

# 国家图书馆数据资源管理的实践与思考

国家图书馆研究院 申晓娟

2022年5月12日



全国科学技术名词审定委员会

数据资源

大数据新闻第一批发布试用

<https://blog.csdn.net/gyy77>

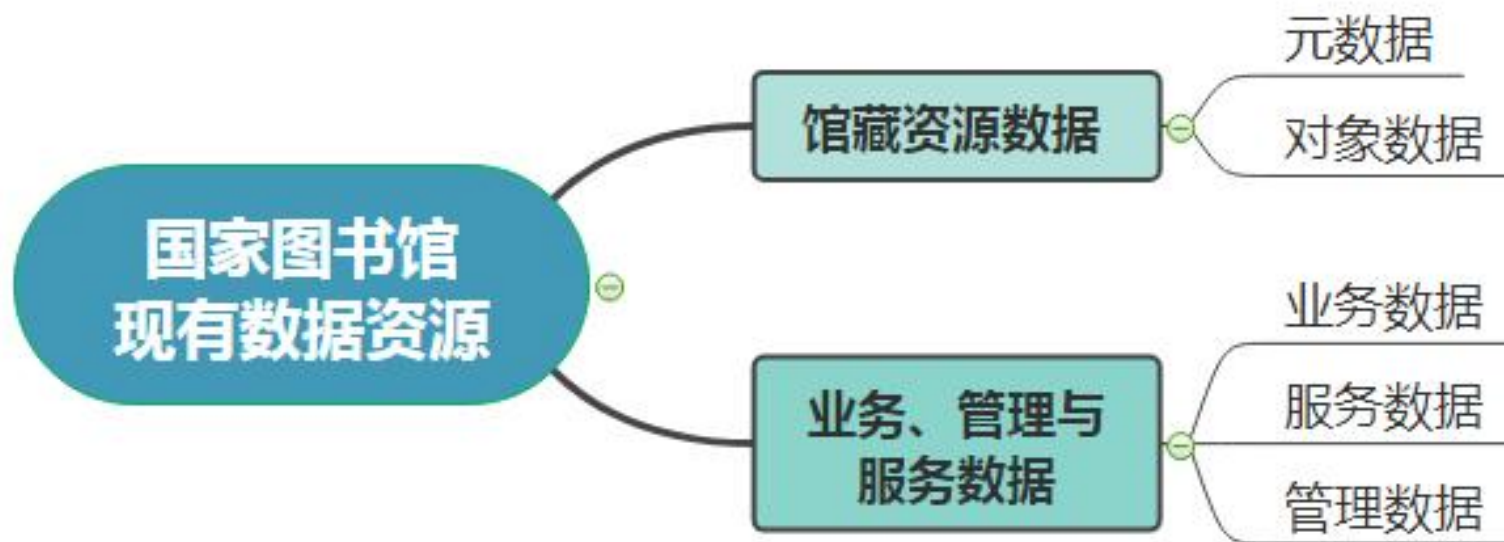
# 数据资源

- 广义上是指对一个机构而言所有可能产生价值的数据库
- 一些过去被我们所忽略或因技术、观念等原因没有能够深入处理的业务数据、服务数据等，逐渐有了更多的应用场景，也有了待挖掘的潜在价值
- 数据资源已经成为图书馆实现智慧化转型升级所依赖的核心要素之一

# 一、国家图书馆的数据资源管理基础



# 国家图书馆现有数据资源



# 馆藏资源数据

## 元数据（实体文献书目数据、数字资源元数据）

**馆藏书目数据** 截至2021年底，馆藏文献书目数据达1656.7万余条  
馆藏揭示率基本实现100%。

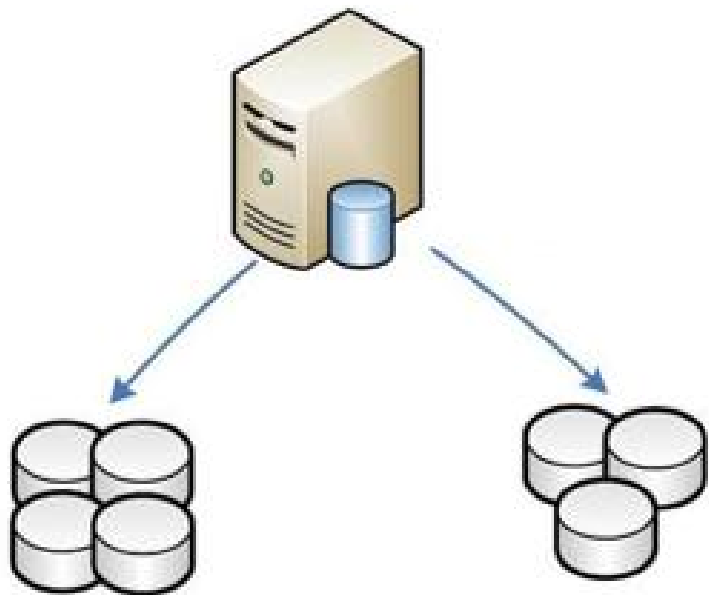
**全国联合目录** 截至2021年底，集成各级图书馆书目数据近1700万条，馆藏数据超过6600万条

**文津搜索系统** 整合元数据近3.8亿条，提供一站式检索发现服务

**专题联合目录建设** 发布260余家单位古籍普查数据82.5万余条；汇集42家收藏单位革命文献与民国时期文献书目数据30余万条、馆藏数据70余万条。

# 馆藏资源数据

## 对象数据（自建、外购）



- 古籍特藏文献和民国时期文献的数字化成果
- 中国记忆、国图公开课、文津经典诵读等项目制作的音视频资源
- 学位论文、中文电子图书等

**自建数字资源总量2497.98TB**

# 馆藏资源数据

## 对象数据（自建）

- 已建数字资源中，除受版权保护限制传播的部分近现代文献及少数特藏品类外，在局域网或互联网发布服务比例大多已达到或超过80%



# 馆藏资源数据

## 对象数据（联建）



“中华古籍资源库”联合发布甲骨、敦煌文献、碑帖拓片、西夏文献、赵城金藏、地方志、家谱、年画、老照片等古籍数字化全文影像数据**10万余部**

全国各级图书馆共建数字资源超过**2.5万TB**

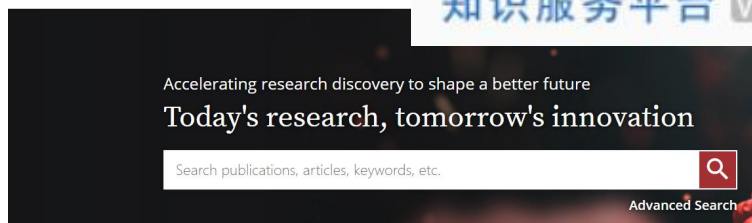
# 馆藏资源数据

## 对象数据（采购）

➤ 订购中外文数据库250余个



Wiley Online Library



WiseNews



# 馆藏资源数据

## 新型馆藏资源建设



# 业务、管理及服务数据

## ➤ 业务数据

文献采访数据、登记数据、编目数据、加工数据、装订数据、分流数据、入藏数据、数字化生产及验收数据、版权登记数据、发布数据、长期保存数据等

## ➤ 服务数据

读者数据、文献流通数据、注册办证数据、文献传递与馆际互借数据、业务咨询数据、各类活动数据、自建及外购数据库使用数据、网站点击数据、到馆读者网络登录数据、手机门户访问数据、新媒体服务数据等

## ➤ 管理数据

公文流转数据、资产使用及处置数据、人员数据、财务数据等

# 业务、管理及服务数据

**对业务、管理和服务数据进行多维度分析监测，为有关业务制度、服务政策、内部管理的优化调整提供数据支撑**

- **根据阅览室室藏文献利用及读者服务数据的统计分析，对服务空间进行优化调整**
- **利用文献借阅量、访问量、检索量、下载量等对馆藏外文资源质量进行分析评价**
- **根据外购数据库使用情况，结合经费优化调整数据库采购方案**



# 面临的主要问题

- 基于内容的细颗粒度数据知识服务能力还比较薄弱
- 对业务数据的价值挖掘利用还很不够
- 对服务数据的深度分析还不够，还做不到精准画像、针对服务
- 资源数据、服务数据、业务数据、管理数据之间还没有完全打通

**数据价值的更多潜力还在沉睡**

## 二、图书馆数据管理面临的新形势



# 01 政策要求

“加大大数据关键技术研发、  
产业发展和人才培养力度，着力  
推进数据汇集和发掘，**深化  
大数据在各行业创新应用**，  
促进大数据产业健康发展。”



——国务院《促进大数据发展行动纲要》，2015





大数据技术和智能传感设备日益普及应用，图书馆在业务管理和用户服务过程中采集到的文献信息流通、系统设备运行、用户线上线下活动等数据将极大丰富



## 03 数据资源精准服务的需求及技术赋予的新可能性

用户对精准高效知识发现  
服务的需求



图书馆传统的文献数据  
加工方式



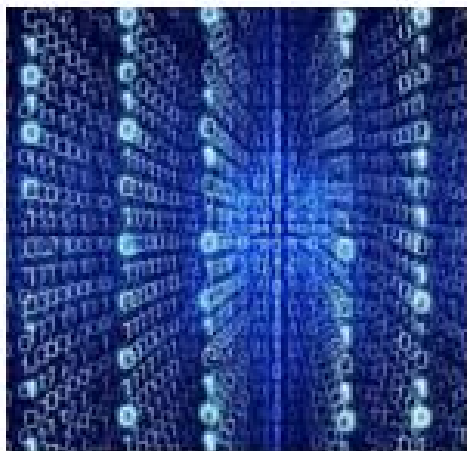
## 数据处理技术快速发展，为图书馆开展大规模的数据资源存储、揭示和关联提供了新的可能性

到2020年底，我国云计算整体市场规模达到**2091亿元**，比2016年增长**306%**，我国自主、安全、稳定的公有云已经形成规模化应用。



# 技术的发展为数据资源的生成、自动聚类与知识关联提供了新的技术工具 也为用户画像的精准描绘提供了新的技术手段

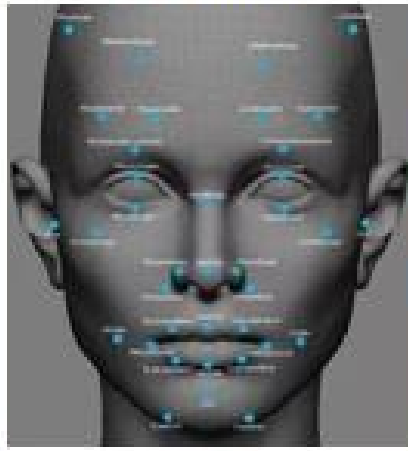
## 智能算法



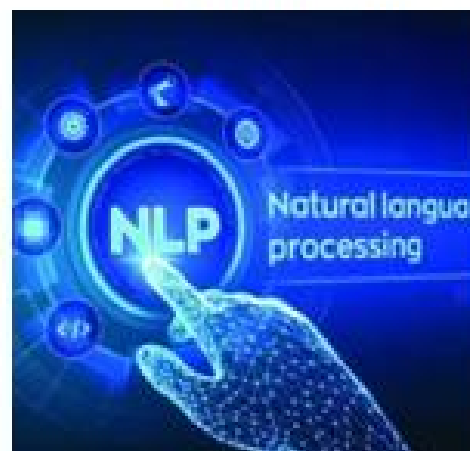
## 语音识别



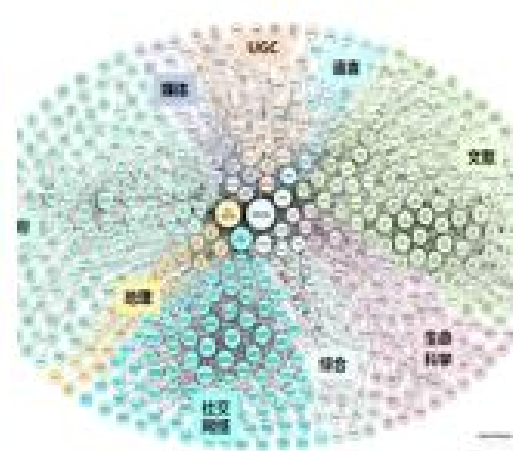
## 图像识别



## 自然语言处理



## 知识图谱



## 04 数据资产权益管理的紧迫性

到2020年，我国数字经济规模达到**39.2万亿元**，比2017年增长了**44.1%**。



资料来源：中国信通院 前瞻产业研究院整理

@前瞻经济学人APP

**可无限复制**是数据资源的最大特点和优势，如果能够将其转化成可单一标识和可追踪的数字资产，对数据的增值开发和服务利用就有了基础。



## 05 数据共享利用与网络信息安全的矛盾日益凸显





我国数据安全管理制度体系不断完善，《网络安全法》《数据安全法》  
《个人信息保护法》三部法律相继出台，构成数据安全保护的法律基础。



(2016年)



(2021年)



(2021年)

# 《互联网信息服务算法推荐管理规定》《网络数据安全管理条例》等政策文件，以及国家标准《信息安全技术大数据安全管理指南》对数据安全保护提出了更具体的要求



## ➤ 图书馆面临的主要数据安全问题

1

数据采集、保存和传输的网络和云基础设施安全

2

用户信息数据采集和分析利用过程中的个人隐私安全

3

数据资源文本挖掘和关联整合过程中的知识产权风险

# 三、关于国家图书馆数据资源管理的思考



## (一) 建设全网知识内容集成仓储，推动实现多源文化数据资源的融汇共享



**我们正在迎来一个知识极大丰富的时代!**

——全球商业太空探索领军人物，Peter Diamandis



**我们同时也处在一个知识焦虑的时代。**

## **(一) 建设全网知识内容集成仓储，推动实现多源文化数据资源的融汇共享**

### **依托云平台，建设全网知识内容集成仓储**

- ✓ 支持各级图书馆的资源采集、生产和加工
- ✓ 支持其他公益性、商业性和非营利性数据资源提供商的开放接入

# (一) 建设全网知识内容集成仓储，推动实现多源文化数据资源的融汇共享

## ➤ 突破图书馆传统的数据资源认知界限



- 将网络原生资源、科学数据资源、开放存取资源、个人创作资源等数据内容纳入图书馆数据资源建设范畴
- 统一揭示和集成管理，形成立体化数据资源仓储管理体系

## （一）建设全网知识内容集成仓储，推动实现多源文化数据资源的融汇共享

### ➤ 突破主要面向行业内部的资源合作共享格局



- 在集成各级图书馆数据资源建设和管理平台的基础上，营造**开放多元的知识运营环境**
- 推动建立贯通知识内容创作、发布、存储、传播、利用等**全域链条的社会化合作机制**
- 支持对全网公益性、非营利性、商业性知识内容的**数据集成管理和多元供给**



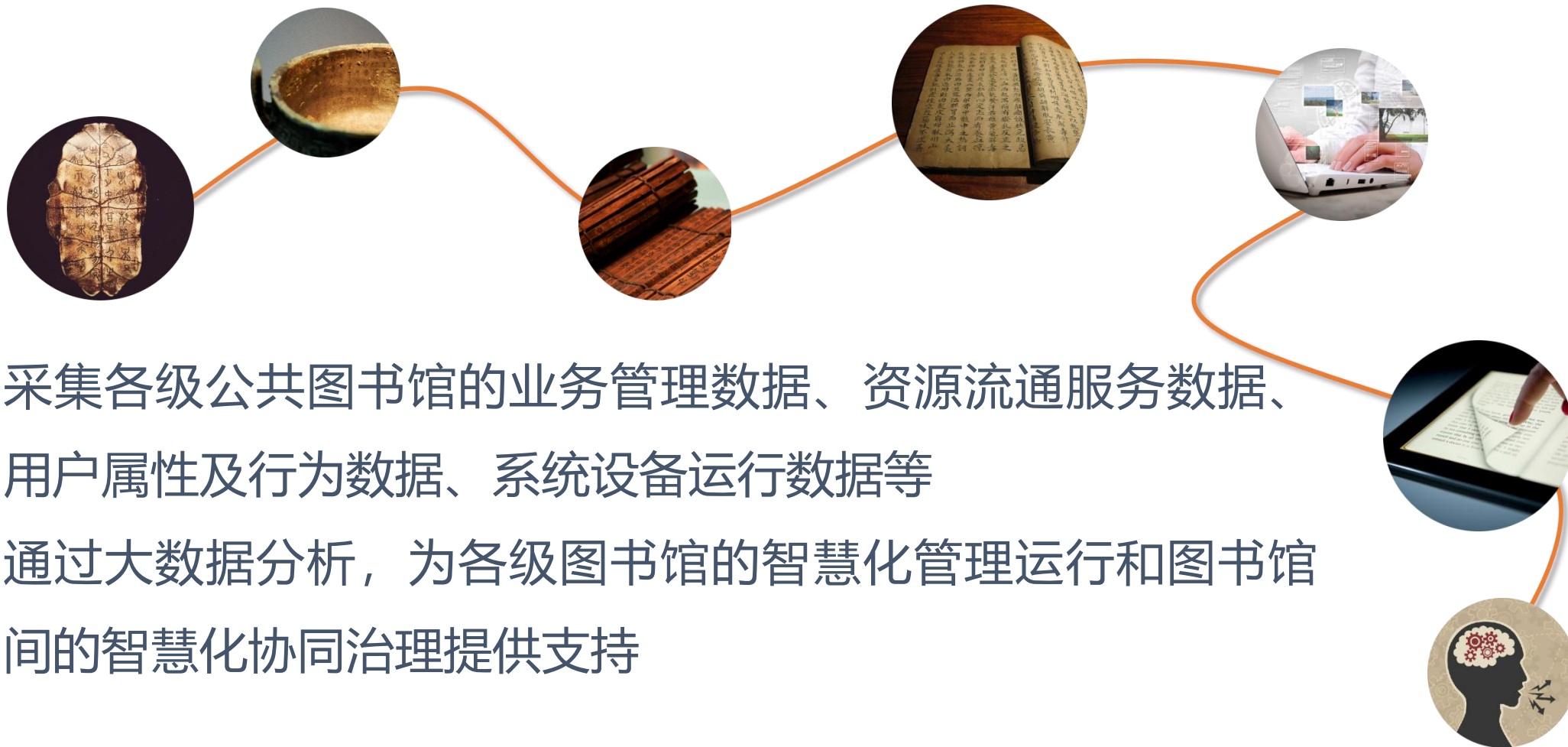
## (二) 加强数据资源的细粒度加工与关联揭示，推动数据资源的增值开发

**沉睡的数据是没有价值的**数据，未来智慧图书馆建设更重要的任务是要面向智慧化知识服务需求，对集成仓储的数据资源作进一步的**深度挖掘和增值开发**

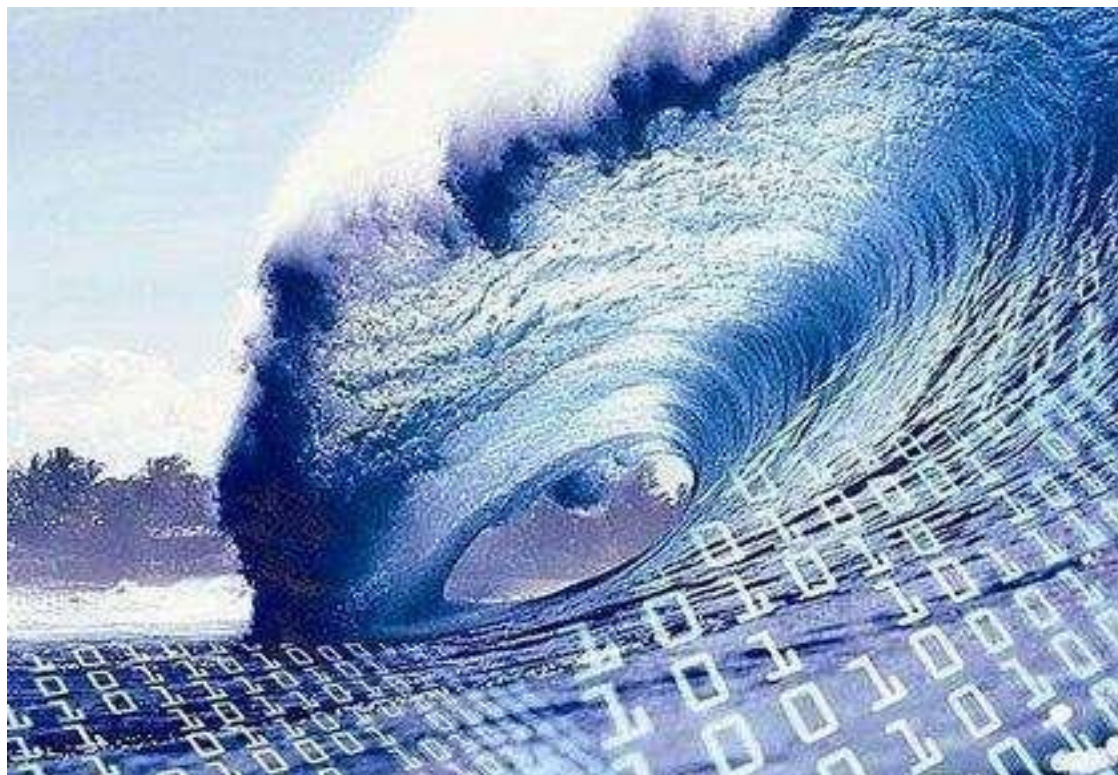


### (三) 建设全国图书馆大数据管理与分析平台，推动形成数据驱动的行业协同治理体系

#### ➤ 打造面向全国各级公共图书馆的大数据管理分析平台



### (三) 建设全国图书馆大数据管理与分析平台，推动形成数据驱动的行业协同治理体系

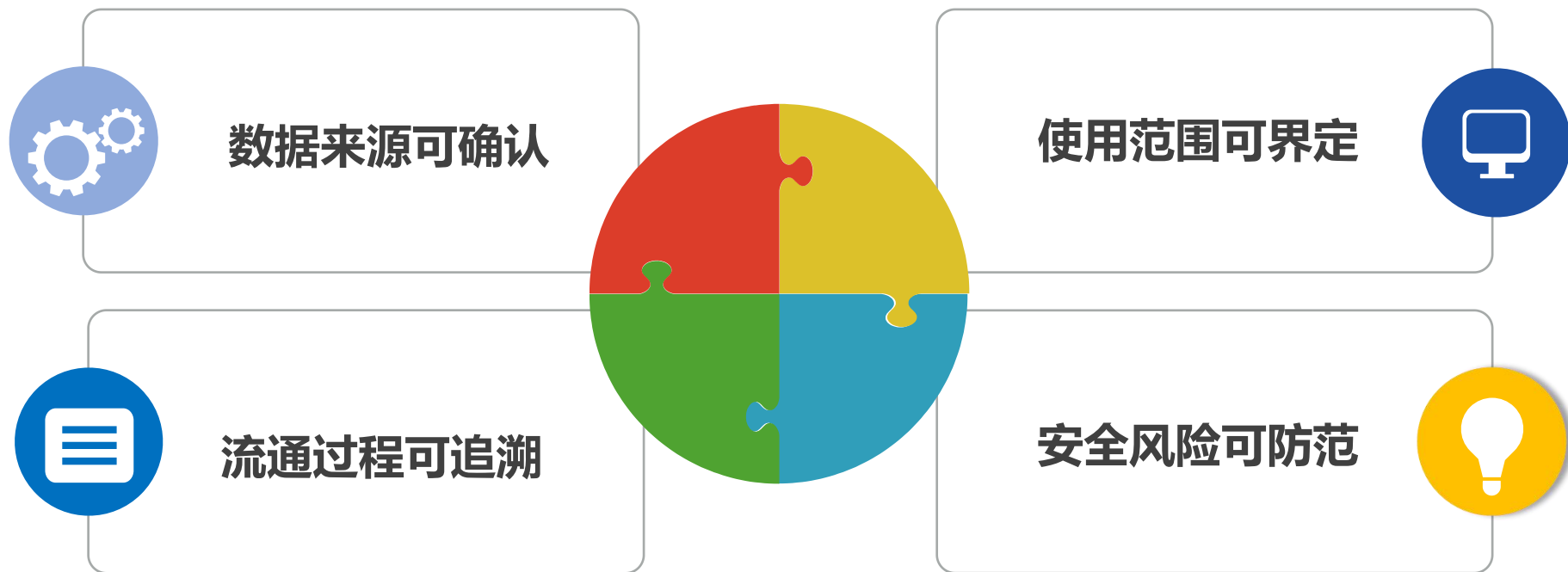


#### 面对数千家图书馆的海量复杂数据：

- ✓ 建立科学的数据采集策略和采集工作机制
- ✓ 选择适用的采集方式
- ✓ 实现对数据的自动识别和动态收割
- ✓ 对来自不同图书馆、不同系统的数据，进一步剔除冗余、建立关联

## (四) 完善数据资产化管理运行机制，打造“数据超市”

- 对图书馆数据资源进行**全流程资产化管理**，建立**数据资产权益管理体系**，是参与数据要素流转交换，促进数据资产保值增值的重要前提和基础



## **(五) 加强数据安全合规管理，支持图书馆数据资源建设、管理和服务的可持续创新**

大数据环境下，图书馆的数据安全管理进一步从数据层面扩展到承载数据的平台、应用、基础设施及开放多元的数据协同合作空间，上升到制度规范、方针政策、组织架构、管理和技术相结合的综合治理层面，更强调覆盖图书馆数据采集、加工、管理、存储、利用等全流程的体系化安全防控

- 通过技术和管理手段，保证图书馆自身控制和管理的数据得以安全存储和使用
- 在数据处理过程中，充分尊重并有效满足利益相关方的数据保护要求
- 遵循保护用户个人隐私、保障弱势群体数据权益、规避算法歧视等道德伦理规范

## (五) 加强数据安全合规管理，支持图书馆数据资源建设、管理和服务的可持续创新

数据安全  
制度

数据管理  
规范

风险动态监控与  
应急响应机制

数据馆员  
培养

技术措施

谢 谢

