

2013 年第 7 期（总第 28 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

2013 年 7 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】	1
数字学术（Digital Scholarship）发布第三版《研究数据保管参考书目》	1
HathiTrust 与 DPLA 达成合作	1
加州大学戴维斯分校公开发布其大学网络档案	2
【动态追踪】	2
2013 年 NDSA 创新奖名单公布	2
DPC 技术观察报告《保存元数据（第二版）》向会员发布	3
【专题报道】	4
2012iPRES 研讨会专题报告：实用的仿真工具及策略	4
【技术与工具】	10
Matchbox 工具集介绍	10
【资料推荐】	10
《透视个人数字存档》报告简介	10
社区数字存档相关资源	11
TIMBUS 第二期新闻通讯发布	13

【信息扫描】**数字学术 (Digital Scholarship) 发布第三版《研究数据保管参考书目》**

数字学术 (Digital Scholarship) 最新发布了第三版的《研究数据保管参考书目》，包括230多篇英文文章和技术报告，内容涵盖了大学及其他研究机构的数字研究数据的保管工作。其中绝大部分都是出版于2000年1月到2012年6月期间，只有很少量的重要资源是2000年以前出版的。详情参见：<http://digital-scholarship.org/rdcb/rdcb.htm>。

编译自：

<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1306&L=digital-preservation&F=&S=&P=6837>

(李文燕编译，齐燕 吴振新校对)

HathiTrust 与 DPLA 达成合作

最近启动的美国数字公共图书馆 (Digital Public Library of America, DPLA) 项目认为数字资源的汇集会创造更大的价值，所以要为全美国的图书馆、档案馆和博物馆所持有的可免费获取的数字材料提供一个统一的在线入口，其目标是，所有分散各处的可共享的美国文化遗产，包括文字、图像、声音等各种对象，都可以通过这样一条统一的路径让公众方便地获取到。

2013年6月18日，HathiTrust与DPLA签署了合作协议。通过知识共享 (Creative Commons) “CC0”协议的许可，HathiTrust的元数据可与DPLA的用户实现共享，同时HathiTrust的近1100万件藏品中的350万件可公开获取的藏品，也可通过DPLA的网站共享。这些藏品依旧由HathiTrust持有和负责长期保存，HathiTrust实际上成为与DPLA合作的多家机构中的数据规模最大的一个数据中心 (content hub) (其它如史密森国家档案馆、纽约公共图书馆)。

在过去的5年中，HathiTrust已经构建了相当规模的基础设施来存储其大学和图书馆合作伙伴扫描的数字资源，与DPLA的合作可以让更多的用户发现这些宝贵的资源，并提供更多的方法获取到它们，提高这些资源的利用率。

更 多 信 息 参 见 ：

<http://www.cdlib.org/cdlibinfo/2013/06/18/uc-libraries-partner-with-dpla-via-hathitrust/>，<http://dp.la/info/2013/06/18/hathitrust-to-partner-with-dpla/>。

编译自：

<http://www.cdlib.org/cdlinfo/2013/06/18/uc-libraries-partner-with-dpla-via-hathitrust/>

<http://dp.la/info/2013/06/18/hathitrust-to-partner-with-dpla/>

(齐燕编译, 刘超 吴振新校对)

加州大学戴维斯分校公开发布其大学网络档案

2013年6月,加州大学戴维斯分校(UC Davis, UCD)发布了其存档的300多个ucdavis.edu域名下的网站,这是启动于2011年的加州大学戴维斯分校大学存档(UCD University Archives)项目的部分成果。该项目旨在抓取大学的管理层、学院和院系、学术部门、行政部门、有组织的研究团体、校际运动会,以及各类学生组织创建的网站。

公众在浏览这些采集自2011年的UCD站点时,既可以看到其四大学院——农业与环境科学学院、生物科学学院、工程学院、文学和科学学院,也可以看到其六个职业学校——贝蒂·艾琳·摩尔护理学院、教育、法律、管理、医药、兽医学等的站点内容。同时,公众还可以浏览到学生会的网站、100多个教学部门的网站,以及刊有大学报告和新闻的行政部门的网站等等。

所有UCD网页档案参见: <http://webarchives.cdlib.org/institutions/UCDL>。

编译自: <http://www.cdlib.org/cdlinfo/2013/06/18/uc-davis-releases-university-web-archives/>

(齐燕编译, 刘超 吴振新校对)

【动态追踪】

2013年 NDSA 创新奖名单公布

6月11日,一年一度的美国国家数字管理联盟(National Digital Stewardship Alliance, NDSA)创新奖评选落下帷幕,2013年获奖名单正式出炉:

- 未来保管员(**Future Steward**)奖获得者:来自Gates存档机构(Gates Archive)的Martin Gengenbach。他的卓越贡献在于为数字取证工具及工作流程进行了完善的文件编制,尤其是他撰写的论文:《我们是这样做的:收藏机构中的数字取证工作流程演示》([The Way We Do it Here: Mapping Digital Forensics Workflows in Collecting Institutions](#)),以及他在DFXML模式编目上取得的成果([his work cataloging the DFXML schema](#))获得了广大评委的高度认可。
- 个人(**Individual**)奖获得者:来自美国韦恩州立大学的Kim Schroeder。她被誉为未来数字保管员的良师益友。在担任美国韦恩州立大学的数字保存讲师期间,她帮

助学生创建了第一个NDSA学生组 ([NDSA Student Group](#)), 为学生们组织的数字保存专题讨论会提供了相关支持, 并促进了数字保管领域的学生们与当地的文化遗产组织之间的协作。

- 项目 (**Project**) 奖获得者: 加州数字图书馆的DataUp。该项目创建了一个开源工具, 能够指导个人用户完成研究数据集的长期保存工作, 包括从数据集的创建和描述, 到将这些数据集缴存在公共仓储库中的整个流程。
- 组织 (**Organization**) 奖获得者: 存档团队 ([The Archive Team](#)), 自称是一个“由一群雄心勃勃的致力于拯救我们的数字遗产的档案管理员、程序员、作家和言论家们组成的自由团体”, 他们积极地从事于保存将要被删除的网站和数字内容这一重要的工作, 同时还积极地推进技术和计算领域内的数字文化的保存。

所有奖项统一在7月23-25日召开的数字保存2013年大会 ([Digital Preservation 2013](#)) 上颁发, 获奖者也会藉此对其工作或项目给予介绍和陈述, 以向人们展示其在探究如何应对数字内容的保存和获取的诸多复杂挑战上的创造力、合作精神以及高涨的积极性。

编译自:

<http://blogs.loc.gov/digitalpreservation/2013/06/and-the-winner-is-announcing-the-2013-ndsai-innovation-award-winners/>

(齐燕编译, 刘超 吴振新校对)

DPC 技术观察报告《保存元数据 (第二版)》向会员发布

继2005年发布第一版《保存元数据》报告后, 今年DPC成员再次编写了《保存元数据 (第二版)》, 记述了8年来保存元数据理论与实践领域的变化和发展, 特别是PREMIS成为事实上的国际标准所带来的影响。

OCLC研究中心和伦敦大学国王学院电子研究中心的从事于数字保存领域电子信息标准工作的专家们, 从第一版报告起就告诉人们, 保存元数据, 尤其是PREMIS数据字典 (PREMIS Data Dictionary) 已经成为支撑数字资源长期可靠管理所不可或缺的重要部分。

该报告论述了PREMIS数据字典发布后所关联的实践中的一系列问题, 包括社区宣传、包装、工具、PREMIS在数字保存系统中的实施过程以及所需相关资源。对保存元数据及其发展现状, 特别是PREMIS的情况感兴趣的读者都可从中受益匪浅。

这是DPC首度发布第二版本的技术观察报告, 具体内容会在第三季度向公众开放。

编译自: <http://www.dpconline.org/newsroom>

(齐燕编译, 刘超 吴振新校对)

【专题报道】

2012iPRES 研讨会专题报告：实用的仿真工具及策略

Dirk von Suchodoletz 等著 刘超 编译

摘要

第9届iPRES会议于2012年10月1日至5日在多伦多大学板栗会展中心（University of Toronto's Chestnut Convention Centre）举办，其中包括一天的“迈向实用的仿真工具和策略”主题研讨。

综述

这次研讨会以“迈向实用仿真工具和战略”为主题，从不同角度审视了作为一种数字保存策略的“仿真”技术，并讨论了新型的复杂数字文物的保存问题的新动向及挑战。这次研讨会汇集了来自近50个活跃在数字保存和长期获取（long-term access）领域的各个机构的从业人员和研究人员，他们中有的来自博物馆，有的来自国家图书馆和各种存档机构，有的来自咨询业务机构，还有的来自数据保管和数字保存服务机构。其中许多人没有接触过仿真技术，但却都深知采取新战略的必要性，因为以往标准的保存和获取策略并不适用于现在大量涌现的新型数字文物。与会人员都十分认同仿真技术作为一种访问策略在数字保存领域日益增加的重要性，并最终甄别出了一些需要协作和跨学科应对的挑战。

此次研讨会的组织者首先从研究人员和从业者那里收集使用案例及正在进行的与“仿真”相关的项目材料。随后分为四个小组对收集内容进行充分讨论，讨论分为四部分：元数据和环境，网络存档和大数据，数字艺术和电脑游戏，以及通过社区参与实现可持续的仿真。

本篇研讨会报告主要对会上提出的不同存储机构当前存在的问题、具有挑战性的使用案例，以及各小组讨论的关键结论进行综述。对其进行进一步的整理，便可识别出需要研发人员与从业者共同合作来解决的即将到来的挑战。

来自OCLC的Titia van der Werf发表了题为“超越争议：作为一种数字保存方法的仿真技术概览”的演讲，从1995年杰夫·罗森堡的著名论文到最近在欧盟资助下的PLANETS和KEEP项目，简短地历数了过去十年中使用仿真技术来解决数字保存问题的情况。

来自不同机构的从业者对不同类型数字材料面临的挑战的谈论拉开了互动讨论的序幕，在这部分，与会者讨论了两种挑战：1) 虽然被保存但使用传统迁移手段有明显缺陷，2) 没被保存但直接对其进行迁移似乎不是一个可行的解决方案。

由于许多国家有法定缴存的要求，因而这些国家的国家图书馆面临着处理各种各样不同种类的新型数字文物的挑战，如数字多媒体百科全书、电脑游戏或软件等。例如，丹麦法定缴存法案第1439条就规定，存储在物理介质上的数字内容也属于缴存范围。目前，丹麦国家图书馆正试图以两种方式将电脑游戏纳入缴存范围：第一，开展协调一致的视频游戏文档管

理实践以确保对游戏相关材料的收集工作的完整性；第二，对在收集和保存便携式媒体（主要是iOS和Android）上“apps（应用程序）”时遇到的技术问题进行了定义和说明。

同样地，按照要求法国国家图书馆必须对国产软件进行保存。作为一个大型的研究型图书馆，德国国家图书馆必须能并提供并确保对多媒体和其它非标准对象的访问。虽然澳大利亚国家图书馆并没明确提出要采用仿真策略，但它是世界范围内少数几个对具有数字考古价值的物理对象进行保存的图书馆之一，包括（可拆卸的）媒体驱动器，特种电缆，控制器，以及计算机软件（如各种应用程序和操作系统）。这些对象将不仅对原始环境的再现大有用处，而且将有助于存储机构处理未来接收的非标准材料。

此外，由于数字革命大大改变了政府部门的工作流程，国家档案馆按规定所接收的材料种类直接受到影响。例如，奥地利国家档案馆所接收到的材料就有通过在线/网络应用程序（如数字税务申报系统）创建和管理的数据。

许多存储机构面临的一个全新挑战是为万维网的归档内容提供访问，而这些机构的归档内容根据不同的呈现应用程序（如专有插件），包含各种不同类型的非标准对象。为了应对这一挑战，澳大利亚国家图书馆的一个工作小组于1996年启动了一个项目，意在保存每年出现的一般通用且较为典型的访问环境。对这些访问环境的虚拟化处理有助于对不同访问环境里的网页内容进行仿真。

对存储机构构成挑战的另一领域是数字艺术。随着创作数字艺术作品的媒体和平台的不断过时和废弃，需要采取新的方法来对其访问入口加以保存。数字艺术品通常无法进行迁移，因为其用户体验要依赖于一定的软件和硬件配置，如作品必须要在一定分辨率的CRT屏幕上才能呈现。此外，由于数字艺术具有实时播放和非标准的人机交互（例如音频、摄像头）等特点，使得其有别于办公室或科学领域中的数字对象。在这方面有所作为的机构之一——卡尔斯鲁厄艺术与设计大学，对从20世纪90年代中期到2000年中期的基于CD-ROM的艺术品进行了访问入口的保存，另外，它的另一个主要目标是对研究生最后一年创作的数字艺术作品进行保存。

在从业者的演示报告展示过后，会议紧接着对最近的仿真及其相关主题的研究项目进行了概述回顾。在这部分，首先提出了原始环境的概念性描述以及缺乏适当元数据框架的问题。作为KEEP项目的成果之一，技术注册表TOTEM能够捕获单独的硬件和软件组件信息，及其之间关系，研讨会上展示了TOTEM基于RDF的XML框架，而且综述了即将发布的PREMIS 3.0元数据标准的讨论意见。由欧盟资助的TIMBUS项目开发出了用以保存业务流程的策略，其中一部分使用了仿真技术来对流程及其环境进行概念和技术捕捉，这部分侧重点也是在于对捕获的和重新配置的流程进行评估。bwFLA是德国巴登-符腾堡州（Baden-Württemberg, Germany）的一个项目，其核心专注于开发可扩展的仿真策略以及设计基于仿真的保存工作流程，与会者以该项目为例提出了一系列关于保存研究数据的挑战。此外，美国印第安纳大学的杰弗里·布朗工作小组正在针对基于MS-Windows原始环境的仿真器和孵化器（emulator

and reproduction) 进行研究。

为进一步对特定主题进行深入讨论,与会人员被分成了四个小组。每个小组针对一个特定的主题,首先试图找出共同的挑战,然后考虑如何应对这些挑战。这四个主题分别是:“元数据和环境”(仿真需要什么样的元数据?在一个特定的环境中,怎么捕捉数字对象的情境信息?);“网络存档和大数据”(怎样应对保存大量不断变化的在线数据的挑战?);“数字艺术和电脑游戏”(如何处理动态和交互式对象具体的技术、法律和组织问题?);“通过社区参与实现可持续的仿真”(在确保社区采取并积极支持仿真工具方面存在哪些问题?解决方案有哪些?)。

元数据和环境

在处理复杂的数字文物时,除了要选择适当的数字保存策略之外,还需要更多的元数据的支持。特别是必须要用原始环境来为数字对象提供呈现的情况下,如果要构建这样一种环境,就需要对所有必要的软件和硬件组件信息进行详细描述。由于对类似的数字文物来说,原始环境应该都是可用的,所以在对其进行描述的时候应尽量明确,以使数字文物的保存需求与环境的功能能够一一对应。此外,还应对软件许可证和文档,以及操作运行的隐性知识进行保存。现有的注册表极为不完整,它无法对软件条目和机器提供标准的描述,也无法体现出对象和软件组件之间的关系,因此不适合于计算用于配置原始环境的对象视图路径。

传统的保存策略通常只是从原始环境(如原始文件系统和软件环境)中提取对象,这样往往会丢失对象的原始环境信息。因而传统策略可能会遗落保存对象的某些重要组成部分,比如一些必要的数字文物、字体集和多媒体对象的编解码器等。业务流程、科研环境以及软件开发系统通常是由专门配置的软硬件网络环境(例如Web服务、传感器)构成的,这就要求这些对象的描述和保存结果要能真实地再现其呈现环境。此外,系统和对象配置问题成为复杂对象的热点话题之一,所谓复杂对象不只简单地由标准软件组件构成,还包括需要特定配置的对象,如数据库、科研环境和业务流程。

网络归档和大数据

对于所使用的网络技术正逐渐被淘汰的存储机构来说,网络归档和大数据成为了一个日益关切的话题。法定缴存的条款规定存储机构须保证用户可长期访问网络归档的材料。但网络存档机构却在见证着归档对象类型(如Flash、Silverlight和JavaScript等)的日益复杂,对于这些机构来说,很难去做采集“计划”或者对接收的对象类型做限定,它们只能紧随技术变革的脚步。这个问题不会因为英国图书馆学界建议的“拒绝老格式消逝”(<http://pressandpolicy.bl.uk/Press-Releases/British-Library-study-suggests-older-web-formats-refusing-to-fade-away-5d0.aspx>) 就被轻易解决。另外,现代的网站的复杂性使得很难对其进行归档,而且因为许多网页是通过用户的请求创建的,网络爬虫也不能抓取到完整的内容。归档机构需要考虑网络存档的未来使用情况,以及如何将资源合理优化地投入到访问策略中。

很显然,浏览器环境和呈现特性会随时间而改变,尤其是使用Wayback Machine

(<http://archive.org/web/web.php>)对现代网站的网页进行过归档处理,情况会更加明显。网络归档并不意味着要把网站的样子和感觉完全复制下来,在这一领域,仿真似乎是不二选择,因为通过使用这种技术可以对呈现出不同网络时代的标准呈现环境进行保存,比如可以保存包括虚拟网络访问在内的呈现环境,这样就不用再对特殊用途的浏览器进行编程了。如果要访问已弃用的格式文件,如旧的文档,或RealMedia格式的音频流,直接使用当时的原始软件就可以了。为了减轻呈现失真的问题,以及能够在未来的环境中对归档网页访问的效度进行评估,需要收集更多的元数据(例如,流行的浏览器的网页截图)。

这一小组讨论的另一个主要焦点是大数据,这一术语目前还没有一个公认的定义。有些存储机构将其解释为数据仓库(Datawarehousing)和关联数据(linked data),也有其它的一些机构把基础科学研究数据(scientific primary research data)称为大数据。先锋性的存储机构已经开始研究大规模(研究)数据的管理问题,比如使用像Apache的Hadoop技术之类的存储策略。尽管对大数据的定义莫衷一是,但参与讨论的与会者一致认为,有必要采取适当的数据描述和元数据策略来对所谓的大数据进行保存、访问和再利用。

讨论围绕着处理大数据的早期阶段逐步深入,主要讨论点是如何制定迁移或仿真策略。毋庸置疑,科研数据如果脱离了环境(如源应用程序或相关流程)往往就会变得没有意义,而仿真技术可以帮助保持源环境的完整,并确保原始工具可以重新运行,这样一来就相当于将创建数据的过程进行了复制再现。利用这一技术,可对证实科学、同行评议数据和随相关论文一起发布的数据的保存问题进行初步探究。

数字艺术和电脑游戏

数字艺术和游戏由于包含互动元素和动态内容,并具有实时性的特点,通常比保存普通的办公文档或传统应用更具挑战性。数字文物在创建时由于通常会充分利用硬件的特性,和/或利用特殊硬件的特殊效果,导致对其的保存也充满挑战,而数字艺术和游戏的情况与之相似。具体的挑战包括:特殊的画面或视觉效果(例如,因为CRT缓慢的刷新率);一些应用程序(比如为了对媒体的复制保护)对软盘驱动器的不规范使用;使用先前尖端的技术,如3D功能。

电脑屏幕技术的进步还带来另一个问题:屏幕类型、分辨率和颜色深度往往成为至关重要的因素,因为同一个图像可能在屏幕比为4:3的老式(黑白)PAL电视机上,在1024x768分辨率的15寸CRT显示屏上,以及在现代的1920x1080像素分辨率的24寸宽幕TFT显示屏上呈现出的效果完全不同。此外,数字艺术作品可能需要非标准类型的音视频输入,而不同的电脑游戏也可能会需要不同类型的操纵杆。一些艺术作品和游戏还要求在与用户互操作时精准地控制时间,因为输入和屏幕响应之间的时间滞后,或对艺术作品和游戏过快的执行速度,都可能造成画面呈现的不真实,甚至使其完全无法使用。所以,不同于传统的数字对象,数字艺术和电脑游戏需要增添额外的技术元数据来对上述设置进行描述。

世界各地的博物馆花费数百万来收集数字艺术作品,而这些作品往往是动态的、交互式

的,它们广泛运行于不同的硬件和软件平台之上,对其进行直接的迁移是不可行的,因为数字艺术作品和它原来的环境是分不开的,而且将其迁移到当前平台还会花费过高的成本。因而博物馆和艺术画廊需要有能够长期保存并展示这些数字资产的通用策略。另外,许多存储机构开始接收到著名作家、政治家、研究人员或富有的捐助者的数字遗产,这种现象在美国尤为普遍,这无疑又扩大了个人存档接收品的范围。

除了保存数字文物本身带来的挑战之外,还有来自体制上的挑战,即对保存的资金匹配不足,而且现阶段的侧重点仍在于对传统对象的保存。同时,一些存储机构还抱怨实际的研究与数字艺术特定的保存需求间依然存在差距。除了技术问题,存储机构在其他一些问题上也缺乏指导意见,比如:在保存数字艺术和游戏的过程中什么样的损失是可接受的;未来的用户在访问一件数字艺术品或视频游戏时,他们期望得到什么样的体验。

还有另一个来自非技术领域的挑战,即怎样保存艺术家的“创作意图”——艺术家打算让他/她的作品成为“永恒的”还是“短暂的”(指是否考虑过技术有可能会过时)。依据不同的意图,艺术家可能会选择是否“允许”对其作品进行仿真处理(或一般性的保存)。正因为如此,收集机构会希望从艺术家那里拿到一个弃权证书,令其同意对作品的修改以便于长期保存。但什么样的修改是能够接受的,什么样的又是不能接受的呢?这个问题可能需要更多的元数据和文档的支持。为了给后期的保存提供方便,在迁移原始数字文物的时候可能需要特殊的技术设备,包括原始的输入和输出设备的适配器,以及原始媒体的阅读器。

对于视频游戏来说法律问题特别具有挑战性:游戏的版权可以从出版商那里取得,但这并不包括游戏中音乐的版权。通常情况下,很难找出一款游戏中所有的权利持有者,因为视频游戏行业的波动很大,一个公司一旦不复存在就会由其它公司接管,同时个人的版权也会被售出。

通过社区参与实现可持续的仿真

基于仿真的保存和访问策略并不仅仅只关乎是否有合适的仿真器或虚拟机。归档的原始环境由各种不同的软件组件构成,需要不断地对其维护,以保证与硬件驱动程序集之间的匹配,但这却是一个很严峻的考验(例如微软操作系统的保存)。此外,例如在创建磁盘映像的时候,如果原始媒介上的数字对象在稍后某个时间点变得可获得的话,还需要一些硬件工具,如电缆和适配器等。

现阶段有两类仿真工具包。一类是质量和可用性都参差不齐的种类繁多的开源硬件仿真器。这类仿真工具的出现通常是由游戏社区的需求驱动的,而正是因为总是游戏爱好者使用这些工具,一些存储机构只把其当做“黑客程序”。另一类是商业发行的仿真器和虚拟机,公司通常倾向于使用这类仿真工具,但在一般情况下,这些公司的业务范围并不包括长期保存。目前,很少有存储机构应用仿真技术来进行长期保存,大部分原因是由于这么做缺乏利益驱动。只有少数的存储机构,如荷兰的国家图书馆和国家档案馆,以及丹麦皇家图书馆对这种方法进行了投资。这些机构中又形成了三个不同的群体,但它们之间缺乏互动和相互间

的联系,显然在这方面需要不断改善:它们之间不仅要明确一系列共同特征而且要就指定社区的定义和范围达成一致意见,以便于质量保证,此外还要一起协作开发仿真工具。

不过大家都意识到了仿真是一种有效的方法,尽管它不能100%复制原始的系统和环境。现有可用的仿真工具,尤其是由游戏爱好者使用的那一类,都是程序员的个人兴趣和感情的产物,设计时并没有考虑到有数字保存的用途,因而这些工具可能会缺乏长期的稳定性,也无法提供适当的API以集成到保存框架。许多仿真器的背后都只有极少数人在做着维护工作,一旦他们对此失去兴趣,这些仿真工具就处于危险之中。此外,对不同计算机平台的热情也是不均匀的,这就不难解释为什么会有那么多不同种类的仿真器可适用于流行的机器,如Commodore 64,却没有或几乎没有适用于罕见的和不受欢迎的系统的仿真器。

研讨会结论

各个小组的讨论结果表明,目前出现的一些挑战只不过是保存领域即将到来的更大挑战的“前菜”。无论是游戏和数字艺术在技术上的复杂性,还是大数据带来的结构和组织方面的挑战,未来在把业务流程、研究数据和政府程序文件等这些当作“严肃”的保存对象的时候都会体现出来。因而利用仿真技术来解决保存数字艺术和游戏问题的方案,也应该适用于大部分其他的仿真案例,正如一位博物馆代表所说的那样:“数字艺术和游戏就是电子书的明天”。

直接的对象迁移战略已到达了极限,而仿真技术能够在一定程度上对这种极限进行突破。最起码在直接迁移对象不成功的情况下,仿真可以作为一个备选的方案。迁移和仿真并不是一种对立关系:仿真仅仅是迁移的一种非常有效的形式。仿真“迁移”的是与保存对象有关的应用程序、操作系统、(或最极端的情形)机器等。然而,要使仿真成为一种通用的保存策略,必须先解决一些如著作权、合理使用豁免等当今面临的复杂的法律问题,而且最好超越国家界限在国际范围内推广这样的解决方案。关于保存对象真实性(即允许一个对象的真实再现)的适用策略的讨论,则强调了对标准化的和全面的元数据方案的需求。

出席会议的从业者提出有必要制定一个仿真日程,并提炼优秀的使用案例来说明仿真的必要性和实用性。在过去的十年中,一些研究项目对仿真在长期保存领域的应用做了一定的探索,但仍旧没能创建出稳定可用的工具和服务。而对现有可用的仿真工具和工作流程需要继续在技术、可用性和经济效益方面加以改进,因为这些工具和服务的可持续性对其在数字保存和访问方面的使用是至关重要的。

编译自：<http://www.dlib.org/dlib/march13/vonsuchodoletz/03vonsuchodoletz.html>

(刘超编译,李文燕 吴振新校对)

【技术与工具】

Matchbox 工具集介绍

Matchbox 工具集是一个开源的工具集，能够为数字图书馆的各种质量保证工作提供决策支持。它可用于评估图像集合的品质属性，对比不同的版本或从集合中找出副本。Matchbox 基于最先进的图像处理技术，不依赖于光学字符辨识（OCR），比先前的方法更加灵活。

Matchbox 基于兴趣点探测技术，该技术已应用于多个可视化计算领域，主要基于对比图像中被探测和统计出的知觉突出点的特定属性。图像的固有属性，包括大小、亮度、旋转不变性等，使得这种方法能够完美地应用于异质文档集合的分析。Matchbox 从这些兴趣点中可以计算出一个视觉词汇和文档指纹，这种方法源于经典的文献检索，现在又被应用于图像的检索中。使用机器学习技术识别出一个集合所有图像的共用的兴趣点，可以计算出一个视觉词汇。接着通过计算这些单词，可以为每幅图像创建一个基于指纹的直方图。这种对一个图像的高度概括性的表示可以用于快速索引和搜索操作。在有效的机器学习算法基础之上，这些指纹可以用于在单个或多个不同的书籍集合中识别可以匹配的图像，一旦确定了匹配对，就可以计算出从相应的兴趣点到准确地缩放、旋转和排列图像的几何变换。这些步骤完成之后，就可以真实地对比这些图像，并计算出一个相似度数来。

在 Matchbox 中，这一方法的实现由一套小型工具集合来完成，因而相比于那些只能解决某个特定任务的程序，Matchbox 具有更大的灵活性，能够处理当前各种问题，比如找出一本书中的重复内容的页面、评估同一本书的两个不同数字版本之间的质量差异、将不同的集合片段组装成一个整体，以及其他诸多未知的问题。

编译自：

<http://www.openplanetsfoundation.org/blogs/2013-06-20-matchbox-toolset-webcast>

（李文燕编译，齐燕 吴振新校对）

【资料推荐】

《透视个人数字存档》报告简介

当子孙后代想要了解我们在 21 世纪早期的生活故事时，恐怕需要借助于大量的原生数字资源，比如网站、电子邮件、数码照片以及其他一些仅有电子格式的资源，因为它们现在已经充斥于我们的生活、工作、学习、娱乐等等各个方面。因此，实施原生数字资源的长期保存工作意义重大，不仅是现在就应该开始，而且更应该成为一种长期的持续性的努力。

然而,与书籍、论文等传统信息载体有着本质的不同,数字内容难以捕捉、组织和使用,而且非常容易因技术发展而被废弃。所以,专门的收藏机构在实施数字资源长期保存之前要进行一系列周密的筹划,并制订一套缜密的规划。同样的,对于想要延续其个人或家庭数字记忆的个人来说,实施个人数字资料的长期保存也需要一系列的准备工作。

数字时代,每个人都需要成为其个人数字资料的档案保管员。在这方面,国会图书馆从2007年起就在其网站上提供一些基本的指南和建议,几年来内容日益丰富和专业,尤其是近两年启动的数字保存博客平台——The Signal,涵盖了以下三个方面的内容:

- 讨论如何应用国会图书馆网站上“保存你的数字记忆”部分列出的指南;
- 记述在个人数字保存实践中所遇到的一系列问题;
- 描述国会图书馆开展的一系列推广活动,包括如何与其他机构和个人合作以促进个人数字存档的开展。

此份报告正是选取了上述三个方面中实用性最强的一些文章,以期能够为个人以及相关机构的个人数字存档工作的开展提供有力的支持。

报告分为三个部分。在个人数字存档指南部分,主要包括数码照片、扫描图像、手机短信等的保存方法,以及对如何选择图像的分辨率、云存储服务的使用、数字保存行动规划等具体方面的讨论。在个人数字存档的实践部分,不同的作者描述了自己或家庭所开展的数字存档行动,以及在这个过程中所经历的一些困难和挑战,向人们分享了宝贵的经验教训。另外还有一篇是某位学者对个人数字存档规划的调查结果的展示。在个人数字存档的推广部分,除了对国会图书馆举办的保存周活动和研讨会,以及与ALCTS、PLA合作开展的一系列活动的记述外,还有个人数字存档工具包的发布、堪萨斯城公共图书馆专家的访谈、青少年对数字保存的认知调查,以及开展个人存档教育所面临的挑战等方面的内容。

详细内容参见报告原文:

http://www.digitalpreservation.gov/documents/ebookpdf_march18.pdf。

编译自: <http://www.digitalpreservation.gov/multimedia/publications.html>

http://www.digitalpreservation.gov/documents/ebookpdf_march18.pdf

(齐燕编译,刘超 吴振新校对)

社区数字存档相关资源

国会图书馆的The Signal平台上有很多关于个人数字存档的资料,尤其是关于公共图书馆如何开展个人存档的推广和培训的主题。但是,公共图书馆及其相关组织也都应意识到社区数字存档的需要和重要性。

社区藏品可能会包括原生数字或数字化的图像、音频、文本,以及其他类型的可以记录地方历史的文化遗产材料。通过这些社区档案,社区成员们能够更好地了解他们生活的环境,

而且人们往往渴望能够贡献出自己的个人材料和藏品。

正如美国图书馆协会所建议的,对于公共图书馆来说,将数字特藏纳入未来战略规划将会有助于未来这些工作的开展和顺利进行。另外,目前有以下10条资源分别涉及到了社区数字存档的发展、管理及获取方面,值得公共图书馆相关人员进行参考:

美国数字公共图书馆的数字中心项目 ([Digital Public Library of America Digital Hubs Program](#)), 该项目将美国40多个州/地区的数字图书馆和无数个大型数字图书馆组成了一个国家级的合作网络,为终端用户把全国各地的数字化的内容汇集到一个单独的接入点,同时也为开发者提供了一个开放平台。

休斯敦地区数字档案馆 ([Houston Area Digital Archives](#)), 经由Internet Archives访问,人们可以自由地在线获取到休斯顿公共图书馆和其他社区档案馆丰富的馆藏。

丹佛公共图书馆社区创建计划 ([Denver Public Library, Creating Your Community](#)), 丹佛公共图书馆的使命是向人们提供信息、思想和经验,让人们能够获得更大的乐趣,丰富个人生活,并增强社区的凝聚力。其社区创建计划 ([Creating Your Community project](#)) 致力于引导社区利用各种技术、合作关系以及宣传推广来以新的方式记录和分享其社区的历史。

弗吉尼亚州阿灵顿县公共图书馆的阿灵顿社区存档 ([Arlington County \(Va.\) Public Library, Arlington Community Archive](#)), 该档案集合包含了阿灵顿县重要的历史文档,其中涉及到了该地民众生活、组织机构及一系列社会问题。研究人员可以找到个人文件、照片和有关当地组织、社团及协会的档案资料。

社区档案及遗产工作组 ([Community Archives and Heritage Group](#)), 这是一个国家级组织,致力于支持和促进英国和爱尔兰社区存档的工作。该网站提供了大量的相关建议和指南,以及个人档案的目录。

马萨诸塞州阿默斯特市琼斯图书馆,数字阿默斯特 ([Jones Public Library, Amherst, MA, Digital Amherst](#)), 通过收集能够反映阿默斯特市民的独立精神、创新能力以及文化底蕴的所有图像、多媒体和文档资料,生动地重现了阿默斯特的历史。

开始存档 ([START AN ARCHIVES!](#)), 是宾夕法尼亚州费城的一位档案保管员撰写的博客文章,论述了在创建社区存档之初所进行的尝试和所遭遇的困难。

好时公司的(宾夕法尼亚州)社区档案馆 ([Hershey \(Pa.\) Community Archives](#)), 当所有的合作伙伴意识到好时公司的历史档案正处危险之境时,他们开始了存档工作。今天,这一档案馆不仅是好时公司、好时信托公司、好时娱乐和度假村公司三家公司共同的档案室,同时也是社区及地方机构的档案的仓储库。

[History Grand Rapids.org](#), 这是美国大急流域历史委员会 (Grand Rapids Historical Commission) 的一项计划,旨在收集当地家庭私人收藏中的图像、公共机构和私营企业的本地档案资料,以及其他一些当地资源,以期能够反映和揭示19世纪中期到20世纪期间大急流域的家庭生活和社区生活,以及人们的工作、旅游、教育及娱乐的各种方式。

麻省大学的波士顿分校路演记忆 ([The Mass. Memories Road Show at UMass Boston](#)), 这是一个基于事件的公共历史项目, 内容是对马萨诸塞州人民分享的家庭照片进行数字化和存档。大学档案馆的员工与当地的规划团队共同组织了多次免费的公共活动, 邀请市民将自己的家庭照片带到现场进行扫描并缴存到数字档案库中。

编译自 :

<http://blogs.loc.gov/digitalpreservation/2013/06/10-resources-for-community-digital-archives/>

(齐燕编译 , 刘超 吴振新校对)

TIMBUS 第二期新闻通讯发布

6月28日, TIMBUS项目发布了其第二期新闻通讯, 主要内容包括:

- 使用案例: 电子健康业务流程保存
- 使用提取工具和推理机来 (填充) 构建和 (访问) 获取语境模型
- 推荐书籍: 《数字资源长期保存的法律问题》
- 精选文章: 《确保基于 Web 服务的业务流程的可持续性》
- 解决第三方对软件托管的依赖性问题
- TIMBUS 实施业务流程保存的方法

详细内容参见:  [TIMBUS Newsletter V2 I1 \(1.57 MB\)](#)

编译自 :

<http://timbusproject.net/news/news/214-the-timbus-project-has-published-its-second-newsletter->

(齐燕编译 , 刘超 吴振新校对)