

# 开放科学条件下科技资源建设与服务 ——以万方数据为例

李旭林  
北京万方数据股份有限公司



# 目录

1 开放科学环境下的信息服务观察

2 万方数据资源建设思路

3 万方数据资源服务模式与探索

WANFANG  
DATA

[www.wanfangdata.com.cn](http://www.wanfangdata.com.cn) | 🔍

# 1 开放科学环境下的信息服务观察

# 1 开放科学环境下的信息服务观察

- 开放科学是“使科学研究中开放透明可访问的知识，通过协作网络共享和发展，使出版和传播科学知识变得容易获取”，是基于**合作工作**、**数字技术**和**协作工具**的知识传播新方法，旨在提高研究成果的**公共可用性和可及性**。

**知识类型：**科技出版物、科学数据、教育资源、软件代码、硬件等等。

**关联过程：**关联整个研究过程，并倡导越早越好

**参与对象：**传统科学界+科学界之外的社会行为者

**支撑基础：**数字技术+协作工具、基础设施

**关键目标：**参与研究过程的所有参与者之间的开放和无缝合作；社会参与者在研究过程中的有意义参与



1. European Commission, Directorate-General for Research and Innovation, Hessels, L., Koens, L., Diederens, P., Perspectives on the future of open science : effects of global variation in open science practices on the European research system, Publications Office of the European Union, 2021, <https://data.europa.eu/doi/10.2777/054281>

2. UNESCO Recommendation on Open Science. <https://unesdoc.unesco.org/ark:/48223/pf0000379949>

## 知识内容的变革

- 开放环境，首先在**内容生产端发生了重大变革**，各种以传播、交流为目标，以各种新型平台为依托的新型“出版”正在颠覆传统信息服务的内容基座，一次知识内容的载体形态、类型、产生环节、产生方式均发生着重大的变化，由此引起的二次、三次内容生产也面临着变革。

研究过程中的各类内容成为学术交流与传播的对象，以Open Research Europe为例，支撑**全领域14类出版**。Octopus一类平台尝试**研究要素**的出版。

“出版”成为**交流的过程**，而不是一个结果，可**动态更新**内容，内容的关联、回溯成为重要组成

学术交流与传播的对象的多元化，引发了后续**评价体系中知识内容的多元化**，多元评价有了支撑点。

研究论文	短评	注册报告	指南出版
视频出版	数据出版	软件出版	方法出版
案例出版	临床案例	同行评议	.....
研究问题	研究假说	研究方法	研究结果
分析过程	结果解释	实际应用	
预印本	eLife	.....	

引用（论文、数据、软件、新闻提及、政策提及等）	
使用行为（下载、浏览、点评、转发等）	.....

## ➤ 用户需求的变革

- 开放科学推进传统科学界之外的社会行为者参与到科技创新中，开放教育加速信息素养的提升，由此信息服务的需求向着更大规模、更高性能、更加智能、更加个性化持续发展，用户面临可使用资源规模与信息获取成本的博弈，市场面临传统付费墙与新生“墙”的博弈，这些都需要先进技术手段、更适配的服务生态的调和，“以精确数据了解用户，以良性交互聚拢用户，以个性关怀留住用户”。

关键词检索



语义检索-认知检索

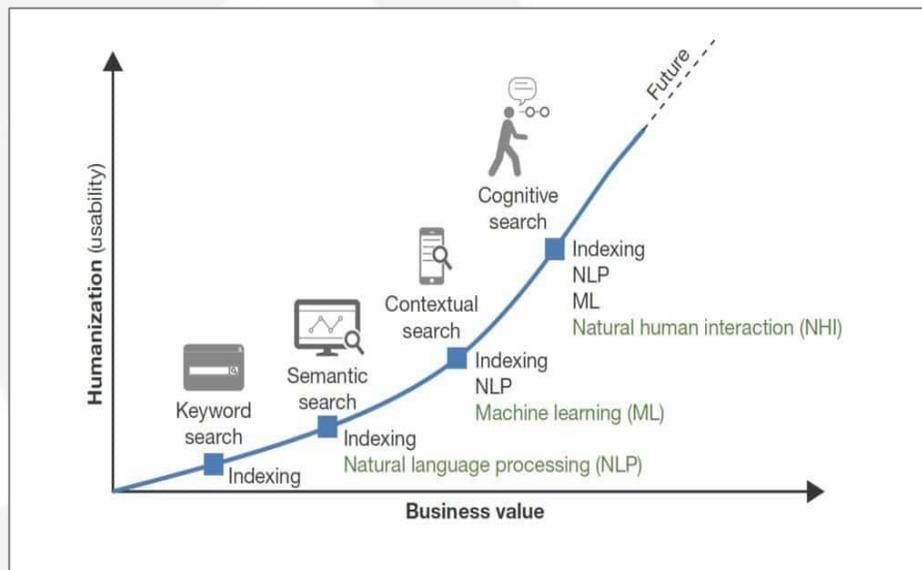


生成式



Etzioni, O. Search needs a shake-up. Nature 476, 25–26 (2011). <https://doi.org/10.1038/476025a>

大规模汇聚



Cognitive Search. <https://marionoioso.com/2019/04/10/cognitive-search/>

大规模汇聚、知识化处理

15 利用python快速实现上述的公式/算法，并且提供一个具体的例子。

好的，以下是Python代码来计算一个研究人员的h指数：

```
python
def calculate_h_index(citations):
    # 将引用次数从高到低排序
    citations.sort(reverse=True)
    # 初始化h指数为0
    h_index = 0
    # 遍历每篇论文
    for i in range(len(citations)):
        # 如果引用次数大于等于当前的i指数，则更新h指数
        if citations[i] >= i + 1:
            h_index = i + 1
        else:
            # 如果引用次数小于i指数，则当前i指数已经是最大的
            break
    return h_index
```

该函数接受一个列表，其中包含每篇论文的引用次数，并返回该研究人员的h指数。

以下是一个例子，展示了如何使用该函数计算一个研究人员的h指数：

利用ChatGPT快速复现论文的算法与模型，让你的研究更高效. <https://mp.weixin.qq.com/s/ux43b3yZHqleY1u2rFKUxQ>

大规模训练、人机交互

## ► 行业生态的变革

- 信息技术与人工智能技术的急速发展，快速打破传统出版中发行、传播、服务的壁垒，原有的出版、发行、传播链条发生极大变化，出版即服务的趋势越来越明显，国际大型信息服务商逐渐形成覆盖整个科研生命周期的闭环，实现了从出版商到服务商再到问题解决方的角色转换，国内信息集成商由于出版环节的缺失或不足，正面临着挑战。

出版

发行

集成

传播

技术

.....



**角色转换：**科技出版商→信息服务商→问题解决方案提供方

**链路延展：**聚焦研究成果→覆盖产学研用整个周期，重视“链路+场景”

**竞争加剧：**技术商有了开源资源的加持，加入赛道并抢得优势；内容端借助开源技术，直接提供服务

**合作生态：**新的合作生态正在形成，传统科学界+科学界之外的社会行为者参与到信息服务链中，需要协调多个利益攸关方。

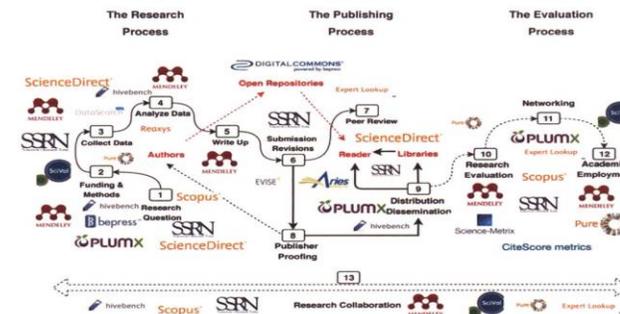
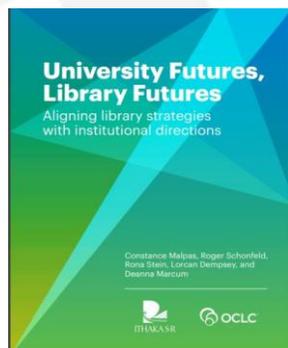


图5 爱思唯尔的业务延伸到科研价值链的每个环节<sup>[14]</sup>



## 图书馆角色升级

- 高校作为开放科学运动中的重要力量，高校图书馆在此环境下的角色也发生了重要变化：图书馆在传统的文献信息保障职能外，逐步成为信息素养教育、学术研究、校园文化建设、大学发展战略规划、产学研社区协同的重要基础保障，是贯穿科研生命周期的重要支撑。



University Futures, Library Futures:  
Aligning library strategies with  
institutional directions. (2018)



教育部关于印发《普通高等学校  
图书馆规程》的通知

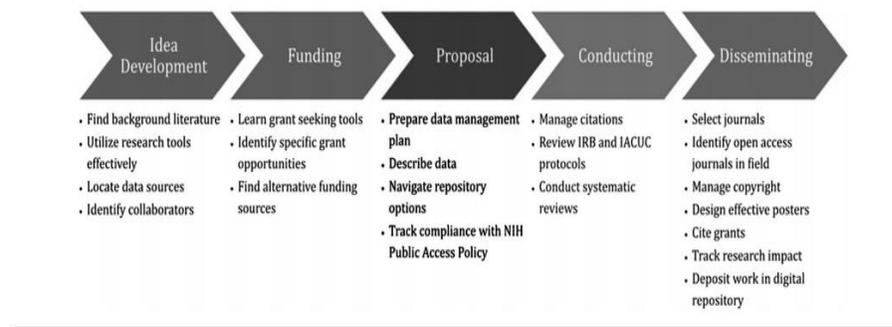
《普通高等学校图书馆规程》(2015)



The role of academic and research libraries  
as active participants and leaders in the  
production of scholarly research (2021)



Figure 1  
Library services across the research lifecycle



WANFANG  
DATA

[www.wanfangdata.com.cn](http://www.wanfangdata.com.cn) | 🔍

## 2 万方数据资源建设思路

## 万方数据资源建设层次结构

### 应用服务层

资源发现

团队识别

主题监测

竞争分析

战略情报

专利分析

标准应用

### 分析工具层

JCR

ESI

检测

诚信

评价

查新

选题

### 知识/组织层

论文引文文献库

项目库

机构本体库

专家作者库

奖项库

专利分析库

术语词表库

标准库

### 数据加工层

元数据采集与融合系统

科研实体识别规范系统

知识单元标引揭示系统

大规模语料模型

### 资源合作层

期刊

学位

会议

预印本

专利

标准

方志

视频

### 源头支撑层

期刊采编发系统

论文审稿系统

学位论文呈交系统

会议征文发布系统

作者授权与会员注册系统

### 素质素养层

论文写作技巧

信息素养培训

学术规范普及

科研诚信教育

学术评价报告

行业标准引领

## 2 万方数据的资源建设思路

- 1、推进细粒度数据资源建设。建立可计算数据仓储和特定数据集。
- 2、尽可能实现更大范围的数据关联。基于知识图谱、知识计算构建多源异构数据的关联网络。
- 3、加强数据的扩展应用。通过工具的设计和开发，进一步提升数据价值，形成面向用户具体需求的扩展应用。
- 4、强调工具与用户的应用场景紧密嵌入。打造研学支持，科技评价，诚信监测等工具应用，在工具基础上进一步构建平台服务。
- 5、通过数据库、工具箱、服务集等不同层次的打造，从以文献服务为主向数据驱动型组织转型。

WANFANG  
DATA

[www.wanfangdata.com.cn](http://www.wanfangdata.com.cn) | 🔍

## 3 万方数据资源服务模式与探索

### 3 万方数据的资源服务模式与探索

#### 市场营销板块

品牌市场推广

客服/用户分析

高校市场

企业市场

公图市场

政府市场

海外市场

#### 技术储备板块

语义语料模型

自动文摘综述问答

自动分类标引

全文检索技术

可视化图谱

#### 产品功能板块

知识服务平台

系统：医药系统、基教系统、文化遗产系统

工具：检测系统、诚信系统、评审系统、指标系统、专家系统

#### 评价分析板块

审读系统

检测查重

诚信规范

计量评价

高被引排行

学科分析

#### 数据加工板块

元数据采集

元数据加工

元数据融合

实体识别

主题标注

数据关联

#### 资源合作板块

期刊

学位

会议

视频

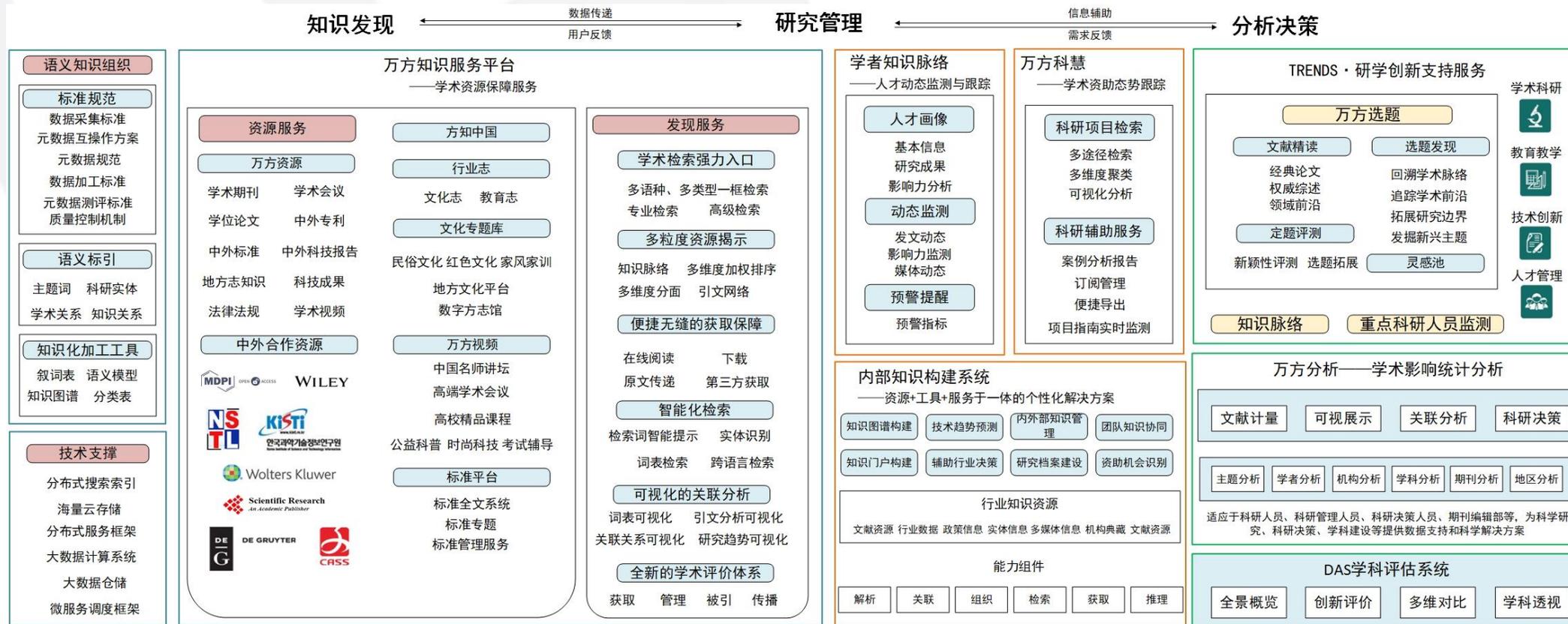
专利

标准

网络资源

## 面向科研周期的工具支持

以用户需求为驱动，以用户实际业务场景为切入点，在知识发现服务的基础上，对数据进行深度挖掘、分析和利用；针对学术研究和科研管理过程中的关键环节，打造紧密嵌入科研生命周期的服务支持，沉淀标准化、可复用、可共享的场景化服务能力。





# 面向科研周期的工具支持

耦合科研管理过程的循证服务：面向科研管理过程中的实际问题和决策场景，提供科技决策、成果管理、科研项目管理、人才管理与引进、科研诚信管理、学科发展评价、成果转移转化、科学数据管理、学术成果认定等服务。

**TRENDIS-科研趋势感知系统**

跟踪领域发展态势 支撑科研决策创新

案例分析：2022中国医学领域发展态势分析报告、2022中国教育学领域发展态势分析报告、2022中国医学技术领域发展态势分析报告

**学者知识脉络**

886万+ 44万+ 0万+

研究热点：信息技术、乡村振兴、COVID-19、核心素质、临床疗效、深度学习、建筑工程、教学改革、课程思政、施工技术、大数器、生活质量、教学策略、智慧教育、阅读教学

高影响力学者：李可法、赵志红、曹均波、王瑞峰、王梦君、刘玉、李素英、李学军、孙守敬、李敏

**学科发展评估平台**

清华大学

基本信息	发文量	被引量	篇均被引量	合作期刊数	作者数
2017-2021	58476	168934	2.9	4263	81679

整体研究概况：2017-2021

学科影响力仪表盘：学科影响力、学术影响力、合作影响力

**标准管理服务系统**

分类检索：国家标准、行业标准、地方标准、团体标准、企业标准

高级专业检索：关键词、分类号、作者、机构、年份、地区

有效性查询：标准状态、检索条件、导入导出

标准预警：标准发布、标准实施、标准废止

标准编写：标准编写、标准实施、标准废止

卫星导航标准专题：标准助力导航 卫星带你远航

行业标准化：国家标准67个，行业标准159,073个

**云计算**

总文献量：75525 | 总被引频次：280693 | 总下载频次：2759881 | 核心论文量：10824 | 基金论文量：11084

学科分布：理学、工学、医学、农学、教育学、文学、法学、哲学、历史学、艺术学、管理学、经济学、军事学、体育学、农学、医学、工学、理学、文学、法学、哲学、历史学、艺术学、管理学、经济学、军事学、体育学

影响力指数：280693 | 3.71 | 4884

**追踪全球科研动态 助力科研创新**

服务案例：全球顶尖学者研究动态追踪、2016-2018年中国人工智能领域发展态势分析、全球顶尖学者研究动态追踪

## 面向科研周期的工具支持

**面向出版源头的精准服务：**面向编辑部提供审稿人推荐、组稿选题、科学数据、期刊影响力评价等服务。同时，打通与期刊编辑部的联系，与编辑部的投稿系统对接，同时支持编辑部对期刊基本信息进行管理，以及可以上传征稿信息。

This screenshot shows the '学者知识网络' (Scholar Knowledge Network) interface. It features a search bar at the top and a main content area displaying search results for '冠状动脉疾病'. The results include a list of articles with titles like '1. 冠心病性——从社区到临床的防控策略' and '2. 冠状动脉粥样硬化性心脏病中的中医学'. Each result includes the author's name, the journal title, and a brief abstract. There are also navigation buttons like '查看详情' (View Details) and '收藏' (Bookmark).

This screenshot displays a detailed analysis page for '数字图书馆论坛'. It includes several data points and charts:
 

- 总发文量:** 3544
- 总被引:** 1,564
- 总被引次数:** 0.62
- 平均被引次数:** 0.20
- 引用期刊数:** 7.00
- 总被引次数占比:** 932 (24.47%)

 Below these statistics are line charts showing trends over time (2004-2020) for '发文量' (Volume) and '被引次数' (Citations). There are also smaller charts for '期刊影响力' (Journal Influence) and '期刊影响力指数' (Journal Influence Index). A '合作网络' (Cooperation Network) diagram is visible at the bottom left.

This screenshot shows the '学术期刊投稿分析系统' (Academic Journal Submission Analysis System) interface. It features a search bar and a list of recommended journals. The interface includes filters for '期刊匹配' (Journal Match), '核刊优选' (Core Journal Selection), and '征稿启事' (Call for Papers). The main content area displays a list of journals with their cover images and key metrics:
 

- 机床与液压:** 北大核心, CSTPCD, 一致性指数: 10.634, 影响因子: 0.734, 活跃指数: 0.21, 传播及时指数: 0.17, 学科融合指数: [Progress Bar]
- 自动化与仪器仪表:** CSTPCD, 一致性指数: 10.63, 影响因子: 0.963, 活跃指数: 0.31, 传播及时指数: 0.26, 学科融合指数: [Progress Bar]

 Each journal entry includes a '去投稿' (Submit) button and an '加入对比' (Add to Compare) button. The interface also shows a '共匹配到 1682 本期刊' (Total 1682 journals matched) and a '重新编辑匹配条件' (Re-edit matching conditions) button.

# 面向公共文化的数字人文服务

## 资源服务

整合旧方志资源十二万多卷、新方志五万余册，建成包含地方志、地情文献等资源的地情数字内容服务平台。



## 数据挖掘

挖掘方志文献，开展行业志、特色文化库开发，完成教育志、交通志、民俗文化、家风家训等专题产品建设。



## 项目定制

面向客户定制门户网站、数字方志馆等系统，实施杭州数字方志馆、芜湖党史方志网、掌上萧山等多个项目。



## 数字人文

基于方志内容，运用大数据技术，打造细颗粒度标引展示、妫川文化、武汉城市文化、视听资源等数字人文产品。



## ▶ 面向环境治理的科研诚信服务

科研失信案件频发，给我国科研学术生态造成不良影响，国家先后出台多项政策措施加强诚信建设，当前，诚信建设在快速落地、不断深化。应对科技监督与科研管理工作中的科研诚信建设需求，万方数据围绕科研活动的“事前-事中-事后”各阶段环节，开展科研诚信服务建设，打造教育培训、风险监测预警、问题筛查核验三大核心模块，提供科研诚信知识资源、学习培训、监测预警、检测查重、信息核查等五类系统产品与服务，并通过一站式综合服务平台为机构科研环境建设提供服务支撑。



# 结语

## 开放科学条件下科技资源建设与服务应关注的几个问题

### 1、基础设施建设能力

突破图书馆传统资源认知界限，将资源建设从狭义走向广义：从文献资源到数据资源、工具资源。重视内部专业资源的组织，实现内外部资源的统一揭示和集成管理。

建设开放灵活、易于拓展维护的全域数据架构，实现可追溯、端到端的一体化数据管理体系。

应用智能标引、知识图谱、语义挖掘等技术，以及更多的工具绑定，加强对数据的增值开发和高效利用，进一步提升数据价值。

### 2、场景化服务能力

专业图书馆不是旁观者，而是科研创新的重要要素之一。

打造紧密嵌入科研生命周期关键环节的服务支持，支持本地的开放科学活动，对出版流程、科学管理、科研进程、知识产权等，提供手段和工具支持，沉淀标准化、可复用、可共享的业务能力。

## 结语

### 3、人的能力

提升图书馆员数据素养，提升读者参与能力，建立互相关联、紧密交互的服务链条。

在专业资源、工具和服务汇聚的基础上，专业图书馆对智慧信息服务的特殊作用和必要性愈发重要，是走完最后一公里的驾驶员，也是实现知识和服务增值的执行人；既是具体应用场景的落地，也是运用通用知识工具解析专业知识内容的必要环节。未来专业图书馆资源服务的职能也许会逐渐弱化，但咨询服务与决策支持的作用会进一步增强，愈发融入科研和创新的实际场景。

期望未来能够与各专业图书馆在资源建设与服务方面建立更密切的合作机制，共同实现内外源数据的融合、最终知识服务产品的交付和专家智慧对知识服务的赋能，共同推进数智时代下信息服务的跨越式可持续发展。

谢谢！  
如有不当之处请指正！

传播知识，创造未来  
我们：永远在倾听您的声音，永远为您做好准备