

# 推动多源数据融合 释放联盟数据价值

北京大学图书馆项目管理中心 冯英

2024.5.16



# 目录

1. 图书馆及图书馆联盟数据情况
2. 图书馆联盟数据融合的思考

# PART 01

---

## 图书馆及图书馆联盟数据情况

---

# 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**
- 图书馆数据资源是图书馆在资源建设与业务管理过程中选择、采集、组织、保存、管理的**各类数据资源**，包括馆藏资源数据、业务数据、读者数据、管理数据等
- 新技术环境下，数据资源不断拓展，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、用户活动等数据将极大丰富<sup>1</sup>
- 信息（数据和知识等）是构成交融式数字图书馆<sup>2</sup>的三大要素之一

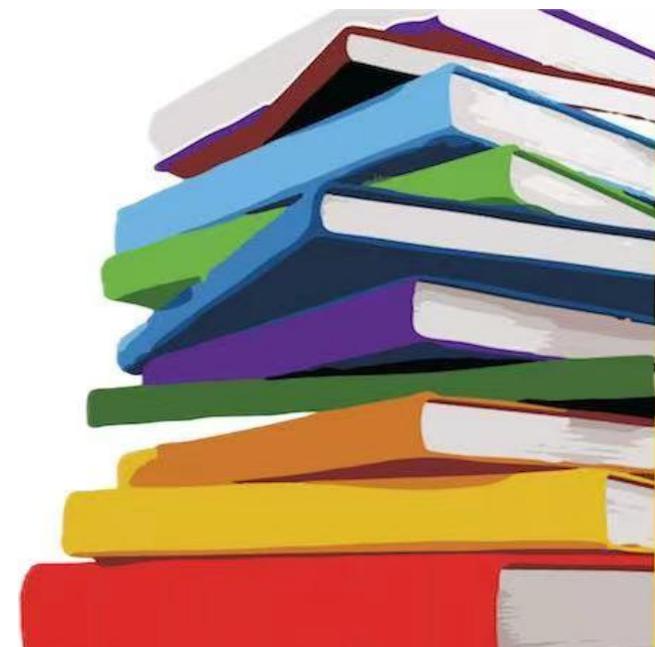
1. 申晓娟 《国家图书馆数据资源管理的实践与思考》

2. 陈建龙. 新时代大学图书馆现代化论纲[M]. 北京大学出版社, 2023

# 图书馆数据资源

- 文献资源数据：图书馆自建或购买的各类文献资源及相关数据，是图书馆数据中最重要的组成部分
  - 元数据：实体文献书目数据、数字资源元数据
  - 对象数据：自建或购买的全文资源、音视频、封面、图片等

**主要支持资源描述、发现与内容展示**



# 图书馆数据资源

- 业务运行数据：图书馆的业务、服务、空间、用户等图书馆运行过程中产生的数据
  - 业务数据：如采访数据、编目数据等
  - 服务数据：用户数据、流通数据、馆际互借与文献传递数据、读者到馆数据、网站访问数据、空间使用数据等
  - 管理数据：财务数据、资产使用与处置数据等

**对图书馆业务与管理进行监测与评估，支撑数据驱动的管理决策与业务优化**

## 1

### 实现资源共享

- 整合联盟内成员馆的资源，实现包括文献、人力和技术等资源的共享
- 利用联盟机制，联合采购与建设文献资源，降低建设成本
- 弥补自身馆藏资源的不足，获取其他图书馆的优质资源

## 2

### 提高服务水平

- 通过协作与合作，共享信息、技术与服务经验
- 实现区域或学科领域内的联合服务，丰富服务内容，提升服务质量

## 3

### 促进合作发展

- 通过联盟，图书馆可以克服资源和经费的局限性，实现更大的服务范围和服务深度
- 为图书馆提供一个交流和合作的平台，促进图书馆事业的发展

# 图书馆联盟数据资源

## 图书馆联盟的数据资源更多是.....

### 文献资源元数据

成员馆馆藏资源  
专题/领域文献  
开放获取资源

### 对象数据

自建或购买的全文资源

### 服务数据

用户数据、网站访问数据、馆际互借与文献传递数据等

### 公共数据

规范数据、知识库 (KB) 等

# 目前图书馆联盟提供的服务

## 资源与数据共享

联机合作编目、数据质量提升、采编一体化

## 资源发现与获取

联合目录、馆际互借与文献传递

## 数字资源集团采购

集团采购、资源与服务评估

---

## 软件服务与技术支持

如CALIS软件服务、PASCAL的共享服务平台

## 数字资源长期保存

如NDPP数字资源长期保存

## 业务培训、标准规范支持等

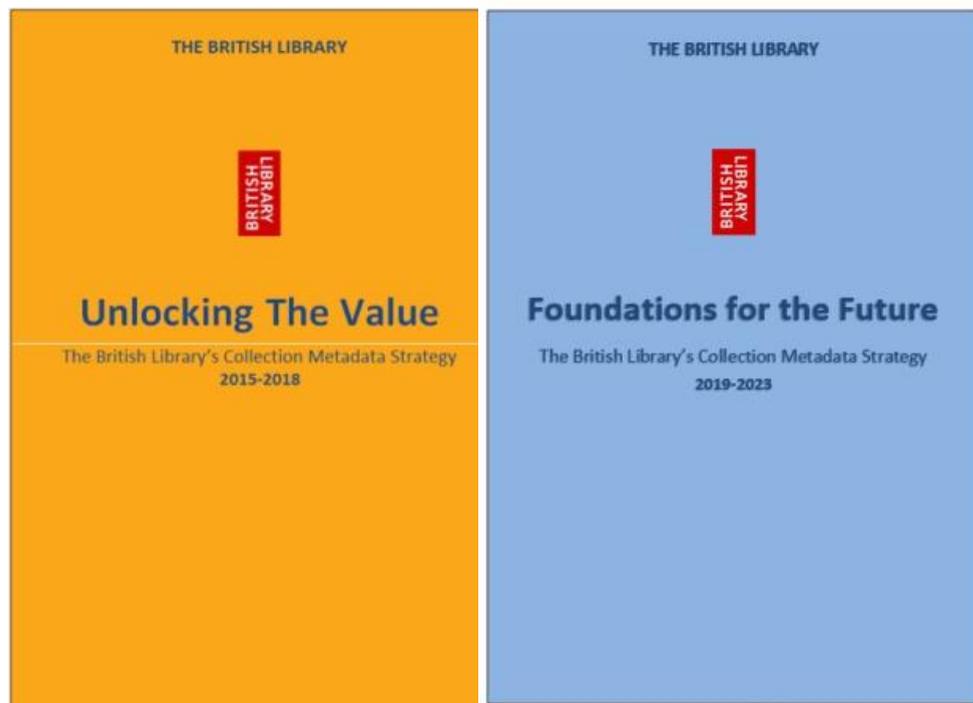
# 联合编目机构寻求更加广泛的合作

- 联合编目项目（Program for Cooperative Cataloging, PCC）是一个由从业人员、服务提供者和工具开发人员组成的多样化联盟，共同努力激发元数据的全部潜力促进和支撑知识组织与发现
- 使命：通过支持图书馆和其他文化遗产社区的元数据生产者，以及与具有共同目标的伙伴结成联盟，促进世界知识的发现和使用
- 战略方向之一是与更广泛的元数据利益相关者社区合作：扩大与图书馆软件和元数据提供者（包括供应商和开源社区）的协作，与GLAM（画廊、图书馆、档案馆和博物馆）元数据社区以及GLAM之外的其他元数据社区进行交流（出处：<https://www.loc.gov/aba/pcc/>）



美国PCC（联合编目项目）  
2023-2027  
战略方向

# 图书馆发布馆藏关联数据寻求合作



## 大英图书馆馆藏元数据策略 2019-2023

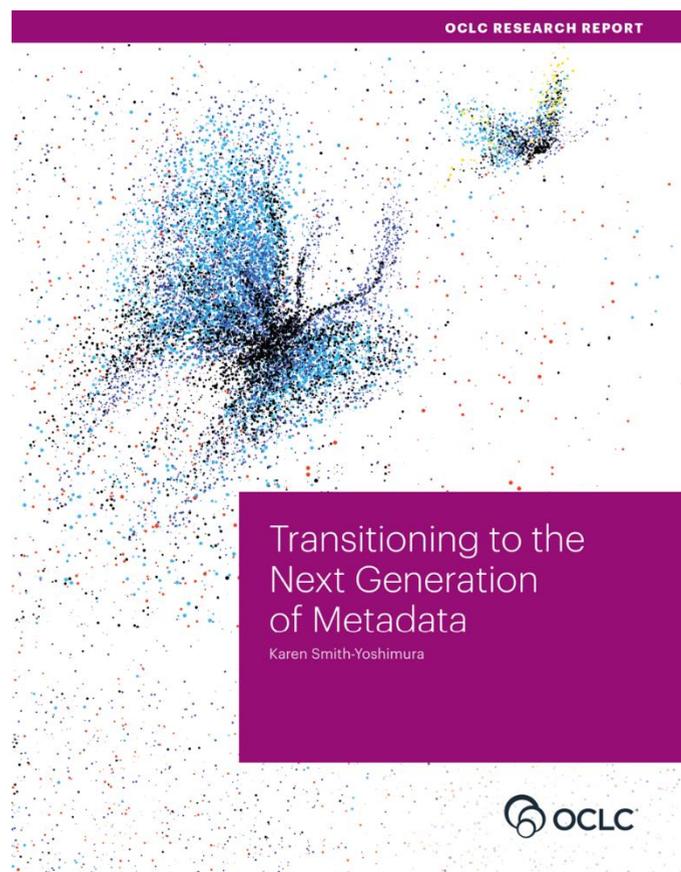
- 数据的有效利用是用户服务的基础
- 数据是一种关键资源
- 自动化工作流程，以最少的干预将出版商电子书元数据升级到图书馆标准

## 大英图书馆NBK长期路线图

- 帮助简化元数据市场（M计划）
- 印刷专著的协调与共享管理
- 为发布者创建的元数据提供更直接的传播渠道
- 结合图书馆和学习者分析
- 支持创建全面的内容层面的关联书目数据元

# 下一代元数据的关联与聚合

## OCLC向下一代元数据过渡



## 元数据是所有发现的基础

- 元数据与图书馆服务息息相关
- 图书馆的创新给元数据管理带来了挑战和压力
- 元数据是不断发展的

## 向关联数据和标识符过渡

- 扩大使用永久标识符
- 从“规范控制”走向“身份管理”
- 关联数据挑战

## 元数据即服务

- 超越传统图书馆目录
- 利用图书馆元数据提供新服务

# PART 02

---

## 图书馆联盟数据融合的思考

---

# 多源数据融合 增强联盟服务——助力馆藏建设

## 1

### 提供馆藏对比分析，指导本馆资源建设工作

- 支持按学科进行与对标图书馆的馆藏数据对比，帮助图书馆了解本馆资源建设的情况
- 需获得对标图书馆的授权允许

持续采集与整合图书馆馆藏资源

建立科学的数据采集与汇交机制

实现数据的自动汇聚与高质量处理

## 2

### 为馆藏质量分析与评估提供数据支撑

- 研究图书馆馆藏质量评价方法及指标体系
- 完善与细化数据处理，推进效用导向的馆藏质量评估

在馆藏质量评价的理论指导下，建设评价基准书目元数据  
进一步提升资源数据质量

## 多源数据融合 增强联盟服务——丰富数据服务

### 3

#### 丰富数据内容，对外提供更多的数据服务

- 通过数据标引、数据关联、数据挖掘以及外部数据融合，探索提供深度的数据服务
- 采集和整合开放存取资源，进行统一管理与维护
- 合作建设公共数据资源，如规范数据、资源百科、增强型书目数据、分类映射等，对外提供开放的数据服务

对数据进行深度处理、智能组织

数据服务平台化，通过数据重用，释放数据价值

## 多源数据融合 增强联盟服务——提升图书馆服务

### 4

#### 进行资源画像，对个性化、精准化的服务提供支持

- 根据文献资源的收藏情况，关联联盟内的资源检索、请求与借阅、推荐等使用情况进行资源画像

### 5

#### 推动数据资源的本地化保存与服务

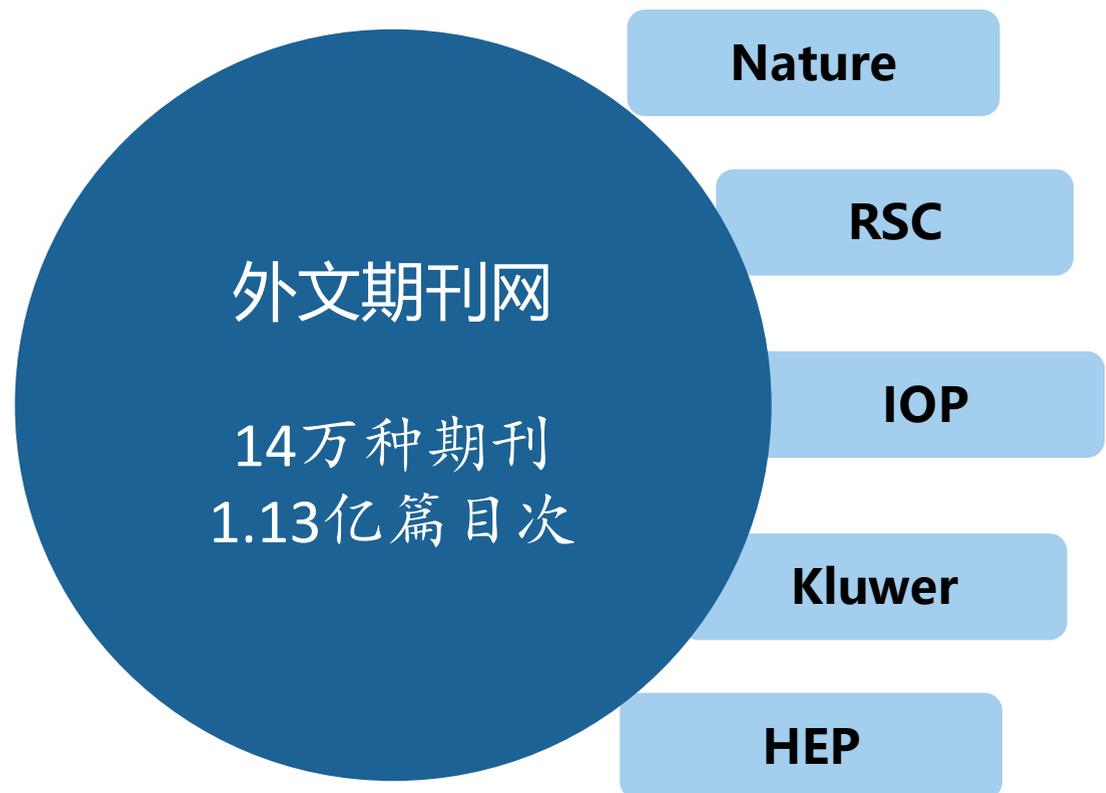
- 帮助图书馆与数据库商通过协商，综合运用长期保存、本地保存、镜像保存等形式，为成员馆设立国内本地化站点，提高数据容灾能力
- 根据用户需求进行本地化的组织管理，提供数据服务

## 6

## 拓展更大范围内的数据合作

- 与行业上游合作，基于资源订购与使用数据提供分析数据，帮助出版社更了解用户需求
- 与博物馆、档案馆等合作，实现元数据的拓展

## CALIS数据的融合案例——外文期刊



### ● CALIS外文期刊网（CCC）与本地库的数据融合

- CCC无全文，通过资源调度系统提供全文链接或文献传递服务；CALIS本地库有目次数据与全文
- CCC对CALIS成员馆提供服务，CALIS本地库对DRAA集团采购相关机构提供本地服务
- 数据融合后，前端访问依旧严格遵照与数据库商的服务约定提供

## CALIS与CASHL数据的整合

- 对CALIS、CASHL两大平台的应用和技术体系进行统一规划，构建统一的应用与技术体系，支撑统一的数据管理和服务管理体系，满足项目各自的资源和服务需求，实现两大平台的统一开发、运维、管理和服务，实现资源共享和最大化利用



## CALIS与CASHL数据的整合

- CALIS与CASHL数据在服务上相对独立
- 主要对CALIS和CASHL馆藏印本资源进行整合，以支持馆藏资源的分析与评估



北京大学图书馆  
PEKING UNIVERSITY LIBRARY



CALIS