

从“数字仓库”到“智能基座” 文化记忆机构数据开放的三重跃迁

夏翠娟 上海图书馆 图情学研究中心

2025年11月27日

Content

目录

01 多模态资源数智化开发利用的三重跃迁
——数字化、数据化、智慧化

02 数据开放模式的三重跃迁
——数字仓库、数据要素、智能基座

03 用户形态的三重跃迁
——人、应用程序、智能体

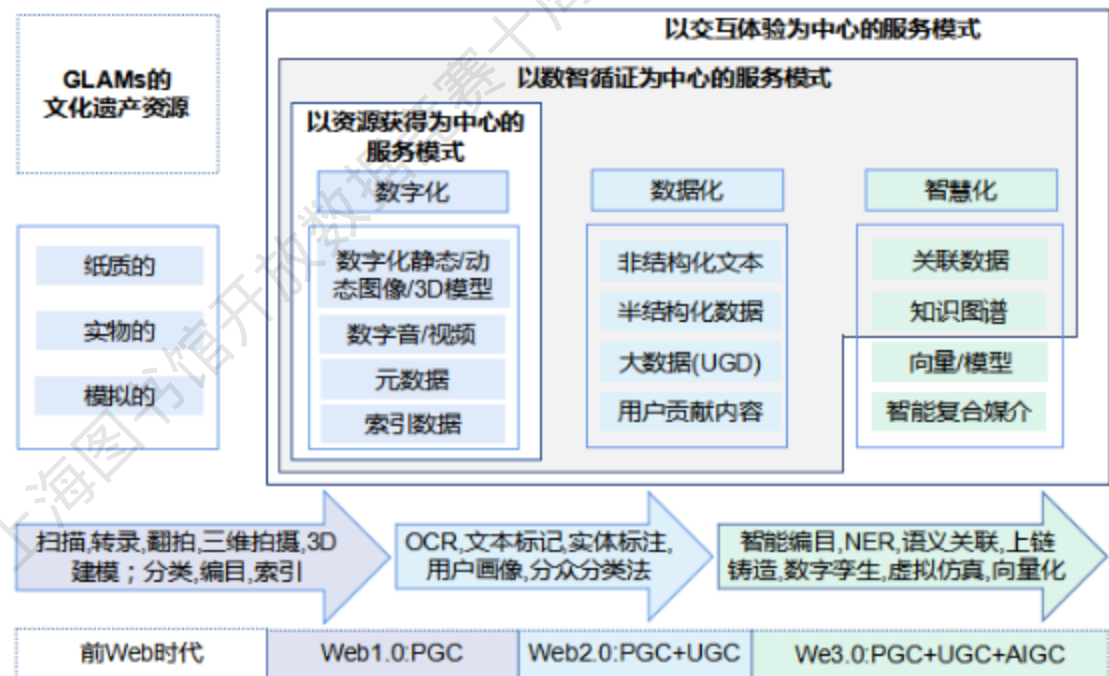
01 多模态资源数智化开发利用的三重跃迁

➤ 加工:

- 数字化、数据化、智慧化

➤ 服务:

- 以资源获得为中心的服务: 数字化
- 以数智循证为中心的服务: 数字化+数据化
- 以交互体验为中心的服务: 数字化+数据化+智慧化



夏翠娟.多模态文化遗产资源的智慧化服务模式研究——从可获得到可循证和可体验[J].信息资源管理学报,2023,13(05):44-55.

01 一部古籍的多重生命

静态镜像（数字化）、元数据（结构化）、关联数据（语义化）

->>> 数字藏品、训练语料.....



图像 (tiff)

為一書以復朱子之舊而更以黃氏之言附益焉
世之通人以吾言為然否
左 衛成公殺叔武一段解
衛侯先期入
杜注不信叔武僭叔武之死成公意也前驅
逆知公意故一見而速殺之下文元兩之語正詁
叔武之無罪見殺也征南註此句最吃 先期
八者間道疾入不使人知
子先長君守門以為使也與之乘而入
杜注詳衛大夫 子惠公之欲速故先入
人僭叔 子知公之欲與叔武為難也
而先入然已在公後長君不知公已
故猶以 子為公使左氏下此三語欲
公之入國人始猶莫知
公子邴大舉仲前驅
杜注衛侯速驅 子未幾二子衛大夫僭按
公知衛子之即在後衛其至則後已
叔武將沐聞公至喜捉髮走出
僭叔問道疾入均至於出不意故貪卒出迎
前驅射而殺之公知其無罪也杖之殺而哭之 夫
十六

文本 (txt)

文獻名稱	文門	收藏單位名稱	館藏索取號	全文索取號
欽定四庫全書	漢文	上海圖書館	線書812441	812441
欽定四庫全書	漢文	上海圖書館	線書812494	812494
漢文古籍	漢文	上海圖書館	線書829929-31	829929-31
漢文古籍	漢文	上海圖書館	線書830012-17	830012-17
漢文古籍	漢文	上海圖書館	線書753995-4002	753995-4002
漢文古籍	漢文	上海圖書館	線書758489	758489
漢文古籍	漢文	上海圖書館	線書754003-06	754003-06
漢文古籍	漢文	上海圖書館	線書829046-53	829046-53
漢文古籍	漢文	上海圖書館	線書829822-30	829822-30
漢文古籍	漢文	上海圖書館	線書828904-06	828904-06
漢文古籍	漢文	上海圖書館	線書734011-15	734011-15
漢文古籍	漢文	上海圖書館	863326-35	863326-35
漢文古籍	漢文	上海圖書館	線書754016-21	754016-21
漢文古籍	漢文	上海圖書館	線書795539-48	795539-48
漢文古籍	漢文	上海圖書館	線書795549-51	795549-51
漢文古籍	漢文	上海圖書館	線書794182-331	794182-331
漢文古籍	漢文	上海圖書館	線書829070-99	829070-99
漢文古籍	漢文	上海圖書館	線書758543-46	758543-46

结构化数据 (csv)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:skos="http://www.w3.org/2004/02/skos#"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <rdf:Description rdf:about="http://data.library.sh.cn/gj/renwu/instance/gj/renwu812441">
    <dc:title>欽定四庫全書</dc:title>
    <dc:creator>乾隆</dc:creator>
    <dc:publisher>上海圖書館</dc:publisher>
    <dc:subject>經史子集</dc:subject>
    <dc:format>PDF</dc:format>
    <dc:language>中文</dc:language>
    <dc:country>中國</dc:country>
    <dc:coverage>全書</dc:coverage>
    <dc:temporal>1795-1807</dc:temporal>
    <dc:spatial>北京</dc:spatial>
    <dc:contributor>上海圖書館</dc:contributor>
    <dc:rights>CC0</dc:rights>
    <dc:identifier>812441</dc:identifier>
    <dc:ispartof>http://data.library.sh.cn/gj/renwu/instance/gj/renwu812441</dc:ispartof>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu812494</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu829929-31</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu830012-17</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu753995-4002</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu758489</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu754003-06</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu829046-53</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu829822-30</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu828904-06</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu734011-15</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu863326-35</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu754016-21</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu795539-48</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu795549-51</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu794182-331</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu829070-99</dc:haspart>
    <dc:haspart>http://data.library.sh.cn/gj/renwu/instance/gj/renwu758543-46</dc:haspart>
  </rdf:Description>
</rdf:RDF>
```

语义化数据 (rdf)

02 开放模式的三重跃迁1：从“数字仓库”到“数据要素”

痛点：数字副本 (JPEG/PDF) + 元数据，无法直接计算，支持检索以供人读。

- 处理技术：自然语言处理 (NLP)、命名实体识别 (NER)、关联数据 (Linked Data)...
- 应用技术：地理信息系统 (GIS)、社会网络关系分析 (SNA)、IIIF...
- 价值：从数字图书馆时代的信息孤岛到开放关联的语义知识库。

破解数据孤岛

数字仓库中的资源和元数据仅支持网页发布、检索和浏览以供人读，无法被机器理解、计算和分析，制约内容价值的释放。

构建知识网络

利用 NLP、NER、LD 技术，将静态的数字化资源对象和二维的元数据记录转化为网状互联和开放调用、可计算的要素。

激活创新应用

支持数智循证、文化分析、记忆展演等专业研究和大众传播，展现数据要素的科研支撑、文化传承、终身教育等社会价值。

02 开放模式的三重跃迁2：从“数据要素”到“智能基座”

-知识生产
-数据共享
-未来共创



核心转移

FROM 解释历史

TO 构建记忆 & 生成未来

- 从对过去的静态解释
- 转向构建动态记忆
- 和生成可能的未来



需求变革

FROM 远读

TO 多模态智能交互

- 相似性计算
- 跨模态检索
- 自然语言交互



角色转换

FROM 数据生产者

TO 共创赋能者

- 路径：保管者 → 生产者 → 赋能者
- 从机械化的生产转向主动赋能与协作



技术升级

模型 & 架构进化

- 大语言模型 (LLMs) & 多模态大模型 (MLLMs)
- 向量数据库 & 检索增强生成 (RAG)
- 微调、继续预训练 & 智能体 (Agent) 开发



范式转向

AI4H →

AI4H & H4AI

- 支持广泛参与 + 迭代部署
- 解决不可解释性、幻觉和价值对齐问题
- 促进 AI 向善的良性发展

03 用户形态的三重跃迁：人、应用程序、智能体（LLMs）

-打包下载
-RestAPI
-LLM API
-MCP

1.0 传统 API 时代

2010年代 - 2022



核心交互范式
预定义端点和参数

数据接入技术

- REST / GraphQL 接口
- OpenAPI (Swagger)

痛点：每个集成需单独开发和维护，成本高昂，扩展性差。

消费端角色转变

操作者 (OPERATOR)

人工点击界面 -> 应用程序调用 API

2.0 早期 LLM 集成

2023 - 2024



核心交互范式
函数调用 / 工具集成

数据接入技术

- LLM OpenAPI 规范
- 向量数据库 (RAG)

痛点：生态碎片化，模型平台标准不一，数据实时性差。

消费端角色转变

监督者 (SUPERVISOR)

人工输入指令 -> LLM 模型选择工具

3.0 MCP 智能体时代

2025年及以后



核心交互范式
上下文暴露

数据接入技术

- Model Context Protocol (MCP)
- Stdio / SSE 标准传输

优势：一次开发，全平台、跨Agent通用，数据实时可见。

消费端角色转变

意图下达者 (INTENT SETTER)

智能体自主感知 -> 规划 -> 执行

一起来、更精彩！！！（2016-2025）

人智协同才有未来... ..

夏翠娟：策划+内容要点+监督生成+反馈优化

Deepseek、通义千问等大模型：演讲大纲、优化完善、提供建议、PPT生成、
布局画图、排版设计等

业绩是AI的，问题是我的；xtykc@yeah.net