

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 | 🔍

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”。

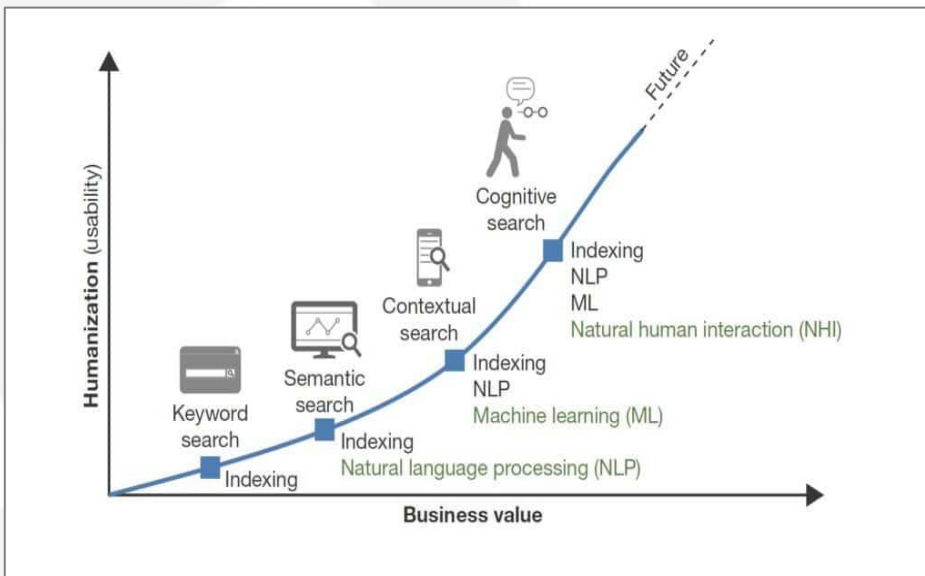
关键词检索



语义检索-认知检索



生成式



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

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>

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

大规模汇聚

大规模汇聚、知识化处理

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。

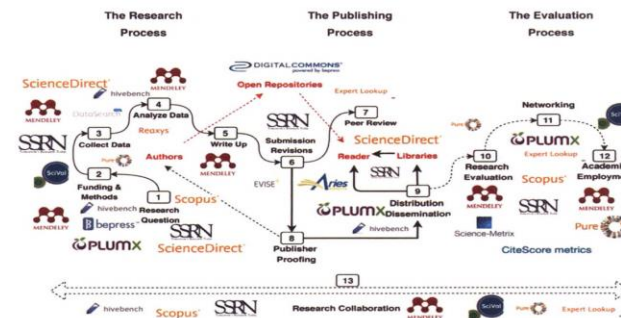
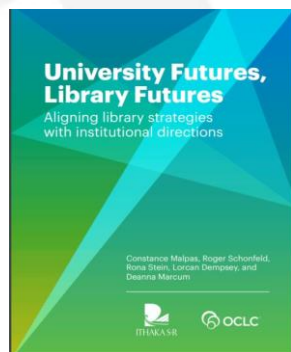


图5 爱思唯尔的业务延伸到科研价值链的每个环节^[14]



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)



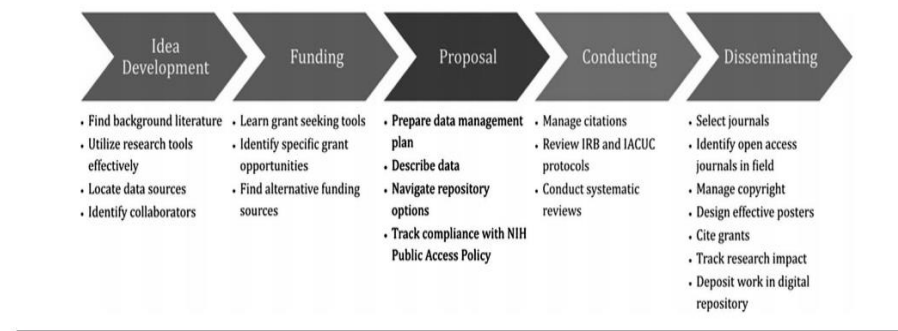
The role of academic and research libraries
as active participants and leaders in the
production of scholarly research (2021)



BIRMINGHAM CITY
University June 2021



Figure 1
Library services across the research lifecycle



WANFANG
DATA

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 | 🔍

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

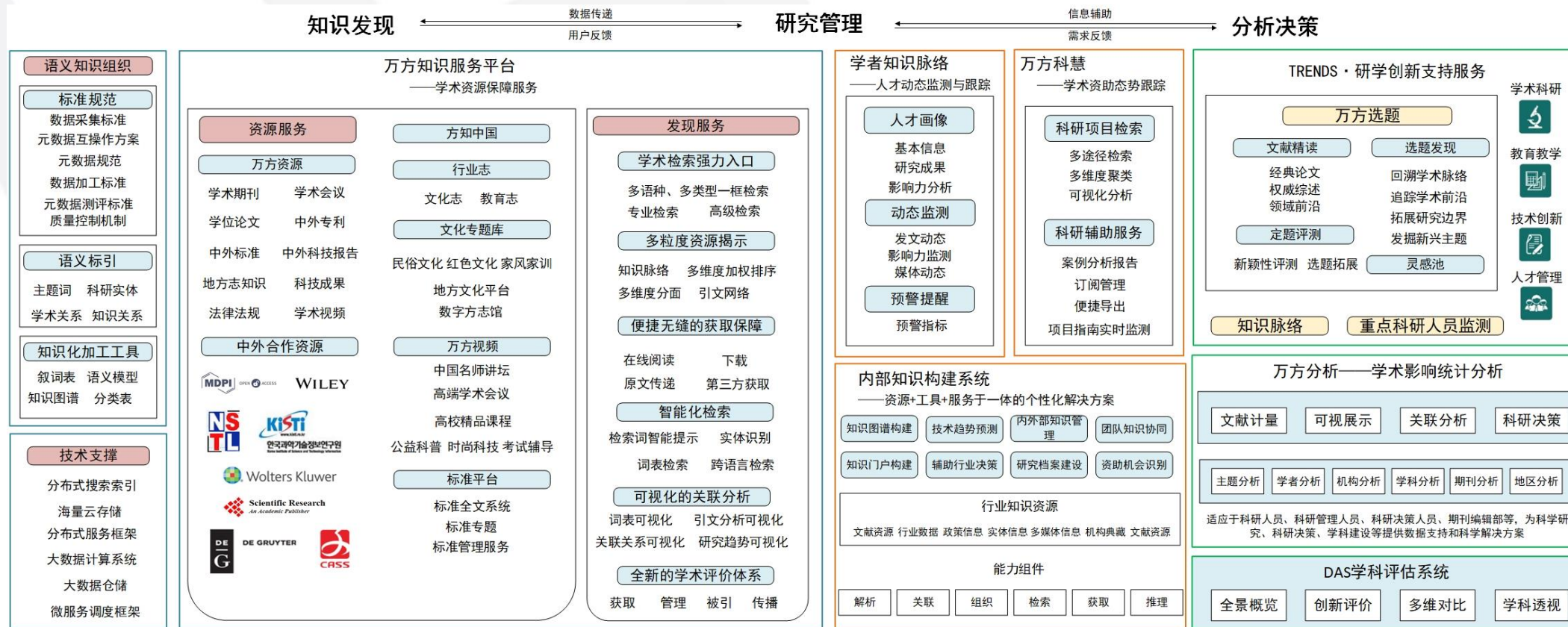
专利

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

贯穿学术研究链的支持服务：以知识发现为主线，在资源检索和阅读过程中，嵌接主题探索-文献管理-论文精读-学术交流-科研选题-趋势跟踪-智能投稿等一系列支撑服务，同时通过移动APP实现随时随地的资源发现与优质内容推送，在学术研究链条中的每一个重要环节为科研人员提供支持。

万方智搜 数据挖掘

全部 期刊 学位 会议 专利 科技报告 成果 标准 法规 地方志 视频 更多>>

高级检索 >> 检索 高级检索 >> 检索历史 >>

资源类型

- 期刊论文 (67470)
- 学位论文 (39955)
- 专利 (8442)
- 更多...
- 年份
- 2023 (384)
- 2022 (6078)
- 2021 (8630)

题名: 作者: 关键词: 起始年: 结束年: 结果中检索

排序: 相关性↓ 出版时间 被引频次 获取范围 显示 20 条 < 3 / 6295 >

批量选择 (已选择 38 条) 清除 批量引用 批量分析 已购全文 找到 125881 条结果

41. 卫星系统在役考核对数据挖掘的应用需求

期刊论文 | 钱照勇 曹福华 张雷 等 - 《科技与创新》 2023年1期

摘要: 针对装备在役考核对数据分析处理的各种需求非常复杂等列问题,结合航天装备在役考核的总体特征,将卫星系统在役考核划分为筹划设计、组织实施、分析评估3个主要阶段,立足这3个阶段,从卫星系统在役考核的基本流程展开,厘清其各个关键环节存在的主要...
数据挖掘 卫星系统 在役考核 需求

在线阅读 下载 引用

国内外文献保障服务

智能扩展

研究趋势

复制 导出TXT 导出XLS 导出DOC 编辑列表

参考文献

查新格式

NoteExpress

RefWorks

NoteFirst

EndNote

Bibtex

默认格式 + 新建格式

格式: 引用格式 列表格式 注: 拖动字段可改变字段顺序

期刊: 作者单位 摘要 关键词 ISSN DOI

会议: 作者单位 会议名称 母体文献 年份 摘要 关键词 DOI

导出时展示全部作者

[1] 当凡超,张磊,刘知远,等. WantWords:基于神经网络技术的反向词典[J]. 辞书研究,2022(4):41-47. DOI:10.3969/j.issn.1000-6125.2022.04.003.
作者单位:清华大学计算机科学与技术系 北京100084

万方选题

文献精读 选题发现 定题评测

您需要更专业的选题助手

- 海量学术资源
- 专业算法模型
- 深度数据挖掘
- 直观图表展示

文献精读 选题发现 定题评测

开始使用

新一代信息技术 智能制造 数字化转型 服务化转型

数字孪生 物联网 云计算 大数据 人工智能 区块链 5G 虚拟现实 增强现实 网络安全

智慧图书馆 智慧服务 智慧管理 智慧应用 智慧数据 智慧资源 智慧空间 智慧环境 智慧文化 智慧生活

公共图书馆 图书馆事业 图书馆治理体系 智慧图书馆

数字图书馆 数字资源 数字服务 数字应用 数字数据 数字资源 数字空间 数字环境 数字文化 数字生活

文献融合 研学旅行 公共图书馆 融合时代 融合服务

安全存储 数字图书馆 区块链 智慧图书馆

知识情报学 人才培养 数字人文 美国情报学...

新兴主题	突发强度	突发起始年	2013-2022
十四五+建设规划	9.96	2020	—
应用场景+数字孪生	9.96	2020	—
智慧图书馆+现代信息技术	8.81	2021	—
可视化分析+知识数据率	8.62	2021	—
智慧服务+知识图谱	8.18	2020	—
新兴技术+智慧服务	7.31	2020	—
应用场景+智慧化	6.22	2020	—
digital+数字孪生	5.91	2021	—
新兴技术+智慧图书馆	5	2020	—
智慧图书馆+高质量发展	5	2020	—
数字孪生+数字技术	4.5	2021	—
人工智能+智慧化服务	4.38	2020	—
数字孪生+智慧图书馆	3.78	2020	—

— 突发年份 — 有发文年份 — 无发文年份

精准投稿

为您的论文寻找最合适的期刊

智能匹配 精准匹配 智能推荐

期刊选择

智能匹配 精准匹配 智能推荐

开始使用

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+ 位学者创建了学术履历

44万+ 个机构的学者产出知识影响力

0万+ 篇学术成果

学者最新动态：张炜 新增文献《创业学习对大学生创业意...》、刘玉 新增文献《艾司唑仑与野苧大皂苷...》、王梦君 新增文献《信息化条件下高职公共...》、李素英 新增文献《中医专科护士持续培养...》、许可 新增文献《高职教育信息化建设的重...》、孙守敬 新增文献《高中英语核心素养视域...》、李敏 新增文献《健康中国视域下高职体育...》

高影响力学者：李可法 浙江大学、赵红 中国农业大学、曹均波 复旦大学附属中山医院、王瑞峰 东北大学

学科发展评估平台

清华大学

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

整体研究概况 (2017-2021)

成果产出能力、学术影响力、合作研究能力

标准管理服务系统

标准助力导航 卫星带你远航

行业标准化：本系统关联标准674个，共159,073个标准

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、专业名称、标准号、发布日期、发布单位、实施日期

有效性查询：标准名称、标准状态、导入导出

标准预警：标准变更、自动提醒、批量提醒邮件

标准编写：标准编写、查询标准全文

标准体系管理：标准体系管理、标准体系、标准体系、标准体系

内容发布：发布标准、发布标准、发布标准

标准收藏：标准收藏、标准收藏、标准收藏

全文互助：全文互助、全文互助、全文互助

云计算

2023 - 2022

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

科研趋势：趋势图显示2003-2022年趋势

影响力指数：280693 总被引频次, 3.71 篇均被引频次, 4884 篇均下载次数

追踪全球科研动态 助力科研创新

服务案例：展示多个科研项目的服务案例

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 '学者知识脉络' (Scholar Knowledge Network) interface. It features a search bar at the top and a main content area with several sections:

- 研究结果 (Research Results):** Lists search results for '冠心病' (Coronary Heart Disease) with metrics like 2107 articles and 17,394 citations.
- 研究趋势 (Research Trends):** Includes a line chart showing trends from 2006 to 2020 and a list of top authors like 魏来 (Wei Lai).
- 研究前沿 (Research Frontiers):** Highlights key research areas such as '冠心病发病机制' (Pathogenesis of Coronary Heart Disease) and '冠心病危险因素' (Risk Factors of Coronary Heart Disease).
- 合作网络 (Cooperation Network):** Displays a network diagram of researchers and their collaborative relationships.

This screenshot displays a detailed analysis for '数字图书馆论坛' (Digital Library Forum). It includes:

- 期刊概况 (Journal Overview):** Key statistics such as 3544 total articles, 1,864 new articles, and an impact factor of 0.82.
- 发文趋势 (Article Trends):** A line chart showing the number of articles published from 2004 to 2020.
- 期刊指数 (Journal Index):** Metrics like 10961 total citations and 153 new citations.
- 期刊影响力 (Journal Influence):** A line chart showing the journal's influence index over time.
- 合作网络 (Cooperation Network):** A network diagram showing relationships between authors and institutions.

This screenshot shows the '学术期刊投稿分析系统' (Academic Journal Submission Analysis System) interface. It features:

- 期刊匹配 (Journal Matching):** A section for matching journals based on specific criteria, showing 1682 matches.
- 核心收录 (Core Collection):** A list of journals with their core collection status (e.g., '全部期刊', '核心期刊').
- 卓越行动计划 (Excellent Action Plan):** A list of journals participating in the plan.
- 期刊分析 (Journal Analysis):** Detailed analysis for '机床与液压' (Machine Tools and Hydraulics), including metrics like 一致性指数 (Consistency Index) of 10.634, 影响因子 (Impact Factor) of 0.734, and 学科融合指数 (Disciplinary Integration Index).

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。



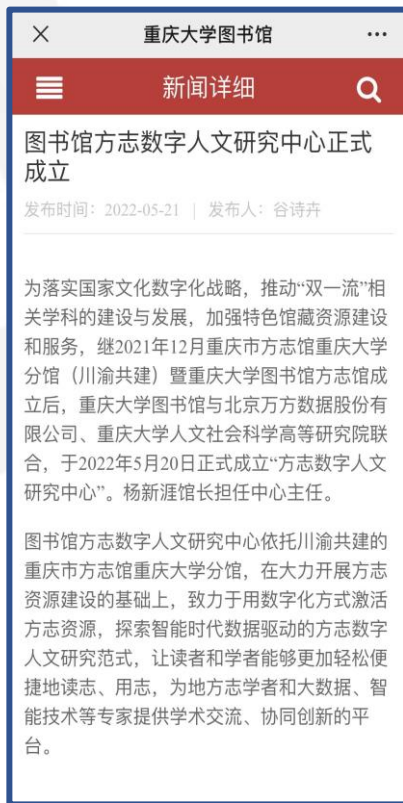
数据挖掘

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