of Understanding), 以促进这两个组织实现他们的共同利益——鼓励创建和使用广泛的视听和文化资产来支持教育、研究和公共使用。

Chronicle 将为经过身份验证的用户提供包括北爱尔兰和“北爱尔兰问题”在内的数字化的新闻和时事材料副本, 同时还提供基于 Web 的工具, 让用户可以进行搜索、浏览和注释。开放对这些资料的访问同时也意味着促进了教育界对“北爱尔兰问题”的理解和赞赏, 也就有机会用创新的技术建立新的研究途径。BBC 将致力于尽可能有效地利用其巨大的存档, 寻求利用档案资料来为学术界和 BBC 更广泛的公共目的服务的方法, 该项目在这方面迈出了重要的一步。

该项目的其他作用也很重要。除了一些组织和学者需要上述资料以外, 从存档需求的角度上来讲, 1963年至1981年间的新闻资料也需要数字化, 这些资料记录在16mm厚的胶片上, 有很大一部分需要保存, 因为1981年之后的资料都被记录在 Beta 或 Digibeta 的磁带上, 恶化的风险小得多。

编译自: <http://www.jisc.ac.uk/news/stories/2012/04/BBC.aspx>

(么媛媛编译, 李红培 吴振新校对)

云预报: JISC Curation 在云研讨会上的报告

2012年3月7-8日, 伦敦举行了云研讨会, 在 JISC Curation 讨论期间, 与会者考虑了云计算在保存任务和策略中所能发挥的潜在作用和效益。该研讨会的重点是学习云服务能够满足保存的什么目的, 讨论怎样在云中生成和分析数据。

草案白皮书《云中的数字保存》探讨了云计算的服务模式、供应模式和利用云计算的一般利益及风险。David Wallom 指出计算和比较云成本以及法律责任的困难是其前进的障碍。从云供应商那里得到的标准化的统计信息不能使机构更容易地衡量成本和效益, 如何满足法律义务仍存在问题。草案白皮书中认为数据或服务水平的损失、法律与管理的不相容、传输

瓶颈是在数字保存任务中利用云的规模化效益和资源快速供应中的三大风险,并进一步阐述了采用云技术涉及到的法律问题。

互联网档案馆 (IA: Internet Archive) 的 Kris Carpenter 预计未来硬件会被完全外包出去,而现在的做法主要是保持对内容的控制和实施混合云存储模型。现在超过 7PB 的 IA 数据可以被公开访问,另外有几个 PB 的数据不公开。IA 在云中存储衍生数据,而原始数据大多存储在本地。

Kindura 项目 (http://kindura.cerch.kcl.ac.uk/http://kindura.cerch.kcl.ac.uk/?page_id=2) 可以让研究人员将其研究成果保存数年。在亚马逊、Rackspace 公司和 iRODS 中, DuraCloud 作为混合云配置,使用了一个自定义的接口,并且这些公司提出了一个与大范围的研究人员进行交互的解决方案,包括一个基于跟踪和用户输入的成本计算引擎。Kindura 项目的一个主要目标是提出一个基于内容需求和使用配置文件来对底层的云存储决议进行有效和半自动化管理的方法。

DuraSpace 的技术领导和 DuraCloud (<http://DuraCloud.org>) 的开发者 Andrew Woods 对既是托管服务又是开源软件的 DuraCloud 做了一个综述。综述的重点是三年来实施基于云的长期保存平台的经验教训。

在分组座谈会上,与会者对云保存进行了一个预想,包括在成本方面变得更加透明的愿望、SLA 和能源效率、互操作性和可移植性的标准、打破“孤岛思维”、整合云选项的 workflows、供应商提供“云机柜”、云服务的认证、云服务的第三方审计以及高配置案例研究的更大的可用性。

DuraSpace CSO 的 Jonathan Markow 进行了案例研究,关于非盈利的开源项目怎样重点关注通过强调影响和交流以及社区能力建设活动和计划来维持运作。

当采用 DuraCloud 和 EDUSERV (<http://www.eduserv.org.uk/>) 提供的云解决方案来增加机构的技术堆栈云能力时,架构建设团体也开始研究云。研究人员正在考虑利用云的科学 workflows,机构也正在考虑如果云供应商没有机构本身存在的时间长,他们如何进行数字资源长期保存。

编译自: <http://duraspace.org/duratest/cloud-forecast-report-jisc-curation-cloud-workshop>

(么媛媛编译,李红培 吴振新校对)

Chronopolis 赢得“值得信赖的数字仓储库”称号

Chronopolis, 一个大型的数据保存网络, 被研究型图书馆中心(CRL)认证为“值得信赖的数字仓储库”, 满足了数字仓库管理的最佳实践要求。Chronopolis由加州大学圣迭戈分校图书馆和圣地亚哥超级计算机中心(SDSC)领导, 在国家大气研究中心和马里兰大学高级计算机研究所有合作伙伴。

CRL在评估中使用的主要指标摘自可信赖仓储库审核和认证(TRAC)标准清单, 该清单是由美国研究型图书馆团体与美国国家档案和记录管理局组成的联合工作小组在2003年开发的, 旨在制定明确的数字存储库评估标准。TRAC标准, 包括组织的基础设施、数字对象管理、技术与工艺的基础设施、安全, 代表了当前值得信赖的数字仓储库的组织和技术需求的最佳实践和思考。TRAC标准涉及到组织架构、数字对象管理、技术及其基础设施和安全等方面, 描述了当前的最佳实践, 并考虑到了值得信赖的数字仓储库所需具备的组织和技术条件。

TRAC认证是数字资源长期保存领域的‘好管家认可印章’, 目前已有Portico 和HathiTrust获得这个信任的称号。为了收集和保存处于危险中的数字信息, 美国国会图书馆于2008年设立了Chronopolis项目, 它能够保存几百TB的任何类型或者大小的数字数据, 满足数据提供者的最低要求。该项目充分利用高速网络、大规模存储能力和合作伙伴(包括加州圣迭戈图书馆和SDSC, 国家大气研究中心和马里兰大学高级计算机研究所)的专业知识, 提供地理上分布式、异构的、高度冗余的归档系统。

认证是基于CRL的审查, 包括审查大量的证明文件, 有的是CRL认证顾问小组自行从开放的第三方来源收集到的文件, 有的是Chronopolis管理层提供的数据和文件。审查还包括CRL的审核人员对SDSC的现场视察。认证顾问小组得出的结论是, Chronopolis的实践和提供的服务是合理的, 适于存档内容, 并能够满足指定Chronopolis服务团体的基本需要。

Chronopolis还是加州大学圣迭戈分校的校园RCI(Research Cyberinfrastructure)的数据存储仓库, 该RCI旨在为研究人员提供存储、管理和共享数据所需要的计算、网络和人力资源。

编译自: http://www.sdsc.edu/News%20Items/PR040912_chronopolis.html

(李红培编译, 齐燕 吴振新校对)

DCC 工具和服务目录

DCC创建了一个工具和服务目录,来帮助人们了解、评估并从中选择能够直接用于其保存和管理任务的工具和服务。

该目录现在有近60个条目,每一个条目都有详细的描述,并有一个汇总表,能够帮助人们对具有相似功能的几个工具的成本收益进行比较。除了绝大多数人关注的免费和开源资源,该目录还提供了一些订购服务供需要外包服务的机构选用。同时,该目录也会定期更新和扩展。

根据预期用户和工具在数据生命周期中最有使用价值的阶段,DCC将这些资源分成五个大类:

- managing-active-research-data 主动管理研究数据目录

<http://www.dcc.ac.uk/resources/external/tools-services/managing-active-research-data>

对于研究人员和其他数据生产者来说,该目录涵盖了项目计划阶段使用到的资源,关注诸如存储、工作流程、ID 分配和数据管理计划等问题。

- sharing-output-and-tracking-impact 成果分享和影响力追踪目录

<http://www.dcc.ac.uk/resources/external/tools-services/sharing-output-and-tracking-impact>

展示了一些新的工具和服务,能够帮助研究人员发现新资源、挖掘与其自身领域的联系,并超越传统的引用指标跟踪其成果的影响。

- depositing-and-ingesting-digital-objects 存缴和摄入数字对象

<http://www.dcc.ac.uk/resources/external/tools-services/depositing-and-ingesting-digital-objects>

对于数据图书馆员和管理者来说,该目录展示的工具和服务能够帮助他们将新内容顺利地纳入原有数字仓库中,包括从馆藏建设到数据迁移,再到登记入册等流程。

- archiving-and-preserving-information 存档和保存信息包目录

<http://www.dcc.ac.uk/resources/external/tools-services/archiving-and-preserving-information>

面向数字资源生命周期的存档阶段,介绍的工具主要关注在执行可靠的存档与保存策略以及向其指定用户提供访问的过程中所产生的技术问题。最后,

- managing-and-administering-repositories 仓储库管理和行政问题目

<http://www.dcc.ac.uk/resources/external/tools-services/managing-and-administering-repositories>

提供的资源用于帮助保证数字仓储库的组织和财务健康运行,包括评估目前的状态、制定未来的战略和计划,发布相关政策和协议的公文以保证其得以实施等。

编译自: <http://www.dcc.ac.uk/news/reviewed-tools-and-services-catalogue>

(李红培编译, 齐燕 吴振新校对)

数字资源长期保存管理工作组成立十周年纪

到 2013 年, 数字资源长期保存管理(DPM)工作组就成立十周年了, 目前各项庆祝活动都已在规划中。DPM 工作组由美国康奈尔大学图书馆在 2003 年成立, 并于 2007 年移交给校际政治与社会研究联盟(ICPSR)管理至今。2012 年夏季之后, 将转由麻省理工学院继续接管。2003 年发布的在线指南(<http://dpworkshop.org>)不仅是对数字资源长期保存的介绍, 也是该工作组的运行基石。该指南目前已有三种语言版本(英语、法语和意大利语), 将来会更多。

迄今为止, 已有 1,000 多位相关人员参与到 DPM 工作组的 40 多个研讨活动中, 他们分别来自超过 35 个国家的图书馆、档案馆、博物馆等 300 多个组织机构。除了跨越美国大陆的多个地区, 研讨会也在以下 5 个国家中各举办过至少一次: 加拿大、澳大利亚、新西兰、南非和德国。DPM 工作组构筑的数字资源长期保存矩阵模型(包括一个可持续发展计划的三大支撑和五个发展阶段)和其他一些核心概念都已经被数字资源长期保存培训计划(Digital Preservation Training Programme, DPTP) <http://www.dptp.org/>、DigCCurr 研究所和南非数字馆藏培训分支项目 <http://digi.nrf.ac.za/publ/Managing%20Digital%20Collections.pdf> 所采用, 并为美国国会图书馆开发的数字保存推广和教育(Digital Preservation Outreach and Education, DPOE) <http://www.digitalpreservation.gov/education/>基础培训模块项目提供了重要素材。

为了纪念即将接手的这一工作组成功运行 10 周年, 麻省理工学院将会针对参与者的工作成果进行调查, 对原有指南进行更新和完善, 开发一个网络空间存储工作组的相关资料, 发布有关数字保存矩阵模型的涵义和应用的白皮书。并有希望举办一个工作会议, 来回顾十年来长期保存管理在培训和实践上的进展, 并展望未来十年间工作组和相关培训项目的发展蓝图。工作会议中他们会做好会议记录, 并出版相关论文集, 这也是反映矩阵模型的组织实践的重要实例。

编译自: <http://www.dpconline.org/newsroom/whats-new/842-whats-new-issue-44-april-2012>

(齐燕编译, 李红培 吴振新校对)

NDSA 发布其网络存档调查结果

美国国家数字管理联盟(National Digital Stewardship Alliance, NDSA)的五个工作组之一——NDSA 内容工作组,致力于对已存储的内容进行标识,为遴选重要内容制定指导方针,发现处于危险中的数字内容或资源集合,并为其合作伙伴寻找所需的孤品内容并进行保存和提供获取。作为这项使命的一部分,在2011年10月3日至31日期间,该工作组针对全美国正在积极开展或计划启动网络内容存档项目的所有组织机构展开一项调查,目的是通过了解涉及到的组织、保存内容的类型、所采用的工具和策略,以及所提供的获取方式等方面来更好地掌握美国当前网络存档活动的概况。

此处仅为调查初步结果,国际互联网保存联盟([International Internet Preservation Consortium](#), IIPC)在美国国会图书馆举办的2012年全体大会上也会发布相同内容。报告全文将很快面世。

该调查精选了28个问题,得到了77个有效回应,涵盖了各类型组织,主要包括22家文化遗产机构(29%, 22/77)、17家政府部门(22%, 17/77)和36家高校团体(46%, 36/77)。在这些受访机构中,31%(24/77)是NDSA成员,8%(6/77)是IIPC的成员。

网络存档活动

受访机构开展网络存档活动的现状如下:

- 63%(49/77)的机构已经有一个网络存档项目在运行中;
- 16%(12/77)的机构正在积极地测试一个网络存档项目;
- 17%(13/77)的机构计划在不久的将来启动一个网络存档项目;
- 4%(3/77)的机构以前开展过网络存档的项目,但没有继续做下去。

有趣的是,在71家确定了自己的网络存档目标的受访机构中,有80%(57/71)的机构是对来自其它组织机构或个人的支撑未来研究的内容进行存档,69%(49/71)的机构是存档自身机构的网络内容,49%(35/71)的机构则是同时涵盖上述两类内容。

在审查完整的调查结果时,人们发现了以下几点值得注意的问题:

1、近几年网络存档活动兴盛,尤其是在学术机构领域

一个令人吃惊的结果是,在过去5年间启动网络存档项目的机构中,高校占了绝对多数。在68家能够确定其开始进行网络存档活动年份的受访机构中,近1/3(32%, 22/68)的机构是在近两年开始他们的网络存档项目的,这个数字与在1989至2006年的17年间开展网络内容存档的机构数目(22, 32%)是一样的。而从2006年至2011年的5年间,开展网络

存档活动的机构犹如雨后春笋——占总受访机构的 68% (46/68)，其主要原因就是很多高校启动了网络存档项目。

2、不协调的保管工作

调查还发现, 已经将存档数据从其外部存档服务方转入自身机构内部存储系统的机构所占比例很低。只有 18% (9/49) 的受访成员机构已经将数据转为内部存储, 这其中包括 12 家受访政府部门中的 2 家和 15 家受访高校中的 4 家。在所有使用外部存档服务的机构中, 共计有 82% 的机构没有将存档数据转为内部存储。从针对这一问题的开放性回答上可以看出, 在将外部收割的数据转存到内部系统时, 机构会有“复制成本”、缺乏相关基础设施、机密性、编目和可获取性等方面的担忧和顾虑。

3、相关政策缺失以及许可获取指导不明确

对于许多机构来说, 内部政策文件仿佛是最需要持续改善的地方。某些项目已经将网络资源纳入现有政策和工作规程中, 其它则还没有, 而其中有一些好像还未确定其机构针对网络内容的当前政策导向。

调查还发现, 不论是在内容的收集还是提供获取上, 当前都严重缺乏向内容创建者寻求授权许可的明确指导。下面两张图分别展示了从内容创建者那里寻求内容收割的授权许可状况和寻求对存档网站提供获取的授权许可相关的一些策略状况:

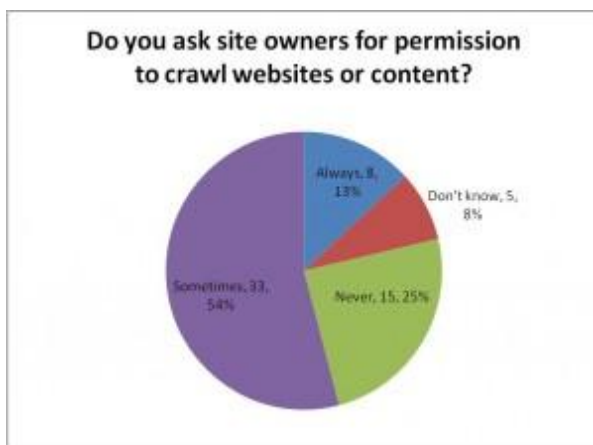


图 1 向网站所有者提出爬行收割其内容的许可申请的策略现状

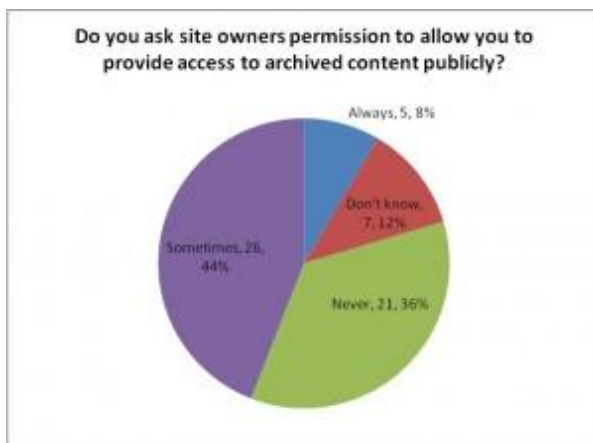


图2 寻求对存档网络内容提供获取的授权许可的相关策略现状

采集的趋势及合作的潜力

被采集的内容类型包括网站、博客和社交媒体:

- 78% (60/77) 的机构已经或计划将网站纳入存档内容类型范围中。
- 57% (44/77) 的机构已经或计划将博客纳入存档内容类型范围中。
- 38% (29/77) 的机构已经或计划将社交媒体纳入存档内容类型范围中。

从对于“简要描述贵机构网络存档资源集合的范围”问题的回答可以总结出, 主要涵盖以下三类: 机构自身文档、原有收藏更新以及围绕一定主题的收集等。

合作的潜力也是调查结果中值得注意的一方面。目前, 只有 23% 的机构在网络存档上有合作行为。但是当问及是否有兴趣参与未来的协作式存档活动时, 有 96% (64/67) 的机构回答“是”(34, 51%) 或者“可能有”(30, 45%)。从这些数字可以看出, 人们对于联合网络存档的合作机会还是很感兴趣的, 只是当前还未有大行动。

虽然调查显示, 在保存网络原生内容以及与提供对存档内容的获取有关的授权许可和管理上还会继续面临各种挑战, 但同时调查也指出了, 对于许多机构来说, 网络存档对其馆藏发展发挥着越来越重要的核心作用。这一点, 再加上越来越开放的合作机会, 相信经过整个领域的共同努力, 在报告中列出的这些挑战都会在适当的时候逐一被攻破。对于那些致力于存档和保存网络资源的组织机构来说, 像 IIPC 全体大会这样的活动以及 NDSA 这样的联盟组织, 是知识共享和协同合作至关重要的纽带。

编译自:

<http://blogs.loc.gov/digitalpreservation/2012/05/web-archiving-arrives-results-from-the-ndsaweb-archiving-survey/>

(齐燕编译, 李红培 吴振新校对)

手机短信存档

手机短信虽然不相连贯但短小精悍,也是一种重要的通信方式。人们重视诸如信件和电子邮件的文本通信信息并实施了很多保存措施,同理,人们也会想要保存一些手机短信。问题是脱离手机,存储短信就不太容易或不太方便了。

手机可以大致分为两种类型:“基本型”和“智能型”。两种类型的手机都可以存储短信、电话簿和其他信息,要么是在一个驱动器上,要么是在一张可拆卸的SIM卡上。如果想要保存文本/SMS消息,必须将它们从手机上卸除然后放到其它存储媒介上。

一些手机短信软件可以帮助人们找回删除的短信。这不是不可能的,因为当人们删除一条短信时,虽然手机提示已经删除,但实际上还会对其保存一段时间并为其留出可用空间。如果手机接收到新的短信,并且需要所谓的被删除的短信所占用的空间,那么新的短信就会覆盖旧的短信。相反,如果新的短信可以找到其它空间进行存储,并不需要“已删除”短信的存储空间,那么被删除的短信就会继续存在,一直到最后终有一些东西覆盖了这些内容,占用了这一空间。这对于文本恢复和文本恢复软件来说是个很大的市场。所以,网络上会有很多面向从事执法、安全、私人调查和其他法律工作的目标群体的商业数据恢复产品。

手机用户要求手机服务提供商提供短信的副本,通常会遭遇一些法律障碍。对于电话文件的获取,美国有一些非常严格的法律法规,包括“2006年消费者电话记录保护法案([2006 Consumer Telephone Records Protection Act](#))”,各个州相关法律的复杂和严厉程度也会各有不同。

长远来看,为短信存档制定计划是更好的方法。如果想要保存短信,就要养成不删除短信的习惯,尤其是一些非常重要的值得珍藏的会谈,并且定期地进行备份。理想的情况是:每次将手机连到电脑上时,相关软件就可以自动地检测到有一些新的短信,并且能够将其自动地导入到电脑进行存档。如果软件能够同时将内容副本上传到云存储备份服务商处,那就更好了。但是,在廉价、易用的自动保存软件出现之前,人们还是需要手动地进行短信保存。

智能手机用户可以查找适合的短信保存程序或软件,而使用基本型手机的用户则需使用相配的SIM卡阅读器。作为一种临时解决方案,人们可以经常将短信发送到自己的电子邮箱,并一条条地单独保存在自己的电脑上。不过,该方法很快就会让人觉得乏味。理想的情况下,技术应该在后台处理好短信的存档细节问题,而不会给人们带来烦扰。

编译自: <http://blogs.loc.gov/digitalpreservation/2012/04/archiving-cell-phone-text-messages/>

(齐燕编译,李红培 吴振新校对)

2012年ALA保存周

美国图书馆协会(American Library Association)保存周活动始于2010年,因为当时在收藏机构中有6.3亿条资料亟需关注和保管。在这些机构中,80%的机构没有负责典藏维护的专职人员,22%的机构则完全没有典藏维护人员。大约26亿条的资料没有应急保护预案。近年来自然灾害频发,人们意识到这些资源都处于危险之中,无论个人、家庭还是集体典藏,风险面前,全体平等。2011年的保存周中,全美国共举办了超过65个活动。今年的保存周活动是在4月22-28日,国会图书馆以及其它图书馆同样举办了很多活动,相关活动和事件的详细信息参见:(ALA's [Preservation Week event map](#)) <http://www.ala.org/alcts/confevents/preswk/cal>。

此次保存周,国会图书馆关注的主要是个人存档方面的问题。个人档案资料的保存和传承越来越受到人们的重视。在当今这个信息技术飞速发展的时代,存储媒介和文件格式没几年就会过时并无法使用。数字信息的这种脆弱性,需要我们采取与之前管理纸质资料完全不同的方法来对其进行保存。藉由保存周的契机,国会图书馆计划在华盛顿地区举办一系列活动,为那些想要保存个人数字资源的人们提供面对面的指导和帮助,主要涉及主题包括:

4月23日——“电影:投资于未来”

4月24日——“保管数码照片:组织和保存数字记忆的策略”

4月24日——“保存数字资料:长期保存数字资料的实用策略”

4月25日——“保管书籍、纸质的文件和艺术作品、影印资料”

4月26日——“长期保存个人数码照片”

活动完整列表及其介绍参见:

<http://www.digitalpreservation.gov/meetings/presweek2012.html>、

<http://www.loc.gov/today/pr/2012/12-069.html>。

对于想要在保存周期间举办个人数字存档活动的机构,国会图书馆提供了相关指南和来源:

<http://www.digitalpreservation.gov/personalarchiving/padKit/index.html>、

<http://www.digitalpreservation.gov/personalarchiving/>。

此次保存周活动中还包括两个网络研讨会:

- 1) “保管家庭纺织品”(<http://www.ala.org/alcts/confevents/upcoming/webinar/pres/042412>)。
- 2) “保存个人数码照片”: 有超过570人参与,共同讨论了如何确定个人数码照片都存放

在了哪里、如何鉴别哪些照片是最重要的以及如何对其进行组织整理。相关会议记录及文件副本参见：<http://www.ala.org/alcts/confevents/upcoming/webinar/pres/042612>。

对于想要在家里学习如何保管个人藏品的人们，国会图书馆提供了一些启蒙资料：

- 1) “美国国会图书馆保存委员会：保管家庭藏品
(<http://www.loc.gov/preservation/family/>)”：让人们了解管理和保护日常家庭藏品的重要性，提供了相关信息以及其它资源的链接。
- 2) “老照片修复的数字化原则
(<https://familysearch.org/techtips/2011/07/beginning-digitization-principles-restoring-old-photographs>)”：来自 Familysearch.org 上的一篇博客，讨论了在扫描老照片时的注意事项。
- 3) “个人数字归档(<http://www.digitalpreservation.gov/personalarchiving/index.html>)”：该站点提供了一些可以长期保存个人的原生数字信息的小建议。

之前的保存周活动中的一些网络广播资料现在可以免费获取了，包括：

- 1) “意外事故偶有发生：保护&保存家庭藏品”
(<http://www.ala.org/alcts/confevents/upcoming/webinar/pres/042611>， Nancy Kraft)
- 2) “立即开始保护未来的访问：文化遗产组织保存数字化图书和其它内容的模型”
(<http://www.ala.org/alcts/confevents/upcoming/webinar/pres/042711>， Amy Kirchhoff)
- 3) “保存个人数字记忆”
(<http://www.ala.org/alcts/confevents/upcoming/webinar/pres/042811>， Bill LeFurgy)
- 4) “存档 101”(<http://www.ala.org/alcts/confevents/upcoming/webinar/pres/051110arch>)
- 5) “霉变的预防和处理”
(<http://www.ala.org/alcts/confevents/upcoming/webinar/pres/051310mold>)
- 6) “防灾准备”(<http://www.ala.org/alcts/confevents/upcoming/webinar/pres/051210disast>， Nancy Kraft)

编译自：

<http://blogs.loc.gov/digitalpreservation/2012/04/preserving-your-personal-digital-photographs-library-of-congress-presents-online-session/>

<http://blogs.loc.gov/digitalpreservation/2012/04/it%e2%80%99s-preservation-week-how-are-you-celebrating-it/>

<http://www.ala.org/alcts/confevents/preswk/cal>

(齐燕编译，么媛媛 吴振新校对)

《技术与工具》

新的开源视频数字化工具

Kate Murray 著 么媛媛 编译

华盛顿地区国家档案馆数字化服务科 (<http://www.archives.gov/preservation/products/>) 近期完成了一个拓宽文化遗产团体格式化视频的技术基础设施选择范围的项目,并推出了其他机构和公众可以使用的最终版本。该项目得到了音像保存解决方案 (AVPS: AudioVisual Preservation Solutions, 网址: <http://www.avpreserve.com/>) 的大力支持,并与联邦机构的数字化准则倡议 (FADGI: Federal Agencies Digitization Guidelines Initiative, 网址: <http://www.digitizationguidelines.gov/>) 共同开发。国家档案馆希望这些工具和应用程序不仅有用,而且能随着文化机构团体的扩大而得到增强。

reVTMD 是一个专门针对文化遗产团体的格式化视频需求的 XML schema。该 schema 复用了几个现有的 schema,包括 videoMD 和 PB core。reVTMD 包含记录捕获历史数据 (包括如采集和回放设备的品牌和型号等关键信息)的解析结构,给该团体增加了价值。reVTMD 地址: <http://www.archives.gov/preservation/products/reVTMD.xsd>。

MediaInfo (<http://mediainfo.sourceforge.net/en>) 支持从多种视频文件格式和封装中提取技术元数据,填充 reVTMD schema,并将数据作为 reVTMD schema 的 XML 文件输出。

AVI MetaEdit 基于 AVPS 和 FADGI 于 2010 年开发的一款很流行的开源 BWF MetaEdit (<http://sourceforge.net/projects/bwfmetaedit/>) 工具的功能而建成。除了 BWF MetaEdit 对 RIFF-based Broadcast WAVE 格式 (BWF) 音频文件提供的功能,AVI MetaEdit 还能支持在 RIFF-based AudioVisual Interleave 格式 (AVI) 视频文件中嵌入和验证元数据。在 NARA 的数字化服务部门中,创建保存者现在会选择 AVI 格式 (<http://www.archives.gov/preservation/products/reformatting/video.html>) 作为保存的主存档。NARA 正与 FADGI 合作完成 AVI MetaEdit 支持的格式化视频的嵌入元数据准则。AVI MetaEdit 现可从 NARA 的 GitHub 网站 (<https://github.com/usnationalarch>) 上获取。

编译自: <http://duraspace.org/new-open-source-video-digitization-tools-national-archives>

(李红培 吴振新校对)

知识库与内容管理系统的区别

虽然机构仓储库 (Institutional Repositories, IR) 和内容管理系统 (Content Management Systems, CMS) 有很多相同的特点, 但是它们的根本目的和所能满足的需求还是有区别的。

简要概括如下:

内容管理系统 (CMS)

- 一般而言, CMS 会被看作是一个数字内容的创建和出版系统
- 面向内容创建/生产和在线出版
- 面向内容的合作创建和修改
- 面向一般性的使用 (任何面向数据内容的一般性使用)
- 面向网站建设和网站内容的创建

机构仓储库 (Institutional Repositories)

- 一般而言, 机构仓储库被看作是一个数字“归档”系统
- 面向长期存储、数字保存和内容完整获取
- 面向确保和维持已完成或已出版内容的起源信息。
- 主要用于学术和/或已出版的内容 (虽然可能也用于一般性内容)
- 趋于跟随最新的图书馆/档案馆的最佳实践 (围绕元数据、长期保存、永久 URL

等)

选择 CMS 或仓储库的依据如下:

CMS 的几个基本优点:

- 更有利于合作创建内容和文档
- 通常更适合高度动态内容 (“活文档”)
- 通常更适合网站建设 (尤其是需要及时更新/演进的网站)

仓储库的几个基本优点:

- 通常会提供更多数字资源长期保存的工具或这类工具的集成 (比如文件格式的识别/校验、完整性检查与杀毒软件的集成等等)
- 通常会为所有的数字内容分配永久的 URL (Handles, DOIs 和/或 URLs) 以确保其长久可访问。
- 趋于跟随最新的图书馆/档案馆的最佳实践, 包括元数据 (比如 Dublin Core、MODS、METS 等)、数字资源长期保存 (OAIS、TRAC、PREMIS 等) 和互操作

能力 (OAI-PMH、 SWORD protocol、 OAI-ORE 等)

- 通常更适合于对已完成或已出版文档的长期保存和获取

机构可以根据自己的需要,在文件的创建和保存工作流程中分别使用不同的系统。例如,CMS 可用于合作创建数字内容,一旦数字内容完成了,就可以归档/保存在仓储库中。另外,CMS 也可以将外部仓储库的存档内容摄入,以支持内容重用/混合,或者仅仅是在相同的界面上动态/协作展示存档的内容。很多 CMS 提供一些模块和插件允许将其直接整合到流行的仓储库平台中。

同时使用仓储库和 CMS 的案例:

- Islandora 是建立在 Drupal CMS 平台之上的,同时将其内容存储在 Fedora 仓储库中
- Drupal 的 DSpace 模块可以把 DSpace 仓储库中的内容和元数据迁移到 Drupal CMS 中
- Joomla 的 DSpace 模块 (J-CAR)可以把 DSpace 仓储库中的内容和元数据转移迁移到 Joomla CMS 中

编译自: http://www.sdsc.edu/News%20Items/PR040912_chronopolis.html

(李红培编译,齐燕 吴振新校对)

数字保存中的 XML Schema

Steve Knight 著 么媛媛 编译

摘要:

在某个早晨,新西兰国家图书馆的一台服务器在一小时内尝试连接一个外部服务器高达 150 万次。原来,这些访问是该图书馆的数字资源长期保存系统尝试验证一些 XML 文件的 XML 结构定义 (XSD: XML Schema Definition) 声明时产生的,这些 XML 文件来自该图书馆正在加载到 Rosetta 的在线报纸应用——Papers Past。这篇文章介绍了新西兰国家图书馆针对这一情况做出的反应,列出了一些找到合适的解决方案之前需要细究的问题,包括这些 XSD 的数字保存状态、它们对诸如 JHOVE 的验证工具的影响和这些数字保存进程中的材料应该存在哪里。

值得关注的初始原因

服务器在这一个小时内尝试连接 217.111.76.181 地址 150 万次,最后被图书馆的本地防火墙阻止了。这个 IP 映射到哪里?为什么服务器要连接它呢?

发生了什么?

馆员们正在测试 Papers Past 的迁移过程,涉及将纳入数字资源长期保存计划中的 61 个

主题资源 (30 万份出版物、420 万份文件, 3.6TB)。这其中包括一些 ALTO XML, 很多都带有 XML 声明:

`xsi:noNamespaceSchemaLocation=http://schema.ccs-gmbh.com/metae/alto-1-2.xsd`

217.111.76.181 这个地址表示的是主机: `schema.ccs-gmbh.com`。为了进行格式验证和元数据提取, 应用程序 (JHOVE) 试图与存储服务器联网, 下载 XSD 并根据相应的 XSD 验证 ALTO。

图书馆花了三天时间对整个 Papers Past 仓储库中的 XML (ALTO 和 METS) 进行分析, 得出了如下 XSD 列表:

`http://schema.ccs-gmbh.com/docworks/alto-1-2.xsd`

`http://schema.ccs-gmbh.com/docworks/mets-metae.xsd`

`http://schema.ccs-gmbh.com/metae/alto-1-2.xsd`

`http://schema.ccs-gmbh.com/metae/mets-metae.xsd` `http://www.loc.gov/standards/mets/mets.xsd`

需要注意的是, 在某些情况下, 一个 XSD 中会有指向另一个 XSD 的指针。

怎么办?

图书馆工作人员出于本能的、也是最简单的选择就是打开相应的端口允许其连接到互联网, 这样就可以使用标识位置上相应的 XSD 验证 ALTO。

然而, 还有很多基本问题需要解决。

首先, 在将一个文件存储到永久存储库前的准备过程中, 允许进行多大程度的干预? 最好进行多大程度的干预? 是不干预还是想怎么干预都行? 跟大多数组织一样, 新西兰国家图书馆也经常在这两种极点中变换干预模式。一方面, 图书馆无法为保存准备工作付出太多, 准备工作可能包括编辑删除不需要的资料, 这些资料没有保存价值或对知识实体无内在价值但却阻碍保存摄入。另一方面, 这种编辑可被视为对源文件的篡改, 因此就涉及到对象的真实性和完整性问题。

编辑 XML 本身来改变 XSD 中的参考位置也有其相应的问题:

- 有大约 240 万的 XML 文件, 这意味着需要存储和基础设施来处理如此大量的资源;
- 有将其他问题 (如被忽视的错误) 引入 XML 的可能性;
- 如何验证每个更改的准确性;
- 何时更改 XML: 是在其处于源位置时还是在元数据提取时, 这两种做法都有自身的一些问题;
- Papers Past 应用、网站和分发的变换所带来的潜在影响, 每个都需要具体分析和测试, 并追踪对于 XML 数据产生的变换。

第二个关键问题跟 XSD 声明的作用有关。XSD 是否具有保存价值呢? 如果有的话, 是什么价值? 图书馆目前在做的工作怎样阐明这个价值并解决所有可能引起的相关问题?

负责新西兰数字资源长期保存计划的国家数字遗产档案 (NDHA: National Digital Heritage Archive) 团队一直认为, 格式的相关配套文件, 如格式规范, 应作为计划的一部分

加以保存, 最合适的存储位置是格式库。为什么 XSD 会有不同呢?

维基百科对 XSD 的定义是: “……XML 文档类型的描述, 通常表示该类型文件在结构和内容上的限制, 超出 XML 本身的基本语法的制约……”“……检查一个 XML 文件是否符合一个结构称为验证, 是独立于 XML 句法格式良好的核心概念之外的。”

因此, 在长期保存计划的背景下新西兰国家图书馆决定将 XSD (验证文件) 与原始的 XML 文档放在一起保存。

XML 目录: 怎样定位并获取 XSD?

决定将 XSD 作为存入永久存储库中的对象的一个保存属性后, 下一个需要详细讨论的问题就是该将这些对象保存在哪里。显然, 这些对象已通过 schemaLocation 提示进行设计, 以便访问声明中指定的主机的相应 XSD。但是, 这是否同样适用于数字资源长期保存计划呢? 该主机仍然可用并可访问的把握有多大? 该主机上的 XSD 本身仍可用并可访问的把握有多大?

考虑到这些问题, 新西兰国家图书馆决定为这种类型的资料建立一个本地的 XML 目录。该图书馆在其数字化计划中曾有这种相关经验。XML 目录提供了一个标准化的方法来将“外部标识符和 URI 引用映射到 (其他的) URI 引用上。目录处理器的主要任务是在目录中为一条输入找到匹配的条目并返回相应的 URI 引用作为输出。实际操作中总是使用找到的第一个匹配, 不需要目录处理器再去寻找额外的匹配。”这种方法的主要优势是能够将 XSD 的时间点版本作为 Rosetta 系统的一部分进行保存, 使其在图书馆的控制之下得到正规的、有管理的备份。值得注意的是, 虽然图书馆正在处理的具体问题都跟 XSD (XML 结构定义) 有关, 但 XML 目录方法也同样适用于所有的 DTD (文件类型定义)。

接下来的问题就是将这个目录存放在哪里。该图书馆的数字资源长期保存系统 Rosetta 的格式库中目前没有保存这类文件资料的功能, 而永久存储库中的对象急需此项功能。上面提到的格式规范也存在同样的问题。

短期的解决办法是给这些资料提供一个内部位置。希望 Rosetta 格式库经过不断演化, 将来可以为这类证明文件的保存提供恰当的方法。

XML 目录方法只需对整体进行较少的变化, 包括:

- JHOVE 中的目录的一个允许时间
- 提供转换机制的 XML 目录文件的创建时间
- 为 Rosetta 环境下的目录和相关的 XSD 提供一个本地站点

该方法需要持续费用来支持。引进某一新的 XSD (如 ALTO1.3) 时, 图书馆需要利用一项技术监测功能通过对整个 Rosetta 的目录文件进行更新来识别和响应新的 XSD。然而, 这项开支有利于维持对数字资源长期保存计划的控制。因为 XML 目录方法不需要开放一个新的端口, 而且引用任何一个新的、未列入目录的 XSD 都将被防火墙标记, 否则就会失败。

JHOVE 能否支持 XML 目录方法?

新西兰国家图书馆与领导 JHOVE2 项目的加州数字图书馆 (CDL) 的团队讨论了以下问题:

- 确认 JHOVE1 不支持 XML 目录
- 通过验证 JHOVE1 中 NDHA 的定制代码开发来建立 XML 目录
- 确认 JHOVE2 支持 XML 目录

与 CDL 的讨论还证实了一些对象引用 XSD 作为其保存属性是有价值的, 并且证明了 XML 目录是一个有效的方法。

XSD 和加强数字资源长期保存过程

NDHA 一个新的担忧是数字资源长期保存的记录水平不够成熟和全面。例如, 为外部和内部引用的 XSD 建设本地的 XML 目录解决方案, 数字资源长期保存系统需要了解:

- 是否有源含义, 如果有的话其性质和程度是怎样的?
- 出处说明应附在哪一层次结构上? 例如, 是知识实体层次还是文件层次?
- 这两种选择的数据模型各有什么影响?
- 如果采用内部定制版的 JHOVE (提供 XML 目录功能), 怎样区分用 JHOVE1、NLNZ 定制的 JHOVE 和 JHOVE2 验证的材料?
- 如果从 NLNZ 定制的 JHOVE 切换回 JHOVE1 的话, 怎样衡量和管理已被定制的 JHOVE 验证的资料的后继影响?
- 对永久存储库中的资料用一个以上 JHOVE 版本进行验证, 数据模型对其的影响是什么?
- 是否应该有一个设备能运行同一工具的多个版本或一个特定工具的不同类型(如用于病毒检查的 ClamAV 和 ESET Nod32)?

事实证明, 定制的 JHOVE 不能附加任何元素到文件的数据模型中, 要实现这一目的就需要改变 Rosetta 代码中的 JHOVE 插件。同样, 也不能给 Rosetta 中的文件数据模型添加出处说明, 没有设计出能插入出处说明的插件。这些工作可以在提取元数据时做, 但不太理想。

解决方法

将 XML 文件纳入 Rosetta 时, 应用 JHOVE 来验证与其 XSD 声明是否一致。如果格式库不支持的话, XSD 就需要定位于本地 Rosetta 环境中, 而不是定位于 NDHA 之外。那么, 该怎样配置 JHOVE 才能选择存放 XSD 的位置呢?

在测试环境下的两个开发用服务器中将 /exlibris/dps/d4_1/system.dir/thirdparty/jboss/server/default/deploy/repository.ear/lib/ 目录下的 jhove-module.jar 换成支持 XML 目录的 JAR 版本。

下面的目录文件和下载的 schema 被复制到相应的目录:

- CatalogManager.properties: /exlibris/dps/d4_1/system.dir/conf/
- xml_catalog.xml: /net/unua/natlib_data/elscratch/XmlCatalogues/

- A bunch of downloaded schema files:
/net/unua/natlib_data/elscratch/XMLCatalogues/schema_files.

然后重启两个服务器。

结果

- 定制的 JHOVE 参照引用 schema 文件 (如 alto-1-2.xsd, mets-metax.xsd) 的本地副本, 能定位初始的 XSD。
- 定制的 JHOVE 未能定位初始 XSD 文件内部引用的其他 XSD 文件。
- 测试的 SIP 内的 XML 文件显示“无效”, 很可能是因为 JHOVE 无法定位引用原始的 XSD 文件。
- 幸运的是, 事实证明, 这些其余的 XSD 的引用是错误地输入到目录中的, 可以得到修复。
- 测试的 SIP 中的 XML 文件终于准确地被测试为“有效”了。

Where to From Here? (将来如何?)

如上所述, 主要的决策制定都围绕着 XSD 的保存价值和 XSD 能够证明其出处 XML 文件的完整性这一作用。

然而, 真正的问题涉及到格式库的持续发展和其支持的服务的性质。从数字资源长期保存角度来看, 应通过本地 XSD 目录来维持与 XSD 的关系而不能通过外部的。

这也解释了本地实例为什么应该出现在格式库中, 与其它你期望的文档、规范、schema、DTD 和支持的其它文件格式保存在一起。

与 CDL 的对话证明了这仅是 NDHA 独有的问题, 因此将其作为对保存进程的核心支持嵌入到持续发展的 Rosetta 中, 特别是格式库的更广泛的规范中是非常明智的。

这需要 Ex Libris 用两种方式提供援助:

- 首先, 支持 JHOVE (即新的 JHOVE2 目录功能。虽然 JHOVE2 可以安装到 Rosetta 中, 但也无法马上使用), 特别是要能在一个系统内运行和管理多个版本的 JHOVE;
- 扩展格式库, 扩大其在技术和说明方面的能力, 支持更多的单一存在和共同存在的文件格式;
- 考虑到扩展数据模型, 使其包括多个 JHOVE 用法的更精细的信息。

如上所述, 这一级别的功能将在即将发布的病毒检查环境下的 Rosetta 中提供。在所有的第三方产品中, 这种灵活性是大家所乐见的, 如 JHOVE, 在第三方工具的使用和管理方面, 承诺能够运行通过改变数据模型来进行更好的记录保存的多个版本的 JHOVE。

编译自: <http://www.ijdc.net/index.php/ijdc/article/view/205>

(李红培 吴振新校对)

《专题报道》

个人数字存档专题报道

美国国会图书馆著 齐燕编译

大约五年前,就有人思考人们是如何管理自己的数字资产的。例如,随着手机、相机和其他消费级数码捕获设备的普及,个人数字资产集合的规模也在迅猛增长,人们将在哪里存放它们?同时,这些数字资产通常是来自不同的媒介,人们是如何进行区分保管的?有没有一些个人数字资产处于危险状态?

2010年,Jeff Ubois组织了首届“个人数字存档会议”来研究这些问题。参会者来自不同的机构、不同的行业,比如家庭电影、数字资产、卫生、金融、谱系学等领域。首次会议取得了成功,并使之成为每年一次的活动,第三届个人数字存档国际会议已在今年2月底召开。这些努力促进了一个新实践社区的诞生,并使得数字保存领域的注意力由机构存档转向了个人存档。

当前,数字保存工作者面临的一大挑战就是如何让公众意识到个人数字资产的处境及面临的威胁。“照片存在鞋盒中”这种善意忽视的年代已经结束,个人数字资产的存档需要更高水平的责任。

数字时代赋予了普通大众分享和保存其数字资产的可能,有些个人数字存档确实也会有巨大的文化价值。然而,存储仍然是一个复杂而且昂贵的难题。目前有很多商业存储解决方案,但都不具有文化机构长期存档的承诺和可靠性。不过,要是由机构提供个人存档帮助服务,从哪里获得相关的财政支持?

Ubois建议,或许个人可以在某一文化机构中购买1TB的存储空间,或者内容缴存者支付一笔一次性的存储费用。在南加利福尼亚大学已有类似事情发生,但是进展缓慢。因为机构建立之初并没有考虑到这一问题,恪守永久保存的承诺对于它们来说也是有一定风险的。当前,数字保存社区应该让公众了解到他们需要立即处理这个问题,因为个人数字存档有可能会是非常昂贵和有很大技术难度的。

美国德克萨斯大学奥斯汀分校信息学院的Sarah Kim在2010年6月至2011年12月期间采访了不同行业背景的20个人,了解了人们在个人数字存档方面的需求。有些人想要删除其所有的文件,因为他们觉得这是他们自己的东西,对于别人来说可能毫无用处。有些人想要在老年时创建一个其选定文档的浓缩集合。有些人想要对其个人数字资产进行分类,以

交付给不同的人使用,比如一些家族文档交给特定的家庭成员,而一些工作文档则交给其同事。有些人还没有特别的计划。有些人则希望交由管理者或其他人来决定某些资料是否需要存档。他总结了人们考虑对其个人数字文档进行长期保存的原因:

- 对于他人可能具有潜在的用处,比如家族病史或者其它调查研究;
- 想要被家庭成员和朋友们纪念;
- 对某些文档、照片、录音、视频或者其他数字资产有些特殊的感情。

今天,每个人都是自己的档案保管员。我们可以自己来决定都要保存哪些个人数字资产以及如何保存。但是有个很严重的问题是,在管理个人数字资源上,绝大部分人都做得不够好。很少有人花时间去组织我们的数字化生活,很多人并没有充分理解他们的存档角色。这意味着,很多人的家庭数字资料集合是非常凌乱的,或者可能跟网络上数十亿甚至数万亿的其他人的照片混合在一起。在我们将一些照片下载到电脑上时,通常会自我承诺稍后返回来对其进行整理,给它们设置有意义的文件名字、标签以及重要的元数据,但是很多时候我们不一定能够做得到。这些不良习惯都会使得我们的个人数字资产处于危险状态。

针对这一情形,有学者认为应该从以下两方面来解决。第一,图书馆和档案馆必须更多地介入个人数字存档领域,增强人们保存其个人数字资产的意识,并提供基础性的指导。其次,技术社区也应该提供更好的工具和服务来为个人数字存档提供更有力的支持。

在第一点上,NDIIPP已经有所作为。国会图书馆工作人员认为,当前可供人们使用的有关个人数字存档的学习资料还非常少,不过地方图书馆和档案馆可以填补此项空白。为此,他们编制了一套资源工具包([resource kit](#))来帮助地方机构提供个人数字存档相关信息的链接。该资源工具包同时提供了一些资源来支持地方上的“个人数字存档日”活动,以使机构与用户直接面对面交流。

该资源工具包面向图书馆、档案馆和其他文化机构的工作人员,提供了如何组织活动的基础性指导,包括组织活动的一些技巧提示,涵盖了NDIIPP的多种资源和讲义,详细地介绍了公众保存其个人数字文件、照片、视频、电子邮件和社交媒体账户的重要技巧和策略,另外还有一系列附加资源,包括一些特定格式的建议等。具体如下:

一、 保管个人数码照片

与其它任何类型的个人数字信息相比,照片通常会具有更丰富的涵义,而且照片是独一无二的,一旦失去,相关信息就永远无法找回了。肯定有一些数码照片,你是想要保留很长一段时间。那么就注意将重要的照片进行分组归类以便集中管理。

存档提示:

1. 确定你的数码照片都在哪里

- 在照相机、电脑和诸如存储卡的可移动媒介上找出你所有的数码照片。
- 还包括您存储在 Web 上的照片。

2. 判断出哪些照片是最重要的

- 挑出你觉得特别重要的图像。
- 你可以挑几张，也可以挑很多。
- 如果同一个重要的照片有多个版本，要保存具有最高画质的那个。

3. 组织整理挑选出来的照片

- 为每张照片定义描述性的文件名称。
- 为照片添加人物名称及描述性主题信息。
- 在您的计算机上创建一个目录/文件夹结构，来存放所选择的图像。
- 为该目录结构及其所包含照片创建一个简要说明。

4. 制作副本，并将它们存储在不同的地方

- 针对您所选择的照片，至少要制作两个副本，当然越多越好。
- 一个副本可以保留在您的计算机或笔记本电脑上，其它的副本应该存放在单独的媒介上，比如 DVD、CD、便携式硬盘、指状储存器或网络存储等。
- 副本存储的不同位置在物理上要相隔足够远的距离。一旦某个地方发生毁灭性的灾难，存储在其它地方的照片必须能够保证是安全的。
- 将这些照片详细目录的副本与其它重要文件一起放在一个安全的位置。
- 每年要至少有一次对您保存的照片进行检查，以确保其可读性。
- 每五年或者在必要的时候，创建新媒介格式的副本，以避免数据丢失。

二、 保管个人数字音频

您可能有许多数字音频文件，包括音乐、讲座和其他录音资料等。其中一些会有个人的、商业的或其他的价值，使您想要对其进行长期保存。您必须确保您挑选出来用作保存的音频文件是一种开放的文件格式，这样才能保证有供未来使用的最大灵活性。

存档提示：

1. 识别出您所有的数字音频文件

- 在电脑、音频播放器、电话，以及诸如存储卡和 DVD 光盘等可移动媒介上找出您的所有音频文件。

- 还包括您通过音频软件管理的那些音频文件。
2. 判断出哪些录音资料具有长期价值
- 挑选出您觉得重要的录音资料。
 - 如果同一个重要录音资料有多个版本,要保存具有最高音质的那个文件。
3. 导出选定的音频资料
- 如果只是保存几个录音文件,您可以使用 Web 浏览器或者软件程序上的“另存为”命令将其导出为单独的文件。
 - 如果是要保存很多录音资料,则需考虑使用软件自动导出。
 - 如果可能的话,将录音资料保存为一种开放的文件格式。
 - 保存音频文件的元数据,包括其被导出的日期。
4. 组织整理选定的音频资料
- 为每个音频文件定义描述性的文件名称。
 - 为文件添加有关录音内容的描述信息。
 - 在您的计算机上创建一个目录/文件夹结构,来存放所选择的录音文件。
 - 为该目录结构及其所包含的录音资料创建一个简要说明。
5. 制作副本,并将它们存储在不同的地方
- 针对您所选择的录音资料,至少要制作两个副本,当然越多越好。
 - 一个副本可以保留在您的计算机或笔记本电脑上,其它的副本应该存放在单独的媒介上,比如 DVD、CD、便携式硬盘、指状储存器或网络存储等。
 - 副本存储的不同位置在物理上要相隔足够远的距离。一旦某个地方发生毁灭性的灾难,存储在其它地方的录音资料必须能够保证是安全的。
 - 将这些资料的简要描述的一个副本与其它重要文件一起放在一个安全的位置。
 - 每年要至少有一次对您保存的音频文件进行检查,以确保其可用性。
 - 每五年或者在必要的时候,创建新媒介格式的副本,以避免数据丢失。

三、 保管个人数字视频

如果您正在使用相机或其他设备来录制数字视频,那么您是在追随制作个人和家庭电影的悠久传统。您可能想要长期保存其中一些影像资料。

技术处理后的文件质量是数字视频保存时的重要考虑因素。比如,Web 上发布的很多视频文件都是有粒状效果的,所包含的信息通常远不及其原始版本,所以应该保存最高质量

的视频文件版本，并对其相关内容进行详细的描述。

存档提示：

1. 确定您所有的数字视频的位置

- 在相机、电脑、电话，以及诸如存储卡等可移动媒介上找出您的所有数字视频文件。
- 还应包括您存储在 Web 上的视频文件。

2. 决定哪些视频文件是最重要的

- 挑选出你觉得最重要的视频文件。
- 您可以挑选几个，也可以挑选许多。
- 您可以仅保存最后编辑的版本，或者您可以同时保存未经剪辑的片段。
- 如果一个重要视频有多个版本，要保存具有最高质量的那一个。

3. 组织整理选定的视频文件

- 为每个视频文件定义描述性的文件名称。
- 为文件添加人物名称及相关的描述性主题信息。
- 在您的计算机上创建一个目录/文件夹结构，来存放所选择的视频文件。
- 为该目录结构及其所包含的视频资料创建一个简要说明。

4. 制作副本，并将它们存储在不同的地方

- 针对您所选择的视频文件，至少要制作两个副本，当然越多越好。
- 一个副本可以保留在您的计算机或笔记本电脑上，其它的副本应该存放在单独的媒介上，比如 DVD、CD、便携式硬盘、指状储存器或网络存储等。
- 副本存储的不同位置在物理上要相隔足够远的距离。一旦某个地方发生毁灭性的灾难，存储在其它地方的视频文件必须能够保证是安全的。
- 将这些资料的简要描述的一个副本与其它重要文件一起放在一个安全的位置。
- 每年至少要有一次对您保存的视频文件进行检查，以确保其可读取性。
- 每五年或者在必要的时候，创建新媒介格式的副本，以避免数据丢失。

四、 保管个人电子邮件

如同纸质的信件一样，您的电子邮件也记录了重要的事件、事务和人际关系信息等，您可能想要保存其中一些或者是许多内容。对电子邮件的保存意味着要保持其与电子邮件程序的分离，因为电子邮件程序的职责不是要长期保存信息，它们随时会改变或者停止提供相关支持。

存档提示:

1. 确定您的电子邮件的所有来源

- 确定您的所有个人电子邮件账户。
- 在每个账户中, 查找所有的文件夹或其他单独的信息分组; 还要包括所有“已存档”的信息。

2. 判断哪些邮件具有长期保存的价值

- 挑选出您觉得特别重要的邮件信息。
- 您可以挑选几封邮件, 也可以挑选许多邮件。
- 要同时保存作为所选邮件组成部分的相关附件。

3. 导出选定的邮件信息

- 如果只是保存几封邮件信息, 您可以使用电子邮件浏览器或者软件程序上的“另存为”命令将其导出为单独的文档。
- 如果是要保存很多封电子邮件信息, 则可以考虑使用电子邮件程序实现自动导出。
- 如果可能的话, 将电子邮件信息保存为一种开放的文件格式。
- 保存电子邮件信息的元数据, 包括“信件标头”(主题、发送方、接收方、时间和日期等)。

4. 组织整理选定的电子邮件信息

- 为每个电子邮件信息文件及相关附件定义描述性的文件名称。
- 在您的计算机上创建一个目录/文件夹结构, 来存放所保存的电子邮件信息及其相关附件。
- 为该目录结构及其所包含的文件创建一个简要说明。

5. 制作副本, 并将它们存储在不同的地方

- 针对您选定的电子邮件信息文件及其附件, 至少要制作两个副本, 当然越多越好。
- 一个副本可以保留在您的计算机或笔记本电脑上, 其它的副本应该存放在单独的媒介上, 比如 DVD、CD、便携式硬盘、指状储存器或网络存储等。
- 副本存储的不同位置在物理上要相隔足够远的距离。一旦某个地方发生毁灭性的灾难, 存储在其它地方的电子邮件及其附件的副本必须能够保证是安全的。
- 将这些资料的简要描述的一个副本与您的重要文件一起放在一个安全的位置。
- 每年要至少有一次对您保存的电子邮件及其附件进行检查, 以确保其可读性。
- 每五年或者在必要的时候, 创建新媒介格式的副本, 以避免数据丢失。

五、 保管个人数字档案

您可能有一些个人简历、学校论文、财务电子表格、演示幻灯片或其他数字文件，您还会有一些对于诸如信件、地图或家谱的原始硬拷贝文档的数字拷贝文件，其中一些会有永久的价值。

对于这种类型的信息，决定保存哪些文档是非常重要的。必须考虑到不同的版本，比如草稿和早期的副本，有些草稿通常会含有最终版本所没有的重要细节信息。

存档提示：

1. 确定您的数字文档文件的位置

- 在您的电脑以及诸如 CD-ROM 和软盘等可移动媒介上找出您的所有数字文档。
- 还包括您存储在相关网站上的所有个人文档。

2. 判断哪些文档具有长期保存的价值

- 挑选出您觉得特别重要的文档。
- 您可以挑选几篇，也可以挑选许多。
- 您可以选择仅保存文档的最终版本，也可以选择同时包含草稿和修订版。

3. 组织整理选定的文档

- 为每篇文档定义描述性的文件名称。
- 在您的计算机上创建一个目录/文件夹结构，来存放所选择的文档。
- 为该目录结构及其所包含的文档创建一个简要说明。

4. 制作副本，并将它们存储在不同的地方

- 针对您所选择的文档，至少要制作两个副本，当然越多越好。
- 一个副本可以保留在您的计算机或笔记本电脑上，其它的副本应该存放在单独的媒介上，比如 DVD、CD、便携式硬盘、指状储存器或网络存储等。
- 副本存储的不同位置在物理上要相隔足够远的距离。一旦某个地方发生毁灭性的灾难，存储在其它地方的您的重要文档必须能够保证是安全的。
- 将简要描述的一个副本与您的重要文件一起放在一个安全的位置。
- 每年至少要有一次对您保存的文档文件进行检查，以确保其可读性。
- 每五年或者在必要的时候，创建新媒介格式的副本，以避免数据丢失。

六、 保管个人网站、博客和社交媒体

如果你有一个博客、网站、Facebook 网页或其他类型的互联网信息共享途径，那么你

还需要考虑对这一丰富信息源的长期保存事宜。

对于这类信息的保存,首要的步骤是要识别出您都有些什么。您可能有多处信息共享群落,那么您需要对每一处都考虑到。

存档提示:

1. 确定所有与您相关的信息源

- 在 Web 上,包括个人网站、社交媒体网站和其它网络服务站点找出您的所有相关内容。
- 需确保既包含最新的信息,也包括所有早期的(已存档)的信息。

2. 判断哪些信息具有长期价值

- 挑选出您觉得特别重要的信息。
- 您可以挑选一些独立的信息片段,也可以选择整个页面或者网站。

3. 导出选定的信息

- 如果只是保存少量的信息,您可以使用 Web 浏览器上的“另存为”命令将其导出为单独的文档。
- 如果是要保存完整的页面或者网站,则需要相关软件将其自动地导出为一系列关联文件。
- 保存 Web 内容信息的元数据,比如站点名称或者创建日期等。

4. 组织整理选定的信息

- 为每篇文档定义描述性的文件名称。
- 在您的计算机上创建一个目录/文件夹结构,来存放所保存的信息。
- 为该目录结构及其所包含的文件创建一个简要说明。

5. 制作副本,并将它们存储在不同的地方

- 针对您所选择的信息,至少要制作两个副本,当然越多越好。
- 一个副本可以保留在您的计算机或笔记本电脑上,其它的副本应该存放在单独的媒介上,比如 DVD、CD、便携式硬盘、指状储存器或网络存储等。
- 副本存储的不同位置在物理上要相隔足够远的距离。一旦某个地方发生毁灭性的灾难,存储在其它地方的重要 Web 内容的副本必须能够保证是安全的。
- 将简要描述的一个副本与您的重要文件一起放在一个安全的位置。
- 每年至少要有一次对您保存的文件进行检查,以确保其可读性。
- 每五年或者在必要的时候,创建新媒介格式的副本,以避免数据丢失。

总之, 这些个人数字存档指南都是来自于实践经验的总结, 为想要实施个人数字存档计划的个人提供了切入点。但在制定具体计划时, 通常还需要参考其它一些资源(参见<http://www.digitalpreservation.gov/personalarchiving/padKit/resources.html>), 主要包括:

有关活动策划的额外资源:

- ALA-ALCTS 保存周 ([ALA-ALCTS Preservation Week](#)): 有关保存周的信息, 及辅助活动策划的一些资源。

有关个人数字存档的额外资源:

1) 基本信息:

- JISC数字媒体 ([JISC Digital Media](#)): 该文档涵盖了创建和使用数字媒介资源的所有方面。
- 个人档案创建者的范例指南 ([Paradigm Guidelines for Creators of Personal Archives](#)): 英国范例项目编制的帮助创建者保管其个人数字档案的非常实用的一系列技巧指南。
- The Signal: 数字资源长期保存博客 ([The Signal: Digital Preservation Blog](#)): NDIIPP计划的博主会经常写一些有关个人数字存档的挑战。

2) 照片

- 保存数码照片 ([Preserving Digital Photographs](#)): 新西兰国家图书馆编制的保存数码照片的技巧。
- 文件格式迁移: dpBestflow.org ([File Format Migration: dpBestflow.org](#)): 来自于美国媒体摄影师协会开展的数码摄影的最佳实践和工作流程(dpBestflow®)项目, 研究了如何优化和改进商业数码艺术作品的加工、创作和保存等活动。这个网站提供了数字图像的文件格式迁移的信息。

3) 个人数字档案

- 数据解放阵线 ([Data Liberation Front](#)): 介绍了用户如何将个人数据导入 Google 产品中或者导出在 Google 产品里的个人数据。

4) 视频

- 数字视频保存: 确定其容器和编解码器 ([Digital Video Preservation: Identifying Containers and Codecs](#)): 来自史密森尼学会档案馆的博客, 讨论了用于长期保存的数字视频格式。

5) 音频

- 保存录音资料 ([Preserving Sound Recordings](#)): 新西兰国家图书馆编制的保存录音资料的技巧提示。
- JISC 的数字媒体: 选择一种数字音频格式 ([JISC Digital Media: Choosing a Digital Audio Format](#)), 这个网络资源提供了建议、指导和培训, 讨论了在选择一种音频格式前必须要考虑的一些关键因素, 并为不同的应用建议了适合的文件格式。它也提供了有关开放格式和专有格式的相关信息。

6) 电子邮件

- 电子邮件记录的过去 ([MUSE](#)): Muse (Memories Using Email) 是一个帮助人们使用长期保存的电子邮件档案来唤起回忆的项目。
- 你已持续收到很多邮件 ([You've Still Got Mail](#)): 来自史密森尼学会档案馆的博客文章, 讨论了电子邮件管理的一些策略。

7) 社交媒体

- HTTrack 网站复印机 ([HTTrack Website Copier](#)): HTTrack 是一款免费的 (GPL, 自由/免费软件)、易用的离线浏览器工具, 利用它可以将一个万维网网站从 Internet 上下载到本地目录, 并递归地构建好所有的目录, 以及从服务器端获取 HTML、图像和其他文件到您的计算机上。
- 网文快捕 ([Scrapbook](#)): 这是一款 Firefox 插件, 一旦安装, 就可以帮助您保存和管理网页。它有多种捕捉选项可供选择, 还配备相应的组织和编辑功能。
- 另存为... ([Save as....](#)): 存储网页到电脑硬盘时最常用的方法。如果您使用的是 Mac, 苹果的 Safari 浏览器上会有“另存为...网页”的功能。

8) 格式综述

- 保护 CD 以延长其寿命 ([Caring for CDs and Their Longevity](#)): 来自加拿大保护学院, 有关处理和保护光学媒体的常见问题解答。
- JISC 数字媒体: 选择一个文件名 ([JISC Digital Media: Choosing a File Name](#)): 文件命名是数字化的首要步骤。这篇文档来自 JISC 数字媒体, 为英国的继续教育和高等教育团体提供了建议、指导和培训, 验证了在管理数字文件时规划和使用一个有效的文件命名系统的必要性。该文重点介绍了使用 8.3 约定的优势, 并考虑了为衍生产品和代理文件命名时的一些可选项。
- 媒体复原: 丢失的记忆 ([Retro Media: Memory \(and Memories\) Lost](#)): 这是一个联机的媒介存档展览馆, 来自纽约州立大学水牛城分校图书馆, 记录了各种媒介格

式的简史、谈及了计算机数据及存储,以及在媒介保存和迁移中图书馆的作用。

数字保存相关资源

- 数字保存手册 ([Digital Preservation Handbook](#)): 数字保存联盟编制的关于数字资源长期管理及持续获取的指南。
- 面向北卡罗来纳州政府雇员的数字保存培训 ([Digital Preservation Education for NC State Government Employees](#)): 这一网站的资源可以帮助北卡罗来纳州政府雇员,以及数字资料的普通用户来学习有关数字保存方面的必要知识。
- 数字磁带保存策略: 保存数据还是视频? ([Digital Tape Preservation Strategy: Preserving Data or Video?](#)): 这篇文章描述了同时包含原生数字和基于磁带的大型视频资料集合所适用的保存理念和策略。
- 数字资料的脆弱性 ([The Fragility of Digital Materials](#)): “数字历史: 收集、保存,以及在 Web 上呈现历史的指南”的一个运用,详细解释了数字格式的脆弱性及处于危险状态的原因。
- JISC 数字保存博客指南 ([JISC Beginner's Guide to Digital Preservation Blog](#)): 为英国 JISC 项目工作人员制定的数字保存指南。
- NEDCC 保存工具包 ([NEDCC Preservation Toolkit](#)): 由美国博物馆与图书馆服务研究所 (Institute of Museum and Library Services, IMLS) 赞助的一个项目。该网站记录了数字保存意愿调查结果,同时提供了适合于文化机构的工具包。
- 团队数字资源长期保存 ([Team Digital Preservation](#)): 这是由欧洲数字保存社区创建的一系列视频资料,描述了数字资源长期保存的挑战。

数字保存相关视频文件下载地址:

Digital Preservation Website: <http://www.digitalpreservation.gov/multimedia/videos.html>

YouTube: <http://www.youtube.com/user/LibraryOfCongress#g/c/EA69BE43AA9F7E68>

iTune U: <http://itunes.apple.com/us/itunes-u/digital-preservation-webcasts/id386017836>

编译自:

<http://blogs.loc.gov/digitalpreservation/2012/02/jeff-ubo-is-and-personal-digital-archiving/>

<http://blogs.loc.gov/digitalpreservation/2012/02/hey-libraries-and-archives-personal-digital-archiving-kit-now-available/>

<http://blogs.loc.gov/digitalpreservation/2012/03/ndiipp-at-sxsw-digital-immortals-and-the-hipster-halo-effect/>

<http://blogs.loc.gov/digitalpreservation/2012/03/the-results-of-one-scholars-survey-what-are-your-plans-for-your-personal-digital-archives/>

http://www.digitalpreservation.gov/personalarchiving/documents/PADKit_v1.pdf

<http://www.digitalpreservation.gov/personalarchiving/padKit/handouts.html>

<http://www.digitalpreservation.gov/personalarchiving/padKit/resources.html>

(李红培 吴振新校对)

【资料推荐】

《内存块：如何保持历史资料的“鲜活”》

这本“白皮书”产生于一个由弗吉尼亚大学牵头、由安德鲁·梅隆基金会赞助，有来自斯坦福大学、耶鲁大学和英国赫尔大学的数字档案保管员共同参与的合作项目。它关注项目合作伙伴，也是绝大多数图书馆和档案馆的共同需求——找到一种面向积压多年的原生数字档案资源的管理方法或持续性框架。在过去，只要有适宜的物理条件，就可以将纸质资源保存数百年，但是现在，创建和存储数字资源的工具日新月异，让档案保管人员每隔几年就会面临一个新的挑战。数字化地产生、数字化地存储，以及没有其它形式的副本，给保存工作带来了巨大的困难。有些档案馆已经应对这一问题多年，但也有一些未曾真正开始甚至害怕迈出第一步，这本白皮书就是要向他们分享经验，帮助他们行动起来。总之，高校档案管理员一直试图找到一种处理数字化资料的通用方法，这本书正是一个突破。

详见：<http://www2.lib.virginia.edu/aims/whitepaper/>

编译自：

<http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=419382&c=1>

(齐燕编译，李红培 吴振新校对)

NDIIPP 的报告、白皮书和文章

NDIIPP成立十一年以来，该计划的数字化保存网络规模已有大幅度扩展，从最初的36个合作机构到现在的超过280个来自世界各地的组织。该共同体管理了超过1,400个数字内容集合，开发了不同的工具和服务来支持这些内容的保管、迁移、验证和保存，并产生了很多项目报告、白皮书和文章等出版物，NDIIPP在网站上将这些成果分享给更多受众，包括：

1. MetaArchive Cooperative宪章(2011年)

MetaArchive Cooperative是一个致力于保护原生数字的或数字形式的文化遗产。此为宪章的2011年更新版,支持描述MetaArchive Cooperative的宗旨和目标的文件。

- MetaArchive 合作章程

(http://www.metaarchive.org/public/resources/charter_member/2011_MetaArchive_Charter.pdf, PDF, 133KB)

- MetaArchive 技术规范

(http://www.metaarchive.org/public/resources/charter_member/2011_Technical_Specifications.pdf, PDF 格式, 65KB)

- MetaArchive 会员协议

(http://www.metaarchive.org/public/resources/charter_member/2011_Membership_Agreement.pdf, PDF 格式, 62KB)

2. 为什么网站会丢失(和为什么它们有时又会被发现)(2009年)

该文载于《ACM通讯》,52卷,第11期(2009年11月),作者Michael Nelson,来自自己保存的网络工具项目(Tools for a Preservation-Ready Web project)。

其预印本链接为:

<http://www.harding.edu/fmccown/pubs/lost-website-survey-cacm-all-in-one.pdf>(PDF, 149KB)

3. 数据泛滥? 信息时代数据保存指南(2008年)

作者是来自圣迭戈超级计算机中心的Fran Berman。该文载于《ACM通讯》51卷总第12期,讨论了帮助从业人员管理海量数据的一些工具。

详细参见:

<http://www.sdsc.edu/about/director/pubs/communications200812-DataDeluge.pdf>(PDF, 6.16MB)。

4. 美国国会图书馆的数字保管:以美国记忆和存档采集与处理测试为例(2007年)

该文是DigCCurr2007会议的会议论文,讨论了关于美国国会图书馆的美国记忆项目的调查结果和分析结论,以及捐赠的适合数字保存的数字档案。该文阐述了数字保管课程发展的机遇。详细内容参见:http://www.ils.unc.edu/digccurr2007/papers/madden_paper_6-2.pdf。

5. 存档采集与处理测试(AIHT)总决赛报告(2005年)

该报告记录了AIHT的发展、管理和结论,是各种数字资源长期保存制度的一个多方测

试。报告中描述了NDIIPP和AIHT的成因、AIHT各阶段的细节、测试过程中得到的经验,并提出了今后工作很有潜力的领域。

详细内容参见:

http://www.digitalpreservation.gov/partners/documents/ndiipp_aiht_final_report.pdf(PDF, 552KB)。

编译自:

<http://blogs.loc.gov/digitalpreservation/2012/03/reports-white-papers-and-articles-document-the-work-of-ndiipp-partners/>

(么媛媛编译, 李红培 吴振新校对)