



国家数字文献资源长期保存发展与挑战

National Digital Preservation Program

NSTL NDPP项目组

2020.10.13

张晓林

使用许可: CC_BY_NC_ND

国家数字文献资源长期保存发展与挑战

- 1 战略规划
- 2 核心机制
- 3 实施成效
- 4 发展与挑战

1 战略规划

- 1.1 面临趋势
- 数字文献已经成为学术信息的主流和首要形态
 - 许多重要文献资源已经以e-first甚至e-only形式出版
 - 开放获取出版物、预印本库、数据存储库等
- 数字文献已经成为学术机构的主流和首要“馆藏”
 - 经费上占主要份额、使用上占主要份额
- 数字文献已经成为用户使用的主要环境和主要资源
 - 使用“数字化”和“网络化”和新常态新习惯
- 数字文献仅仅是数字知识的一小部分
 - Evolving digital information ecosystem
 - 数据、多媒体、教育资源、社交媒体、智慧生活圈.....
- 数字文献及其数字信息环境提供了巨大使用优势
 - 发现、关联、整合、重用...新常态、新业态

1 战略规划

■ 1.2 面临挑战

■ 数字内容实质上处于濒于损毁的状态

■ 数字知识内容长期保存已成为关键战略需求

■ 内在挑战：

- 介质的长期可靠性和可用性
- 内容的完整性和可使用性
- 内容的长期可获得性
 - 技术系统可靠性、网络链接可靠性、市场健壮性
 - 提供者管理可靠性、自然风险
- 地缘政治的可靠性：卡脖子风险

■ 长期保存挑战

- 订购模式只获得网络使用权
- 永久获取权的不可靠性、本地镜像站点的可控性
- 采购合同的限制（禁止系统化下载和拷贝）

■ 1.3 多种可能机制

■ 国家级图书馆（KB、BL、LC...）

- 基于法定存缴制度、侧重“本国出版物”
- 公共投入、有限服务

■ Portico

- A digital DARK ARCHIVE service provided by ITHAKA
- 社群管理机制（Board of Trustees，图书馆代表和出版社代表）
- 经济机制：出版社和图书馆均付费
- 公共服务机制：触发事件

■ CLOCKSS

- 社群管理机制（Board of Directors，图书馆代表和出版社代表）
- 经济机制：出版社和图书馆均付费
- 保存体系：分布式系统（12个，包括香港大学）
- 公共服务机制：触发事件

■ 1.4 中国选择

■ 是否长期保存？

- 创新驱动发展、创新支撑竞争
- 数字学术文献是关键战略资源，又是**濒于损毁**资源
- 必须采取可靠措施保证其长期可获得性

■ 长期保存什么？

- 在有限能力和有限时间下：
- 资源重要性、资源使用度、资源可用性风险……
- 权益可获得性、技术成熟性、管理成熟性……
- 可逐步覆盖其他资源

■ 如何长期保存？

- 本土保存、自主可控（覆盖国外文献、防止恶性环境）
- 程序合法、权益平衡（符合市场规律、建立可信赖机制）

- 1.4 中国选择
- 是否有权长期保存？
- 长期保存权：
 - 保存权：Right to archive for long term use
 - 保存处理权：Right to processing for archiving
 - 保存服务权：Right to providing service under certain conditions
 - 合作保存权：Right to collaborative preservation and succession
- 长期保存权来源
 - 采购资源是采购长期可用的知识内容，本身内含长期保存权
 - 专门法律保障
 - 著作权保护限制与例外→图书馆保存权
 - 网络信息传播保护条例→图书馆保存权
- 长期保存权的获得
 - 市场博弈+合同安排（与出版社的长期保存协议）
 - 通过具有法律效力的机制落实权益和保障权益
 - 诚信尽责保护出版社的合法权益

1.5 中国战略

- 作为国家科技信息资源战略保障，长期保存我国科技创新用户所需的主要数字文献资源，同时积极参与国家教育文化社会各领域所需的其他数字文献资源的长期保存。
- 主要选择使用价值高、学术价值高、风险高、保存可操作性强的数字文献资源，优先部署长期保存，并创造条件保存其他具有保存价值和消除风险作用的数字文献资源。
- **使用价值**：当前及长期需求量（采购机构和实际使用量）
- **学术价值**：在科学研究和国际发展中的长期学术价值、且保证系统性历史地覆盖重要资源；
- **风险程度**：资源依赖度、资源内容或使用渠道不可替代性及其风险程度；
- **可操作性**：权益安排可接受度、内容格式规范化程度、所需保存技术的掌握程度、保存机制的可负担性、合作协调能力等。

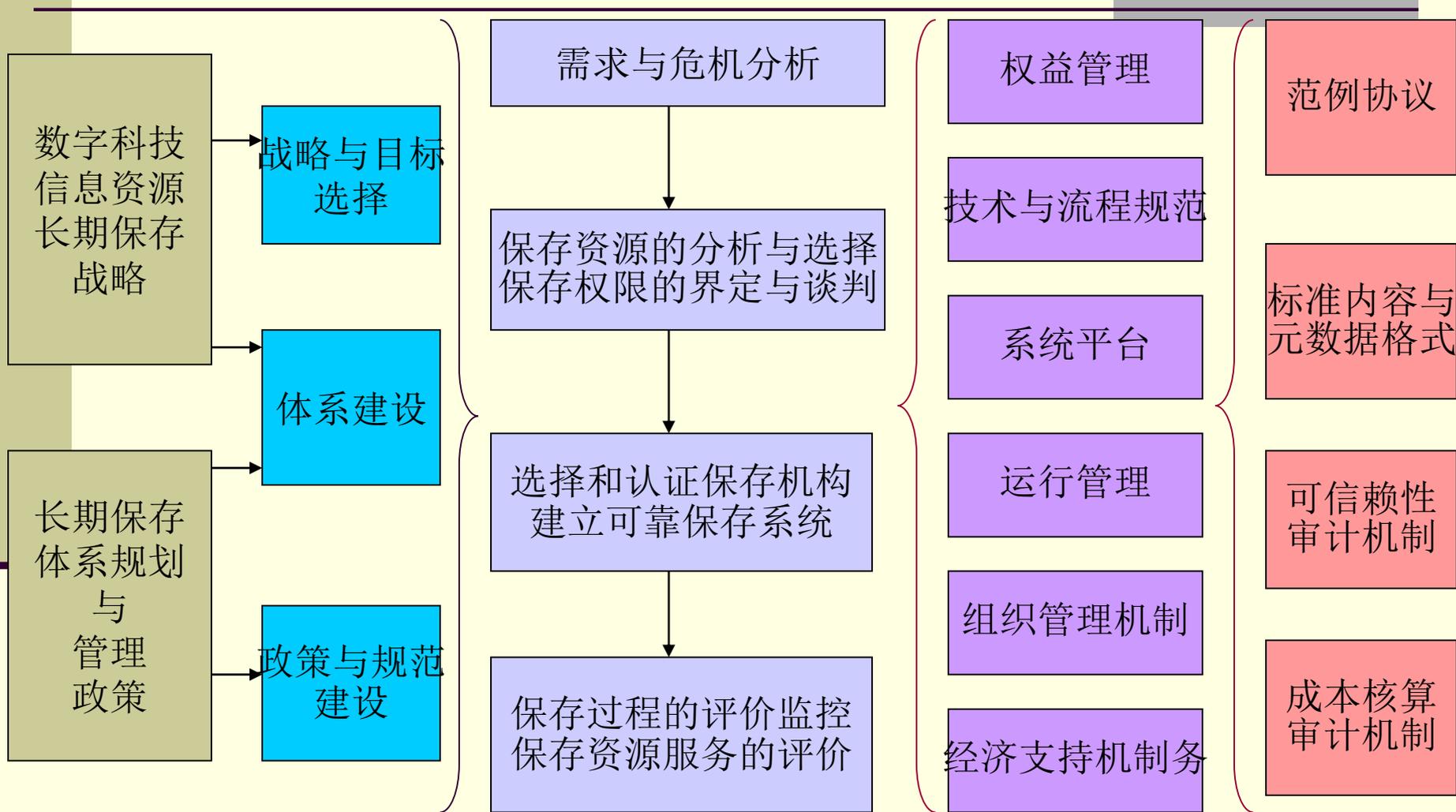
1.5 中国战略

- 首要目标范围：学术期刊、学术专著、会议论文、学位论文等
- 其他目标范围：密切相关的管理、经济、行业、市场等资源领域
- 目标资源类型：
 - 1) 主要国际综合出版商的期刊和会议录数据库及这类数据库的回溯内容；
 - 2) 主要国际专业学会/协会和大学等专门出版商的期刊和会议录数据库及这类数据库的回溯内容；
 - 3) 主要国际学位论文数据库
 - 4) 主要国内期刊、会议和学位论文数据库；
 - 5) 主要的开放获取期刊、会议录、专业知识库等数据库；
 - 6) 重要数字学术专著数据库和重要数字工具书资源；
 - 7) 其他重要数字知识资源。

国家数字文献资源长期保存发展与挑战

- 1 战略规划
- 2 核心机制
- 3 实施成效
- 4 发展与挑战

2.1 规划体系



2.2 运行原则

- **国家主导**：作为国家文献保障与服务平台的有机部分，由国家投资，由NSTL管理；
- **联合参与**：动员国内图书馆合作参与，并选择若干图书馆作为“合规保存机构”联合承担保存任务；
- **可靠管理**：保证重要数字文献资源在大陆本地可靠长期保存，保证各个长期保存系统的可靠运行和可靠监管，保证与文献出版市场的稳定有机结合；
- **公共服务**：长期保存体系及其服务作为公益性公共服务由国家支持（NSTL/NDPP年度预算）

2.3 组织机制

- 通过分工合作的合规保存机构作为保存节点具体实施
- 合规机构条件：
 - 公共事业法人；
 - 本身采购数字资源并提供服务；
 - 支持NDPP长期保存战略目标和运行目标；
 - 参加或承担被保存资源的保存谈判；
 - 承担长期保存的法律责任、接受公共认证审计；
 - 提供被保存资源在触发事件下的公共获取服务；
 - 遵守长期保存体系的法律、技术和管理规范；
 - 提供长期保存系统的基础设施、人员和管理支持；
 -
 - 适当兼顾行业和地区分布。
- 当前保存节点
 - 中科院文献情报中心、中国科技信息所、北京大学图书馆

2.4 可信赖性机制

■ 如何保障长期保存本身的可信赖性

- 你说保存的内容确实得到保存、确实能在需要时提供服务
- 整个保存过程和系统能长期可靠地做到上述要求

■ 长期保存可靠性的系统观

- 权益管理可靠性
- 技术流程可靠性
- 系统过程可靠性
- 管理机制可靠性
- 经济支持可靠性
- 人员队伍可靠性……

**Trusted Digital Repositories:
Attributes and Responsibilities**

An RLG-OCLC Report

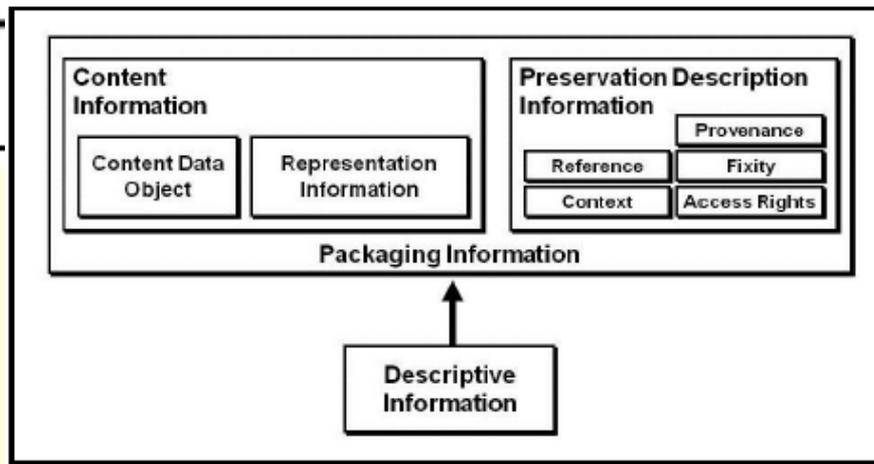
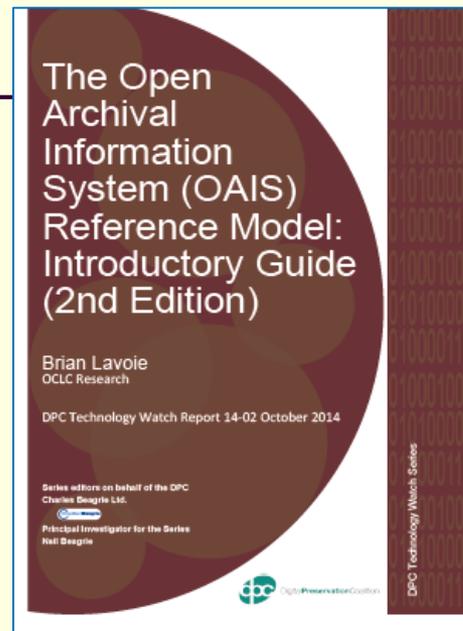
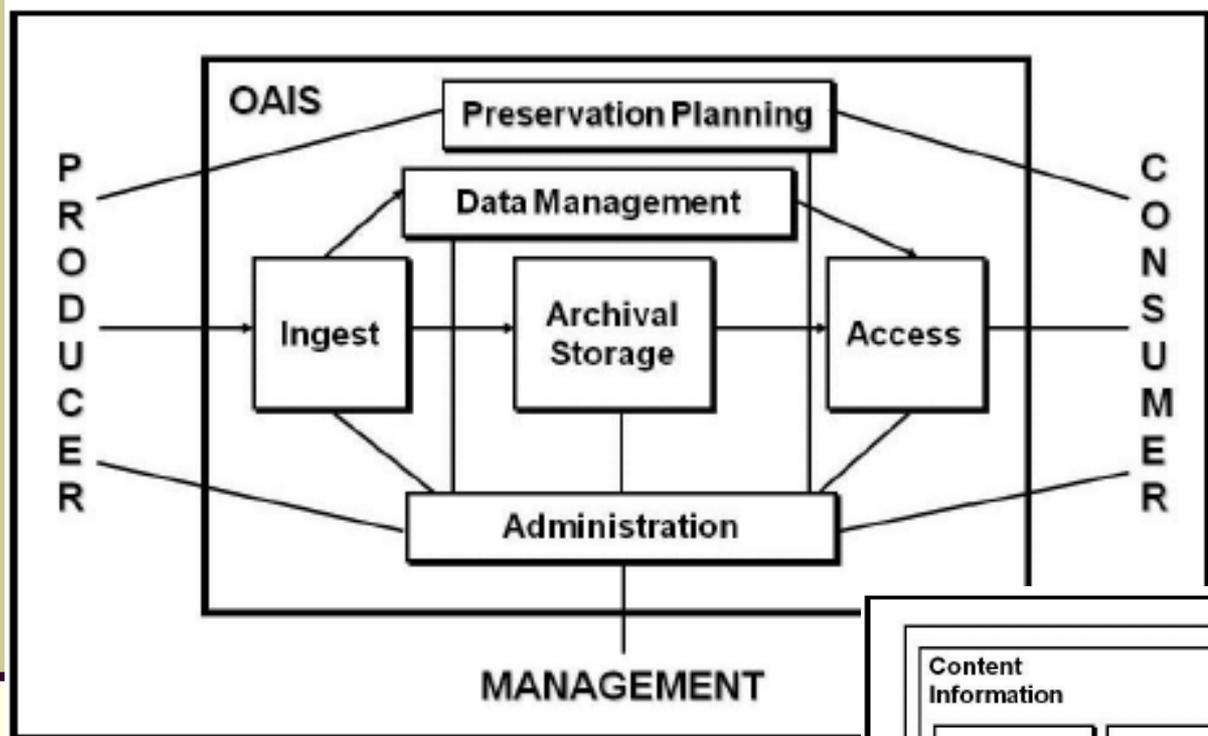
■ 长期保存服务的可靠性审计

- 审计什么？（保存审计和保存机制审计；审计标准）
- 谁来审计？（自查审计、第三方审计、用户审计…）
- 审计细粒度？（期刊/卷/期/篇/元数据和内容数据？）
- 保存审计与保存机制认证

2.5 公共服务条件

- NDPP本质上是Dark Archive，只在资源提供者无法提供正常服务时才由保存者向订购者提供检索和获取服务
- 服务触发条件（Trigger Events）
 - **Publisher No Longer in Business**
 - **Title No Longer Offered**
 - **Back Issues No Longer Available**
 - **Catastrophic or sustained failures**
- 触发事件后的服务：
 - 确认触发事件
 - 对授权用户开放公共服务（检索与获取）
 - 用户授权管理、使用统计、异常与非法使用监测
- 触发条件消失后处置：
 - 触发条件消失的确认
 - 关闭公共服务

2.6 技术体系（遵循OAIS模型）



2.6 技术体系（可靠流程与系统）

确保可靠、安全、高效、可管理的长期保存系统

- 保证数字媒介本身的可靠性
- 保证数字内容本身的完整性
- 保证摄入和保存过程的准确、高效和可管理
- 保证存档系统的可靠和可管理
- 保证公共服务系统的可靠和可管理
- 保证整体过程的规范性和可审计性

保证整体系统的安全性

- 物理空间的可控性（Access Control）
- 保存系统的可控性（stand-alone & authorized only）
- 高风险操作的可管理性（authorization & monitoring）
- 鲁棒和冗余的软硬件配置
- 周期性重授权和实时监测……

2.7 保存审计与可信赖性认证

远程保存审计

- 10所代表性机构（数据库授权机构）、指定数据库
- 指定授权审计人员、指定审计时间范围
- 随机选取审计范围特定量的刊、卷、期、篇，对比期刊网站
- 严格监测使用状况
- 填报审计情况

现场可信赖性认证

- 管理机制、人员队伍、经费投入
- 保存资源的权益管理系列文件
- 保存系统的配合与管理现状
- 保存系统的安全管理现状
- 现场保存审计……

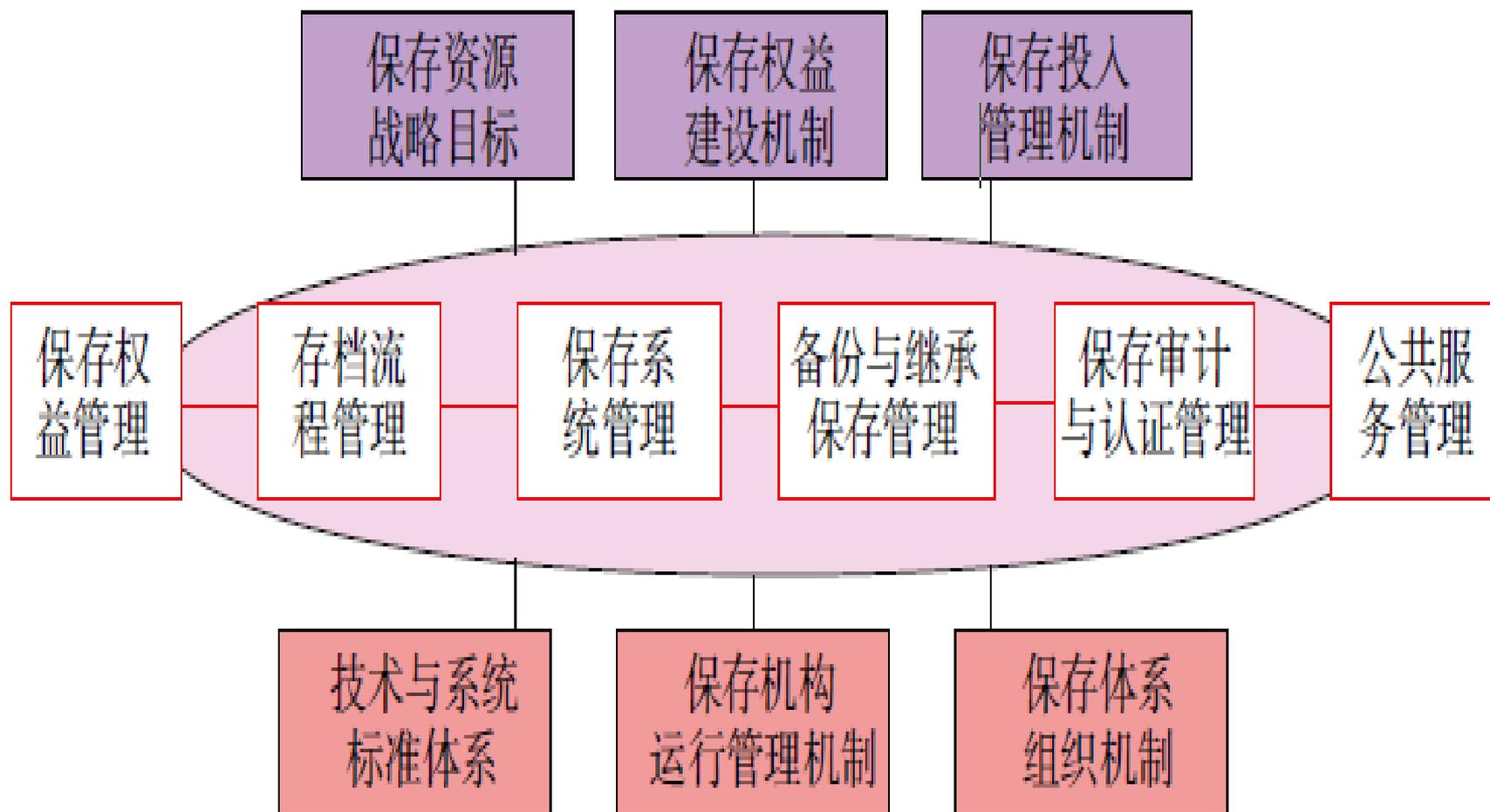
国家数字文献资源长期保存发展与挑战

- 1 战略规划
- 2 核心机制
- 3 实施成效
- 4 发展与挑战

3.1 发展过程

- 2004 中科院和NSTL举办国际iPres会议
- 2004 NSTL启动数字化科技信息资源长期保存研究项目
- 2007 中科院启动数字科技文献长期保存服务系统建设，研究和试验数字期刊与图书长期保存
- 2009 中科院与Springer 签署长期保存协议并开始存储其电子期刊，随后又与NPG、Wiley等签署长期保存协议
- 2013 科技部正式同意由NSTL牵头组织NDPP
- 2014 NSTL开始国家数字科技文献长期保存体系建设
 - 资源保存、系统建设、规范规则
 - (1) 示范系统，2014，中科院节点和中信所节点
 - (2) 基本系统，2015-2016，扩展建设若干节点
 - (3) 体系完善，2017-2020，覆盖所有重要资源

3.2 总体实施体系



3.3 全国协同的权利声索和权利管理

■ 2015年《数字文献资源长期保存共同声明》

- 图书馆对所采购数字文献资源拥有长期保存的权利。
- 图书馆采购数字文献资源实质上是采购知识内容为用户提供当前和未来服务，因此资源的长期可靠利用是这种采购行为的内在含义和基础条件。
- 没有可靠的长期保存，图书馆对所采购知识内容的长期利用权就不完整和不可靠，因此长期保存是图书馆持续进行资源采购的必要基础，是对出版社销售资源的重要支持。
- 长期保存权包括：数据存档权、数据处理权、数据服务权、合作保存权及继承受保存权。后两者是保存权的自然延伸，也是长期保存权利平等实现的基础。

3.3 全国协同的权利声索和权利管理

- NDPP受NSTL委托、代表国家和签署共同声明的图书馆行使上述长期保存权
- NDPP充分维护出版社的合法权益，诚信尽责保护出版商的合法权益，建立可信赖机制保证相关参与方遵守长期保存中的权利与义务。
- NDPP充分维护图书馆界的合法权益，诚信尽责保护参与图书馆合法权益，建立常规化机制接受参与图书馆的监督与检查。
- NDPP积极推动建立行使长期保存权的制度化机制
 - 推动国家确立“使用公共资金采购数字文献资源机构”长期保存责任，机构应要求所采购资源在中国本土得到长期保存。
 - 建立并支持权益谈判机制。
 - 根据分工向出版商提供提出长期保存要约。
 - 结合资源订购谈判进行长期保存协议谈判。
 - 建立详细和可靠的长期保存权益管理机制，NDPP登记；
 - NDPP对长期保存协议的执行进行检查和保存审计。

3.4 技术体系

NDPP Network

Management Node

用户认证 审计报告 资源报告

权限管理

命名空间分配

节点注册管理

全局PID解析

元数据收割汇集

节点备份/复制管理

注册、日志、元数据，报告

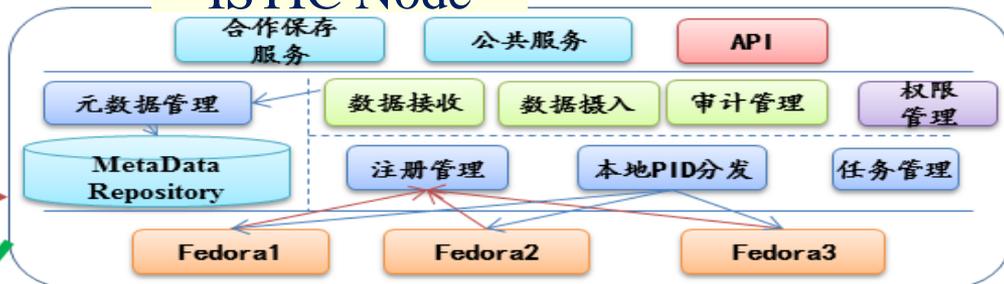
备份复制

注册、日志、元数据，报告

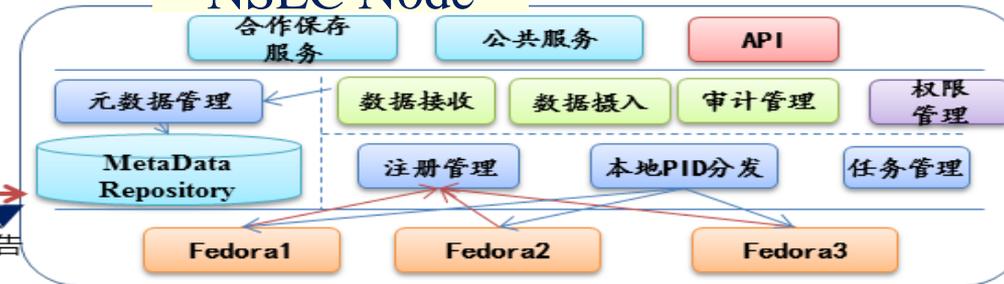
备份复制

注册、日志、元数据，报告

ISTIC Node



NSLC Node

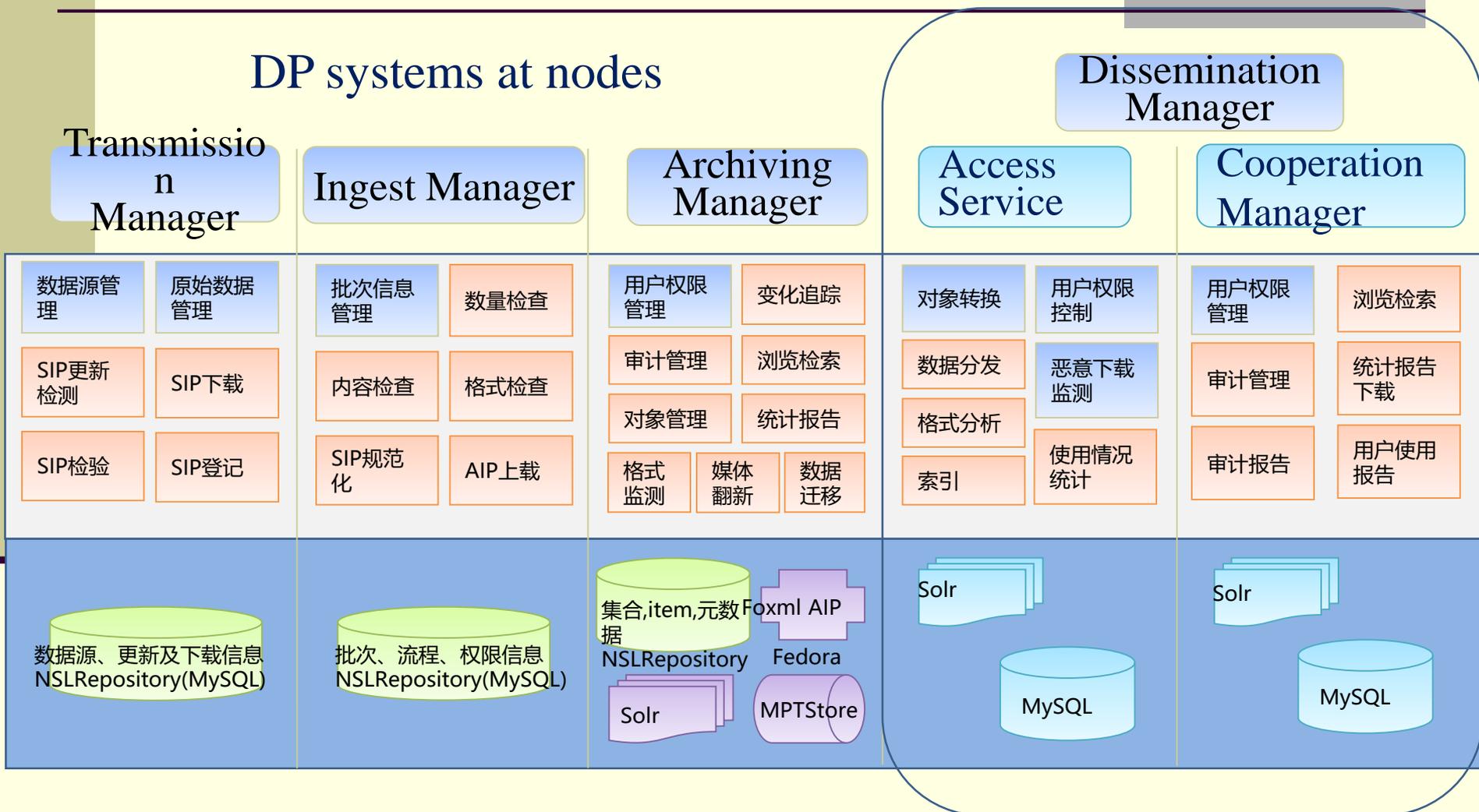


PU Node



3.4 技术体系

DP systems at nodes



3.5 已保存资源

- 万方科技资源库
- Springer期刊库
- IOP期刊库
- Elsevier期刊库
- Wiley期刊库
- PNAS期刊库
- AGU期刊库
- Springer科技图书库
- Springer回溯图书库
- IOP图书库
- Springer Protocols库
- ProQuest论文库
- CUP期刊库
- 中小型学协会期刊库（45家、800多种期刊）
-
- 维普中文期刊库
- NPG期刊库
- RSC期刊库
- Elsevier图书库
- BioMed Central期刊库
- OUP期刊库
- Taylor & Francis期刊库
- RSC图书库
- Wiley图书库
- Springer Protocols回溯库
- arXiv预印本库
- Emerald期刊库

3.6 可信赖审计与检查

■ 保存机构保存进展报告

- 月度报告：应保存多少、每次接收多少、实际保存多少

■ 保存机构自我保存审计报告

- 年度自我保存审计报告：针对每种资源，系统自动统计

- 应保存数量、接收数量、实际保存数量、累积保存数量

■ NDPP组织第三方机构保存审计

- 遴选非保存机构的10家合作保存机构

- 远程随机抽取审计对象（期刊、卷期、篇）

- 实时对比期刊网站

- 统计分析多少应该保存的对象得到可靠保存

■ NDPP组织对合规保存机构认证审计

- 按照长期保存系统可信赖认证标准

- 组织非合规保存机构和其他保存机构专家

- 现场进行保存审计、检查管理制度、检查保存日志等

国家科技数字文献资源长期发展与挑战

- 1 战略规划
- 2 核心机制
- 3 实施成效
- 4 发展与挑战

4.1 营造Preservation-enforcing环境

- 加强市场博弈、推动长期保存
 - 中国是出版大国、贡献大国
 - 出版社：DP-Friendly、DP-Resistant
 - 加强落实《互联网信息服务管理办法》
 - 加强落实公共订购机构长期保存责任
 - 国家资金和激励措施优先支持DP-friendly出版社
- 加强所有订购机构对长期保存的责任和推动力
 - 了解和支持NDPP
 - 订购时理直气壮、持之以恒、理利兼施地要求保存权
- 加强长期保存权利的联合谈判
 - NSTL+NDPP+合规保存机构+订购联盟共同谈判

4.2 扩大NDPP长期保存资源

- 扩大学协会出版社资源
- 扩大重点综合出版社资源
- 扩大电子图书出版资源
- 继续试验重点工具书数据库

- 适当扩大合规保存机构
 - 必须获得尚未得到保存的资源的保存权利
 - 或者获得特殊资源的保存权利

4.3 扩展长期保存研发工作

- OA出版资源、预印本资源
- 重要工具书数据库资源
- 重要IR集合资源
- 公共科学数据资源
- 科技机构网络资源
- 科研教育数字环境组织资源
-

4.4 加强长期保存系统建设

- 优化长期保存流程管理
- 优化自动化智能化安全管理
- 强化技术基础设施能力
-
- 加强DNPP机制化建立，保障可持续性运行与发展

共同的挑战、共同的命运、共同的努力

每个机构都面临数字信息资源长期保存的挑战

- 你已经面临数字内容为主的文献资源体系
- 你如何规划基于数字环境的文献资源建设政策？
- 你如何规划基于资源生命周期的数字内容可持续保存利用机制？
 - 谁来保存、如何保存、如何监督、如何可持续...
- 你如何参加NDPP：
 - 参与规划、参与谈判、参与培训、参与宣传
 - 参与对合规保存机构的保存审计和可信赖机构认证审计

国家科技数字文献资源长期发展与挑战

■ 谢谢！

- <http://www.ndpp.ac.cn/>
- <http://dpsm.las.ac.cn/>