

2013 年第 8 期（总第 29 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 8 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
数字资源保存基准（Benchmark DP）项目介绍.....	1
国会图书馆关联数据服务新增 21 个保存词汇.....	2
【动态追踪】	2
美国国会图书馆网络存档内容实现了与网站其他内容的集成	2
【重要文献摘译】	3
《DPC 技术观察报告：网络归档》摘译.....	3
【专题报道】	8
数字保存领域术语词汇汇总.....	8

【信息扫描】**数字资源保存基准（Benchmark DP）项目介绍**

数字资源保存基准（Benchmark DP）项目是由维也纳科学技术基金（[Vienna Science and Technology Fund](#), WWTF）资助的一个为期三年的研究项目，于2012年11月启动，主要目的是开发用于评估和比较数字资源保存的流程、系统和组织能力的一种集成化的、系统化的方法。

目前信息通信技术（ICT）领域的理论框架、工程方法和技术还不足以有效地维持数字编码信息的长期真实性、可理解性和可用性。在文化遗产和eScience领域，数字资源的长期保存一出现就成为信息系统的关键挑战。但是，在如何应对数字寿命的问题上，尚未有足够的理论支持及可靠的框架指导，很多数字保存应用研究都遇到了诸多障碍。

1. 从研究对象和内容的维度看，数字保存领域缺乏用于实验研究的客观的、标准的、可对比分析的度量指标和基准集合。而且由于 DP 领域发展时间有限，也无法提供这些基准的可靠性的有力证据，然而，这些指标又是实现定量研究和创新的关键所在。
2. 从系统、流程和组织维度看，一方面现有的模型没有充分考虑到数字资源的寿命和信息长期保存的需要，另一方面数字保存领域的相关知识尚未嵌入到系统架构和设计理念中。因此造成系统化的对比分析，以及流程、系统、组织能力的改进都是几乎不可能的事情。
3. 从工程维度来看，越来越有必要保存复杂的系统和流程而不只是单单保存信息本身。然而，到目前为止，数字寿命尚未纳入信息系统生命周期的基本设计考虑中。

Benchmark DP将会通过在以下3个互连的维度创建理论框架和定量基准，来促进信息技术领域对数字资源寿命的基本认识和能力提升：

1. 在内容维度，项目将会根据充分了解的证据事实创建一个模型驱动的基准生成框架，以提供近似实际的数字信息集，用于系统的定量实验和测量。
2. 在系统和流程维度，项目将会为数字保存创建和评估一个能力成熟度模型，从寿命的角度，改进系统的流程和管理 ICT 系统。
3. 最后，项目将会奠定数字资源寿命的重要性的基础，推进其成为信息系统生命周期的基本设计点。

总之，通过提供用于测度、控制和改进DP流程和组织能力的首个系统化的、定量化的方法，Benchmark DP将会增强组织在信息管理和信息系统设计方面的控制和改进流程的能力，能够使相关系统达到数字资源的预期寿命的要求。

编译自：<http://benchmark-dp.org/>

(李文燕编译, 齐燕 吴振新校对)

国会图书馆关联数据服务新增 21 个保存词汇

国会图书馆和PREMIS编辑委员会发布了21个新词汇, 它们可以用于国会图书馆的关联数据服务的保存元数据。每个词汇映射了一个在PREMIS数据词典2.2中规定的语义单元的受控值列表。此前在国会图书馆的关联数据服务(id.loc.gov)中, 只有3个保存词汇, 如今发布了21个新词汇后, 可以覆盖更大范围的PREMIS语义单位。此外PREMIS编辑委员会最近发布了一个改进的PREMIS OWL本体, 该本体基于PREMIS2.2, 可以利用id.loc.gov上的24个取值词汇(value vocabularies)。

词汇详情参见: <http://id.loc.gov/preservationdescriptions/>

编译自:

<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1306&L=digital-preservation&F=&S=&P=5931>

(李文燕编译, 齐燕 吴振新校对)

【动态追踪】

美国国会图书馆网络存档内容实现了与网站其他内容的集成

历时一年多的时间, 国会图书馆的部分网络存档内容终于与网站的其他部分实现了集成融合, 从此, 这些网络存档内容同网站上的书籍、照片、期刊等一样, 可以被用户方便地检索和使用了。

目前, 已集成了八套网络存档集合, 其余藏品的集成工作还在继续, 预计还会包括原网络存档网站([Library of Congress Web Archives](#), 所有内容迁移完毕之后此网站才会停止运行)上尚未涉及的网络内容。

新集成后的网络存档界面被称为“已存档的网站([Archived Web Site](#))”, 其所具有的特性包括:

- **可以从图书馆的主页直接搜索网页档案:** 当在loc.gov站点上进行搜索时, 可以通过选择“Archived Web Site”这一选项来直接搜索已存档的站点, 而不需要到达另外一个单独的界面。
- **已存档网站的内容可与图书馆其他内容同时被检索到:** 当进行综合检索时, 返回的结果中不仅有报刊、照片等普通类型的记录, 还会夹杂有网络存档的内容。

- **分面搜索:** 用户可以通过限定日期、名称、资源编著者、主题、位置和语言等条件缩小搜索结果的范围。
- **组合式记录:** 当URL发生变化或者同一个URL隶属于多个藏品时, 之前多个、分散的目录数据现在被组合起来, 用户可以很容易地找到某特定网址包含的所有过往的内容。
- **缩略图浏览和特色藏品专栏:** 项目组为所有的存档网站制作了缩略图, 并将其集成到了搜索页面, 内容保管员还为8个藏品集合分别挑选了一些缩略图做成了特色藏品专栏。
- **新的浏览工具:** 除了网络存档用户熟知的通过目录数据和Wayback日历视图外, 国会图书馆同时提供了一个新的浏览工具, 即以一种新的方式展现Wayback Machine上的内容。

处理网络档案不像处理数字化图像那么简单, 所以项目组为之付出了很多努力, 未来还会继续进行反复的修改和调整, 以期提供更好的用户体验。

编译自:

<http://blogs.loc.gov/digitalpreservation/2013/06/moving-on-up-web-archives-collection-has-a-new-presentation-home/>

(齐燕编译, 刘超 吴振新校对)

【重要文献摘译】

《DPC 技术观察报告：网络归档》摘译

Maureen Pennock 著 刘超编译

该报告列出了从事网络归档活动的机构当前共同面临的一些问题, 并就实施网络归档的一些技术方法给出了建议, 更重要的是, 为机构可利用的资源提供了宝贵的信息。报告主体分为六个部分, 分别是: 引言、问题、标准、软件、集成系统、案例研究以及结论和建议。现对具体内容进行摘译:

1 引言

1.1 为什么要归档网站?

(略)

1.2 网络归档的挑战

要捕获和归档一个简易的小型个人网站相对简单, 可以直接从服务器上复制或下载并进行离线存储, 即使没什么技术专长的人也可以操作。但如果要长期且大规模地从事网络归档

活动却要复杂的多,这不仅需要循环性地捕捉庞大而复杂的网站,还要维护和明晰同一个网站的不同版本之间的关系,并同时管理“提取”的网站集合中人为划定的界限,所有这些都更需要更加复杂的解决方案的支撑。再加上是以“国家范围”的规模进行收集,以及技术过时这个长期难解决的问题,更增加了挑战的难度。

虽然已开发出一些可以应对许多挑战的技术解决方案,使得小型的组织更容易开展网络归档的行动,但有些问题还没有完全解决,比如时间的连贯性问题:在网络归档的时候,一个大型网站的内容可能在其被采集过程中就已经发生了变化,因而其时间的完整性就无法保证了。类似地,在网络归档的时候,不同网站间的链接可能会造成网站被采集的次数不均,使用户难以判断一个时间点什么样的信息才是可用的。网络的快速变革意味着网络归档的目标也会永远地变化着。网络归档的解决方案需要能够快速响应,并能尽可能多地采集网络内容材料,为未来的用户提供“身临其境”式的体验。

然而所谓“身临其境”式的体验是什么样的还不清楚,还有一些学术问题有待解决:例如如何体现存档网站的“真实”性?如果一个存档网站中的一些链接损坏了,那么它是真实的,还是算作内容缺失呢?由于网站不是静态对象,它是动态生成的,而且对不同的用户可能存在呈现差异,在这种情况下,其真实性的标准又如何制定呢?这些问题仍在探讨中。

其它方面的挑战和问题包括法律的不确定性(如版权,数据保护和诽谤)、质量保证(要确保已捕获了所有必要的文件,并且这些文件能够被呈现出来)、数字资源的长期保存(即使技术发生了变革,这些资源也可以持续到未来并能被呈现出来),以及关于怎样确保网络归档技术与网络演变的步伐保持同步的问题。报告第4部分对这些问题进行了更详细地讨论。

1.3 重要的网络归档行动

从事大规模网络归档活动的组织和机构包括:互联网档案馆(Internet Archive)、国际互联网保存联盟(IIPC)、大英图书馆、法国国家图书馆、美国国会图书馆、澳大利亚国家图书馆以及瑞典国家图书馆。从事较小规模但偏重于解决网络归档活动中的实际问题的组织有:英国网络归档联盟(UKWAC)和北欧网络档案馆(Nordic Web Archive)。提供商业网络存档服务的机构有:互联网档案馆(Archive-It服务)、互联网记忆基金会(IMF Archive.net网络归档服务)。此外,由欧盟委员会资助的进行网络归档研究和实践的项目有:

- SCAPE项目
- LAWA项目
- ARCOMEM项目
- BlogForever项目
- LIWA项目

1.4 网络归档的技术方法

- 1) 客户端归档(Client-side archiving)
- 2) 事务归档(Transactional archiving)

3) 服务器端归档 (Server-side archiving)

2 问题

2.1 法律

(略)

2.2 选择

(略)

2.3 当前网络爬虫的局限性

网络爬虫在捕获某些类型的内容时存在局限性, 这些有问题的内容包括:

- 数据库/动态驱动的内容 (即通过数据库为响应用户请求而生成的网页);
- 流式多媒体文件;
- 只能通过本地站点搜索才能获取的内容——网络爬虫几乎不能分析其脚本代码;
- 受密码保护的内容——如果提供密码的话, 网络爬虫就能克服这个问题, 如果没有密码的话就很难获取这些内容;
- 某些由Javascript驱动的菜单——如URL是动态生成的情况。

2.4 真实性、完整性和质量保证

(略)

2.5 Web 2.0的归档

Web 2.0的网站如维基、博客 (带有评论功能)、社交网站以及媒体共享网站, 对其进行归档是非常耗时的, 某些情况下甚至是一个不可能完成的任务。

2.6 时间连贯性

(略)

2.7 病毒和恶意软件

(略)

2.8 重复数据的删除 (De-duplication)

(略)

2.9 搜索限制

(略)

2.10 长期保存

(略)

3 标准

网络标准中最核心的是HTML和相关的HTML/XML技术, 此外还有robots.txt协议。《ISO 28500:2009》和《ISO技术报告14873》为网络归档活动提供了标准, 《ISO 28500:2009》常被称作WARC标准, 而《ISO技术报告14873: 网络归档的统计和质量问题》(定于2013年出版) 中提供了大量有意义的馆藏建设统计资料, 并确立了一些与内容持续获取有关的定量指标。

此外,还有英国中央信息办公厅(COI)的网站归档标准,一份关于管理URLs的类似文件明确要求管理URLs的网站管理员要有管理持久且有意义的URLs的实践经验。长期保存的标准有《ISO 14721:2012》,通常被称作开放档案信息系统(OAIS)参考模型。

一个标准中如果没有提到元数据将是不完整的。Marc 21、ISAD(G)以及都柏林核心元素集可以当作网络存档的描述性元数据使用,而国会主题词表(LCSH)和/或杜威十进分类法(DDC)可以用来辅助资源发现。另外,PREMIS可用于保存元数据的创建,METS则可提供描述、管理和结构元数据。

4 软件

4.1 集成系统

4.1.1 PANDAS

PANDAS(PANDORA项目的数字存档系统)是第一代集成网络归档系统之一。它支持选择、设置权限、计划安排、收割、质量保证、归档以及访问,但不是一个开源软件。目前,由国家图书馆同盟(NLA)使用PANDAS进行有选择的网络归档。

4.1.2 网络管理工具(Web Curator Tool, WCT)

网络管理工具是一个开源的工作流程工具,可用于管理挑选出的网络归档流程。它支持选择、设置权限、描述、收割、质量保证等,有单独的访问界面。

4.1.3 NetarchiveSuite

NetarchiveSuite是用Java语言编写的一个开源的网络归档应用程序,可用于管理选择性的和跨领域性的网络归档内容。它支持选择、权限设置、计划安排、收割、质量保证和访问。

4.2 第三方/商业服务

Archive-It是一个由互联网档案馆提供的网络归档订阅服务。Archivethe.Net是一个由互联网记忆基金会(IMF)提供的基于网络的网络归档服务。作为加州数字图书馆的一部分,加州大学管理中心(The University of California's Curation Centre)可为选出的网络归档集合提供完全托管的网络存档服务。Hanzo Archive可提供商业网站的归档服务,以满足法规遵从、电子发现和记录管理等相关的商业经营需求。提供商业网站归档服务的还有Reed Archive,它可以满足用户对法规遵从、诉讼保护、电子发现以及记录管理的需求。

4.3 网络爬虫

HTTrack是早期开发的一个免费的开源离线浏览器和爬虫程序。Heritrix也是一个开源的爬虫程序。WGget是一个支持从网络上下载文件的免费实用工具,经过配置可作为网络爬虫使用,用户可指定从网上下载文件的深度。Miyamoto是一款成熟且具高性能的网络爬虫工具,它由Hanzo Archive设计用来捕获其它爬虫程序常会出问题的网络内容。而最近推出的由互联网记忆(Internet Memory)设计的Memorybot爬虫工具,则是专门用来满足大规模爬行需要,与其它爬虫程序相比,其成功捕获的URLs数量会显著增加。

4.4 搜索

NutchWAX是一款用来索引和搜索网络归档集合的开源工具，它需要配合访问接口使用，如Wayback。SOLR也是一个开源的搜索平台。

4.5 访问

支持访问的组件和工具有：开源组件Wayback，以及Memento框架（曾获得2010年数字保存大奖）。此外，许多分析访问工具正被改进增强功能，以更好地发掘和利用集合的时空特性。UK Web Archive为用户提供了一款NGRAM浏览器，在响应用户搜索请求的时候，它可以以图形的形式来显示搜索字词随时间在网络存档中的出现频率。在此基础上进一步的工具也已经被开发出来，包括tagclouds、学科层次结构的可视化工具，以及基于邮政编码的访问工具。另外，美国一个项目已开发出了6个可视化工具。

4.6 其它软件

SiteStory是Memento团队开发的一个用于支持网络事务归档的新的解决方案；Twittrervane是一个可按主题或事件对网站进行选择的新方法。

5 案例研究

（略）

6 结论和建议

在过去的十年中，网络归档技术有了明显的进步，人们对其中涉及的问题的理解也加深了。在归档和保存网上文化记忆方面，我们有广泛的工具和服务可供选择，并且都能符合捕获和保存网上记录的要求。此项工作仍在不断进行中，因为只要互联网还在不断发展，那么网络归档技术就必须能跟上其步伐。令人鼓舞的是有许多研发项目都在利用现有的网络技术设计开发不同的解决方案，尤其推动和促进了数据集合的访问和使用。

除了技术的发展，从实际部署和使用网络归档工具的实践中获得的知识和经验，也使人们对网络归档的最佳实践，将网络归档嵌入组织环境的运营策略，业务需求和利益，使用案例以及资源选择有了更好的理解。一个希望开展网络归档计划的组织在开始之前必须要对自己的业务需求非常清楚，因为业务需求才是任何网络归档活动的根本推动力，而且它还会对网络归档策略及选择政策的细节的制定产生显著的影响。事实上，各种商业服务和技术的应运而生就已标志着网络归档作为一种业务需要和一门学科的成熟。

然而，尽管在这些方面不断的做着努力，网络归档仍然面临着重大的挑战：归档社交和协作平台的研究和实践仍在进行中；为确保网络归档是按照预先设定来进行收割并存档的“质量保证”，是在过去十年中取得最少进展的领域之一；需要投入更多资金和注意力来开发更可靠的工具，以验证网络爬虫不仅捕获了所有必要的文件和内容，而且可以适当地对其加以呈现；通过提高爬虫的可信赖性，以及减少目前已知对爬虫来说是“风险”的领域，来设计更加智能的爬虫软件。如果不能很好地解决这些挑战，就有可能面临集合和归档的网站是不完整的风险，尤其是以“国家范围”的规模进行网络归档的情况。数据的不完整会使得其潜能无法被完全发掘出来，学者在做领域级的研究时无法从不完整的集合中得出权威的结

论。而且由于人们已经在开发能够支持大规模集合的研究用途的分析访问工具，因而处理好这些挑战就显得更为重要了。

能够支持网络归档长期保存的工具同样需要进一步的开发。在按需求进行迁移，以及将仿真作为处理大规模、异构的网络归档的策略方面也需要投入更多的研发努力。由从业者主导的研究必须明确要以业务需求为中心，且最终交付的可行方案必须是全面的机构保存策略。澳大利亚国家图书馆在2009年进行的一些试验项目以及SCAPE项目中的网络归档成果，都可作为开展网络归档活动的一个基点。人们日益达成的共识是：网络归档可以支持法律、电子发现，以及提供业务/商业/国家情报服务，未来几年要将这种共识转化为对网络归档和保存技术更大力度的投资。

最后，我们绝不能忽视法律的问题。法律不仅给收集过程设置了障碍，也限制了访问，一直以来都是收集机构面临的最严峻的挑战之一。如果法律问题不解决，人们只能坐视资源继续流失。据大英图书馆估计，如果没有法定缴存法律，2012年伦敦奥运会相关的网上活动的保存内容将不到实际内容的1%，与英国重大自发事件相关的归档内容也会微乎其微，如2005年伦敦爆炸案、2009年英国议会报销丑闻、2011年伦敦骚乱等，无疑，这将会对未来的研究非常不利。虽然解决这些挑战是执行机构和政治家的责任，但是他们需要用户和领域专家来传达其实际需要和要求。不能，也不应该各自独立地来解决网络归档的技术挑战。

查看原文可点击链接：<http://dx.doi.org/10.7207/twr13-01>

编译自：<http://dx.doi.org/10.7207/twr13-01>

<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1305&L=digital-preservation&F=&S=&P=7457>

(刘超编译，李文燕 吴振新校对)

【专题报道】

数字保存领域术语词汇汇总

齐燕 吴振新 刘超 李文燕 编译校对

对于任何学科或领域来说，若想构建一个有生命力的理论体系，首先需要建造标准规格的砖瓦——具有规范定义的概念。拥有一套标准化的术语和概念不仅是为满足学科或领域科研人员沟通和交流的需要而生，还是其成熟和规范的标志，更是促进其继续发展和演进的基石。

数字资源保存领域发展迄今，美国国会图书馆、数字资源长期保存联盟（Digital Preservation Coalition, DPC）、国家数字化管理联盟（National Digital Stewardship Alliance, NDSA）等国家或国际性的组织机构都为之做出了诸多贡献。在术语上，DPC于2008年12月

出版了《DPC 数字资源长期保存管理手册》(<http://www.dpconline.org/publications/digital-preservation-handbook>), 有一个专门的章节给出了手册中使用的定义和概念; NDSA则是在今年2月份发布了一个专门的术语表(<http://www.digitalpreservation.gov/ndsa/ndsa-glossary.html>); 另外, 由美国空间数据系统咨询委员会(Consultative Committee for Space Data System, CCSDS)制定、于2003年作为ISO的标准(ISO 14721:2003)颁发的OAIS(Open Archival Information System, 开放档案信息系统)也有自己使用的一套术语, 在其2009年8月出版的红皮书《REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS)》(<http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/CCSDSAgency.aspx>)上也有专门的章节给出了相关描述。

本专题即是将这三套术语进行汇总整合, 期望能够让读者对数字保存领域的术语和概念有一个全景观式的视图。

1) 可获取性 (Access)

在《手册》中被定义为数字资源可继续地并持续地被使用, 同时保持着资源的可靠性, 准确性以及功能性, 这三种性质被视为数字资源创建和/或获得的基本目的。

(来源: DPC)

2) 访问辅助工具 (Access Aid)

帮助用户(Consumers)定位、分析、订购或检索感兴趣的归档信息包(Archival Information Package)的一个软件程序或者文档。

(来源: OAIS)

3) 访问集合 (Access Collection)

一个 AIP 集合, 如果档案存储(Archival Storage)中没有该集合的封装信息(Packaging Information), 那么就会由集合描述(Collection Description)来对其进行定义。

(来源: OAIS)

4) 访问功能实体 (Access Functional Entity)

一种 OAIS 功能实体, 用户(Consumers)借此获取归档的信息并使用相关的服务。

(来源: OAIS)

5) 访问权限信息 (Access Rights Information)

指与内容信息(Content Information)的访问权限相关的信息, 包括法律框架、许可术语、访问控制, 以及提交协议(Submission Agreement)中规定的与保存(OAIS)和最终使用(用户)相关的访问和分配条款。此外, 还包括对权限管理执行措施的具体说明。

(来源: OAIS)

6) 访问软件 (Access Software)

能够以人/系统可理解的形式将信息对象(Information Object)的部分或全部信息内容诚

信给用户的一类软件。

（来源：OAIS）

7) 散单 (Adhoc Order)

用户 (Consumer) 提出的访问 OAIS 中当前可用信息的请求。

（来源：OAIS）

8) 管理功能实体 (Administration Functional Entity)

一种 OAIS 功能实体，用于对 OAIS 的其他功能实体的日常管理和控制。

（来源：OAIS）

9) 归档信息集合 (Archival Information Collection, AIC)

聚合了其他归档信息包 (Archival Information Package) 的内容信息 (Content Information) 的一个归档信息包 (Archival Information Package)。

（来源：OAIS）

10) 归档信息包 (Archival Information Package, AIP)

保存在 OAIS 中的由内容信息 (Content Information) 及相关保存描述信息 (Preservation Description Information, PDI) 构成的信息包。

（来源：OAIS）

11) AIP 修订版本 (AIP Edition)

以改进而非以保存信息为目的，对 AIP 的内容信息 (Content Information) 或保存描述信息 (Preservation Description Information) 进行更新或完善后生成的一个新的 AIP。该修订版本的 AIP 并不能视为迁移 (Migration) 操作的产物。

（来源：OAIS）

12) AIP 备选版本 (AIP Version)

对源 AIP 的内容信息 (Content Information) 或保存描述信息 (Preservation Description Information) 进行了转换，可作为替换源 AIP 时的一个候选。一个改进版本的 AIP 可被视为一次数字迁移 (Digital Migration) 后的产物。

（来源：OAIS）

13) 归档信息单元 (Archival Information Unit, AIU)

一种 AIP，可能由多个数字对象（比如多个文件）构成，但是为保持其原始的完整性，存档机构选择不对其进行分解。

（来源：OAIS）

14) 档案原件 (Archival original)

代替“接收的版本 ([Received Version](#))”使用。

（来源：NDSA）

15) 档案存储功能实体 (Archival Storage Functional Entity)

一种 OAIS 功能实体，用于 AIP 的存储和检索。

（来源：OAIS）

16) 档案馆 (Archive)

保存信息并为特定用户社区 (Designated Community) 提供存取服务的组织。

（来源：OAIS）

17) 相关描述 (Associated Description)

特定的访问辅助工具 (Access Aid) 中对一个信息包的内容进行描述的信息。

（来源：OAIS）

18) 验证 (Authentication)

及时地在特定时间点确立数字资源真实性的机制。例如，数字签名。

（来源：DPC）

19) 真实性 (Authenticity)

反映数字对象的可信任程度的一种机械特性，利用与数字对象相关的元数据可以明确地证明数字对象与它的声明一致。

（来源：NDSA）

20) 备份 (Backup)

为防止由于数字资产的意外破坏或损坏所造成的损失而另外制作的副本。备份副本的基本特性就是要能够保证在主本访问失败的情况下相关信息可以恢复如初。

（来源：NDSA）

21) 最佳版本 (Best Edition)

在缴存日期之前的任意时间在美国出版发行的一件作品的某个版本，经美国国会图书馆确定为最适合用于存档的版本。

（来源：NDSA）

22) 位保存 (Bit Preservation)

最基础的保存方法，它能够确保数字对象及其相关元数据以其原始的格式长期保持完整性，即使存储它们的物理媒介在不断发展和改变。也被称为“bit preservation”。

（来源：NDSA）

23) 原生数字资源 (Born Digital)

是指数字形式为唯一存在形式的资源，没有模拟形式的等价物作为信息来源或是由其转变而来。手册中所使用的这一术语不包含以下两种情况：1) 由模拟形式原稿转换而来的数字资源；2) 来源于数字形式，但却已经印成纸质的，如一些电子记录。

（来源：DPC）

24) 保管链 (Chain of Custody)

用于按时间顺序维护和记录数字对象的处理历史，包括所有权的转移、任意数字文件从

被创建到最终正式版本的过程等。请同时参考术语“provenance”。

（来源：NDSA）

25) 校验和 (Checksum)

一个通过算法来计算的数值，用于检验单个文件或者某个文件集合的状态和内容，以确定在其传输或存储过程中是否发生了意外错误。在之后的任意时间，都可以重新计算这个数值，并与之前存储的数值进行比较，据此可判断数据的完整性，如果两个校验和能够匹配，那么几乎可以完全确定数据没有发生改变。请同时参考术语“[Fixity Check](#)”。

（来源：NDSA）

26) 集合描述 (Collection Description)

一类信息包描述 (Package Description)，专门为访问辅助工具 (Access Aids) 提供 AIC 的相关信息。

（来源：OAIS）

27) 通用服务 (Common Service)

OAIS 必需的支持性服务，例如进程间通信、命名服务、临时存储分配、例外处理、安全保障，以及目录服务。

（来源：OAIS）

28) 用户 (Consumer)

这里的用户不止包括人，还包括客户端系统，通过与 OAIS 的服务进行交互查找和获取感兴趣的存档信息。需要特别指出的是，其他的 OAIS，以及存档机构内部的 OAIS 员工或系统也都可以担任这一角色。

（来源：OAIS）

29) 内容数据对象 (Content Data Object)

与相关联的表示信息 (Representation Information) 一起构成内容信息 (Content Information) 的数据对象 (Data Object)。

（来源：OAIS）

30) 内容信息 (Content Information)

计划进行长期保存的信息，或者是其中一部分信息。它是由一系列内容数据对象 (Content Data Object) 及相关表示信息 (Representation Information) 组成的一个信息对象 (Information Object)。

（来源：OAIS）

31) 环境信息 (Context Information)

描述内容信息与其所处环境关系的信息，包括创建内容信息的原因，及其与其它内容信息对象关联的方式。

（来源：OAIS）

32) 协作档案馆 (Co-operating Archives)

其特定用户社区拥有利益相关关系的档案馆群。他们可能会互相订购或摄入来自对方的数据。通常,为实现馆际请求,协作档案馆至少要能够支持一个通用的提交信息包(SIP)和分发信息包(DIP)的传输和处理。

(来源: OAIS)

33) 数据 (Data)

信息的一种可重新解释的表示,以规范化的形式存在,可以用于通信、编译和处理。诸如比特序列、数字表、页面的文字、一个人讲话的录音,或者一个月球岩石标本等等都是数据。

(来源: OAIS)

34) 数据字典 (Data Dictionary)

用来描述数据的正式的术语库。

(来源: OAIS)

35) 数据分发会话 (Data Dissemination Session)

通过媒介传递或一个单独的远程通信会话将数据提供给用户的过程。数据分发会话的格式/内容遵循 OAIS 和用户在签署的请求协议中商定的数据模型。该数据模型描述了 OAIS 使用的逻辑结构,及其在媒介传送或者远程通信会话中的展现方式。

(来源: OAIS)

36) 数据管理数据 (Data Management Data)

在数据管理过程中创建并进行永久存储以支持档案馆各类数据管理操作的数据,包括用户付费和授权的会计数据、政策数据、重复请求的基于事件的订单(订阅)数据、保存过程的历史数据,以及用于生成档案管理报告的统计数据等。

(来源: OAIS)

37) 数据管理功能实体 (Data Management Functional Entity)

一种 OAIS 功能实体,能够填充、维护和访问各种各样的信息,比如支持档案存储(Archival Storage)中检索功能的目录和清单、在检索出的数据上运行的处理算法、用户访问统计数据、用户付费记录、基于事件的订单管理数据、安全控制记录,以及 OAIS 的计划、政策、和规程等。

(来源: OAIS)

38) 数据对象 (Data Object)

物理对象(Physical Object)或者数字对象(Digital Object)。

(来源: OAIS)

39) 数据提交会话 (Data Submission Session)

通过媒介传递或一个单独的远程通信会话将数据提供给 OAIS 的过程。数据提交会话的

格式/内容遵循 OAIS 和用户在签署的请求协议中商定的数据模型。该数据模型描述了数据生产者使用的逻辑结构，及其在媒介传送或者远程通信会话中的展现方式。

（来源：OAIS）

40) 衍生版本 (Derivative)

原始源文件转换后的版本，通常被称为“服务”、“访问”、“分发”，“查看”或“输出”文件，用来方便访问获取或其他用途。

（来源：NDSA）

41) 派生 AIP (Derived AIP)

从一个或者多个源 AIP 抽取或者聚合相关信息而生成的 AIP。

（来源：OAIS）

42) 描述信息 (Descriptive Information)

主要由信息包描述 (Package Descriptions) 组成的信息集合，被提供给数据管理 (Data Management) 从而支持用户对 OAIS 存档信息的查找、订阅和检索。

（来源：OAIS）

43) 特定用户社区 (Designated Community)

能够理解特定信息的一组潜在用户，可由多个不同的用户社区组成。通常，这一概念由档案馆来定义，并且其内涵可能会随时间而变化。

（来源：OAIS）

44) 数字存档 (Digital Archiving)

不同机构对这一术语的使用大不相同。图书馆和档案馆通常将它与“数字资源保存 (digital preservation)”同等使用。计算机专业人员更倾向于用数字存档来表示备份和日常维护的过程，而不是将其作为数字资源长期保存的战略来看待。数字资源长期保存是此术语的扩展性定义，在本手册中被广泛使用。

（来源：DPC）

45) 数字内容 (Digital Content)

以数字格式创建、发布或分发的任意条目，包括但不限于文本、数据、录音、照片和图像、电影和软件等。可与术语“数字材料 (Digital Materials)”交换使用。

（来源：NDSA）

46) 数字材料/资料 (Digital materials)

以数字格式创建、发布或分发的任意条目，包括但不限于文本、数据、录音、照片和图像、电影和软件等。可与术语“数字内容 (Digital Content)”交换使用。

（来源：NDSA）

47) 数字迁移 (Digital Migration)

出于长期保存的目的对数字信息进行在 OAIS 内部的转移。与一般意义的转移相比，数

字迁移有以下三个特征:

- a) 关注于完整信息内容的保存;
- b) 认为迁移后的信息替代了原来的信息;
- c) 认可由 OAIS 对迁移的所有方面负有完全的控制和管理权责。

(来源: OAIS)

48) 数字对象 (Digital object)

一个概念化的术语,描述了由一个或多个相关的数字文件组成的数字内容聚合单元,这些文件可能包括元数据,衍生版本和/或相配套的封装工具。

(来源: NDSA)

49) 数字保存 (Digital preservation)

通过一系列的管理活动、政策、战略和行动来应对介质故障和技术变革的挑战,确保数字内容在所需的长久时期内都能够精确呈现。

(来源: NDSA)

50) 数字出版物 (Digital Publications)

已公布可供公众查阅的原生数字资源,采取免费或收费使用和传播的形式。包括经由通信网络获得的网络出版物、以软盘或光盘形式传播的物理格式出版物,它们可能是静态的也可能是动态的。

(来源: DPC)

51) 数字记录 (Digital Records) 同 电子记录 (Electronic Records)

机构日常业务中创建的记录,是机构较正式的传播形式。主要包括文字处理文件、电子邮件、数据库或内部网页等。

(来源: DPC)

52) 数位资源 (Digital Resources) 同数字资源 (Digital Materials)

一个广泛的定义,包括由模拟形式数据转化成计算机可识别的数字形式(数字化)而产生的数字代用品、永远没有也不打算制作模拟形式复制品的“原生数字资源”,以及电子记录。

(来源: DPC)

53) 数字签名 (Digital signature)

通过使用被检验文件的加密摘要来检验数字对象的方法。该摘要是利用算法对文件内容进行计算,产生的数值随后被加密成公钥/私钥对中的私钥。要想检验文件是否被篡改,文件接收者可以使用公钥将签名进行解密还原成原始的摘要,同时根据传输后的文件重新计算一个新的摘要,进而比较二者是否匹配。如果匹配,那么说明文件在传输过程中没有受到任何攻击或破坏。可参考术语"[Checksum](#)"和"[Fixity Check](#)."。

(来源: NDSA)

54) 数字化 (Digitisation)

指通过扫描或其它手段将模拟形式的数据转化成数字形式的过程。由此产生的数字拷贝、数字代用品可以被归类为数字资源,其可获取性同“原生数字资源”一样面临广泛的挑战。

(来源: DPC)

55) 分发信息包(Dissemination Information Package)

OAIS 应用户请求向其发出的信息包,可能由一个 AIP 也可能由多个 AIP 组成。

(来源: OAIS)

56) 文件材料 (Documentation)

指创建者提供的知识和相关的资源库,其中包含的信息可以确立知识的起源、历史和语境,可以确保知识为他人所用。另见**元数据 (Metadata)**。“至少,文件材料应提供关于数据集合的内容、起源、数据结构等方面的信息,以及与其匹配的使用限制和条件。需要充分详细的记录,可以让数据创建者在未来数据创建过程被遗忘时,还能利用此资源。同时也需要足够全面的文档,使其它人能充分挖掘和利用资源,并且文档内容要足够详细,使未参与数据创建的人理解数据集合以及整个创建过程。”(History Data Service)

元数据 (Metadata)

是描述资源重要方面的信息。迄今最多的讨论已倾向于强调元数据用于资源发现的目的。本手册重点强调成功的数字资源长期保存和管理需要什么样的元数据,以及哪种元数据可以保证基本的上下文、历史和技术性信息与数字对象一起被保存。

(来源: DPC)

57) 仿真 (Emulation)

通过开发相关技术以实现在未来计算机上模拟过时系统的方法,以此来克服软硬件技术过时。

(来源: NDSA)

58) 基于事件的订单 (Event Based Order)

用户创建的请求,旨在能够基于某一或某些事件而周期性地获取到相关信息。

(来源: OAIS)

59) 联合档案馆 (Federated Archives)

一致同意通过一个或多个公用的资源发现辅助工具来提供其归档内容获取的档案馆群。

(来源: OAIS)

60) 文件格式 (File format)

能够作为数据文件存储的信息包,该包按照某一特定信息模型进行固定的字节序列化的编码,和/或根据某一物理存储结构具体格式进行的固定编码。

(来源: NDSA)

61) 查找工具 (Finding Aid)

访问工具的一种，使得用户能够搜索并确定感兴趣的归档信息包（AIP）。

（来源：OAIS）

62) 不变性检查 (Fixity check)

校验数字对象是否被非法改变的一种机制，校验和、消息摘要和数字签名等都是进行不变性检查的工具。

这些不变性检查所产生的不变性信息可以作为数字对象的完整性和真实性的证据，是获得信任的必要因素。另请参见“校验和（[Checksum](#)）”和“数字签名（[Digital Signature](#)）”。

（来源：NDSA）

63) 验证信息 (Fixity Information)

为保证内容信息对象不会在没有被记录的情况下被改变，而对该验证机制进行记录的信息。例如，一个文件的循环冗余检验（Cyclical Redundancy Check, CRC）码。

（来源：OAIS）

64) 格式迁移 (Format Migration)

一种克服技术过时的方法，以一系列通用格式来保存数字内容，或者以原始格式保存，只在呈现时转换为所需格式。格式迁移的目的是保存数字对象，并保持用户在技术不断演变的情况下仍能对数字对象的内容进行检索、演示或其他使用。

（来源：NDSA）

65) 全局社区 (Global Community)

在联合档案馆的背景下对用户社区（Consumer community）概念的延伸，全球社区的用户可通过一个或者多个公用的查找工具来访问多个档案馆的内容。与本地社区（Local Community）相对应。

（来源：OAIS）

66) 独立可理解性 (Independently Understandable)

信息的一个特点，因为这一特点信息可被特定的社区直接进行编译、理解和使用，而无需借助包括专家在内的并不易获取的资源。

（来源：OAIS）

67) 信息 (Information)

能够用于交流的任何类型的知识。在交流中，它是以数据的形式呈现的。例如，在观测温度时，一个比特字符串（data），以及对如何将其解读为以摄氏度表示的温度数字的说明（Representation Information）即为一信息。

（来源：OAIS）

68) 信息对象 (Information Object)

数据对象 (Data Object) 连同其表示信息 (Representation Information) 合称为信息对象。

(来源: OAIS)

69) 信息包 (Information Package)

一个概念性的容器, 由内容信息 (Content Information) 和/或相关联的保存描述信息 (Preservation Description Information) 组成。与其紧密关联的概念有用于界定和识别内容信息的封装信息 (Packaging Information), 以及用于帮助内容信息搜索的封装描述信息 (Package Description information)。

(来源: OAIS)

70) 信息产权 (Information Property)

信息产权说明中描述的内容信息 (Content Information)。信息产权是信息内容的一部分, 其具体表达或确切涵义可在内容数据对象 (Content Data Object) 及其表示信息 (Representation Information) 的某些部分找到。

(来源: OAIS)

71) 信息产权说明 (Information Property Description)

对信息产权的描述。即针对某一目的而对内容信息对象的部分信息内容进行的描述。

(来源: OAIS)

72) 摄入 (Ingest)

数字对象被添加入保存管理环境的过程。

(来源: NDSA)

73) 摄入功能实体 (Ingest Functional Entity)

一种 OAIS 功能实体, 负责从信息生产者那里接收提交信息包 (Submission Information Packages), 准备用于存储的归档信息包 (Archival Information Packages), 并确保 AIP 及其相关的描述信息都存储到了 OAIS 中。

(来源: OAIS)

74) 知识库 (Knowledge Base)

由人工或系统整合出来的信息集合, 可帮助人或者系统理解接收到的信息。

(来源: OAIS)

75) 实例 (Instance)

数字文件、对象或集合的任一特定实体。

(来源: NDSA)

76) 完整性 (Integrity)

参见术语 "[Fixity Check](#)"。

(来源: NDSA)

77) 生命周期 (Life Cycle)

涵盖数字内容的创建、获取、遴选、描述、可持续性、访问和保存等一系列迭代的、模块化的管理进程。

(来源: NDSA)

78) 本地社区 (Local Community)

联合档案馆中的一个档案馆最初的服务对象——某一特定用户社区 (Designated Community)。

(来源: OAIS)

79) 长期 (Long Term)

一段足够长的时间, 在这期间技术的变革、新的媒体和数据格式, 以及特定用户社区的变化等都有可能对 OAIS 持有的信息产生一定的影响。这个时期可以扩展到无限期。

(来源: OAIS)

80) 长期保存 (Long Term Preservation)

以确保信息的真实性且可以被特定用户社区独立理解为宗旨对信息进行长期维护的行为。

(来源: OAIS)

81) 管理 (Management)

由那些确保 OAIS 总体政策与更高层次的政策框架 (如更大组织的政策) 协调一致的人员所发挥的职能。

(来源: OAIS)

82) 成员描述 (Member Description)

描述集合成员的相关描述。

(来源: OAIS)

83) 元数据 (Metadata)

关于数据的数据。

(来源: OAIS)

84) 元数据: 管理型 (Metadata: Administrative)

管理元数据由技术元数据和保存元数据组成, 一般用于数字资源的内部管理。

(来源: NDSA)

85) 元数据: 描述型 (Metadata: Descriptive)

标识资源或描述其知识内容的元数据, 可用于内容发现和识别。

(来源: NDSA)

86) 元数据: 保存型 (Metadata: Preservation)

执行、记录以及评估确保数字内容的长期保存和可访问性的相关工作所需要的语境信息。保存元数据记载了与保存活动相关的技术流程, 阐明了权限管理信息, 确立了数字

内容的真实性，并记录了保管链及数字对象的起源信息。

（来源：NDSA）

87) 元数据：权限管理（**Metadata: Rights Management**）

属于管理型元数据，阐明了版权、用户限制及相关许可协议的信息，这些信息都会影响到数字内容（包括元数据文件）的最终使用。

（来源：NDSA）

88) 元数据：结构型（**Metadata: Structural**）

用于描述包含有复杂数字对象的内容文件的逻辑或物理类型、版本、关系或其他特性的元数据。

（来源：NDSA）

89) 元数据：技术型（**Metadata: Technical**）

描述创建文件的过程及其技术信息的元数据。通常与文件格式或者创建文件的原始软件密切相关，比如创建或修改一个数字对象的扫描设备及其设置信息。

（来源：NDSA）

90) 迁移（**Migration**）

可代替“格式迁移（[Format Migration](#)）”或“存储迁移（[Storage Migration](#)）”。

（来源：NDSA）

91) 不可逆转换（**Non-Reversible Transformation**）

不能确保过程可逆的转换。

（来源：OAIS）

92) 开放档案信息系统（**Open Archival Information System**）

一个 OAIS 就是一个档案库，由组织、人员和系统组成，其中组织可能隶属于更大的组织，而系统则承担着保存信息的职责，并且向特定的社区（**Designated Community**）提供服务。由于一个 OAIS 可履行一系列职责，因而使得其超越了“archive”一词的涵义。OAIS 中的“O”（即 Open）隐含的意义是，该推荐标准以及将来的推荐标准都是在公开讨论的基础上开发的，但并不意味着对归档资料的获取是没有限制的。

（来源：OAIS）

93) 订单协议（**Order Agreement**）

档案馆和消费者之间就明确传送媒体类型和数据格式等操作细节而签订的协议。

（来源：OAIS）

94) 订购辅助工具（**Ordering Aid**）

帮助用户了解订购其感兴趣的 AIP 的花费并进行订购的一个应用程序。

（来源：OAIS）

95) 组织单元（**Organizational Unit**）

负责保管或保存数字资源集合的科室、业务部门、管理部门、规划部门、技术部门或其他工作组。

(来源: NDSA)

96) 概要描述 (Overview Description)

一种专门的集合描述 (Collection Description), 将集合作为一个整体进行描述。

(来源: OAIS)

97) 包 (名词) (Package (Noun))

指任意类型的数字数据的容器。请同时参阅["Package \(Verb\)"](#)。

(来源: NDSA)

98) 打包封装 (动词) (Package (Verb))

创建任意类型的数字数据容器的活动。请同时参阅["Package \(Noun\)"](#)。

(来源: NDSA)

99) 信息包描述 (Package Description)

用在访问辅助工具上的信息。

(来源: OAIS)

100) 封装信息 (Package Information)

用来约束和识别一个信息包 (Information Package) 的组成部分的信息。例如, CD-ROM 上的符合 ISO 9660 标准的容量和目录信息, 这些信息提示了包含着内容信息和保存描述信息的文件的内容。

(来源: OAIS)

101) 保存描述信息 (Preservation Description Information, PDI)

实施内容信息 (Content Information) 的保存所必需的信息, 可以分为: 起源信息、参考信息、验证信息、环境信息和访问权限信息。

(来源: OAIS)

102) 许可权限 (Permissions)

在计算环境中对特定角色的系统用户设置的访问权限, 同时也是计算机系统对特定对象的一种访问管理机制。根据不同的系统或应用, 可以为某一特定用户、用户组或者所有用户, 可以为一个角色或一组角色, 或者基于一个或多个用户的特征设置不同的许可权限。

(来源: NDSA)

103) 物理对象 (Physical Object)

诸如月亮岩石、生物样本和显微镜载片等具有物理上可观测到的一些性质的对象, 而且这些性质的呈现信息非常适合记录下来进行保存、发布和独立使用。

(来源: OAIS)

104) 保存副本 (Preservation copy)

用于保存的数字内容，通常被认为是数字资源所包含的知识内容的主版本。除数字内容本身之外，保存主文件还会包含与原件相关的其他信息。由于是按照高标准创建而来，因此在原件被破坏、损坏或没有保留的情况下，保存主文件可以代替原始记录来使用。通常，不会对保存主文件进行重大的处理或编辑，多是对其进行复制来制作分发副本等。

（来源：NDSA）

105) 保存计划功能实体 (Preservation Planning Functional Entity)

一种 OAIS 功能实体，用以监视 OAIS 环境并提供相关建议和保存规划，以确保即使在原先的计算环境过时的情况下，OAIS 中存储的信息仍能被特定用户社区用户获取和理解。

（来源：OAIS）

106) 处理/流程 (名词) (Process (noun))

一个定期发生的持续性活动，或者以某种方式发生或执行的一系列动作，拥有特定的结果或输出；一个持续的操作或者一系列操作。

（来源：NDSA）

107) 处理 (动词) (Process (verb))

记录或解释（信息、数据等）；通过程序来对（数据）进行特定操作。

（来源：NDSA）

108) 信息生产者 (Producer)

提供被保存的信息的人或者客户端系统，包括其它的 OAIS 或者 OAIS 的内部人员或系统。

（来源：OAIS）

109) 起源 (信息) (Provenance)

有关数字对象起源的信息，还包括其生命周期过程中发生过的任何变化的信息。

（来源：NDSA）

110) 起源信息 (Provenance Information)

用来记录内容信息 (Content Information) 的历史的信息。这些信息给出了内容信息的原始形式或者来源，以及从生成后发生的所有变动和保管者信息。档案馆从摄入过程开始创建和保存来源信息，而此之前的来源信息应由信息生产者来提供。来源信息可辅助证明信息的真实性。

（来源：OAIS）

111) 接收 (的) 版本 (Received Version)

真实的且唯一的重要资源条目，要么是原件，要么是现存的与图书馆最初获得的条目最接近的替代品或副本。请同时参考["Preservation Copy."](#)

（来源：NDSA）

112) 参考信息（Reference Information）

用作内容信息（Content Information）的标识符的信息，包括允许外部系统明确引用特定内容信息（Content Information）的标识符，例如 ISBN 号。

（来源：OAIS）

113) 参考模型（Reference Model）

用于理解某个环境中实体之间的重要联系，以及开发支持该环境的一致标准和规范的框架。通常，一个参考模型只包含少量的标准的概念，可以作为向非专业人员提供相关教育和培训的基础。

（来源：OAIS）

114) 格式变换（Reformatting）

将资源从一种存储媒介复制到另一种不同的存储媒介（媒介变换），或从一种文档格式转换到另一种文档格式（格式变换）。

（来源：DPC）

115) 媒质更新（Refreshing）

将信息资源从一种存储媒介复制到相同的存储媒介上。

（来源：DPC）

116) 更新（Refreshment）

一次数字迁移，其结果是用一个完全精准的副本替代一个媒体实例，保证档案存储（Archival Storage）的所有硬件和软件的运行与之前完全一样。

（来源：OAIS）

117) 重新封装（Repackaging）

一次数字迁移，过程中 AIP 的封装信息会发生变化。

（来源：OAIS）

118) 复制（Replication）

一次数字迁移，过程中封装信息、内容信息和 PDI 均没有发生改变。在转移到同类或者新的介质的过程中，用于表示信息对象（Information Objects）的比特被保存了下来。

（来源：OAIS）

119) 呈现信息（Representation Information）

把一个数据对象（Data Object）映射到多个有意义的概念的信息。例如：格式为一个 FITS 文件的比特序列的表示信息可能由 FITS 标准（定义了使用的格式）和一个字典（解释了文件中不属于标准的关键词的意思）组成。

（来源：OAIS）

120) 呈现网络（Representation Network）

充分描述一个数据对象的含义的一系列呈现信息 (Representation Information) 所组成的网络。数字形式的呈现信息 (Representation Information) 需要附加的呈现信息 (Representation Information), 以使得其数字形式能够长期被理解。

(来源: OAIS)

121) 呈现渲染软件 (Representation Rendering Software)

以人们易于理解的方式来显示一个信息对象的呈现信息 (Representation Information) 的一类软件。

(来源: OAIS)

122) 受限使用 (Restricted Use)

数字内容的使用受到一定的限制, 原因可以是版权的限制、捐赠协议的要求、安全检查所需、存在个人身份信息 (PII), 或者是内容仅供内部使用。

(来源: NDSA)

123) 检索辅助工具 (Retrieval Aid)

一种使得授权用户能够检索到信息包描述 (Package Description) 所描述的内容信息和 PDI 的应用程序。

(来源: OAIS)

124) 可逆转换 (Reversible Transformation)

在这种转换过程中, 以新的呈现形式存在的结果实体的集合 (或子集) 的与原始形式定义的结果实体是对等的, 这意味着新的呈现与原始呈现及基础实体集之间存在一对一的映射。

(来源: OAIS)

125) 模式 (Schema)

数据结构的形式化描述。比如 XML, 一种定义结构、元素和属性的常用方法, 可以用于 XML 文档使其结构符合某个特定模式。

(来源: NDSA)

126) 搜索会话 (Search Session)

用户向档案馆发起的一次会话, 在这过程中, 用户使用档案查找工具 (Finding Aids) 来识别和调研感兴趣的档案资料。

(来源: OAIS)

127) 语义信息 (Semantic Information)

比结构信息 (Structure Information.) 更进一步描述信息含义的呈现信息 (Representation Information)。

(来源: OAIS)

128) 提交信息包 (Submission Information Package, SIP)

信息生产者 (Producer) 交付给 OAIS 的信息包 (Information Package), 用于一个或者多个 AIP 及其相关的描述性信息 (Descriptive Information) 的构建或更新。

(来源: OAIS)

129) 存储迁移 (Storage Migration)

复制数字数据的内容的过程, 将数字数据从一种存储技术环境或配置环境转移到更新的技术环境或配置环境中。

(来源: NDSA)

130) 存储: 存档 (档案级存储) (Storage: Archival)

数字存储的一种类型, 它能够为数字对象的长期存储、维护和检索提供相关服务和功能。

(来源: NDSA)

131) 存储: 近线 (近线存储) (Storage: Nearline)

在计算机科学领域中使用的一个术语, 描述了介于在线存储 (支持数据的频繁快速的访问获取) 和离线存储/存档 (用于备份或者长期存储, 访问量较少) 之间的一种中间类型的数据存储方式。“Nearline”是“near-online”的缩写。请同时参阅术语“离线存储 ([Offline Storage](#))”和“在线存储 ([Online Storage](#))”。

(来源: NDSA)

132) 存储: 离线 (离线存储) (Storage: Offline)

在这种方式下, 所有的存储媒介都必须首先连接到计算设备上才可被计算系统访问。离线存储所使用的介质包括磁带驱动器, 固定媒体 (CD、DVD 光盘、闪存盘), 或者硬盘驱动器等, 它们都不支持连续不断的网络访问。也被称为可移动的存储。另请参阅“近线存储 ([Nearline Storage](#))”和“在线存储 ([Online Storage](#))”。

(来源: NDSA)

133) 存储: 在线 (在线存储) (Storage: Online)

本地或网络可获取的存储方式, 所有应用程序都可以即时获取数据, 不需要访问任何的低层次的存储环境。另请参阅“近线存储 ([Nearline Storage](#))”和“离线存储 ([Offline Storage](#))”。

(来源: NDSA)

134) 结构信息 (Structure Information)

关于信息是如何组织的信息。例如, 比特流如何映射为一般的计算机数据类型, 包括字符、数字、点阵, 以及聚合后的字符串和数组等。

(来源: OAIS)

135) 提交协议 (Submission Agreement)

OAIS 和信息生产者之间达成的协议, 规定了一个数据模型以及数据提交会话所需的其它一些规范。该数据模型详细描述了信息生产者 (Producer) 使用的格式/内容和逻辑结

构, 及其在每次媒体传送或通信会话中的表示方式。

(来源: OAIS)

136) 持续计划 (Succession Plan)

规划了 OAIS 存档资源的管理权、所属权和/或控制权何时以及如何转交给后续的 OAIS 以确保这些存档资源能够得到持续有效的保存。

(来源: OAIS)

137) 转换 (Transformation)

一次数字迁移, 在这个过程中, 档案信息包 (Archival Information Package) 的内容信息 (Content Information) 或者 PDI 会发生变化。例如, 在保存文本文档时, 将编码由 ASCII 变成 UNICODE 就是一种转换。

(来源: OAIS)

138) 转换信息产权 (Transformational Information Property)

一种信息产权, 对它的保存是判断不可逆转换 (Non-Reversible Transformation) 充分保存了信息内容的必要非充分条件。转换信息产权与提交信息真实性 (Authenticity) 的证据一样重要。诸如此类的信息产权需要与特定的表示信息 (Representation Information), 包括语义信息, 联系起来表示其内涵和意义。“重要产权 (significant property)”一词在文学中有多重定义, 但有时表达的就是转换信息产权这一意思)。

(来源: OAIS)

139) 唯一标识符 (Unique Identifier)

根据某一标识模式对特定对象设置的字符串, 保证对象被唯一地标识。

(来源: NDSA)

140) 单元描述 (Unit Description)

一类信息包描述, 专门为访问辅助工具 (Access Aids) 提供其所需的有关档案信息单元 (Archival Information Unit) 的信息。

(来源: OAIS)

141) 校验 (Validation)

根据一组数据校验规则对数据进行校验以验证其正确性和有用性的流程。可能会涉及封装规则、文件结构或者特定文件格式的配置文件。

(来源: NDSA)

142) 验证 (Verify)

检查数据文件的副本以确定其与原始数据文件完全等同, 或随着时间的推移未发生变化的流程。

(来源: NDSA)

编译自: <http://www.dpconline.org/publications/digital-preservation-handbook>

<http://www.digitalpreservation.gov/nds/nds-glossary.html>

<http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/CCSDSAgency.aspx>