

# 数字资源长期保存专题

---



中国人民大学信息资源管理学院 钱毅

2018-11

# 演讲者介绍



**钱毅 博士/副教授**

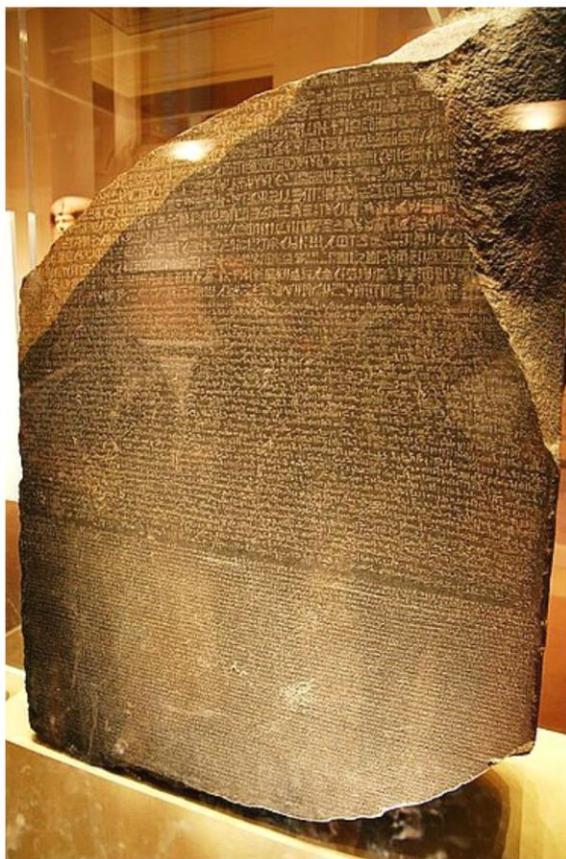
- 国家标准GB/T 29194-2012《国家数字资源管理功能需求规范》主持人
- 国家社科基金项目“数字资源标准体系建设研究”(2008)主持人
- 国家社科基金项目“基于组织的数字资源管理成熟度研究”(2014)主持人
- 国家社科重大基金项目“历史文化村镇数字化保护与传承：理论、方法与应用”子课题负责人(2016)

# 目录

- 1 数字资源长期保存概念理解
- 2 数字资源长期保存发展历程
- 3 数字资源长期保存主要项目
- 4 数字资源长期保存策略技术
- 5 数字资源长期保存标准规范
- 6 数字资源长期保存系统设计



# 从“罗塞塔”谈起



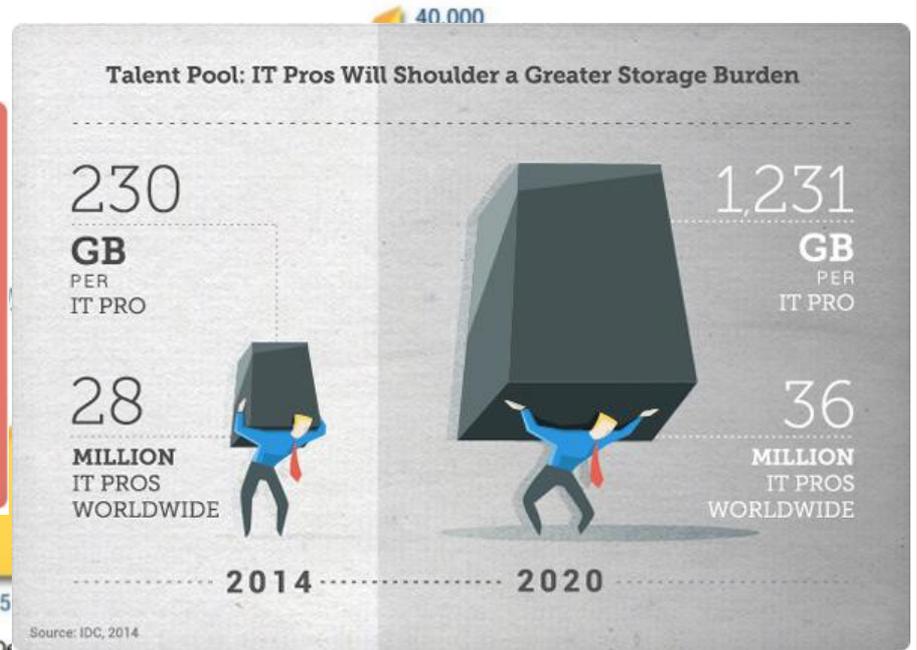
# 背景-数字资源大爆发

## The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020

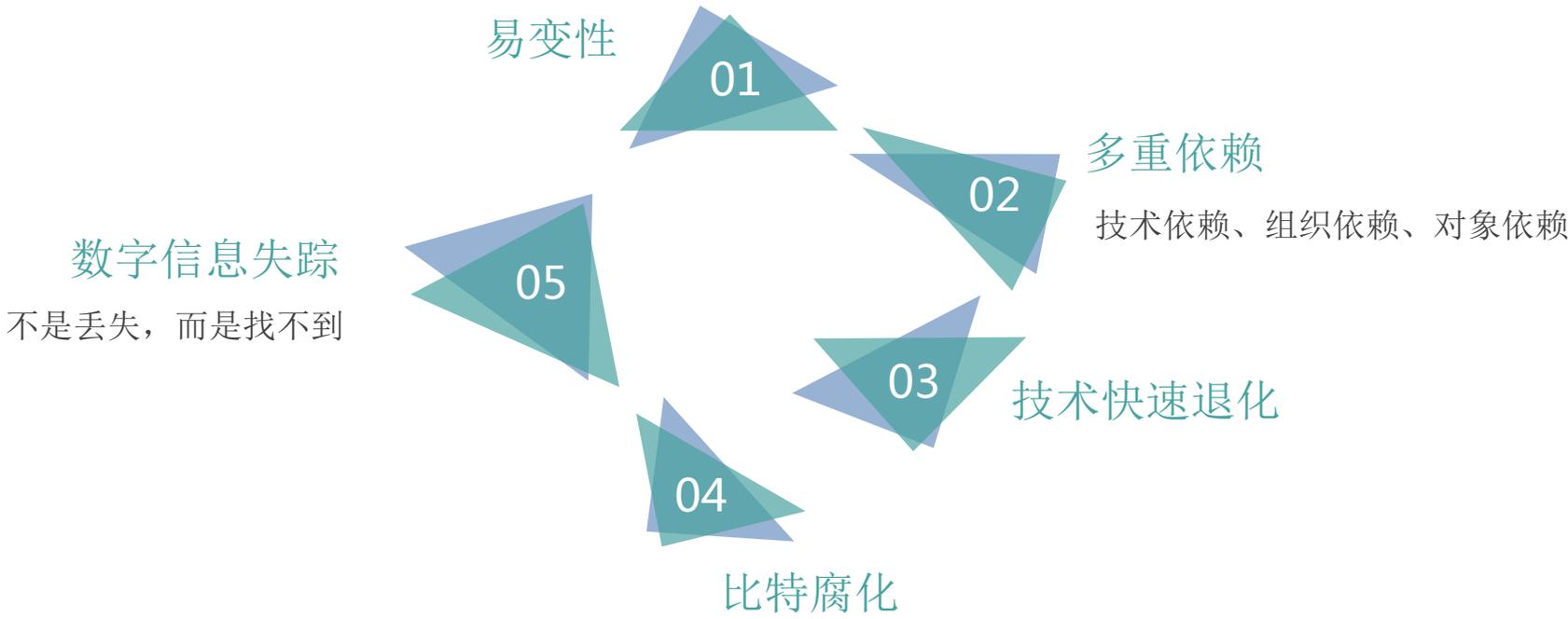


2009 2010 2011 2012 2013 2014 2015

Source: IDC's Digital Universe Study, sponsored by EMC, December 2013



# 背景-数字资源威胁加剧



# 1

## 数字资源长期保存概述

---

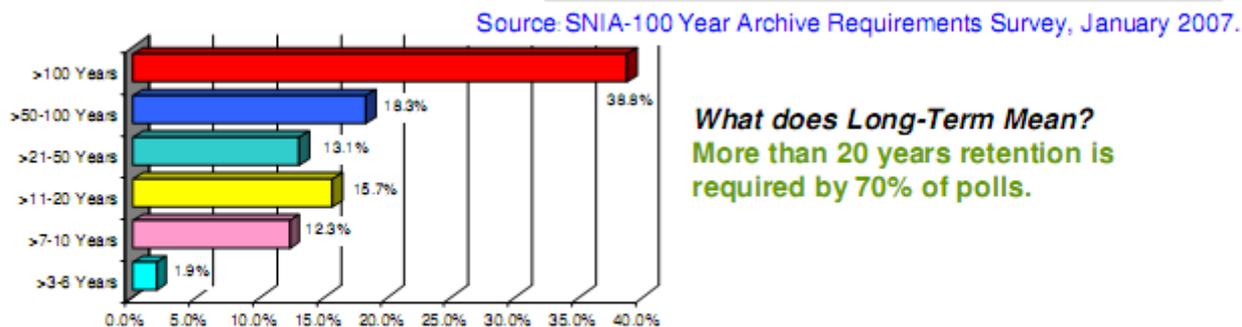
概念  
层次  
比较

# 概念-理解

## 长期保存

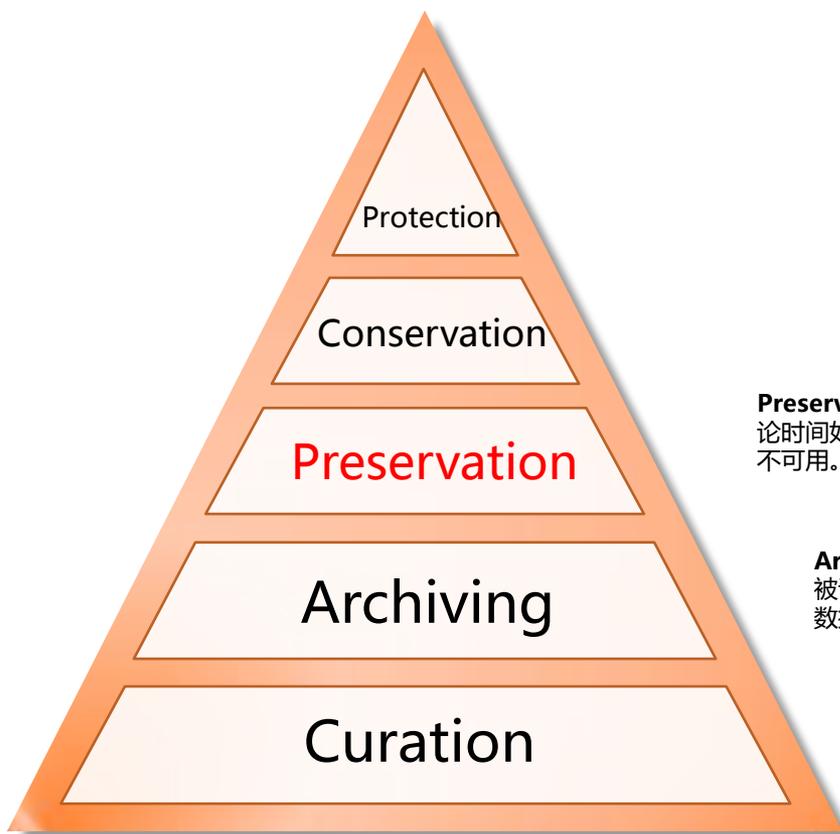
数字资源信息保持利用性和真实性的期限。

注：这个时间可能是几年至几百年，主要取决于机构的需要与要求。对于一些机构，保存时间由管理者的需要、法定需要和业务需要而决定。对于档案馆这样保存公共文件的机构，保存数字资源信息的期限通常为几百年。【ISO/TR 18492】



**What does Long-Term Mean?**  
More than 20 years retention is required by 70% of polls.

## ◆ 概念理解：什么是保存 **PRESERVATION**



**Preservation**：是在Archiving中的一种行为，它要求对特定的数据进行维护，不论时间如何流逝，这些数据始终能够被存取和理解，并不会随着技术的更新而变得不可用。

**Archiving**：是一种Curation行为，它确保数据合理地选择和存储，能够被访问，并且数据的逻辑和物理完整性随着时间的流逝能够得到维护，保障数据的安全和真实可信。

**Curation**：是指在数据产生之时就进行的对数据进行管理和促进数据利用的活动，这一活动确保数据符合当前的应用目的，并且能够被发现和重用。。更高层面的Curation不但包括数据的管理，还包括对相关数据标注内容及其他数据内容关系的管理。

## DP概念

### OAIS

一种长期地对信息进行维护，保持这些信息能被指定团体独立理解，并对其真实性提供证据支持的活动。

### DPC

英国数字保存联盟

一系列受管控的、确保数字资源能够持续不断被存取应用的活动，只要有需求，这些活动就需要不断持续下去。

### IBM白皮书

长期保存的过程，可以视为降低数字对象对软硬件环境的依赖性的过程。长期保存需要涵盖两个方面：比特保存（物理保存）和信息保存（逻辑保存）

## 与DP相关的概念

### Curation

数据产生之时就进行的对数据进行管理和促进数据利用的活动。不但包括数据管理，还包括数据标注及其他数据内容关系的管理。（更多的是一种积极的保存行为）

### Archiving

一系列受管控的、确保数字资源能够持续不断被存取应用的活动，只要有需求，这些活动就需要不断持续下去。

### Preservation conservation

保存：一种Archiving 行为，要求对特定数据进行维护，无论时间多久，始终能被存取和理解。

保护：一种preservation 行为

### Protection

防护：一种conservation行为

## DP概念

### 长期保存

保证可以持续地对数字资源的存取应用提供支持，或者至少保证数字资源中的信息内容可以无限制传递下去。

### 中期保存

保证在一个有期限的时间段内，即使经过技术变迁，仍可持续地实现对数字资源的存取应用。

### IBM白皮书

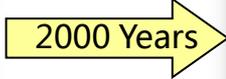
在一个可以预见的时间段内，能够实现对数字资源的存取应用。这个时间段内不会出现技术变迁。

# 比较：二者都存在物理保存与逻辑保存部分

The Dead Sea Scrolls

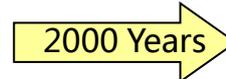


bits (data)



כי לוא רמה תורה למה ולוא המספר המרבה תולעה  
 חי וחי יודה למה יודו למה מול מוטמי רגל מדויקעה  
 המרבה להמה רבוקתמה תשמילם כי בודים נפש מול  
 חי נשמת מול בשרי אתה נתתה עשה עמנו  
 מנומה מרוב רחמימה וברוב גדוקתימה שמע  
 בקול אורבי שמו ולוא עזב הסדר מוחמה  
 ברוך עושה גדוקת סעמי חסדיו  
 חסד רחמיים שאנת נפשי לחלל שמחה להודות ברינה  
 המרבה להגיד אמורתמה לתהלתמה אין חקי למות

information

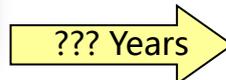


1. Surely a maggot cannot praise thee nor a grave worm recount thy loving-kindness.
2. But the living can praise thee, even those who stumble can laud thee. In revealing
3. thy kindness to them and by thy righteousness thou dost enlighten them. For in thy hand is the soul of every
4. living thing; the breath of all flesh hast thou given. Deal with us, O LORD,
5. according to thy goodness, according to thy great mercy, and according to thy many righteous deeds. The LORD
6. has heeded the voice of those who love his name and has not deprived them of his loving-kindness.
7. Blessed be the LORD, who executes righteous deeds, crowning his saints
8. with loving-kindness and mercy. My soul cries out to praise thy name, to sing high praises
9. for thy loving deeds, to proclaim thy faithfulness--of praise of thee there is no end. Near death

12 Mech 10 2 10

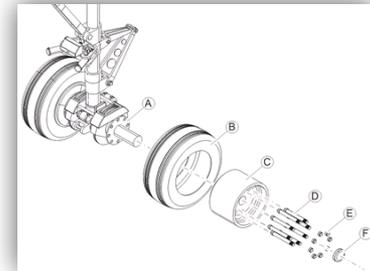
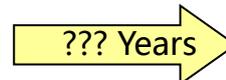


bits (data)



```
0010110100100110010
1001111001010010011
1100110100101110101
0100110010010010100
1001001011001011100
1101001011001011100
1101001010000101001
001
```

information

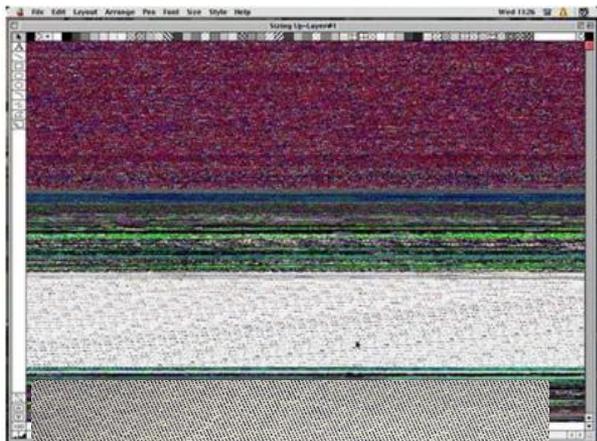


**Digital Preservation** is the means of ensuring usability in spite of obsolescence of everything hardware, software, formats, etc. **Archiving** addresses getting the bits into the future.

# 概念-数字资源长期保存

长期保存的过程，可以视为降低数字对象对软硬件环境的依赖性的过程。

比特保存与逻辑保存并重



Pits on CD

```
Ffd8ffe000104a46494600010201 SOI
008300830000ffed0fb050686f74 APP0 JFIF 1.2
6f73686f7020332e30003842494d APP13 IPTC
03e90a5072696e7420496e666f00 APP2 ICC
00000078000000000004800480000 DQT
000002f40240ffeeffee03060252 SOF0 183x512
0347052803fc0002000000480048 DRI
0000000002d80228000100000064 DHT
000000010003030300000001270f SOS
0001000100000000000000000000 ECS0
0000600800190190000000000000 RST0
0000000000000000000000000000 ECS1
000000000000000000000000003842 RST1
494d03ed0a5265736f6c75746966f ECS2
6e0000000010008313a300020001 RST2
008313a3000200013842494d04 ...
```

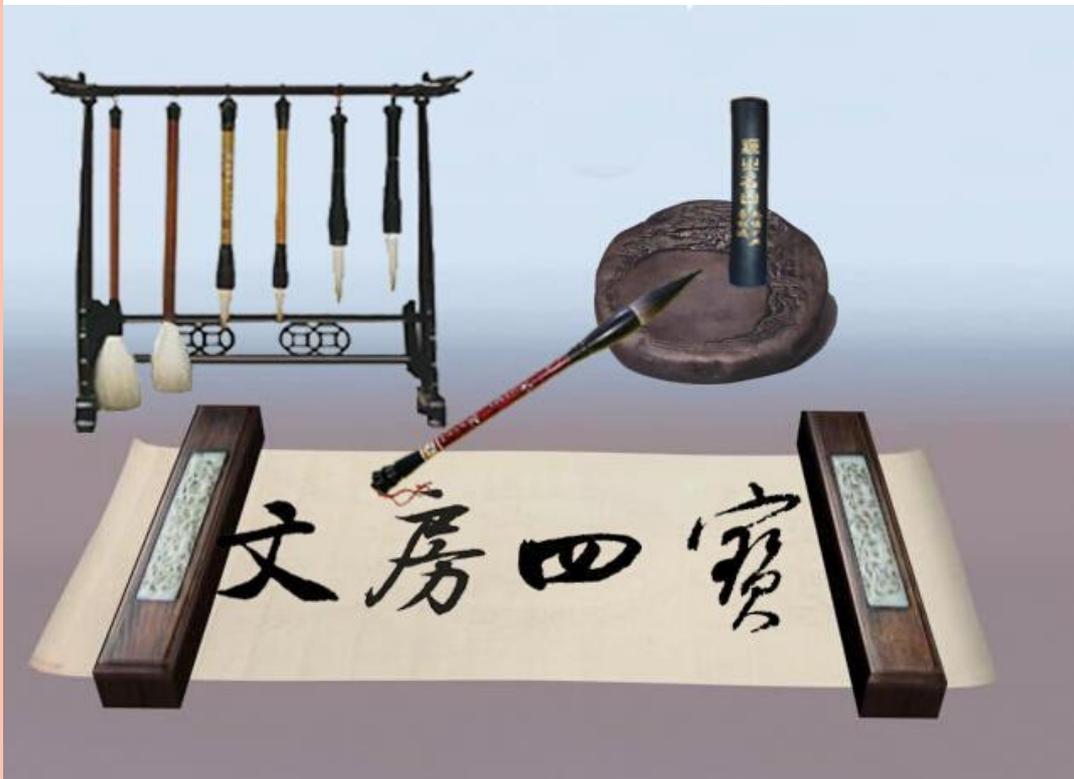


# 传统保存特点

- 总体上是狭义的保存概念，传统保存重视物理实体保管和外围环境防护。
- 传统资源要素保护内容具有较强稳定性



# 传统资源保护关注要点



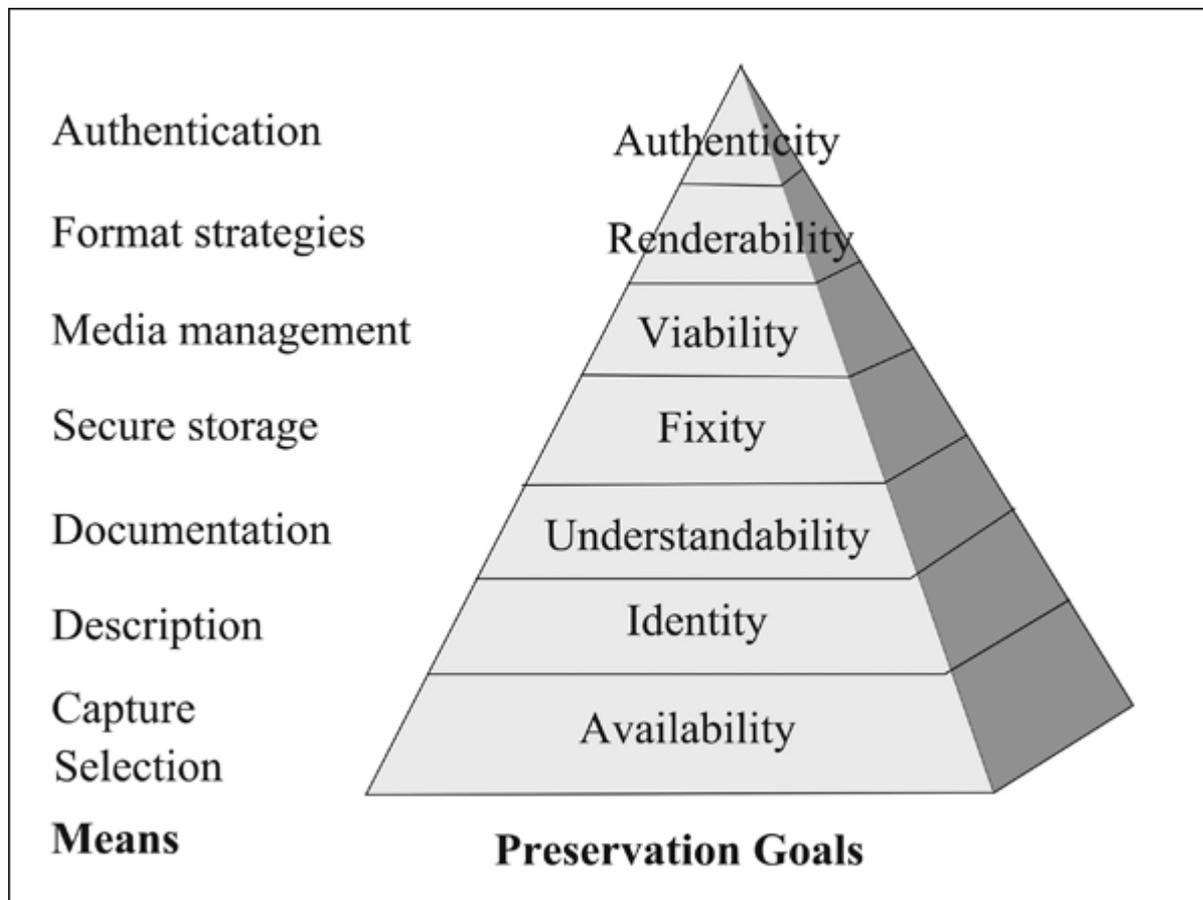
微观分析

- 防火



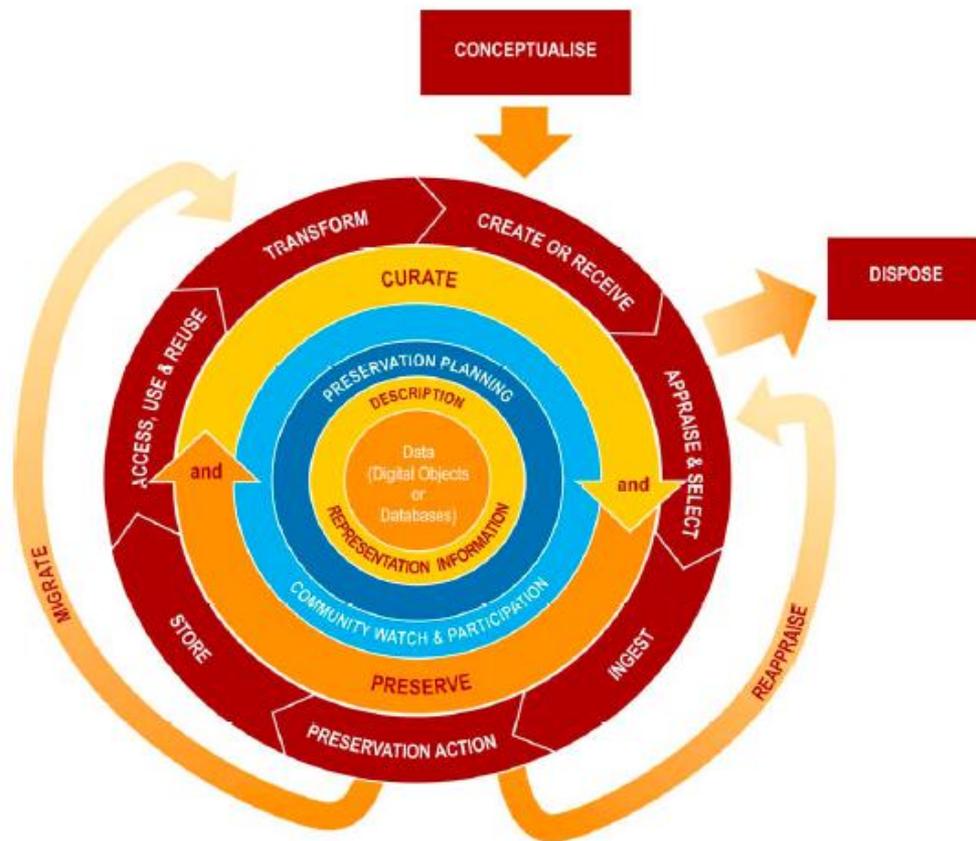
外部环境

# 层次-数字资源长期保存体系

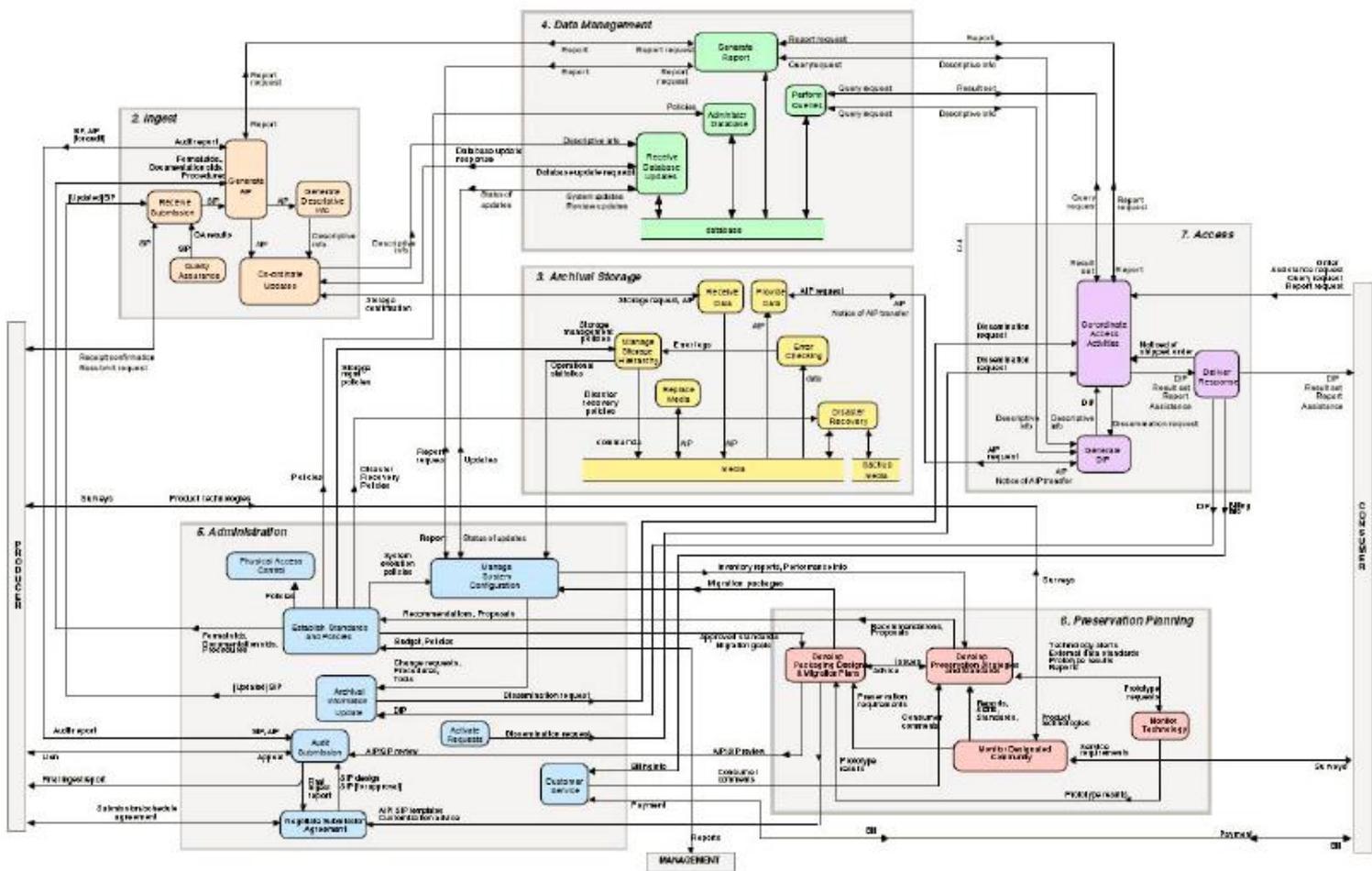


认证管理  
格式策略  
媒体管理  
安全存储  
文档记录  
资源描述  
归档选择

## ◆ DCC数字资源长期保存生命周期模型

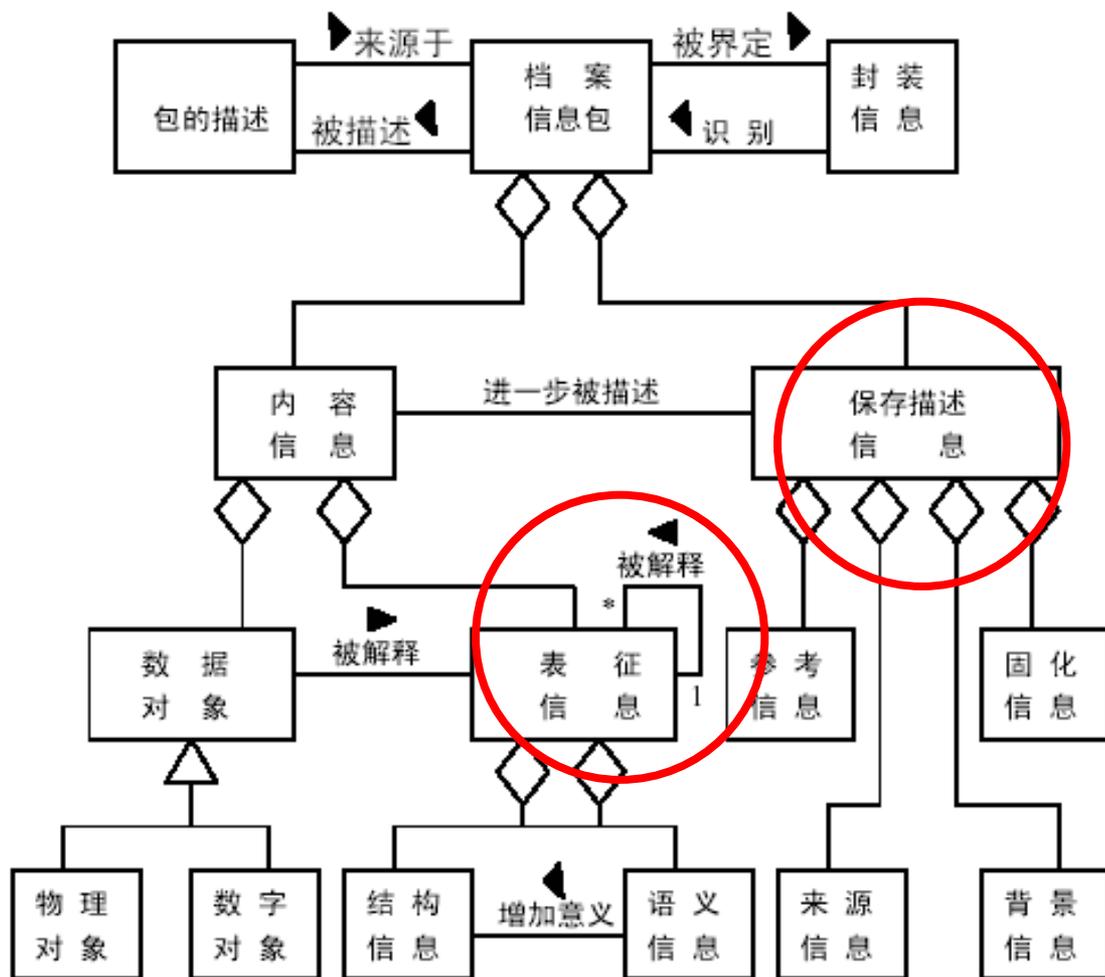


# 数字资源长期保存技术体系：OAIS功能模型



# 微观模型-数字资源长期保存

微观——信息模型



- OAIS认为信息是数据和表征信息（RepInfo）的结合体。
- RepInfo 具有迭代性，每一层RepInfo就意味着一定的依赖性
- 长期保存的核心就是维系是RepInfo循环
- 对于不同的用户，所需要的RepInfo可能不一样

# 长期保存本质理解：OAIS信息模型

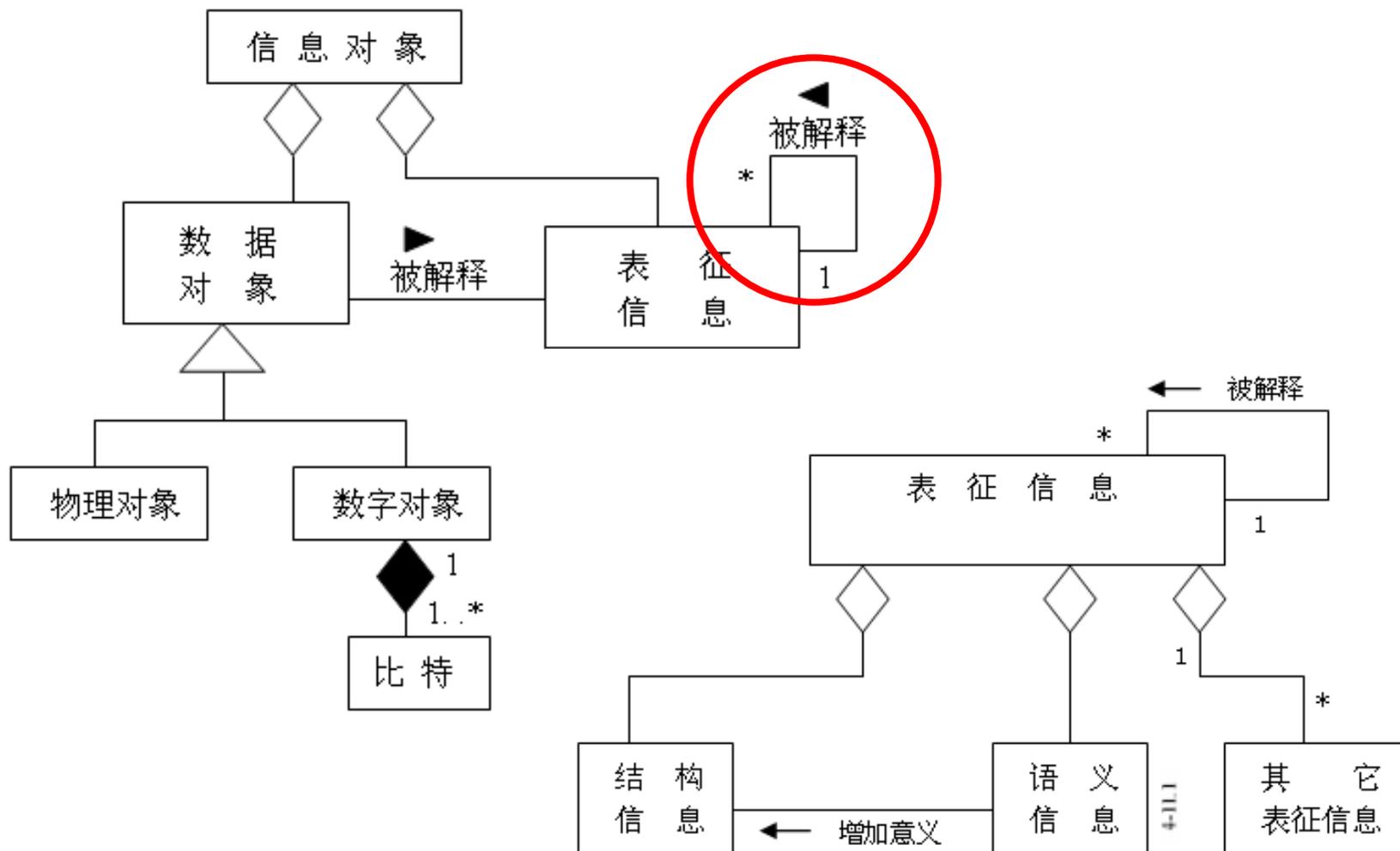


图 1.1.1: 表征信息对象

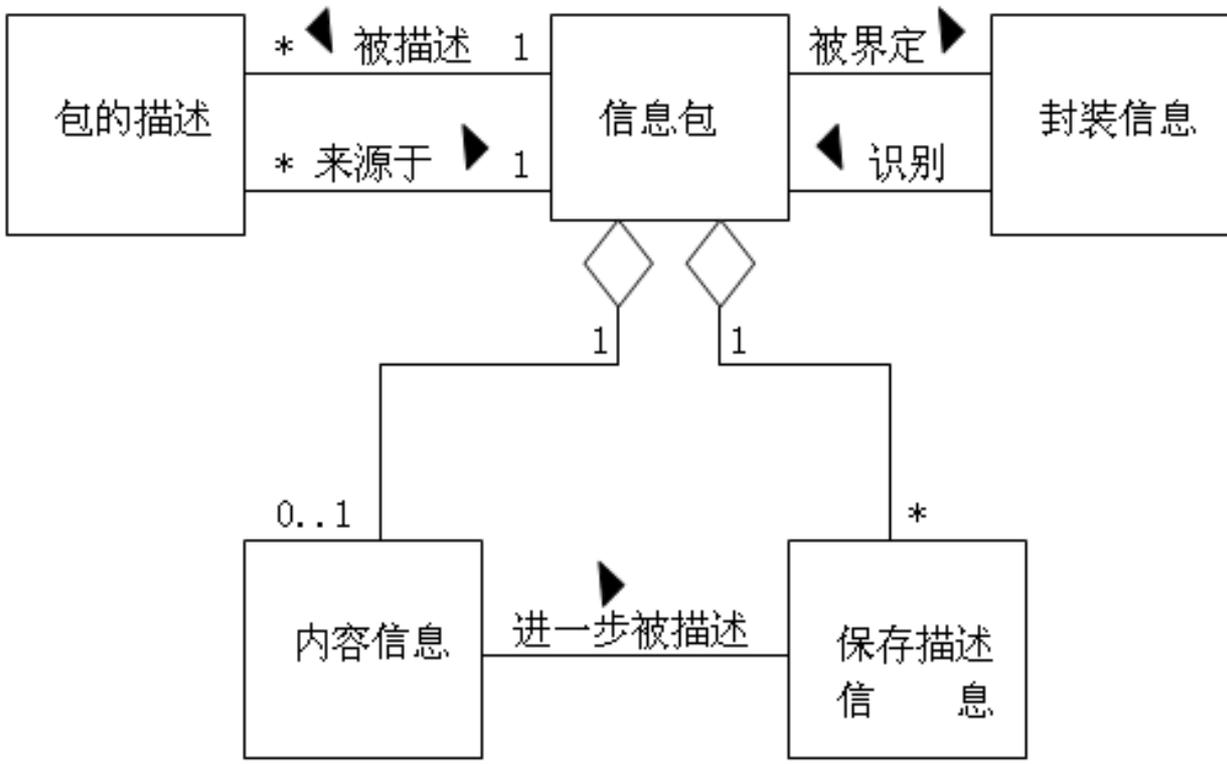


图 1-1-1: 信息包的内容

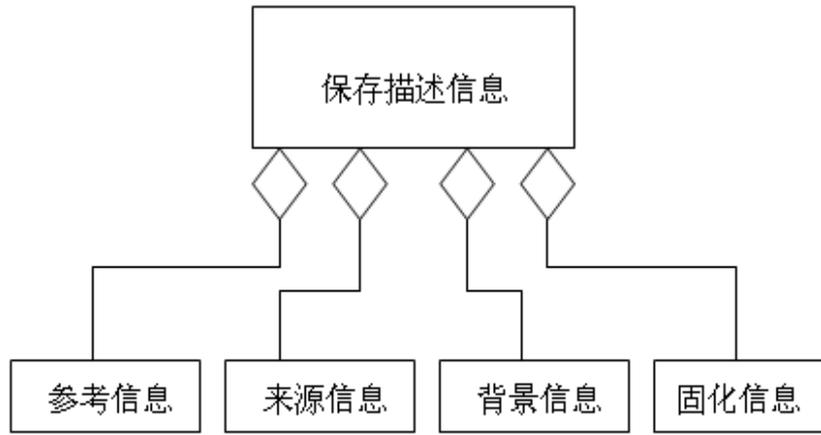


图 1-1-2: 保存描述信息(PDI)

# OAIS信息模型解读

- OAIS认为信息是数据和表征信息（RepInfo）的结合体。
- RepInfo 具有迭代性，每一层RepInfo就意味着一定的依赖性
- 长期保存的核心就是维系是RepInfo循环
- 对于不同的用户（designed community)而言，所需要的RepInfo可能不一样

# 2

## 数字资源长期保存发展历程

---

# Timeline: Digital Technology and Preservation

View All

General Developments

Protocols & Formats

Networks

Hardware & Software

Media

Crisis & Obsolescence

Organizational Response

go to >>

[1950](#)

[1960](#)

[1970](#)

[1975](#)

[1980](#)

[1985](#)

[1990](#)

[1995](#)

[2000](#)

[2005](#)

[2010](#)

2012

- ■ New and improved version of the [OAIS](#) standard released.
- ■ [PLANETS](#) wins the DPC Award for Research and Innovation for permanently changing the digital preservation landscape by "by shifting the focus to practical, sustainable solutions that are soundly supported by practice." 2013
- ■ NDIIP develops [Per: public](#) regarding personal
- ■ One year after the [DataUp](#) project is born to
  - ■ Tenth anniversary of the [Digital Preservation Management Workshop!](#)
  - ■ [Archive Team](#), a self-described "loose collective of rogue archivists, programmers, writers and loudmouths dedicated to saving our digital heritage," wins an NDSA Innovation Award for its work advocating for the preservation of digital culture within the technology and computing sectors.
  - ■ [Digital Public Library of America](#) launched.
  - ■ Tenth anniversary of the [International Internet Preservation Consortium](#).
  - ■ New [National Digital Stewardship Residency](#) sends first ten residents to the Washington, D.C. area.

Source:<http://www.dpworkshop.org/dpm-eng/timeline/popuptest.html>

## ◆ 概念提出阶段 ( ~1996 )



美国记忆项目启动

90年代中期

1994

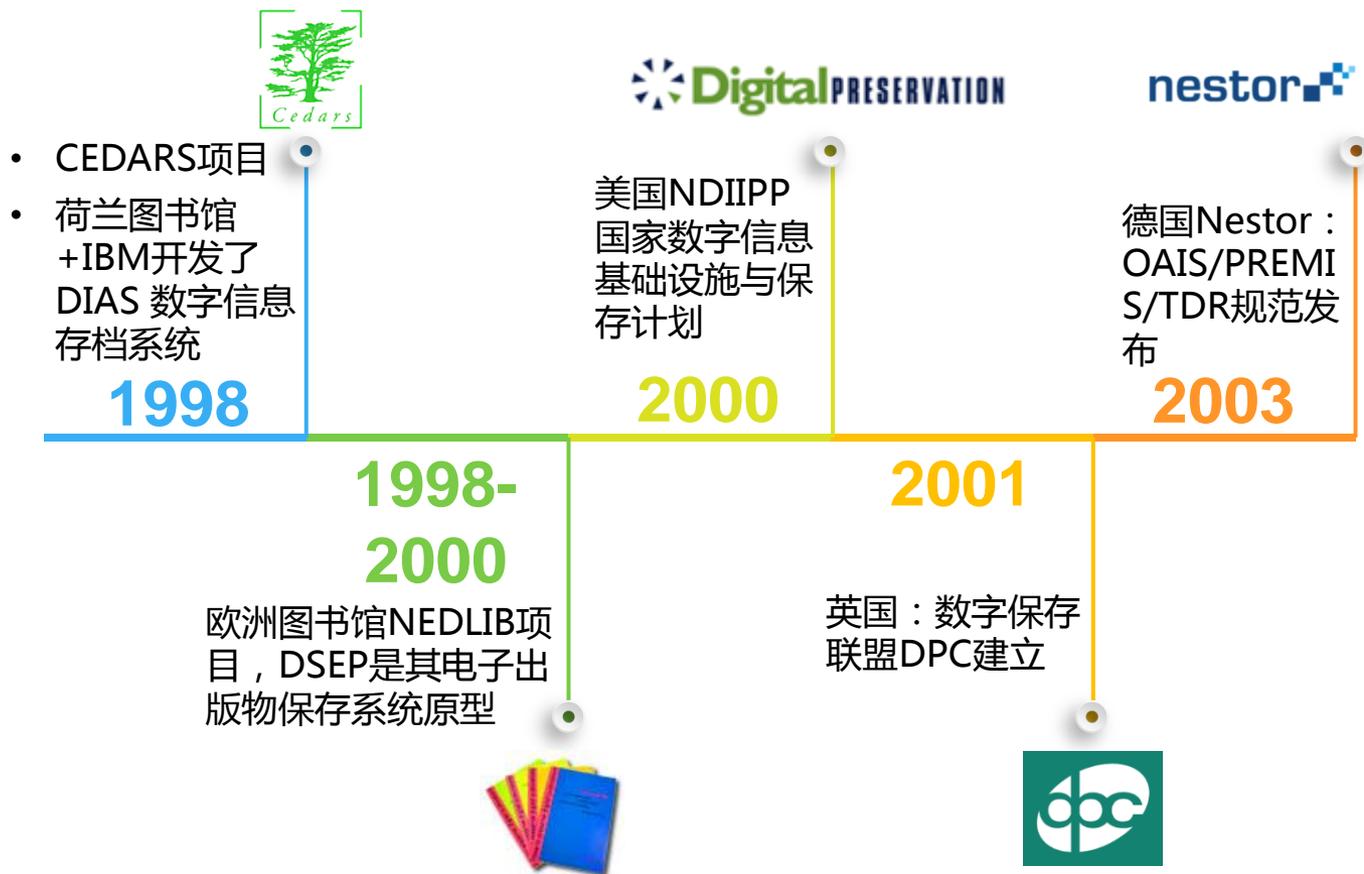
- ◆ NARA Paul Conway发表 digitizing preservation
- ◆ RLG 出版 digital imaging technology for preservation

《保存数字信息：数字信息存档特别工作组报告》公认的奠基性文献

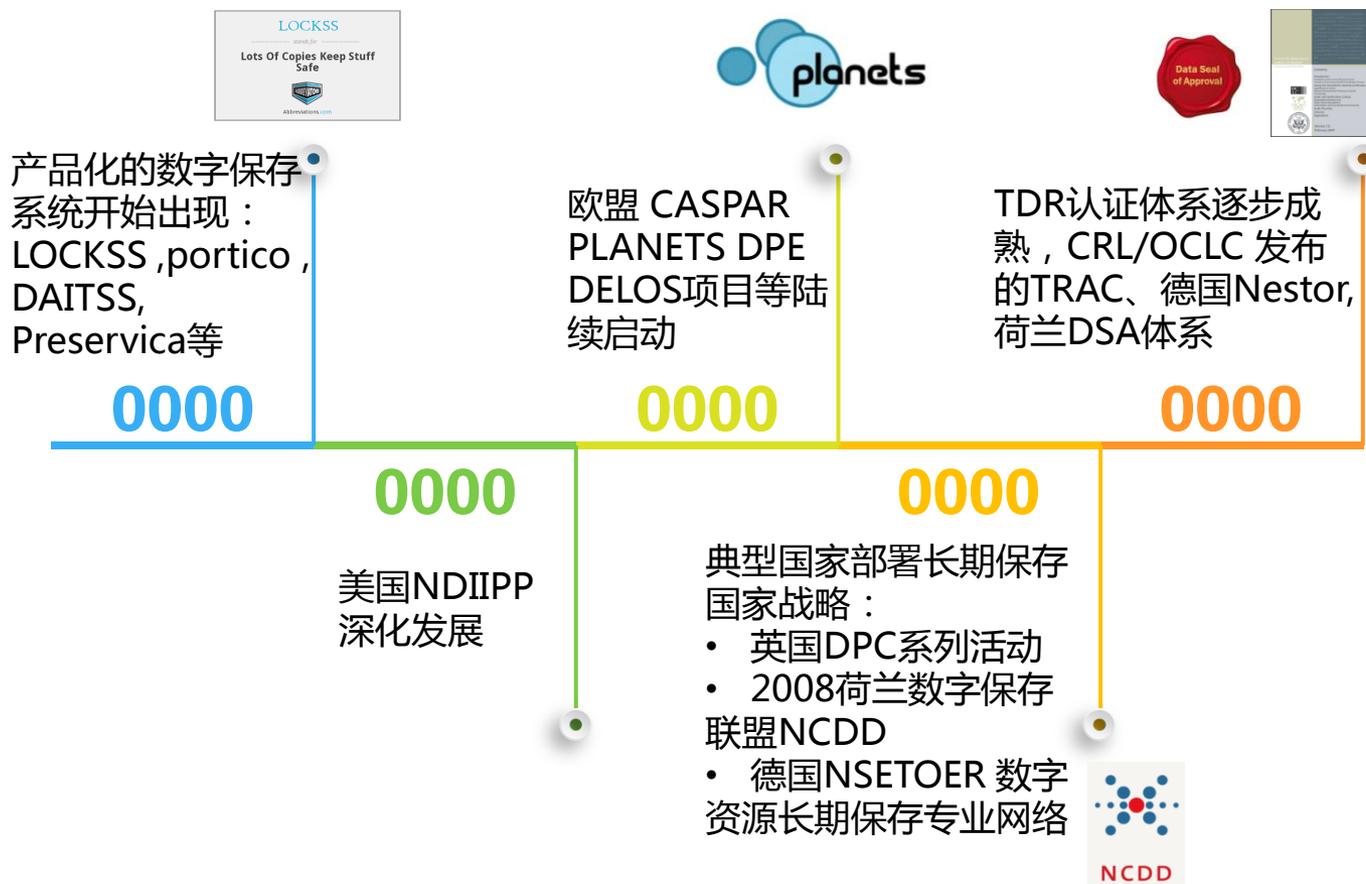
1996



## ◆ 数字保存理论基本成型阶段



## ◆ 数字保存深入发展阶段 ( 2000+ )



# 3

## 数字资源长期保存主要项目

---

典型项目

基于OAIS 进行组织的项目体系

# 典型项目

## InterPARES

InterPARES项目即“保障数字资源永久真实性国际合作项目”是一个由加拿大社会科学和人文研究委员会、美国国家历史出版和文件委员会、联合国教科文组织等联合资助的大型国际合作项目，旨在研究如何长久保障数字资源的真实性。

## OAIS

OAIS参考模型 (ISO 14721)是致力于长期保存和维护数字信息可存取档案系统的一个基本概念框架，框架描述了系统中与数字资料的长期保存相关的环境、功能组件和信息对象。包括功能模型、信息模型、互操作模型以及长期保存基本策略等内容。

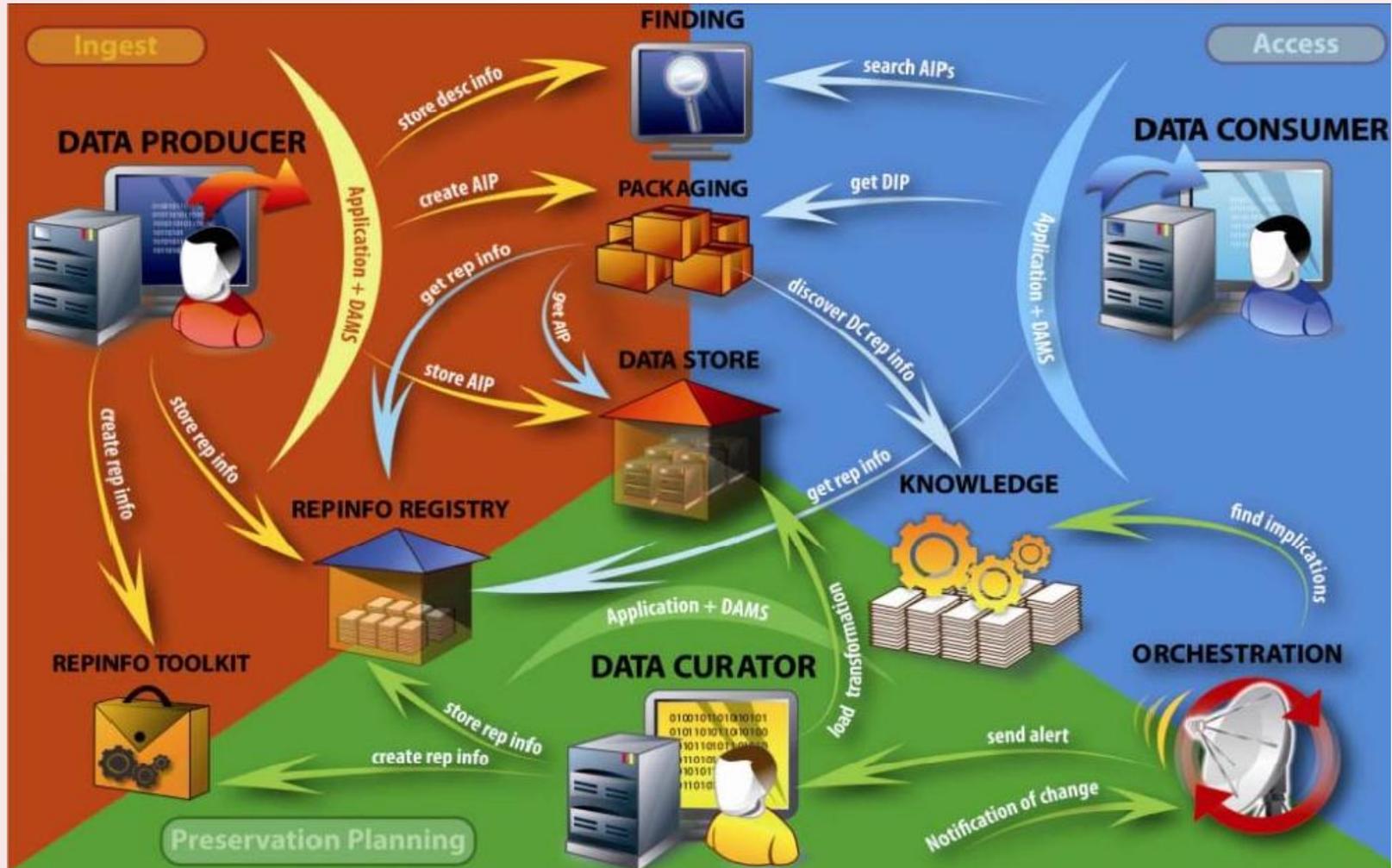
## NEDL IB

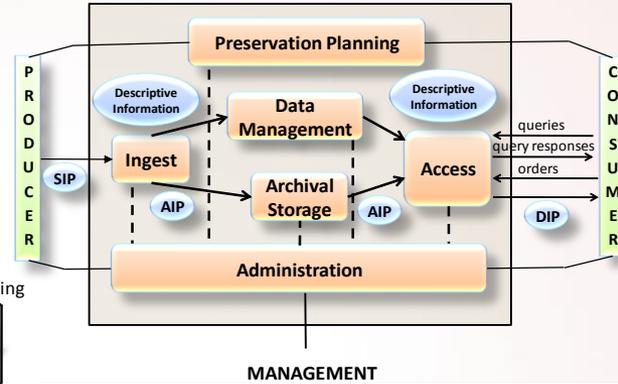
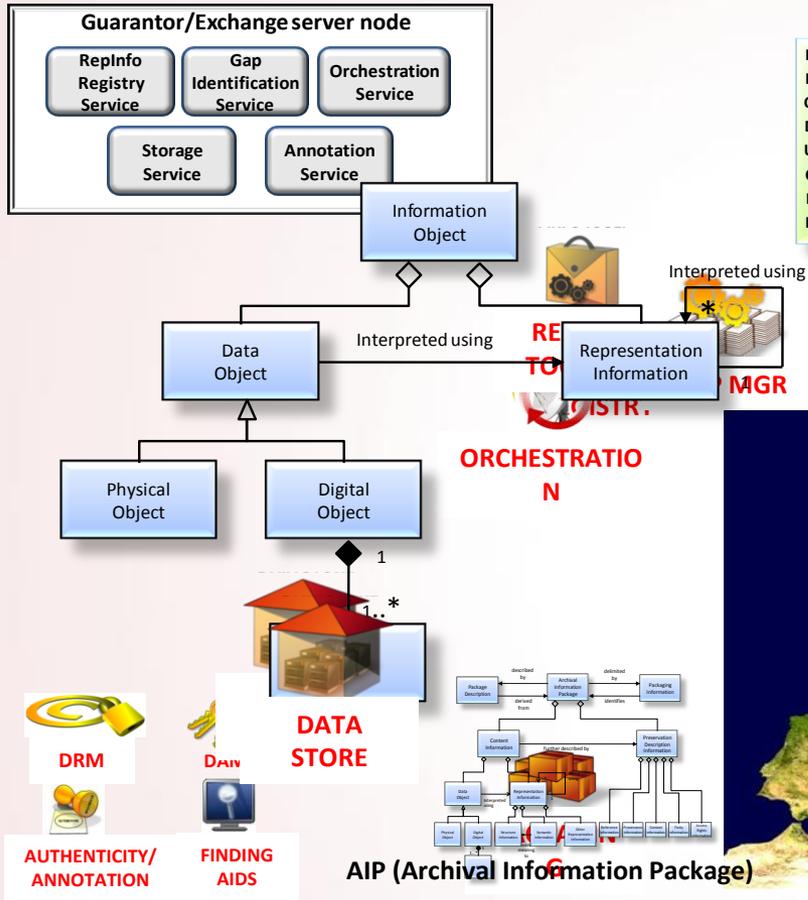
网络化欧洲存储图书馆项目。由欧洲数个国家图书馆出版社参加，主要研究数字版本存缴机制和长期保护系统机制，其目标是构建一个基于网络的欧洲保存图书馆的基础框架，确保在目前可用的电子出版物在将来同样可供利用。

## CASPAR

由欧盟资助的文化、艺术和科学数据长期保存、利用与检索项目。

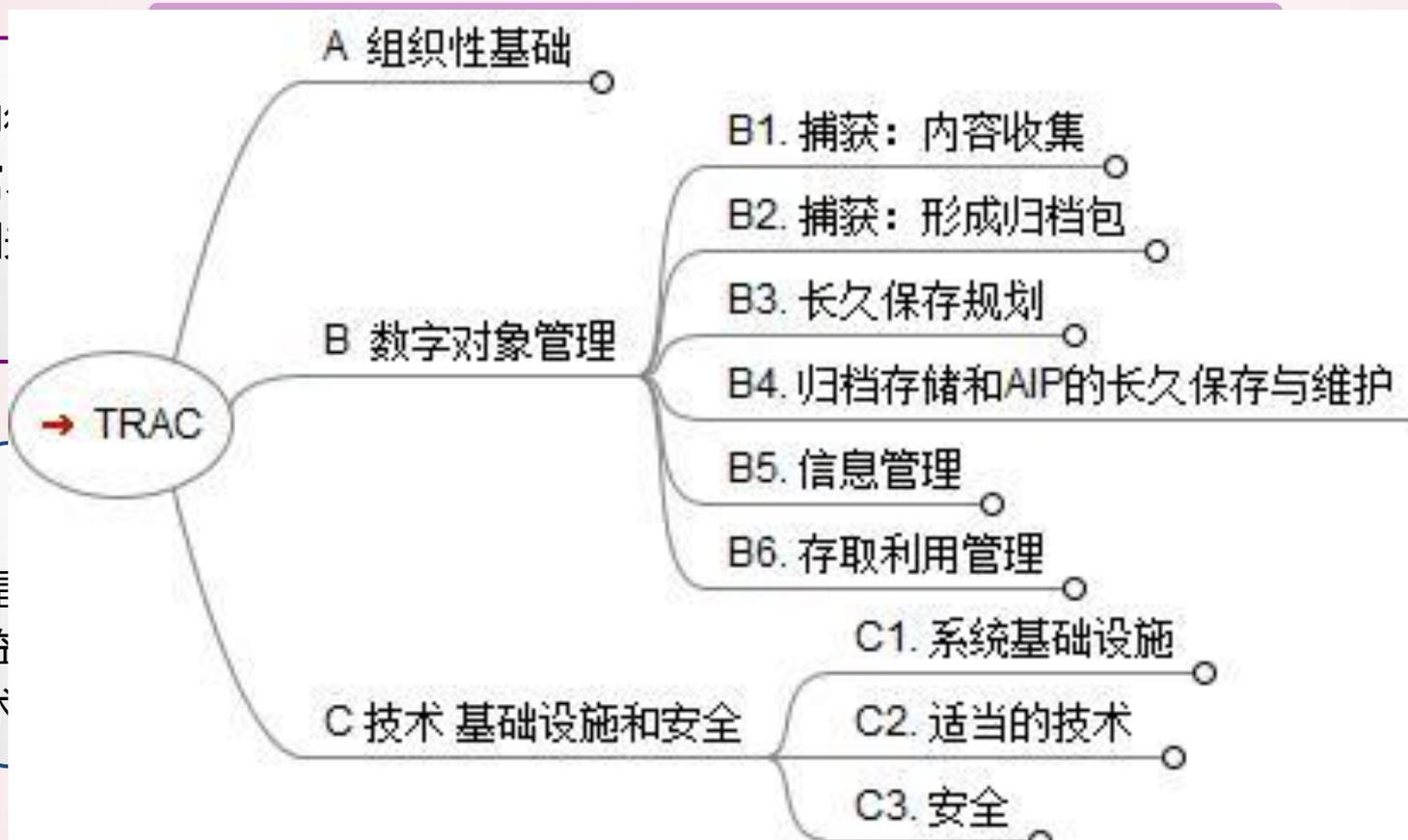
# CASPAR key component





即  
館  
相

是  
臨  
術



案  
及

目标  
其面  
从学

# 康乃尔大学注目的十个项目/文件

1. **OAIS** 开放档案信息  
系统参考模型
2. **PREMIS** 保存元数据
3. **Audit Checklist**  
认证可信任数字仓储  
的审核清单
4. **PADI** 国际数字保存  
资源主题门户网站
5. **NSF Data Report**
6. **Tufts/Yale reports**
7. **PDF/A**
8. **DCC Manual**
9. **RLG DigiNews**
10. **DPM reader**

Source: Nancy Y. McGovern, 2005

[“10 x 10: Tracing Cornell’s Digital Preservation Efforts”](#)

# 基于OAIS组织的项目体系分析

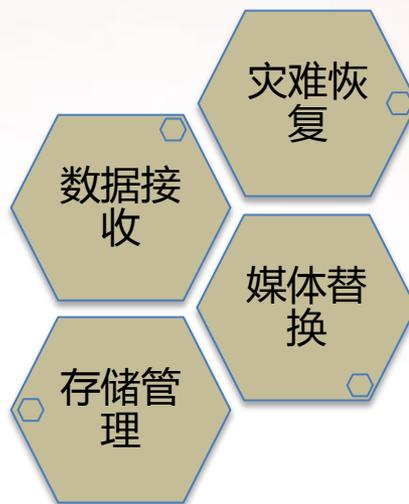
摄入职责INGEST



典型项目与工具

- 对象封装技术：METS、JATS FOXML
- 唯一持久标识技术：Handle DOI PURL ARK
- 完整性验证技术：MD5, SHA
- 元数据抽取：MET PDFBox XENA

## 档案存储功能



## 典型技术

- 载体选择
- 存储技术：RAID+
- 分层存储体系 HSM
- 网格存储管理：SRB iRODS
- 分布式存储：HDFS，Duracloud
- 灾备技术

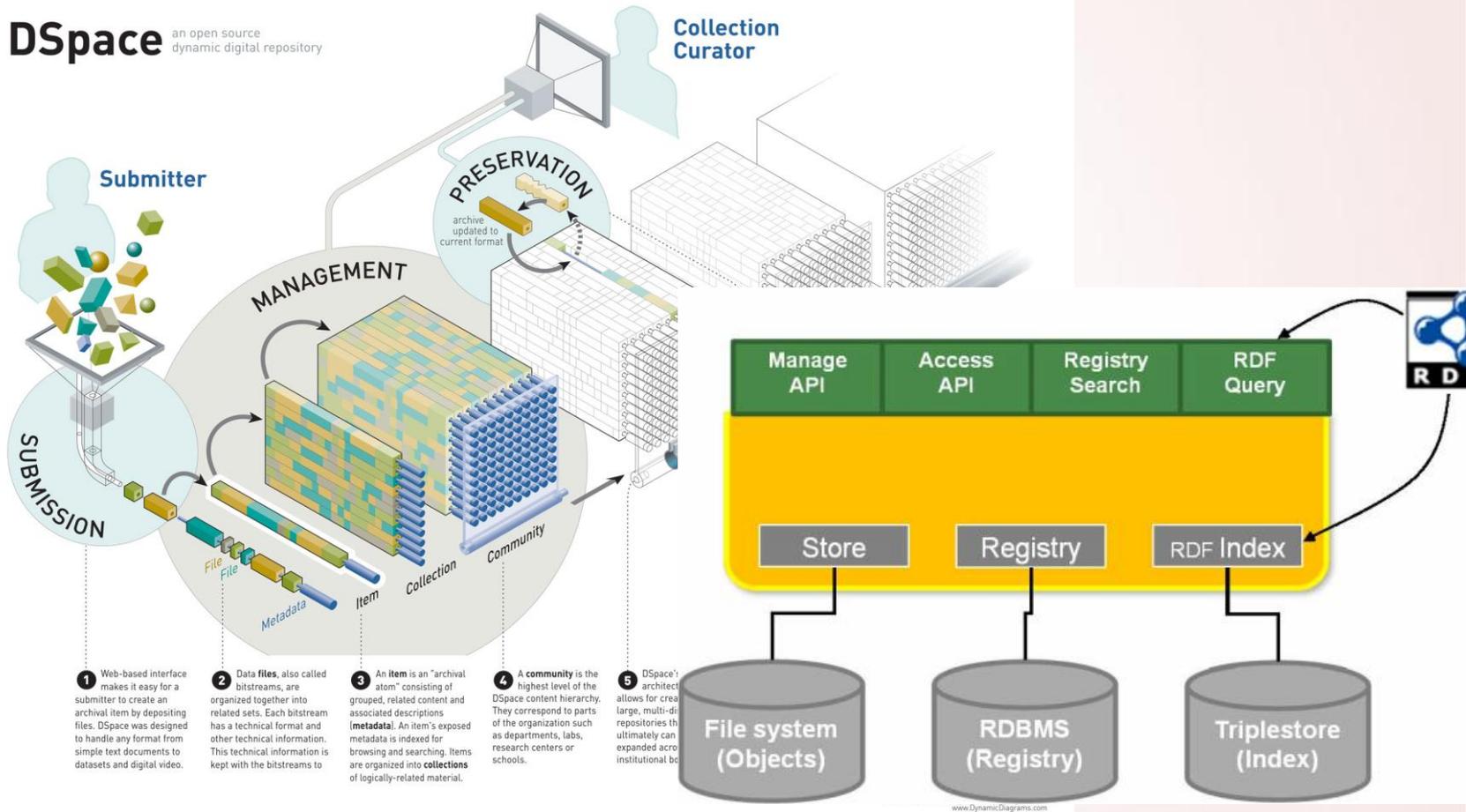
## 数据管理功能实体



## 典型技术

- 元数据存储方案：
  - 文件系统+RDBMS；
  - 文件系统+XMLDB；
  - 文件系统+RDBMS+元数据备份等
- 数据管理功能主要由数字仓储承担
  - Dspace
  - Fedora

**DSpace** an open source dynamic digital repository

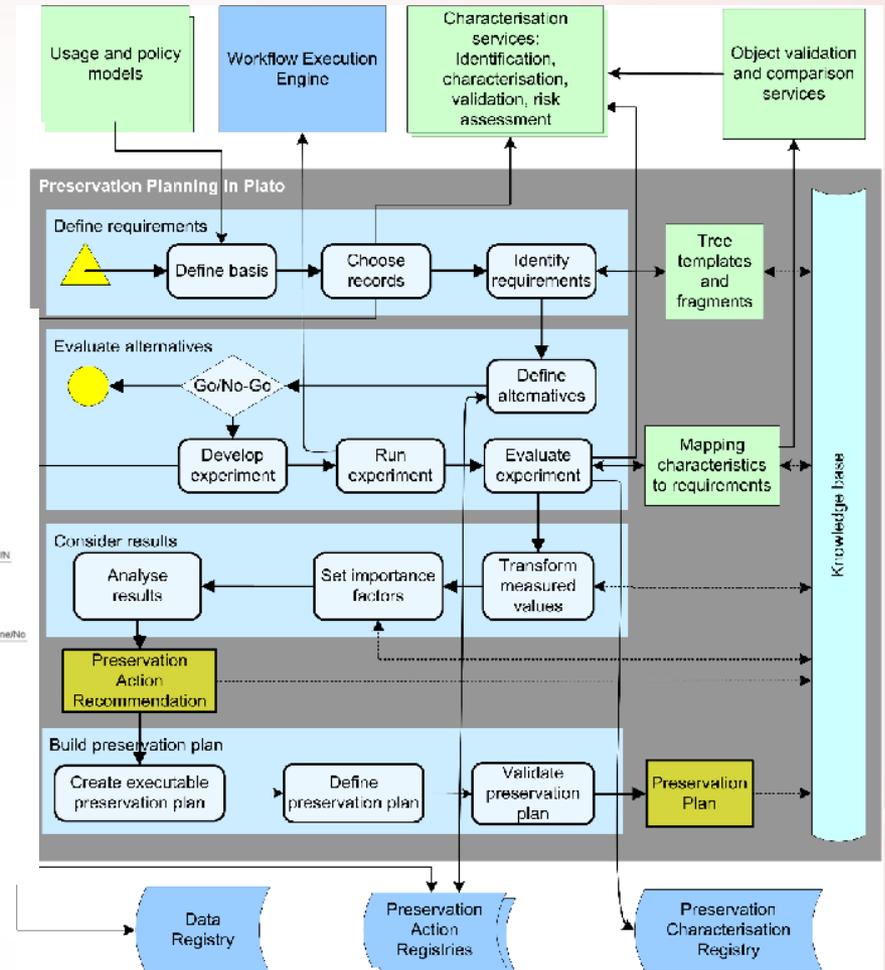


## 保存规划



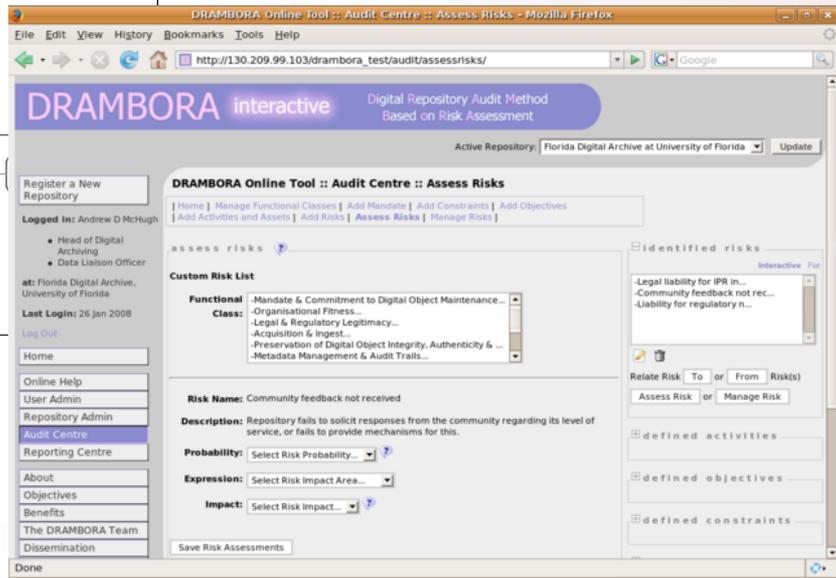
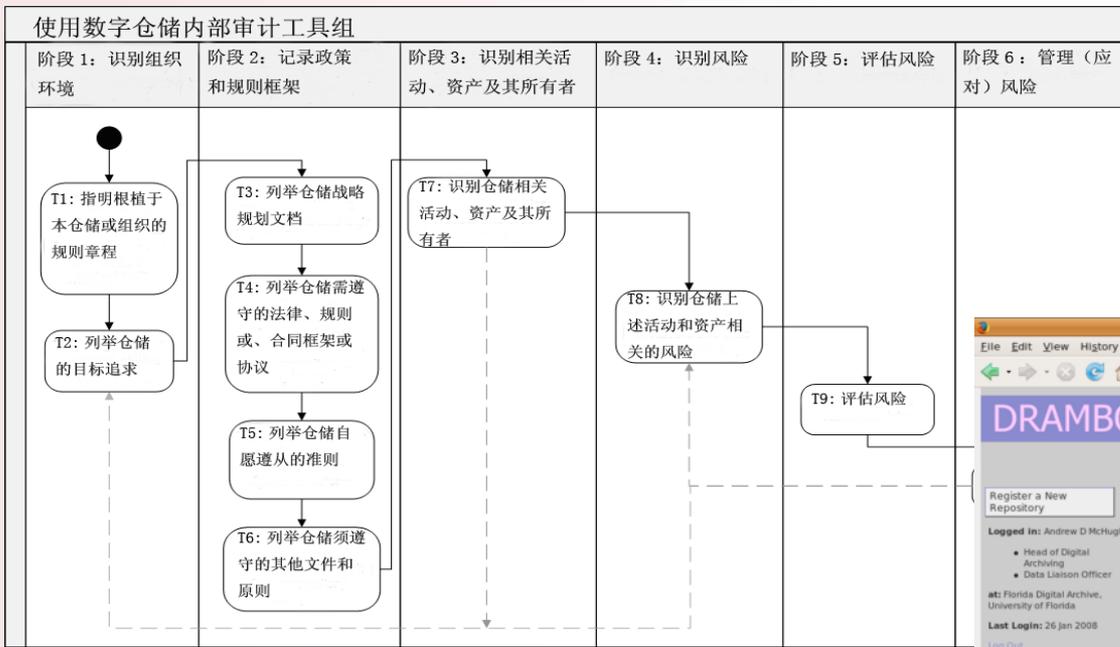
## 典型项目及工具

- PLANTS保存规划
- 数据管理规划：DCC 的DMPonline
- 风险评估工具：DRAMBORA
- 可信任规划工具：DEP 的 PLATTER

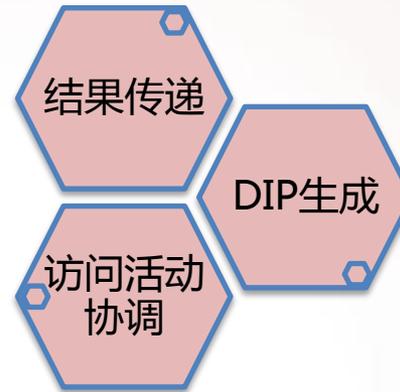


# DRAMBORA interactive

## Digital Repository Audit Method Based on Risk Assessment



## 访问功能



## 典型技术

- 访问控制技术：DAC/MAC/RBAC
- 身份认证技术
- 案例：NLA的DCM

## 行政管理功能



## 典型工具及案例

- 工具实例：C3PO
- 案例：UKDA TNA Dspace IASA

## 欧盟第六和第七框架长期保存项目按主题分类表

研究主题	资助项目
数字保存成本问题	4C
数字保存技术、框架、方法	DuraArk, ForgetIT, KEEP, PERICLES, PROTAGE, SHAMAN, SCAPE, TIMBUS, WF4Eve, DELOS
可信赖仓储	APARSEN
数据保存	CASPAR, SHAMAN, WF4Ever, APARSEN, DIACHRON, ENSURE
网络归档	LiWA, ARCOMEM, BlogForever
音视频内容的数字保存	PRESTOSPACE, DAVID, PrestoPRIME
联盟协作	APARSEN, DPE, PLANETS, DELOS

# 4

## 数字资源长期保存策略技术

---

- 宏观策略
- 管理策略
- 技术方法

### 备份与迁移

- 多重备份+异地备份
- 硬件迁移、软件迁移、载体迁移、格式迁移、版本迁移
- 典型：DATISS，Dspace，NDAD，IA，Kulturarw3

### 仿真

- IBM提出的UVC
- 案例：荷兰KB

### 开放描述与注册

- 格式登记、服务登记、元数据登记、永久标识符登记
- 典型：PRONOM、MIME、PREMIS、EPICUR

### 风干

- 是迁移与仿真的辅助策略
- CDL(加州图书馆) RFCs 仓储



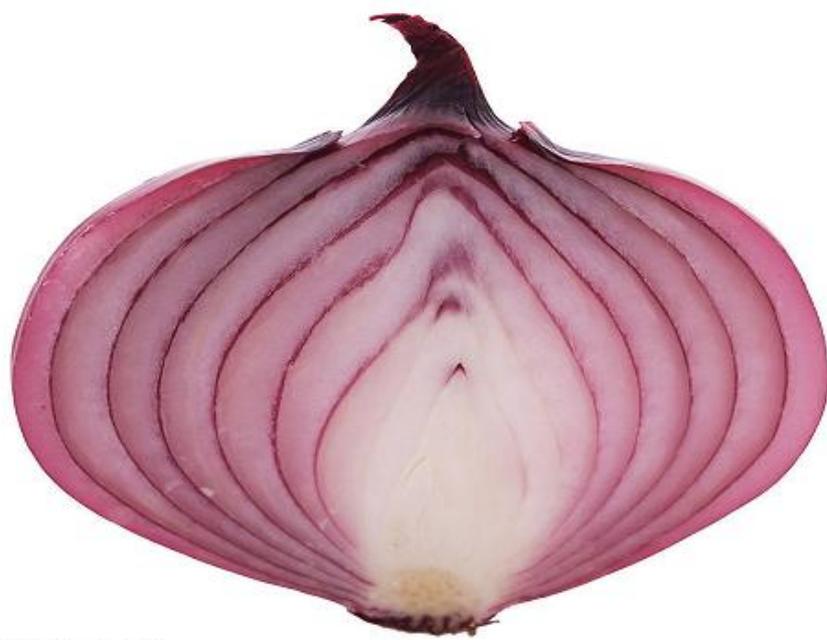
# 宏观策略

- 加强多方合作，建立合作伙伴关系
- 确立数字资源责任制
  - 数字资源的每个处理环节都要有相应的责任人承担数字资源的在本阶段应该保持的特征
- (档案) 异地异质备份

# 封装

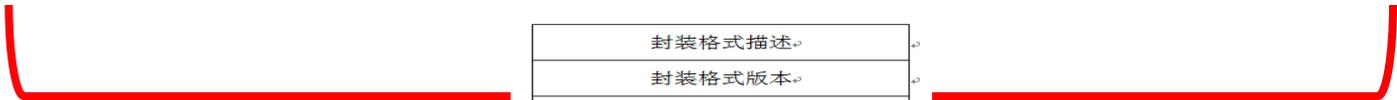
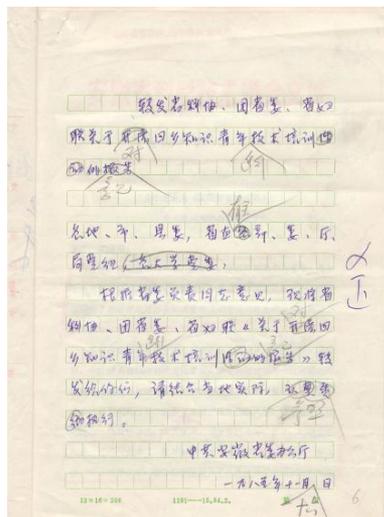
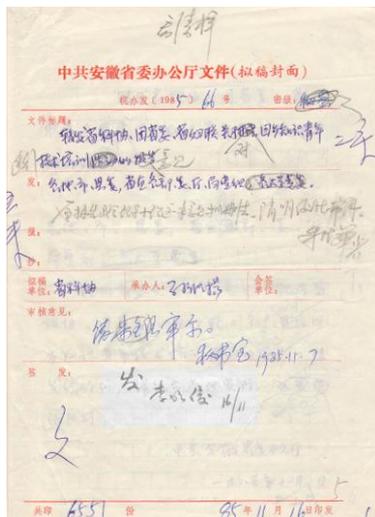
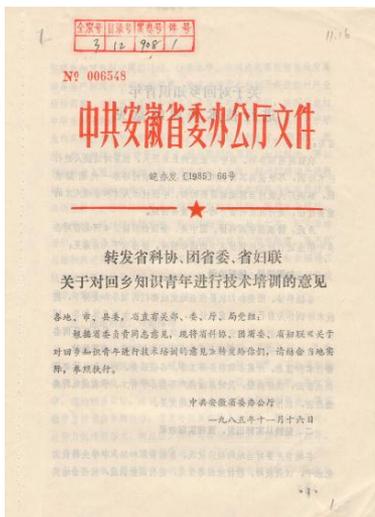
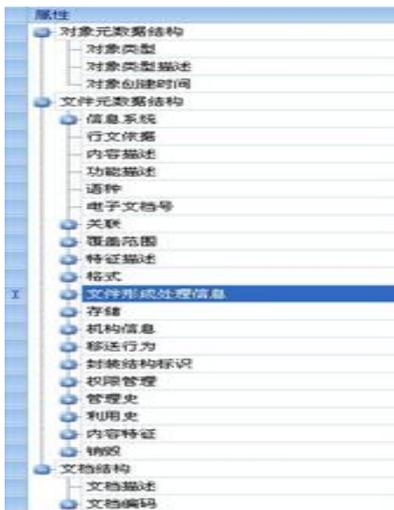
- 封装是将数字资源及其元数据按指定结构打包的过程，被普遍认为是保持数字资源内容数据和元数据可靠联系的最有效方法。
- 数字资源封装技术可以做到封装格式和计算机的软硬件无关，利于数字资源长期保存、交换和利用，实现数字资源的自包含、自描述和自证明，通常采用XML方式。
- 封装方式并没有解决数字资源本身的长期可用问题，当外部环境发生变化，封装包中的数字资源依然存在不可用的风险，需要和其他技术结合使用。

- 典型案例：澳大利亚 VERS
- [标准解读：DA/T 48](#)



www.cqpts.com/

# 元数据封装



# 案例： 台灣



圖 2-3 數位內容檔案封裝檔

資料來源：機關檔案管理資訊化作業要點

電子封裝檔	
封裝檔電子簽章	Signature 簽章時戳
封裝檔內容	
封裝檔資訊	
詮釋資料	
補簽追認?	憑證對應 Signature 簽章時戳 借卡
電子檔案	
檔案管理單位點收簽章	Signature 簽章時戳 Signature 簽章時戳
線上簽核流程	簽核流程+ 分會點* 簽核流程+
線上簽核資訊	簽核點定義 Signature 簽章時戳 Object 異動資訊 簽核人員 異動別 次位簽核人員* 簽核資訊 簽核文件夾 (細項另參附圖)
簽核點定義	Signature

# 技術措施——格式管理

- 什么是格式
  - 文件格式是一种电脑计算机用语，指计算机为了存储信息而使用的对信息的特殊编码方式，用于识别内部储存的资料。
  - 为了支持数字资源的长期保存，文件格式包括文件版本、压缩算法、比特流编码、软硬件运行环境等更特殊、粒度更细的表示信息。

## • 格式需求管理

### – DA/T47 版式数字资源长期保存格式需求规范 解读

1. 支持真实性

2. 格式透明

3. 不绑定软硬件

4. 格式自描述

5. 格式自包含

6. 固定显示

7. 持续可解释

8. 可转换

9. 易存储

10. 持续可用

# 分类型管理



可接受转让格式

内容类型

推荐存储格式

Office 套件: MicrosoftOffice

EPUB (电子书的基础标准)

可扩展超文本标记语言(XHTML)

MPEG-1 3 层, MPEG-2 3 层 (MP3)

联合图像专家组 (JPEG)

数字视频特征:

可接受转让格式

内容类型

推荐存储格式<sup>1</sup>

dBase 格式 (DBF)

结构化数据-  
数据库

软件独立归档的关系数据库 (SIARD) \*

文本: 数据描述分隔平面文件

SAS

结构化数据-  
统计和定性分  
析数据

文件倡议 (DDI) 版本 3.0 \*

统计数据和元数据交换 (SDMX) \*

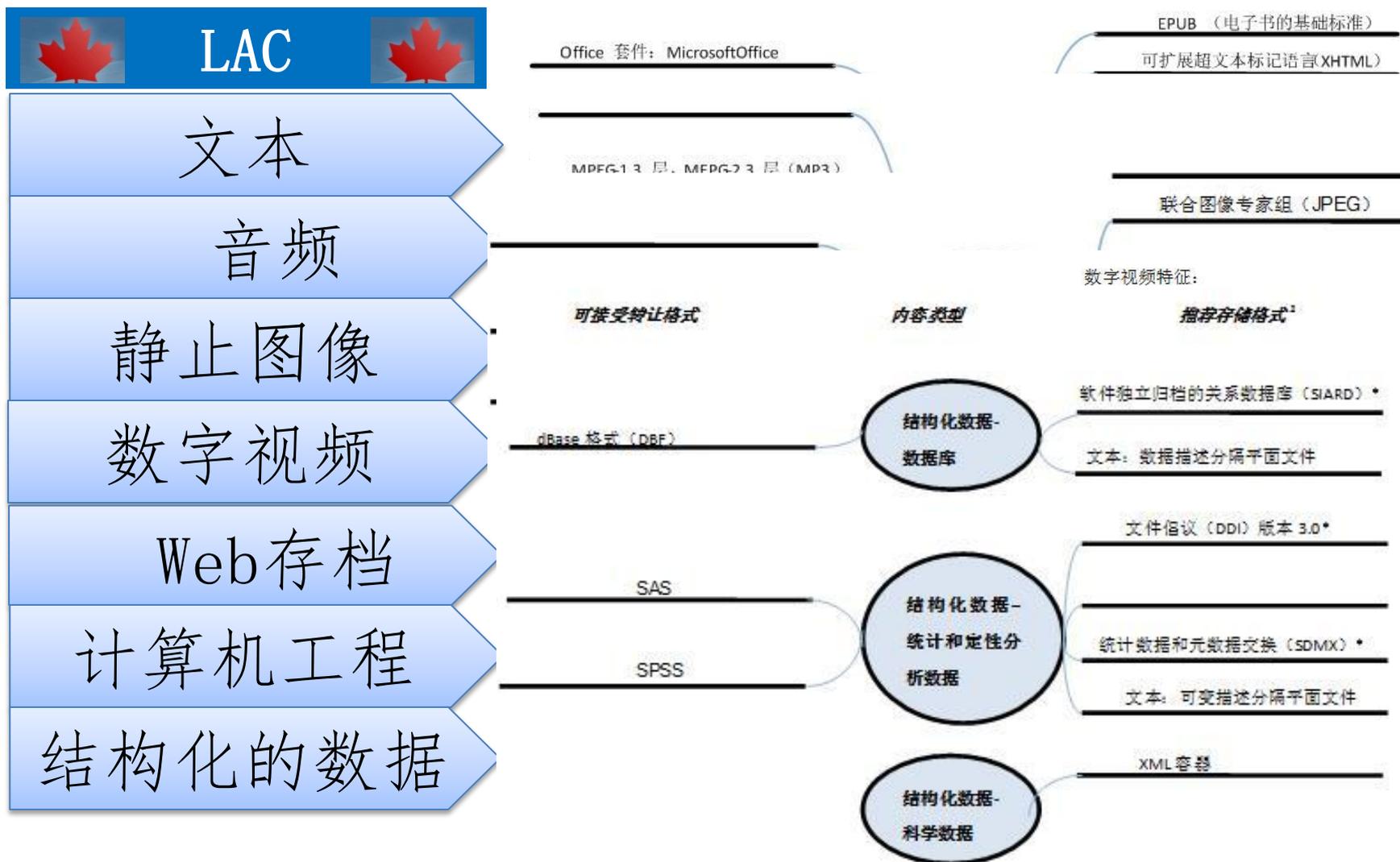
文本: 可变描述分隔平面文件

SPSS

结构化数据-  
科学数据

XML 容器

# 分类型管理



# 数字文件格式登记系统

## PRONOM

英国国家档案馆（TNA）建立  
PRONOM技术登记系统  
DROID格式识别工具  
PUID永久唯一标识符

## MIME MTR

使用最广泛的格式登记系统，登记的格式属性：MIME类型名称、MIME子类型名称、必备参数、可选参数、编码要求、安全要求、互操作要求、发布说明、所支持的应用、联系人等。

## GDFR

将文件格式定义为一种固定的以位序列编码的信息模型，登记的格式属性包括：标识符、别名、创建者、所有者、维护者、分类、关系、定义说明、签名、工具、状态、来源和注释。

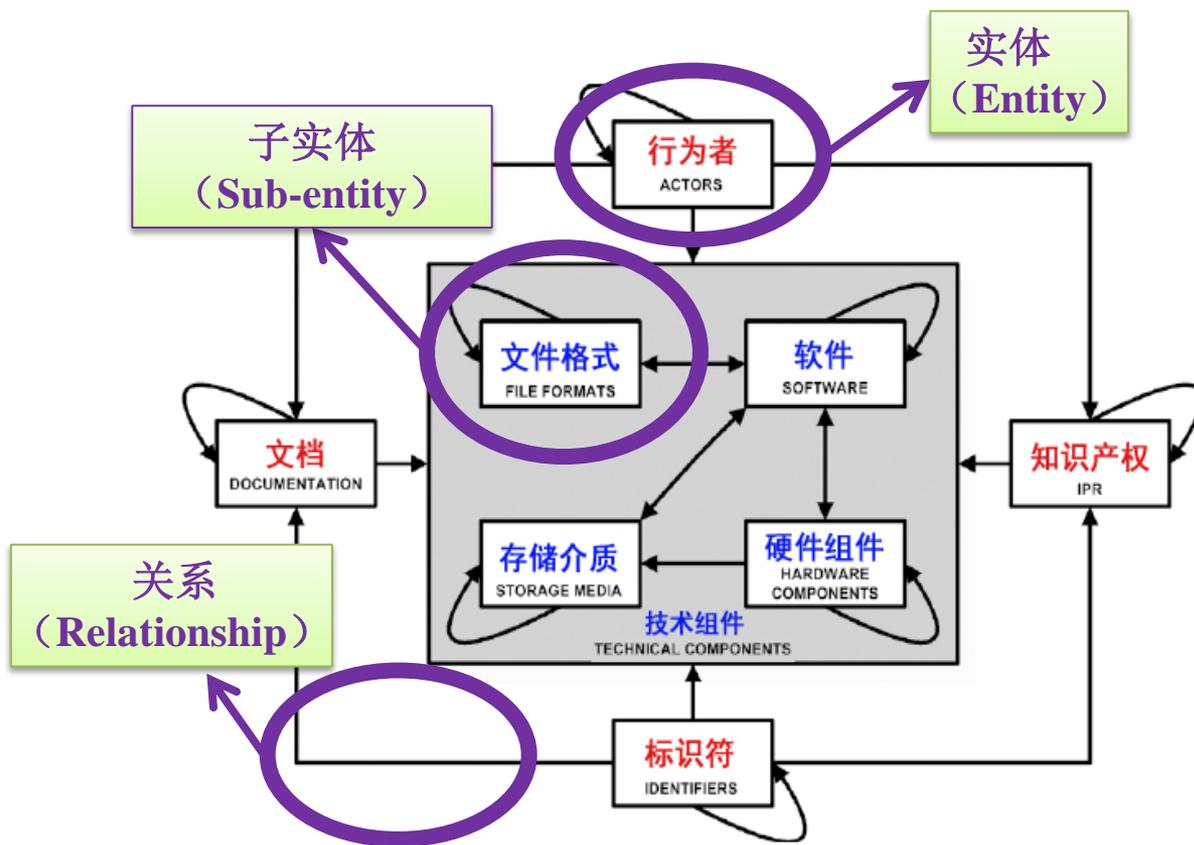
## UDFR

统一的数字格式注册系统  
融合GDFR和PRONOM两者优点



PRONOM项目的核心成果

# 格式结构



# 开发预研格式管理工具

在面对大量格式类型进行长期保存时需要准备和研制一些必要的管理工具，以有效提升格式管理的整体水平，包括格式识别、验证、转换等类别。



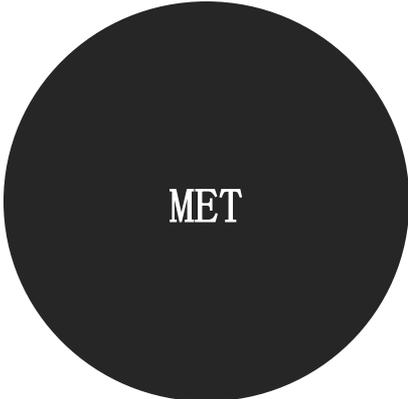
DROID

自动识别格式  
记录文件格式类型、版本  
以xml或csv格式输出  
评估文件使用格式的过时风险  
推荐采取迁移行动



JHOVE2

自动识别、校验文件格式  
标识信息、验证条件和验证标准  
格式标识、验证和特征描述操作



MET

自动化元数据抽取工具  
以XML的标准格式输出  
以PREMIS为基础

# 我国的归档格式管理——18894修订标准

文本、  
位图

RTF、WPS、DOC、JPEG、JPEG-2000、TIFF、PNG、PDF、

计算  
机辅  
助

SVG、SWF、WMF、EMF、EPS、DXF、  
JPEG、JPEG-2000、TIFF、PNG

数据  
库

ET、XLS、DBF、XML、JSON

珍贵且  
需永久  
保存的  
数字资源

照片  
类

MXF JPEG、TIFF

录音  
类

WAV、MP3、MP4

录像  
类

MPG、FLV、AVI

公务  
电子  
邮件

EML

网页

HTML

专用软  
件产生  
的数字  
资源

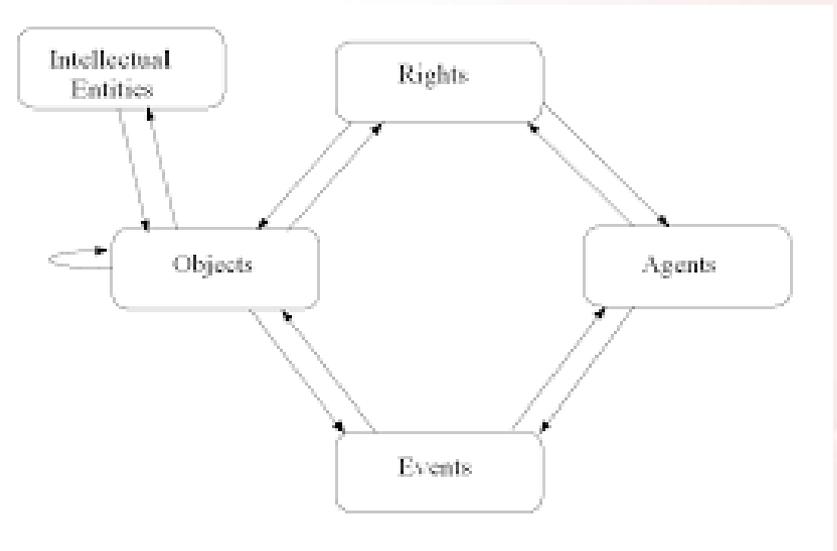
转换成通用格式归档

元数据选择ET、XLS、DBF、XML、  
JSON

# 保存元数据

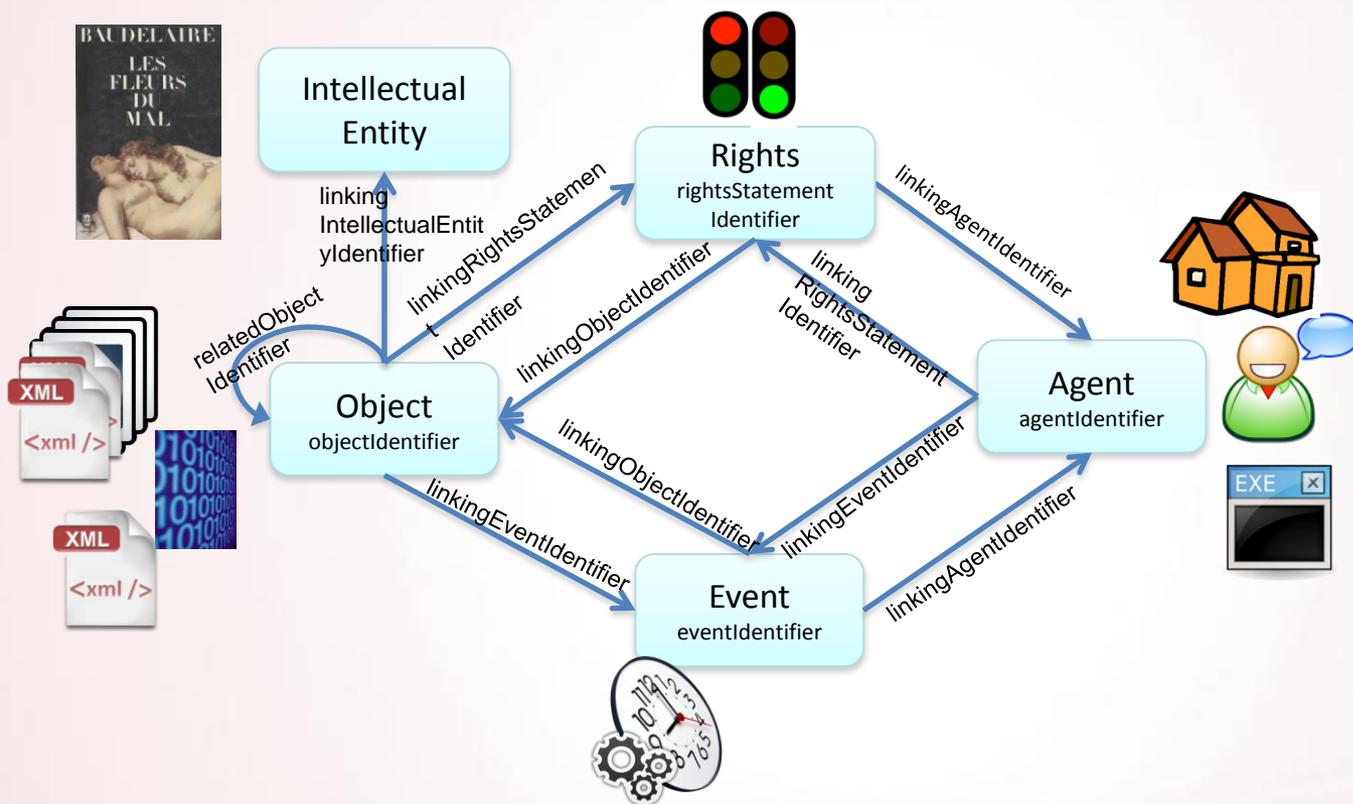
基于OAIS信息模型开发了保存元数据框架

- PREMIS
- 英国CEDARS
- NEDLIB
- NLA保存元数据
- 应用：PREMIS+MET



# PREMIS PRESERVATION METADATA IMPLEMENTATION STRATEGIES

## PREMIS IDENTIFIERS IN ACTION



# 仿真

- 仿真是用一个计算机系统模拟另一个计算机系统，使前者的功能完全与后者相同，即前者接收与后者相同的数据，执行与后者相同的程序。它可以使一个计算机系统执行为另一个计算机系统编写的程序，而不必重新编写程序。
- 仿真技术方法主要包括
  - 硬件模仿
  - 软件模仿
  - 模仿另一种操作系统
  - 通过一层中介或虚拟机器来模仿

## 技術措施——再生

- 再生性保護技術是將過時的某些電子信息適時地轉換到縮微品或紙介質上的一種保護措施。如轉移到縮微品上。
- 再生性保護技術存在的問題是：
  - 部分有聲信息、多媒體信息無法轉移到膠片或紙張上去；
  - 信息如果轉移到膠片或紙張上之後，由於這兩種載體表現形式的局限性，會使電子信息失去原有的風格和魅力。

## 技術措施——遷移



- 遷移是持續地將電子信息從一種技術環境轉換到另一種技術環境的過程,以進行電子信息的長久保存。主要包括存儲介質的遷移和運行環境的遷移,遷移的同時根據需要完成格式轉換。
- 硬件遷移
  - 將數據轉到新的容器中
- 軟件遷移
  - 轉移到同樣格式的新版本
  - 轉移到新格式
- 在使用時才做遷移



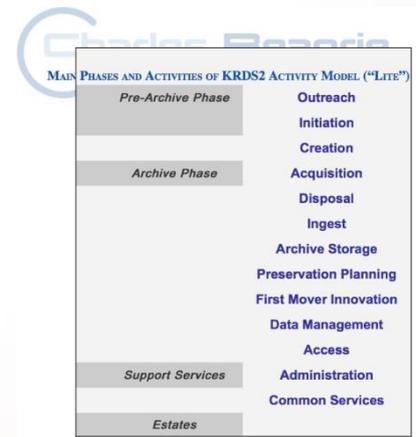
# 成本分析

## 📍 数字文献长期保存成本模型研究

- LIFE
- KRDS

## 🔧 数字成本分析工具

- NASA-CET



## THE LIFE MODEL v1.1



# 5

## 数字资源长期保存标准规范

---

# 国际长期保存标准

- 国际同类研究日趋成熟
  - ISO 14721
  - ISO 16363
  - ISO 13008
  - ISO 26102
  - ISO 17068
  - ISO 18492
  - ISO 19005
  - ...

## ISO16363

《可信数字仓储审核与认证：标准与审核表》确定了可信数字仓储的基础定义，明确了必须考虑和评价的定义组成成分。它还讨论了本文档的利用展望以及该标准潜在的基本原则，并罗列了可信赖仓储所需要满足的标准，并对其进行了解释和举例。

## ISO 13008: 2012

《数字文件的转换和迁移程序》本标准规定了数字文件（包括数字文件本身及其元数据）在转换和迁移中的规划事项、要求和程序，以实现数字文件的真实性、可靠性、完整性和可用性，从而确保其在业务活动中的证据价值。

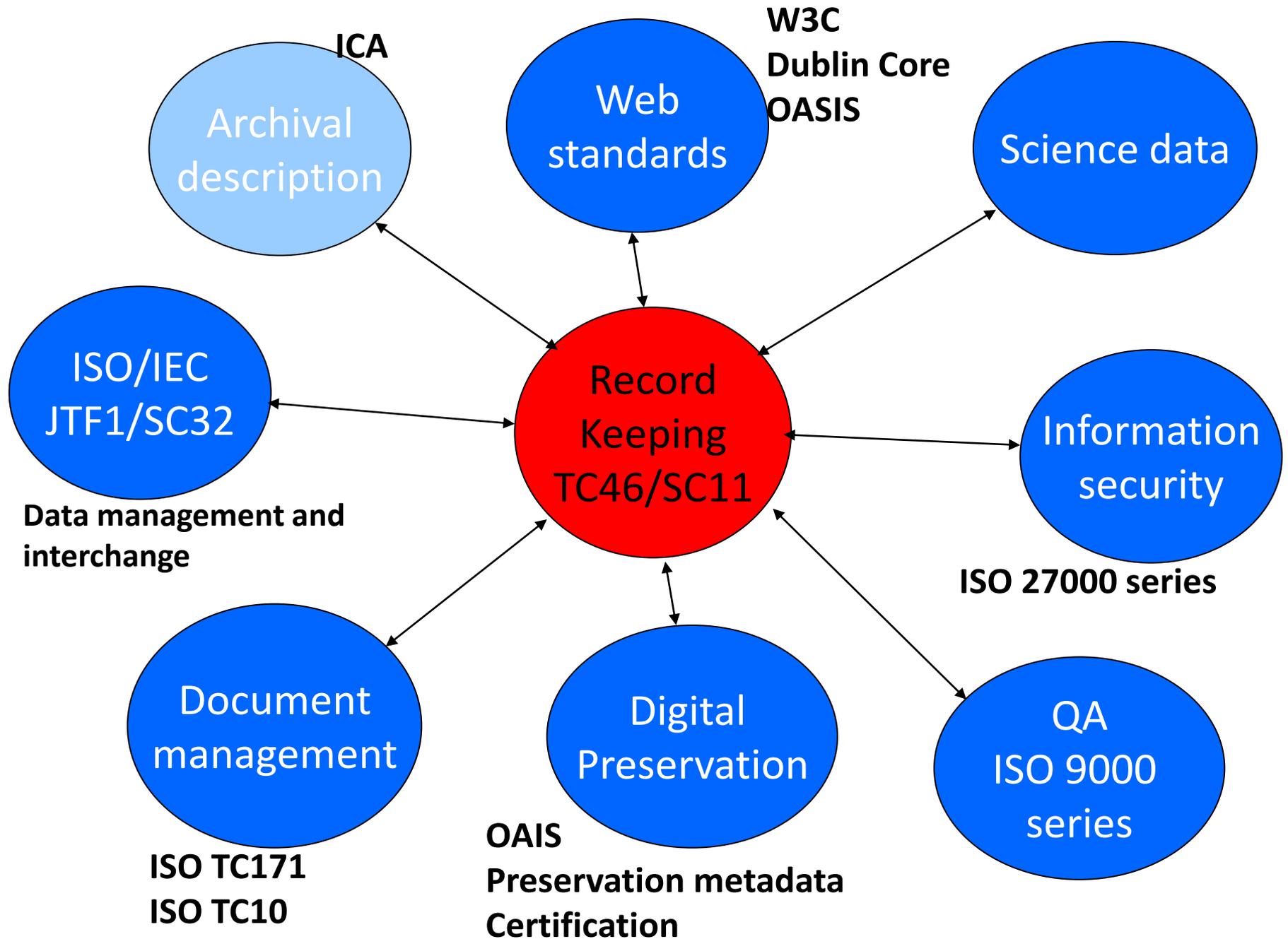
## ISO 17068: 2012

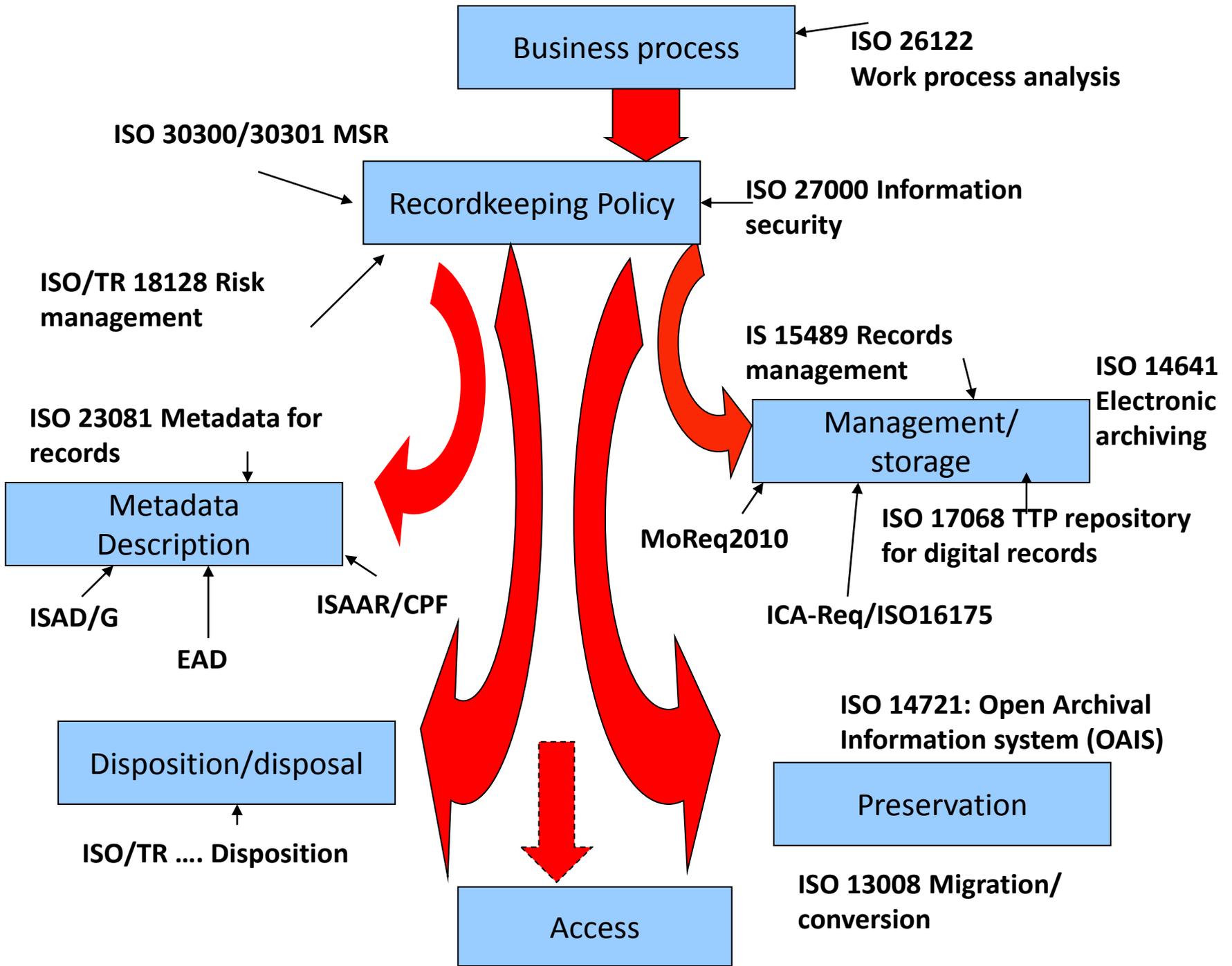
《数字文件可信第三方仓储标准》未正式发布，尚处在投票阶段。该技术报告主要介绍了可信第三方仓储（TTPR）提供的服务和流程，这些服务和流程可以确保TTPR能够证明他们受委托管理的数字文件的可靠性和真实性。该报告还介绍了真实性的基本原则、TTPR服务模型和服务水平协议的合同项，以实现上述目的。

## ISO 26102

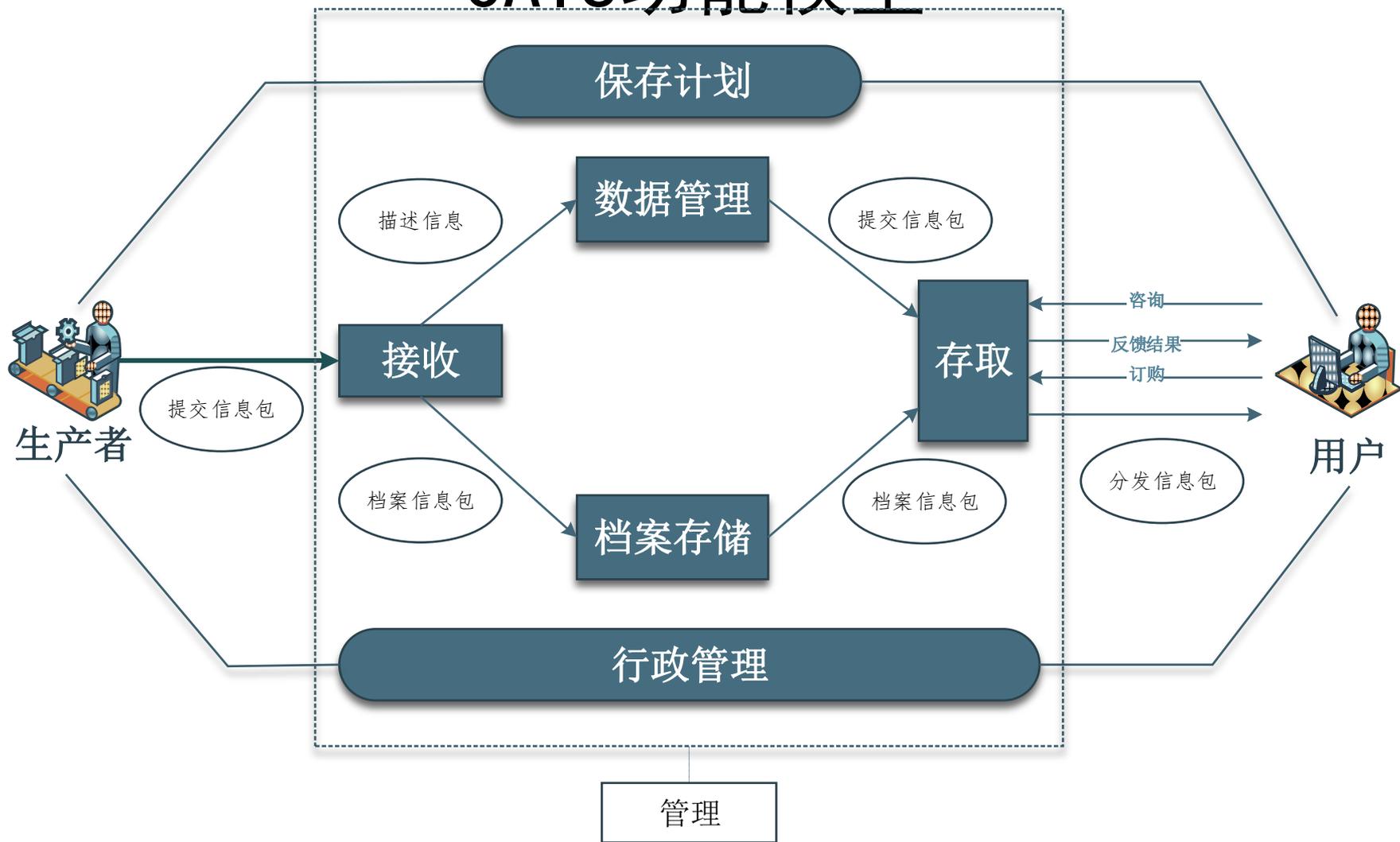
《数字资源长期保存需求》尚未正式发布，该技术报告对ISO15489在数字文件的长期保存方面做出了进一步的指导。它对技术环境限制下的管理问题做出了明确的阐述。

## Relationships to other domains





# OAIS功能模型



连接各实体的线条是确定信息双向交流的通讯路径

# 功能模型

## 接收

该实体从生产者接收提交信息包（SIPs），并为档案馆的内容存储和内容管理作好准备。

## 档案存储

该实体为存储、维护和检索档案信息包提供服务。

## 数据管理

该实体为两类描述信息提供保管、维护和存取服务。这两类描述信息分别是：识别和证明档案资源的描述信息；识别和证明行政管理数据的描述信息。

## 行政管理

该实体向档案系统的所有工作提供服务。

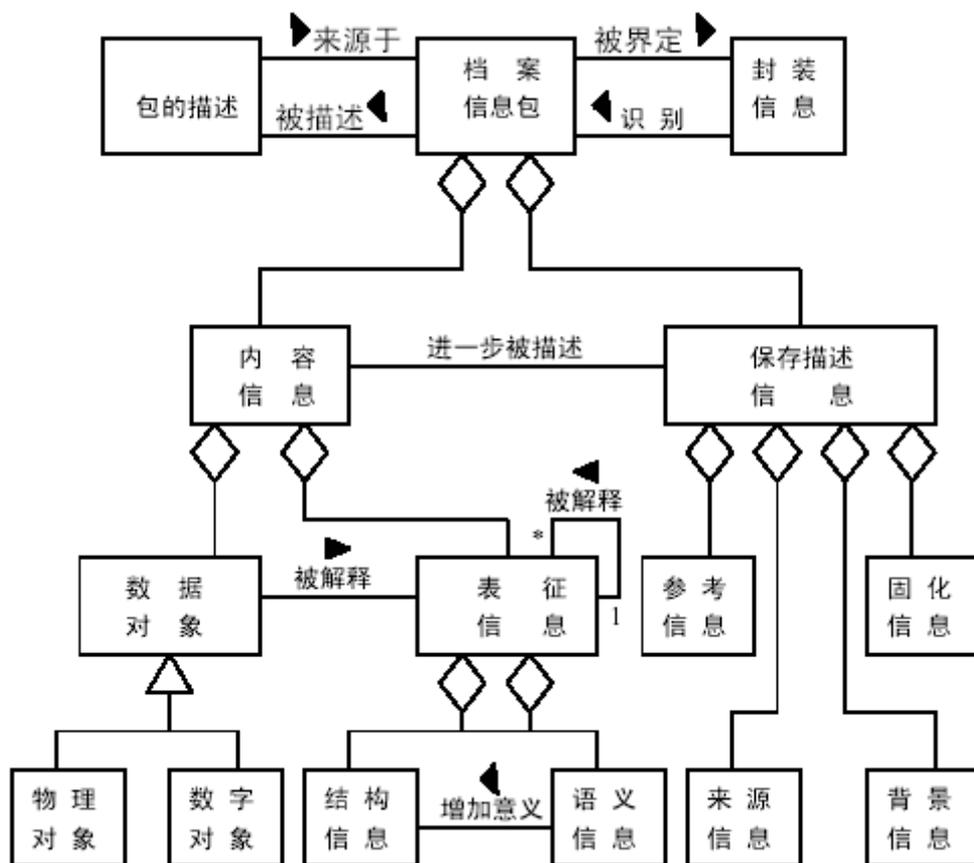
## 保存计划

该实体为监控OAIS环境和提供推荐书而服务。提供推荐书的目的是为了确保持存的信息能够长期被指定用户群存取——即使原来的计算机环境已经过时。

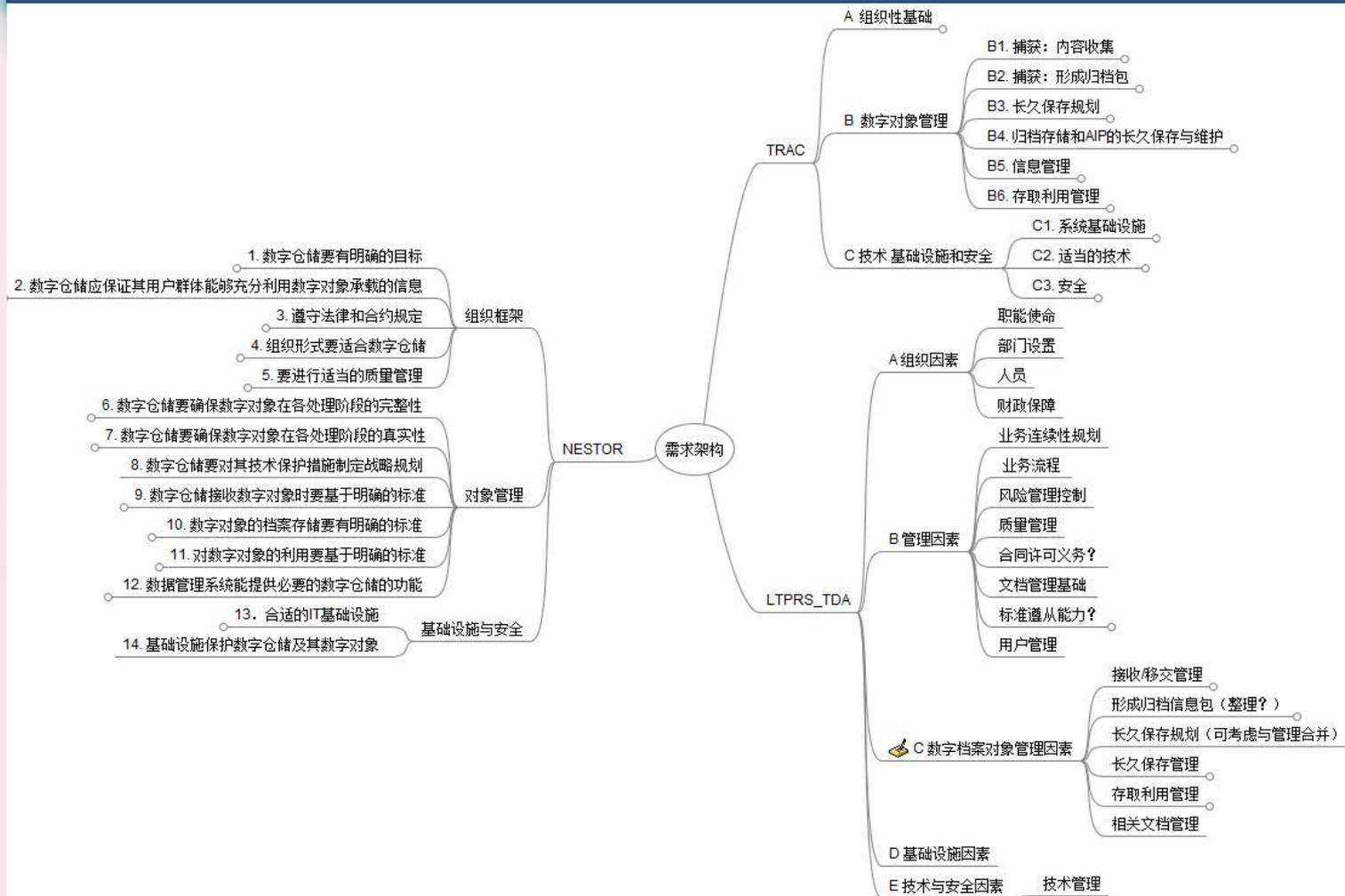
## 存取

该实体为支持用户而服务，它支持用户对OAIS所存信息的存在、描述、位置和可用性做出测定，并且允许用户请求和接收信息产品。

# OAIS 信息模型



# 规范结构



组织框架

对象管理

基础设施与安全

- 6.1 接收：数字仓储要保证数字对象的完整性
- 6.2 档案存储：数字仓储要保证数字对象的完整性
- 6.3 获取：数字仓储要保证数字对象的完整性
- 7.1 接收：数字仓储要保证数字对象的真实性
- 7.2 档案存储：数字仓储要保证数字对象的真实性
- 7.3 存取：数字仓储要保证数字对象的真实性

6. 数字仓储要确保数字对象在各处理阶段的完整性

7. 数字仓储要确保数字对象在各处理阶段的真实性

8. 数字仓储要对其技术保护措施制定战略规划

- 9.1 数字仓储要指定其传输对象（提交信息包，SIPs）
- 9.2 数字仓储要识别数字对象的哪些特征对信息保存具有重要意义
- 9.3 数字仓储要对数字对象采取技术控制以实施长期保存的措施

9. 数字仓储接收数字对象时要基于明确的标准

- 10.1 数字仓储要定义其归档对象（归档信息包，AIPs）
- 10.2 数字仓储要负责将传输对象(SIPs)转变为归档对象(AIPs)
- 10.3 数字仓储要保证归档信息包的存储和可读性
- 10.4 数字仓储要实施相应的策略来保证对归档信息包的长期保存

10. 数字对象的档案存储要有明确的标准

- 11.1 数字仓储要定义其利用对象（分发信息包，DIPs）
- 11.2 数字仓储要保证AIPs向DIPs的转换

11. 对数字对象的利用要基于明确的标准

- 12.1 数字仓储对数字对象及它们的关系进行唯一和永久的标识
- 12.2 数字仓储需要足够的元数据来对数字对象进行形式和内容上的描述与标识
- 12.3 数字仓储需要足够的元数据来对数字对象进行结构上的描述
- 12.4 数字仓储需要足够的元数据来记录由DR造成的数字对象的改变
- 12.5 数字仓储需要足够的元数据来对数字对象进行技术描述
- 12.6 数字仓储需要足够的元数据来记录相应的使用权利和条件
- 12.7 在任何时候都要保证数字对象的元数据的记录

12. 数据管理系统能提供必要的数字仓储的功能

- 13.1 IT基础设施要实现对象管理的要求
- 13.2 IT基础设施要实现IT安全系统的安全要求

13. 合适的IT基础设施

14. 基础设施保护数字仓储及其数字对象

# 基于TRAC的数字档案长保规范设计简介

- 组织因素
- 管理因素
- 数字档案管理因素
- 技术与安全因素

## 规范内容介绍——组织因素

### — 职能声明与业务连续性保障

- 档案馆应在有关的法律、法规、政策中声明其对数字档案信息长期保存、管理和存取利用的职责。

职能声明  
与业务连  
续性保障

- 档案馆应保障数字档案信息的可持续管理，有效降低法律环境、社会环境、技术环境、行政管辖权、经费资助规模的变动对数字档案信息长期保存的影响。

## — 组织机构与人员设置

a) 档案馆应根据需要设置内部组织机构，明确目标、分工和流程，配备专职人员，合理进行资源配置。

b) 档案馆应根据需要规划人力资源，并根据组织结构进行人员角色设定、岗位职责和发展前景的说明。

### 组织机构与 人员设置

c) 档案馆应提供满足需要的工作人员，具备档案管理、系统管理、利用审核等技能。

d) 档案馆应向人员提供继续教育机会，以满足应对数字档案长期保存的技术发展、法律环境变化的需要。

## — 财政保障

a

档案馆具备长期保存数字档案信息的财政基础。由国家拨款的档案馆应在说明资金来源和规模。

b

档案馆应对财务活动进行持续管理，具备完整、合规的财务计划，财务活动符合标准，各类财务收支可追溯。

c

档案馆应对形成的财政收入进行说明和公示。

## 业务连续性计划

- 提供数字档案信息长期保存的长远规划，定期修订和完善。
- 预测风险和突发因素，并制定应急预案。
- 制作数字档案信息备份，确定备份管理机制。
- 档案馆之间宜开展协议互助，共同支撑业务连续性。

## 业务流程管理

- 档案馆应明确保管范围内的数字对象，明确必要的业务流程和环节，并遵从相应的业务规范。根据保管对象对业务流程进行区分，识别业务节点，进行业务描述，实行业务流程标准化管理等。

## 风险管理控制

- 档案馆应针对数字档案信息进行风险管理
- 档案馆要不间断进行风险监控，要对技术、管理带来的风险进行区分，识别潜在的威胁。
- 档案馆应根据不同的风险采取适当的应对措施，宜综合采取管理和技术的措施来应对，并根据环境和要求的变化进行策略调整。

## 合同许可义务

- 应通过协议明确有关术语、角色、责任等。协议及实施方案需第三方审计。
- 对数字档案信息对象的收集、移交和利用应遵从相关法律法规要求。
  - a) 制定相关的协议或方案，保护合法利益。
  - b) 档案馆与数字档案生成者签订移交对象相关协议，对知识产权和所有权转移作出说明。若受协议限制不便提供利用的，应作出解释。
  - c) 数字档案信息符合档案保管的要求，涉密信息应满足保密要求。
  - d) 档案馆应确保与长期保存对象相关的操作和处置行为应符合协议规定，遵从法律法规要求。
- 档案馆应以正式或非正式协议的方式，与数字档案信息提供者共同确认移交范围、利用条件、档案馆的保存权利和义务，以及涉及知识产权、隐私保护方面的约定。

## 规范内容介绍——管理因素

质量管理

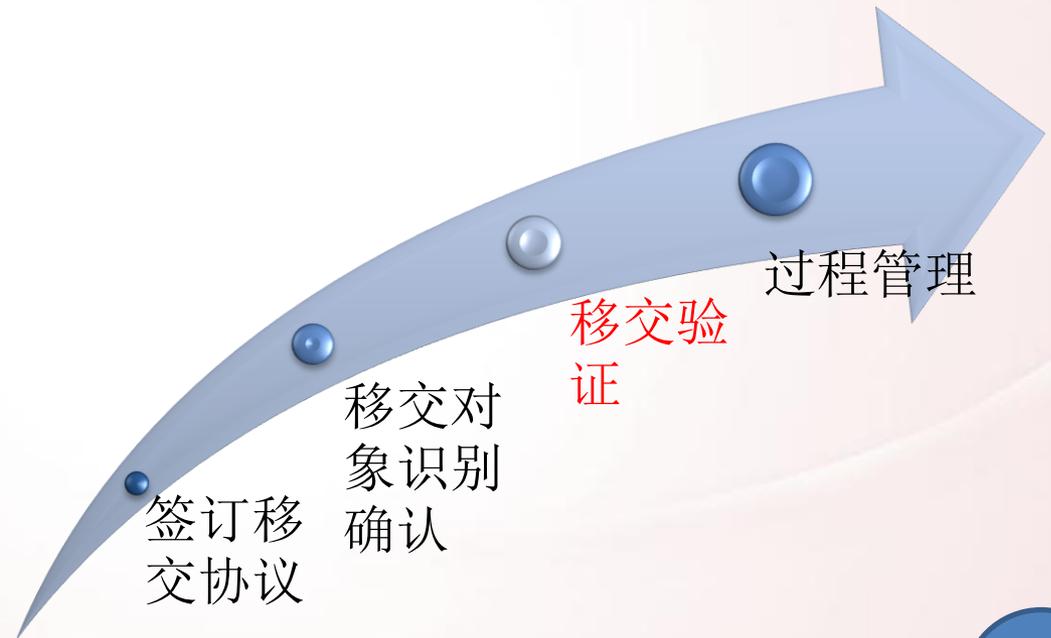
文档管理

标准遵从能力

用户/利用管理

# 数字档案管理

- 接收管理
- 形成AIP
- 长期存储管理
- 长期保存规划
- 存取利用管理
- 综合管理

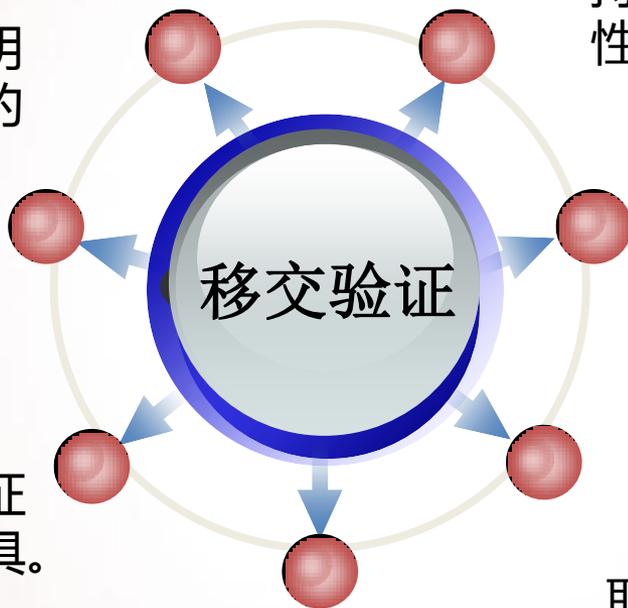


## — 移交验证

【格式验证】档案馆明确长期保存数字对象的推荐格式，遵从DA/T 47。列入移交协议

档案馆宜提供格式验证、真实性验证、完整性验证的程序标准和必要的工具。

档案馆应具备格式验证机制，检查格式是否符合要求。



档案馆应具备完整性验证程序，以检查移交对象的完整性

【真实性验证】档案馆应支持验证数字档案信息的真实性。

【准确性验证】档案馆就移交数字档案对象所展示信息与移交单位所描述的信息进行对比。验证对象的准确性。

【完整性验证】档案馆应采取措​​施保证移交数字对象的完整性。

# 规范内容介绍——数字档案对象因素（其他）

## 形成AIP

规范准备

AIP制作

## 存取利用管理

存取利用政策

存取利用实施

## 长期存储管理

常规检测

灾难恢复

## 长期保存规划

制定长期保存策略

技术监控机制

信息包设计与迁移管理

## 综合管理

标准政策

元数据管理

审计管理

# 6

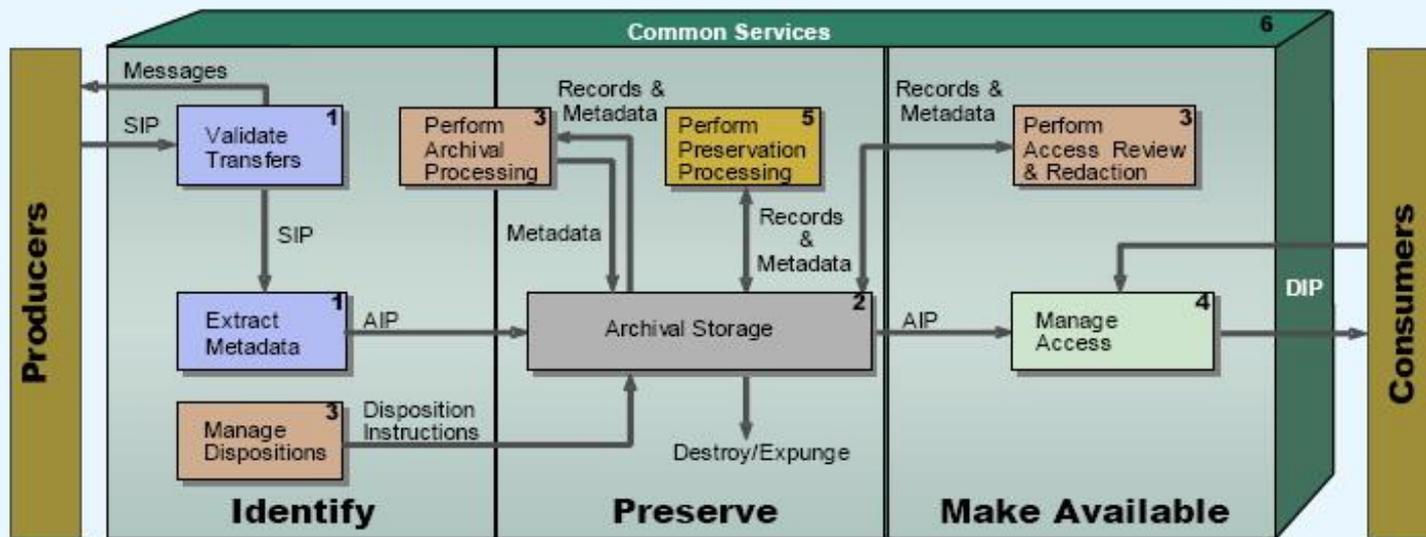
## 数字资源长期保存系统设计

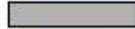
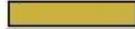
---

# 保存系統



# 案例：ERA-基于OAIS架构

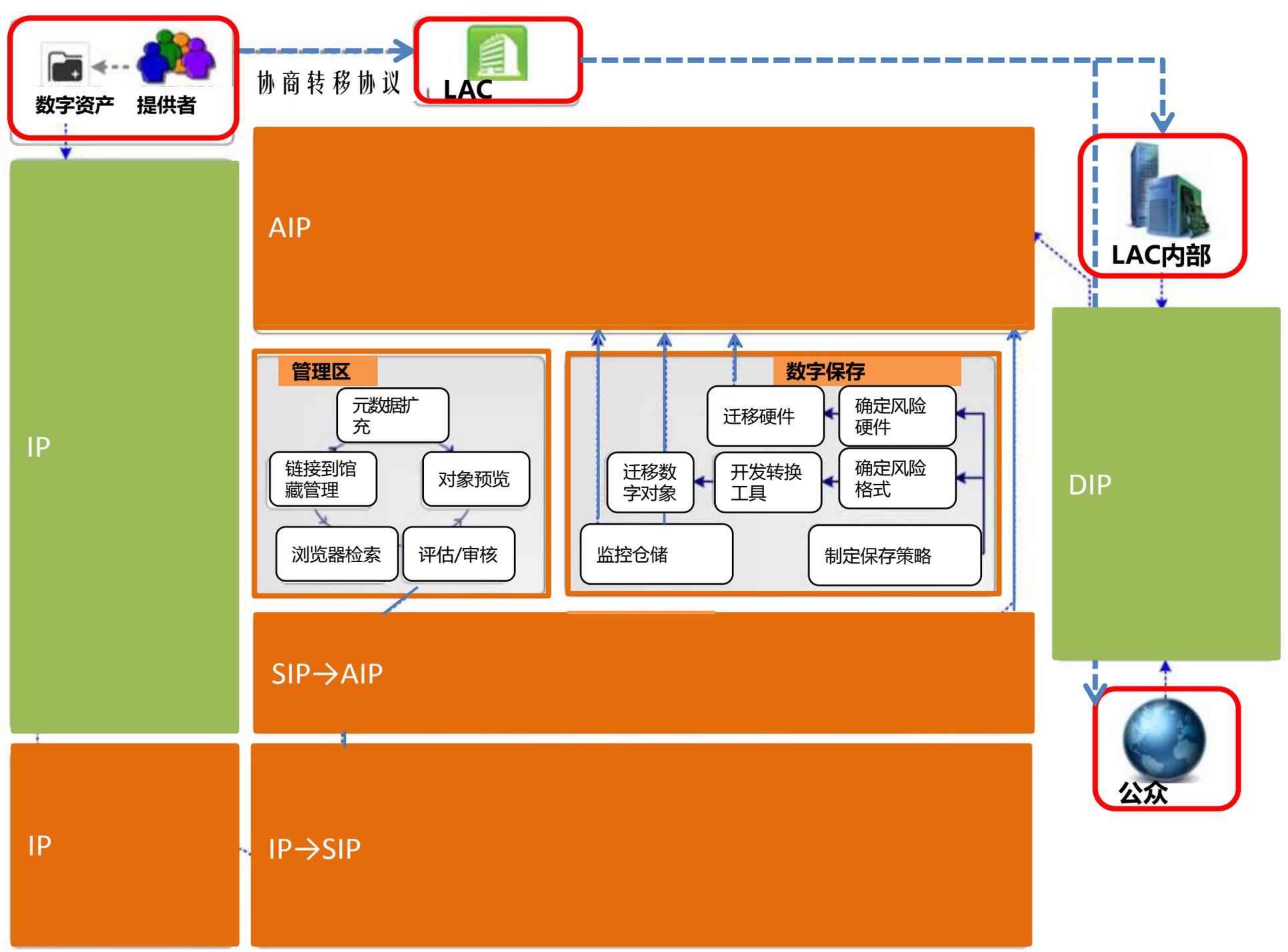


Legend:	OAIS Functions	ERA System -Level Packages	Service Oriented Architecture
	1. Ingest	Ingest	Business Application Services
	2. Archival Storage	Archival Storage	
	3. Data Management	Records Management	Common Infrastructure Services
	4. Access	Dissemination	
	5. Preservation	Preservation	
	6. Common Services	Local Services & Control ERA Management	

ERA\_ENG\_081c

# 加拿大LAC TDR

- 为了实现加拿大政府数字资源的长期保存，LAC 建立了**基于OAIS参考模型的TDR**。
- 通过对LAC TDR的研究，**加拿大图书档案馆自身就要成为一个可信任数字仓储**。2008年，加拿大图书档案馆正式发布了可信任数字仓储计划。
- **LAC 开始检查现存政府电子馆藏**，对其进行评估以便继续保存，并为把它们集成到TDR 准备好所要求的元数据。
- **LAC 研制了一套基于风险的评估方法**，该方法可以帮助政府各部门处理大量电子遗留系统文件，减少可以被删除的文件并自动为留下的文件补充元数据。
- 为了应对遗留系统数字资源问题，LAC在许多地区都成立了**数字办公室**，鼓励各类信息专业人员在电子领域开展能力建设



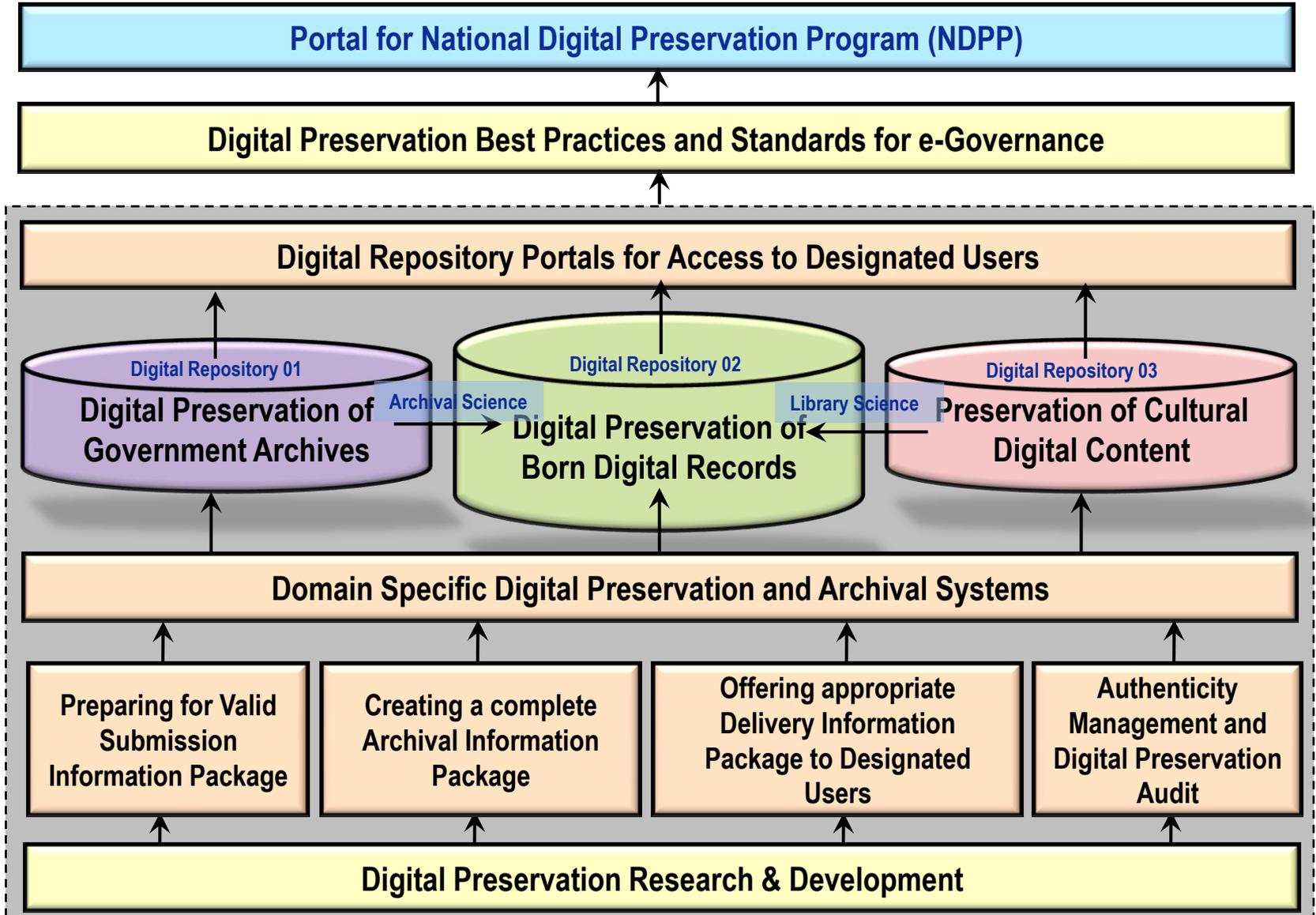
# 澳大利亚

- <http://www.naa.gov.au/collection/preserving/digital-records/index.aspx>



### Scope of CoE-DP Project

T  
O  
O  
L  
S  
&  
T  
E  
C  
H  
N  
O  
L  
O  
G  
I  
E  
S



# 谢谢!



钱毅 [qianyi1973@126.com](mailto:qianyi1973@126.com) [Y.COM](http://www.y.com)