



北京大學圖書館
PEKING UNIVERSITY LIBRARY

图书馆的数据生态建设

— 以北京大学图书馆为例

童云海

2021年4月23日



- **数据生态建设的缘起**
- 数据生态建设的概念
- 数据生态建设的架构
- 数据生态建设的行动

开展数据生态建设是落实《2035年愿景与2019-2022年行动纲领》的重要举措

五大使命

- 信息资源保障
- 信息服务创新
- 信息文化培育
- 人类文明弘扬
- 行业发展推进

四大行动

- 固本应变行动
- 服务转型行动
- 综合改革行动
- 标杆引领行动

数据生态建设是图书馆服务转型、综合改革的客观要求



面临挑战

- 资源结构和形式的变化
- 客户结构和需求的变化
- 细分领域竞争对手出现
- 合作伙伴不断更迭
-



应对措施

- 馆内机构重组
- 业务流程重组
- 人员发生流动
- 提供创新服务
-



如何应对创新服务

带来的业务快速重组和变化？



- 数据模型的改变要比业务模型的改变缓慢的多
- 牢牢抓住数据模型，开展数据生态建设
 - 适应快速重组和变化的有力支撑
 - 创新服务的重要驱动力

数据生态建设是提高对数据及其流动的认知，提升信息化水平的客观要求

北大图书馆的信息化建设是一个复杂巨系统

馆藏丰富、类型各异

电子、纸质；文本、多媒体；图书、期刊；中文、外文；普通、古籍和特藏；数字化加工的自有资源；……

总分馆体系

本馆+医学馆+44个院系分馆+...+马克思楼、昌平新校区

面向全国服务的CALIS+CASHL

独立机构的沙特馆

众多支持馆内日常运行的合作伙伴

馆内目前正在运行的系统158个



- 容易导致数据不清、流向不明
- 容易导致信息孤岛频现
- 容易导致数据不一致、无关联
-

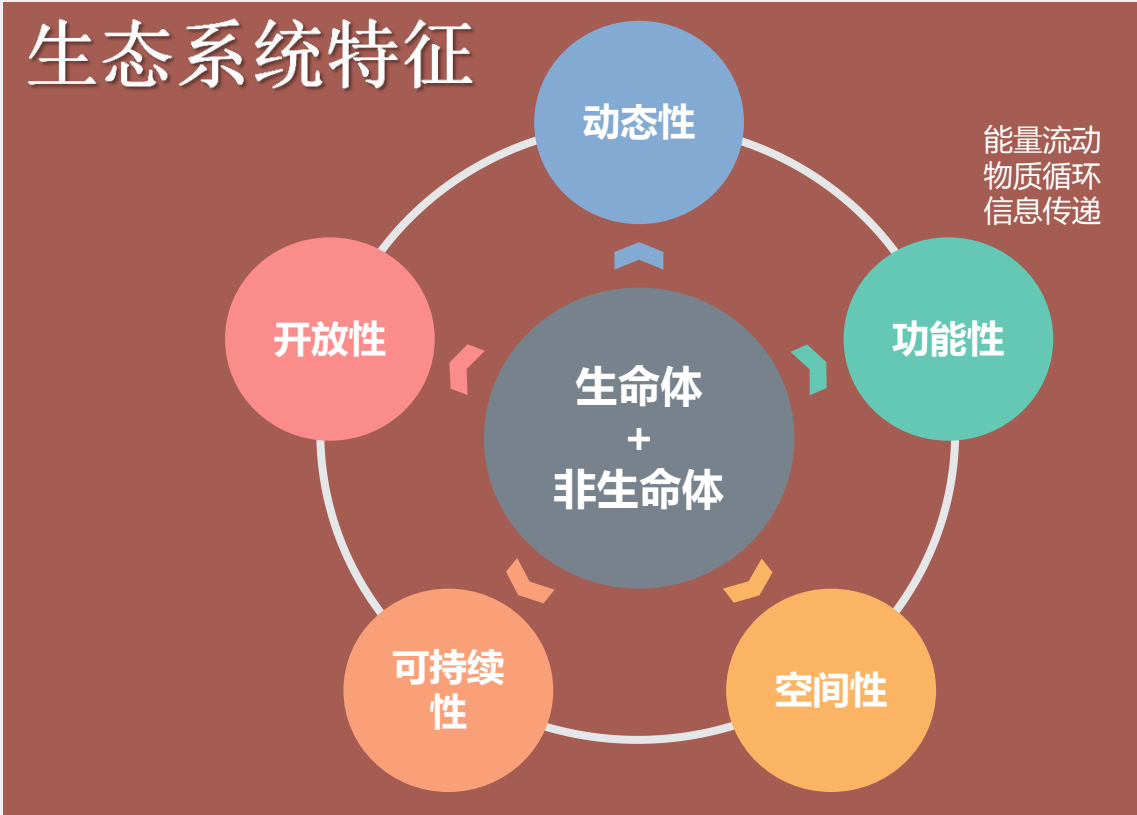
讨论提纲



- 数据生态建设的缘起
- 数据生态建设的概念
- 数据生态建设的架构
- 数据生态建设的行动

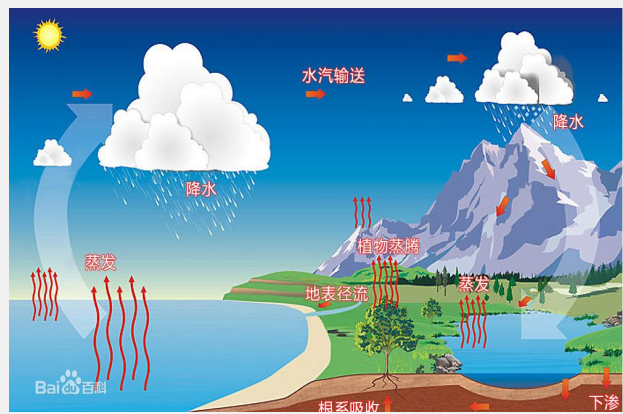
生态系统

生态系统是指在自然界的一定的空间内，生物与环境构成的统一整体。在这个统一整体中，生物与环境之间相互影响、相互制约，并在一定时期内处于相对稳定的动态平衡状态。



数据生态系统

数据生态系统，在一系列标准的支撑下，由人和具有可扩展性的硬件基础设施、数据设备、软件（含服务）、数据构成的统一整体。可系统地实现、控制和评估机构信息的采集、存储、加工、处理、流通、服务。是机构摸清家底、规范业务、提高业务运营效率的必要前提；是机构应对变化、数据驱动创新服务、提升决策能力的重要基础；是推进机构数据治理体系和治理能力现代化的重要抓手。



颠覆式创新：

被动的、自然选择式基因遗传，转型为自主选择式基因重组

互联网+：

全面感知用户、数据与产品服务结合

强调客户体验，渠道整合，实时互动交流

数据驱动：

对数据和数据驱动的认知力

开放创新：

辩证对待数据生态系统；内部开放+外部开放

数据生态系统：为人-机-物融合系统提供条件



物与物相连、人与物的互联，智能融入万物，实现无缝连接，协同计算

形成人类社会、虚拟空间、自然空间、机器物理空间的联通互动、数字孪生、虚实交融的景象



围绕个人兴趣组织利用数据资源，进行生产、生活和创造，在终身教育、创新创业、社交休闲、购物娱乐等方面享受更加精准、舒适、优质服务

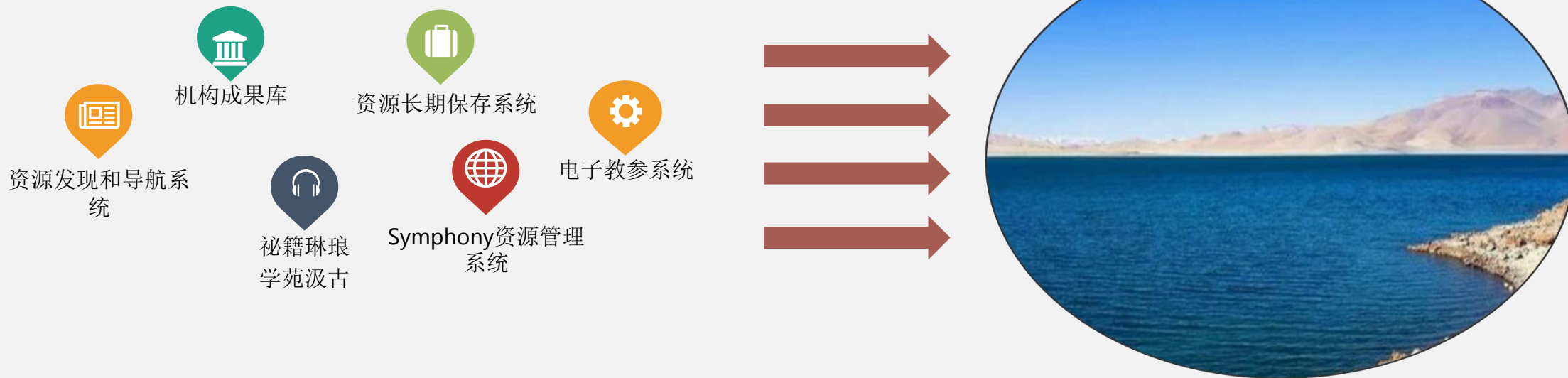
讨论提纲



- 数据生态建设的缘起
- 数据生态建设的概念
- **数据生态建设的架构**
- 数据生态建设的行动

图书馆数据生态建设的架构

数据要素：文献信息资源、空间资源、馆员、数据设备、服务产品、用户、……



核心业务平台 + OA平台

目标：保证用户并发环境下，数据的一致性、完整性、及时性

方法：数据集中化、业务协同化

大数据平台

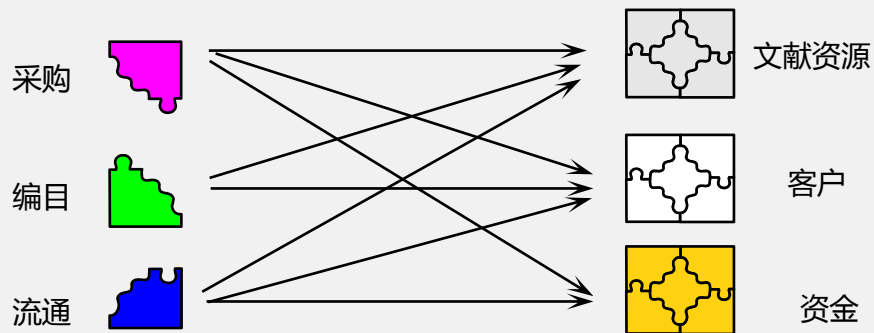
目标：保证数据的全面性（全量数据）

方法：数据集中化、业务协同化

两种不同的数据环境：操作型环境和分析型环境

操作型(数据库)

分析型(数据仓库、大数据)

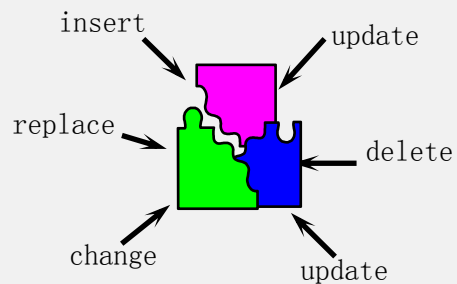


按照业务流程进行组织

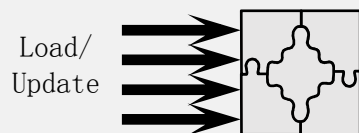
按照分析主题进行组织

操作型(数据库)

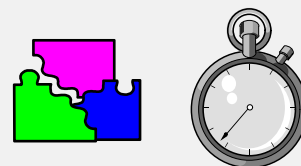
分析型(数据仓库、大数据)



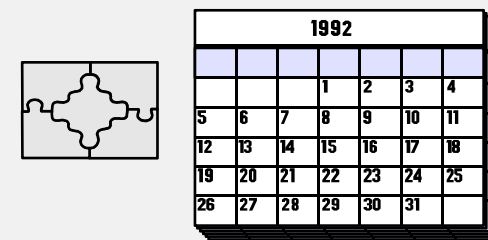
实时更新



定期加载
在某个时间点保持不变



主要关心当前数据



通常关心历史数据

两种不同的数据环境：数据特性的差别

原始数据/操作型数据	导出数据/分析型数据
面向应用	面向主题
详细的	综合的，或提炼的
在访问瞬间是准确的	代表过去的数据，快照
是为日常工作服务	为管理者服务
可更新	不更新
重复运行	启发式运行
处理需求预先可知	处理需求事先不知道
非冗余性	总是存在冗余
对性能要求高	对性能要求宽松
一次访问一个单元	一次访问一个集合
静态结构：可变的内容	结构灵活
访问频繁	访问很少或不多

讨论提纲



- 数据生态建设的缘起
- 数据生态建设的概念
- 数据生态建设的架构
- 数据生态建设的行动

行动：推进数据治理体系和治理能力的现代化

- 对接国际和国家标准，结合我馆实际，建立系列数据定义和描述标准
- 建立数据质量评估标准
- 构建数据安全防护体系，增强版权管理意识，保护用户个人隐私

■ 夯实根基，固本应变

- 梳理系统，数据确权
- 填补空白，补齐短板
- 数字加工，有序推进
- 心中有数，摸清家底
- 开放融合，互融互通

数据驱动

- 填补缺失服务
- 提升现有业务
- 创新新型服务

■ 感知用户，增强粘性

- 倾听数字原生代的需求
- 增强用户的参与度
- 留住用户的轨迹
- 保护用户隐私
- 发展长期伙伴关系

具体行动之一：基于FOLIO的“双流通”和“流通+”

- 现役图书馆集成系统（ILS）集成采访、编目、流通、检索等模块，为读者提供稳定的纸本馆藏检索、预约等服务，但存在一定的局限性，且对新功能的扩展支持上的局限性较高
- 问题驱动和需求驱动相结合，追求新发展

- 基于FOLIO的基础平台，结合现役图书馆集成系统Symphony实现“双流通”模式
 - 在保障现役系统稳定服务的同时为新平台逐步尝试开展服务奠定基础
 - 重新打造用户管理系统，实现全馆用户的一体化管理
 - 台口馆员通过FOLIO的用户界面完成借、还等操作，系统后台进行处理完成两个系统的“双流通”

- 以微信小程序、APP为交互界面，开发移动服务“流通+”
 - 研发移动端“一站式”服务功能，实现无实体卡交互
 - 微信小程序
 - 北京大学APP

建设内容：从用户分类角度



用户



校园卡用户

校园卡用户扫码借书，提升用户体验，推广一站式



校外用户

- 无校园卡用户通过用户注册或提交申请，馆员审核完成用户的创建和管理
- 校外无校园卡用户扩展借阅功能，通过扫码完成借书



场景

- 用户通过移动端应用生成身份二维码
- 馆员或自动借书机扫码获取用户身份信息，完成借书过程

截止到2021年4月18日（试运行三个月）

- 现有用户共计72930位；图书共导入复本数据7,284,939条，题名数据2,748,082条；
- 完成17620册图书借出，12068册图书归还。四月份在系统中借还量约占总借还量的50%左右。

具体行动之二：构建图书馆数据服务平台

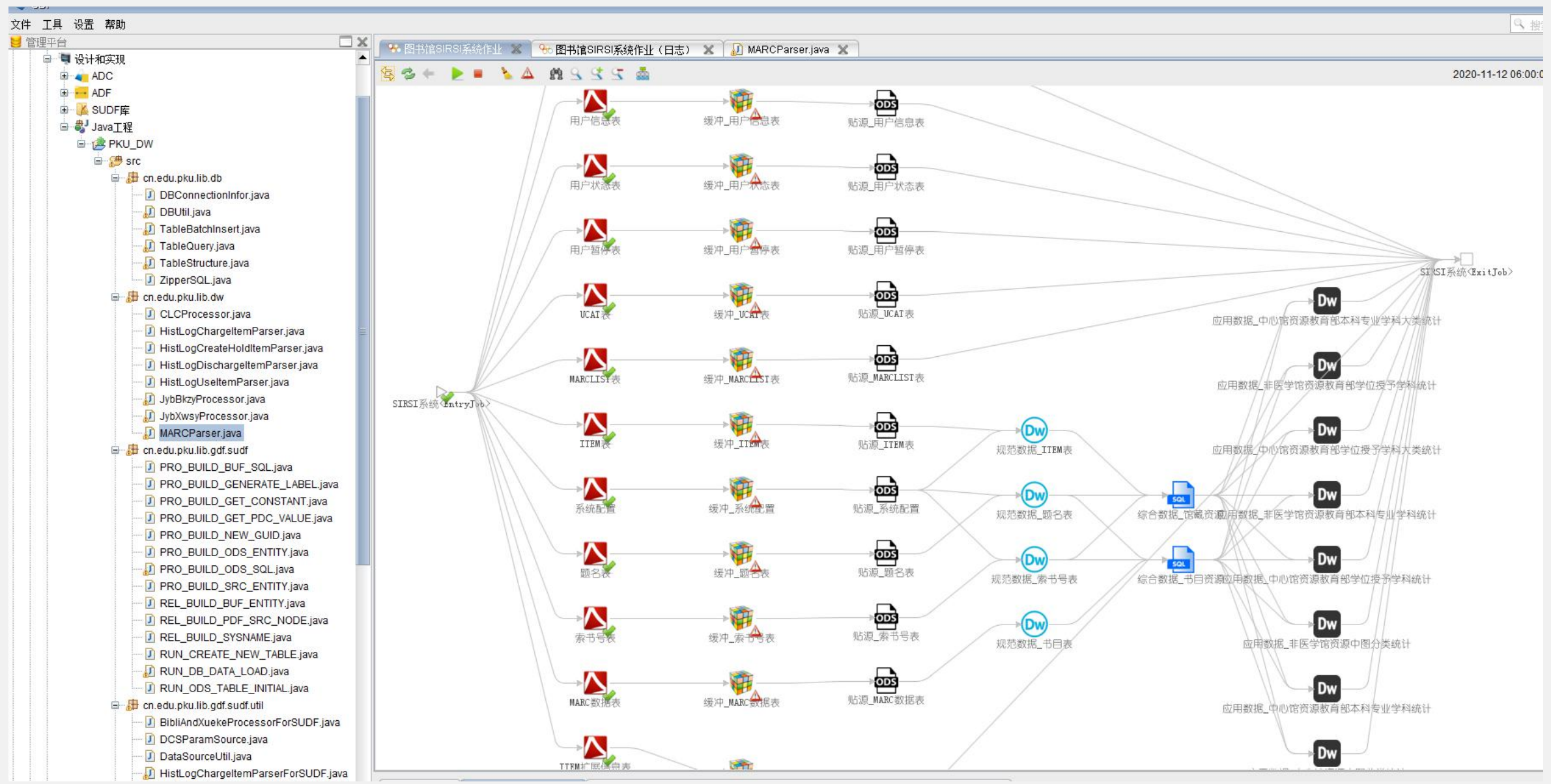
- 将数据标准为软规则，开展图书馆相关数据标准的制订和施行
- 通过对业务系统的数据抽取、转换和加载，面向主题组织数据，构建北京大学图书馆数据仓库（大数据的组成部分）
- 数据驱动和业务驱动相结合，构建分析模型，开展面向部门和馆领导的服务

■ 数据仓库模型的构建

- 文献资源
- 客户
- 服务产品
- 渠道、空间
- 资产
-

- 建立和完善文献资源、客户、服务等方面的评价指标体系
- 构建多维度、多层次的分析模型
- 可视化展现
-

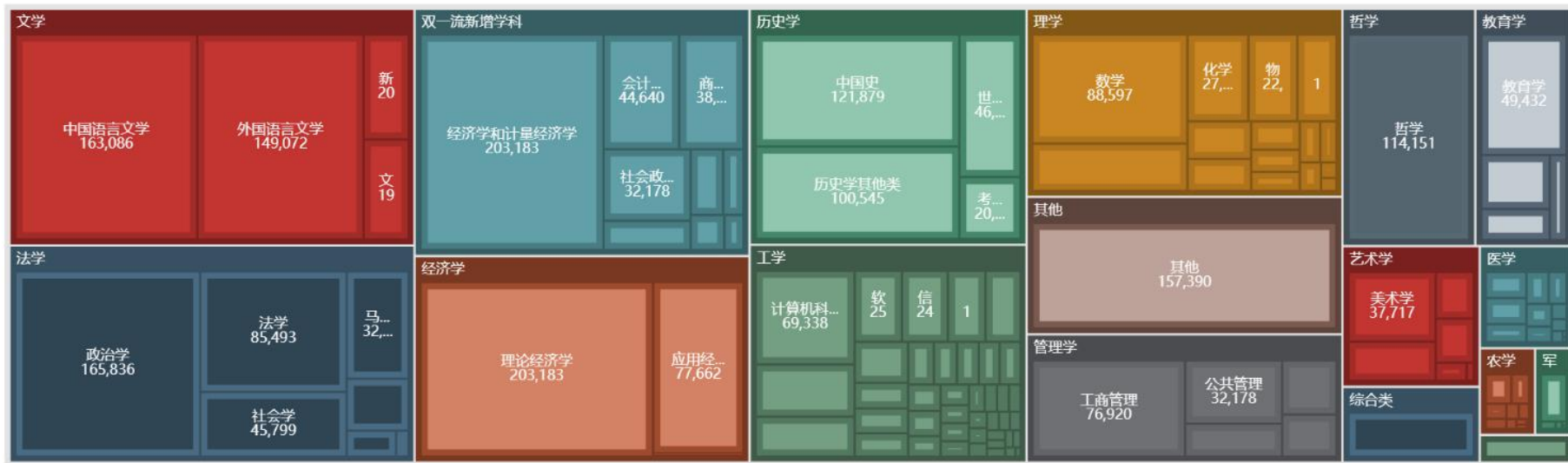
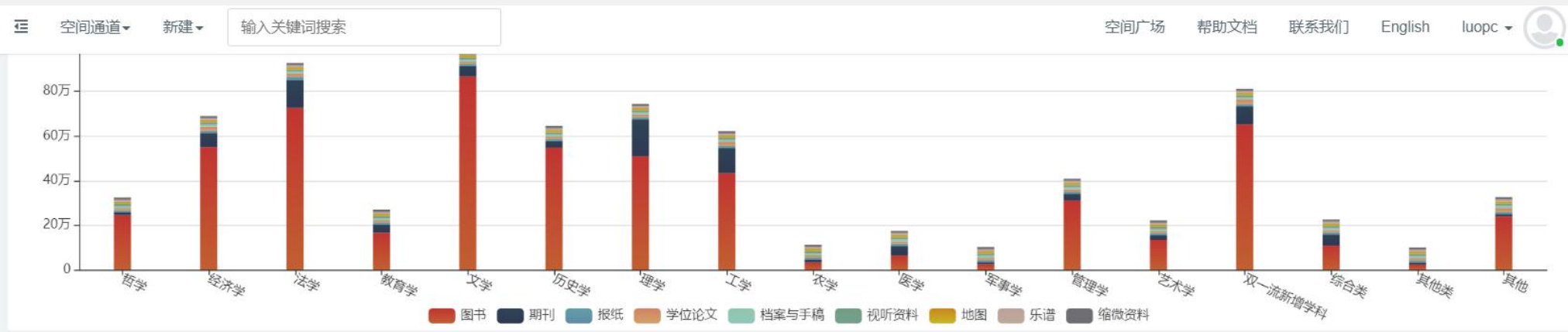
图书馆数据服务平台的数据抽取、转换和加载



图书馆数据服务平台的可视化展现

图书馆数据共享平台

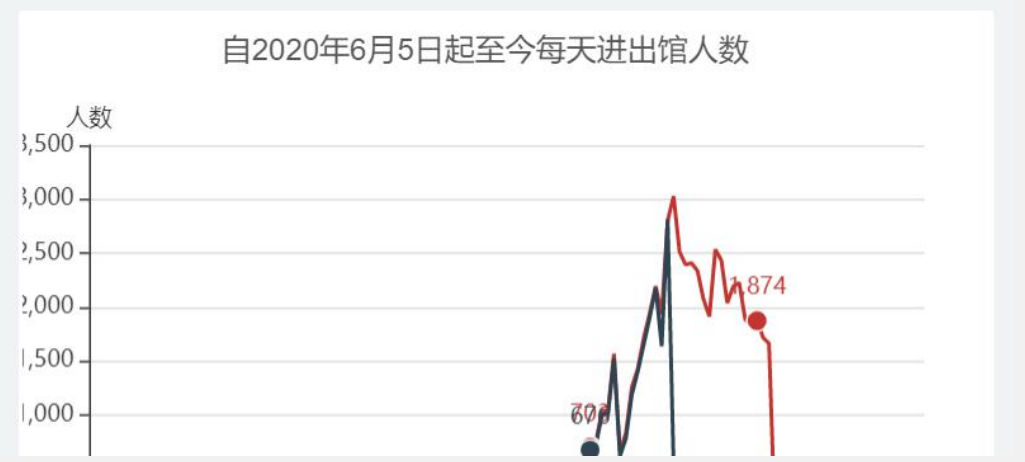
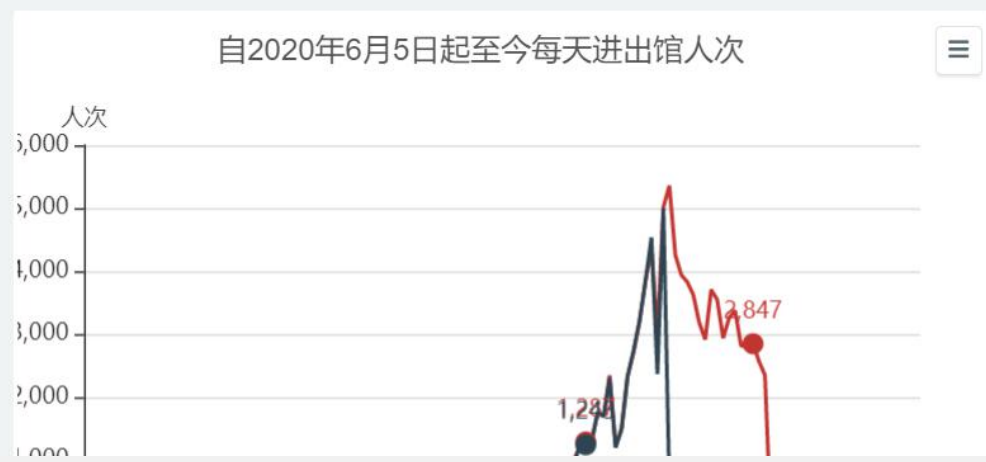
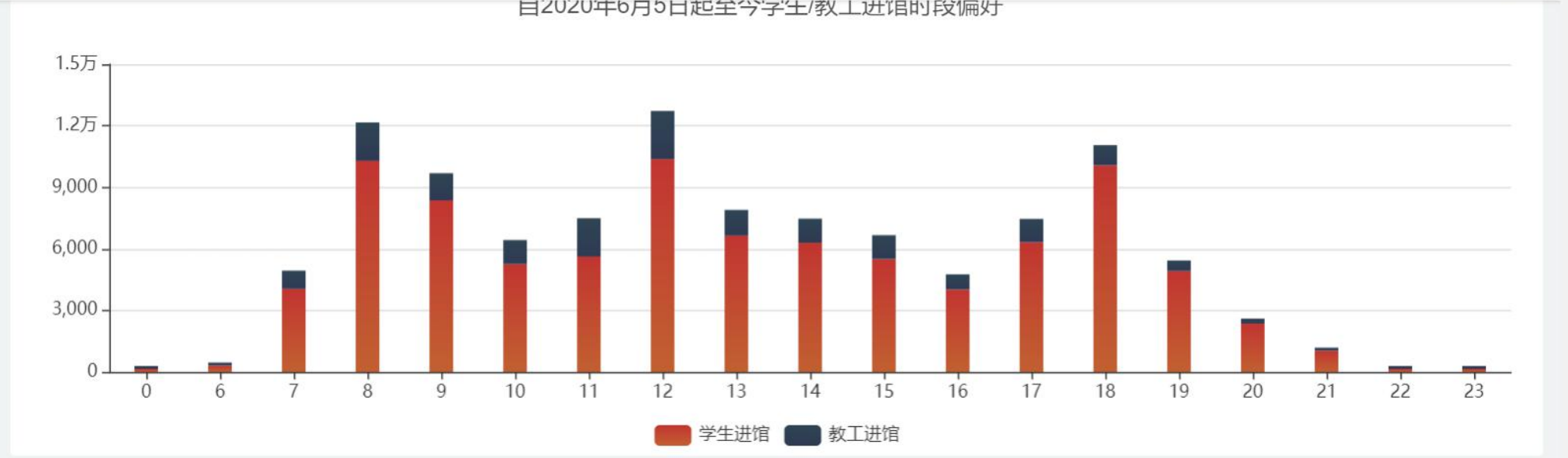
- 概况
- 数据大屏 新建
- 报表 新建
- 资源面板
- 中心馆纸质资源情况
- 按中图法学科分布
- 按教育部学位授予学科分布
- 按教育部本科专业目录分布
- 按排架号和馆藏址分布
- 中心馆资源类型分布及使用量情况
- 中心馆学科资源分布及使用情况
- 中心馆特藏资源
- 1. 资源总览
- 2. 图书



图书馆数据服务平台的可视化展现

- 图书馆数据共享平台**
- 纸质资源外借统计
- 进馆-人脸识别
- 进馆-人脸识别-2**
- 预约表统计
- 室内阅览表统计
- 财务面板 >
- 备份-待删除 >
- 报表管理
- 大屏管理
- 数据管理 >
- 数据源 HOT
- 数据模型 HOT
- 数据表预览

空间通道 ▾ 新建 ▾
空间广场 联系我们 English ▾ 





整体谋划、系统重塑、全面提升

让我们携手共同努力！

谢谢大家！