

# “数智时代高校图书馆用户数据治理体系研究”项目

## 结题材料目录

一、结题项目验收单

二、结题报告

三、支撑材料

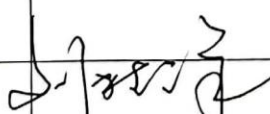
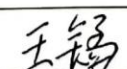


附件 1、高校图书馆用户行为数据治理体系研究报告

附件 2、2020 学年江苏大学图书馆用户行为记录数据集（发表于《图书馆杂志》，2022, 41(10)）

附件 3、用户数据视角下高校图书馆用户行为特征研究（发表于《图书馆学研究》，2023(02)）

## 项目结题验收单

专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	数智时代高校图书馆用户数据治理体系研究		
主持人	张晓阳	职务/职称	研究馆员
所在单位	江苏大学图书馆（加盖单位公章）		
专 家 意 见	<p>课题以高校图书馆用户数据为研究对象，以 WSR 系统分析方法和数据治理理论为基础，综合运用文献调研、网络调研、访谈和案例分析等研究方法，构建 WSR 数据治理新框架，以 A 大学图书馆为例，结合框架分析其用户行为数据治理的思路与方法，基于 Streamlit 设计治理平台，为实现用户精准服务提供路径和实践参考。</p> <p>课题理论运用扎实，调研充分，案例分析恰当，所提出的数据治理框架体系对高校图书馆用户数据治理具有重要的参考价值，拓展了图书馆学理论应用新领域；A 大学图书馆实证分析对当前图书馆进行多源数据治理具有重要的指导意义。</p> <p>专家组一致同意课题结题，并建议今后进一步加强应用研究，促进图书馆数据治理的标准化。</p>		
专家签字		刘竟	
职务/职称		副教授	



项目编号：  
2022031

## CALIS 全国农学文献信息中心研究项目 结题报告

项目名称：数智时代高校图书馆用户数据治理体系研究

项目关键词：用户数据 数据治理 治理体系 高校图书馆

项目单位(盖章)：江苏大学图书馆

通信地址：江苏省镇江市学府路 301 号 (212013)

项目主持人：张晓阳

联系电话：13914559916

电子邮件：zhxy@ujs.edu.cn

提交日期：2023 年 5 月 23 日

# 数智时代高校图书馆用户数据治理体系研究

## 1 研究背景、目的及意义

### 1.1 研究背景

近年来，云计算、物联网、人工智能、5G 等新技术迅速发展，正对各行各业产生着革命性的影响，人类进入数据富足供给的时代<sup>[1]</sup>。在这个时代，数据已经成为了一种极其宝贵的资源，而对于各行各业来说，掌握数据的能力已经成为了提高竞争力和创新力的重要保障。高校图书馆亦不例外，来自全球范围内的图书馆组织、协会、学者与馆员正积极投身于数据工作建设之中，力图抓住这前所未有的机遇与挑战，为图书馆学研究及图书馆事业建设贡献力量。

#### 1.1.1 时代背景

数智时代，人工智能技术在各行业掀起浪潮，ChatGPT 横空出世，基于“数据”的技术革命席卷而来，数据要素在市场资源配置中扮演着重要角色。

2020 年 12 月，教育部高等学校图书情报工作指导委员会召开会议，从宏观、中观、微观三个层面深入探讨高校图书馆“十四五”规划的路径与方法，指出高校图书馆要迈向高质量发展规划阶段，满足用户多样化与个性化的信息需求，并加快智慧化、个性化和特色化的智慧图书馆建设进程<sup>[2]</sup>。2021 年 3 月 11 日，十三届全国人大四次会议通过了关于“十四五”规划和 2035 年远景目标纲要的决议<sup>[3]</sup>，从此，全国各地图书馆陆续响应，争相开展“图书馆十四五战略规划”的编制工作，积极响应党和政府的号召，将加强人工智能、大数据、云计算等新技术的应用融入自身的发展规划中去。2023 年 2 月 27 日，中共中央、国务院印发了《数字中国建设整体布局规划》，指出数字中国的建设是推动中国式现代化的重要引擎，也是构筑国家新竞争优势的有力支撑。数字中国的建设离不开政府、社会组织与广大人民的参与和奉献，图书馆与相关组织也应在数据化和智慧化的道路上积极展开探索。

积极响应党和政府在各项政策中提出关于数字化建设的号召，有益于各图书馆顺应时代发展潮流，把握自身建设与发展的要点与趋势，从而加速其数据化、规范化与智慧化的进程。

### 1.1.2 行业背景

近年来, IFLA(国际图联)等国际图书馆组织与协会预测了图书馆的研究热点与趋势, 强调了数据资源的重要意义, 号召相关专家与馆员参与到信息技术理论与数据科学等相关领域的研究工作中。

2020年, 研究图书馆协会(ARL)发布《研究型图书馆和新兴技术的未来主题和预测》<sup>[4]</sup>, 指出图书馆与相关机构应该计划采用新兴的信息技术, 如数据科学、物联网和云计算等技术, 以高效管理和利用在研究、学习中生成的各类数据, 实现数据增值, 积极推动图书馆的现代化和智慧化建设工作开展。2020年, 信息科学与技术协会(ASIS&T)发布《信息科学与技术2020-2025战略规划》<sup>[5]</sup>, 报告指出要广泛吸引全球范围内具备信息科学技术与相关学科背景且对数据科学感兴趣的人才, 协同研究数据有关的理论与技术, 促进信息技术理论和实践相结合。2021年, 国际图书馆协会与机构联合会(IFLA)发布《趋势报告》<sup>[6]</sup>, 提出了图书馆领域未来10年发展的20个总体趋势, 其中之一是“数据治理”, 要求对图书馆各系统中存在的数据加以开发和利用。吴丹<sup>[7]</sup>对全球战略蓝图展开分析, 展望“十四五”时期图情学科愿景, 指出新时代下, 图书情报学科可以在社会关注的文化服务、技术创新、数据治理以及可持续发展等方面进行深化研究, 以不断升级图书情报服务, 加强数据开发与利用, 从而推动社会的可持续发展。

不难发现, 与“数据”有关的图书馆学与图书馆事业研究是近年来学科建设发展的重要内容, 推进信息技术的应用与加强数据开发利用是图书馆数字化与智慧化发展的重要引擎。

### 1.1.3 现实背景

随着信息技术的不断发展, 高校图书馆的业务系统面临着升级与替换的抉择, 无论是旧系统的迁移, 还是新系统平台的接入与使用, 图书馆都不可避免的投入财力、人力和设备资源等以妥善管理各类用户行为数据。此时, 用户与图书馆各类服务交互活动留下的数据已积累到一定程度, 来源也更加广泛, 这些行为数据存储在不同的业务系统中, 数据规模、增长速度与结构复杂度均在不断提升, 伴随而来的是数据质量、数据安全、数据标准等问题, 且业务系统间的关联性较差, 最终形成数据孤岛和系统孤岛, 增加了用户行为数据管理工作的难度。

目前，各馆尚处于数据治理的初始阶段，且对行为数据的重视程度不高，所以在用户行为数据治理的组织、流程与程序等方面缺乏相关探索。再加上一系列的数据问题，高校馆亟待开展用户行为数据资产的管理。

综上所述，全球范围内的代表性图书馆组织、协会、学者与专家均视信息技术的创新应用与数据的治理和价值增值为未来图书馆学研究和图书馆事业建设的重要发展趋势，“数据”与“智慧”理念已深入人心，且伴随着高校馆业务系统的更新与升级，数据治理似乎已成为图书馆数字化与智慧化发展的必然趋势。

各类用户行为数据的可信度和价值的提升要求高校馆实施数据治理，高校馆理应积极探索数字化之路，使“乱数据”变成高价值的“智慧数据”，“人工管理数据”变成“自动化、智能化管理数据”，“被动式的用户服务”变成“主动的、个性化的、智慧化”的服务。

## 1.2 研究目的

以高校图书馆用户行为数据的治理为研究对象，分析典型性高校馆用户行为数据治理现状，归纳数据治理的要素和问题；运用 WSR 系统分析法构建数据治理框架，结合数据治理规范、权威数据治理框架和专家的意见进一步完善框架；运用新框架分析 A 大学图书馆开展用户行为数据治理的思路与方法，构建数据治理平台；最终提出高校图书馆用户行为数据治理的对策与建议。

## 1.3 研究意义

数智时代下，数据已成为一项重要资产，开展用户行为数据治理方能彰显其价值和推进图书馆数字化与智慧化进程。数据治理是图书馆发展的重要机遇<sup>[8]</sup>，研究数据治理是智慧图书馆理论拓展与延伸的重要探索，亦是图书馆实现自我转型与创新发展的契机，更是对党和国家提出建设数字中国目标的积极回应。

### （1）理论意义

数据治理理论广泛应用于政府、银行和商企等领域，目前的研究尚未针对高校图书馆的用户行为数据治理展开系统分析。本文将数据治理与数据科学相关理论融入智慧图书馆理论研究，在分析典型性高校图书馆用户行为数据治理的要素基础上，运用 WSR 系统分析方法构建融合物理、事理和人理要素的数据治理新框架，拓展了

图书馆学研究领域。

## （2）实践意义

本文以 A 大学图书馆用户行为数据治理为例，运用 WSR 数据治理框架系统分析了其治理目标、治理机制和利益相关方，并基于 Streamlit 设计了治理平台。数据治理框架理论的实践应用对于提升图书馆服务水平与用户满意度具有重要的意义与价值，为实现用户精准服务提供路径，也为同行开展用户行为数据治理提供参考和借鉴。

## 2 研究内容及方法（思路、方法、具体内容）

### 2.1 研究内容

#### （1）相关概念与理论分析

对高校图书馆用户行为数据治理的概念进行了界定，并解释了数据治理与数据管理之间的区别与联系；阐述了代表性数据治理框架的主要内容，包括数据治理概念框架和国家数据治理规范中的框架，引入并分析 WSR 系统分析方法的原理与适用范围，为本研究打下理论基础。

#### （2）高校图书馆用户行为数据治理要素分析

通过网络调研和访谈法调研典型性高校图书馆用户行为数据治理的经验与做法，归纳数据治理要素和存在的问题。

#### （3）WSR 视域下高校图书馆用户行为数据治理框架体系构建

运用 WSR 系统分析方法从物理维度、事理维度和人理维度归纳用户行为数据治理的元素，结合国家规范、数据治理概念框架和专家的意见对框架进一步完善，并解释各要素及其关系的内涵。

#### （4）A 大学图书馆用户行为数据治理实证研究

运用 WSR 数据治理框架分析 A 大学图书馆用户行为数据治理的目标、治理机制和利益相关方，并构建数据治理平台；通过对数据治理框架的应用以再次验证新框架的科学性与可行性。

#### （5）数据治理框架应用的对策与建议

结合数据治理新框架和实证分析的经验，提出物理、事理和人理维度下的应用

建议。

## 2.2 技术路线

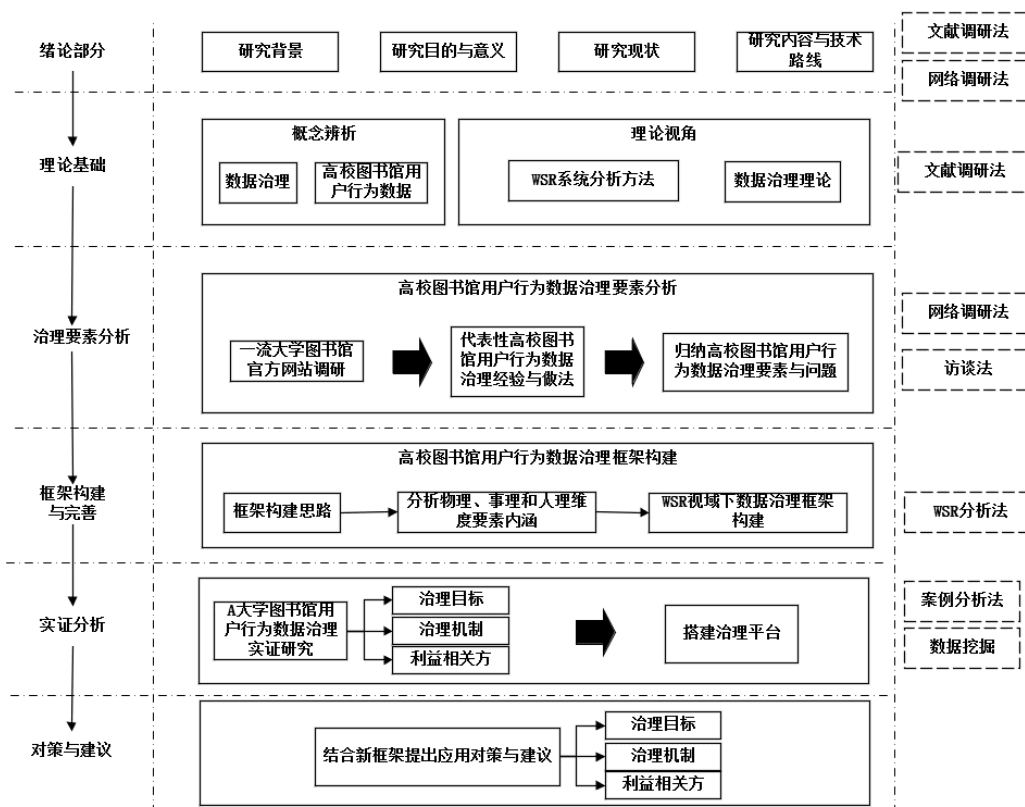


图 1 技术路线图

## 2.3 研究方法

### (1) 文献调研法

通过在 CNKI 和 Web of Science 中检索、收集相关文献并对主要文献深入阅读，分析高校图书馆用户行为数据治理的研究现状，探究构建用户行为数据治理框架的理论基础，为本文的研究思路与方法提供参考与借鉴。

### (2) 网络调研法

通过网络调研法调研国内 39 所一流大学图书馆官方网站，分析用户行为数据治理实践情况，根据调研结果选取典型性的高校图书馆作为访谈对象。

### (3) 访谈法

通过访谈典型性高校图书馆的相关专家了解其用户行为数据治理的经验和做



法，归纳总结数据治理过程中的相关要素。

#### （4）WSR 分析法

在分析典型性高校图书馆用户行为数据治理要素基础上，运用 WSR 分析法将各治理要素进行有机融合，以构建数据治理新框架。

#### （5）案例分析法

以 A 大学图书馆用户行为数据治理为例，运用 WSR 数据治理框架从物理维度、事理维度和人理维度分析其治理目标、治理机制和利益相关方，并设计出治理平台。

#### （6）数据挖掘

在实证分析阶段，使用 Python 设计 A 大学图书馆用户行为数据的质量、安全和洞察自动化程序。

### 3 结论与建议

数字中国背景下，大数据、人工智能等技术在各个领域迅速发展，加快了数字化转型进程。为了顺应时代发展潮流和促进图书馆的现代化与智慧化发展，高校图书馆积极接受和采纳新的设备与技术平台，这进一步保障了其用户服务水平和效率的提升，然而，新技术的使用使一把双刃剑，在图书馆服务期间，各业务系统留下了大量的用户行为数据，伴随而来的是新的质量、安全等管理问题，这不仅增加了馆员的工作负担，还阻碍了图书馆数字化进程。因此，高校图书馆有必要实施数据治理来解决相关问题，通过种种举措提升数据的可信度和价值，以提升服务质量和效率，并主动为读者提供个性化与智慧化的服务。本文首先对典型性高校图书馆用户行为数据治理的关键要素进行了分析与归纳，然后在国家数据治理规范中的数据治理框架和数据治理概念框架的基础上运用 WSR 系统分析方法从物理（W）、事理（S）和人理（R）的角度将相关要素进行有机融合，以构建新框架；接下来，运用 WSR 数据治理框架对 A 大学图书馆用户行为数据治理进行分析，提出数据治理的目标、机制和相关方，并基于 Streamlit 设计数据治理平台；最终提出 WSR 数据治理框架的应用对策与建议。具体研究成果如下：

（1）总结归纳当前高校图书馆用户行为数据治理的重要要素。采用网络调研法和访谈法调研典型性高校图书馆的用户行为数据治理经验与做法，归纳相关要素，凸显当前数据治理工作的重点。

(2) 基于 WSR 系统分析方法构建新数据治理框架。通过 WSR 系统分析方法从物理、事理和人理的维度融合高校馆用户行为数据治理的要素，并进一步结合国家规范、数据治理概念框架和专家的意见对模型进行补充与完善，阐述各要素的内涵与关系。

(3) 以 A 大学图书馆用户行为数据治理展开实证分析。运用 WSR 系统分析方法，分析 A 大学图书馆用户行为数据治理的目标、机制和相关方，经过分析，数据治理目标是数据质量、数据安全、元数据和数据洞察，然后围绕数据治理目标明确了数据战略，设计了相关角色与任务和决策权的分配，并分析了现有的相关政策内容，还设计出用户主数据、进出馆数据、座位预约数据和借还数据的管理流程、程序和数据字典等，最终基于 Streamlit 设计出数据治理平台。

(4) 提出了 WSR 框架的应用对策与建议：明晰数据治理的多元目标，包括多目标的设定依据与视野扩展；建立健全数据治理机制，包括岗位设置与组织建设、规范的治理流程与程序和健全的关系机制；在数据治理生命周期内推进多元主体协同治理。

## 4 项目成果（发表的文章、开发的软件、取得的实践效果等）

### 4.1 发表的文章

[1] 张晓阳,刘洋,苗松,张朝霞.2020 学年江苏大学图书馆用户行为记录数据集[J].图书馆杂志,2022,41(10):97-103.

[1]张晓阳,许欣,刘洋.用户数据视角下高校图书馆用户行为特征研究[J].图书馆学研究,2023(02):82-94.

## 5 参考文献

[1] 朝乐门. 数据科学理论与实践[M]. 清华大学出版社, 2017.

[2] 关志英, 王波. 从宏观、中观、微观三个层面深入研讨,科学制定高校图书馆“十四五”发展规划——“2018-2022 年教育部高等学校图书情报工作指导委员会第三次工作会议”综述[J]. 大学图书馆学报, 2021, 39(01): 8-14.

[3] 新华社. 十三届全国人大四次会议表决通过关于“十四五”规划和 2035 年远景目标纲要的决议 \_ 滚动新闻 \_ 中国政府网 [EB/OL]. [2021-07-17]. [http://www.gov.cn/xinwen/2021-03/11/content\\_5592248.htm](http://www.gov.cn/xinwen/2021-03/11/content_5592248.htm).

[4] ARL. Future Themes and Forecasts for Research Libraries and Emerging

- Technologies[EB/OL]. [2021-07-17].  
<https://www.arl.org/wp-content/uploads/2020/08/2020.08.21-future-themes-and-forecasts-for-research-libraries-and-emerging-technologies.pdf>.
- [5] ASIS&T. ASIS&T Strategic Plan 2020-2025[EB/OL]. [2021-07-17].  
<https://growthzonesitesprod.azureedge.net/wp-content/uploads/sites/946/Final-2020-2025-Strategic-Plan.pdf>.
- [6] 图书馆杂志. 图情报告 | 国际图联发布 2021 年《趋势报告》[EB/OL]//微信公众平台. [2022-02-10].  
[http://mp.weixin.qq.com/s?\\_\\_biz=MzAxMjExMzkwMQ==&mid=2653671686&idx=1&sn=6345477ec2a00d330a76f4dcb416927b&chksm=8069b384b71e3a924721332d43356ecef6bc0cb3fcf082c63dab9404cfe38c8301a2fd90c30#rd](http://mp.weixin.qq.com/s?__biz=MzAxMjExMzkwMQ==&mid=2653671686&idx=1&sn=6345477ec2a00d330a76f4dcb416927b&chksm=8069b384b71e3a924721332d43356ecef6bc0cb3fcf082c63dab9404cfe38c8301a2fd90c30#rd).
- [7] 吴丹, 郭清玥. “十四五”时期图情学科愿景展望——基于全球战略蓝图的分析[J]. 图书情报知识, 2021, 38(03): 14-26.
- [8] 顾立平. 数据治理——图书馆事业的发展机遇[J]. 中国图书馆学报, 2016, 42(05): 40-56.



项目编号：2022031

# 数智时代高校图书馆用户数据治理体系 研究报告

主持人：张晓阳

成 员：罗敏 苑成生 苗松 刘洋 许欣 丁雨桐 朱艳蓉

江苏大学图书馆

2023年5月

## 摘要

本研究以 WSR 系统分析方法和数据治理理论为基础，综合运用文献调研、网络调研、访谈和案例分析等研究方法，在数字中国建设背景下，将高校图书馆用户行为数据作为研究核心，分析其治理要素，基于调研结果和数据治理理论构建 WSR 数据治理新框架，通过实证分析验证框架的科学性与合理性，并提出框架的应用对策与建议。主要研究结果包括：

(1) 构建了 WSR 视域下的数据治理框架体系，包括治理目标、治理机制和利益相关方三个维度，治理目标分为数据管理体系和数据价值体系，治理机制包括结构机制、程序机制和关系机制，利益相关方分为数据供给者、数据管理者和数据消费者。

(2) 以 A 大学图书馆为例，提出了用户行为数据治理为例展开实证分析，分析出其治理目标包括数据质量、数据安全、元数据和数据洞察，主要的利益相关方包括馆务会、相关馆员、软件服务商和校级信息化中心，并设计出相关的数据战略、组织结构、人员角色、流程和程序等，最终基于以上分析研发数据治理平台工具。

(3) 结合新框架提出物理维度下明晰数据治理的多元目标，事理维度下建立健全数据治理机制和人理维度下推进多元主体协同治理的应用建议。

本研究的创新点和意义之处在于：(1) 构建 WSR 视域下高校图书馆用户行为数据治理新框架，拓展了图书馆学理论应用新领域，丰富智慧图书馆研究理论体系。(2) 以 A 大学图书馆为例，结合框架分析其用户行为数据治理的思路与方法，并基于 Streamlit 设计治理平台，为实现用户精准服务提供路径和同行开展数据治理提供借鉴与参考。

**关键词：**WSR 系统分析法；高校图书馆；用户行为数据；数据治理；案例研究

# 目 录

<b>1.绪论</b> .....	<b>6</b>
1.1 研究背景.....	6
1.1.1 时代背景.....	6
1.1.2 行业背景.....	6
1.1.3 现实背景.....	7
1.2 研究目的与研究意义.....	7
1.2.1 研究目的.....	7
1.2.2 研究意义.....	7
1.3 国内外研究现状.....	8
1.3.1 数据治理研究.....	8
1.3.2 研究述评.....	10
1.4 研究方法.....	11
1.5 研究内容与技术路线.....	11
1.5.1 研究内容.....	11
1.5.2 技术路线.....	12
<b>2.相关概念与理论基础</b> .....	<b>13</b>
2.1 相关概念.....	13
2.1.1 数据治理概念界定.....	13
2.1.2 高校图书馆用户行为数据治理概念界定.....	13
2.2 理论基础.....	13
2.2.1 WSR 系统分析法.....	13
2.2.2 数据治理框架模型.....	13
2.2.3 相关理论的应用思路.....	15
<b>3.高校图书馆用户行为数据治理要素分析</b> .....	<b>16</b>
3.1 典型高校图书馆用户行为数据治理现状分析.....	16
3.1.1 调研对象与调研步骤.....	16
3.1.2 调查结果与分析.....	16

3.2 高校图书馆用户行为数据治理要素归纳和问题分析 .....	17
3.2.1 数据治理要素归纳 .....	17
3.2.2 数据治理问题分析 .....	18
<b>4.WSR 视域下的数据治理框架构建 .....</b>	<b>19</b>
4.1 框架构建方法 .....	19
4.2 治理要素分析 .....	19
4.2.1 物理维度要素分析 .....	19
4.2.2 事理维度要素分析 .....	19
4.2.3 人理维度要素分析 .....	20
4.3 数据治理框架构建 .....	20
<b>5.A 大学图书馆用户行为数据治理实证分析 .....</b>	<b>22</b>
5.1 实证研究的思路与方法 .....	22
5.2A 大学图书馆用户行为数据治理现状与问题分析 .....	22
5.2.1 用户行为数据现状与问题分析 .....	22
5.2.2 用户行为数据管理现状与问题分析 .....	22
5.2.3 用户行为数据治理相关方分析 .....	23
5.3 A 大学图书馆用户行为数据治理目标分析 .....	23
5.4 A 大学图书馆用户行为数据治理机制分析 .....	23
5.4.1 数据战略 .....	23
5.4.2 决策权的分配 .....	23
5.4.3 角色与责任 .....	24
5.4.4 政策 .....	25
5.4.5 流程 .....	26
5.4.6 程序 .....	29
5.4.7 合约协议 .....	38
5.4.8 绩效衡量 .....	38
5.4.9 合规性监督 .....	38
5.4.10 问题管理 .....	39

5.4.11 沟通 .....	39
5.4.12 培训 .....	40
5.4.13 决策的协调 .....	40
5.5 A 大学图书馆用户行为数据治理利益相关方分析 .....	40
5.6 A 大学图书馆用户行为数据治理平台设计 .....	41
5.6.1 功能简介 .....	41
5.6.2 数据的质量审查与加工功能 .....	41
5.6.3 统计指标分析功能 .....	45
5.6.4 数据合并与格式转换功能 .....	46
<b>6.WSR 数据治理框架应用对策与建议 .....</b>	<b>48</b>
6.1 物理维度下明晰数据治理的多元目标 .....	48
6.2 数据治理多目标的设定依据 .....	48
6.3 数据治理多目标的视野拓展 .....	48
6.4 事理维度下建立健全数据治理机制 .....	48
6.4.1 完善用户行为数据治理的岗位设置与组织建设 .....	48
6.4.2 建立用户行为数据治理的规范流程与程序 .....	49
6.4.3 健全数据治理的关系机制 .....	49
6.5 人理维度下推进多元主体协同治理 .....	49
<b>7.结论与展望 .....</b>	<b>50</b>
7.1 研究结论 .....	50
7.2 研究不足与展望 .....	50
参考文献 .....	52



# 1.绪论

在数字中国建设的背景下，数字化和智慧化成为了两个重要的研究议题，数据也成为了市场要素的重要组成部分，数据治理建设在政府、银行、医院和高校等领域如火如荼。随着高校图书馆接受和使用新的信息技术，产生了大量的行为数据。治理这些数据将有益于加速高校图书馆的数字化与智慧化进程。

## 1.1 研究背景

近年来，云计算、物联网、人工智能、5G 等新技术迅速发展，正对各行各业产生着革命性的影响，人类进入数据富足供给的时代<sup>[1]</sup>。在这个时代，数据已经成为了一种极其宝贵的资源，而对于各行各业来说，掌握数据的能力已经成为了提高竞争力和创新力的重要保障。高校图书馆亦不例外，来自全球范围内的图书馆组织、协会、学者与馆员正积极投身于数据工作建设之中，力图抓住这前所未有的机遇与挑战，为图书馆学研究与图书馆事业建设贡献力量。

### 1.1.1 时代背景

数智时代，人工智能技术在各行业掀起浪潮，ChatGPT 横空出世，基于“数据”的技术革命席卷而来，数据要素在市场资源配置中扮演着重要角色。

2020 年 12 月，教育部高等学校图书情报工作指导委员会召开会议，从宏观、中观、微观三个层面深入探讨高校图书馆“十四五”规划的路径与方法，指出高校图书馆要迈向高质量发展规划阶段，满足用户多样化与个性化的信息需求，并加快智慧化、个性化和特色化的智慧图书馆建设进程<sup>[2]</sup>。2021 年 3 月 11 日，十三届全国人大四次会议通过了关于“十四五”规划和 2035 年远景目标纲要的决议<sup>[3]</sup>，从此，全国各地图书馆陆续响应，争相开展“图书馆十四五战略规划”的编制工作，积极响应党和政府的号召，将加强人工智能、大数据、云计算等新技术的应用融入自身的发展规划中去。2023 年 2 月 27 日，中共中央、国务院印发了《数字中国建设整体布局规划》，指出数字中国的建设是推动中国式现代化的重要引擎，也是构筑国家新竞争优势的有力支撑。数字中国的建设离不开政府、社会组织与广大人民的参与和奉献，图书馆与相关组织也应在数据化和智慧化的道路上积极展开探索。

积极响应党和政府在各项政策中提出关于数字化建设的号召，有益于各图书馆顺应时代发展潮流，把握自身建设与发展的要点与趋势，从而加速其数据化、规范化与智慧化的进程。

### 1.1.2 行业背景

近年来，IFLA（国际图联）等国际图书馆组织与协会预测了图书馆的研究热点与趋势，强调了数据资源的重要意义，号召相关专家与馆员参与到信息技术理论与数据科学等相关领域的研究工作中。

2020 年，研究图书馆协会（ARL）发布《研究型图书馆和新兴技术的未来主题和预测》<sup>[4]</sup>，指出图书馆与相关机构应该计划采用新兴的信息技术，如数据科学、物联网和云计算等技术，以高效管理和利用在研究、学习中生成的各类数据，实现数据增值，积极推动图书馆的现代化和智慧化建设工作开展。2020 年，信息科学与技术协会（ASIS&T）发布《信息科学与技术 2020-2025 战略规划》<sup>[5]</sup>，报告指出要广泛吸引全球范围内具备信息科学技术与相关学科背景且对数据科学感兴趣的人才，协同研究数据有关的理论与技术，促进信息技术理论和实践相结合。2021 年，国际图书馆协会与机构联合会（IFLA）发布《趋势报告》<sup>[6]</sup>，提出了图书馆领域未来 10 年发展的 20 个总体趋势，其中之一是“数据治理”，要求对图书馆

各系统中存在的数据加以开发和利用。吴丹<sup>[1]</sup>对全球战略蓝图展开分析,展望“十四五”时期图情学科愿景,指出新时代下,图书情报学科可以在社会关注的文化服务、技术创新、数据治理以及可持续发展等方面进行深化研究,以不断升级图书情报服务,加强数据开发与利用,从而推动社会的可持续发展。

不难发现,与“数据”有关的图书馆学与图书馆事业研究是近年来学科建设发展的重要内容,推进信息技术的应用与加强数据开发利用是图书馆数字化与智慧化发展的重要引擎。

### 1.1.3 现实背景

随着信息技术的不断发展,高校图书馆的业务系统面临着升级与替换的抉择,无论是旧系统的迁移,还是新系统平台的接入与使用,图书馆都不可避免的投入财力、人力和设备资源等以妥善管理各类用户行为数据。此时,用户与图书馆各类服务交互活动留下的数据已积累到一定程度,来源也更加广泛,这些行为数据存储在不同的业务系统中,数据规模、增长速度与结构复杂度均在不断提升,伴随而来的是数据质量、数据安全、数据标准等问题,且业务系统间的关联性较差,最终形成数据孤岛和系统孤岛,增加了用户行为数据管理工作的难度。

目前,各馆尚处于数据治理的初始阶段,且对行为数据的重视程度不高,所以在用户行为数据治理的组织、流程与程序等方面缺乏相关探索。再加上一系列的数据问题,高校馆亟待开展用户行为数据资产的管理。

综上所述,全球范围内的代表性图书馆组织、协会、学者与专家均视信息技术的创新应用与数据的治理和价值增值为未来图书馆研究和图书馆事业建设的重要发展趋势,“数据”与“智慧”理念已深入人心,且伴随着高校馆业务系统的更新与升级,数据治理似乎已成为图书馆数字化与智慧化发展的必然趋势。

各类用户行为数据的可信度和价值的提升要求高校馆实施数据治理,高校馆理应积极探索数字化之路,使“乱数据”变成高价值的“智慧数据”,“人工管理数据”变成“自动化、智能化管理数据”,“被动式的用户服务”变成“主动的、个性化的、智慧化”的服务。

## 1.2 研究目的与研究意义

### 1.2.1 研究目的

以高校图书馆用户行为数据的治理为研究对象,分析典型性高校馆用户行为数据治理现状与问题,归纳数据治理的要素和问题;结合权威数据治理框架和专家的意见,运用 WSR 系统分析法构建并完善用户行为数据治理体系框架;运用数据治理框架分析 A 大学图书馆开展用户行为数据治理的思路与方法,构建数据治理平台;最终提出高校图书馆用户行为数据治理的对策与建议。

### 1.2.2 研究意义

数智时代下,数据已成为一项重要资产,开展用户行为数据治理方能彰显其价值和推进图书馆数字化与智慧化进程。数据治理是图书馆发展的重要机遇<sup>[2]</sup>,研究数据治理是智慧图书馆理论拓展与延伸的重要探索,亦是图书馆实现自我转型与创新发展的的重要契机,更是对党和国家提出建设数字中国目标的积极回应。

#### (1) 理论意义

数据治理理论广泛应用于政府、银行和商企等领域,目前的研究尚未针对高校图书馆的用户行为数据治理展开系统分析。本文将数据治理与数据科学相关理论融入智慧图书馆理论

研究，在分析数据治理的要素基础上，运用 WSR 系统分析方法构建融合物理、事理和人理要素的数据治理体系新框架，拓展了图书馆学研究领域。

### (2) 实践意义

本文以 A 大学图书馆用户行为数据治理为例，运用 WSR 数据治理框架系统分析了其治理目标、治理机制和利益相关方，并基于 Streamlit 设计了治理平台。数据治理框架理论的实践应用对于提升图书馆服务水平与用户满意度具有重要的意义与价值，为实现用户精准服务提供路径，也为同行开展用户行为数据治理提供参考和借鉴。

## 1.3 国内外研究现状

### 1.3.1 数据治理研究

以“数据治理”为关键词在 CNKI 中检索相关文献，检索时间限制为 2012-2021 年，经过数据清洗后得到 1676 条检索结果，近 5 年以来相关论文年发表数量逐年上升；根据知网检索结果可视化功能，发现主要主题词有“数据治理、大数据、数据质量、数字化转型、智慧校园、数据安全”等；笔者继续以“data governance”为主题词在 Web of Science 数据库核心合集中检索相关文献，检索时间限定为 2012-2021 年，共得到 28215 条结果，近 5 年以来，每年相关文献数量超过 2200 篇。图 1.5 描绘了 2012-2021 年间 CNKI 和 Web of Science 核心合集中有关“数据治理 (Data Governance)”文献年发表篇数。

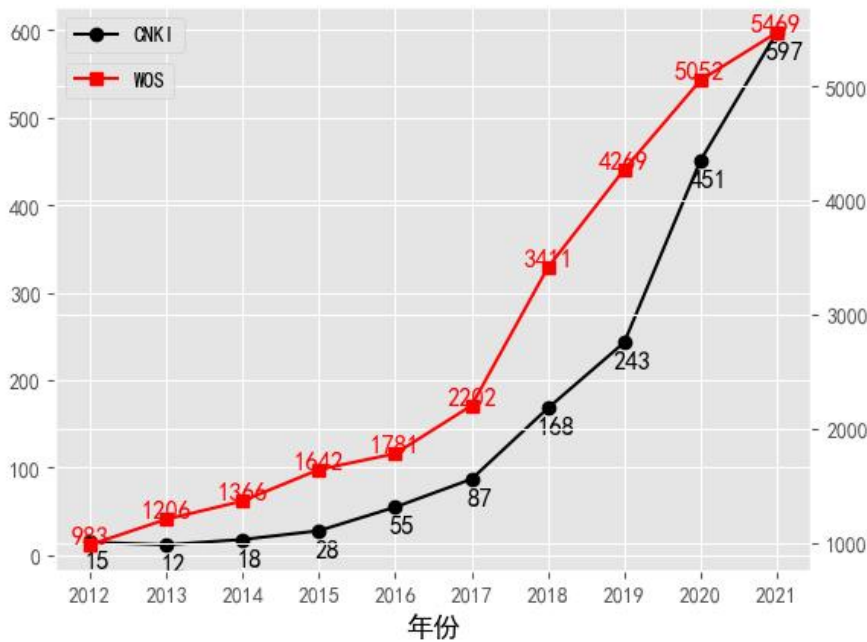


图 1.5 国内外数据治理研究趋势

通过对比发现国内学者在数据治理领域公开的文献数量和国际学者存在一定差距，近几年来，“数据治理”在国内的研究发展迅速，年发表文献数量于 2020 年达到 600 篇，约为国际学者已公开文献篇数的 1/5。

#### (一) 数据治理内涵研究

目前，数据治理的概念尚未有一个明确的标准与界定。DGI 认为数据治理是对数据相关事项的决策和权力的行使。这是一个与信息相关的流程的决策权利和责任系统，根据商定的模式执行，这些模式描述了谁可以在什么情况下使用什么方法对什么信息采取什么行动<sup>[35]</sup>。DAMA 认为数据治理是指对数据资产管理行使权力和控制的活动集合（包括计划、监督和执

行)<sup>[36]</sup>。Cohen 认为<sup>[37]</sup>数据治理是一个公司管理数据的数量、一致性、可用性、安全性和可用性的过程。Thoma-s 认为<sup>[38]</sup>数据治理指人和信息系统在执行与信息有关的程序时的组织机构、规则、决策权和责任。包冬梅<sup>[39]</sup>从数据治理的目标、职能、核心和本质四个方面来概括数据治理的概念内涵。张邵华<sup>[40]</sup>认为数据治理的本质是就是对一个机构(企业或政府部门)的数据从收集融合到分析管理和利用进行评估、指导和监督(EDM)的过程,通过提供不断创新的数据服务,为企业创造价值。

### (二) 数据治理要素研究

Marinos 认为数据治理的要素有职责与战略责任、标准、管理盲点、迎接复杂性、跨部门问题、计量、合作、战略控制点的选择、合规监测、意识与培训<sup>[41]</sup>。Sonia 认为强劲启动、快速致胜、领导支持、设置资助、效果度量、坚持不懈是数据治理的六个至关重要的因素<sup>[42]</sup>。Stockdale 总结了数据治理的 5 个要素:治理结构,角色和职责,数据分类,政策、标准、指南和实施<sup>[43]</sup>。Ibrahim 总结了数据治理的 9 个要素:业务驱动的数据整合策略,员工的数据能力,灵活的数据工具和技术,明确、包容的数据要求,明确的数据过程和程序,有重点和切实的数据战略,既定的数据角色和责任,明确数据可访问和可用权限责任,有效的数据监测和反馈<sup>[44]</sup>。

### (三) 数据治理框架研究

国内外主要的的数据治理模型包括 DGI 数据治理框架<sup>[35]</sup>, IBM 数据治理框架<sup>[45]</sup>, DAMA 数据治理框架<sup>[46]</sup>, CALib 高校图书馆数据治理框架<sup>[39]</sup>、大数据治理模型<sup>[40]</sup>和 HAO 数据治理模型<sup>[47]</sup>。

### (四) 数据治理的成熟度模型

数据治理成熟度模型有利于评估组织当前数据管理和控制的现状,是实施数据治理非常关键的环节<sup>[48]</sup>。Gartner 将数据治理成熟度划分为 6 个阶段:无意识、有意识、被动反映、主动反应、可管理、有效果<sup>[49]</sup>。IBM 认为“初始级、可管理、可定义、定量管理、优化级”是数据治理成熟度的 5 个阶段<sup>[45]</sup>;张宇杰将政府大数据治理的成熟度等级分为 7 级:未启动、非正式、有文件、有计划、有部署、有影响和有重大影响<sup>[50]</sup>。

### (五) 数据治理体系研究

数据治理体系是指一个组织内部的一系列规则、标准和流程,旨在确保数据的质量、安全、可用性和可信度,并提供清晰的决策框架和责任分配机制。目前,数据治理体系的研究集中在健康数据、政府数据、商业数据、高校数据和图书馆数据等主题。

表 1.3 数据治理体系研究

领域	作者	研究内容
健康数据	傅昊阳 <sup>[51]</sup>	从大数据治理保障体系、大数据管理与应用平台、大数据智能处理技术与方法、大数据创新应用四个方面构建健康医疗大数据治理体系。
	刘刚 <sup>[52]</sup>	概括俄罗斯国家数据治理体系建设主要措施:健全法律法规制度、构建组织管理体系、建立“统一信息平台”、知识实施国家实验、强化网络安全保障。
政府数据	谭必勇 <sup>[53]</sup>	分析英国以规章制度为保障、以组织要素为主体、以技术治理为途径的政府数据治理体系建设路径。
	安小米 <sup>[50]</sup>	提出政府大数据治理体系通用要素框架,包含治理目标、治理主体、治理客体、治理活动和治理工具五个关键要素。
商业数据	以娜 <sup>[54]</sup>	认为商业银行数据治理体系由上至下包括数据治理、数据管理、数据应用和技术支持四个层面。

领域	作者	研究内容
	李福娟 <sup>[55]</sup>	提出航空公司数据治理，包含数据治理战略、机制、专题和实现4个层面。
高校数据	宋苏轩 <sup>[56]</sup>	构建具有“过程”、“调和”、“多元和“互动”特征的”高校数据治理统筹管理体系。
图书馆数据	卢凤玲 <sup>[57]</sup>	提出融合数据治理体系的智慧图书馆框架，包括基础设施层、数据服务层和交互应用层三个层次。

#### (六) 数据治理应用领域研究

从已有文献来看，数据治理研究领域主要包括企业数据治理、政府数据治理、图书馆数据治理、高等教育数据治理等方面。

表 1.4 数据治理研究领域

应用领域	研究人员/机构	研究内容概述
企业数据治理	DAMA <sup>[46]</sup>	提出由功能子框架和环境要素子框架组成的 DAMA 数据治理框架。
	DGI <sup>[35]</sup>	提出 DGI 数据治理框架
	IBM <sup>[45]</sup>	提出由支撑域、核心域、促成因素和成果四个层面构成的数据治理框架。
政府数据治理	许阳 <sup>[58]</sup>	对政府数据治理的概念、应用场景和困境进行总结概括。
	张宇杰 <sup>[59]</sup>	构建了政府大数据治理的成熟度评测指标体系。
图书馆数据治理	包冬梅 <sup>[39]</sup>	提出 CALib 高校图书馆数据治理框架。
	顾立平 <sup>[8]</sup>	探索数据治理在图书馆事业应用的可行性。
	严昕 <sup>[60]</sup>	构建了公共图书馆治理框架。
高等教育数据治理	许晓东 <sup>[61]</sup>	将高等教育数据治理划分为数据的获取和抽取、整合和分析、解释和预测三个阶段。
	董晓辉 <sup>[62,63]</sup>	设计了高校教育大数据治理框架；并对高校教育治理体系构成要素展开探索。
	谢娟 <sup>[64]</sup>	从价值、向度与路径三个层面探讨教育数据治理的伦理框架。

### 1.3.2 研究述评

目前，国际上已形成比较成熟的数据治理框架，且广泛应用于银行、政府、商企和高校等领域，其主要关注数据治理生命周期内的要素组成，较少关注相关因素的层次性且并鲜有在高校馆用户行为数据治理领域开展深入研究，更未从物理、事理和人理交融的角度归纳关键治理要素。

用户行为数据的管理需要系统科学的数据治理框架来指导。随着图书馆业务系统的更新与迭代，各类用户行为数据产生了质量、安全等问题，为了提升数据的可信度和价值和推进图书馆数字化进程，高校图书馆亟需一个系统科学的数据治理框架。因此，本研究提出运用 WSR 系统分析方法从物理、事理和人理三个层面归纳高校图书馆用户行为数据治理的要素以构建新框架，并提出其应用对策与建议，对于高校图书馆提升用户行为数据治理能力具有重要的现实意义。

## 1.4 研究方法

### (1) 文献调研法

通过在 CNKI 和 Web of Science 中检索、收集相关文献并对主要文献深入阅读,分析高校图书馆用户行为数据治理的研究现状,探究构建用户行为数据治理框架的理论基础,为本文的研究思路与方法提供参考与借鉴。

### (2) 网络调研法

通过网络调研法调研国内 39 所一流大学图书馆官方网站,分析用户行为数据治理实践情况,根据调研结果选取典型性的高校图书馆作为访谈对象。

### (3) 访谈法

通过访谈典型性高校图书馆的相关专家了解其用户行为数据治理的经验和做法,归纳总结数据治理过程中的相关要素。

### (4) WSR 分析法

在分析典型性高校图书馆用户行为数据治理要素基础上,运用 WSR 分析法将各治理要素进行有机融合,以构建数据治理新框架。

### (5) 案例分析法

以 A 大学图书馆用户行为数据治理为例,运用 WSR 数据治理框架从物理维度、事理维度和人理维度分析其治理目标、治理机制和利益相关方,并设计出治理平台。

### (6) 数据挖掘

在实证分析阶段,使用 Python 设计 A 大学图书馆用户行为数据的质量、安全和洞察自动化程序。

## 1.5 研究内容与技术路线

### 1.5.1 研究内容

#### (1) 相关概念与理论基础

对高校图书馆用户行为数据治理的概念进行了界定,并解释了数据治理与数据管理之间的区别与联系;阐述了代表性数据治理框架的主要内容,包括数据治理概念框架和国家数据治理规范中的框架,还介绍了 WSR 系统分析方法的原理与适用范围,为本研究打下理论基础。

#### (2) 高校图书馆用户行为数据治理要素分析

通过网络调研和访谈法调研典型性高校图书馆用户行为数据治理的经验与做法,归纳数据治理要素和存在的问题。

#### (3) WSR 视域下高校图书馆用户行为数据治理框架构建

运用 WSR 系统分析方法从物理维度、事理维度和人理维度归纳用户行为数据治理的元素,结合国家规范、数据治理概念框架和专家的意见对框架进一步完善,并解释各要素及其关系的内涵。

#### (4) A 大学图书馆用户行为数据治理实证研究

运用 WSR 数据治理框架分析 A 大学图书馆用户行为数据治理的目标、治理机制和利益相关方,并构建数据治理平台;通过对数据治理框架的应用以再次验证新框架的科学性与可行性。

#### (5) 数据治理框架应用的对策与建议

结合数据治理新框架和实证分析的经验,提出物理、事理和人理维度下的应用建议。

## 1.5.2 技术路线

本研究的技术路线图如图 1.6 所示。

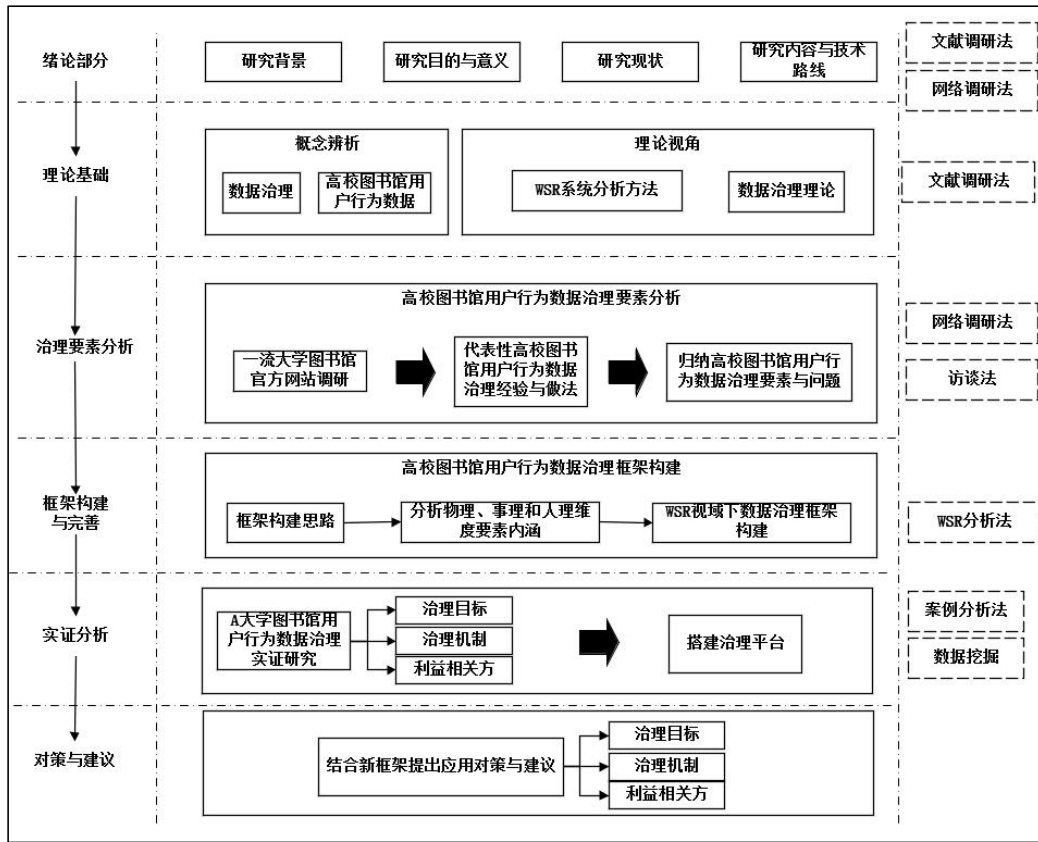


图 1.6 技术路线图

## 2. 相关概念与理论基础

### 2.1 相关概念

#### 2.1.1 数据治理概念界定

数据治理的概念在不同研究领域的定义存在一定差别，以下是一些代表性组织与学者提出的相关定义。DGI<sup>[35]</sup>认为数据治理是对数据相关事项的决策和权力的行使。这是一个与信息相关的流程的决策权利和责任系统，根据商定的模式执行，这些模式描述了谁可以在什么情况下使用什么方法对什么信息采取什么行动。DAMA<sup>[46]</sup>则认为数据治理是指对数据资产管理行使权力和控制的活动集合（包括计划、监督和执行）。包冬梅<sup>[39]</sup>从数据治理的目标、职能、核心和本质四个方面来概括数据治理的概念内涵。

本研究认为数据治理是在利益相关方协作基础上围绕数据治理目标而实施的一系列管理活动，包括数据治理生命周期内的战略、组织结构与决策权分配、角色与责任、政策、流程、程序等的制定。

#### 2.1.2 高校图书馆用户行为数据治理概念界定

用户行为数据是指用户在接受服务时各种行为产生的数据，如用户图书借阅、活动参与的数据，访问数字资源的数据<sup>[66]</sup>。大数据时代读者行为数据主要由阅读服务需求产生、阅读活动相关信息采集、阅读内容与服务模式选择、开展阅读活动和用户阅读反馈 5 个活动步骤产生<sup>[67]</sup>。阮光册等人认为数据驱动的用户行为研究更多地聚焦于用户与各类信息系统平台及信息本身交互过程中产生的数据<sup>[68]</sup>。因此，本研究将高校图书馆用户行为数据（Library Users' Behavior Data, LUBD）定义为用户和高校图书馆各类用户服务的交互活动产生的数据。高校图书馆用户行为数据治理则是指对高校馆的用户行为数据资产管理行使权力和控制的活动集合，包括决策权与责任的指定，正式化的政策、程序、流程等的制定和利益相关方的关系协调等活动。

### 2.2 理论基础

#### 2.2.1 WSR 系统分析法

WSR（物理——事理——人理）系统分析法是顾基发提出<sup>[69]</sup>，用以解决相对复杂问题的工具。WSR 方法论强调了人在系统中的重要性，从物、事、人三个方面对系统展开研究。物理是指客观存在的物质，不以人的意志为转移；事理反映了是人对物的响，侧重于系统机理和组织等内容分析。人理则是对组织中的人及其关系在系统中的作用的研究，强调人的主观作用。WSR 方法广泛应用于框架与体系构建<sup>[70,71]</sup>、评价研究<sup>[72,73]</sup>、影响因素<sup>[74,75]</sup>探究等方面。WSR 系统分析法可以从物、事和人三个维度分析高校图书馆用户行为数据治理涉及的多个因素，并将数据治理有关的要素关联成一个系统的框架，能够提供用户行为数据治理的整体视角和新思路，为高校图书馆实施用户行为数据治理提供系统科学的理论指导。

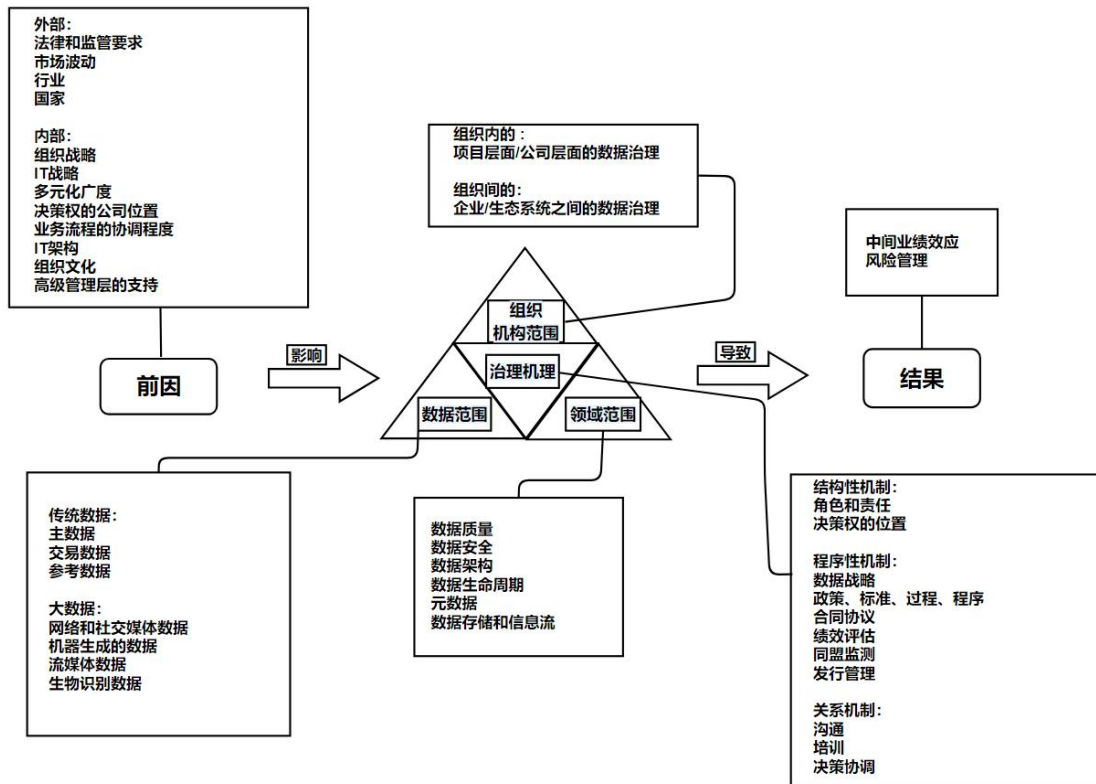
#### 2.2.2 数据治理框架模型

数据治理框架模型是包含各类治理要素的系统框架，为数据活动的开展与实施提供系统科学的理论指导。本研究主要参考借鉴了国家规范中的数据治理框架和数据治理概念框架。



### (1) 数据治理概念框架

2019年，Rene Abraham等人提出数据治理概念框架<sup>[76]</sup>，认为治理机制是框架的核心维度，将被应用于组织机构范围、数据范围和领域范围，其包括连接业务、IT和数据管理功能的正式结构、用于决策和监控的正式流程和程序，以及支持相关方积极参与和协作的实践；强调每个数据治理计划必须指定重点关注的数据类型，不同类型数据对数据治理的要求不同，并将数据区分为传统数据和大数据；领域范围是数据治理计划实现的目标，一般的数据治理项目都包含2至3个目标；组织范围限制了治理数据的范围与层次，区分为组织内部与组织间的数据治理。在数据治理计划与实施之前，还需要关注内外部因素，这对数据治理的实施有重要影响。在数据治理项目实施的部分阶段应及时评估数据治理的结果，即评估数据治理对数据利用水平的积极影响程度和数据相关的风险是否有所缓解。



### (2) 国家数据治理规范

2018年，由中国国家标准化管理委员会发布的《信息技术服务 治理 第5部分：数据治理规范》<sup>[77]</sup>（GB/T 34960.5-2018），提出指导性数据治理框架，包含顶层设计、数据治理环境、数据治理域和数据治理过程四个部分；顶层设计是实施数据治理的基础，包括战略规划、组织构建和架构设计；内外部环境 and 促成因素是数据治理实施的保障；数据治理域是数据治理实施的对象，包括数据管理体系和数据价值体系，数据管理体系包括数据标准、数据质量、数据安全、元数据管理、数据生存周期，数据机价值体系包括数据流通、数据服务和数据洞察；数据治理过程则是数据治理实施的方法，由统筹和规划、构建和运行、监控和评价、改进和优化组成。

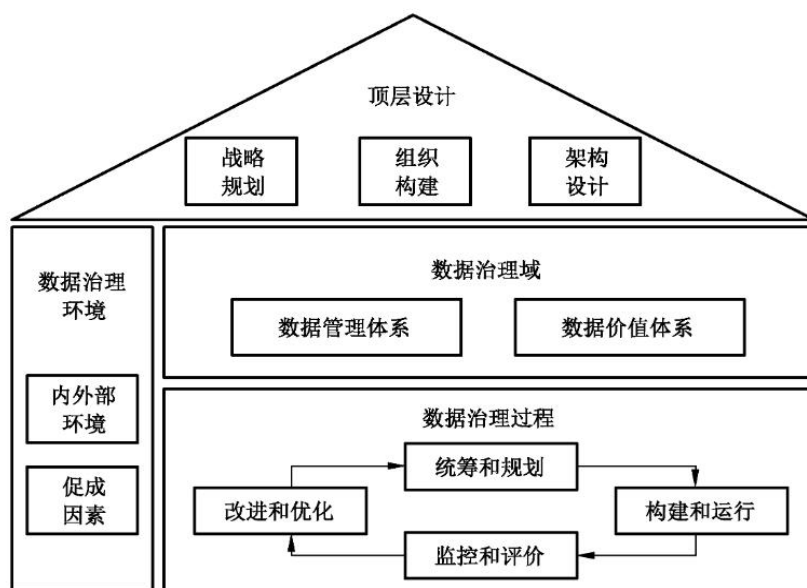


图 2.2 数据治理框架<sup>[77]</sup> (作者重绘)

### 2.2.3 相关理论的应用思路

本研究在归纳高校图书馆用户行为数据治理的要素基础上，运用 WSR 系统分析方法从物理、事理和人理的维度将相关治理要素有机融合，以构建一个系统科学的数据治理理论框架，并结合数据治理概念框架、国家规范和专家的意见进一步完善新构建的框架。

## 3.高校图书馆用户行为数据治理要素分析

目前，各高校对用户行为数据的重视程度不同，部分馆已对其采取了相应的管理措施。为了进一步了解当前高校图书馆用户行为数据治理的现状，将通过网络调研选择代表性图书馆，并基于与代表性高校馆的深度访谈归纳和分析数据治理的相关要素与问题。

### 3.1 典型高校图书馆用户行为数据治理现状分析

#### 3.1.1 调研对象与调研步骤

为了了解高校图书馆用户行为数据治理情况，对国内 39 所一流大学图书馆官方网站进行调研（调研时间：2022 年 10 月），通过检索与“用户”、“读者”与“数据”相关信息，发现相关的公告和新闻较少，更多的是关于用户行为数据的可视化报表，例如图书馆年度报告、读者阅读报告等，主要通过网页、微信公众号和视频的方式展示。

在文献调研和网络调研的基础上，选取了 11 所代表性高校图书馆，其官网络拥有较详细的阅读报告或已在数据管理平台公布过相关的用户行为数据集，通过电话和邮箱咨询的方式，访谈了来自于武汉大学、南京大学、厦门大学、南方科技大学、复旦大学和江苏大学图书馆的六位专家，通过咨询了解了其用户行为数据治理的经验与做法，以发现用户行为数据治理的相关要素。

#### 3.1.2 调查结果与分析

从调研的结果来看，重点治理的用户行为数据包括纸本图书借还数据、座位预约使用数据和进出馆数据，以下从数据质量、数据安全、数据流通、数据洞察、数据标准、元数据管理以及利益相关方七个方面分析用户行为数据治理现状。

##### （1）数据质量管理情况

在数据质量方面，各馆主要依托图书馆业务系统的自动化流程控制生成的用户行为数据质量。当前高校馆的用户行为数据来源于多个业务系统，例如图书流通借还系统、图书馆门禁管理平台 and 座位预约系统，为了保证多个业务系统共有的用户基本属性数据的一致性，部分高校馆已和校级的信息化中心等部门展开合作。同时，受限于各个业务系统的独立性，各业务系统一般都由相关的负责人完成数据采集、清洗和质量监控等多份工作，这些工作由具备相关数据素养的馆员在馆内部约定形成一致的流程。数据质量工作的发起人一般是由分管馆长担任，由其完成数据质量工作的各项决策活动。数据质量管理的战略在各馆没有形成正式的文件说明，但其目标通常是通过数据清洗工作等提升数据的可信度以满足各项活动的需求，且部分高校由信息化中心等部门颁布了《数据管理办法》等政策，明确提出有关数据质量等管理要求。针对常见的数据质量问题，如缺失值、异常值和重复值等质量问题，各馆均有馆员完成相应的工作。每当有新的相关政策与规章制度或重要数据管理任务，馆内部都会进行培训工作，并通过小型会议等形式加强不同层级馆员之间的沟通与交流，以及时发现和反馈问题，提升工作的效率。

##### （2）数据安全情况

在数据安全方面，一般由图书馆系统保障部门等负责相关安全工作。国务院颁布的《个人信息保护法》、《网络信息安全法》等法规已受到各高校馆的重视，且部分高校已有校级部门制定并颁布《数据安全管理办法》等规章制度，用户行为数据反映了用户个人的在馆活动规律，这类数据的安全是图书馆数据工程重要的工作内容。当前主要通过业务系统与云端数据隐私控制和多端备份等方法提升数据安全能力。

### （3）数据流通管理情况

在数据流通方面，部分高校馆已有成文的用户行为数据申请、审批等流程规定，要求申请者严格遵照规定完成数据使用的申请手续等，包括填写使用者个人或组织的基本信息以及数据使用的目标和方法等，待相关部门的领导或负责人完成审批工作后，申请者才会接收到数据。在数据库服务商处存储的部分数据库使用数据主要通过 API 接口和申请使用等方式获取，部分高校馆和服务商之间有合约协议，规定了流通数据的方式与内容和双方的责任与义务等。数据流通战略目标旨在促进用户行为数据共享的合法性与安全性，以进一步提升图书馆系统安全性和用户个人敏感数据的保障力度。

### （4）数据洞察管理情况

在数据洞察方面，部分高校馆对用户进出馆数据、座位预约使用数据、纸本图书预约与借还数据等按年或月进行统计分析形成读者年度报告等，分析的结果以网页、微信公众号文章、短视频等方式发布至各平台。部分微信公众号推送文章和短视频里包含了动态的图表，给用户带来更加惊艳的视觉效果。部分馆还基于用户行为数据开展了毕业生画像和实时馆情分析等活动。目前数据洞察的方法以统计分析为主，诊断性分析和预测性分析运用较少。

### （5）数据标准管理情况

在数据标准方面，高校馆主要使用业务系统本身的流程和标准来规范生成的用户行为数据，部分字段需要参照校级相关部门颁发的数据管理办法，如来源于不同业务系统的用户基本属性数据会存在不一致的情况，包括身份类型、班级等字段，需要参照规范的编码规则；高校馆也参考了外部已有的规范，包括书籍、电子资源等元数据规范，如用户借阅借还数据中涉及的书籍名称、作者等字段需要参考外部数据标准。

### （6）元数据管理情况

在元数据方面，高校馆主要依靠各业务系统来管理元数据，在部分拥有先进信息技术与平台的馆会有相关馆员负责元数据模板的创建与管理等活动。

### （7）利益相关方

用户行为数据治理生命周期内涉及的主要相关方包括相关馆员、馆内领导、校级相关部门、软件服务商和服务对象（主要是在校师生）等。师生用户与高校图书馆用户服务的交互活动留下的数据是用户行为数据的主要来源；校级信息化部门通常会提供标准的用户基本属性数据，系统服务商在得到相应的请求后会提供数据库使用数据等；数据馆员一般负责数据的采集、加工和分析等工作，业务馆员负责对数据背后的业务问题深入分析，部分学校会建立数据中心，由校级信息化中心等部门负责对全校所有数据的管理，也有部分馆和系统服务商合作协同管理数据；馆内和校级的领导会通过年度阅读报告等方法了解图书馆的业务运行情况。

## 3.2 高校图书馆用户行为数据治理要素归纳和问题分析

### 3.2.1 数据治理要素归纳

综上所述，将典型性高校图书馆用户行为数据治理的要素进行归纳，汇总如表 3.1 所示。有“●”标记的表示代表馆已存在对应的管理活动。

表 3.1 高校图书馆用户行为数据治理要素分析

	角色责任	政策	流程	程序	合约协议	问题管理
数据质量	●	●		●		●
数据标准	●	●				
数据安全	●	●				●

	角色责任	政策	流程	程序	合约协议	问题管理
元数据	●	●				
数据流通	●	●			●	
数据洞察	●	●	●	●		

目前，高校图书馆用户行为数据治理目标主要包括数据质量、数据安全、数据流通和数据洞察四个方面，其次关注的是数据标准、元数据。从具体的数据管理活动来看，角色与责任和政策关注较多，而与流程、程序、合约协议和问题管理相关的管理活动较少或尚未开展。从利益相关方来看，主要包括用户、馆内相关领导与馆员、校级信息化中心和软件服务商等，协调的措施主要包括沟通与培训。

### 3.2.2 数据治理问题分析

目前，针对于各馆重点关注的治理要素仍然存在一定的问题，以下围绕数据质量、数据安全、数据流通和数据洞察四个方面进行阐述。

#### （一）数据质量管理有待标准化和自动化

部分馆虽然有相关馆员胜任相关的数据质量管理工作，但是这些工作的流程与程序尚未标准化与正式化，且仅凭借相关馆员的经验完成，因此，数据质量工作的要求有待明确，包括数据质量检查的指标（缺失率、重复率、冗余性等）、数据质量控制的方法等。此外，针对一些数据质量问题，仍依靠人工来完成，并未针对常见的质量问题设计自动化的清洗程序。

#### （二）数据安全意识需要提升

目前，数据安全的控制仅限于业务系统的安全管理，包括数据采集、访问权限等。然而，在日常的数据生产活动中，数据尚存在一定的安全隐患，需要相关馆员提升数据安全意识，例如，在采集用户行为数据时应遵采集规范，注意隐私信息泄露；在数据报表中应选取合适的加密工具对相关字段进行脱敏处理。

#### （三）数据流通范围需扩大

部分馆针对本校内的数据流通已有明确的规定，数据的申请使用必须要遵循相关规定，且仅对本校部分师生开放，因此，各类行为数据流通度低，为了提升数据的价值和可信度，应对各类数据进行加工，形成数据产品，从而面向更多的受众，通过举办数据竞赛等措施对数据进行深度挖掘，以实现数据增值服务。

#### （四）数据洞察自动化工具有待开发

部分馆已针对各类型行为数据进行统计分析，并生成了馆年度数据报告和用户画像等，但是数据范围、统计指标和展现方式不尽相同，为了提升馆员工作效率，应建立规范的指标字典和设计自动化的报表工具。

## 4.WSR 视域下的数据治理框架构建

通过分析典型性高校图书馆用户行为数据治理的经验做法,归纳出数据治理的相关要素,为了构建系统科学的数据治理框架,需要运用合适的理论方法并参考已有的权威的框架。为此,本文通过 WSR 系统分析方法将数据治理的相关要素有机融合,并结合数据治理理论对各要素及其关系进行解释说明。

### 4.1 框架构建方法

在分析高校图书馆用户行为数据治理要素基础上,结合国家治理规范、数据治理概念框架和专家的意见,从物理维度(W)、事理维度(S)和人理维度(R)对相关要素进行分类,以构建系统科学的用户行为数据治理框架。物理维度关注的是数据治理的客体,人理维度需要考虑数据治理过程中有关人的因素,而事理则需要关注数据治理的机理与机制。

### 4.2 治理要素分析

#### 4.2.1 物理维度要素分析

物理维度对应着数据治理的客体,是数据治理的目标。国家数据治理规范<sup>[77]</sup>认为数据治理域是数据治理的实施对象,包含数据管理体系和数据价值体系,其中,数据管理体系包括数据标准、数据质量、数据安全、元数据管理和数据生存周期,而数据价值体系包括数据流通、数据服务和数据洞察。Abraham R 将数据治理的目标命名为数据治理决策域,包括数据质量、数据安全、数据架构、数据生命周期、元数据以及数据存储和基础设施<sup>[76]</sup>。综合以上分析,图书馆用户行为数据治理的目标可以概括为数据标准、数据质量、数据安全、元数据、数据生存周期、数据服务、数据流通、数据洞察和数据架构。

数据标准用来确保整个组织的数据表示和数据相关的活动的执行是一致的和规范的<sup>[36,78,79]</sup>。数据质量是指数据在特定情况下满足其使用要求的能力<sup>[80]</sup>。数据安全是指维护有关数据的可访问性、真实性、可用性、保密性、完整性、隐私和可靠性的安全要求<sup>[80-82]</sup>。元数据管理指的是一组管理元数据的流程、程序和工具<sup>[83]</sup>,通过对描述数据的数据进行管理,以实现数据的全生命周期管理<sup>[84]</sup>。数据架构包括企业数据对象的定义<sup>[35]</sup>和在概念、逻辑和物理层面开发企业数据模型<sup>[36]</sup>。数据生存周期指数据获取、存储、整合、分析、应用、呈现、归档和销毁等各种生存形态演变的过程<sup>[77]</sup>。数据流通指通过数据开放共享、交换和交易等方式流通数据<sup>[77]</sup>。数据服务是一系列可重用的操作,可用于访问和转换数据,提供一种可通过 Web 服务进行访问或可对其运行 SQL 查询的统一数据模型<sup>[35]</sup>。数据洞察以满足业务需求为目的,采用机器学习和统计学方法对数据进行分析,如回归分析、分类分析、聚类分析和关联规则分析等<sup>[35]</sup>。

#### 4.2.2 事理维度要素分析

事理维度用于分析数据治理的机理与机制,即数据治理机制,数据治理机制是指建立标准和统一的规则、价值观、程序等,使行为者能够为集体行动建立共识<sup>[85,86]</sup>,其作为数据治理方法的一部分,有助于计划和控制数据管理活动<sup>[36]</sup>。Abraham R 将数据治理机制分为结构机制、程序机制和关系机制。结构性机制决定报告结构、治理机构和问责制<sup>[87]</sup>,包括角色和责任以及决策权的分配。程序性治理机制旨在确保数据被准确地记录、安全地保存、有效地使用和适当地共享<sup>[87]</sup>,包括数据战略、政策、标准、流程、程序、合约协议、绩效衡量、合规性监测和问题管理。关系机制促进了利益相关者之间的合作<sup>[87]</sup>,强调了利益相关方的关系

协调，包括沟通、培训和决策的协调。综合上述分析，本研究认为图书馆用户行为数据治理机制包括结构机制、程序机制和关系机制，结构机制包括角色与责任和决策权的分配，程序机制包括战略、政策、流程、程序、合约协议、绩效衡量、合规性监测、问题管理，关系机制包括沟通、培训与决策的协调。

数据战略是组织开展数据工作的愿景和高阶指引<sup>[36,77]</sup>。政策为数据的创建、获取、存储、安全、质量和访问等提供高层次的指导方针和规则<sup>[36,88,89]</sup>。清晰的数据流程是用于管理数据的标准化、文档化和可重复的方法<sup>[90]</sup>。程序是“文档化的方法、技术和步骤，用于完成特定的活动和任务”<sup>[36]</sup>。合约协议是指数据提供与数据共享是内部部门与外部组织之间达成的合作协议<sup>[91]</sup>。绩效测量旨在通过测量目标实现的水平来评估数据治理的有效性<sup>[81,92]</sup>。合规性监测包括对数据相关专家的监督以及对数据管理项目与服务的监督<sup>[36]</sup>。问题管理指数据相关问题的识别、管理和解决<sup>[36]</sup>。沟通、培训与决策的协调发生在数据治理的主体内部或其之间，高效的沟通机制和反馈机制提供了强有力的保障。

### 4.2.3 人理维度要素分析

人理维度与数据治理的利益相关方有关。DAMA<sup>[36]</sup>将数据治理相关方分为供给者、参与者和消费者。DGI<sup>[35]</sup>认为数据治理中的利益相关者包括创建数据的群体，使用数据的群体以及为数据制定规则 and 要求的群体。综合以上分析，本研究将图书馆用户行为数据利益相关者分为数据供给者、数据管理者和数据消费者，由数据供给者提供数据，数据管理者负责数据生存周期的相关活动，由数据消费者完成利用或决策活动。

## 4.3 数据治理框架构建

结合高校图书馆用户行为数据治理的经验与做法和以上治理要素的分析，构建了初步的治理框架，之后又访谈了来自高校图书馆与数据治理领域的学者与专家，经过深入交流，对框架的细微之处进行了完善。

最终构建出如图 4.1 所示的 WSR 数据治理框架。图中高亮表示的要素表明其是部分高校馆已经关注的重要内容。

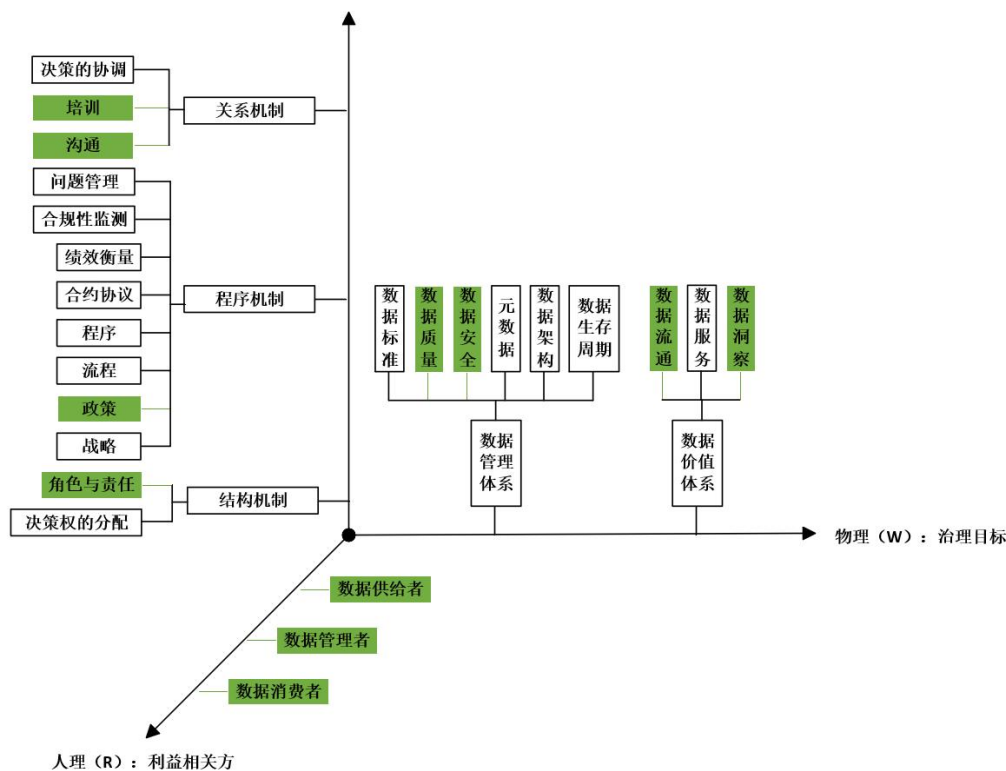


图 4.1 WSR 视域下高校图书馆用户行为数据治理框架

WSR 数据治理框架中各维度的关系可描述为利益相关方 (R) 通过一系列的治理机制 (S) 实现了治理目标 (W)，结构机制与程序机制反映了为实现数据治理目标而设定的一系列规范与规则等，具有一定的强制性；关系机制则与利益相关方密切相关，用以协调数据供给者、数据管理者和数据消费者之间的关系，富有弹性，相对比较灵活。DGI 认为每个数据治理项目都可能涉及多个目标，每一项新工作都应该围绕数据治理生命周期来执行<sup>[35]</sup>，即每一个数据治理目标都有相关的治理机制和利益相关方，例如，数据质量需要考虑政策、数据质量管理岗位设置和数据质量控制的程序等，而在实现数据质量目标的过程中，离不开相关馆员等利益相关方关系的协调，其它数据治理目标的实现亦是如此。



## 5.A 大学图书馆用户行为数据治理实证分析

为了进一步验证 WSR 数据治理新框架的科学性与实践应用的可行性，本章节将以 A 大学图书馆为例展开实证分析，结合其用户行为数据治理现状，分析其治理目标、治理机制和治理相关方，并设计出数据治理平台。

### 5.1 实证研究的思路与方法

A 大学图书馆是一家规模较大、设施完善和服务优质的大学图书馆，拥有先进的双向人脸识别门禁系统、座位预约系统和新一代图书馆服务平台等系统与平台，为广大师生提供便捷的进出馆服务、座位使用服务和图书借还服务等。凭借着先进的管理理念，A 大学图书馆已通过微信公众号和视频等形式对部分行为数据进行可视化分析和故事化描述，并尝试对主要的行为数据进行质量和安全管理，彰显了数据的价值，提升了自身的吸引力和用户满意度。

以 A 大学图书馆用户行为数据治理为例，分析其治理现状与痛点，结合 WSR 视域下的数据治理框架，从物理维度、事理维度和人理维度探究其治理目标、治理机制和利益相关方，以确立清晰的数据治理思路和方案。

### 5.2A 大学图书馆用户行为数据治理现状与问题分析

笔者对 A 大学图书馆运行保障部门进行现场调研并与多位专家和资深馆员深度访谈，提取了主要用户行为数据，并对其清洗与分析，以此深入了解主要用户行为数据及其治理情况。接下来，将从用户行为数据现状、管理现状和利益相关方三个方面进行分析。

#### 5.2.1 用户行为数据现状与问题分析

A 大学图书馆所管理的用户行为数据包括纸本图书的借还数据、用户进出馆数据、座位预约使用数据、数据库资源使用数据、纸本图书预约数据、图书馆主页访问数据和微信公众号文章数据。其中，主要统计分析的数据是图书的借还数据。在这些数据中，借还数据、座位使用数据和进出馆数据的记录数据较为丰富，可以视为主要的用户行为数据资产；而数据库资源使用数据由于技术原因无法关联用户 IP，其他行为数据的规模相对较小。

#### 5.2.2 用户行为数据管理现状与问题分析

针对数据质量方面，目前各业务系统本身承担控制生成数据质量的主要责任。尽管信息化中心部门颁布的《数据管理办法》中明确提出了数据质量的控制要求，但由于各部门尚未充分落实到位且部分馆员缺乏数据治理的意识，并且质量控制的流程与程序尚未标准化，因此各业务系统的用户行为数据可能会存在缺失、错误和异常等问题。

在数据安全方面，存在一定的隐私泄露风险。具体而言，馆员在分析和展示数据报表时，对包含用户个人敏感信息（例如姓名、学号等）的字段的加密与脱敏缺乏控制，这可能导致个人信息泄露。

在元数据方面，不同业务系统的用户基本属性数据存在不一致性的问题，缺乏元数据标准。例如，同一 ID 的用户身份类型和班级可能不一致。此外，各业务系统之间的数据缺乏关联与共享，形成了数据冗余和业务孤岛，增加了数据存储、分析和使用的成本。

针对数据洞察方面，研究支持中心的学科服务馆员主要采用 SQL 和 Excel 对年度数据进行统计分析，并将分析结果以微信公众号的形式发布。但由于数据分析范围有限，同时缺乏标准化的分析指标体系，因此存在提升分析深度和广度的空间。

### 5.2.3 用户行为数据治理相关方分析

用户行为数据治理的主要参与者包括馆务会成员、相关馆员、部分校级信息化中心人员和软件服务商人员。馆务会和校级信息化中心成员在数据治理过程中扮演主要决策者的角色，数据治理项目的发起与重大变更，都需要由其承担相应的责任与风险。相关馆员是数据治理工作的实际完成者，负责各项数据治理活动的实施。软件服务商主要扮演数据供给者的角色，为图书馆提供各类行为数据。

## 5.3 A 大学图书馆用户行为数据治理目标分析

A 大学图书馆数据治理的主要重点是用户进出馆数据、图书借还数据和座位预约使用数据。当前，应将数据质量、数据安全、元数据和数据洞察视为主要治理目标。随着数字基础设施的建立和完善，应逐步关注数据生命周期、数据架构、数据服务和数据流通等目标。

## 5.4 A 大学图书馆用户行为数据治理机制分析

### 5.4.1 数据战略

基于各类用户行为数据及其管理存在的主要问题，制定以下的数据战略。

(1) 明确数据质量、数据安全、元数据和数据洞察等管理者的角色责任和义务，建立高效的决策与沟通机制，以保证数据治理的高效性和有效性。

(2) 建立标准化和正式化的数据治理流程和相关程序，形成正式化的文件，以确保数据治理流程的规范化和持续性。

(3) 扩展用户行为数据分析的思路与方法，尤其是数据分析与人工智能等信息技术的应用，以优化图书馆资源和服务内容的管理，提升用户个性化服务和精准化服务的水平，以满足用户需求。

(4) 监督和评估数据治理活动的效果、效率和效益，根据情况调整和变更计划方案，以保持数据治理活动的可持续性和适应性。

(5) 建立常见问题的知识库，提高工作效率，提升数据治理工作的质量和效率。

### 5.4.2 决策权的分配

A 大学图书馆的用户行为数据治理活动牵涉到多个利益相关方，包括馆务会、相关业务馆员、数据馆员、软件服务商和校级信息化中心。一般情况下，馆务会负责数据治理重大事项的决策和管理，由校级信息化中心的工作人员为相关活动提供指引和协助。考虑到数据治理活动的重要性和规模，可以适当地授权下属部门承担责任，以提高工作效率。目前，A 大学图书馆用户行为数据治理工作处于起步阶段，应根据实际情况建立数据治理办公室或非正式的临时组织，该组织应由各类业务数据的主要负责人组成，包括相关业务馆员、技术馆员和软件服务商等人员。该组织的职责包括整理和规范化相关政策文件，制定正式的用户行为数据治理流程和程序等文件。在日常活动中，应根据实际情况，由运行保障中心、研究者支持中心等部门负责主要的用户行为数据治理活动和工作，由软件服务商等外部组织协助完成相关活动。目前，运行保障中心的馆员负责数据质量、数据安全和元数据等活动，而研究支持中心的馆员负责各业务系统的用户行为数据统计和分析工作。

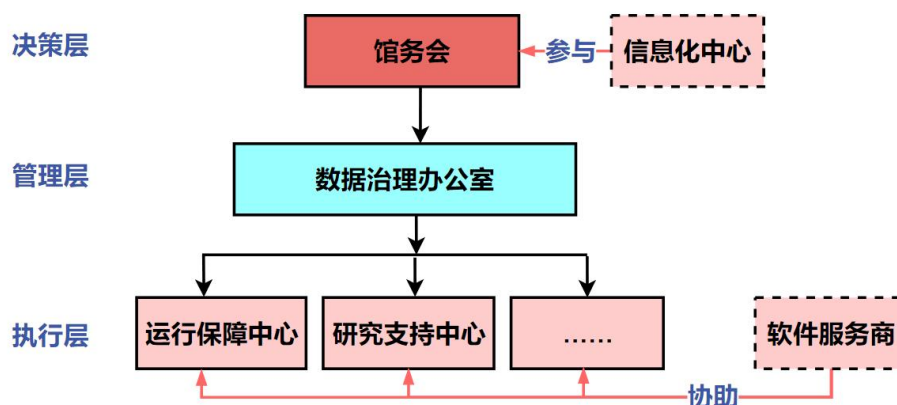


图 5.1 A 大学图书馆用户行为数据治理组织机构

### 5.4.3 角色与责任

为了解决用户行为数据管理存在的问题，数据质量、数据安全、元数据和数据洞察的管理者应承担以下的责任与义务。

表 5.1 用户行为治理的相关角色与责任

Table 5.1 Roles and responsibilities related to user behavior governance

角色	责任
数据质量管理者	(1) 监督用户行为数据的真实性、完整性、规范性与及时性。 (2) 数据的编码应该遵循学校的统一编码规则。 (3) 使用标准化的流程与程序评估和提升数据质量。 (4) 从其它部门获取的数据应在不改变原始数据内容的前提下依据标准和定义来共享、引用和衍生数据，当发现数据质量问题时，应及时反馈。 (5) 形成日常工作的定期报告，包括数据质量问题与改进措施和取得的绩效等内容。
数据安全管理者	(1) 监督用户行为数据安全，及时备份数据，记录、报警和阻断数据库的违规操作行为。 (2) 保障数据的公开与应用的安全，保护用户个人隐私。 (3) 形成日常工作的定期报告，包括数据库异常使用记录等数据安全问题。
元数据管理者	(1) 监督元数据及其模板（制定规范的数据字典）的规范性，定期报告元数据状态。 (2) 监督各业务系统元数据的一致性，发现问题应及时追究原因，及时上报问题并与软件承包方沟通以解决问题。 (3) 形成日常工作的定期报告，包括元数据及其管理问题和工作进展。
数据洞察管理者	(1) 依照规范化的用户行为数据统计分析流程与程序完成日常的报表工作，以新闻、微信公众号文章和短视频等形式发布读者阅读数据报告。 (2) 对数据进行探索性分析，发现用户行为规律，形成分析报告，以拓展用户个性化与精准化服务内容，包括图书推荐和馆藏优化等管理与服务。 (3) 形成日常工作的定期报告，包括数据分析的问题和工作进展等内容。

#### 5.4.4 政策

A 大学图书馆的用户行为数据治理政策与规章制度主要以国家、地方和校级相关政策为基础，并参考了以下表格中列出的相关政策和内容。

表 5.2 用户行为数据治理的相关政策

政策文件名称	公布时间	主要相关内容
《中华人民共和国网络安全法》	2016	(1) 要求制定安全管理制度和操作流程,明确主要机构的负责人及其责任(第二十一、三十四条)。
		(2) 要求采取数据分类、备份和加密等措施保障数据安全(第二十一条)。
		(3) 要求对收集的用户信息严格保密,建立健全用户信息保护制度(第四十、四十一、四十二条等)。
		(4) 建立网络安全监测预警和信息通报制度(第五十一、五十四条等)。
		(5) ……
《中华人民共和国个人信息保护法》	2021	(1) 要求对于个人信息的处理要有合理的目的,处理规则要公开、透明,要提升个人信息的质量(第五、六、七条等)。
		(2) 在实施新闻报道等行为时,应在合理范围内处理个人信息(第十三条等)。
		(3) 制定个人信息的管理制度和操作流程,对个人信息进行分类管理,采取加密、去标识化等安全技术措施(第五十一条等)。
		(4) ……
《高等学校数字校园建设规范(试行)》	2021	(1) 在应用服务方面,要重视用户体验,实现以用户为中心的服务建设与扩展;应加强事务分析型应用系统建设,充分利用数据资源和数据分析、人工智能技术,为用户提供更加智能化的服务。(第 8 小节等)。
		(2) 在网络安全服务方面,要注重基础设施、信息系统和数据等的安全(第 9.5 小节等)。
		(3) 在保障体系方面,要加强组织机构、人员队伍、标准规范和评价体系等的建设(第 10 小节等)。
		(4) ……
《数据管理办法》	2022	(1) 各部门在申请数据使用服务时,应提交申请,按照“按需最少数据字段”原则审核。(第十四条)。
		(2) 在数据质量方面,采集的数据应该真实、完整、规范、及时;数据编码要遵循校级规范(第五章)。
		(3) 数据应用应注意保护个人隐私,只能用于申请

政策文件名称	公布时间	主要相关内容
		授权范围以内的用途（第六章）。
		(4) ……

以上政策文本内容针对 A 大学图书馆用户行为数据治理提出了数据质量、数据安全和数据洞察三个方面的要求。

在数据质量与元数据方面，需要明确数据质量管理者的责任，建立标准化的控制流程。在用户基本信息出现质量问题时，应及时反馈；用户行为数据和相关元数据的编码应遵守信息化中心的规则和方案，以确保不同业务系统之间数据的一致性。

在数据安全方面，需要明确关键数据的主要安全负责人的责任，并建立数据安全的保障制度。在数据使用、共享和应用过程中应注意保护个人隐私，对相关数据进行分类管理，必要时采取加密和脱敏等技术手段。

在数据洞察方面，应注重用户体验，勇于应用数据分析和人工智能等技术，实现以用户为中心，以数据为驱动的图书馆资源管理优化，提升个性化和精准化服务水平。

### 5.4.5 流程

流程设计包含用户行为数据治理流程和各类型用户行为数据的治理流程，数据治理的一般流程是对数据治理项目的整体把握，明确了项目实施过程的关键环节和任务，协调了利益相关方得到主要责任与义务。各类型用户行为数据治理的流程在此基础上又进一步细化与深入，为项目的顺利实施提供保障。

#### (1) A 大学图书馆用户行为数据治理流程设计

该数据治理流程结合了 DGI 等数据治理框架以及高校用户行为数据治理经验，以确保对 A 大学图书馆的用户行为数据进行规范管理。

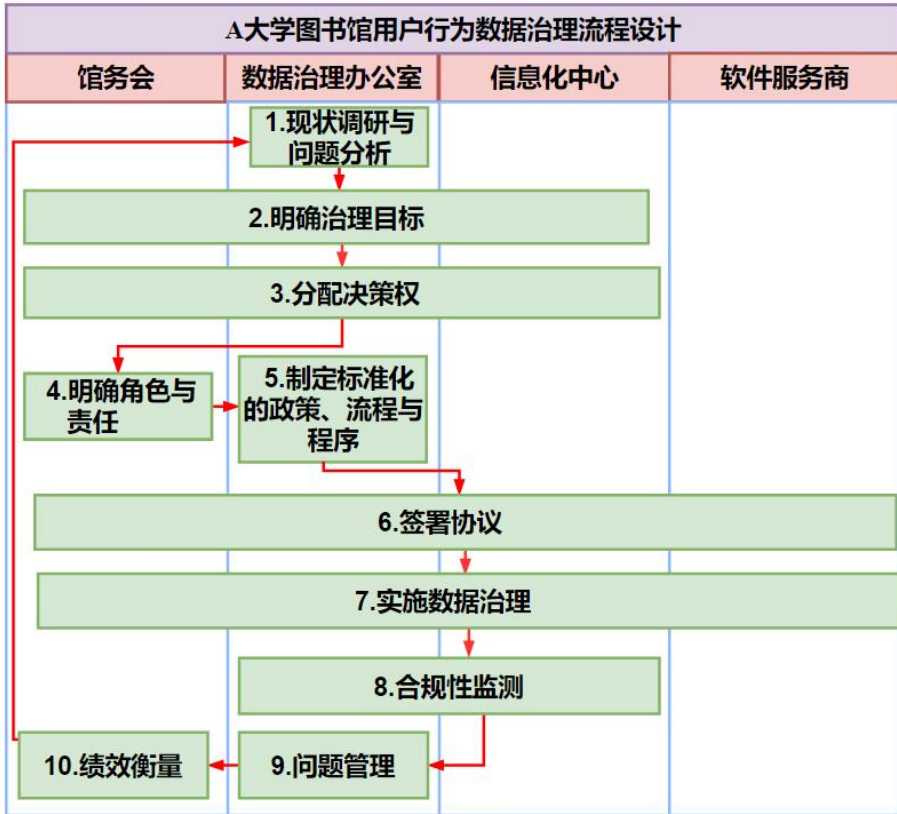


图 5.2 A 大学图书馆用户行为数据治理流程设计

A 大学图书馆的用户行为数据治理应该遵循图 5.2 的一般流程。首先，需要对用户行为数据及其管理现状展开调研分析，数据治理办公室负责归纳整理存在的用户行为数据及其管理问题，形成调研报告，由馆务会审核。接下来，由用户行为数据治理项目的发起人，即馆务会明确治理目标，数据治理办公室和信息化中心相关科室提供建议和对策，以把握治理的关键要素。在理清现实问题和需求后，再由馆务会分配决策权并设立相应的用户行为数据管理角色岗位，为数据治理项目实施提供组织制度保障。之后由数据治理办公室负责调研已有的政策文件，对于与用户行为数据治理相关的内容与要求应尽可能地落实在具体的设计方案和实践工作中去，倘若当前政策不能够满足治理项目的需求，则需要制定相关政策。对于规范化流程与程序制定，应该由具备各类型用户行为数据管理经验的相关馆员负责。在明确数据治理的工作内容后，在必要时，需要和信息化中心、软件服务商等组织与部门签署合同协议。协议规定了双方的权利与义务，有利于提升数据治理项目开展的效率。在数据治理项目实施的生命周期内，应该进行合规性监测，以保障项目的每一项子活动实施没有偏离设定好的目标和已制定的政策、流程与程序等。针对过程中常见的问题以及对应的解决办法，可以建立馆内部的共享知识库，实现经验共享，形成馆内约定俗成的一套问题管理体系，以便于快速解决常见的业务问题，即使馆员换岗或调离后也能够继续使用。在项目实施过程中，还应该对馆员工作绩效和业务效益进行阶段性评估，并建立基于绩效的馆员奖惩制度。最后，用户行为数据治理是一项长期的基础性工作，应随着图书馆用户行为数据规模与结构的变化与增长对数据治理活动相关的计划、政策、流程、程序等进行调整，让数据治理实施更加科学化、合理化、智能化和人性化。

## （2）用户主数据治理流程

用户主数据治理主要包括元数据管理、数据质量管理和数据安全三个模块。在初始阶段，应确认用户数据采集的字段范围，保障仅有得到信息能够为后续的数据分析和挖掘提供强有力的支持，同时也不能采集过多的用户信息，以避免用户信息的泄露和不可预估的损失。接下来，在相关数据规范的基础上，制定详细的数据字典，包括字段的名称、含义、数据类型和计算公式等，对于分类变量应枚举其取值范围，例如，参考 A 大学用户一卡通编码规范，可以建议学号与身份类型的映射规则，进一步设计该字段的自动化清洗程序。接下来，通过自动化的数据质量检查程序检查主数据质量，包括缺失值数量、字段之间的冗余性、异常值和重复值等的检查，如果存在相关字段冗余的情况，在经过协商一致后，将确认冗余字段的取舍，缩小数据的采集范围。待数据自动化清洗完成后，检查是否还遗留有之前未出现的质量问题，此时需要由人工进一步检查数据存在的问题，并由相关人员调查问题产生的原因，最终提出解决问题的措施，倘若有必要，可由技术馆员将常见问题的解决方案设计成自动化处理程序。接下来，需要检查数据集是否存在敏感信息，如果存在的话，应严格遵照人员权限管理数据的查询和使用，甚至是加密数据。

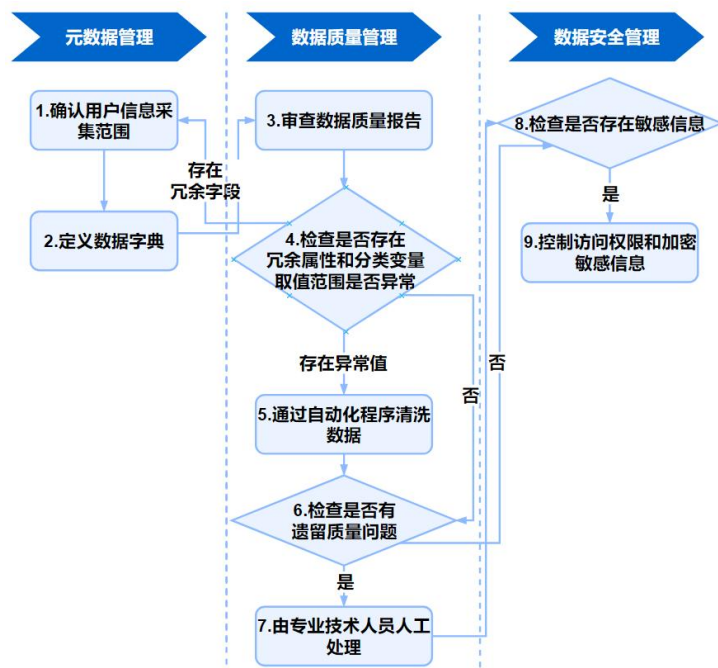


图 5.3 用户主数据治理流程

### (3) 用户行为数据治理流程

用户行为数据治理流程是对用户进出馆数据、用户图书借还数据和用户座位预约数据的治理流程的概括和总结，其相比较于用户主数据的治理，增加了数据洞察这一模块，数据流向和处理步骤也变得更加复杂。

如图 5.4 所示，在元数据管理阶段，在数据字典基础上，还应制定数据统计分析的指标体系，包括指标的含义、计算公式和单位等信息。接下来进入数据质量管理阶段，与主数据质量管理的流程相似，但由于数据信息和结构的不同，具体的数据质量控制的重点和数据清洗的程序等内容会有很大差异，例如主数据需要关注用户身份类型和单位等字段，用户进出馆数据则需要关注进（出）馆时间字段，其间的取值范围和结构不同导致自动清洗的程序设计也会不同。此外，需要关注被自动清理程序过滤的字段和数据，由相关馆员对涉及图书馆与用户等相关方安全利益的异常数值产生的原因展开深入调研。在数据洞察管理阶段，应围绕设定好的统计指标和分析目标，采集相关的数据并进行适当的数据转换，针对部分转换后的数据字段，应再次检查其数据质量，以减少异常数值对数据分析和可视化带来的负面影响，因为隐藏在原始数据中的异常值可能在被计算之后才容易被检测到，待完成质量清洗工作后，应进行数据集成工作，确保相关的数据能够获取和可以使用，倘若有更多字段需求，应及时与信息化中心和软件服务商部门沟通，以获取必要的字段。接下来，应该围绕指标展开维度建模，如按照年、月、日、小时、用户所在单位、身份类型等维度或角度描述用户的行为规律或业务过程，并设计与各维度关联的度量（例如，进（出）馆人次与时间、用户等维度关联）。紧接着，应通过统计分析方法和相关方法对各项数据与指标展开分析，将分析结果转换为可视化图表，最终撰写完成阅读报告或年度数据报告等。注意，如果要在图表中与用户相关的敏感信息，需要采取必要的脱敏手段，例如学号、姓名等字段。

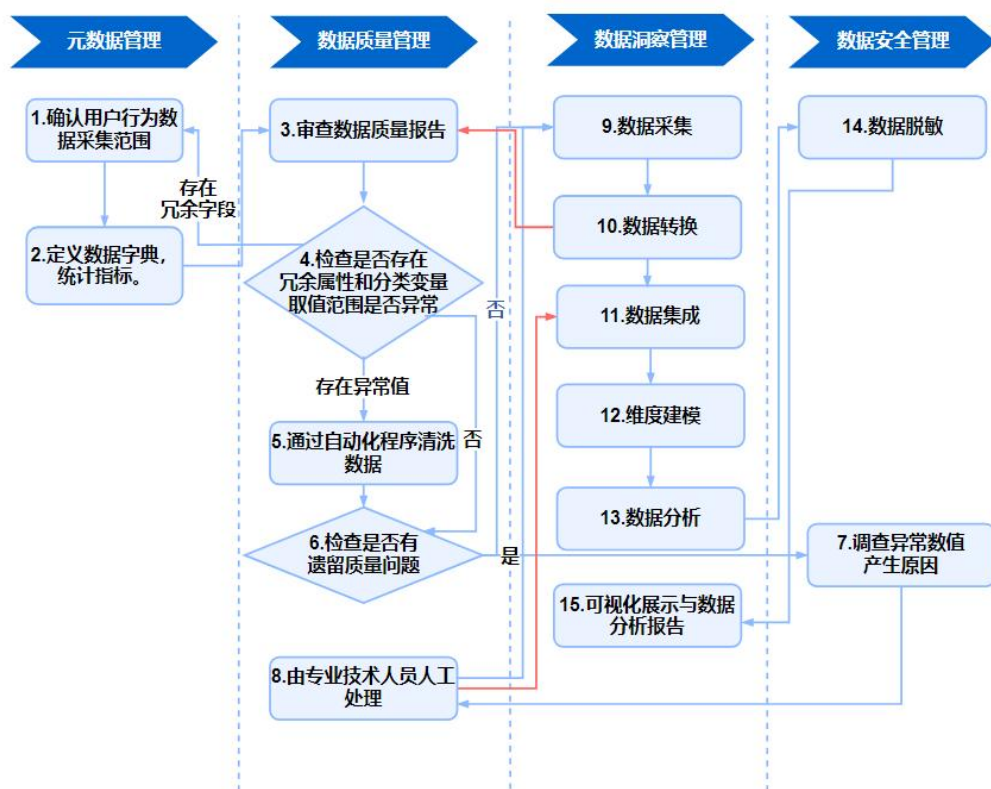


图 5.4 用户行为数据治理流程

### 5.4.6 程序

A 大学图书馆主要收集用户进出馆数据、图书流通借还数据和座位预约使用数据，这些数据需要经过质量评估和加工处理，同时需要进行数据加密和脱敏来保护用户隐私，最终通过数据洞察程序进行分析和挖掘。以 2020 年 9 月 1 日至 2021 年 1 月 20 日期间的 A 大学图书馆用户行为数据为例，进行规范化程序设计。在规范化程序设计的过程中，需要考虑数据质量的标准和要求，对数据进行清洗和转换，以确保数据的一致性和准确性。同时，还需要为数据加密和脱敏设计相应的程序，以保护用户隐私和数据安全。最后，将运用数据洞察程序来挖掘数据中的有价值信息，为图书馆的管理和服务提供有力支持。

#### (1) 用户主数据治理程序

用户主数据治理程序由元数据管理、质量管理和安全管理三个模块组成。

##### ① 用户主数据的元数据管理

A 大学图书馆的用户基本信息数据存在于各业务系统中，主要包括用户证件号、姓名、单位、性别和身份类型等基本信息，但是各业务系统的基础信息存在不一致的情况，例如，与用户“证件号”相同含义的字段还有“一卡通号”，“读者账号”，‘USER\_ID’等；同一字段的取值范围不同，例如，参照的规范不同导致同一用户的身份类型在不同系统中不一致，同时，不同业务系统的用户身份类型唯一值总数不同。为了应对以上问题，需要建立规范统一的用户主数据字典，其中，数据类型以 MySQL 参考，业务含义对字段进行了解释，并枚举了分类变量的唯一值。

表 5.3 用户主数据字典

序号	字段名称	示例值	数据类型	业务含义
1	证件号	1111111111	CHAR(10)	用户的一卡通号，由 10 位数字组成。编码



序号	字段名称	示例值	数据类型	业务含义
2	性别	男	VARCHAR(2)	规则参考 A 大学一卡通编码规范。 用户性别信息，取值范围：男、女。
3	单位	法学院	VARCHAR(60)	用户所在单位信息，取值范围较广，包括图书馆、法学院、机械工程学院、海外教育学院。
4	年级组	2001	CHAR(4)	用户所在年级组，由年份和班级组成，例如，2001，表示 2020 年入学，该用户属于 1 班。
5	读者类型	硕士研究生	VARCHAR(20)	用户身份类型，参照 A 大学一卡通编码规范，最终将用户识别为 10 类用户，分别是教职工、博士研究生、硕士研究生、本科生、京江学院、国际交流学院、外教、教职工-劳务租赁、教职工其他类人员和临时人员。

### ② 用户主数据的质量管理

基于以上的用户主数据字典，通过编写 Python 程序控制分类变量的唯一值范围的一致性，同时，为了保证各业务系统用户信息的一致性，将单独存储用户数据信息，在必要时，采用数据库查询连接的方式补充用户基本信息数据。

以下函数通过字典映射的方式控制用户身份类型，根据用户一卡通号起始数字前 2 至 3 位数字划分用户身份类型，最终将用户身份类型限制为 10 大类，需要使用 Numpy 库。

```
def map_user_type(x):
    if x.startswith('1'):
        return '教职工'
    elif x.startswith('211'):
        return '博士研究生'
    elif
x.startswith('222')|x.startswith('223')|x.startswith('224')|x.startswith('225')
|x.startswith('226')|x.startswith('227')|x.startswith('228'):
        return '硕士研究生'
    elif x.startswith('3'):
        return '本科生'
    elif x.startswith('4'):
        return '京江学院'
    elif x.startswith('510'):
        return '国际交流学院'
    elif x.startswith('55'):
        return '外教'
    elif x.startswith('598'):
        return '教职工-劳务租赁'
    elif x.startswith('599'):
        return '教职工其他类人员'
    elif x.startswith('9'):
        return '临时人员'
    else :
        return np.nan
```

图 5.5 用户身份类型数据质量控制

### ③ 用户主数据安全控制

用户主数据的安全控制主要从字段采集和访问权限控制两个方面考虑。在采集字段方面，尽可能使用更少的字段满足数据分析相关的需求，同时限制不同用户查询、修改等操作的权限；做好用户一卡通号、姓名、身份证号等重要字段的安全保障措施。

## (2) 用户进出馆数据治理程序

### ① 用户进出馆数据元数据管理

从 A 大学门禁管理平台导出的进出馆数据记录包含 15 个字段，通过 `Pandas_profiling` 初步审查数据质量报告，以检测各字段之间的相关性，如果相关性过高，可由馆员再次凭借业务知识判断两字段之间是否冗余，最终保留其中一个字段即可，在这个数据集中，事件和闸机编号表现出高度的相关性，这里选择保留闸机编号字段。除了用户基本属性字段外剩 6 个字段（保留一卡通号字段），各字段名称及其业务含义如表 5.4 所示。

表 5.4 用户进出馆数据字段

序号	字段名称	示例值	数据类型	业务含义
1	时间	2021-08-31 21:57:00	DATETIME	用户进馆或出馆的时间。
2	学号	1111111111	CHAR(10)	用户一卡通号或学工号。
3	类型	进门	VARCHAR(4)	用户进馆与出馆事件类型，取值范围：进门、出门。
4	验证方式	人脸	VARCHAR(6)	用户进馆与出馆认证方式，取值范围：人脸、二维码和一卡通。
5	闸机编号	E01	CHAR(4)	用户进馆或出馆经过的闸机编号。E01-E06 表示进馆的闸机号，E07-E12 表示出馆闸机号。

### ② 用户进出馆数据质量管理

用户进出馆数据重点关注的字段是时间字段，其它分类属性需要通过唯一值范围检测，借助 `Pandas_profiling` 工具，可以快速实现这一目标，它还可以检查字段缺失值及其之间的关联情况，图书馆管理人员通过自动化工作生成的数据质量可视化报告，可以发现质量问题。

针对时间这一字段，将结合 A 大学图书馆对外提供的进馆服务的时间范围和同一时间范围内用户进馆与出馆记录数量相同的特点对原始数据进行过滤和筛选。为了应对用户进出馆数据在同一天内进馆记录数和出馆记录数量不相等的问题，将借助 `Pandas` 库来预处理数据，具体代码如图 5.6 所示。

首先，将原始数据按照“学号”和“时间”字段进行升序排列，如果是正常的记录，用户的进馆记录和出馆记录将顺序排列，即上一条是用户进馆记录，紧接着是出馆记录、下一条进馆记录等。接着，借助 `Pandas` 的 `shift` 函数，将“时间”、“类型”，“学号”和“索引”列整体向上移动一行，将获得当前记录的下一条记录相关字段信息，这样便于使用比较运算符判断相邻的两条记录是否为同一用户同一天生成且上一条为进馆记录，下一条为出馆记录。最终，删除不必要的字段，将正常记录和存在安全隐患的记录分别存储在两个文件。

```

def filter_records_by_time(df):
    df = df.sort_values(by=['学号', '时间'])
    df['时间1'] = df['时间'].shift(-1)
    df['类型1'] = df['类型'].shift(-1)
    df['学号1'] = df['学号'].shift(-1)
    df['索引1'] = df.index.to_series().shift(-1).dropna().astype('int64')
    dfuser = df[((df['类型'] == '进门') & (df['类型1'] == '出门')) & (df['时间'].dt.date == df['时间1'].dt.date) & (df['学号'] == df['学号1'])]

    # 保存符合条件记录的索引
    df_index = pd.concat([dfuser.index.to_series(), dfuser['索引1']], axis=0).dropna()
    # 保留有效字段
    df = df[['时间', '学号', '类型', '验证方式', '闸机编号']]
    df_safe = df.loc[df_index]
    df_unsafe = df.drop(index=df_index)

    df_safe.sort_values(by='时间', ascending=True, inplace=True)
    df_safe.reset_index(inplace=True, drop=True)

    df_unsafe.sort_values(by='时间', ascending=True, inplace=True)
    df_unsafe.reset_index(inplace=True, drop=True)
    return df_safe, df_unsafe

```

图 5.6 用户进出馆时间数据预处理

在完成时间字段的数据预处理后，将通过可视化的方式检查进馆与出馆记录的年、月、日、小时人次分布，如图 5.7 所示，负责进出馆数据质量管理的馆员可以拖动图表下方的滑块来选择日期范围，以快速发现时间字段被清洗后的情况，在每个图表的右上角提供了交互式按钮，支持下载和数据查看功能。

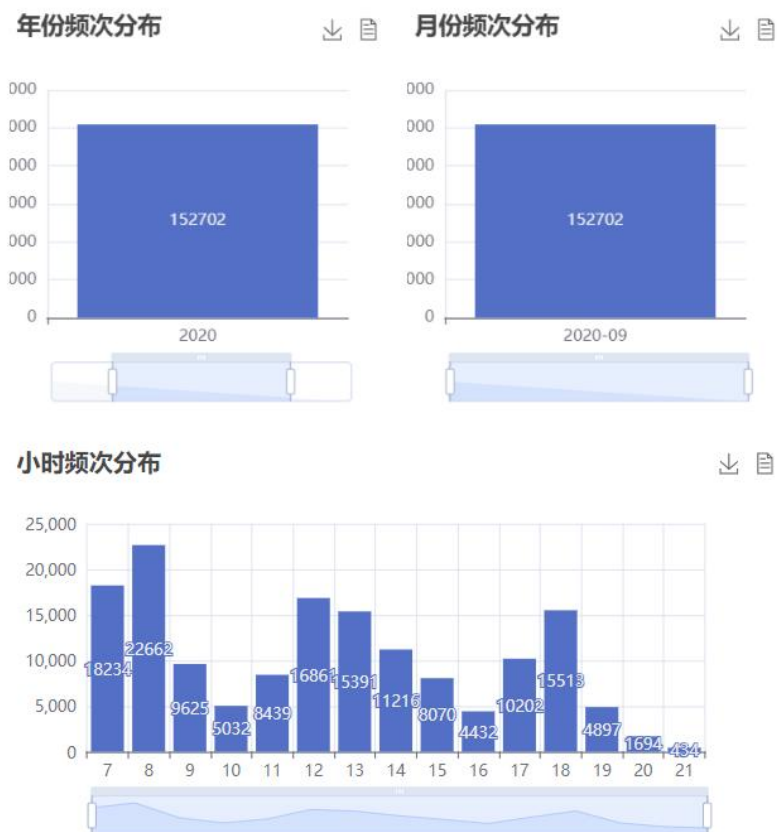


图 5.7 进馆与出馆记录时间的频次分布

④ 用户进出馆数据洞察管理

用户进出馆数据的洞察以统计分析为主，参考北京大学、武汉大学等高校的阅读报告，制定以下的统计指标，随着图书馆业务与技术的拓展与提升，统计指标应进行适当的调整和扩充。

表 5.5 用户进出馆数据统计指标

序号	指标名称	单位	含义与计算方法
1	总进（出）馆人次	人次	统计一段时间范围内的用户总进（出）馆人次。
2	每年（月、日）进（出）馆人次	人次	按年（月、日）统计用户进（出）馆人次。
3	每年（月）男生和女生进（出）馆人次	人次	按年（月）和性别分组统计用户进（出）馆人次。
4	进（出）馆总人次 Top10 学院	人次	按学院分组统计进（出）馆总人次。
5	月进（出）馆人次 Top3 学院	人次	按月份和学院分组统计进（出）馆总人次。
6	各类型用户进（出）馆人次	人次	按用户身份类型分组统计进（出）馆总人次。
7	个人在馆时长 Top10	小时	统计每位用户在馆累计时长，取前 10 位读者。
8	平均在馆时长 Top10 单	小时/人	统计各单位的平均个人在馆时长，计算公

序号	指标名称	单位	含义与计算方法
	位		式: 各学院在馆总时长/各学院进馆人数。
9	各时刻进馆人次	人次	统计各时刻进馆人次, 时间范围为 7:40 至 21:40。
10	各时刻出馆人次	人次	统计各时刻出馆人次, 时间范围为 7:40 至 21:40。

#### ⑤ 用户进出馆数据安全

用户进出馆数据涉及的用户敏感信息是学号, 需要合理控制访问用户进出馆数据的访问权限; 个别指标需要展示用户学号或姓名等信息, 可以采用加密的措施, 例如, 学号字段, 可以用“\*”替代第 3 至 6 位数字, 最终展现在用户眼前的是“20\*\*\*\*2003”; 如果是姓名字段, 可以采用用户姓名加称呼的方式, 例如“刘同学”, “刘老师”。

#### (2) 用户座位预约数据治理程序

##### ① 用户座位预约数据元数据管理

从 A 大学图书馆座位预约管理平台导出的数据文件包含 17 个字段, 除了用户基本属性信息外 (保留一卡通字段), 还包括 13 个字段, 各字段名称及其业务含义如所示。

表 5.6 用户座位预约数据字典

序号	字段名称	示例值	数据类型	业务含义
1	学号	1111111111	CHAR (10)	用户一卡通号。
2	预约时间	2020-09-03 13:06:38.263	DATETIME	用户预约使用座位的时间。
3	开始时间	2020-09-03 13:06:38.263	DATETIME	用户预约开始使用座位的时间。
4	结束时间	2020-09-03 13:06:38.263	DATETIME	用户预约结束使用座位的时间。
5	签到时间	2020-09-03 13:06:38.263	DATETIME	用户使用座位签到时间。
6	签退时间	2020-09-03 14:06:38.263	DATETIME	用户离开座位签退时间。
7	使用时长	60.0	DECIMAL (6, 2)	用户使用座位时长=签退时间-签到时间, 单位为分钟。
8	空间名称	1	VARCHAR (4)	用户使用的座位编号, 各个空间区域座位数量不同。
9	区域名称	自然科学图书 借阅 A 区	VARCHAR (40)	用户使用的座位所在空间区域, 取值范围: 自然科学图书借阅 A 区, 自然科学图书借阅 B 区, 二楼电子阅览室, 社会科学图书馆借阅 A 区, 社会科学图书馆借阅 B 区, 社会科学图书馆借阅 C 区-1, 社会科学图书馆借阅 C 区-2, 外文图书阅览室, 教师研究生阅览室, 学术期刊阅览室, 特色文献阅览室, 工具书阅览室。
10	使用状态	已使用	VARCHAR (10)	用户使用座位状态, 取值范围: 已使用、已关闭、用户取消。

序号	字段名称	示例值	数据类型	业务含义
11	预约方式	座位现场预约	VARCHAR(20)	用户预约使用座位方式，取值范围为座位网上预约、座位现场预约。
12	违约类型	未违约	VARCHAR(20)	用户违约的类型，取值范围：未违约、未签到、临时离开超时、临时离开未刷卡、未签退。
13	违约时间	2020-09-03 13:06:38.263	DATETIME	用户违约的时间。

#### ②用户座位预约数据质量管理

用户座位预约数据重点关注的字段是时间相关字段，例如“签到时间”，“签退时间”。为了控制数据质量与安全，需要初步筛选签到时间与签退时间的时间范围，记录产生的范围应该与图书馆座位开放使用的时间范围一致，即保留原始数据集中 7:40 至 21:40 产生的记录，并限制年份、月份和日期范围。在完成自动清洗后，应以可视化的方式展示年、月、日和小时等时间维度下的用户签到（离）次数分布图，与用户进出馆数据的“时间”字段检查类似，但是增加了更多的 DATETIME 类型字段，包括“签到时间”、“签退时间”、“预约时间”、“开始时间”和“结束时间”。

#### ③用户座位预约数据洞察管理

用户座位预约数据的洞察以统计分析为主，在参考北京大学、武汉大学等高校的阅读报告后，制定以下的统计指标。

表 5.7 用户座位预约数据统计指标

序号	指标名称	单位	含义与计算方法
1	座位预约次数 Top10 用户	预约次数	统计一段时间范围内的用户预约座位的总次数。
2	每年（月、日）座位预约人次	人次	按年（月、日）统计用户座位预约人次。
3	各时刻座位签到（退）人次	人次	按年（月、日）统计用户座位签到人次。
4	座位使用时长 Top10 用户	小时	统计一段时间范围内的用户使用座位总时长，计算公式：签退时间-签到时间，注意单位换算。
5	使用次数 Top10 座位	人次	统计一段时间范围内各个区域座位的使用次数，由区域名称和空间名称形成组合键，再分组统计。
6	各区域座位预约使用次数	人次	统计一段时间范围内图书馆各区域座位预约使用的人次。
7	座位预约次数 Top10 单位	预约次数	统计一段时间范围内各单位座位预约次数。
8	男（女）用户座位预约次数	预约次数	统计一段时间范围内男（女）用户座位预约次数。
9	各身份类型用户座位预约次数	预约次数	统计一段时间范围内各类型用户座位预约次数的总和。

#### ④用户座位预约数据安全

用户座位使用数据安全与进出馆数据安全类似，重点关注的一卡通号和姓名等

字段。对于在质量清洗阶段被自动过滤的数据，应该调查异常数值产生的系统原因，如果是硬件系统和系统软件造成的，需要及时和软件服务商沟通以修复和更新相关软件与硬件；造成异常数值的原因还包括图书馆特殊阶段所开展的服务内容的变更、用户违反了图书馆的服务规定。

#### (4) 用户借阅数据治理程序

##### ① 用户借阅数据元数据管理

用户借阅数据主要包含题名、作者、索书号、财产号、出版社、读者账号、出版社、借书时间、还书时间、应还时间、续借时间、操作类型(0 借, 1 续借, 3 还)、事件创建时间等字段，经过初步的数据质量审查，将操作类型、事件创建时间字段不纳入采集范围，因为操作类型字段的数值分布极不均衡，且存在一定的错误，但是对于后续的分析 and 挖掘工作几乎没有影响；事件创建时间是用户的借书时间或还书时间，造成了字段间的冗余。数据字典如表 5.8 所示。

表 5.8 用户借阅数据字典

序号	字段名称	示例值	数据类型	业务含义
1	题名	交通管理与控制	VARCHAR(50)	用户借阅书籍的题名。
2	作者	罗霞,刘澜,刘海旭	VARCHAR(40)	用户借阅书籍的题名。
3	索书号	U491/L98=2	VARCHAR(20)	索书号用于标识书籍的馆藏位置,方便图书馆管理人员对书籍分类、定位和管理。
4	财产号	2017056313	CHAR(10)	每一本书籍的唯一标识码,由 10 位数字组成。
5	出版社	人民交通出版社股份有限公司	VARCHAR(40)	用户借阅书籍的出版社。
6	读者账号	1111111111	CHAR(10)	借还图书的用户账号,和一卡通号相同,由 10 位数字组成。
7	借书时间	2020-09-01 09:01:00	DATETIME	书籍借阅时间。
8	还书时间	2020-11-17 16:12:49	DATETIME	用户归还书籍的时间。
9	应还时间	2020-11-14	DATETIME	用户所借阅图书的到期时间。
10	续借时间	2020-10-15 16:06:18	DATETIME	用户在借阅图书到期之前可以延长借阅时长的时间。

##### ②用户借阅数据质量管理程序

用户借阅数据重点关注的字段是作者，出版社，读者借书时间，还书时间，续借时间。时间字段的清洗与检查和其它行为数据的时间字段处理方法相似，即通过 Pandas 的 `to_datetime` 函数将相关字段转化为时间类型，再通过布尔表达式筛选数据记录。

在原始数据中，作者字段包含“编著”、“主编”、“...”、“[等]”和“著”等信息，这些信息的实际意义价值不大，可以将其删除，以节约存储空间，并为后续的数据分析与挖掘工作打下基础。借助正则表达式，将这些无关信息替换为空字符串，还要注意清除字符串两边的空白字符，具体代码如所示。

```

def process_author(df):
    df["作者"] = df["作者"].str.strip().replace(r'编著|主编| |\.\.\.\| \[等\]', '', regex=True).str.replace(r'著$', '', regex=True)
    return df

```

图 5.8 用户借阅数据作者字段清洗

出版社字段的部分数值末尾存在“、”、“:”等干扰信息，同样需要使用正则表达式替换为空白字符串。

```

def process_publisher(df):
    df['出版社'] = df['出版社'].str.replace(r',$|:$', '', regex=True)
    return df

```

图 5.9 用户借阅数据出版社字段清洗

Figure 5.9 User borrowing data publisher field cleaning

③用户借还数据洞察管理

通过阅读其它馆的年度阅读报告，发现用户借还数据是这些馆重点分析和描述的对象。主要的统计指标包括热门图书、热门作者、热门出版社、书籍借阅次数 Top10 用户等指标。具体的统计指标如所示。

表 5.9 用户借阅数据统计指标

序号	指标名称	单位	含义与计算方法
1	借阅次数 Top10 用户	借阅次数	统计一段时间范围内用户借阅书籍次数。
2	借阅次数 Top10 书籍题名	人次	统计一段时间范围内每种书籍的借阅人次。
3	借阅次数 Top10 书籍作者	人次	统计一段时间范围内每位作者所出版书籍的借阅人次。
4	借阅次数 Top10 出版社	人次	统计一段时间范围内每个出版商所出版书籍的借阅人次。
5	每年（月、日）借阅人次	人次	按年（月、日）分组统计书籍借阅次数。
6	每年（月、日）还书人次	人次	按年（月、日）分组统计还书人次。
7	平均借阅时长 Top20 书籍	天	统计一段时间范围内每种图书的借阅时长，计算公式：平均借阅时长=sum(还书时间-借书时间)/借阅次数。
8	借阅次数 Top10 单位	人次	统计一段时间范围内各单位借书次数。
9	男（女）用户图书借阅次数	借阅次数	统计一段时间范围内男（女）用户书籍借阅次数。
10	各身份类型用户书籍借阅次数	借阅次数	统计一段时间范围内各类型用户图书借阅次数的总和。

④用户借还数据安全



用户借还数据的管理应关注“一卡通号”和“姓名”等敏感信息，和借阅数据、座位预约数据的处理方式相似。

#### 5.4.7 合约协议

数据治理的生命周期内离不开利益相关方的参与，A 大学用户行为数据治理涉及的相关方主要是校园的信息化中心和门禁系统、座位预约管理系统以及流通系统的开发商（系统软件服务提供方）。其中，用户信息数据相关字段的获取与补充需要通过和信息化中心部门协商，数据质量、安全和洞察相关的要求已经在信息化中心颁布的数据管理规定中有所体现，与此同时，图书馆和信息化中心有着共同的目标，即提升数据的质量与价值，推进智慧校园的建设与发展，所以图书馆很容易和信息化中心等相关部门达成一致意见，待数据基础设施逐渐成熟后，应考虑制定规范的合约协议，以明确双方的权利与义务，保障用户基本信息数据共享和使用的透明和合规性。

对于各业务平台产生的行为数据的治理，需要软件服务商的参与和支持，因为用户行为数据记录的原始定义、数据结构和取值范围等都是其设定的，这些设定对数据质量产生了极大的影响，后续可能会产生字段冗余、数值取值范围不一等问题，甚至有些数据问题是由于硬件与设备及其现场环境导致的。以上问题的出现，均需要由相关馆员和服务商的技术人员或相关代表及时沟通，以从源头治理数据，减少数据的计算、存储和人力资源等成本。随着图书馆业务、服务内容相关政策的调整与变更，原先的数据模型也应该及时得到修正，以适应现行的标准和规范。所以，无论是旧的管理系统还是新一代服务管理平台，都应与服务商签订合同协议，当出现任何数据质量、分析和安全等方面的问题时，才会有所保障。

#### 5.4.8 绩效衡量

A 大学图书馆在进行用户行为数据治理时，主要从治理效果、效率和效益三个方面来考虑绩效衡量。在治理效果方面，可以采取定量的方法来评估治理后的数据质量，例如检查缺失率、重复值和异常值等；同时也可以从社交媒体的视角进行评估，通过对用户行为数据的可视化分析和故事化阐述来考核图书馆官方微信公众号、视频号等媒体平台的粉丝数量和阅读量是否有提升。在效率方面，可以采取定性和定量相结合的方式进行评估，以工作总结的形式进行汇报评估。在效益方面，可以通过问卷调查用户的满意度，也可以比较图书馆进馆人次、图书借阅次数和座位预约次数等指标的变化来评估效益。这些绩效评估的结果将有助于制定和调整治理策略，推动图书馆的发展和服务水平提升。

#### 5.4.9 合规性监督

在数据治理项目的实施过程中，对项目的执行过程进行定期的检测和评估是至关重要的。这有助于确保项目按照设定好的流程和程序进行实施，如果发现项目的实施偏离目标，需要及时反馈意见，以获取上级的支持，及时弥补错误，并根据需要做出适当的调整。

此外，在数据治理过程中，还需要检测各类治理工具和平台的操作和使用是否符合规范。不规范的操作和使用可能会导致数据质量和安全问题，严重影响数据治理项目的效果。因此，要加强对各类治理工具和平台的规范化管理，避免因操作不当带来的潜在风险。

除此之外，数据治理的绩效还受到各原始业务系统操作规范的影响。因此，需要重视各原始业务系统的操作规范，从源头减少错误数据的产生，以此降低数据治理的计算和人力成本等。这可以通过加强对业务系统的监督和管理来实现，确保数据的质量和可靠性，进一步提高数据治理的效率和效果。

总之，在数据治理项目的实施过程中，定期检测和评估项目的执行过程、加强对治理工

具和平台的规范化管理，以及重视各原始业务系统的操作规范，这些措施可以帮助保证数据治理的绩效。此外，在实施过程中，还可以采取一些其他的措施，如建立有效的风险管理体系、提高数据质量和安全性的意识和能力、加强员工培训等，以进一步提高数据治理的效果和效率。

#### 5.4.10 问题管理

在数据治理项目实施的过程中，问题管理也是至关重要的一环。A 大学图书馆主要从以下三个方面考虑。

首先，针对常见的数据操作问题，可以通过规范化解决方案，提高解决问题的效率。常见的数据操作问题包括数据质量问题、数据安全问题等。通过对这些问题的解决方案进行规范化，可以形成图书馆内部一套约定俗成的解决方案，为数据治理项目提供有力支撑。此外，设计容易操作的低代码平台，也能够提高解决问题的效率，让非技术人员也能够快速解决一些常见的数据操作问题。

其次，在数据管理层面也需要类似的规范化解决方案。数据管理问题可能涉及到权利、责任方面，例如，当出现数据问题时，应该由谁负责，处理问题的流程与程序如何。通过规范化解决方案，可以让数据管理更加高效、标准化。

最后，建立简易的问答系统或常见问题手册，能够解决数据治理项目中的大部分常见问题。一些常见的问题，如数据格式错误、数据缺失、数据重复等，可以通过问答系统或常见问题手册得到解决。这不仅能够提高解决问题的效率，还能够为数据治理项目中的新人员提供基础知识和操作指南，帮助其更快速地适应工作。

总之，问题管理是数据治理项目中不可或缺的一环。通过规范化解决方案、设计容易操作的平台、建立问答系统或常见问题手册等方法，可以提高数据治理项目的效率，降低出错率，提升整个图书馆数据治理的水平和质量。

#### 5.4.11 沟通

在数据治理中，沟通是非常重要的环节，不同层面的沟通能够确保数据治理的有效实施。

首先，图书馆内部人员之间的沟通非常关键。在数据治理过程中，会涉及多个部门，包括运行保障中心和研究支持中心，当用于可视化分析的行为数据出现质量、安全问题时，研究支持中心与运行保障中心馆员应及时进行沟通，使得数据能够满足业务需求；此外，运行保障中心也应该在出现质量、安全等隐患时，给与研究支持中心人员适当的提示。

其次，领导与馆员之间的沟通也是非常重要的。馆领导应该传达用户行为数据治理项目价值与意义的重要性，并听取馆员的反馈与建议，对计划与方式做出适当的调整；馆员需要理解数据治理的政策、流程与程序等，并遵守相关规定。通过双向沟通机制和反馈机制推进数据治理进程。

同时，图书馆还需要与信息化中心、服务商等外部组织进行沟通。在这种情况下，图书馆需要确保双方能够理解彼此的需求和限制，并在合作过程中进行充分的沟通和协作。

最后，图书馆还应该积极收集来自图书馆用户的反馈信息。广大师生作为图书馆各类服务的消费者，其提出的各项建议与请求，相关馆员应尝试通过对各类行为数据的治理以满足其个性化的需求，通过用户的反馈机制使得图书馆的各项政策与服务得以调整和优化，实现用户满意度的持续提升。

总之，在数据治理中，沟通是确保有效实施的关键环节。

#### 5.4.12 培训

培训对于保障数据治理的有效实施至关重要。

首先，馆领导和相关管理人员应向下属传达数据治理项目的意义和价值所在，并详细介绍与展示相关政策、流程与程序，减少因操作不当带来的数据问题，并提升项目实施的规范性。

其次，在培训中还需要强调注意点和常见问题，讲解数据治理过程的重要环节和操作要点，提醒培训对象注意事项和应对策略。同时，还可以根据馆内数据治理实践中出现的常见问题和管理问题，进行针对性的周期性培训，提高馆员的数据素养和数据治理能力。这些培训可以通过讲座、培训班、在线学习等多种方式进行，让馆员掌握更多的数据治理知识和技能，为数据治理的有效实施提供有力保障。

总之，数据治理培训应该从数据治理项目的意义、流程与程序、操作规范、注意点与常见问题等方面入手，为培训对象提供全面、系统的数据治理知识和技能，为数据治理的有效实施提供有力保障。

#### 5.4.13 决策的协调

决策协调在数据治理项目的实施和重大变更中扮演着至关重要的角色。然而，A 大学图书馆用户行为数据治理项目目前处于起步阶段，缺乏专业的数据治理组织，因此，相关的决策权主要由馆务会承担与分配。为确保项目的顺利开展，需要采取一系列的决策协调措施。

首先，需要加强馆务会成员与馆员之间的决策协调。在项目开展过程中，难免会面临各种问题和困境，如果馆员无法承担变更所带来的影响，他们应该及时上报领导，并经过共同协商后，做出适当的安排和调整。

其次，馆务会还需要与信息化中心等组织和代表进行干涉。这些组织和代表在项目实施中扮演着重要角色，可以为用户行为数据治理活动提供建议和技术协助，因此与他们的紧密合作和沟通十分必要。领导应该建立一个开放性的交流渠道，定期召开会议，促进双方之间的沟通和合作。

### 5.5 A 大学图书馆用户行为数据治理利益相关方分析

A 大学图书馆用户行为数据治理涉及的相关方主要包括馆务会成员、相关馆员、图书馆用户、信息化中心和软件服务商。

数据供给者主要包括图书馆用户、信息化中心和软件服务商。首先，高校馆中的各类用户行为数据来源于用户和各项服务的交互活动，用户是行为数据的最大贡献者。其次，为了实现数据洞察的目标，需要获取更加规范和全面的用户基本信息，由此不得不向校级信息化中心提出数据使用申请。最后，部分行为数据的获取需要根据合约协议由软件服务商传输给相关馆员。

数据管理者主要包括相关馆员和软件服务商。来自运行保障中心和研究支持中心的相关馆员负责数据的日常质量、安全、元数据和洞察工作，在必要时，软件服务商应按照合约协议对相关数据的质量、安全进行管理。

馆务会、图书馆用户和相关业务馆员扮演了数据消费者的角色。首先，由数据洞察阶段生成的数据报告反映了图书馆一段时间内的业务情况，馆务会和相关业务馆员可由此了解馆内相关服务情况，师生用户则可以通过报告了解学院与学校的阅读情况。其次，在进行数据质量与安全控制时，相关业务馆员可以根据数据问题推测出各项服务存在的问题，并及时汇报领导，并做出适当的优化与调整。

## 5.6 A 大学图书馆用户行为数据治理平台设计

结合以上对 A 大学用户行为数据治理目标、机制和利益相关方的分析，基于 Streamlit 设计出简易数据治理平台。Streamlit 可以快速构建 Web 应用，提供了多种交互式组件，通过这些组件，用户可以轻松完成文件上传、交互式图表多种功能。此外，Streamlit 可以轻松的部署到云服务器上，可以与相关馆员分享设计好的 APP。

### 5.6.1 功能简介

基于 Streamlit 的数据治理平台提供了丰富的数据治理功能，能够满足多项核心治理需求，该平台提供了用户基本信息、借阅、进出馆和座位预约等数据的质量审查功能，结果以可视化的方式呈现。通过查看数据报告，可以清晰地了解各种数据的质量情况。此外，该平台还提供各种数据的质量加工功能，并以交互式图表的形式展示数据清洗后的结果。如果您发现数据仍存在问题，可以再次选择相应模块进行数据清洗。平台还汇集了各种用户行为数据的主要统计指标，馆员只需要上传相应的数据文件，便可一键实现自动化图表。这些图表提供了交互功能，可以根据需要选择相应指标的图表下载。



图 5.10 数据治理平台功能简介

如图 5.10 所示，数据治理平台包含七个模块，分别是简介、用户基本属性数据、用户借阅数据、用户进出馆数据、用户座位预约数据、统计指标分析和实用小工具。简介模块对数据治理平台的功能进行了简要概述，用户基本属性数据、用户借阅数据、用户进出馆数据与用户座位预约数据模块提供了对应数据的质量审查、加工功能，同时支持下载存在安全隐患数据。统计指标分析模块提供各类行为数据统计指标的计算与可视化的一键运行，图表具备查看数据和下载等交互式功能。实用小工具模块提供了便捷实用的数据合并与数据转换功能。

### 5.6.2 数据的质量审查与加工功能

用户主数据与各类行为数据的质量审查与加工的操作步骤类似，但是各侧重点不同，以下以用户主数据的审查与加工为例展开详细论述。

#### (1) 模块简介

用户基本属性数据模块包含简介、数据质量审查、加工、报告查看等功能。简介部分对原始数据表数据示例、重点治理内容、一卡通编码规范做了简要介绍。如图 5.11 所示。原始数据表格示例包含了用户上传的主数据文件的字段要求与规范，一卡通编码规范建立了用户一卡通号与身份类型的映射规则。

## 用户主数据治理（用户基本信息数据治理）

用户主数据原始格式如下表：重点关注的字段是读者单位和读者类型，根据一卡通编码规范，可以将读者身份类型与一卡通对应位置数字建立映射关系。

	证件号	性别	单位	年级组	读者类型
0	XXXXXXXXXX	男	图书馆	2001	研究生
1	XXXXXXXXXX	女	法学院	2002	本科生
2	...	...	...	...	...

参考一卡通编码规范，将本馆用户分为以下10类：

	用户一卡通号	身份类型	用户一卡通号.1	身份类型.1
0	1XXX-XXXX-XX	教职工	228X-XXXX-XX	硕士研究生
1	211X-XXXX-XX	博士研究生	3XXX-XXXX-XX	本科生
2	222X-XXXX-XX	硕士研究生	4XXX-XXXX-XX	京江学院
3	223X-XXXX-XX	硕士研究生	510X-XXXX-XX	国际交流学院
4	224X-XXXX-XX	硕士研究生	55XX-XXXX-XX	外教
5	225X-XXXX-XX	硕士研究生	598X-XXXX-XX	教职工-劳务租赁
6	226X-XXXX-XX	硕士研究生	599X-XXXX-XX	教职工其他类人员
7	227X-XXXX-XX	硕士研究生	9XXX-XXXX-XX	临时人员

图 5.11 用户主数据治理简介

### （2）审查原始主数据质量报告

主数据质量报告的查看需要遵照如图 5.12 所示的步骤，首先，选择待上传文件的类型，暂时支持“CSV”和“XLS”格式，建议选择“CSV”格式，数据文件格式之间的转换可以借助实用小工具模块实现。接着，需要勾选“查看原始数据质量报告”复选框，然后点击“Browse files”（浏览文件）以上传文件，最后点击“原始数据质量报告”选项卡。

主数据质量治理程序 原始数据质量报告 清洗后数据质量报告

1. 请根据您的需要做出选择

请您选择您上传的文件类型

csv  xls

请您根据需要选择以下功能

查看原始数据质量报告  是否清洗数据  查看清洗后的数据质量报告

2. 请上传您的文件

请选择一个文件！

Drag and drop file here  
Limit 1GB per file

Browse files

图 5.12 原始主数据质量报告查看

图 5.13 至图 5.16 展示了部分结果。图 5.13 简要概括了数据集的情况，包括字段的数量（Number of variables）、记录条数（Number of observations）、缺失率（Missing cells (%)）、重复率（Duplicate rows (%)）、分类变量（Categorical）与数值变量（Numeric）数量。

## Overview

Dataset statistics		Variable types	
Number of variables	4	Numeric	1
Number of observations	34304	Categorical	3
Missing cells	15		
Missing cells (%)	< 0.1%		
Duplicate rows	0		
Duplicate rows (%)	0.0%		
Total size in memory	1.0 MiB		
Average record size in memory	32.0 B		

图 5.13 原始主数据质量报告概览

图 5.14 绘制了各字段之间的关系图，颜色越深，表示两个字段之间的关系越强，图中表明“性别”和“单位”字段之间存在较强的相关性。

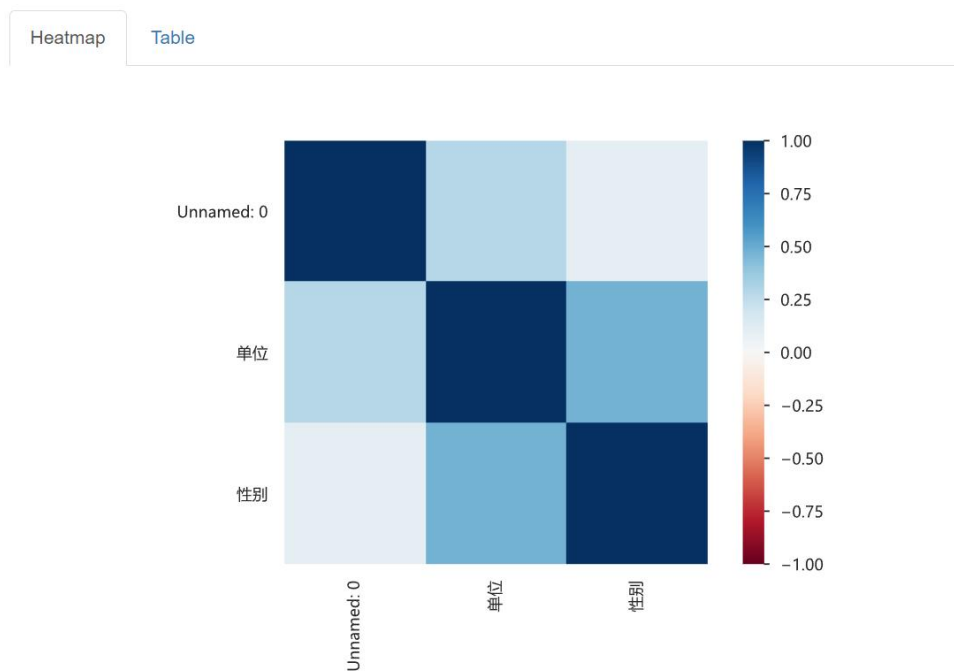


图 5.14 原始主数据各字段相关性检查

图 5.15 展示了各字段非空数值情况。图 5.16 详细描述了“性别”字段，它是一个分类变量，水平柱状图显示了其取值范围和各类别的频次分布，在下方有“Overview”，“Categories”，“Words”和“Characters”四个选项卡，“Overview”展示了该变量的值长度、字符数量和数据示例等；“Categories”与“Words”选项卡展现了各类别（短语）的频数与频率；“Characters”选项卡从单个字符与标点符号等角度展开分析，例如，“院”出现了 55227 词，频率为 11.3%，“( )”出现了 17205 次，当出现特殊字符时，应筛选包含特殊符号的数据记录，由人工确认数据取值是否符合规范。

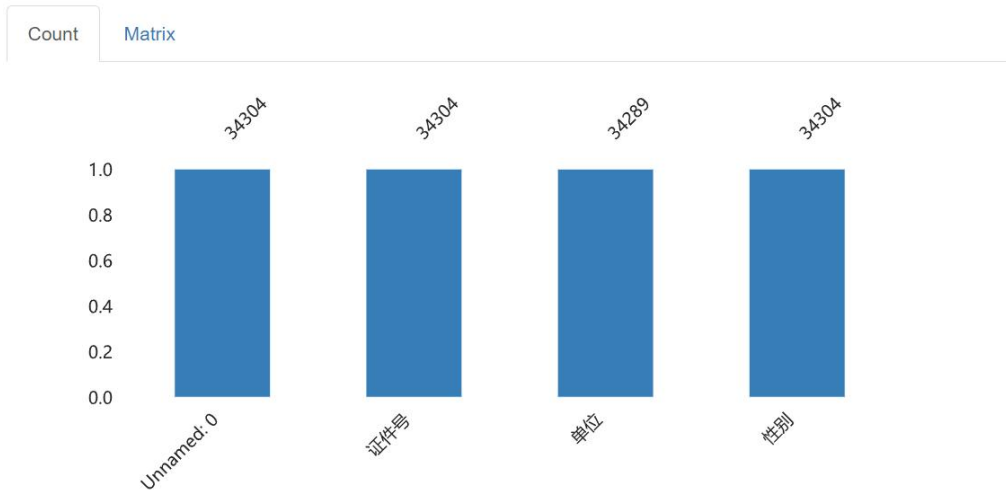


图 5.15 原始主数据各字段缺失值检查

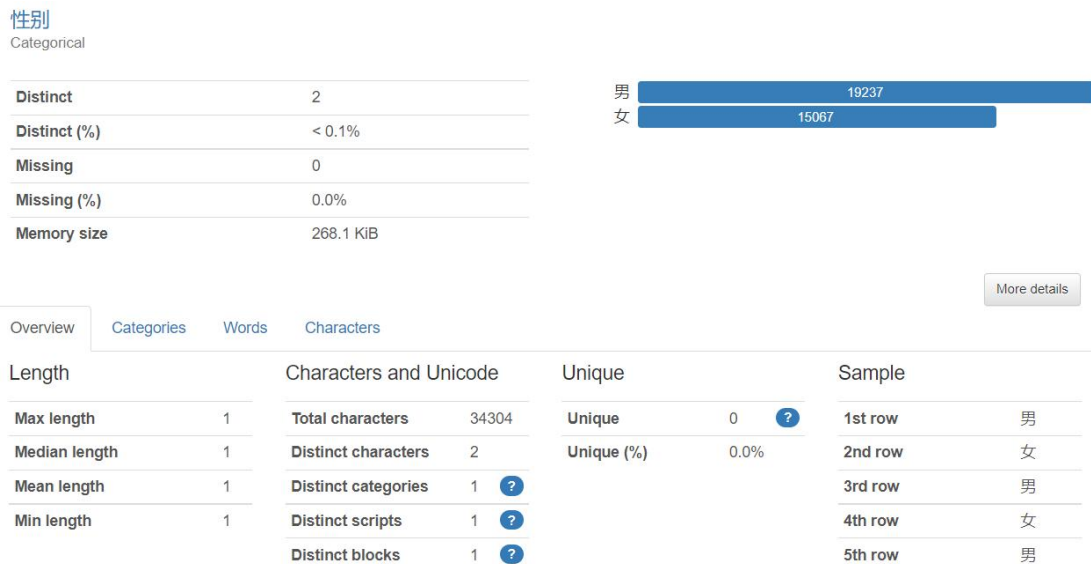


图 5.16 原始主数据性别字段详细信息

### (3) 用户主数据自动清洗

如果需要实现用户主数据的自动清洗，只需要勾选“是否清洗数据”复选框，然后上传数据文件即可，当数据清洗完成后，可以查看清洗后的数值分布。图 5.17 通过柱状图展示了清洗后的各用户身份类型频次分布，同时反映了该字段清洗后的取值范围，易于数据管理者发现遗留下的问题。图表下方的按钮提供了清洗后数据文件的下载功能。如果想审查清洗后数据的详细质量报告，可以勾选“查看清洗后的数据质量报告”复选框。

## 1.请根据您的需要做出选择

请选择您上传的文件类型

csv  xls

请您根据需要选择以下功能

查看原始数据质量报告

是否清洗数据

查看清洗后的数据质量报告

主数据开始清洗

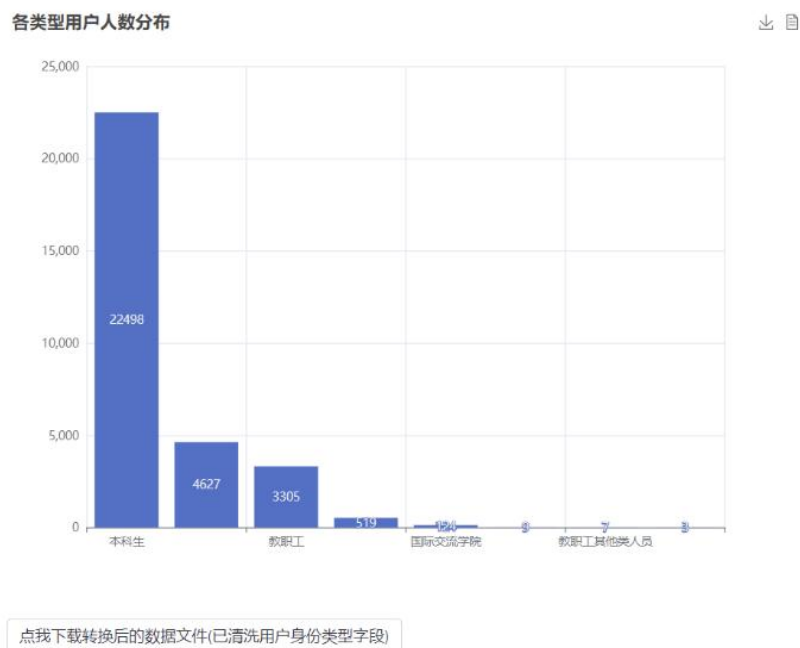


图 5.17 用户主数据质量清洗

### 5.6.3 统计指标分析功能

#### (1) 模块简介

统计指标分析功能支持借阅数据、座位预约数据和进出馆数据的统计分析，需要用户同时上传完成质量审查与加工后的用户主数据和用户行为数据文件，且需要按照约定好的名称命名文件，例如 Info.csv 对应于用户主数据文件，Access.csv、Seat.csv 和 Read.csv 分别对应于进户馆数据、座位预约数据和借还数据。接下来，以借阅数据分析为例详细讲解步骤。

#### (2) 借阅数据统计指标分析

在使用自动化指标可视化分析功能之前，数据管理者需要选择即将要分析的用户行为数据类型，在这里需要选择“用户借阅数据”；接下来需要上传规范命名后的数据文件，如图 5.18 所示，此时已上传“Info.csv”和“Read.csv”两个文件，完成这两步操作后，等待系统运行完毕，预先设定的各项指标与可视化图表便会自动显示在当前页面。



## 1.请您选择用户行为数据类型。

请您选择其中一个

- 用户进出馆数据  用户借阅数据  用户座位预约数据

## 2.请您上传Info.csv和Read.csv

请您上传相关文件!

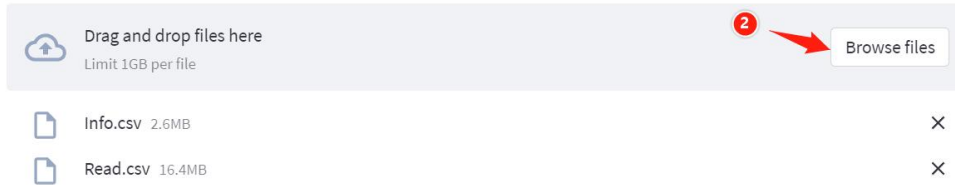


图 5.18 用户借阅数据统计分析操作步骤

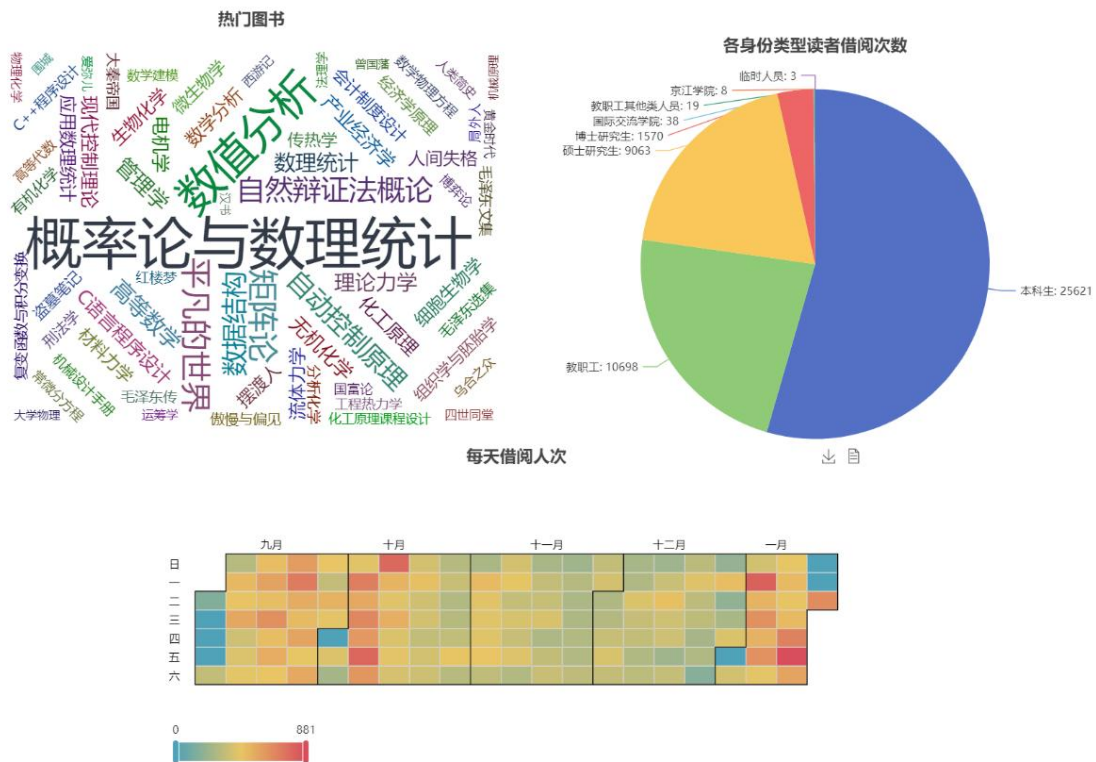


图 5.19 用户借阅数据可视化分析

图 5.19 展示了用户借阅数据部分可视化图表，包括热门图书、各身份类型用户借阅次数和每天用户图书的借阅次数，分别以词云图、饼状图和日历图展示，每个图表均具备用户查看绘图数据和下载功能的交互式按钮。

### 5.6.4 数据合并与格式转换功能

多个数据文件的合并与数据文件格式之间的转换是数据治理过程中常见的问题，“实用小工具”模块为数据管理者提供了上述功能，如图 5.20 所示，其支持用户将同一类型的多个“CSV”文件或“XLSX”文件进行合并；同时提供这两种数据文件格式的相互转换功能。

## 1.数据合并

请您选择相同类型的多的文件。

csv xlsx

请选择您需要合并的文件。

 Drag and drop files here  
Limit 1GB per file Browse files

## 2.数据转换

请选择上传的文件类型

转换为-->

请选择转换后的文件类型

xls ▼

CSV ▼

请选择一个文件!


 Drag and drop file here  
Limit 1GB per file Browse files

图 5.20 数据合并与格式转换功能

## 6.WSR 数据治理框架应用对策与建议

在 WSR 数据治理框架基础上,结合当前高校图书馆用户行为数据治理的重要要素和 A 大学图书馆实证分析的结果,从物理维度、事理维度和人力维度提出数据治理框架应用的对策与建议。

### 6.1 物理维度下明晰数据治理的多元目标

高校图书馆用户行为数据治理目标体现了数据治理项目的意图与导向,明晰多元治理目标有利于充分调动高校图书馆公共资源、汇集各方力量以推动目标实现。数据治理目标源于多方面的总结,包括政策与规章制度的强制性要求和现实问题的迫切需要,各馆需要结合实际情况确定不同发展阶段的多元化数据治理目标。

### 6.2 数据治理多目标的设定依据

#### (1) 政策与规章制度

国家、地方政府、学校的相关政策法规会对高校图书馆的用户行为数据治理提出强制性的要求,例如《个人信息保护法》<sup>[93]</sup>明确规定分析、评估境内自然人的行为等情形需要遵守此法;《网络安全法》<sup>[94]</sup>提出数据分类、重要数据备份和加密等措施以保障网络运行安全,要对收集的用户信息严格保密并建立健全用户信息保护制度;高校信息化中心等部门会颁布相关规章制度,如数据安全管理办法等。因此熟悉用户行为数据相关政策法规是确保数据治理项目内容科学合理性的重要前提。

#### (2) 现实需求

高校图书馆用户行为数据治理目标可能源于多方面的现实需求,包括利益相关方的要求与需要。主要考虑来自馆内中高层管理者、校级单位等的决策需求,也包括师生用户的信息需求。相关馆员在服务用户与处理用户行为数据时,也会面临数据质量问题,如数据缺失和不同系统数据不一致等问题。

### 6.3 数据治理多目标的视野拓展

目前图书馆用户行为数据治理还处于初步阶段,只有局部的用户行为数据治理活动与项目且缺乏体系化治理,再加上不同馆所处环境的差异导致了用户行为数据管理方法各不相同,所以高校图书馆应拓宽用户行为数据治理的目标视野,制定各阶段的数据治理目标。目前各馆重点的治理目标是数据质量、数据安全、数据洞察和数据流通,而数据标准和元数据主要依托服务商的系统平台来管理,数据架构、数据生命周期、数据服务由于人才与设备的缺失而关注较少,未来需要加强数据基础平台的规划与建设工作。

### 6.4 事理维度下建立健全数据治理机制

对于数据治理项目的每个目标都应该有相应的治理机制,治理机制为整个数据治理项目提供统一的规则与程序指导等。

#### 6.4.1 完善用户行为数据治理的岗位设置与组织建设

目前部分高校图书馆设置有数据管理员等相关岗位,但大多数馆都是以馆员多重身份兼职的方法完成行为数据的管理活动,如系统管理员负责多个系统技术支持,并兼管各系统数据的质量、安全和流通等,缺乏系统的数据治理组织体系和明确的分工协作机制,如数据治

理工作组,以及建立不同环节岗位数据质量与安全责任制等。在用户行为数据治理初始阶段,可以考虑建立敏捷组织,以具体的任务为中心,实现项目的高效运行,达到快速解决需求的目标。待有一定的数据技术与平台基础后,在校级或者图书馆整体数据治理组织架构下,可以设定相关角色与任务以突出对用户行为这一主题数据的重视与管理。

#### 6.4.2 建立用户行为数据治理的规范流程与程序

目前主要参照国家、地方和学校颁布的相关政策进行数据治理相关计划与活动的调整,为保障数据治理目标实现,高校图书馆和相关部门应制定正式化与标准化的战略、政策、流程、程序、合约协议并衡量绩效、监测合规性和管理问题。其中,数据战略代表数据治理的行动方向,如以数据洞察为主目标的用户行为数据治理的数据战略:以用户在馆行为数据为基础,实现馆内相关资源的配置优化并实现用户服务的精准化与个性化;政策是规范组织成员行为而规定的行动准则,如高校图书馆用户行为数据安全管理办法、高校图书馆用户行为数据质量管理办法等;清晰的数据流程包括记录用户行为数据生命周期的过程等,如数据质量控制流程、数据安全控制流程等;程序是处理问题的例行方法,如用户行为数据服务申请使用办事手续等;合约协议在高校图书馆和系统平台服务商之间签署,以明确双方的责任;绩效评估需要构建评价指标体系,以衡量数据治理工作带来的效果、效率和效益。另外,在项目实施阶段,需要监测其是否满足相关政策、标准和流程等,以及时发现问题并通过反馈、评估等方法调整计划与方案,并针对治理过程中常见的问题制定相应的解决方案。

#### 6.4.3 健全数据治理的关系机制

健全高校图书馆用户行为数据治理的关系机制有利于提升项目的运行与反馈效率,包含三个方面:培训、沟通和决策的协调。一是通过专业培训提升馆员数据素养。高校图书馆在接入新的信息系统平台与设备或新项目时,应该对相关部门的馆员进行培训以提升其操作过程的合规性,减少由人工操作失误引发的原始数据记录的错误。当有新的政策、程序、流程等发布时,也要进行培训以培育数据治理组织文化,将数据治理上升至一种组织内部共享的精神价值。二是充分沟通优化数据治理方案。在数据治理项目设计与运行过程中,应建立广泛的沟通机制,减少由不充分沟通造成的错误指示和重复性工作等。三是加强数据治理项目的决策协调工作,兼顾国家、学校、社会和用户的要求,建立决策协调机制,以兼顾多方利益并提升信息上报与反馈的效率。

### 6.5 人理维度下推进多元主体协同治理

在用户行为数据治理项目运行的不同生命周期,会有不同的利益相关方协同参与项目。对于作为数据供给者的师生用户,图书馆应该广泛收集其对各项用户服务的反馈建议,并结合数据探究异常数据和行为规律产生的深层次原因,以及时发现问题并改进服务,提升用户对服务的满意度;图书馆还应配合校级相关单位,参与校级数据治理项目,同时完善自身的用户行为数据治理建设。受限于馆系统与设备的信息技术水平和馆员数据素养的差异,各馆可以和校内的信息化中心、计算机学院和校外的专业IT团队等进行合作,推进数据生命周期内各项管理活动的顺利开展。数据消费者是数据治理的服务对象,包括,其要求与需求往往是数据治理项目开展的目标,高校馆应密切关注并监督项目运行状况,与数据使用者及时沟通交流,以发现潜在风险并做出适当的计划变更与调整。所以,高校馆在治理生命周期的不同阶段应积极推进与相关方合作交流,以推进协同治理,进而高效实现提升数据质量和数据增值等治理目标。

## 7.结论与展望

### 7.1 研究结论

数字中国背景下,大数据、人工智能等技术在各个领域迅速发展,加快了数字化转型进程。为了顺应时代发展潮流和促进图书馆的现代化与智慧化发展,高校图书馆积极接受和采纳新的设备与技术平台,这进一步保障了其用户服务水平和效率的提升,然而,新技术的使用使一把双刃剑,在图书馆服务期间,各业务系统留下了大量的用户行为数据,伴随而来的是新的质量、安全等管理问题,这不仅增加了馆员的工作负担,还阻碍了图书馆数字化进程。因此,高校图书馆有必要实施数据治理来解决相关问题,通过种种举措提升数据的可信度和价值,以提升服务质量和效率,并主动为读者提供个性化与智慧化的服务。本文首先对典型性高校图书馆用户行为数据治理的关键要素进行了分析与归纳,然后在国家数据治理规范中的数据治理框架和数据治理概念框架的基础上运用 WSR 系统分析方法从物理(W)、事理(S)和人理(R)的角度将相关要素进行有机融合,以构建新框架;接下来,运用 WSR 数据治理框架对 A 大学图书馆用户行为数据治理进行分析,提出数据治理的目标、机制和相关方,并基于 Streamlit 设计数据治理平台;最终提出 WSR 数据治理框架的应用对策与建议。具体研究成果如下:

(1) 总结归纳当前高校图书馆用户行为数据治理的重要要素。采用网络调研法和访谈法调研典型性高校图书馆的用户行为数据治理经验与做法,归纳相关要素,凸显当前数据治理工作的重点。

(2) 基于 WSR 系统分析方法构建新数据治理框架。通过 WSR 系统分析方法从物理、事理和人理的维度融合高校馆用户行为数据治理的要素,并进一步结合国家规范、数据治理概念框架和专家的意见对模型进行补充与完善,阐述各要素的内涵与关系。

(3) 以 A 大学图书馆用户行为数据治理展开实证分析。运用 WSR 系统分析方法,分析 A 大学图书馆用户行为数据治理的目标、机制和相关方,经过分析,数据治理目标是数据质量、数据安全、元数据和数据洞察,然后围绕数据治理目标明确了数据战略,设计了相关角色与任务和决策权的分配,并分析了现有的相关政策内容,还设计出用户主数据、进出馆数据、座位预约数据和借还数据的管理流程、程序和数据字典等,最终基于 Streamlit 设计出数据治理平台。

(4) 提出 WSR 数据治理框架的应用对策与建议。在调研与实证分析的基础上,结合 WSR 新框架提出物理维度下明晰数据治理的多元目标,事理维度下建立健全数据治理机制和人理维度下推进多元主体协同治理的应用建议。

### 7.2 研究不足与展望

本研究虽然对高校图书馆用户行为数据治理框架的构建与实证展开研究,但受限于时间和条件,本研究仍有一定的局限性。

(1) 本研究基于典型性高校图书馆用户行为数据治理的经验做法分析与归纳治理要素,由于在调研时符合条件的馆并不多,且很多馆尚未给与用户行为数据重视,所以相关的措施也未实施,因此访谈的专家数量有限,仅访谈了武汉大学、南方科技大学、复旦大学、厦门大学等高校馆。相信随着高校馆信息技术的应用与提升,用户行为数据会成为宝贵的财富,能为图书馆服务质量与效率提升和个性化与智慧化服务赋能。

(2) 本研究以 A 大学图书馆用户行为数据治理展开实证分析,对用户主数据、进出馆数据、座位预约数据和借还数据的质量、安全、元数据和洞察的流程与程序进行了设计,并

基于 Streamlit 设计出数据治理平台，但是受限于技术与资源等条件，电子资源使用数据等仍然未进行有效的治理，且如何使用这些数据为图书馆优化管理服务和用户提供智慧化服务的具体方案仍有待深入研究，数据治理仅是实现这些蓝图的前提。

(3) 本研究提出的 WSR 数据治理新框架是符合当前高校馆的实际情况的，且实证分析的思路对同行有一定借鉴作用，但是各馆情况不一，侧重有所不同，需要根据实际情况合理运用理论框架，且当今世界，技术发展日新月异，治理要素与思路需要及时做出调整，灵活转变。

## 参考文献

- [1] 朝乐门. 数据科学理论与实践[M]. 清华大学出版社, 2017.
- [2] 关志英, 王波. 从宏观、中观、微观三个层面深入研讨,科学制定高校图书馆“十四五”发展规划——“2018-2022 年教育部高等学校图书馆情报工作指导委员会第三次工作会议”综述[J]. 大学图书馆学报, 2021, 39(01): 8-14.
- [3] 新华社. 十三届全国人大四次会议表决通过关于“十四五”规划和 2035 年远景目标纲要的决议\_滚动新闻\_中国政府网[EB/OL]. [2021-07-17]. [http://www.gov.cn/xinwen/2021-03/11/content\\_5592248.htm](http://www.gov.cn/xinwen/2021-03/11/content_5592248.htm).
- [4] ARL. Future Themes and Forecasts for Research Libraries and Emerging Technologies[EB/OL]. [2021-07-17]. <https://www.arl.org/wp-content/uploads/2020/08/2020.08.21-future-themes-and-forecasts-for-research-libraries-and-emerging-technologies.pdf>.
- [5] ASIS&T. ASIS&T Strategic Plan 2020-2025[EB/OL]. [2021-07-17]. <https://growthzonesitesprod.azureedge.net/wp-content/uploads/sites/946/Final-2020-2025-Strategic-Plan.pdf>.
- [6] 图书馆杂志. 图情报告 | 国际图联发布 2021 年《趋势报告》[EB/OL]//微信公众平台. [2022-02-10]. [http://mp.weixin.qq.com/s?\\_\\_biz=MzAxMjExMzkwMQ==&mid=2653671686&idx=1&sn=6345477ec2a00d330a76f4dcb416927b&chksm=8069b384b71e3a924721332d43356ceefa6bc0cb3cf082c63dab9404cfe38c8301a2fd90c30#rd](http://mp.weixin.qq.com/s?__biz=MzAxMjExMzkwMQ==&mid=2653671686&idx=1&sn=6345477ec2a00d330a76f4dcb416927b&chksm=8069b384b71e3a924721332d43356ceefa6bc0cb3cf082c63dab9404cfe38c8301a2fd90c30#rd).
- [7] 吴丹, 郭清玥. “十四五”时期图情学科愿景展望——基于全球战略蓝图的分析[J]. 图书情报知识, 2021, 38(03): 14-26.
- [8] 顾立平. 数据治理——图书馆事业的发展机遇[J]. 中国图书馆学报, 2016, 42(05): 40-56.
- [9][10][11][12][13][14][15][16][17][18][19][20][21][22] [23][24][25][26][27][28][29][30][31][32][33][34]
- [35] DGI Data Governance Framework - The Data Governance Institute[EB/OL]. [2021-08-29]. <https://datagovernance.com/the-dgi-data-governance-framework/>.
- [36] 数据管理协会[国际]. DAMA 数据管理知识体系指南 (原书第 2 版) [M]. DAMA 中国分会翻译组, 译. 机械工业出版社, 2020.
- [37] Cohen R. What's in a name? Data governance roles, responsibilities and results factors[J]. DM Review, 2006, 8.
- [38] Thomas G. Alpha males and data disasters: the case for data governance[M]. Brass Cannon Press, 2006.
- [39] 包冬梅, 范颖捷, 李鸣. 高校图书馆数据治理及其框架[J]. 图书情报工作, 2015, 59(18): 134-141.
- [40] 张绍华, 潘蓉, 宗宇伟. 大数据治理与服务[M]. 上海科学技术出版社, 2016.
- [41] Marinos G. We're Not Doing What?: The Top 10 Corporate Oversights In Data Governance[J]. Information Management, 2004, 14(9): 62.
- [42] Six Critical Success Factors for Data Governance[EB/OL]//Kalypso. [2021-10-27]. <https://kalypso.com/viewpoints/entry/six-critical-success-factors-for-data-governance>.
- [43] Stockdale S. Deconstructing Data Governance[J]. Posters and Presentations, 2015.
- [44] Alhassan I, Sammon D, Daly M. Critical success factors for data governance: a telecommunications case study[J]. Journal of Decision Systems, 2019, 28(1): 41-61.
- [45] Wrobel A, Komnata K, Rudek K. IBM data governance solutions[C]//2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC). Krakow: IEEE, 2017: 1-3.
- [46] Earley S, Henderson D, Data Management Association, 等. DAMA 数据管理知识体系指南: 原书第 2 版 [M]. 2020.
- [47] 吴信东, 董丙冰, 堵新政, 等. 数据治理技术[J]. 软件学报, 2019, 30(09): 2830-2856.
- [48] 张宁, 袁勤俭. 数据治理研究述评[J]. 情报杂志, 2017, 36(05): 129-134+163.
- [49] Newman D, Logan D. Gartner introduces the EIM maturity model[J]. Gartner Research Publication, ID, 2008(G00160425).
- [50] 安小米, 白献阳, 洪学海. 政府大数据治理体系构成要素研究——基于贵州省的案例分析[J]. 电子政务, 2019(02): 2-16.
- [51] 傅昊阳, 徐飞龙, 范美玉. 论医院健康医疗大数据治理及体系构建[J]. 中国中医药图书馆情报杂志, 2019, 43(03): 1-5.
- [52] 刘刚. 俄罗斯国家数据治理体系建设及启示[J]. 情报杂志: 1-7.
- [53] 谭必勇, 刘芮. 英国政府数据治理体系及其对我国的启示:走向“善治”[J]. 信息资源管理学报, 2020, 10(05): 55-65.
- [54] 以娜. A 商业银行数据治理的思路与实施方式研究[D]. 广西师范大学, 2020.
- [55] 李福娟, 施国强. 航空公司数据治理体系及应用实践[J]. 信息与电脑(理论版), 2020, 32(06): 145-147.
- [56] 宋苏轩, 杨现民, 宋子强. 高校数据治理统筹管理体系的构成与实践路径[J]. 中国远程教育, 2021(11): 58-67.
- [57] 卢凤玲. 融合数据治理体系的智慧图书馆框架研究[J]. 图书馆, 2021(05): 74-78.
- [58] 许阳, 胡月. 政府数据治理的概念、应用场域及多重困境: 研究综述与展望[J]. 情报理论与实践: 1-13.
- [59] 张宇杰, 安小米, 张国庆. 政府大数据治理的成熟度评测指标体系构建[J]. 情报资料工作, 2018(01): 28-32.
- [60] 严昕. 公共图书馆数据治理框架构建研究[J]. 图书馆, 2020(05): 58-63.
- [61] 许晓东, 王锦华, 卞良, 等. 高等教育的数据治理研究[J]. 高等工程教育研究, 2015(05): 25-30.

- [62] 董晓辉, 郑小斌, 彭义平. 高校教育大数据治理的框架设计与实施[J]. 中国电化教育, 2019(08): 63-71.
- [63] 董晓辉. 活动理论视角下高校教育数据治理体系构成要素研究[J]. 中国电化教育, 2021(03): 79-87.
- [64] 谢娟. 教育数据治理的伦理框架: 价值、向度与路径[J]. 现代远程教育研究, 2020, 32(05): 15-24.
- [65] 曾凯, 高亮, 王新颖主编. 大数据治理及数据仓库模型设计[M]. 成都: 电子科技大学出版社, 2017: 202.
- [66] 嵇婷, 吴政. 公共文化服务大数据的来源、采集与分析研究[J]. 图书馆建设, 2015(11): 21-24.
- [67] 陈臣. 基于大数据的图书馆个性化服务用户行为分析研究[J]. 图书馆工作与研究, 2015(2): 28-31.
- [68] 阮光册, 任金玥, 涂世文. 数据驱动下的图书馆用户行为研究[J]. 图书馆杂志, 2019, 38(12): 28-37.
- [69] 寇晓东, 顾基发. 物理-事理-人理系统方法论 25 周年回顾——溯源、释义、比较与前瞻[J]. 管理评论, 2021, 33(5): 3-14.
- [70] 夏立新, 陈欢, 夏彦彦. 基于物理-事理-人理系统方法论的我国文献信息资源保障体系建设: 内容框架与实施路径[J]. 情报科学, 2022, 40(2): 4-10.
- [71] 杨峰, 姚乐野. WSR 描述下的快速响应情报体系: 一个综合集成的框架[J]. 情报资料工作, 2017(3): 11-17.
- [72] Wang Q, Li S. Shale gas industry sustainability assessment based on WSR methodology and fuzzy matter-element extension model: the case study of China[J]. Journal of Cleaner Production, 2019, 226: 336-348.
- [73] 杨洋洋, 谢雪梅. 基于 WSR 方法论的政府舆情治理评价研究[J]. 东北大学学报(社会科学版), 2021, 23(3): 62-70.
- [74] 李锋, 柳浩. WSR 视域下政府数据治理影响因素与路径研究[J]. 河海大学学报(哲学社会科学版), 2021, 23(6): 44-53+110.
- [75] 蒋甲丁, 肖潇, 张玲玲. 知识生态视角下基于 WSR 的大型工程项目知识共享影响因素及实证研究[J]. 管理评论, 2021, 33(10): 171-184.
- [76] Abraham R, Schneider J, vom Brocke J. Data governance: A conceptual framework, structured review, and research agenda[J]. International Journal of Information Management, 2019, 49: 424-438.
- [77] GB/T 34960.5-2018. 信息技术服务 治理 第 5 部分: 数据治理规范[Z]. 国家市场监督管理总局; 中国国家标准化管理委员会, 2018.
- [78] Kim H Y, Cho J S. Data governance framework for big data implementation with a case of Korea[C]//2017 IEEE International Congress on Big Data (BigData Congress). IEEE, 2017: 384-391.
- [79] Palczewska A, Fu X, Trundle P, et al. Towards model governance in predictive toxicology[J]. International Journal of Information Management, 2013, 33(3): 567-582.
- [80] de Abreu Faria F, Maçada A C G, Kumar K. Information governance in the banking industry[C]//2013 46th hawaii international conference on system sciences. IEEE, 2013: 4436-4445.
- [81] Carretero A G, Gualo F, Caballero I, et al. MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000[J]. Computer Standards & Interfaces, 2017, 54: 139-151.
- [82] Donaldson A, Walker P. Information governance—a view from the NHS[J]. International journal of Medical informatics, 2004, 73(3): 281-284.
- [83] 邵峻, 李博, 梁丽华, 等. 国内元数据研究共词图谱分析[J]. 情报杂志, 2013, 32(10): 106-110+105.
- [84] 李雨霏, 刘海燕, 闫树. 面向价值实现的数据资产管理体系构建[J]. 大数据, 2020, 6(3): 45-56.
- [85] Zhang Q, Sun X, Zhang M. Data Matters: A Strategic Action Framework for Data Governance[J]. Information & Management, 2022, 59(4): 103642.
- [86] Benfeldt O, Persson J S, Madsen S. Data governance as a collective action problem[J]. Information Systems Frontiers, 2020, 22(2): 299-313.
- [87] Borgman H, Heier H, Bahli B, et al. Dotting the i and crossing (out) the T in IT governance: New challenges for information governance[C]//2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 2016: 4901-4909.
- [88] Thompson N, Ravindran R, Nicosia S. Government data does not mean data governance: Lessons learned from a public sector application audit[J]. Government information quarterly, 2015, 32(3): 316-322.
- [89] Alhassan I, Sammon D, Daly M. Critical success factors for data governance: a theory building approach[J]. Information Systems Management, 2019, 36(2): 98-110.
- [90] Al-Ruithi M, Benkhelifa E, Hameed K. Data governance taxonomy: Cloud versus non-cloud[J]. Sustainability, 2018, 10(1): 95.
- [91] Abraham R, Schneider J, vom Brocke J. Data governance: A conceptual framework, structured review, and research agenda[J]. International Journal of Information Management, 2019, 49: 424-438.
- [92] Al-Ruithi M, Benkhelifa E, Hameed K. A systematic literature review of data governance and cloud data governance[J]. Personal and Ubiquitous Computing, 2019, 23(5): 839-859.
- [93] 国务院办公厅. 中华人民共和国个人信息保护法[EB/OL]. [2022-09-03]. [http://www.gov.cn/xinwen/2021-08/20/content\\_5632486.htm](http://www.gov.cn/xinwen/2021-08/20/content_5632486.htm).
- [94] 新华社. 中华人民共和国网络安全法[EB/OL]. [2022-09-03]. [http://www.gov.cn/xinwen/2016-11/07/content\\_5129723.htm](http://www.gov.cn/xinwen/2016-11/07/content_5129723.htm).



# 2020 学年江苏大学图书馆用户行为记录数据集\*

张晓阳<sup>1</sup> 刘洋<sup>2</sup> 苗松<sup>1</sup> 张朝霞<sup>1</sup>

(1 江苏大学图书馆 2 江苏大学科技信息研究所)

**摘要** 本数据集采集了江苏大学图书馆门禁管理平台、流通借还系统和座位预约系统中 2020 年 9 月 1 日至 2021 年 8 月 31 日的用户进出馆数据、用户借阅数据和用户座位预约使用数据, 将三个业务系统中的数据加工后通过时间和用户一卡通号进行关联, 共有 5 个数据文件。加工后的数据文件较规范完整, 可以通过数据挖掘来分析和预测用户在不同时空维度上的在馆行为特征和规律, 以优化图书馆信息服务和管理工作, 提升服务水平。

**关键词** 高校图书馆 门禁数据 借还数据 座位预约数据 数据集 用户信息行为

DOI: 10.13663/j.cnki.lj.2022.10.011

## Dataset of User Behavior Records of Jiangsu University Library in 2020 Academic Year

Zhang Xiaoyang<sup>1</sup>, Liu Yang<sup>2</sup>, Miao Song<sup>1</sup>, Zhang Zhaoxia<sup>1</sup>

(1 Jiangsu University Library; 2 Institute of Science and Technology Information of Jiangsu University)

**Abstract** This dataset collects the user entry and exit data, user borrow and return data and user seat reservation data from September 1, 2020 to August 31, 2021 of Jiangsu University Library access control management platform, circulation system and seat reservation system. The data in the three business systems are processed and associated with the users' ID through time, resulting in a total of five data files. The processed data files are more standardized and complete, and the characteristics and rules of users' behavior in the library in different time and space dimensions can be analyzed and predicted by data mining, so as to optimize the library information service and management, and improve the service level.

**Keywords** University library, Library access data, Borrow and return data, Seat reservation data, Dataset, User information behavior

### 数据集基本信息

数据集中文名称	2020 学年江苏大学图书馆用户行为记录数据集
数据集英文名称	Dataset of User Behavior Records of Jiangsu University Library in 2020 Academic Year
数据作者	张晓阳, 刘洋, 苗松, 张朝霞
通信作者	刘洋
作者单位	江苏大学图书馆, 江苏大学科技信息、研究所
版本号	V1.0

\* 本文系 CALIS 全国农学文献信息中心研究项目“数智时代高校图书馆用户数据治理体系研究”(项目编号: 2022031)的研究成果之一。

通信作者: 刘洋, lyyjll@163.com

(续表)

数据集中文名称	2020 学年江苏大学图书馆用户行为记录数据集
版本时间	20220704
基金项目类型	CALIS 全国农学文献信息中心研究项目
基金项目名称	数智时代高校图书馆用户数据治理体系研究
基金项目编号	2022031
国家	中国
语种	中文
数据覆盖时间范围	20200901-20210831
地理区域	江苏大学
经纬度	北纬 N32°12'33.90" 东经 E119°30'43.68"
数据格式	.csv
数据体量	用户主数据 34 304 条, 共 3.31MB; 借还记录 104 271 条, 共 19.3MB; 座位预约和使用记录 248 786 条, 共 52.5MB; 进出馆记录 1 894 109 条, 共 232MB; 用户行为记录 1 040 013 条, 共 117MB
主题分类	图书情报; 数据计量; 用户信息行为
全球唯一标识符	hdl:20.500.12304/10713
数据获取地址	http://hdl.handle.net/20.500.12304/10713
数据集组成	数据集共包括 5 个数据文件: 用户主数据、2020 学年用户借阅数据、2020 学年用户进出馆数据、2020 学年用户座位预约数据、2020 学年用户每日行为汇总数据
数据使用条款	本数据集可以通过在网站注册登录获取, 可以用于科学研究和教学目的, 需标注引用信息, 禁止二次分发和商业演绎

## 0 引言

近年来, 云计算、物联网、人工智能、5G 等新技术迅速发展, 人类进入数据富足供给的时代<sup>[1]</sup>, 这对各行业的变革产生了重要影响。2020 年 12 月, 教育部高等学校图书情报工作指导委员会召开会议, 从宏观、中观、微观 3 个层面, 深入探讨高校图书馆“十四五”规划的路径与方法, 即高校图书馆要迈向高质量发展规划阶段; 面向用户信息需求, 加大开放共享力度; 加快智慧化、个性化和特色化图书馆建设<sup>[2]</sup>。吴丹等对全球战略蓝图展开分析, 展望“十四五”时期图书情报学科可以在文化服务、技术创新、数据治理以及可持续发展等方面深化研究, 不断升级图书情报服务, 响应“新文科”建设部署, 加强数据开发与利用, 促进社会可持续发展<sup>[3]</sup>。在信息时代, “数据”“智慧”成为图书馆界的热点词, 如何借助先进的信息技术管理、挖掘和分析图书馆用户数据已成为一种趋势。本数据集采集了江苏大学图书馆门禁管理平台、流通借还系统和座位预约系统中 2020 年 9 月 1 日至 2021 年 8 月 31

日的用户进出馆数据、用户借阅数据和用户座位预约使用数据, 将 3 个业务系统中的数据加工后通过时间和用户 ID 进行关联, 共有 5 个数据文件。数据集字段丰富, 可用于探究和预测用户的在馆行为规律和特征, 为优化图书馆管理与服务提供参考依据。

## 1 数据采集和处理方法

数据文件分别从江苏大学图书馆门禁管理平台、流通借还系统和座位预约系统导出, 使用 Python 进行清洗、规整、脱敏, 形成较规范的数据文件, 再将 3 个业务系统中的用户每日行为数据进行汇总。

### 1.1 数据采集

数据采集主要步骤: 分别进入江苏大学图书馆门禁管理平台、流通借还系统、座位预约管理系统, 查询时间范围为 2020 年 9 月 1 日至 2021 年 8 月 31 日的记录, 将原始数据导出。

### 1.2 数据加工

#### 1.2.1 主数据加工

用户的一卡通号、身份类型、学院和性别

等字段为3个业务系统共用,所以将其集中处理,采用江苏大学流通借还系统中的规范数据,作为用户主数据,后期可以通过用户ID进行关联匹配。用户一卡通号关系到个人信息安全,所以采用非对称加密算法MD5对其进行脱敏处理,如表1所示。

表1 数据脱敏处理

字段名	脱敏后
USER_ID	2e83f71e9b27f0be26fef5b408eafc40

### 1.2.2 用户进出馆数据加工

(1) 原始数据文件中的“读者姓名”“身份证号”字段信息较为敏感,所以将这两个字段的数据删除,这有效保护了读者的隐私,也不会影响后续的数据清洗和挖掘工作。

(2) 将原始数据文件中的字段名称规范处理。

(3) 将一卡通号字段为“-”的无效记录删除。

(4) 原始数据文件中还存在较多的异常记录,需要处理以下2种情况:用户在一天内只存在进馆记录或者出馆记录;用户一天内存在多条进馆记录和出馆记录且进馆记录数量不等于出馆记录数量。解决措施:将所有记录按照用户ID和日期排序,将上下两条记录的一卡通号和时间日期(提取原始时间中的年月日)相同且上一条记录为用户进馆数据,下一条记录为用户出馆数据的所有匹配记录筛选出。

(5) 计算用户在馆的时长,单位为小时。

(6) 删除临时变量、冗余字段和重复记录,保留有价值的字段。

### 1.2.3 用户借还数据加工

(1) 将原始数据文件中的字段名称规范处理。

(2) 根据用户借阅和归还图书的时间计算用户的图书实际借阅时长,单位为天。

(3) 删除临时变量、冗余字段和重复记录,保留有价值的字段。

### 1.2.4 用户座位预约和使用数据加工

(1) 将原始数据文件中的字段名称规范处理。

(2) 计算用户预约使用座位时长和实际使用座位时长,单位为小时。

(3) 删除临时变量、冗余字段和重复记录,保留有价值的字段。

### 1.2.5 用户每日行为数据汇总

(1) 用户进出馆数据加工:将数据记录按照学号和日期聚合,统计用户一天的在馆总时长,单位为小时。

(2) 用户借阅数据加工:将数据记录按照学号和用户图书借阅日期聚合,将同一天的多条记录合并;使用字典存储一条记录中的书名、作者、出版社和借阅时长;使用列表存储用户一天内多条借阅记录。

(3) 用户座位预约使用数据加工:将数据记录按照学号和用户现场签到日期聚合,将同一天的多条记录合并,并统计用户一天内预约和实际使用图书馆座位的总时长,单位为小时;使用字典存储一条记录中的区域、空间名称、预约方式和违约方式;使用列表存储用户一天内多条座位使用记录。

(4) 通过时间和用户ID将各业务系统中的数据关联,汇总用户每日行为数据。

## 2 数据字典和数据样本

数据集共有44个字段,具体如表2所示。

数据字典符合元数据国际标准规范DDI。

## 3 数据质量控制

本数据集采集于江苏大学图书馆门禁管理系统、流通借还系统和座位预约系统,包含用户进出馆数据、图书借还数据和座位预约使用数据。原始数据文件中存在较多的质量问题,都能够使用Python逐一处理。针对用户主数据,包括一卡通号、身份类型、专业和性别等字段,存在各个业务系统中不一致的情况,所以统一采用流通借还系统中用户的基本数据,后续可以通过用户ID进行匹配关联。针对各个业务系统中的数据,统计各个字段空值,将缺失值较多的字段删除;检查每个字段的唯一值数量,删除只有一个唯一值的字段;通过直方图直观审查涉及用户时间以及时长的字段,将不符合时间范围要求的数据进行再度加工;删除冗余字段,例如用户进出馆记录中,入口通道名称、类型、闸机编号关联性较强,可以

表2 数据字典

数据	序号	字段	名称	样例值	备注
主数据	1	USER_ID	用户一卡通号	2e83f71e9b27f0be 26fef5b408cafc40	该字段已进行脱敏处理 (具备唯一性)
	2	USER_TYPE	身份类型	研究生	参照《江苏大学一卡通 基础信息统一编码规 范》,用户类型分为10 种,主要包括教职工、 研究生、本科生等,空 值表示一卡通未提供
	3	USER_DEPT	用户单位或 学院	图书馆	该字段表示读者所在单 位或者学院信息,空值 表示一卡通未提供
	4	USER_GENDER	性别	男	空值表示一卡通未提供
用户进出 馆记录	1	USER_ENTRY_TIME	用户进馆时间	2020-09-01 09:35:25	该字段表示用户进馆的 时间
	2	USER_ID	用户一卡通号	2e83f71e9b27f0be 26fef5b408cafc40	该字段已进行脱敏处理 (具备唯一性)
	3	USER_ENTRY_ VERIFICATION	用户入馆认证 方式	人脸	该字段表示用户进馆认 证方式,包含:人脸、二 维码和一卡通
	4	USER_ENTRY_GATE_ NAME	用户入馆通道	入口通道 1	该字段为用户入馆通 道,共有6个人馆通道
	5	USER_EXIT_TIME	用户出馆时间	2020-09-01 09:35:25	该字段表示用户出馆的 时间
	6	USER_EXIT_ VERIFICATION	用户出馆认证 方式	人脸	该字段表示用户出馆认 证方式,包含人脸、二 维码和一卡通
	7	USER_EXIT_GATE_ NAME	用户出馆通道	出口通道 1	该字段为用户出馆通 道,共有6个出馆通道
	8	USER_TIME_IN_ LIBRARY	用户在馆时长	2.22	该字段表示用户在馆时 长,单位为小时
用户座 位预 约和 使用 记录	1	USER_ID	用户一卡通号	2e83f71e9b27f0be 26fef5b408cafc40	该字段已进行脱敏处理 (具备唯一性)
	2	USER_TIME_OF_ APPOINTMENT	用户预约使用 座位时长	2020-09-01 13:06:10.423	该字段表示用户预约使 用图书馆座位的时长
	3	USER_START_TIME	用户预约开始 使用座位时间	2020-09-01 13:06:10	该字段表示用户预约开 始使用图书馆座位时间
	4	USER_END_TIME	用户预约结束 使用座位时间	2020-09-01 23:50:00	该字段表示用户预约结 束使用图书馆座位时间
	5	USER_SIGN_IN_TIME	用户现场签到 时间	2020-09-01 13:06:10	该字段表示用户使用图 书馆座位时间
	6	USER_SIGN_OUT_ TIME	用户现场签退 时间	2020-09-01 16:05:18.110	该字段表示用户现场签 退使用图书馆座位时间

(续表)

数据	序号	字段	名称	样例值	备注
用户座位预约和使用记录	7	USER_USE_TIME_OF_SEAT	用户使用座位的时长	3.21	该字段表示用户使用图书馆座位的时长, 单位为小时
	8	SPACE_NAME	空间名称	1	该字段表示用户使用的座位编号, 不同区域座位编号总数不同
	9	AREA	区域	自然科学图书借阅 A 区	该字段表示用户所在空间区域
	10	USER_WAY_OF_APPOINTMENT	用户预约方式	座位现场预约	该字段表示用户预约使用座位的方法, 包含座位现场预约和座位网上预约两种方法
	11	USER_BREAK_APPOINTMENT_TYPE	用户违约方式	未违约	该字段表示用户违约方式, 包含未违约、临时离开未刷卡、临时离开超时三种方式
12	USER_USE_TIME_OF_APPOINTMENT	用户预约使用时长	10.24	该字段表示用户预约使用图书馆座位的时长, 单位为小时	
用户借还数据	1	USER_ID	用户一卡通号	2e83f71e9b27f0be26fef5b408eafc40	该字段已进行脱敏处理(具备唯一性)
	2	BOOK_TITLE	书籍名称	腐蚀电化学	该字段表示书籍的名称, 空值表示系统未录入
	3	BOOK_AUTHOR	书籍作者	曹楚南编著	该字段表示书籍的作者, 空值表示系统未录入
	4	BOOK_PUBLISHER	出版商	化学工业出版社	该字段表示书籍的出版社, 空值表示系统未录入
	5	BOOK_CALL_NUMBER	索书号	TG174.41/C21	该字段表示书籍的索书号, 空值表示系统未录入
	6	BOOK_BAR_CODE	财产号	9706066	该字段表示书籍的财产号
	7	USER_BORROW_TIME	用户借书时间	2021-04-15 22:02:24	该字段表示用户借阅图书的时间
	8	USER_RETURN_TIME	用户还书时间	2021-05-09 18:00:12	该字段表示用户归还图书的时间
	9	DUE_TIME	应还日期	2021-05-09 18:00:12	该字段表示用户应还图书日期
	10	RENEWAL_TIME	续借日期	2021-05-09 18:00:12	该字段表示用户续借图书日期, 空值表示系统未录入
	11	USER_USE_TIME_OF_BOOK	用户使用图书时长	23.33	该字段表示用户借阅图书的时长, 单位为天

(续表)

数据	序号	字段	名称	样例值	备注
	1	USER_ID	用户一卡通号	2e83f71e9b27f0be 26fef5b408eafc40	该字段已进行脱敏处理 (具备唯一性)
	2	TIME	时间	2020-09-01	该字段表示用户进馆日期 (也可以表示用户图书 借阅日期、用户使用图书 馆座位现场签到日期)
	3	USER_TIME_IN_ LIBRARY	用户在馆时长	2.22	该字段表示用户在馆时 长,单位为小时
	4	USER_USE_TIME_OF_ SEAT	用户使用座位 的时长	3.21	该字段表示用户使用图 书馆座位的时长,单位 为小时
	5	USER_USE_TIME_OF_ APPOINTMENT	用户预约使用 时长	10.24	该字段表示用户预约使 用图书馆座位的时长, 单位为小时
	6	USER_USE_TIMES_OF_ SEAT	用户使用座位 的次数	2	表示用户当天使用座位 的次数
用户 每日 行为 数据 汇总	7	USER_USE_RECORDS_ OF_LIBRARY_SPACE	用户使用图书 馆空间的记录	[{'AREA': '社会科学图书 馆借阅A区', 'SPACE_ NAME': '121', 'USER_ WAY_OF_APPOINTMENT': '座位现场预约', 'USER_ BREAK_APPOINTMENT_ TYPE': '未违约'}, {'AREA': '社会科学图书馆借阅 A区', 'SPACE_NAME': '145', 'USER_WAY_OF_ APPOINTMENT': '座位现 场预约', 'USER_BREAK_ APPOINTMENT_TYPE': ' 未违约'}]	表示用户当天使用图书 馆座位的记录,使用列 表存储多条记录
	8	USER_BORROW_ RECORDS	用户借阅记录	[{'BOOK_TITLE': '数 学分析讲义', 'BOOK_ AUTHOR': '张福保,薛星 美,潮小李编', 'BOOK_ PUBLISHER': '科学出版 社', 'USER_USE_TIME_ OF_BOOK': '38.27days'}, {'BOOK_TITLE': '数学分 析', 'BOOK_AUTHOR': ' 华东师范大学数学系编', 'BOOK_PUBLISHER': '高 等教育出版社', 'USER_ USE_TIME_OF_BOOK': '38.27days'}]	表示用户当天借阅图书 的记录,使用列表存储 多条记录
	9	USER_BORROW_ COUNTS		2.0	该字段表示用户当天借 阅图书的册数

注: 用户行为数据文件包含大量空值, 由不同用户行为差异所导致。

相互推演,只保留入口通道名称即可,以节约存储空间。

#### 4 数据价值

本数据集包含江苏大学图书馆2020年9月份至2021年8月份的一学年用户进出馆数据、用户借还图书数据和用户预约使用座位数据,共包含44个字段,通过分析数据可以探究用户在图书馆的活动时间、空间和图书借还规律特征,以优化图书馆的管理与服务工作,提升图书馆用户数据治理的水平,促进用户服务的智慧化。此外,可以通过数据建立模型来评估用户对图书馆实施相关政策措施和开放服务的满意度。

#### 5 数据使用方法和建议

本数据集可以通过统计分析来探究高校图书馆用户的在馆时间、空间以及图书借还规律。可以针对某一业务系统展开深入分析,如用户进出馆记录可以用来分析用户在不同时间

阶段内进馆与出馆时间分布规律,用户在馆时长分布规律;可进一步通过时间序列分析、机器学习等模型方法预测读者的行为,如建立用户模型,衡量用户对图书馆的利用和依赖程度;预测未来一段时间内,读者的在馆时间规律特征等;也可以和用户借阅数据、座位预约使用数据进行关联分析和挖掘更多的用户行为规律,例如探究用户在馆时长、座位使用时长和图书借阅册数、图书主题之间的关系,为进一步测度图书馆利用率、开展文献阅读推广和学风评价、合理配置空间和文献资源提供研究和决策支持。

##### 5.1 数据引用格式

张晓阳,刘洋,苗松,等.2020学年江苏大学图书馆用户行为记录数据集[DB/OL].[日期].  
<http://hdl.handle.net/20.500.12304/10713>.

##### 5.2 论文引用格式

张晓阳,刘洋,苗松,等.2020学年江苏大学图书馆用户行为记录数据集[J].图书馆杂志,2022(10):97-103.

#### 参考文献

- [1] 朝乐门.数据科学理论与实践[M].北京:清华大学出版社,2017.
- [2] 关志英,王波.从宏观、中观、微观三个层面深入研讨,科学制定高校图书馆“十四五”发展规划——“2018-2022年教育部高等学校图书情报工作指导委员会第三次工作会议”综述[J].大学图书馆学报,2021,39(1):8-14.
- [3] 吴丹,郭清玥.“十四五”时期图情学科愿景展望——基于全球战略蓝图的分析[J].图书情报知识,2021,38(3):14-26.

张晓阳 江苏大学图书馆,副馆长,研究馆员。研究方向:科学计量与数据服务。作者贡献:文章选题与思路构建,论文修改。E-mail:zhxy@ujs.edu.cn 江苏镇江 212013

刘洋 江苏大学科技信息研究所,在读硕士研究生。研究方向:数据治理。作者贡献:撰写论文,加工清洗数据文件。江苏镇江 212013

苗松 江苏大学图书馆,副研究馆员。研究方向:图情技术。作者贡献:数据提取与规范。江苏镇江 212013

张朝霞 女,江苏大学图书馆,馆员。研究方向:系统管理。作者贡献:采集数据。江苏镇江 212013

(收稿日期:2022-01-11 修回日期:2022-07-14)

# 用户数据视角下高校图书馆 用户行为特征研究\*

张晓阳 许欣 刘洋

**【摘要】** 高校图书馆的门禁管理系统、流通借还系统和座位管理系统数据可以用来分析高校图书馆用户的行为习惯和阅读需求。采用 t 检验、斯皮尔曼相关性分析和访谈法研究某高校图书馆 2020—2021 学年不同学历和不同专业用户的进馆行为、借阅行为和座位使用行为的特点。结果发现：考试季各类用户的进馆次数和在馆时长皆为全年最高；文科生更倾向于留馆阅读而理科生更倾向于外借；研究生更倾向于借阅自身专业领域的图书而本科生借阅图书种类却不受限于专业。根据高校图书馆用户入馆、在馆和用馆数据，揭示用户行为特征和差异，可为高校图书馆优化管理和升级服务提供参考。

**【关键词】** 高校图书馆 用户行为 用户数据

**Abstract:** The data of entrance control system, borrowing and returning system and seat management system of university library can be used to analyze the behavior habits and reading needs of university library users. This paper used the T-test, Spearman correlation analysis and interviews to investigate the behavioral characteristics of users with different degrees and professional backgrounds in the academic year of 2020-2021. The results showed that during the examination season, the number of times users entered and the length of time they stayed in the library were both the highest in the year; in terms of borrowing, the students of liberal arts tended to stay in the library to read, and students of science and engineering tended to check out the books from the library; the types of books borrowed by graduate students were more specialized than those borrowed by undergraduates. According to the data of users entering, in and using the library, this paper reveals the characteristics and differences of users' behaviors, which can provide references for the optimal management and personalized service of university libraries.

**Key words:** university library users' behavior user data

DOI:10.15941/j.cnki.issn1001-0424.2023.02.009

## 0 引言

高校图书馆是服务人才培养和科学研究的学术型机构，对涵养校园优良学风和提升大学综合实力有一定的促进作用<sup>[1-2]</sup>。21 世纪以来，互联网和数字信息技术迅速发展，数据库管理系统在图书馆被广泛应用，图书馆的后台管理系统记录了大量数据，这些数据包括用户利用图书馆提供的业务和服务的各类行为信息，如进出馆信息、纸质图书借阅信息、座位使用信息等数据。分析用户的行为特征，并对用户使用图书馆各类行为之间的相关性展开研究，有利于探究不同类型的用户对图书馆提供的不同业务和服务使用习惯的异同，进一步帮助图书馆有针对性地调整管理方式和升级相关服务，充分发挥高校图书馆人才培养、科学研究和文化创新的作用<sup>[3]</sup>。

## 1 图书馆用户行为的相关研究和应用

国内外已有许多学者根据用户利用图书馆开展的业务和服务所产生的数据对用户行为进行研究。余以胜、张玉峰<sup>[4]</sup>通过数据挖掘技术分析图书馆用户信息。王伟等<sup>[5]</sup>认为，对图书馆用户借阅、检索和文献利用信息的挖掘，可以揭示数据之间的关联，分析用户行为。李爱国等<sup>[6]</sup>认为可以通过技术分析，将图书馆的用户信息资源应用于

\* 本文系江苏大学信息分析与文献计量研究生教学案例项目（编号：〔2022〕21-1-30）的研究成果之一。



图书馆决策管理、个性化服务等多个领域。Kanarkard W 等<sup>[7]</sup>通过机器学习算法进行图书馆大数据分析,对图书馆用户行为进行预测。李丹等<sup>[8]</sup>通过图书馆用户信息、图书信息、空间服务信息数据分析用户行为,构建用户画像,提供个性化服务。

高校图书馆用户的行为根据用户对图书馆所提供服务的利用方式有多种分类方法。吴英梅等<sup>[9]</sup>将用户的在馆行为分为到馆频次、在馆时长、借书数量、访问图书馆网站次数、电子资源利用 5 个方面。李荣等<sup>[10]</sup>将图书馆用户行为分为入馆行为、阅览区座位使用行为、借阅行为、纸质资源利用行为、电子资源利用行为、其他行为 6 个方面。李庆丰等<sup>[11]</sup>认为大学生对图书馆的利用主要集中在 5 个方面:(1) 自我学习的场所;(2) 获取图书馆管理员或职员帮助;(3) 检索和数据库的利用;(4) 与课程学习相关的文献利用;(5) 对取得的信息质量进行判断。

图书馆用户的在馆行为,在不同角度、群体中,表现出不同的特征。李庆丰等<sup>[11]</sup>认为不同学生群体利用图书馆业务的方式存在差异。谭云江等<sup>[12]</sup>基于数据挖掘技术对图书馆用户行为进行关联性分析,发现不同类型用户对图书的借阅存在着一定的模式。易明等<sup>[13]</sup>基于人类动力学分析出大学生的图书借阅行为在群体、学院、个体 3 个层面和本科、硕士、博士 3 个层次表现出不同的规律。邢荣华等<sup>[14]</sup>从时间因素和读者因素两方面对高校图书馆用户的借阅规律展开分析。

通过对图书馆用户数据的统计和分析,许多学者对图书馆的服务和运营提出了建议,以改进图书馆的服务质量和管理效率。施国良等<sup>[15]</sup>通过对不同图书馆用户群体的相似性和差异性分析提出个性化服务策略。张晓阳<sup>[16]</sup>通过分析高校图书馆用户的借还书的时间规律和出纳人员的工作效率,重新调整人力资源配置,提高了借还服务效率。尹相权等<sup>[17]</sup>分析了高校图书馆用户使用研究间的行为规律特征,提出了空间改造和推广服务的若干建议。Aguilar - Moreno E 等<sup>[18]</sup>通过地理信息技术对用户在图书馆内的行为轨迹数据进行了收集和分析,改善图书馆用户的在馆体验。

综上所述,国内外学者围绕用户在图书馆的行为特征展开了大量研究,为本研究提供了参考。近年来,随着数字时代的到来,手持终端和电子读物的发展给实体图书馆带来了巨大的冲击,实体图书馆作为长期以来高校学生首选课外学习场所的地位面临挑战。在这样的背景下,高校图书馆需要进一步掌握和分析用户的在馆行为及其特点,如:不同大学生群体花多少时间在实体图书馆的使用上?高校图书馆用户利用实体图书馆空间的特点是什么?在入馆用户中借阅纸本图书的所占比重有多大?针对以上问题,本研究将以某高校图书馆为例,通过统计分析的方法,探索用户在馆行为的特点及其相关性,并提出相应建议。

## 2 数据和方法

笔者获取了某高校图书馆 2020—2021 学年用户的进馆行为、借阅行为和座位使用数据并进行数据的集成和汇总,通过描述性分析和诊断性分析,对用户的在馆行为特点及其相关性进行探索。

### 2.1 数据来源

大学生对高校图书馆空间的使用,包括阅览室座位、公共空间、餐厅等,其中使用阅览室座位的行为可以被座位管理系统记录并产生数据,而其他空间使用行为难以被记录。如果用户产生进出馆数据却没有当日的座位预约使用数据或借阅数据,则可能认为其产生了公共空间使用、餐厅使用等其他图书馆空间使用行为。对于纸质资源的利用,可以体现在借阅数据上(但是仅在馆内阅读却未借出的行为无法被记录),而电子资源和服务的利用则因不受空间所限制,难以统计在馆使用行为。因此,本研究选择数据来源分别是:(1)从图书馆门禁管理平台提取的用户进出馆数据;(2)从流通借还系统提取的用户借阅数据;(3)从座位管理系统提取的座位预约和使用数据。

本研究数据来自于是一所工科特色的综合性大学,在校生 40 000 余人,大部分时间集中在校本部学习。其图书馆 2019 年底装备智慧门禁系统,详细记录用户身份及其进出馆时间;座位管理系统记录座位使用用户的个人身份信息和使用的座位位置、使用时长等;流通借还系统记录用户借还书类型和时间,符合本研究对高校图书馆用户行为研究的数据需求。本研究以该校图书馆 2020 年 9 月 1 日至 2021 年 8 月 31 日之间所有进馆的本科生和研究生为研究对象,数据采集于该高校图书馆门禁管理平台、流通借还系统和座位管理系统,通过 Python 加工缺失值、重复值和错误值、构建标准数据字典等保障措施提升数据质量<sup>[19]</sup>。在原有数据集的基础上,本研究汇总统计了用户月度进馆次数、在馆时长、借阅图书册次(分为自科类图书和社科类图书)、预约座位次数、使用座位时长等数据,具体字段见表 1。

表1 用户行为属性字段

行为属性	字段
入馆行为	入馆时间 (月份)
	在馆时长
借阅行为	借阅次数 (册次/月)
	借阅类型 (自科、社科)
座位使用行为	座位预约次数 (次/月)
	座位使用时长 (小时/月)

## 2.2 研究方法

根据汇总后的数据,本研究首先初步分析不同群体、不同时间段用户进入图书馆及使用图书馆资源的行为特点,如按月份统计用户进馆次数和在馆时长、按学院统计图书借阅和座位使用情况等。然后,将各项数据导入SPSS软件,采用K-S正态检验、相关分析等统计方法,分析不同时间、不同群体进馆行为和借阅行为、座位使用行为间的相关性。本研究侧重于描述性分析和诊断性分析,通过描述性分析可以挖掘2020—2021学年用户的在馆行为规律,通过诊断性分析探究相关关系的内在原因<sup>[20]</sup>。

## 3 基于用户入馆时间的数据分析

本研究把该校图书馆用户按学历分成本科生用户和研究生用户,分别探索不同时间段两类用户的在馆行为,并分析其特点和差异。

### 3.1 2020—2021 学年用户在馆行为总体情况

2020—2021 学年,该校图书馆进馆人次 189 万,进馆人数 34 303 人。其中本科生进馆 22 507 人,约占在校本科生总人数的 56%;研究生进馆 8 396 人,约占在校研究生总人数的 65%。在所有进馆用户中,11 746 位用户借阅过纸质书刊,占总人数的 34%;13 559 位用户使用座位预约系统预约过阅览室座位,占总人数的 39%。

6 月和 12 月是在校生进馆人次最多和在馆时间最长的月份。6 月进馆人次为 22 万次,在馆时长 63 万小时,平均在馆时长 2.85 小时/次;12 月进馆人次为 24 万次,在馆时长 80 万小时,平均在馆时长 3.26 小时/次。2 月是进馆人次最少和在馆时间最短的月份,其次是 3 月、7 月和 8 月。纸质书刊借出册次最多的月份是 10 月,一个月内共借出 1.1 万次;借出册次最少的月份是 2 月和 8 月,各月借出册次未超过 1 000 次。具体数据见表 2。

表2 2020—2021 学年各月份用户的进馆行为和借阅行为数据

月份	进馆人次 (次)	在馆时长 (小时)	平均在馆时长 (小时/次)	借阅册次 (次)
1	75 634	246 319.47	3.26	6 574
2	130	49.87	0.38	134
3	74 165	209 921.84	2.83	5 643
4	142 314	385 114.97	2.71	7 753
5	160 461	440 611.95	2.75	6 096
6	222 702	635 740.87	2.85	4 646
7	68 879	212 749.47	3.09	1 948
8	68 538	248 448.45	3.62	668
9	131 773	365 675	2.78	9 847
10	168 861	499 904.82	2.96	11 023
11	173 044	535 183.88	3.09	7 822
12	246 609	803 449.98	3.26	7 184

### 3.2 基于用户的入馆月份分析用户行为

6月和12月是本科生进馆高峰期,2月、7月、8月入馆人次最低。和本科生相似,研究生的入馆高峰是12月,2月、7月、8月入馆人次最少。而其他时间,研究生与本科生呈现出较不同的趋势,具体情况见图1和图2。(图1的横坐标表示月份,纵坐标表示当月入馆人次占学年总入馆人次的百分比;图2的横坐标表示月份,纵坐标表示当月纸质资源借阅册次占学年总入馆人次的百分比。)使用SPSS进行独立样本t检验,按照本科生和研究生分组,检验结果p值小于0.001,具体检验结果见表3。可以认为,本科生和研究生的入馆时间存在统计学意义上的显著差异。

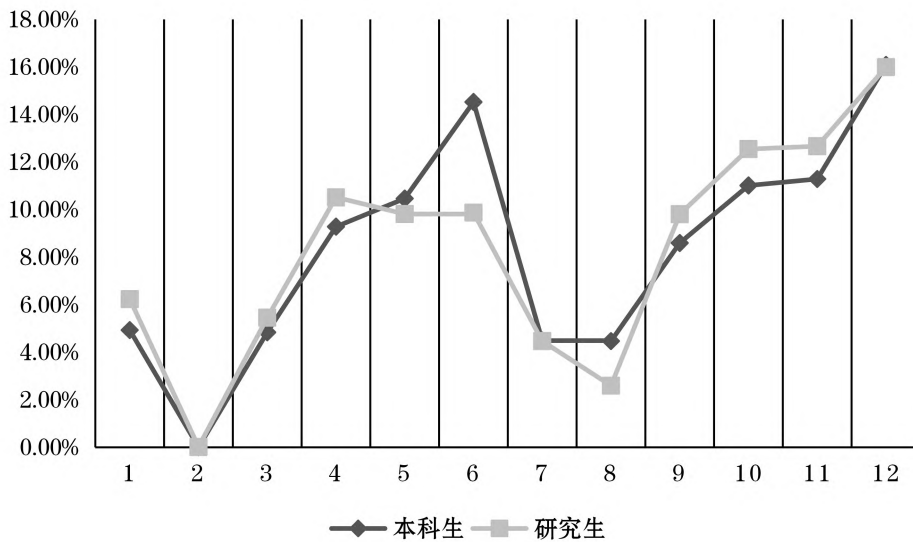


图1 本科生和研究生用户入馆人次随月份变化

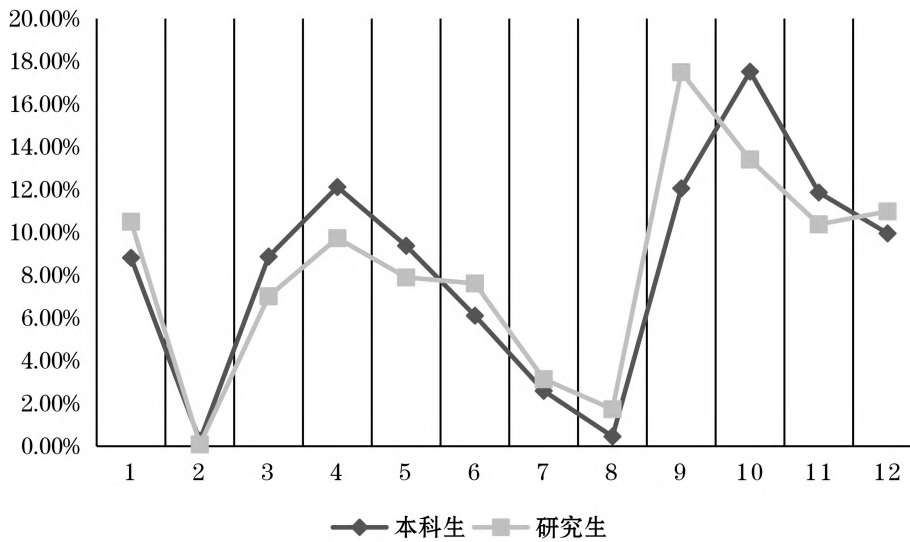


图2 本科生和研究生用户借阅册次随月份变化

表3 本科生和研究生用户入馆时间检验

平均值等同性 <i>t</i> 检验					
	<i>t</i>	自由度	<i>p</i>	平均值差值	标准误差差值
入馆月份	20.447	909 851	0	0.246	0.012
借阅月份	-15.136	905 853	0	-0.147	0.01

根据每月进馆人次和借阅册次的不同特点,将本科生进馆时间分为三个时期。A时期是进馆人次和平均在馆

时长都高、但借阅量中等的月份,如6月、12月;B时期进馆人次中等,但是借阅量达到最高峰,如4月、10月;C时期进馆人次低但平均在馆时间较长、借阅量较低,如2月、7月、8月。研究生则根据进馆人次和借阅册次的不同特点分成4个时期:D时期(11月,12月),进馆人次和平均在馆时长最高,借阅量中等;E时期(4月,5月,6月),进馆人次和借阅量都中等;C时期(2月,7月,8月),进馆人次低,平均在馆时间较长,借阅量低;F时期(9月),进馆人次中等,借阅册次最高峰。具体时期分类及代表月份、特征见表4。

表4 本科生和研究生用户入馆时间分类

类型	月份	特征	进馆人次排行	在馆时长排行	借阅册次排行
本科生 A 时期	6, 12	进馆人次和在馆时间都高,但借阅量中等	前 20%	前 20%	前 20% ~ 60%
本科生 B 时期	4, 10	进馆人次和在馆时长中等,借阅量最高峰	前 20% ~ 60%	前 20% ~ 60%	前 20%
本科生 C 时期	2, 7, 8	进馆人次低,借阅量低	后 30%	后 30%	后 30%
研究生 D 时期	11, 12	进馆人次和平均在馆时长最高,借阅量中等	前 20%	前 20%	前 20% ~ 60%
研究生 E 时期	4, 5, 6	进馆人次和借阅量都中等	前 20% ~ 60%	前 20% ~ 60%	前 20% ~ 60%
研究生 C 时期	2, 7, 8	进馆人次低,借阅量低	后 30%	后 30%	后 30%
研究生 F 时期	9	进馆人次中等,借阅册次最高峰	前 20% ~ 60%	前 20% ~ 60%	前 20%

### 3.2.1 各时期本科生用户的在馆行为特点

A时期的本科生进馆次数多,在馆时间长,是利用图书馆提供的资源和服务最积极的时间段。B时期本科生的月借阅册次(自科类和社科类的总和)全年最高,但这一时期用户的进馆次数和在馆时长只处于中等水平。根据斯皮尔曼相关分析结果,B时期用户的进馆行为和借阅行为间相关性系数与A时期较相近,但是借阅行为和座位使用行为之间的相关性系数远低于A时期,说明B时期用户借阅图书后更倾向于带出馆阅读。C时期用户两种类型(自科和社科)的借阅行为和进馆行为之间皆存在显著负相关,说明该时期的本科生进馆次数越多、在馆时间越长,借阅纸质图书的可能性就越低,反映出这一时期用户进入图书馆主要是为了利用图书馆舒适的空间环境。

具体相关性分析结果见表5。以本科生A时期为例,进馆次数和在馆时长之间的相关性系数为0.959,说明在馆时长的有序值变化值有91%(即0.959的平方)能用进馆次数的有序值来进行预测。

表5 本科生各时期在馆行为间相关性

		进馆次数	在馆时长	自科借阅	社科借阅	座位预约次数	座位使用时长
本科生 A 时期	进馆次数	1	0.959	0.075	0.077	0.392	0.394
	在馆时长	0.959	1	0.05	0.055	0.422	0.428
	自科借阅	0.075	0.05	1	0.154	0.059	0.058
	社科借阅	0.077	0.055	0.154	1	0.052	0.051
本科生 B 时期	进馆次数	1	0.934	0.075	0.075	0.316	0.318
	在馆时长	0.934	1	0.027	0.026	0.354	0.36
	自科借阅	0.075	0.027	1	0.158	0.041	0.039
	社科借阅	0.075	0.026	0.158	1	0.018	0.016
本科生 C 时期	进馆次数	1	0.952	-0.03	-0.026	0.093	0.093
	在馆时长	0.952	1	-0.055	-0.046	0.11	0.11
	自科借阅	-0.03	-0.055	1	0.201		
	社科借阅	-0.026	-0.046	0.201	1	0.026	0.026

注:表中的数值为斯皮尔曼相关性分析的相关性系数,正数表示两种行为间存在正相关关系,负数表示两种行为间存在负相关关系,未填入数字则代表两种行为间的相关性不显著。

### 3.2.2 各时期研究生用户的在馆行为特点

除了 E、F 时期进馆次数和自科借阅间相关性不显著外, 研究生各时期的进馆次数、在馆时长和自科借阅之间都是显著负相关。而社科借阅方面, 在借阅量较高的 E、F 时期与进馆行为的相关性系数皆大于 0.1, 虽然数字本身不高, 但是在所有用户全时期当中都是最突出的。

研究生下半年座位使用情况较为突出。D 时期和 F 时期涵盖了下半年的 9—12 月份, 这段时间研究生的进馆行为和座位使用行为间的相关性系数接近或超过 0.5, 远超过 E、C 时期 (最高 0.251)。这两段时期也分别是到馆人次和借阅人次最高的时期。具体相关性分析结果见表 6。

表 6 研究生各时期在馆行为间相关性

		进馆次数	在馆时长	自科借阅	社科借阅	座位预约次数	座位使用时长
研究生 D 时期	进馆次数	1	0.893	-0.084	0.072	0.419	0.419
	在馆时长	0.893	1	-0.127	0.049	0.437	0.439
	自科借阅	-0.084	-0.127	1	0.064	-0.068	-0.068
	社科借阅	0.072	0.049	0.064	1	0.047	0.048
研究生 E 时期	进馆次数	1	0.827		0.133	0.234	0.234
	在馆时长	0.827	1	-0.019	0.125	0.251	0.251
	自科借阅		-0.019	1	0.093		
	社科借阅	0.133	0.125	0.093	1	0.062	0.062
研究生 C 时期	进馆次数	1	0.826	-0.053		0.107	0.108
	在馆时长	0.826	1	-0.101		0.109	0.11
	自科借阅	-0.053	-0.101	1	0.129		
	社科借阅			0.129			
研究生 F 时期	进馆次数	1	0.842		0.165	0.469	0.469
	在馆时长	0.842	1	-0.052	0.15	0.509	0.511
	自科借阅		-0.052	1	0.063	-0.064	-0.066
	社科借阅	0.165	0.15	0.063	1	0.049	0.049

注: 表中的数值为斯皮尔曼相关性分析的相关性系数, 正数表示两种行为间存在正相关关系, 负数表示两种行为间存在负相关关系, 未填入数字则代表两种行为间的相关性不显著。

## 4 基于用户所属学科的数据分析

### 4.1 各类学科本科生用户的在馆行为特点

2020—2021 学年, 本科生用户分布于 24 个学院。因部分学院学生进馆次数较少, 故首先选取进馆次数前 80% 的学生所在的 11 所学院, 分别是财经学院、医学院、计算机科学与通信工程学院、电气信息工程学院、汽车与交通工程学院、管理学院、机械工程学院、能源与动力工程学院、材料科学与工程学院、数学科学学院和外国语学院。此外, 文学院和艺术学院虽然入馆人次不在前 80%, 但是借阅册次较高, 因此共分析 13 所学院。将这 13 所学院按学科划分为三类: 理工类、人文社科类和医药类。其中属于医药类的有医学院, 属于人文社科类的有外国语学院、文学院、艺术学院, 属于理工类的有管理学院、财经学院、计算机科学与通信工程学院、电气信息工程学院、汽车与交通工程学院、机械工程学院、能源与动力工程学院、材料科学与工程学院、数学科学学院。

#### 4.1.1 理工类本科生用户的进馆人次和借阅册次位居首位

在 3 类学科的学院中, 理工类学院本科生用户进馆人次、借阅册次最多; 人文社科类学院次之, 最少的是医药类学院 (因为只有一所学院, 人数最少)。从各学院用户平均进馆人次来看, 医药类本科生用户进馆人次最多, 理工类其次, 人文社科类最少; 从平均各学院图书借阅册次来看, 理工类本科生用户借阅册次最多, 医药类其次, 人文社科类最少。统计结果见图 3。

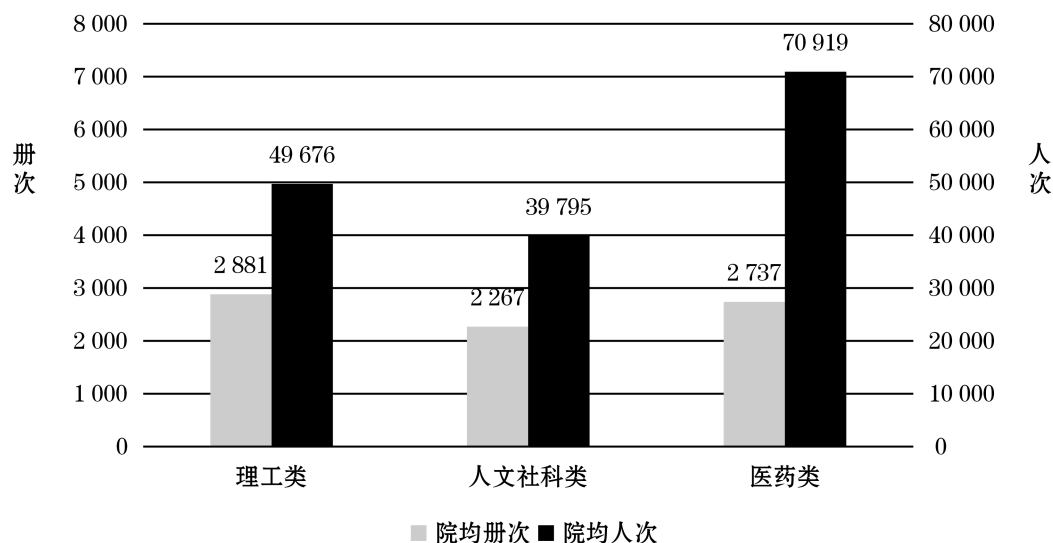


图3 不同类学院本科生平均借阅册次和进馆人次

图4是本科生各学院用户进馆人次(x轴)、借阅册次(y轴)和座位预约次数(气泡大小)的气泡图。可以看出,理工类绝大多数学院的气泡比人文社科类学院大,说明理工类学院用户比人文社科类学院用户预约座位次数更多;医药类学院用户3项数据都处于中等偏上;人文社科类学院都处于图的左侧,说明人文社科类学院用户进馆次数最少。从具体的各个学院来看,计算机学院、电气学院、汽车学院、医学院用户进馆次数和图书借阅次数都较多;财经学院用户虽然进馆人次最多,但是借阅图书数量偏少;数学学院、文学院、管理学院用户进馆次数偏少,借阅图书册次较多。

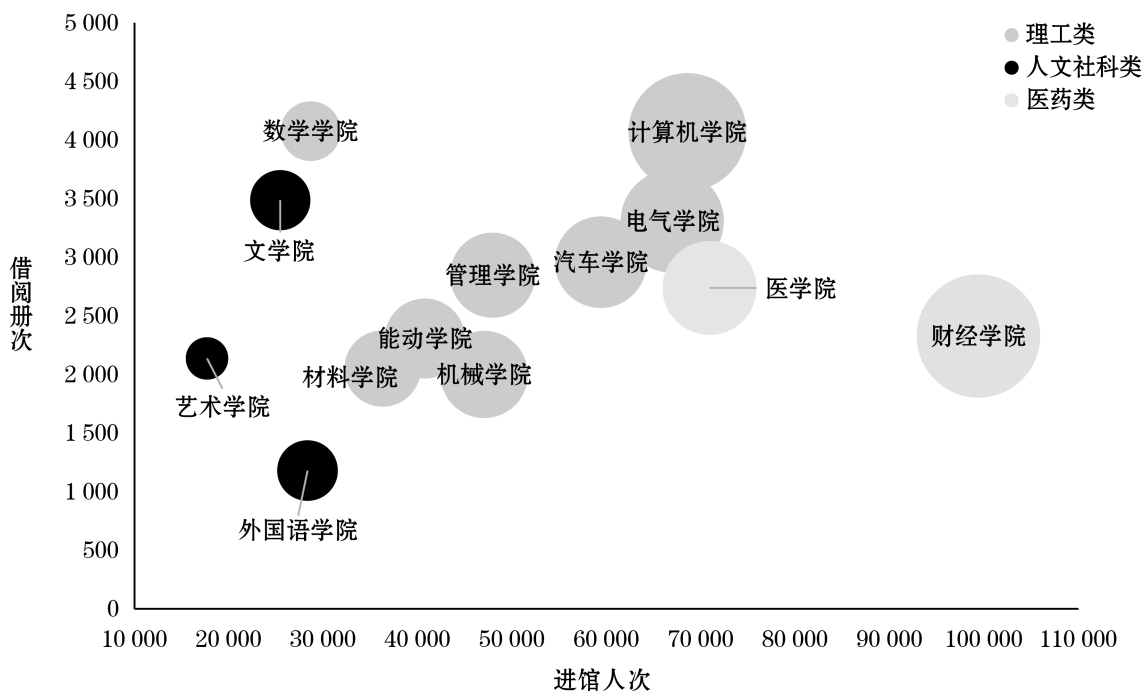


图4 各学院本科生进馆人次、借阅册次、座位使用时长气泡图

#### 4.1.2 3类学科本科生用户的借阅偏好呈现不同特点

3种学科属性学院的借阅呈现出不同的特点。人文社科类学院本科生用户借阅社科类图书的几率远高于自科类图书;医药类学院恰好相反,借阅自科类图书的几率高于社科类图书,且在3类学院所有进馆行为和借阅行为的相关性系数中数值最大;理工类学院的用户则对两种类型的图书有同样的借阅率。相关性分析结果见表7。

表7 不同类学院本科生进馆和借阅行为相关性

学科类型	入馆行为	自科借阅	社科借阅
理工类	进馆次数	0.057	0.056
	在馆时长	0.011	0.021
人文社科类	进馆次数	0.028	0.083
	在馆时长	不显著	0.031
医药类	进馆次数	0.099	0.057
	在馆时长	0.06	0.021

而在众多理工类学院当中,不同学院间两种图书的借阅比例各不相同,最终在总数上形成近似1:1的比例。管理学院和财经学院本科生用户的社科类图书借阅数量大于自科类,而其他学院则相反,自科类图书大于(或近似等于)社科类图书借阅册次。具体情况见表8。

表8 理工类各学院的本科生借阅图书册次及比例

学院	自然科学		社会科学		借阅总数
	借阅册次	占比	借阅册次	占比	
数学科学学院	2 713	66.58%	1 362	33.42%	4 075
计算机科学与通信工程学院	2 792	68.65%	1 275	31.35%	4 067
电气信息工程学院	2 283	69.04%	1 024	30.96%	3 307
汽车与交通工程学院	1 800	60.87%	1 157	39.13%	2 957
管理学院	442	15.53%	2 404	84.47%	2 846
财经学院	306	13.15%	2 021	86.85%	2 327
能源与动力工程学院	1 295	56.16%	1 011	43.84%	2 306
材料科学与工程学院	1 398	68.23%	651	31.77%	2 049
机械工程学院	1 299	64.98%	700	35.02%	1 999

#### 4.1.3 本科生用户借阅和座位使用行为间相关性偏低

在座位使用方面,3类学院本科生用户的进馆行为和座位预约行为间相关性系数都在0.3左右,非常接近,但是在借阅行为和座位使用行为之间表现出了差异性。理工类学院用户的两类图书借阅行为和座位使用行为皆显著相关。其中,自科类图书借阅和座位使用之间的相关性系数更高一些。人文社科类学院用户只有社科类图书借阅和座位使用行为之间显著相关,而医药类学院只有自科类图书借阅和座位使用行为之间显著相关。总的来说,虽然数值具有差异,但是皆小于0.1,实际上体现出的差异可以忽略不计,这也是由于借阅量总体偏低导致的。相关性分析结果见表9。

表9 本科学院用户借阅和座位使用行为相关性

学科类型	借阅行为	座位预约次数	座位使用时长
理工类	自科借阅	0.041	0.04
	社科借阅	0.023	0.022
人文社科类	社科借阅	0.051	0.05
医药类	自科借阅	0.078	0.076

#### 4.2 各类学科研究生用户的在馆行为特点

2020—2021学年,研究生用户分布于26个学院。因部分学院学生进馆次数较少,故选取进馆人次前80%的学生所在的15所学院。将这15所学院按学科划分为3类:理工类、人文社科类和医药类。其中医药类的有医学

院,人文社科类有教师教育学院、图书馆、艺术学院、法学院,理工类有管理学院、财经学院、机械工程学院、电气信息工程学院、材料科学与工程学院、汽车与交通工程学院、食品与生物工程学院、流体机械工程技术研究中心、计算机科学与通信工程学院、农业工程学院。

#### 4.2.1 人文社科类各学院研究生用户平均进馆次数最多

和本科生一样,研究生理工类学院数量最多,同时其用户的进馆人次和借阅册次也是3类学院中最高的。但是从各学院平均数据来看,人文社科类学院研究生用户的进馆人次略高于理工类学院,而院均借阅册次依然是理工类学院最多。具体情况见图5。

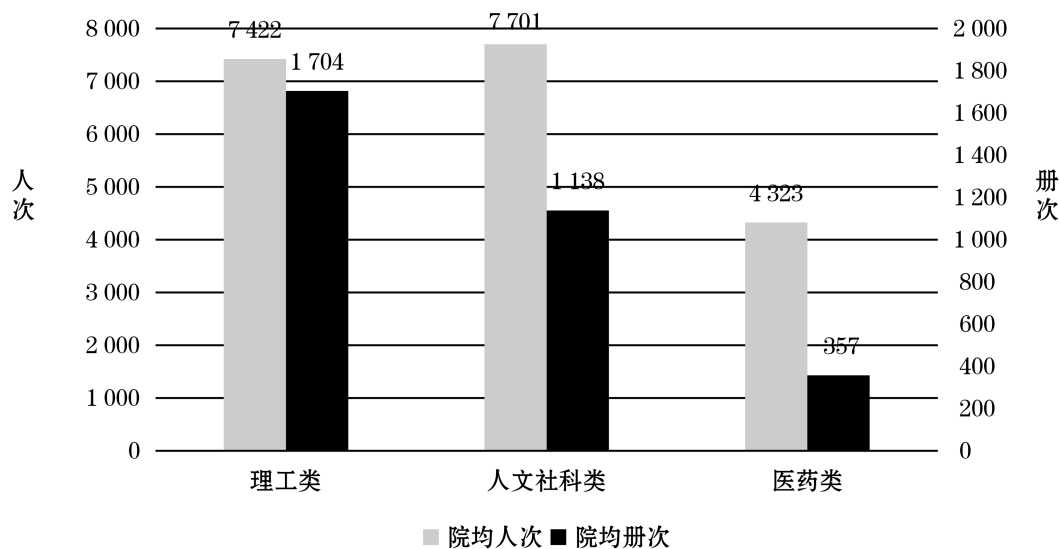


图5 不同类学院研究生平均进馆人次和借阅册次

图6是研究生各学院进馆人次(x轴)、借阅册次(y轴)和座位预约次数(气泡大小)的气泡图。除了人数最少的医药类学院(仅有医学院一所学院),人文社科和大多数理工类学院的气泡大小相仿;理工类学院的气泡分布在图的左右两侧,而人文社科类学院的气泡主要集中在靠中部位置,x轴的平均线附近。从具体的各个学院来看,机械学院、汽车学院、电气学院、材料学院、管理学院用户的进馆频次和图书借阅数量都是比较高的;艺术学院、农业工程学院的图书借阅数量高于所有学院的平均值,但是进馆频次较低;教师教育学院、图书馆、财经学院用户的入馆频次较高,借阅量较低。

#### 4.2.2 人文社科类学院研究生用户在馆时长最为突出

人文社科类学院研究生用户的进馆次数和在馆时长间相关性系数最高,为0.926,说明人文社科类学院研究生用户在馆时长有序值变化的86%能用他们进馆次数的有序值来进行预测。与之相对应的,人文社科类学院研究生用户进馆次数和座位预约行为间的相关性系数也最高。分析结果数据见表10。

表10 不同类学院研究生进馆次数和座位使用行为相关性

	在馆时长	座位预约次数	座位使用时长
人文社科类学院用户进馆次数	0.926	0.335	0.335
理工类学院用户进馆次数	0.83	0.343	0.343
医药类学院用户进馆次数	0.754	0.297	0.297

#### 4.2.3 3类学院自科借阅与进馆行为间相关性不显著或呈负相关

理工类学院用户的自科借阅与进馆次数之间相关性系数皆不显著,但这并不能说明理工类学院的用户不借阅自科类图书,因为从表11可以看出这些学院借阅自科类图书的占比远高于社科类图书(财经学院和管理学院除外)。通过对原始数据进行观察和分析,发现理工类学院的用户经常以较少的进馆次数借阅较多自科类图书,呈现出“少次多量”的特点,所以自科类图书的借阅册次并不是随着进馆次数的增多而增加的。



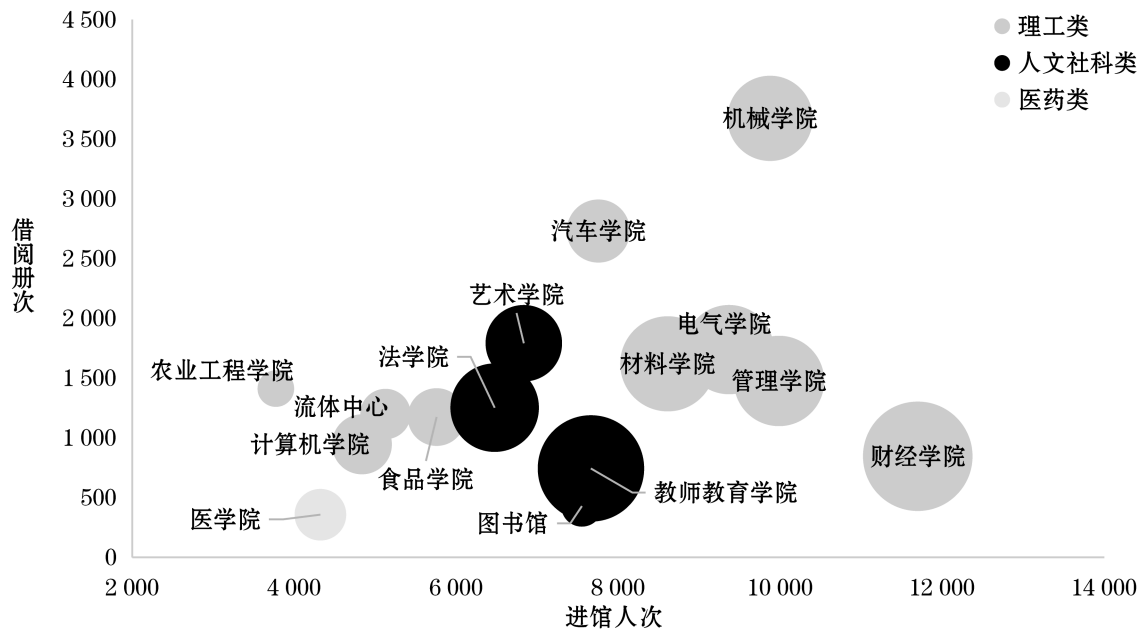


图6 不同学院研究生进馆人次、借阅册次、座位使用时长气泡图

表11 各类学院研究生借阅情况

学院	自然科学		社会科学		借阅总数
	借阅册次	占比	借阅册次	占比	
机械工程学院	3 098	84.32%	576	15.68%	3 674
汽车与交通工程学院	2 299	84.18%	432	15.82%	2 731
电气信息工程学院	1 535	88.27%	204	11.73%	1 739
材料科学与工程学院	1 145	70.68%	475	29.32%	1 620
农业工程学院	1 240	87.82%	172	12.18%	1 412
流体机械工程技术研究中心	937	78.08%	263	21.92%	1 200
食品与生物工程学院	871	74.19%	303	25.81%	1 174
计算机科学与通信工程学院	716	75.69%	230	24.31%	946
财经学院	138	16.31%	708	83.69%	846
管理学院	290	19.65%	1 186	80.35%	1 476

人文社科类学院的自科借阅和进馆次数之间相关性系数皆不显著则主要是因为该类学院研究生用户主要借阅社科类图书；医药类学院的两种图书借阅行为和其他在馆行为皆无显著相关性，可能是因为医药类学院本身较少，来馆用户偏少，借阅数量更是偏少，故难以计算相关程度。相关性分析结果见表12。

表12 不同类学院研究生进馆和借阅行为相关性

学科类型	入馆行为	自科借阅	社科借阅
理工类	进馆次数	不显著	0.088
	在馆时长	-0.032	0.076
人文社科类	进馆次数	不显著	0.069
	在馆时长	-0.034	0.03
医药类	进馆次数	不显著	不显著
	在馆时长	不显著	不显著

## 4.2.4 理工类学院研究生用户更倾向于将纸书借出馆阅读

理工类学院研究生用户的自科类图书借阅和座位使用行为的相关性系数为负,说明借阅的自科类图书越多,留在馆内使用座位的可能性越低。这与前文中得出的结论相互印证,理工类研究生用户来馆次数较少,一次性借阅较多数量的图书并带离图书馆。相关性分析结果见表13。

表13 研究生学院用户借阅和座位使用行为相关性

学科类型	借阅行为	座位预约次数	座位使用时长
理工类	自科借阅	-0.015	-0.016
	社科借阅	0.037	0.036
人文社科类	自科借阅	不显著	不显著
	社科借阅	0.057	0.057
医药类	自科借阅	0.062	0.062
	社科借阅	不显著	不显著

## 5 结论

为了多角度分析图书馆用户在馆行为的特点,同时讨论数据分析得到的结果可能产生的原因,本文采取了访谈的方法,以该校图书馆多位馆员、管理人员和一线服务人员为访谈对象,采用开放性问卷,根据数据分析的主要结论及图表,请受访者回答以下问题:在平时的工作中能否观察到某些/某类用户表现出来的行为特点与数据分析结果相印证?根据工作经验和平时的观察,能否分析数据分析结果可能产生的原因?通过对访谈资料的整理,主要有以下3种情况:

(1) 用户使用阅览室空间的实际情况与数据统计结果是相对应的,人数最少的几个时间点是由于放假期间,学生在校人数少产生的,同时因为人少,阅览室座位不紧张,不需要预约座位也有大量空位,故座位预约行为产生的几率也最低。人数最高点都是由于考试季的到来,如大学英语四六级考试、研究生考试、公务员考试、期末考试等,多数用户利用图书馆的空间和座位进行自己的学习活动、复习和刷题,所以进馆行为与座位使用行为的相关性最高。

(2) 纸质资源借阅情况整体偏少,但是纸质资源的实际利用率并非如借阅率一般低,主要体现在馆内阅读。每天在闭馆时,工作人员清理散落在图书馆各处的未借图书每天约700多册,大于流通借还系统产生的借阅数据。

(3) 2020—2021 学年使用座位预约的用户不到进馆用户的40%,主要原因是:①该馆引入的系统为第一代座位预约系统,操作太繁琐,用户使用不方便;②馆内布局调整较大,造成桌牌更换不及时,造成有号无座;③用户观念未养成;④管理时紧时松;⑤非考试季时,座位大量空余;⑥座位系统使用无针对性。

总体上看,数据分析的结果基本符合实际馆内情况。因为当前数据采集渠道的局限性,没有办法详细采集所有数据(如纸质资源实际使用情况,阅览室座位实际使用情况等),故可能造成部分相关性系数计算结果偏低,但依然可以从中看出一些规律和特点:

(1) 无论本科生还是研究生,理工类学院数量多,因此拥有最多的总进馆人次。但3类学科的院均进馆人次中,理工类学院无论本科生和研究生都只排第二,本科生排第一的是医药类学院,研究生排第一的是人文社科类学院。

(2) 各学院本科生用户在馆时长的差异较小,进馆次数和在馆时长间的相关性系数都在0.9以上,但是研究生用户差异较大。在同样的进馆次数条件下,人文社科类学院研究生用户的在馆时间更长,其次是理工类,最后是医药类。

(3) 本科生和研究生各学院用户进馆行为和借阅行为之间的相关性系数普遍低于0.1,即用户借阅图书册次的有序值变化值只有1%能用他们进馆次数和在馆时间的有序值来进行预测,可以认为所有学院的用户进馆后产生借阅行为的只是极少数。根据工作人员反馈,每天在闭馆时清理散落在图书馆各处的未借图书约700册,高于每天借出图书册次,说明更多用户选择内阅而非外借。内阅的数据难以统计和获取,而外借的数据相对更少,但是依然可以从中看出本科生和研究生各学院用户借阅的不同特点,比如在自科和社科两类图书的借阅比例上有所不同。在本科生用户当中,人文社科类学院的用户借阅社科类图书的数量都远大于自科类,理工类和医药类学院用

户借阅两类书的比例都在 1:1 到 1:2.2 之间,相对来说差距不大。但是在研究生用户当中,两类书借阅数量比例相差最小的学院,其比值就有 1:3,而最大的甚至相差百倍,明显倾向于自身专业领域。

(4) 本科生和研究生有共同的特点——借阅行为和座位预约行为之间的相关性系数普遍低于 0.1。因为进馆的人只有少数借书,借书的人又不一定使用座位,再加上座位管理系统自身存在一定的缺陷,难以严格限制用户使用座位时必须预约,所以借阅行为和座位预约行为之间相关性较低。研究生用户的自科借阅和座位使用行为之间呈现出负相关,体现出了研究生用户在借阅自科类图书时“少次多量”“借出馆阅读”的特点。

## 6 建议

分析用户入馆、在馆和用馆数据,揭示各类用户的行为特征和差异,可以助力完善高校图书馆资源管理的升级优化,也为高校图书馆的个性化服务提供参考。

### 6.1 根据馆内人流量调整资源分配和服务内容

用户进馆时间具有明显的分布特征。6 月和 10—12 月为进馆高峰,该时间段为复习备考时期,用户进馆人次和在馆时间都大幅增加,这段时间应该开放最多的阅览室座位;该时期用户来馆的目的性较为突出,绝大多数用户来馆都是为了复习备考,故图书馆可以联合各学院,将校内开展的与各类考试相关培训、讲座、动员的宣传活动在馆内展开,利用人流量的优势提高宣传效率,让更多用户了解活动的内容并有机会加入进来,进一步体现图书馆作为高校文化中心和交流中心的职能。人力资源的分配也可以根据人流量动态调整。高峰期用户进馆更早离馆更晚,图书馆在调整开馆和闭馆时间的同时也需重新分配馆内服务人员的值班时长和排班方式。在用户进馆高峰时间段,入馆人次和在馆时间的提升幅度都较大,但是图书借阅册次提升相对较小,可以看出进馆用户多数是为了使用空间资源进行自习。该时期图书馆除了应该增加阅览室座位的开放数量、增设工作人员岗位外,还应该适时地提供图书资源推荐服务:根据不同时间段考试的类型,或根据不同年级的开课情况,提供分类清晰的学习书目指南,引导用户使用馆内资源,避免考生在外面到处找不到好资源,真正有用的书籍却躲在图书馆书架上吃灰的情况。

高校 2 月、7 月、8 月是假期,图书馆进馆人次和借阅人次都处于最低谷,仅安排最低限度的人力和物力资源即可。除了上述极端情况之外,4 月、5 月、9 月是最接近月入馆人次平均值的月份,平时可以根据这几个月份的人流量,动态地调整空间资源及人力资源安排。

### 6.2 基于用户数据提供精准化服务

从入馆人次和图书借阅数量的角度来看,一些学院的用户表现出了对图书馆较高的依赖度,如财经学院、医学院、计算机学院、数学学院、文学院的本科用户,还有财经学院、管理学院、机械学院、汽车学院的研究生用户。因为这些学院用户是图书馆资源的最主要使用者,图书馆可以根据这些学院用户的阅读兴趣与倾向、馆藏图书利用情况等行为数据,调整和采取更加科学的文献保障政策,正向引导用户的阅读行为<sup>[10]</sup>;对于图书馆资源使用较少的学院用户,图书馆虽然难以通过他们的行为数据主动分析其资源需求,但是依然要保持足够的关注,保证资源荐购和建议提取的渠道通畅并及时回复,确保“小众”用户的个性化需求也在一定程度上被尽可能满足。

从借阅后的留馆情况来看,借阅了社科类图书的用户留馆阅读的可能性大于借阅自科类图书的用户,这一点在研究生用户当中体现得更加突出。因为借阅自科类图书后不留馆阅读而是直接带走的情况更多,所以图书馆应该适当调整相关服务及设施所在位置,如将自助借还机设置在更接近自科阅览室的位置;与此相对,社科类图书的用户相对倾向于留馆阅读,则自动售卖机、储物箱等设施可以安排在更方便他们使用的位置。

用户在馆内阅读纸质资源并不一定会外借,更多用户离馆前会将图书直接归还,而非刷卡借走。一些学院的借阅行为和座位使用行为之间存在正向的显著相关性(如数学学院、外国语学院等),说明该学院用户倾向于在馆内阅读馆藏纸质资源。因为内阅行为普遍大于借阅,所以这个倾向实际上应该更加突出。图书馆可以根据这类学院用户较常借阅的图书种类,在总座位数不变的前提下适当增加那些阅览室的座位数量,同时减少其他阅览室的座位数量(因为对于仅自习的用户而言,坐在存放哪类图书的阅览室都没有区别),有效分配空间,为用户提供便利,提高服务精准度<sup>[10]</sup>。

### 6.3 加强用户行为数据采集以实现资源智能推送

已有的数据采集系统并不能满足所有的用户行为分析需求,如纸质资源内阅情况、电子资源使用情况、阅览室座位利用情况等。高校图书馆应该加强数字资源使用情况统计,同时增加纸质资源和座位使用情况的获取渠道,如利用 RFID 技术应用实现图书精准定位与高效盘点、引进新一代座位预约和管理系统等<sup>[21]</sup>。通过追踪和充分分

析用户行为,主动向用户推送专业相关专著、优秀文艺著作或科普图书。通过及时高效地满足用户的个性化需求,让用户逐渐对图书馆的个性化服务产生获得感和依赖感<sup>[22]</sup>。

#### 注释

- [1]王凌.大学生利用图书馆与学习成绩的关联性实证研究——以首都医科大学为例[J].图书情报工作,2017(24):39-44.
- [2]梁秀霞.大学图书馆竞争力与大学排名相关性分析——基于65所大学的实证研究[J].现代情报,2016(1):96-101+166.
- [3]中华人民共和国教育部.教育部关于印发《普通高等学校图书馆规程》的通知[EB/OL].[2023-01-04].http://www.moe.gov.cn/srcsite/A08/moe\_736/s3886/201601/t20160120\_228487.html.
- [4]余以胜,张玉峰.数据挖掘技术与用户知识获取[J].情报理论与实践,2003(1):65-66+74.
- [5]王伟,张征芳,王海明.基于数据挖掘的图书馆读者行为分析[J].现代图书情报技术,2006(11):51-54+60.
- [6]李爱国,孟祥保,张长秀.图书馆用户信息资源化:概念解析与理论框架构建[J].图书情报工作,2015(13):6-12.
- [7]Kanarkard W, Seemajaruek C, Pongsuwan T, et al. Predictive Analytic of Library Patron Behavior: Proceedings of the 3rd International Conference on Communication and Information Processing (ICCIIP 2017) [Z]. 3rd International Conference on Communication and Information Processing (ICCIIP), 2017: 1-5.
- [8]李丹,高建忠.基于用户画像的图书馆推荐服务初探[J].图书馆,2019(7):66-71.
- [9]吴英梅,何臻.高校图书馆对学生学业科研影响的实证研究——以北京师范大学为例[J].图书情报工作,2014(20):73-77+90.
- [10]李荣,王庆石,李双,等.基于读者行为大数据的图书馆精准服务与管理创新研究——以东北财经大学图书馆为例[J].图书馆研究,2018(3):19-23.
- [11]李庆丰,王超.高校图书馆利用与学生学习结果关系的实证研究——基于“2017年首都大学生就读经验调查”[J].图书情报工作,2019(21):56-65.
- [12]谭云江,凌云,闫海峰,等.基于数据挖掘技术的高校图书馆读者行为研究[J].图书情报工作,2010(S2):359-362.
- [13]易明,张展豪,李怡.大学生图书借阅行为的人类动力学分析[J].图书馆学研究,2019(22):83-93.
- [14]邢荣华,朱玉珍,张静,等.基于流通日志的大学生借阅规律和影响因素的实证研究[J].现代情报,2017(8):116-121+166.
- [15]施国良,张潇潇,杨小莉.高校读者群体差异对其借阅行为和阅读偏好的影响研究[J].图书馆,2020(4):59-64.
- [16]张晓阳.高校图书馆流通出纳岗位人力资源配置研究[J].图书情报工作,2005(3):108-111.
- [17]尹相权,李书宁,弓建华.基于系统日志的高校图书馆研究间用户利用行为分析[J].现代情报,2018(1):115-120.
- [18]Aguilar-Moreno E, Montoliu-Colas R, Torres-Sospedra J. Indoor Positioning Technologies for Academic Libraries: Towards the Smart Library [J]. Profesional de la Informacion, 2016, 25 (2): 295-302.
- [19]张晓阳,刘洋,苗松,等.2020学年江苏大学图书馆用户行为记录数据集[J].图书馆杂志,2022(10):97-103.
- [20]朝乐门.数据科学理论与实践:第2版[M].北京:清华大学出版社,2019.
- [21]刘莉.RFID技术在图书馆中的应用[J].无线互联科技,2022(11):98-100.
- [22]吴钗芸.大数据环境下高校图书馆读者行为分析——以湖北工程学院图书馆2015年度统计为例[J].湖北工程学院学报,2017(3):120-125.

**张晓阳** 江苏大学图书馆副馆长,研究馆员,硕士生导师。研究方向:科学计量学、用户信息研究。

**许欣** 江苏大学科技信息研究所2020级图书情报专业硕士研究生。研究方向:数据分析、学科服务。通讯作者。

**刘洋** 江苏大学科技信息研究所2020级图书情报与档案管理专业硕士研究生。研究方向:数据科学与治理。