

## 目 录

<b>【重要文献摘译】</b> .....	1
《分布式数字资源长期保存指南》摘译 .....	1
《New Roles for New Times: Digital Curation for Preservation》报告摘译.....	6
<b>【动态追踪】</b> .....	13
“Screening the Future 2011”会议内容介绍 .....	13
欧盟委员会数字资源长期保存研究 2013 展望 .....	15
描述数字报纸生产的 CRL 报告 .....	18
<b>【信息扫描】</b> .....	21
视听资源存档的新战略和挑战 .....	21
地球空间数据长期保存资源中心网站发布 .....	22
新的数字资源长期保存项目- APARSEN .....	23
新的数字资源长期保存项目- TIMBUS.....	24

## 【重要文献摘译】

### 《分布式数字资源长期保存指南》摘译

《分布式数字资源长期保存指南》（以下简称《指南》）由 MetaArchive 的众多成员合著，是介绍成功合作策略以及有助于文化存储组织互利互惠合作的新模式的系列著作之一。

分布式长期保存是文化存储机构的新兴实践领域，该书适用于这一领域的广泛主题。复制和传播可以保证无限的无损保存材料，但是要建立有效的组织和技术过程来运作这种形式的数字资源长期保存却有些困难。机构需要参照一些实践案例，以了解怎样以易于管理的、低成本的方式完成这项任务。



创作过程中，《指南》考虑了众多的受众范围，包括图书馆员、档案馆员、学者、长期保存人员、技术人员、法律人员以及管理人员。利用该指南，读者可以从理论和实践上了解分布式数字资源长期保存的新兴领域。

《指南》有意将每部分章节编排成既可以独立成篇又可以与其它部分合成一体，允许读者根据自身需求选择其阅读路径。

本期我们将摘译其中的 DDP 架构内容，其余各章节将在下期继续刊载。

#### 第一章：DDP 架构

本章概述了 DDP（分布式数字资源长期保存）网络的基本技术架构，主要关注于 PLN（the Private LOCKSS Network）及其核心元素。接着重点阐述了一个支撑分布式数字资源长期保存实践的通用开源技术基础设施：PLN（the Private LOCKSS Network），揭示了 PLN 架构的基本元素，并详细说明了它们是如何相互作用而形成长期保存的网络。

本章提供了后续深入探讨 PLN 框架的基础，并且就其本身来说，值得推荐给所有读者。

#### 什么是 DDP 网络

分布式数字资源长期保存方法论认为可靠的长期保存系统必须将数字文档的复本分配到不同的地理位置。同样，系统也必须在这些地方保存（但不仅仅是备份）这些文档。所有的 DDP 网络因此可以共享特定的资源，包括组成这些网络的多个保存站点。实践证明，在选择站点时需要遵循下列原则，这些原则将减小影响网络的错误发生机率：

**保存相同内容的站点半径距离应该大于 75-125 英里。**这有助于确保当物理环境遭到破坏时两个保存相同内容的站点减少其灾难性的损失。

**保存站点应当分布在典型自然灾害（如飓风、台风、龙卷风）区域之外。**如，如果网络建立在美国墨西哥湾，那么其保存站点应该建立在墨西哥湾区域之外，以避免将其所有的保存复本存放在有飓风风险的地区。

**长期保存站点应该跨越不同电力网分布。**DDP 网络应该考虑支撑其基础设施的电力网分布，并避免将任何内容的所有复本放置在同一电力网区域。

**长期保存站点应该在不同系统管理员的监管之下。**为安全起见，所有长期保存的复本都不能被任何个人或团队访问；相反，每个长期保存站点的控制和监管应该由本地的站点实施操作，以确保网络的内容不会因某一站点人为因素而受损。

**保存在不同站点的内容应该可以进行实况查询，并且应该定期检查字节衰减或其它问题。**因为数字内容生性就是脆弱的，没有任何外在灾难情况或其它易于识别的触发事件时，就会在文件层面中发生退化。因此，应该定期评价内容。如果有迹象显示保存复本已经改变了（如校验和计算不匹配），那么需要将该复本与其它复本比较。一旦经过比较得出权威的版本，退化的文档将会被取代为权威版本的复本，或者如果原始版本仍然可用于再次摄取，则从原始文档中创建复本来取代退化的文档。该过程同任何影响文档的其它过程一样，应该在保存系统中永久记录，以帮助验证保存文档的权威性。

**根据上述规则，应该至少复制三份内容。**这是在面临文档退化/字节衰减时应该保存的权威文档最小复本数。当一个文档不再匹配其它文档时，则有两个文档（除了校验和信息以及其它有关文档的记录）可以相互比较以恢复上面所说的权威版本。只要同时遵循其它规则，大量复本将会分布在跨越足够多的不同地理位置上，确保复本可以在大多数可预知的灾难环境中保存下来。

值得注意的是，现在越来越流行的云计算服务器同样应该遵循这套原则。支撑云计算的服务器也有其物理地址，如果云环境用来存储保存复本，那么就需要考虑其服务器的地理位置。如果网络选择在云环境中存储多个保存站点，则应确保服务器的地理位置满足上述原则。

不管 DDP 网络采用什么样的技术基础设施，网络都需要执行三大任务：内容摄取/收割、内容监管、内容检索。下文简要概述了每一项任务，但是这些程序在不同的技术基础设施中可能根本上都是不同的。

### 内容摄取

通常主要利用两种方式中的一种将内容迁移到 DDP 网络中。网络可以通过站点摄取内容，或者直接通过从存储媒介中就地收割。基于站点的摄取为所有缓存提供了从相同服务器收集内容的方法，而不需要任何硬件或存储媒介的交易。DDPs 由地理上相距甚远的保存服务器组成已被证明是迄今为止最流行的选择。不管框架是什么样的，长期保存网络必须有一种方法来验证所摄取/收割内容的每个复本都和原始内容以及其相互之间相匹配。

DDP 网络，像任何保存系统一样，必须考虑到数字内容经过了高水平的有意改变。个别文档可能会被编辑或另存为权威的版本。文档将会被添加进资源集合（甚至是“封闭”的集合）中。DDP 网络可能想要通过将定期重新摄取长期保存内容列入计划中来预期诸如此类的变化；明确相关机构的职责义务以监管诸如此类的更新和附加物，并在网络中定期共享它们。

### 内容监管

每个长期保存服务器都必须参与某种形式的内容维护和监管。维护和监管可能是通过基于网络的投票（如下面描述的 LOCKSS 案例）、通过迭代校验和比较、或者通过其它相关机制。长期保存网络应该定期向相关机构报告其网络服务器状态的详细情况以及其保存的内容情况。

### 内容恢复

当相关机构需要用保存的复本取代本地复本时，DDP 网络必须为其检索内容的方法。这必须在文档层面和资源集合层面都是可行的，并且长期保存网络应该可以在合理的时间范围内传递权威的文档版本。

本章的下一个部分探索了一个技术方法——the PLN，将其作为解释这些元

素如何运作的案例。

## 什么是 PLN?

迄今为止，在文化保存团体（包括图书馆、档案馆、博物馆以及其它文化保存实体）中，许多 DDP 解决方案都依赖于 PLN（Private LOCKSS Network）网络中的 LOCKSS 软件。PLN 是一组封闭的、在地理上分散分布的服务器（在 LOCKSS 术语中称为“Cache 缓存”），它们被配置用于运行开源的 LOCKSS 程序包。该软件利用因特网将缓存之间以及缓存与保存了有益于长期保存网络内容的站点联系起来。

在 PLN 中，每个缓存都有相同的权利和职责。没有装备特殊能力或特征的最重要缓存之分。缓存运行后，即使其失去了与中央服务器的联系也可以继续运行。这种对等的技术结构具有强健的抗故障能力。如果网络中的任一缓存发生故障，其它缓存则可接管它。如果缓存崩溃了，网络中的任何其它缓存都可以用来修复它。由于所有的缓存都是相似的，维护网络的工作实际上分配给了网络中的所有合作者。这是分布式长期保存方法强大的优势之一。

PLN 的 LOCKSS 缓存执行一组面向长期保存的功能：

- 它们在本地的磁盘上摄取、保存提交的内容。
- 它们进行投票，比较所有内容的缓存复本，以达成关于网络中每个缓存内容单元的真实性和准确性的一致意见。
- 它们修复那些通过网络轮询过程后被认为崩溃的内容。
- 它们从原始位置重新摄取内容以发现新的或改变的片段，并在原始版本的之外同时保存所有的修订。
- 它们检索并为有需要的授权者提供所保存内容的复本（如，如果最初提交内容的捐赠者本身的主要文档由于自然的、技术的或人为原因而变得崩溃或受到损坏，它们可以为其提供保存内容的复本。）
- 它们提供信息、缓存管理工具和 LOCKSS 后台程序用户界面以辅助监管和报告活动。

下文简要介绍了以上这些活动，并在后续章节中进行了详细阐述。

## PLNS 怎样“保证内容的安全性”：功能和组件

### PLN 的组成部分

PLNs 由一组独特的、地理上相距遥远的长期保存站点组成，每个站点都运行一个安装了 LOCKSS 软件的服务器。每个长期保存站点作为一个内容仓储提供服务，其内容是通过基于网络的协议摄取的。PLNs 是安全的网络并且通常作为一个暗箱存档运行，其内容一般不对公众开放。每个服务器都以相同的方式配置。所有服务器在权利和职责方面都是对等的。网络中的缓存通过网络相互连接；这些服务器相互之间通过 TCP 协议，即 LCAP 定期交流其所保存的内容，LCAP 的设计非常独特，难以受到攻击。

目前，LOCKSS 团队建议将在 LOCKSS 网络中保存的内容复制 7 次。如果网络包括多于 7 个站点，则不是每个站点都需要存储每份内容；依据上述 DDP 原则，内容可能会跨越不同成员分布。

和许多长期保存战略（尤其是非 DDP 系列）不同，LOCKSS 不仅仅依赖于长期保存的校验。LOCKSS 环境更为重要的是对保存复本的持续监管。LOCKSS 从网络中摄取内容。在摄取时，LOCKSS 软件会在这些保存站点间实行投票，使用 SHA-1 校验和摄取任何一个特殊的数字文档，以保证其所保存的文档与那些数字文档具有相同的版本。

在内容摄取之后，LOCKSS 缓存定期重复进行投票，以及时发现内容复本的退化或改变。多份内容（至少 7 个复本）使得网络可以通过投票识别哪些版本是权威的、哪些不是（如发生了字节损坏或其它退化）。此外，缓存定期继续重新访问最初摄取内容的站点（只要该站点可访问），无论在什么位置发现内容的更改或更新，缓存都会将这些新的、改变的文档作为独立的版本摄取并保存起来。不管在哪发现内容的删减，缓存都会记录下这些，并且不会删除先前的版本。

这种再次摄取的特征具备了大多数长期保存技术所忽略的东西：数字媒介的好处之一就是一旦出版或生产出来了，作者可以对其进行更新或更改。通过定期重新摄入内容，LOCKSS 就可以不仅仅只是保存一个资源快照——它可以随着时间的推移，保存资源不断变化的形式。

每个 LOCKSS 缓存都执行网络中的核心业务，包括内容摄取、维护、监管和恢复。后续的章节将深入介绍这些程序，详见下期文献摘译。

## **LOCKSS：字节层面的长期保存和格式迁移**

在底层，LOCKSS 软件为所有文档类型和格式的数字对象提供字节层面的长

期保存。字节层面的长期保存是所有长期保存解决方案的基础元素。实质上，它确保了组成给定文档的所有 0 和 1 字节仍然是完整的，因而维护了后续需要获取和使用的文档的完整性。所有的 PLN 都提供字节层面的长期保存。

LOCKSS 同样可以提供一组特定的服务，使得所保存的文档在将来可以获取和使用（包括标准化和迁移文档活动）。2005 年 LOCKSS 首次在网络中创建并测试了格式迁移策略。现在许多 PLN 都开始为濒危的或废弃的特定文档类型创建迁移途径。一旦 PLN 为特定文档类型创建了迁移途径并在 LOCKSS 环境中实施了，则其也可以被其它 PLN 使用。这为不断发展的开源 LOCKSS 使用者和开发人员团体提供了又一额外好处。

## 结论

所有的 PLN 都不是完全相同的。一个 PLN 可能由少量的保存相同内容的缓存组成；或者由大量的保存不同内容的缓存组成。一个 PLN 可能在中心仓储中建立插件，或者在每个相关机构中维护一个仓储。PLN 可能依赖于斯坦福大学 LOCKSS 团队来维护其插件和档案单元列表，或依靠自己来维护这些资源。安装 PLN 有多种方法，下一章我们将介绍一组案例，这些案例通过可操作的、技术上的决策（这些决策是依据其成员的需求而做出的）自定义了其网络。

信息源：

<http://www.metaarchive.org/GDDP>

编译：何莉娜 校对：何莉娜 吴振新

## 《New Roles for New Times: Digital Curation for Preservation》

### 报告摘译

#### 概要

二十一世纪，ARL（the Association of Research Libraries）图书馆一直在进一步探索并采用一系列新角



色来服务于研究机构、研究人员、学者和学生。现在，ARL在研究型图书馆关键角色方面组织发布一份新的报告集的时机已经成熟。《新时代新角色》系列（New Roles for New Times）将识别和描述新兴角色，并总结图书馆成员在发展新角色和提供服务中的早期合作经验。每份报告描述一个新角色，并明确表述受新角色影响的读者以及不同的机构在体验新角色过程中所获得的好处。报告集强调了现有工作，报道了研究者的最新发现，并且提供了对趋势、最佳实践以及关键问题的分析。《新时代新角色：“数字资源长期保存”》报告中探索了研究型图书馆如何尝试在新的研究知识和信息产生链中增加自身价值的问题。

数字资源长期保存，是指人们在数字资源生命周期内为维护 and 增加资源的价值而采取的措施，包括在创建数字化资源的过程中所使用的工作流程。数字保存的重点在于“开展一系列必要的管理活动，确保在必要的时候可以持续获取数字资源”。报告强调了这些活动的共同点，尤其关注于数字资源长期保存如何有利于保存多机构共享的数字化资源。

该报告建议，研究型图书馆重新定位为整个校园内充满活力的知识分支，用以提供数字内容的专业指导和知识（不管其是在什么地方创建或维护的）。图书馆不能再寄希望于研究人员和学者到馆咨询和请求援助，而应该发掘新的方式，主动为他们提供不受地域限制的服务。研究和学习活动越来越多的在机构内和机构间进行，越来越具有合作性、跨学科性、国际化和虚拟化的特征。为了保持活力，图书馆必须调整其服务提供方式来适应新的环境。

在此过程中，报告记录了一系列目前图书馆在数字化领域承担的、有发展潜力的新兴角色。报告强调并讨论了在研究型图书馆社区中相对新趋势（包括研究型图书馆一直以来为大学提供的外包服务）的潜在影响。最后，报告提出了在数字资源长期保存领域的一系列合作案例，并且思考了研究型图书馆之间为实现共同目标而建立约定的积极影响。

报告作者断言，研究型图书馆未来的一个重要部分是在多个数字资源长期保存合作机构间构建可发展的网络基础设施并提供服务。如果各机构选择“独立发展”策略，采用异构的、价格昂贵的、最终无法持续的基础设施，则会有被孤立的危险。报告给读者提供了一个完整的参考，主要关注数字长期保存新兴实践和研究型图书馆如何培养执行实践能力，以确保其上层机构继续实施创建、分发和



保存知识的核心任务。

### 一、研究型图书馆的新角色

当前，图书馆较少关注旧的公共技术服务范式，更多关注于基础设施、内容和服务的构建。在这三者中，基础设施包括设备、技术和适用于组织的专业知识。内容指的是图书馆可以获取到的所有信息资源，包括不断增长的、产生于校园的数字馆藏以及已注册或购买的电子资源。服务包括传统的信息服务和更多虚拟领域中的新兴服务，例如信息生产、获取和传播以及数字资源长期保存（如“嵌入式研究馆员”的角色）。图书馆这三个部分——基础设施、内容和服务——都直接受到研究型图书馆所经历的文化和制度层面的范式转换的影响。

#### 1. 网络基础设施和合作

网络基础设施包括图书馆旨在创建的整个研究环境。这将使研究人员能够有效地沟通，促进其创建和传播工作的能力，而不必在每次开始一个新工程时都要开发相应的工具。图书馆间进行合作时希望建立一个稳定的网络基础设施环境，在产生于社区的解决方案上进行投资，而不是坚持建设个性化的工作流和系统。与其他机构协同工作，将会提供更加强大的基础设施和馆藏集合，这需要采用互操作的工作方式而不是依赖于单方努力。

#### 2. 内容采集和托管

研究型图书馆正在通过提供综合可靠的数字资源长期保存服务来扩大使用范围，这支持了大学里知识内容的生产活动。大学通过托管大范围的内容来达到长期保存的目标，其中包括数字化馆藏、已注册的内容、采集物（例如，Web 存档和手稿集）、研究数据和其它产生于大学的基础资源材料、大学知识产出的电子印制品（例如，出版物、ETDs）、教材、其他多媒体资源和抓取的数字化讲座系列、座谈会和其他校园活动。

#### 3. 数字出版和数字资源长期保存服务

内容制作以及数字资源长期保存的角色补充了内容采集和托管的角色。图书馆和学术界之间的合作关系非常重要。图书馆是提供大学电子出版的有利场所，其与尚存的大学出版社存在合作关系。到目前为止，图书馆的电子出版实验是相当成功的，特别是在开放获取的期刊生产、电子出版物出版以及作为数字社会科学资源出版的学术出版领域。

## 二、研究型图书馆员的新角色

当前研究人员、学者、教师和学生的信息管理需求在不断的改变，为应对这种不断的变化，图书馆员至少要承担七种不同的新角色。

### 1. 内容采集和权限授予者

除了建立的内容要有较高的质量、相关性、馆藏调整和维护的综合平衡性外，研究图书馆现在必须在资源采集时设法建立新标准，如数字文件稳定性、可用性以及可靠性的新标准。在法律顾问的帮助下，图书馆员和档案保存者能够创建和检查数字采集和捐赠的有关政策，确保给图书馆授予足够的权限来适当地进行数字资源长期保存。解决不断增加的知识产权问题，不仅需要法律专业知识，也需要信息管理专业知识，因此越来越多的研究型图书馆雇佣权益管理专家而不再依靠学术型大学的总顾问。

### 2. 学习领域的教学合作者

该角色包括物理的和虚拟的两部分，二者都直接受到日益增长的数字化的科学、社会科学和文化遗产资料的信息管理需求的影响。图书馆员越来越多地参与到校园的整体环境中，包括构建课堂教学设计的伙伴关系、通过在物理的和虚拟的环境中提供参考/帮助服务并在本地社区中提供持续教育的拓展培训。

### 3. 观察者/人类学家（关注于信息生产者和消费者）

图书馆员为学习信息生产者和消费者的习惯，正试图花费时间融入他们的校园环境，即“在本地人中生活”。如果研究型图书馆希望满足大学未来的需求，就必须了解它们的组成成分——管理者、员工、教授和学生。随着技术不断地快速发展，研究机构的需求同样也将迅速转变。

### 4. 系统构建者

该角色不是指服务器机房里的系统管理，而是有关设计和构筑物理空间（如信息共享空间information commons）、虚拟空间（如网站、数字资源库、以及其他信息系统）、分散的系统建设活动（如注册、元数据创建和版权研究）。同时还包括为学者提供新的机制——以电子书和电子期刊以及更加专业的数字资源形式——发布他们的手稿和文章。

### 5. 内容的生产者和传播者

图书馆员过去关注的内容保存——包括选择、管理和长期保存——是研究型

图书馆当前工作的典型组成部分。新兴的角色关注于数字资源长期保存或者数字化对象和馆藏的生命周期管理。角色包括明显不同但是相关的两部分：访问获取和长期保存。在面向访问获取的领域，图书馆要比从前管理更多的信息，最终帮助用户获取他们寻找的内容。数字资源长期保存也要满足对数字内容长期的需求，这与原先的方式有显著区别。这是针对日益增长的有关数字内容的稳健性、真实性和可持续性的需求而演化产生的。

## 6. 组织设计者

组织设计能够提供必要的灵活性和实验氛围来转变体制结构，以适应其新的领域范围。要维持成功的、正转型的研究型图书馆，关于新的组织形式的实验（包括分布式和分散的机构间的合作模型）是必须的。进行这个实验的图书馆员通过组织设计的方法帮助在职的职员发现适应和学习新技能的方法。同时他们还正在努力研究如何重新定义图书馆的劳动分工，更好地适应与其角色相称的数字和物理愿景。

## 7. 合作网络的创建者和参与者

研究图书馆如果想为学术交流建立一个可持续发展的框架，就必须形成一个社团，在其中他们能够管理自己。目前图书馆在已经获得资助的项目期间建立了合作实体。在此意向的支配下，机构一直在努力实施长期的、合作的、可持续的实践来支持他们的共同需求，包括技术开发、仓储管理和长期保存。图书馆员开始认真协调机构内和机构间的联盟，并维护各个部门或机构共同努力制定的有关角色和责任的文档。他们也在学习如何提供一种接合方式，用以帮助在不同设置的机构间建立健康的合作关系。

## 三、数字资源长期保存的合作战略

该报告简要介绍了一系列新兴的、由ARL图书馆创建并参与的数字保存计划。本领域的很多path-breakers都参与到项目中，给项目带来了美好的前景。他们同时也暗示了正在形成的新愿景，包括机构间的合作可能对长期保存起到的潜在的巨大作用。机构列表：the National Digital Stewardship Alliance, Chronopolis, Florida Digital Archive, MetaArchive Cooperative, Council of Prairie and Pacific University Libraries (COPPUL) Private LOCKSS Network, Alabama Digital Preservation Network (ADPNet), HathiTrust, DuraSpace, Data-PASS, and the

University of California Curation Center, California Digital Library。

### 建议

针对任何数字资源长期保存的情况，协作的或基于社区的数字长期保存方式很可能都是更有效而持久的。然而，合作战略和社区管理资源的有效使用需要本地投资和能力建设。下面的建议反映了这一现实：

•**停止等待并开始积极参与本地业务：**通过另外授权的连接方式、数据中心以及本机构中的个别教授和研究者，与国家自然科学基金资助项目办公室、教务长、研究副主席、院长和系主任建立关系。正如校园办事处的政策伙伴关系，全体合作者的伙伴关系是数字长期保存机构的有力推动者。

•**生产周期要求：**考虑通过提供电子出版服务（电子期刊、电子图书、电子会议记录等）和保存数字档案、数据集、数码艺术品及网站等，来建立多媒体和其它数字资产的新服务。其中重要的一步是把这些服务置于组织内部，并抓住它们出现的机会。

•**启动再培训和重用工作人员：**图书馆需要思考将数字长期保存作为图书馆核心职能，并投入相应的财政和其他资源。对相关的工作人员提供深入的、长期的培训计划。在图书馆中引入专家。维护图书馆目前或即将参与的数字保存服务事业的日常事务。

•**尽可能做一个实干家，而不是一个调解人：**将我们的未来交给外部团体是没有必要的，研究型图书馆已充分表明：它有可能进行合作，以一种经济可行的方式来管理数字内容、维系人员和维护技术基础设施。

•**考虑数字长期保存合作：**走近合作伙伴，与他们一起选择一个项目或一个有意义的技术基础设施，并开始建立长期合作关系。

•**实施合作盟约：**协同工作并与其他部分机构建立一个稳步的可持续网络基础设施。早期的经验和目前的趋势表明：多个机构间的协作方式提供了一个成功的组织化环境，以满足庞大的不断转变的挑战。

如果研究型图书馆在整个艺术、人文、社会科学、科学和工程领域都是研究发展和网络基础设施管理的积极参与者，那么其将大有收获。如果意识到自己在研究环境中新兴的角色——人类学家、数字内容的联合生产者和传播者以及研究

合作交流和学术目标管理的系统构建者，那么研究型图书馆和其雇员将成功地转化为二十一世纪早期全球数字研究界高度受重视的服务实体和合作伙伴。

## 附录A：科学与工程的网络基础设施和数据保存

信息专业人员在网络基础设施环境中管理和保存的内容，包括科学、工程、定量社会科学，其角色可以概括如下：

### 1. 研究数据和信息对象的创建者

- 发展和应用数据模型、本体论和分类学
- 初始捕获和管理仪表数据
- 捕获研究团队的网络通讯工具的内容

### 2. 研究数据和信息对象的长期保存和管理者

- 收集/采集、描述、出处跟踪、访问和重用、集成、维护以及数据和信息的数字长期保存
- 提供计算资源和数据存储

### 3. 虚拟社区协作者

•管理协作网络空间，将参与到社区中特定研究领域和研究项目的研究人员连接在一起（即打包并提供相关的内容给社区，设计基于web的社会网络，帮助建立人们之间的关联以及方便查找信息）

角色的转变更早地将信息专业人员，包括图书馆员，带进了数据和信息生命周期，他们的工作重点在于增加知识生成过程中的价值，而不仅是在周期结束时管理信息产品。Anna Gold已经很好地合成了一些新的服务和活动：

“以丰富而强健的方式支持数据重用和整合的关联数据，需要充分了解数据的出处、本体的发展、专家标注和分析的文件。由此，这些活动激活的服务将包括可视化、模拟、数据挖掘和建模以及知识表示和抽取的其他形式。”

报告下载：

[http://www.arl.org/bm~doc/nrnt\\_digital\\_curation17mar11.pdf](http://www.arl.org/bm~doc/nrnt_digital_curation17mar11.pdf)

信息源：

<http://www.metaarchive.org/reports>

编译：李丹丹 校对：何莉娜 吴振新

## 【动态追踪】

### “Screening the Future 2011” 会议内容介绍

3 月 14 至 15 日，荷兰视听研究所（the Netherlands Institute for Sound and Vision）举行了“Screening the Future 2011”会议。会中，22 位专家做了关于数字化、 workflow、数字保存和可持续性、资金、保存策略等主题发言。下面将简要

介绍会议的主旨发言：

1. 大数据、关联数据（Paul Miller）

主要探讨目前不断出现的“大数据”（Big Data）以及“大数据”之间关联的问题。在互联网范围中，亚马逊、谷歌和Facebook的数据中心的有益经验已被用于传统的组织中，同样也可以被应用到欧洲的视听存档中。如果学术界、政府和企业认识到关联数据的价值，那么语义Web十年之久的承诺也将显示出实际的价值。

2. 视听存档中从生产者到消费者的可靠性、真实性和信任（RAT）管道（Seamus Ross）

用户关注于：数字视听仓库存储的数字内容的可靠性、真实性和安全性；数字视听仓库能够确保资源从生产者到消费者的长期完整性和真实性。

近十年的调查显示，可信数字仓库的核心存在多种不同的属性，他们在一定的基准、准则、指标和标准的范围中被具体化。

3. 扩大和缩减规模：尽可能的利用我们所拥有的资源（Brewster Kahle）

基于廉价的数字化馆藏技术和互联网接入技术，本部分主要介绍了the Internet Archive将电影、电视节目、音乐唱片数字化，然后利用这些资源提供免费开放获取服务的实际经验。目前，本项目馆藏包括40万部电影、80万个录音以及近100万小时的电视节目。项目组成员发现：如果被给予体验这些资料的机会，成千上万的人会很愿意参与其中。

4. 保存多少副本？（David Rosenthal）

数字资源长期保存中，考虑到保存馆藏（例如音视频存档）的经济性，如何能够保存尽可能少的复本并达到合理安全的目的呢？这很难提供一个确定的答案。试图提供一个具体的数字又会引出另外的问题。“合理安全”真正的含义是什么？为了实现“合理安全”需要克服什么样的威胁？目前可用的技术又有哪些？如何知道他们能够很好的抵抗威胁？如何选择最低成本损失概率的解决方案？本部分仔细探究了这些问题的具体细节，但没有提供具体回答，只是提供了很好的规则来处理这些问题。

5. PrestoCentre：保持视听内容长期有效（Daniel Teruggi）

欧洲共同资助的项目PrestoPRIME不仅强调基于文件视听馆藏所面临的新闻

题，而且试图解决保存支持的问题。PrestoPRIME提出：建立一个具有强大能力的视听资源保存中心——一个由五大存档机构建立的正式组织，其机构间协同工作以完成Presto项目。PrestoCentre作为一个新组织将继续其工作，不会停止保存支持行为。

#### 6. 建立数字资源长期保存的工作流

本部分的目的在于帮助小型和大型档案馆、图书馆和博物馆保存电影、录音带和录像带，构建新的数字化工作流，以及更好的理解他们在不同解决方案之间的选择和取舍行为。本部分的主讲嘉宾主要是先进档案机构和数字化服务供应商的代表，他们将介绍其一直在做的项目。

#### 7. 编码转换介绍：工具和过程（Skip Elsheimer, Dave Rice）

随着时间的推移，数字资源格式也在不断演变。本部分讲解编码转换的基础知识、实用工具、策略以及有效提供数字视听媒体馆藏存取服务的挑战。内容主要涉及到检查基于软件的工具和应用程序，识别易出错的地方，评估无损、有损转换方法，质量控制和验证。

#### 8. 创建数字保存方案，明确风险和估计费用（Matthew Addis, Richard Wright）

本部分概括叙述了Presto、PrestoSpace和PrestoPrime的经验教训，并向与会者介绍了通过PrestoCentre中心有效获取工具和资源的方法。

信息源：

<http://screeningthefuture.eventbrite.com/>

项目视频下载：

<http://www.prestocentre.eu/news/2011/03/18/video-prestocentre-official-launch>

编译：李丹丹 校对：何莉娜 吴振新

## 欧盟委员会下一个框架计划的数字资源长期保存研究主题

五月伊始，欧盟在卢森堡开展了一次研讨会。研讨会的目的是帮助欧盟委员会确定下一个框架计划（起始时间将为 2013 年）下的数字长期保存研究的重点。



与会人员议定了数字资源长期保存现状的一些共同主题，展望了它在 2020 年的未来蓝图，提出了在下一个框架计划下重点研究领域的建议。

与会者一致认为：由于欧盟的大力支持，某些领域，在工具的研发和方法上取得了不错进展。虽然目前还没有最佳范例可以参考的文集，数字资源长期保存的技能同样得到了扩展。遗憾的是，这些领域之间没有进行良好的沟通，且这些领域局限在承担公共存储职能的机构和研究团体中，没能推广。同时，数字资源长期保存专家与数据生成和管理者之间似乎也缺乏沟通。

无论是一般规模还是大规模商业领域中的社团，它们可能肩负着长期关注数据的职责，都没有从数字长期保存中得到商业灵感。因此，没有成熟的市场，产业也没有动力投资数字资源长期保存解决方案的开发。监管框架不断演化更加注重数据长期保存，提升了其对数字长期保存的兴趣。但是，与会者认为应该将重心从数字资源长期保存的危机管理方法向增值模式转移。

然而，缺少一种能够说服商业组织在这一领域投资的有效经济模式。除非市场对数字资源长期保存产品的需求上升，否则就不会有满足这一需求的产品投入。

有必要从数据保存向知识保存转变，同时，要能够保证所保存数据的真实性。研讨会上展望了数字资源长期保存在 2020 年的未来蓝图，与会者们预测会出现明确的经济模式证明数字资源长期保存的价值。这一模式将使得数字资源长期保存被广泛接纳，在事实上成为拥有大量数据的组织的常规业务。

数据的易逝性和真实性也将受到越来越多的关注。另外，数字资源长期保存将保存整个的业务流程而不仅仅是其中的数据。其次，用户在数据存在的原始环境中与其交互的经验也会被保存。

为了获得深入到个人用户级别的广泛应用，必须要简化数字资源长期保存操作并实现其自动化，还得为所提供的保存服务找到一个能被广泛认可的付费模式。这样的服务将以在线的形式提供。某些组织已经意识到了这一潜在机会，他们通过广告收入来资助保存服务。

需求的增长会刺激产业级保存服务的产生。反过来，这些服务的产生又会导致数字资源长期保存领域产生一个新行业。该行业角色定位明确，拥有自己的资格认证、培训和再培训方式。

现在正困扰数字资源长期保存的法律问题将得到克服。国际通用的数字资源长期保存标准将会出台, 标准中不仅强调保存什么数据以及如何实施保存工作, 还强调舍弃什么数据。

扩展现有的元数据模型, 建立一个单一的、全局的模型。

研讨会还为 2013-2020 的框架提出了研究主题, 研讨小组提出了 14 个领域供参考。

- 长期保存信息的抽取; Extraction of Preservation Information
- 集成访问: 时间-系统-领域; Integrated access: Time – Systems – Community
- 将数字长期保存作为一个计算机科学问题重新表述; Reformulate Digital Preservation as a computer science question
- 集成仿真系统; Integrated emulation systems
- 知识长期保存; Knowledge Preservation
- 质量评估; Quality Assessment
- 复杂对象; Complex Objects
- 自动化; Automation
- 易用性(例如隐式保存)和私有数据; Ease of use (e.g. implicit preservation) and private data
- 将数字资源长期保存集成到数字资产管理中; Integration of Digital Preservation into Digital Asset Management
- 标准; Standards
- 市场驱动和成本效益; Market-Driven and Cost Benefit
- 自保存对象; Self-Preserving Objects

在这些领域中, 有一些共同的关注点。

首先, 与会者普遍认为有必要扩大利益共同体。同样的, 他们还建议计算机科学领域的工程学科应该和在数字资源长期保存研究初期有着重大影响的人文艺术学科深入接触。

与会者还一致认为需要关注自动化和简化工作, 同样, 需要开发出服务处理日益增长的复杂数据对象。有一项提议指出开发保存对象, 将其作为实施保存的

一个备选方式。

大量与会者强调要开发支持数字资源长期保存投资的良好商业模式。他们普遍认为, 欧盟在项目融资模式外还应有额外的奖励措施, 以吸引产业界的合作伙伴参加项目。

许多提议指出采取分阶段的方式进行框架计划研究, 开始几年根据小目标制定行动方案, 随后, 在大的集成项目中并入这些工作。

最后, 与会者提出投资发展培养计算机科学家的新方式, 确保长期保存问题成为大学课程的核心部分。以此来保证将来的信息专家在他们的工作中永久地嵌入长期保存工作。

信息源:

<http://www.dpconline.org/newsroom/latest-news/712-ec-digital-preservation-research-2013-onwards>

[http://cordis.europa.eu/fp7/ict/telearn-digicult/future-of-the-past\\_en.pdf](http://cordis.europa.eu/fp7/ict/telearn-digicult/future-of-the-past_en.pdf)

编译: 丁宁 校对: 何莉娜 吴振新

## 描述数字报纸生产的 CRL 报告



研究图书馆中心为 NDIIPP 项目撰写了“在数字环境下保存新闻: 转型期的报纸行业概述”报告。

该报告首次尝试追踪纸质报纸和在线报纸中报道的新闻内容和信息的生命周期。同时, 报告首次阐述了这两种主要新闻发布渠道之间的关系。

报告分为两个部分:

1. 背景和方法
2. 围绕新闻信息生命周期的三个主要阶段, 概览新闻工作流和新闻系统

### 背景和方法

目前的这份报告基于国会图书馆 NDIIPP 项目于 2009 年 9 月召集的研讨会产生。9 月举行研讨会的目的在于探寻在国家层面收集和保存数字化新闻信息的

策略。

该报告注重以下几个兴趣点：

1. 电子复印版的特性：国会图书馆对以版权登记和存档为目的进行的印刷报纸的电子复印版这一特殊新闻格式如何进行开发十分重视。

2. 网络和印刷形式的新闻报道之间的关系：国会图书馆想了解各个新闻网站公布的内容与纸质版的如何匹敌。

3. 技术性的格式规范和电子新闻成果的传播：用于规定格式、管理和传播新闻内容的现存的或正在制定的技术标准发展到了什么程度。通过印刷和网络传播的新闻内容其现行格式化和编码的做法有哪些？

报告的第二部分以选择的用于研究的报纸实例为基础，概括地讲述了新闻采集、编辑、发行的过程。个别严重偏离这种一般性叙述的报纸，在报告中也有特别说明。

### **新闻工作流和系统概况**

在报告中，将新闻内容的生命周期分为以下几个阶段：

A. 采集：新闻机构从新闻信息和内容产生者、报道者和拥有者手中收集新闻信息。

B. 编辑和出版：编辑、处理和加工新闻内容和信息，制备供各种媒体发布用的产品。

C. 发行：通过印刷和网络媒介传播报道新闻内容、产品及其衍生品。

### **采集新闻内容**

某一特定报刊或者新闻网站的内容都是由一群人和机构共同生产的。印刷型报纸和新闻网站最大的不同在于，新闻网站相较而言包含更多的内容，这些内容进行进一步加工，它们都是由第三方商业服务商提供的，例如广告公司或者广告服务器、谷歌地球（第三方内容）。另外，相对于印刷型报纸，新闻网站吸收了更多由用户提供的内容，无论是直接借助新闻机构控制的博客、调查和其他论坛，还是借助 YouTube、Twitter、FaceBook 等第三方社交媒体站点。这些增长的多样的新闻内容来源和获取渠道增加了工作流的复杂程度，同时，也给新闻机构和新闻信息保存方的权限管理带来了挑战。

报告的该部分分为 A.1 和 A.2 两节。A.1 下的 A.1.1 和 A.1.2 中对新闻内容

的提交和收集以及图片和其他非文本媒介进行了介绍。A.2 阐述了文本和图像规范化编辑标记。

### 新闻编辑和出版

在此处我们所讨论的编辑和出版活动是指在报纸出版者或出版者的隶属机构监管下进行的生产印刷版和网站或者其他对等衍生品以及管理实施编辑出版活动所需要的内容的过程 and 任务。这些活动包括对采集的内容和报纸印刷版合集的审查、修改、编辑、格式化、标记、贴标签、文本内容 feeds 和电子传真聚合器的格式化、网络版报纸内容的开发和格式化。

此部分的报告，揭示了 B.1 出版系统和 B.2 归档和数字资产管理两节内容。B.1 下属小节 B.1.1 中将进一步介绍编辑系统，B.1.2 中将进一步介绍报纸编页或者设计排版。B.2 的下属小节 B.2.1 对主要数字资产管理系统进行了介绍，B.2.2 对数字资产管理的过程进行了介绍。

### 发行新闻

大部分新闻报纸出版者采用印刷版和电子版两种方式发行新闻内容。除了少数出版者，所有出版者的印刷版报纸仍然很大程度上只在当地经营。但是，大量的发行活动由母公司管理。或者，在国家与国际层面他们依赖于 YouTube、WordPress 等专业服务提供商和其他运营机构。用于不同发行渠道的文本、数据和图像的前期预备处理工作都发生在编辑系统。但是，送往网络和移动平台的内容很快和用于印刷和聚合器发行的版本分离。出版者通过一个通用的内部工作流程产生类型设定印刷和电子传真版生产所需要的参数。编辑系统中，网站、移动设备和其他电子内容采用的工作流与印刷和电子传真采用的工作流是分离的。

报告在此部分分别介绍了四种发行渠道：印刷版、电子复制版、文本聚合器、网络。在网络发行途径中进一步讲述了印刷型和网络版的区别、多变的网络内容使用寿命、网络内容的采集、网络出版系统和工作流以及网络发行辅助网络。

信息源：

[http://www.digitalpreservation.gov/news/2011/20110505\\_news\\_CRL\\_report.html](http://www.digitalpreservation.gov/news/2011/20110505_news_CRL_report.html)

[http://www.digitalpreservation.gov/library/docs/CRL\\_digiNews\\_report\\_110502.pdf](http://www.digitalpreservation.gov/library/docs/CRL_digiNews_report_110502.pdf)

编译：丁宁 校对：何莉娜 吴振新

## 【信息扫描】

### PrestoCentre 中心宣告成立

#### ——应对视听资源存档的新战略和挑战

3 月 14 至 15 日，在荷兰西尔沃苏(Hilversum)的荷兰视听研究所（the Netherlands Institute for Sound and Vision）举行了题为“Screening the Future”的会议，会议伊始 PrestoCentre 中心正式宣告成立。



这次会议有超过200名视听资源档案管理员、图书馆员、博物馆专业人员、电影制片人、倡议者和在荷兰视听研究所从事文化遗产保存的政策管理员参与其中。会议为期两天，主要聚焦于视听领域的共同挑战和解决方案。

十几年来，英国广播公司（BBC），the Institut National de l'Audiovisuel en France (INA), Radiotelevisione Italiana (RAI), Österreichischer Rundfunk (ORF)和荷兰视听研究所携手合作寻求这些及其它挑战的解决方案。在欧洲共同资助的Presto项目中，他们与科研机构、提供商和供应商共同合作，开发了系列方法，将长期的经验和理论运用到声音和移动图像内容的实际需求中，以使这些资源得以安全地在能负担得起的分布式和联邦的环境中保存及呈现。

这些机构间合作产生了一个独特的组织——PrestoCentre，这将确保该领域建立的知识和贡献精神得以延续，而不至于“逐渐消失”。并且，PrestoCentre对于有效利用有限的资源和现有可用的基础设施至关重要。

PrestoCentre中心是一个会员制驱动的组织，其为社区中的利益相关者提供视听数字化和数字化保存服务。PrestoCentre通过在线和离线服务、出版物和培训，给保存人员和视听内容的创建者提供分析和建议。并与国际组织和项目，及其专家、志愿者和倡议者密切合作，加强视听部门长期存取文化遗产资源的能力。

PrestoCentre 社区的会员将能够访问一个由世界领先的档案馆、大学和供应商组织组成的专家网络。PrestoCentre 提供培训和服务，较好地满足了目标用户

和领域相关者的需求，并且通过一个动态门户网站促进他们之间的合作。门户网站在 2011 年将不断发展为一个供领域内实践者举行主要社区会议的地方。

信息源：

<http://screeningthefuture.eventbrite.com/>

下载会议报告：

<http://prestocentredownloads.webs.com/Documents/STF2011/Screening%20the%20Future%202011%20-%20Programme.pdf>

编译：李丹丹 校对：何莉娜 吴振新

## 地球空间数据长期保存资源中心网站发布

地理空间数据长期保存资源中心近日发布了新网站。建立该网站是为了帮助负责生产和管理地理空间数据的人员了解实现长期地理空间数据保存和获取的最新方法和工具。该网站目前由 CIESIN（国际地球科学信息网络中心，是哥伦比亚大学地球科学研究所的一部分）负责维护。地理空间数据长期保存资源中心同时是国会图书馆 NDIIPP 项目的一部分。

通过以下链接可链入该网站：<http://geopreservation.org/>



网站主要分教育和培训、工具和软件、政策和效益三大模块，该网站描述各种相关资源并提供相关链接，网站提供按主题、资源类型和自由文本三种搜索功能。在网站首页提供地理空间数字资源数据管理者、系统开发者和研究者的链接，同时还有该专业领域长期保存会议的链接。

信息源：

<http://duraspace.org/geospatial-data-preservation-resource-center-launches-web-site>

编译：丁宁 校对：何莉娜 吴振新

## 欧盟资助新的数字资源长期保存项目- APARSEN

最近，欧盟委员会资助了一个名叫 APARSEN 的新项目。

APARSEN 项目试图建立一个卓越网络。该项目以现存的永久访问联盟为依托，同时引进大量新成员和一系列具体的工作程序。它的核心理念是对数字长期保存研发工作中潜在的存储残片进行核查。这一核查工作试图通过建立一个参与此跨领域项目的成员和机构共享的论坛来实现。该项目提议中包含 28 个不同的工作程序，4 年为期，截至 2014 年全部完成。



卓越网络是欧盟委员会为扩展和支持现存计划而资助的代表性项目。该项目将有助于参与者加强自己的研发。但参与者需要承诺独立资助项目研发。总而言之，APARSEN 网络的初衷是通过交流、意识的提高、职员培训、巩固、传播和合作关系的建立提升数字长期保存研究的价值。项目中大量的工作程序对应着时下欧洲正在广泛开展的数字长期保存研究。

信息源：

<http://www.dpconline.org/newsroom/latest-news/706-two-new-digital-preservation-projects-aparsen-and-timbus>

编译：丁宁 校对：何莉娜 吴振新



## 欧盟资助新的数字资源长期保存项目- TIMBUS

最近，欧盟委员会资助了一个新项目 TIMBUS

TIMBUS 是“Timeless Business Processes”的缩写。该项目的目的是填补现存长期保存方案的空白，提供长期的业务流程和服务。现有的数字迁移工具为长期数据管理和仿真提供方法和服务。这些工具正在进行功能扩展，试图为其他软件提供长期访问的入口，以便实现信息对象租赁服务。业务流程越来越依赖以服务和数据共享为义务的复杂网络。然而，与“提供服务的网站”的依赖关系缺少恰当的文档说明。随着服务间相互依赖关系的增长，风险管理中的风险评估和风险处理工作越来越复杂。如果，业务流程的执行环境评定网络中的服务和数据可用，数字长期保存服务就要为它们提供描述机制，并通过该机制发现已经失效的流程。

TIMBUS 对业务信息环境出现的新趋势积极响应。它承认集中式的内部业务流程逐渐过时，提供服务和基础设施支持将软件 and 平台作为服务提供的‘服务网络’。这种点播式架构之所以盛行，是因为在节省开支方面它提供了更大的灵活性和可扩展性；但是，它又引发了其他的问题——可靠性和持久性。因此，TIMBUS 不得不构建一系列必需的流程和工具确保服务能够持续几十年的可访问。TIMBUS 将开发和扩展一些企业智能风险管理的工具、服务依赖关系监管工具、合法性生命周期管理工具和分布式的相互依存的服务虚拟化工具，以实现持续访问的目标。

信息源：

<http://www.dpconline.org/newsroom/not-so-new/706-two-new-digital-preservation-projects-a-parsen-and-timbus>

编译：丁宁 校对：何莉娜 吴振新