北京大学图书馆
PEKING UNIVERSITY LIBRARY

面向数据增值服务的北京大学图书馆数据管理与服务实践

北京大学图书馆 数据资源服务中心

20220512

北京大学图书馆

目录

北京大学图书馆
PEKING UNIVERSITY LIBRARY

1 背景

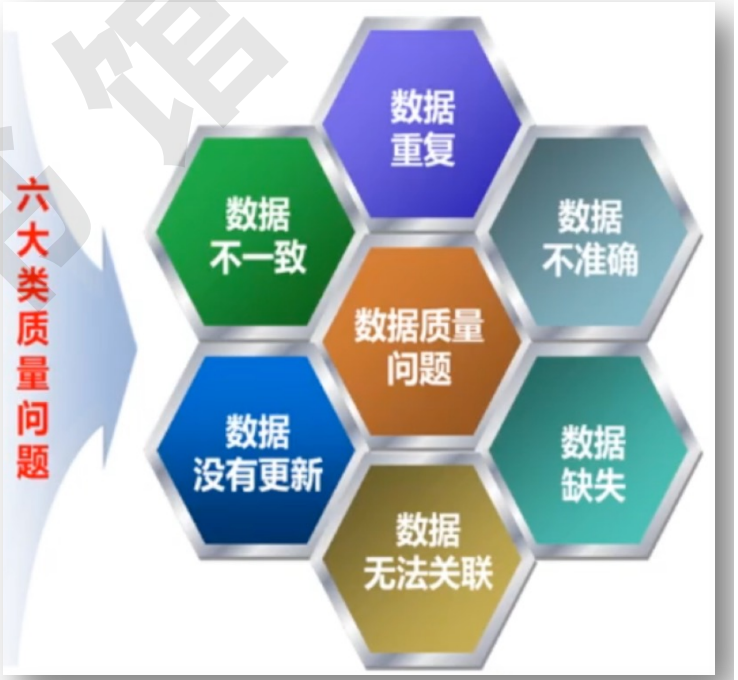
1. 政策的要求
2. 数据资源的不断拓展
3. 数据资源精准服务的需求及技术赋予的新可能性
4. 数据资产权益管理的紧迫性
5. 数据共享利用与网络信息安全的矛盾日益凸显

申晓娟.《国家图书馆数据资源管理的实践与思考》. 2022.5.12.

北京大学图书馆
PEKING UNIVERSITY LIBRARY
北京大学图书馆

1 背景

- 图书馆作为一个机构，抛开图书馆的身份，经历了一段时期的信息化建设后，其业务运行数据存在着各个组织机构普遍存在的问题：
  - 数据孤岛：数据分散在各个系统，没有进行汇总，更没有进行关联，都在各个系统负责人手上，有的甚至在公司手上无法拿到；
  - 数据采集问题：增长过快，种类复杂.....
  - 数据质量问题：
    - 数据重复问题；
    - 数据不一致，不准确；
    - 数据缺失；更新不及时；
    - .....
  - 数据使用：靠领导权威或关系亲疏找系统负责人拿数据，没有规范；
  - 数据安全与隐私存在隐患；
  - 数据不可知、不可控、不可取、不可连、不可信；



北京大学图书馆

1 背景

- 图书馆作为服务高校教学科研的信息中心，和其他组织机构相比又有其特殊性：
  - 文献资源数据是图书馆数据中最重要的组成部分
    - 图书馆的文献资源服务使命；
    - 半结构化，非结构化（文本），量非常大，增长迅猛。

- 图书馆数据：

数据类型	内容	结构特点	存储	增长情况	应用目的
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	结构化	存储在本地	逐年稳定增长	支撑数据驱动的管理决策和业务优化
文献资源数据	文献资源及其相关数据	半结构化、非结构化（文本）	极少量存储在本地	飞速增长	为用户提供深层次的、个性化、精准化、智能化的知识服务

北京大学图书馆

2 要解决的问题

北京大学图书馆
PEKING UNIVERSITY LIBRARY
北京大学图书馆

3 同行的诸多努力

- 图书馆：数据的整理、加工、管理和服务工作
  - 纸质资源的管理、电子资源的管理……
  - CALIS，CADAL等全国性的联盟对某一特定类型资源的全国性管理……
  - ……
- 智慧图书馆背景下
  - 国图、国科图、中国工程院……
  - 南京大学、上海交通大学、重庆大学……
- 图书馆数据管理与增值服务 VS 图书馆数据治理、图书馆数字化转型、图书馆数据资产管理、图书馆数据中台 ……

北京大学图书馆
PEKING UNIVERSITY LIBRARY

4 北大图书馆的实践

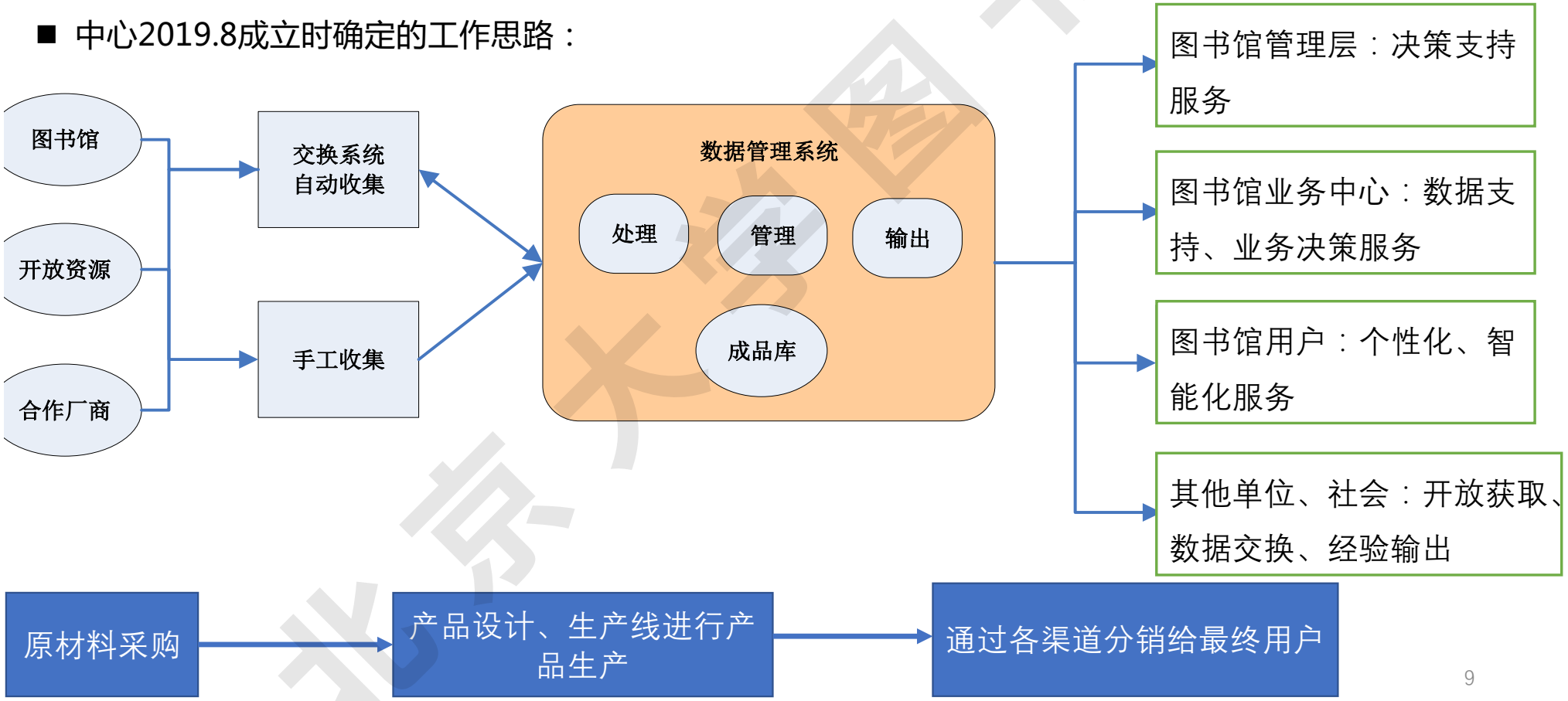
北京大学图书馆 PEKING UNIVERSITY LIBRARY

- 2019年北大图书馆内部机构改革，数据相关的工作汇集到一起：数据资源服务中心
- **中心主要职责：**
  - 全面加强可用**数据汇聚和资源化**工作，努力搭建数据仓储和交换、数字出版和开放获取、长期保存和自助服务等平台，建立健全**数据资源管理及其开放和网络服务体系**，为用户学习、教学、科研活动和信息文化培育提供有力的支撑及保障。
- **目前设置3个组**：
  - 数据开发组、数据质量控制组、数字加工组；
- **主管副馆长：童云海（数据专家，大数据理论知识+社会实际应用经验）**
  - 研究领域：数据挖掘和知识发现；数据安全保护；数字图书馆；
- **主任：张俊娥（图书馆数据专家，对书目数据、业务数据、期刊数据有深厚的理解）**
  - CALIS数据中心创始人，参与CALIS联机编目、e读等项目，主持CCC、DMS建设等项目。

北京大学图书馆

4 北大图书馆的实践

- 中心的工作内容：数据汇集和资源化，数据资源管理及其开放和网络服务体系；
- 中心2019.8成立时确定的工作思路：



4 北大图书馆的实践

北京大学图书馆 PEKING UNIVERSITY LIBRARY

- 实践工作：
  - 首先，进行图书馆数据资源体系梳理；
    - 对我馆数据资源进行全面的盘点，梳理出我馆数据资源目录体系；
  - 然后，建设图书馆数据管理与服务的数据基础设施；
    - 对数据进行深度处理、智能组织和科学化管理；
  - 最后，构建图书馆资源数据增值服务体系；
    - 通过提供精准、可信、细粒度的数据服务，探索数据驱动的图书馆管理决策和服务创新；

北京大学图书馆

4.1 梳理图书馆数据资源体系，打下坚实基础

- 数据盘点
  - 对图书馆的数据进行梳理，包括图书馆购买的数据、运行产生的数据以及图书馆工作需要的外部数据等，

全面深入地了解了图书馆数据的现状，充分认识到了图书馆数据中心建设任务的艰巨性，为中心接下来的数据工作绘制了一份蓝图。

  - 对每类数据的现状、需要进行的加工处理以及可能提供的数据服务、应用场景等进行了梳理，梳理格式如右图，以图书资源馆藏数据为例：

	数据属性		描述	备注
1.	数据含义		图书资源的在北京大学图书馆（包括中心馆、分馆、医学馆等）的馆藏情况。	
2.	原始数据	原始数据来源	图书馆 SIRSI 系统	
3.		原始数据获取方式	每天凌晨 5 点同步来自 l 的数据	
4.		原始数据格式	Oracle 数据库表	
5.		原始数据量	500 多万 2)。	
6.		原始数据采集后存放地址	系统后台数据库（Oracle）	
7.		原始数据采集的更新机制	每天凌晨同步前一天的数据	
8.	数据处理	需要对原始数据做的数据处理	需要对数据表中的若干字段的数字编码转换为可理解的文字；需要将馆藏数据与书目数据进行关联，以便进行分析。	尚未进行
9.		是否需要其他业务部门支持	否	
10.	成品数据	成品数据的主要字段	ID、馆藏图书馆、馆藏主位置、馆藏当前位置、创建时间、类型、文献类别 1、文献类别 2、题名、作者、语种、出版年、出版社、ISBN、分类号、关键词。	预估的
11.		关联的数据	图书的借还、室内阅览、预约等数据	
12.		成品数据格式	Oracle 数据库表	
13.		成品数据量	500 多万(	少量馆藏数据缺失书目信息
14.		成品数据存放地址	DCS 系统后台数据库（Oracle）	
15.		成品数据更新机制	原始数据同步完之后，系统自动进行数据处理	
16.	数据项创建时间		2020-05-29	
17.	数据项修改历史		2020-05-29 初次撰写 2020-06-08 修改字段描述	
18.	其他说明			

4.2 建设数据基础设施，全力提升数据质量

- 数据管理平台
  - 数据生产流水线、数据工作的支撑平台。

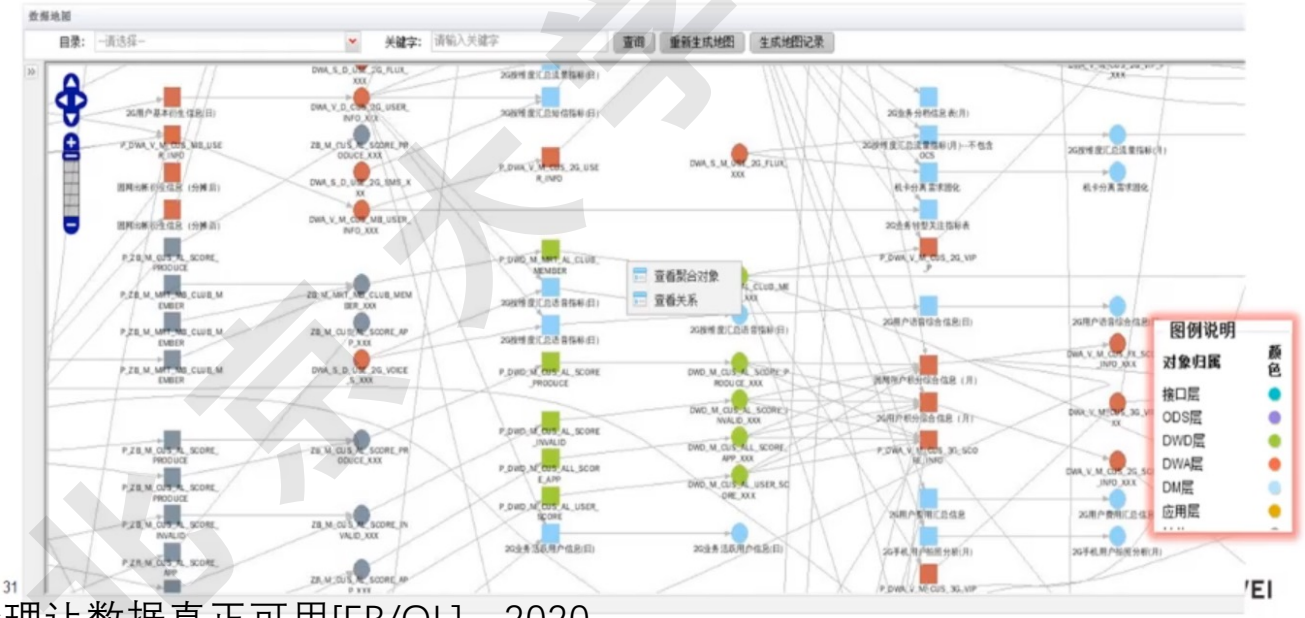
华为孙长森，数据治理让数据真正可用[EB/OL]，2020.

4.2 建设数据基础设施，全力提升数据质量

- 数据管理平台
  - 数据生产流水线、数据工作的支撑平台。

元数据管理——数据地图

数据地图是元数据信息的全景视图，描述所有元数据对象的血缘关系，所处层级覆盖范围由归集库->中心库->基础库->主题库。全面呈现了数据库中不同数据层级之间的关联关系，数据分类归属，数据关系一目了然。

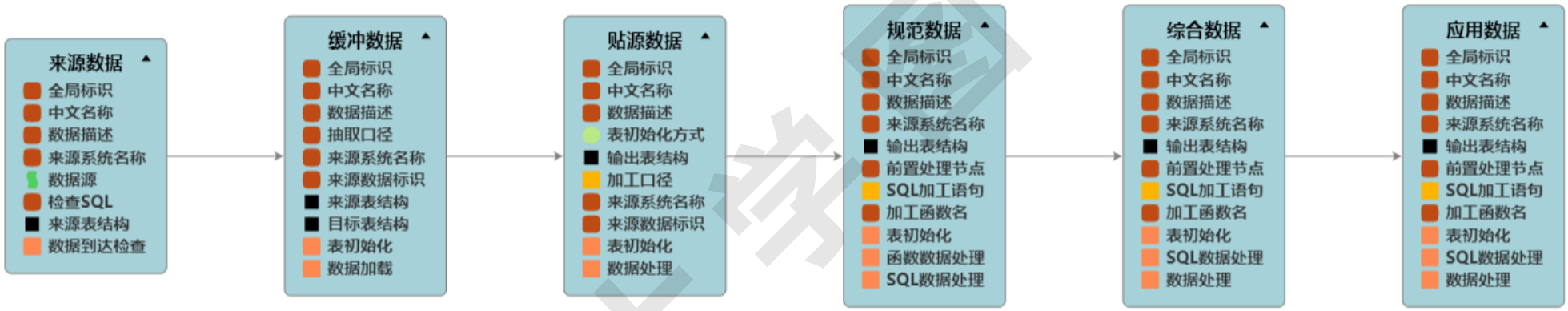


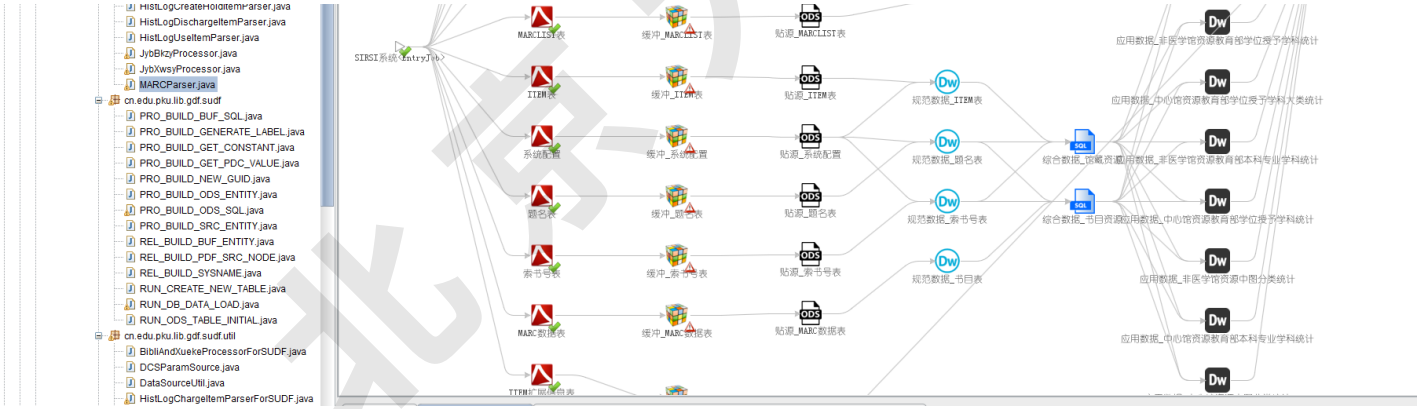
华为孙长森，数据治理让数据真正可用[EB/OL]，2020.

4.2 建设数据基础设施，全力提升数据质量

北京大學圖書館 PEKING UNIVERSITY LIBRARY

- 数据管理平台
  - 数据生产流水线、数据工作的支撑平台。





北京大学图书馆

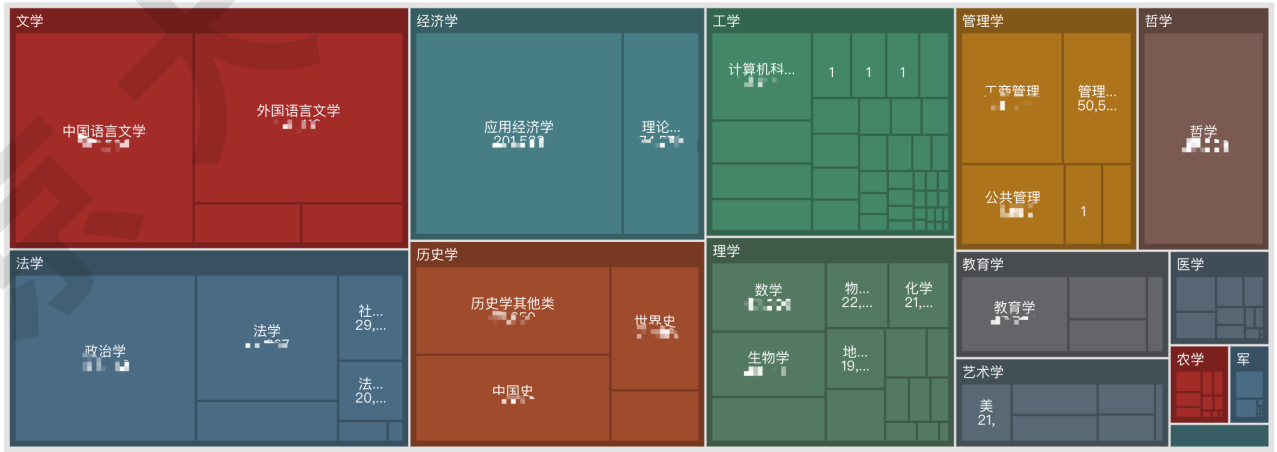
4.3 构建图书馆资源数据增值服务体系，释放数据价值

- 数据利用与服务
  - 构建面向四类服务对象的数据增值服务体系：
    - 管理层：决策支持服务；例如：管理决策数据需求、运行数据月报、运营效能评估。
    - 业务中心：数据支持、业务决策服务；例如：业务决策支持、业务优化支持、数据评估服务等。
    - 用户：个性化、智能化服务；
    - 社会：开放获取、数据交换、经验输出；
  - 数据增值服务的不断深化：
    - （1）丰富数据内容（根源）：通过数据标引、数据关联、数据挖掘、外部数据融合探索提供深度的、智能化的数据服务；
    - （2）拓展服务渠道（途径）：尝试多种数据服务渠道，探索更高速便捷提供数据服务的新方法、新途径；
    - （3）深挖用户需求（内容）：尝试从多维度进行数据建模和分析、探索如何为四类服务对象提供其真正需要的数据服务；

北京大学图书馆

4.3 构建图书馆资源数据增值服务体系，释放数据价值

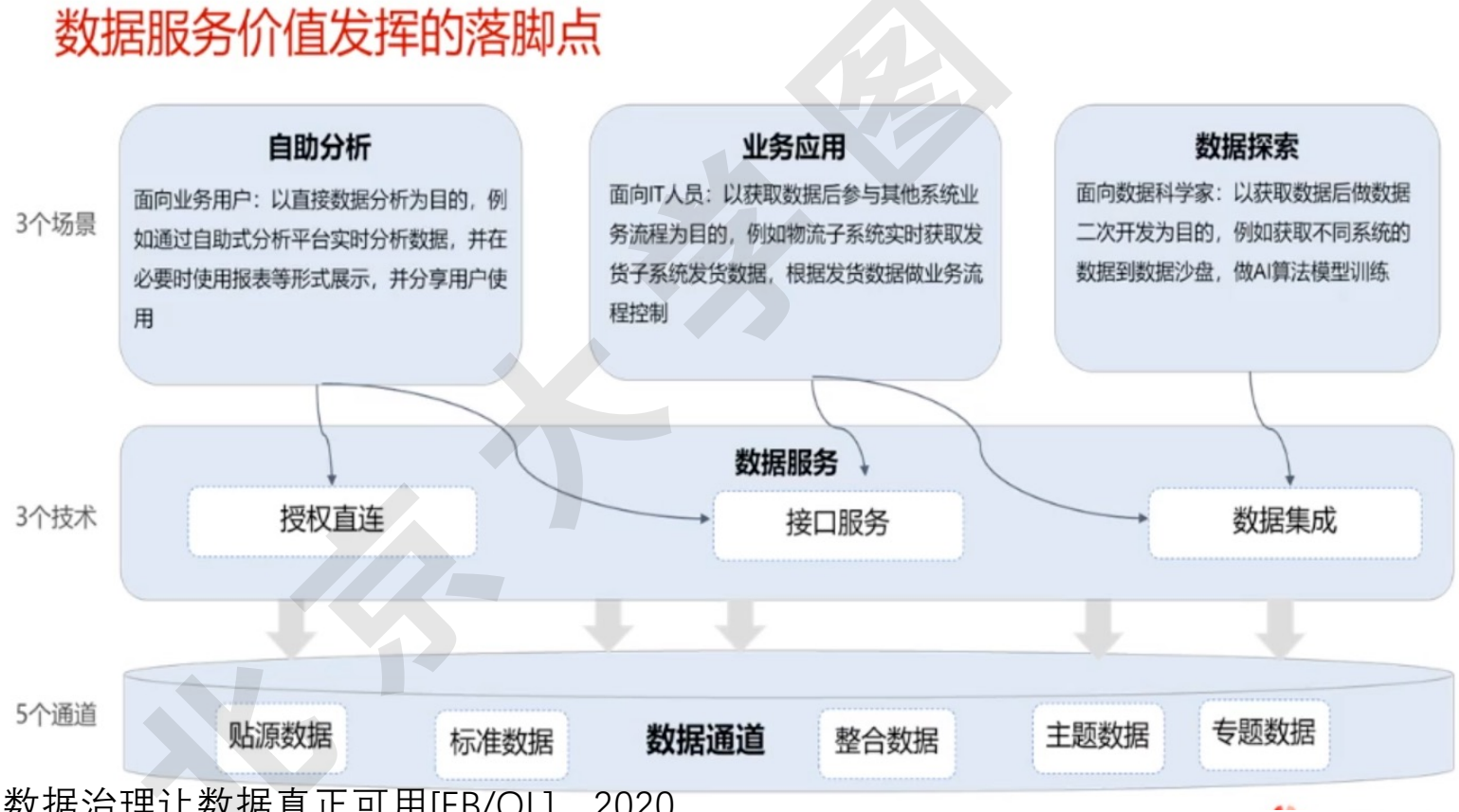
- 数据增值服务的不断深化：
  - （1）丰富数据内容：通过数据标引、数据关联、数据挖掘、外部数据融合探索提供深度的、智能化的数据服务；以纸质资源为例：
    - 通过将图书馆馆藏资源所使用的分类体系中图法、皮号裘号、杜威号和用户相对更熟悉的教育部学位授予学科、学校学科建设更需要的双一流学科进行映射，形成了图书馆馆藏资源数据的多重知识组织体系；同时，对原有馆藏书目数据进行深度挖掘和处理，形成了一套规范化的馆藏书目数据库。
    - 按照新的资源组织方法把分散在各个馆藏址的纸质资源进行了重新组织，为我馆东馆、西馆馆藏资源布局提供决策参考。



4.3 构建图书馆资源数据增值服务体系，释放数据价值

北京大学图书馆 PEKING UNIVERSITY LIBRARY

- 数据增值服务的不断深化：
  - （2）拓展服务渠道：尝试多种数据服务渠道，探索更高速便捷提供数据服务的新方法、新途径；



华为孙长森，数据治理让数据真正可用[EB/OL]，2020.

4.3 构建图书馆资源数据增值服务体系，释放数据价值

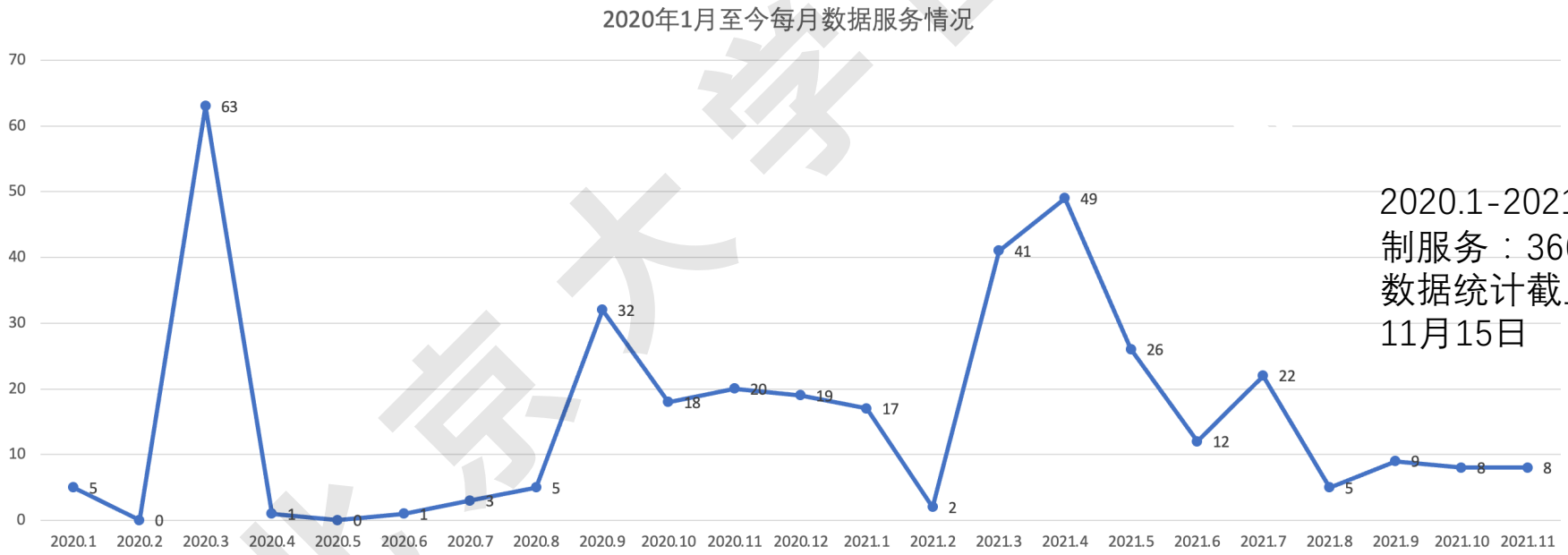
北京大學圖書館 PEKING UNIVERSITY LIBRARY

- 数据增值服务的不断深化：
  - （2）拓展服务渠道：尝试多种数据服务渠道，探索更高速便捷提供数据服务的新方法、新途径；
    - 1）主动服务
      - 图书馆运行数据月报；2021年4月开始，每月一期；
      - 部门面板（文献、知识、特藏），主题面板（资源、财务、用户、资源利用、进出馆）；

4.3 构建图书馆资源数据增值服务体系，释放数据价值

北京大学图书馆 PEKING UNIVERSITY LIBRARY

- 数据增值服务的不断深化：
  - （2）拓展服务渠道：尝试多种数据服务渠道，探索更高速便捷提供数据服务的新方法、新途径；
    - 2）定制服务
      - 根据各业务中心的需求定制化提供数据服务



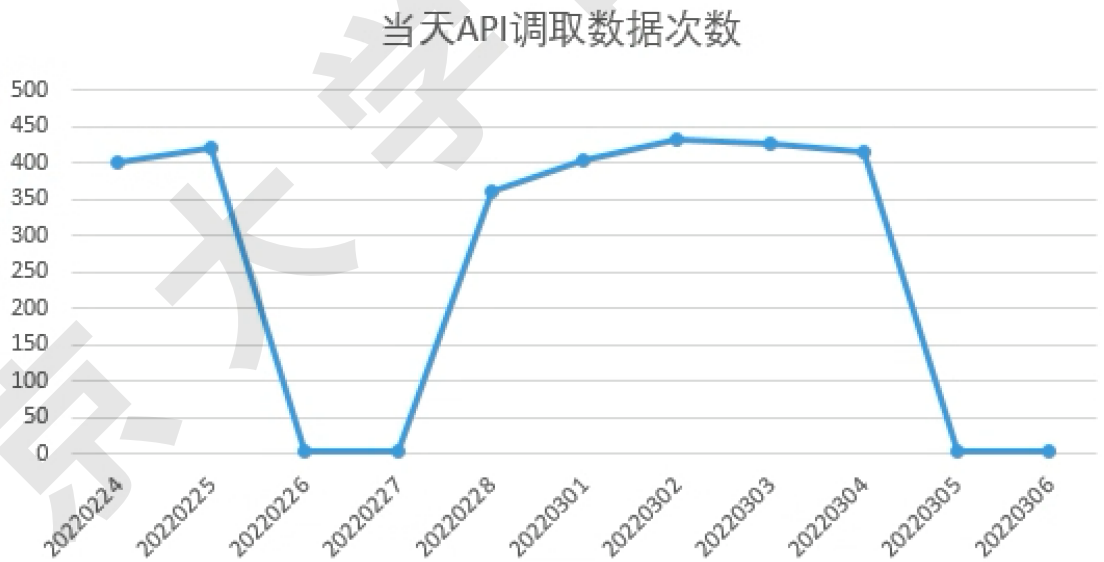
2020.1-2021.11.15 定制服务：366次
数据统计截止2021年11月15日

4.3 构建图书馆资源数据增值服务体系，释放数据价值

北京大学图书馆 PEKING UNIVERSITY LIBRARY

- 数据增值服务的不断深化：
  - （2）拓展服务渠道：尝试多种数据服务渠道，探索更高速便捷提供数据服务的新方法、新途径；
    - 3）API服务
      - 数据源自应用系统，经数据中心规范、关联、挖掘后，服务于应用系统；

时间范围：
2022-2-24至2022-3-6



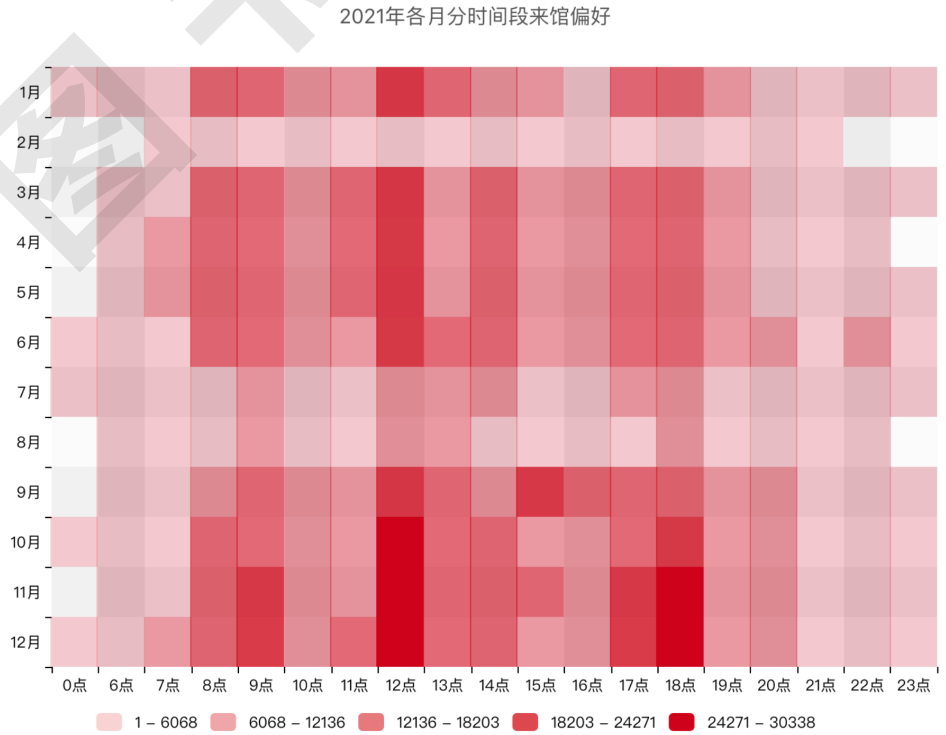
北京大学图书馆

4.3 构建图书馆资源数据增值服务体系，释放数据价值

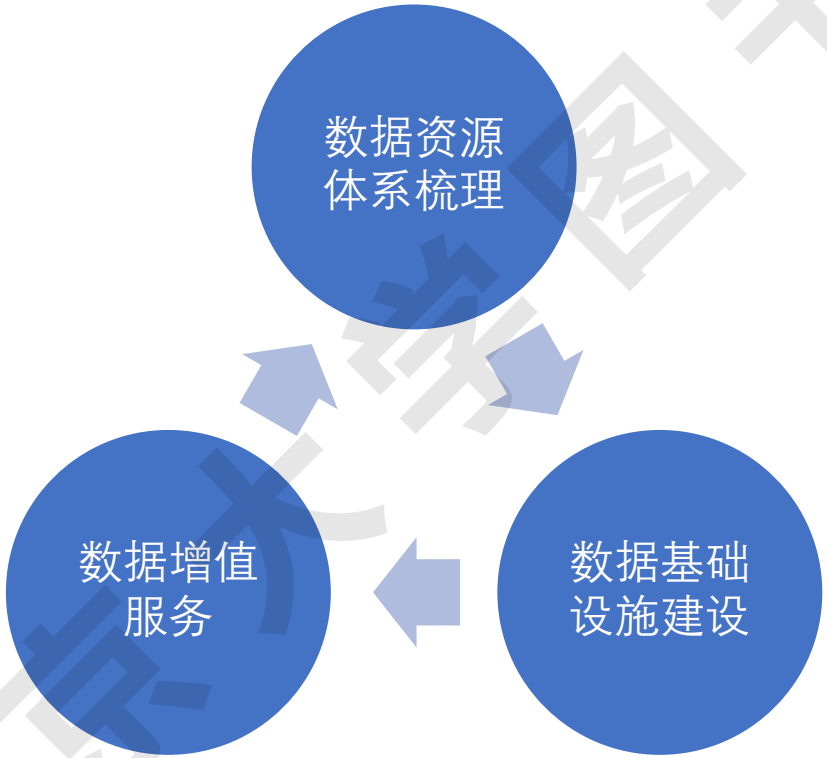
北京大学图书馆 PEKING UNIVERSITY LIBRARY

- 数据增值服务的不断深化：
  - （3）深挖用户需求：尝试从多维度进行数据建模和分析，探索如何为四类服务对象提供其真正需要的数据服务；
    - 不断改进服务形式，例如：每次和服务对象的沟通都加深对服务对象需求的理解深度，尝试将对方的需求归纳总结并转化为固定需求，变被动提供数据为主动推送数据，随需随取；
    - 不断改进服务内容，例如：提供新的数据分析结果或数据呈现形式；基于对数据的深入理解+对需求的深入理解，构建数据核心指标，围绕指标展示统计分析数据；

（4）2021 年各月分时间段来馆偏好



4 北大图书馆的实践



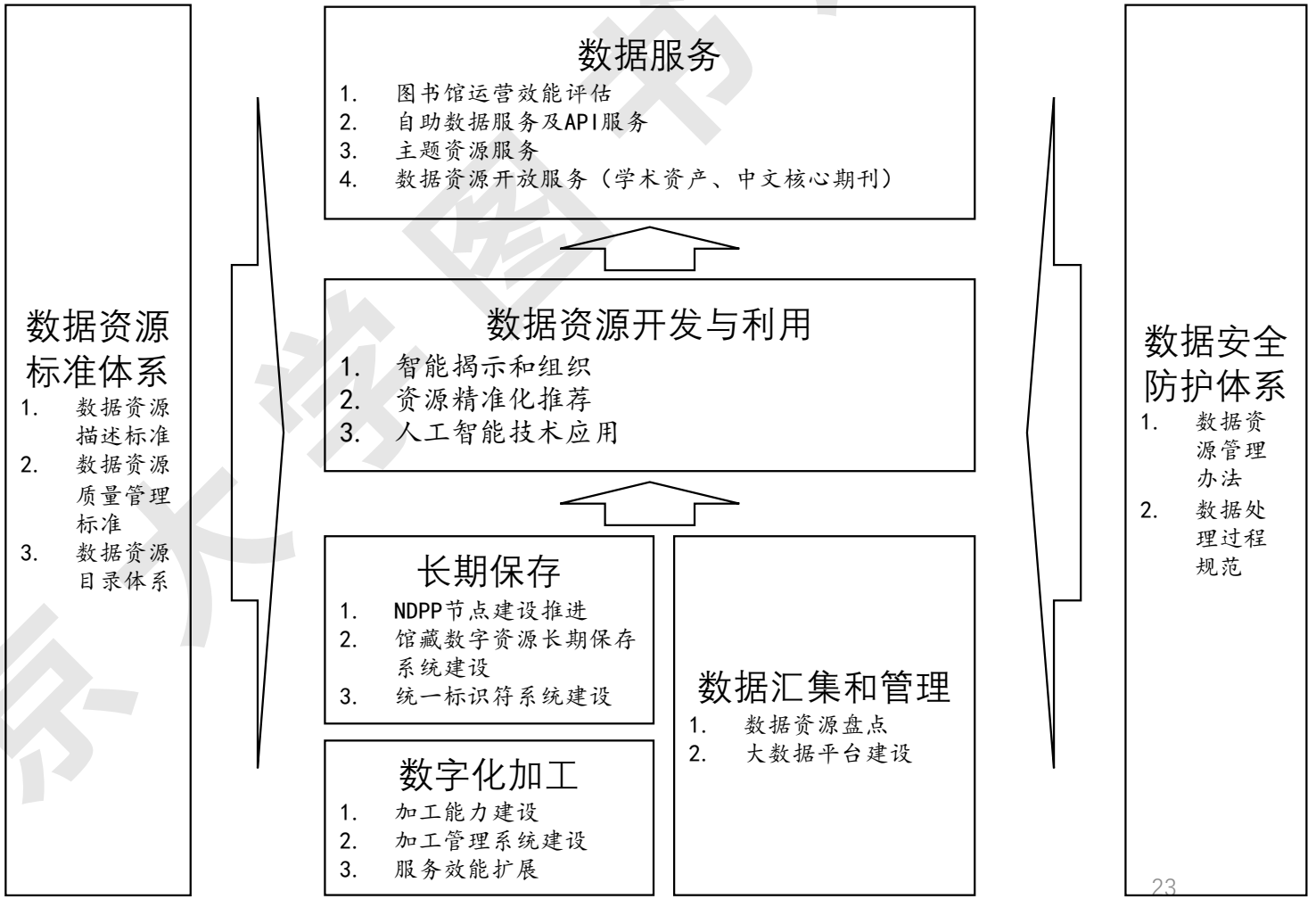
北京大学图书馆
PEKING UNIVERSITY LIBRARY

5 总结与展望

（1）作为高校图书馆，对图书馆数据的管理和服务进行了实践，形成了一套可行的思路、技术方案，可为同行提供参考借鉴。欢迎同行多交流。

（2）宝贵、海量的文献资源数据的梳理和管理需要同行通力合作、共建共享。@CALIS & DRAA

（3）图书馆数据管理和服务工作任重而道远。



请批评指正！

欢迎同行多交流：liud@lib.pku.edu.cn

北京大学图书馆