



# CADAL用户行为挖掘与分析

CADAL项目管理中心 金佳丽

2023.07.04

---

# CONTENTS

- 1 CADAL项目背景
- 2 用户行为数据集
- 3 使用建议

PART ONE

# CADAL项目背景

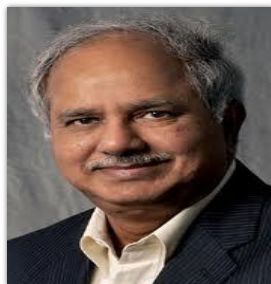


# 缘起

2000年，中美计算机科学家与图书馆联合发起“中美百万册数字图书馆国际合作计划”（China-America Digital Academic Library，简称CADAL），开始了全球最早的大规模数字化资源工程（Mass Digitization）。



潘云鹤  
中国工程院院士



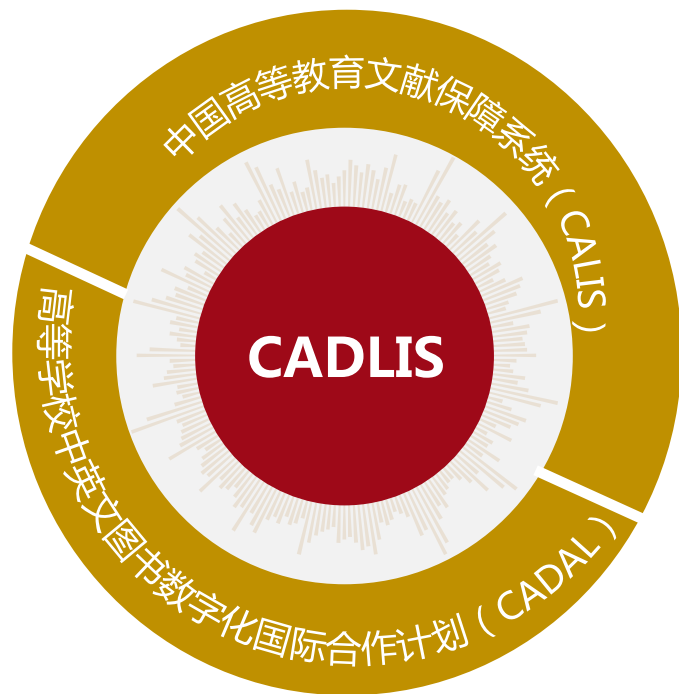
Raj Reddy  
计算机图灵奖得主



# CADAL进入211工程 公共服务体系

计社发[2002]1305号

2002年9月，国家发改委、教育部、财政部下发《关于“十五211工程”建设的若干意见》，支持“高等学校中英文图书数字化国际合作计划（CADAL）”的实施，为高等学校教学科研提供强有力的数字资源支持。



2009年1月，教育部批复浙江大学立项建设“数字图书馆教育部工程研究中心”。8月，基于“中英文图书数字化国际合作计划（CADAL）”的建设成果，教育部高教司通过了《大学数字图书馆国际合作计划（China Academic Digital Associative Library, CADAL）建设可行性研究报告》CADAL项目二期建设正式启动。

2011年2月25日，“数字图书馆教育部工程研究中心”通过教育部组织的验收。





# 发展历程

2000

中美科学家共同倡议，数字化人类1%的知识，即100万册图书。如持之以恒，则10年可以完成。

敢为天下先  
We did it first

2001-2006

项目一期  
100万册图书  
数字化

2007-2012

项目二期，新增150万册图书数字化，标准规范、服务网络、技术服务升级

一览众山小  
We do it better

2013-2018

项目运维阶段

- 整合海量资源
- 融合先进技术
- 泛在个性服务
- 全球开放合作

同时开展：数据合作，联盟建设，特藏建设，数字人文

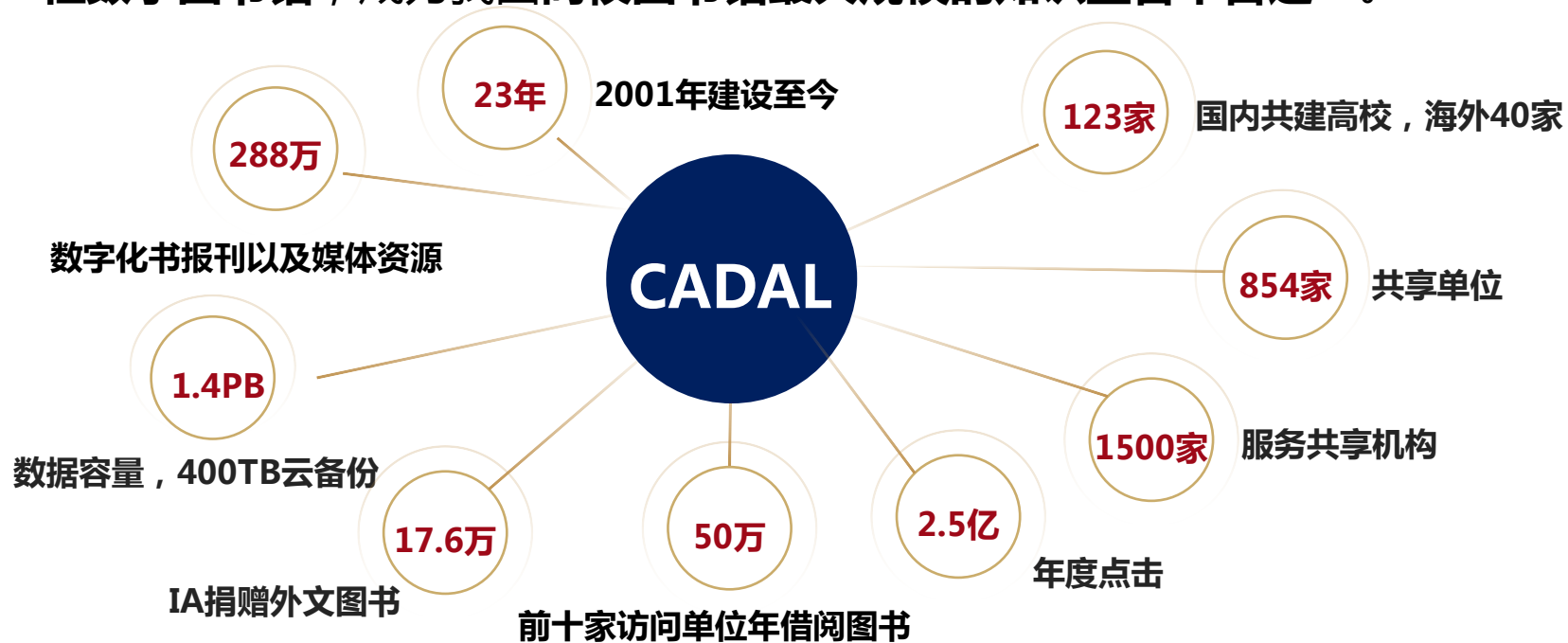
更上一层楼  
We can do better

2019-

数据助力知识服务  
大数据——大知识

# 发展现状

CADAL项目的建设把中国高校图书馆由购买资源共享有效推动到了资源的共建共享，建成了拥有超过288万册中英文电子图书的**全球最大公益性数字图书馆**，成为我国高校图书馆最大规模的知识整合平台之一。





# 资源建设

截至2022年12月，CADAL入库总量**2,888,088**册（件），在线量为**2,720,420**册（件）。



古籍  
( 242939 )




民国图书  
( 176327 )



民国期刊  
( 155134 )



当代图书  
( 820690 )



外文图书  
( 713193 )



满铁  
( 18704 )



靖宇县  
地方志  
( 17381 )



侨批  
( 50065 )



民国文献大全



墓志拓片



中国写本文献数据库



甲骨数字化



老照片



蒋介石资料



侨批



满铁资料



地方志



已建成**13**个特藏库

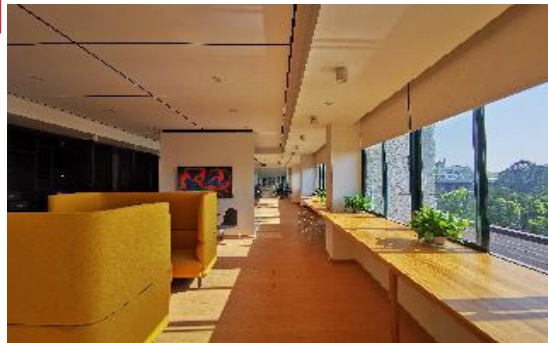


## 愿景

项目以“共建共享”理念为指导思想，以先进的技术手段，全面整合国内高校图书馆、图书情报服务机构、学术研究机构所拥有或生产的各类信息资源及其相关服务，有重点地引进、共享国际相关机构的各类信息资源与服务。对包括书画、建筑工程、篆刻、戏剧、工艺品等在内的多种类型媒体资源进行数字化整合，项目建成的资源覆盖理、工、农、医、人文、社科等多种学科，通过因特网提供一站式的个性化知识服务，向参与建设的高等院校、学术机构提供教学科研支撑。

## 使命

构建拥有多学科、多类型、多语种海量数字资源的，由国内外图书馆、学术组织、学科专业人员广泛参与建设与服务，具有高技术水平的学术数字图书馆，成为国家创新体系信息基础设施之一。目前，CADAL已进入项目运维期，现阶段主要致力于保证已建成的项目成果持续有效地发挥作用，并不断增强高等教育公共服务基础设施建设。





# 用户行为数据集

PART TWO



# 大学数字图书馆国际合作计划

CHINA ACADEMIC DIGITAL ASSOCIATIVE LIBRARY

全部 ▾

请输入搜索内容



# CADAL

网址：<https://cadal.edu.cn>



资源量 **272万**



共建共享单位 **977家**



当天检索量 **4085次**



当天阅读量 **16877次**



当天访问量 **294690次**



当天API访问量 **233137次**



## 大学数字图书馆国际合作计划 (CADAL) 数据监控平台

资源总数

2,848,797

用户总数

614,828

共建共享单位

977

当年新增用户数

24,862

在线登录人数

150

人

未登录访客数

4907

人

当天网站检索次数

4105

次

当天阅读量

16953

页

当天访问量

295405

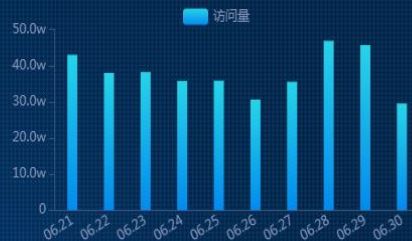
次

当天API访问量

233720

次

近10天网站访问量/近10天API访问量



近30天热门资源Top10

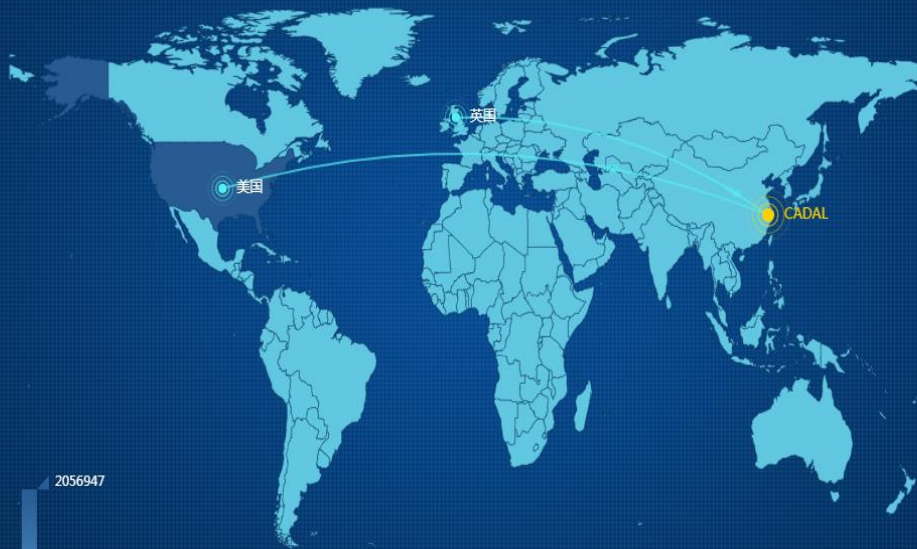


中华教育文化基金董事会报告  
访问量: 1,242



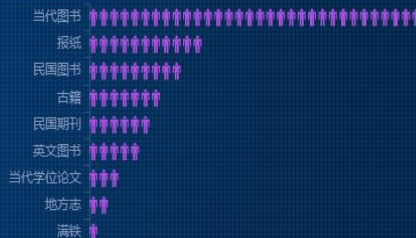
严复集  
访问量: 1,198

近30天搜索热词



中国地图 / 世界地图

近10天资源访问量



当年机构访问量排行

排名	所属机构	访问量
1	西安工程大学	1,807,838
2	南京师范大学	1,747,149
3	南京大学	1,358,031
4	浙江万里学院	1,148,339
5	南开大学	1,147,088
6	南京航空航天大学	1,076,478
7	厦门大学	1,036,993
8	湖南大学	1,035,632

用户行为统计

请选择资源类型

请选择机构

Top: 10

2023-01-01 至 2023-06-29

打开多选 筛选列

序号	用户ID	用户名	所属机构	访问总量	资源访问量	借阅量	借阅本数	阅读页数	阅读本数	登录次数
1	9014			442414	441538	10357	4027	441142	12857	1
2	9014			334336	80746	264	97	14763	599	81
3	9014			302220	300480	7824	2638	300060	7559	4
4	9014			120730	120080	3137	1200	119958	3216	2
5	9014			100247	100112	1987	782	100095	3543	0
6	9014			86576	86509	1967	785	86479	2821	0
7	9014			74595	70143	3807	523	67862	1397	398
8	9014			65934	65736	307	138	65691	481	0
9	9014			62433	62364	1169	508	62349	2266	0



名称

检索记录.csv

借阅记录.csv

浏览记录.csv

用户表.csv

资源表.csv



A	B	C	D	E	F	G	H	I	J	K	L	
id	user_id	user_name	keyWord	create_time	ip		search_record (检索记录)					
626435	1340000	已屏蔽	萨满教	2022/1/1 0:07	36.104.120.**							
626436	1200000	已屏蔽	文学运动史料选·第二册	2022/1/1 0:26	10.186.42.***		字段	类型	初始值	描述	备注	
626437	1160000	已屏蔽	德育原理	2022/1/1 0:51	202.194.12.**		id	int	自增	主键id		
626438	1160000	已屏蔽	德育原理	2022/1/1 0:54	202.194.12.**		user_id	varchar	not null	用户id		
626439	1160000	已屏蔽	德育原理	2022/1/1 0:56	202.194.12.**		user_name	varchar	not null	用户名		
626440	1160000	已屏蔽	教育伦理学	2022/1/1 0:58	202.194.12.**		keyWord	varchar	not null	检索词		
626441	1160000	已屏蔽	教育伦理学	2022/1/1 0:58	202.194.12.**		create_time	date	not null	检索日期		
626442	1160000	已屏蔽	教育伦理学	2022/1/1 1:01	202.194.12.**		ip	varchar	not null	检索ip		
626443	1160000	已屏蔽	公民教育	2022/1/1 1:05	202.194.12.**							
626444	1160000	已屏蔽	公民教育	2022/1/1 1:05	202.194.12.**							

名称

检索记录.csv

借阅记录.csv

浏览记录.csv

用户表.csv

资源表.csv

### 民国经济史

荐购

作者：朱斯煌

出版社：银行学会

馆藏单位：四川大

出版时间：1948-0

资源类型：民国图

标签：民国;经济史

主题：民国;经济

说明：冯子明 汤心

分享 付费

试读

提醒

借阅后有7天阅读时长,本章节剩余 10次

借阅次数,是否借阅此书第 1 章?

暂不借阅

借阅此书

A	B	C	D	E	F	G	H	I	J	K	L	M	N
id	userid	username	resourceisno	resourceisno	resourcechapter	resource	cover_image	author	remaining	borrow_start_time	borrow_end_time	ip	
61786	1220000	已屏蔽	6374711	6374711	鲁迅的文学道路：论	306/374/711/otiff/	C冯雪峰（著）			2022/7/18 18:14	2022/7/25 18:14	10.79.20.**	
99649	1210000	已屏蔽	33078076	33078076	实用阻燃技术	33/078/076/otiff/	C(欧育湘			2022/1/8 17:30	2022/1/15 17:30	10.79.20.**	
114750	1370000	已屏蔽	33146576	33146576	唐诗300首钢笔字帖	33/146/576/otiff/	C王永兴书			2022/2/1 2:53	2022/2/8 2:53	50.76.33.***	
115028	1150000	已屏蔽	6887541	6887541	唤醒中国：国民革命	06/887/541/otiff/	C费约翰（著）；李恭			2022/7/23 10:21	2022/7/30 10:21	10.190.77.***	
118299	1370000	已屏蔽	33055180	33055180	历代草行名作鉴赏	33/055/180/otiff/	C沈鸿根			2022/1/25 4:42	2022/2/1 4:42	50.76.33.***	
118320	1370000	已屏蔽	13264073	13264073	中学生字帖	13/264/073/otiff/	C本社			2022/1/25 2:21	2022/2/1 2:21	50.76.33.***	
124907	1380000	已屏蔽	7111212	7111212	数学物理方法	07/111/212/otiff/	C吴崇试，（编著）			2022/3/5 22:58	2022/3/12 22:58	58.194.168.***	
150162	1150000	已屏蔽	6876730	6876730	苦笑录	06/876/730/otiff/	C陈公博（著）			2022/7/23 10:21	2022/7/30 10:21	10.189.182.***	
251421	1390000	已屏蔽	53000854	53000854	湖南私立大麓中学校	53/000/854/otiff/	C张平子；黄衍钧			2022/3/3 16:44	2022/3/10 16:44	220.170.199.**	
342568	1380000	已屏蔽	2046208	2046208	合并字學篇韻便覽(六	02/046/208/otiff/	C(明)徐孝輯			2022/5/16 10:51	2022/5/16 10:54	106.128.116.**	
420459	1190000	已屏蔽	11417522	11417522	Who's Who in China	11/417/522/otiff/	C00000001.jpg			2022/1/1 20:53	2022/1/8 20:53	123.135.136.**	
429077	1190000	已屏蔽	50003017	50003017	交通大学同学录	50/003/017/otiff/	C[交通大学编			2022/1/1 20:53	2022/1/8 20:53	123.135.136.**	
429112	1190000	已屏蔽	50003344	50003344	交通大学毕业生调查	50/003/344/otiff/	C(胡端行编			2022/1/1 20:53	2022/1/8 20:53	123.135.136.**	
429167	1190000	已屏蔽	53000677	53000677	铁道部职员录	53/000/677/otiff/	C铁道部总务司人事科			2022/1/1 20:53	2022/1/8 20:53	123.135.136.**	
432038	1190000	已屏蔽	6359638	6359638	辛亥以后十七年职官	406/359/638/otiff/	C(刘寿林（编）			2022/1/1 20:53	2022/1/8 20:53	123.135.136.**	

名称

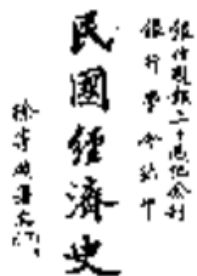
检索记录.csv

借阅记录.csv

浏览记录.csv

用户表.csv

资源表.csv



## 民国经济史

荐购

作者：朱斯煌

出版社：银行学会银行通报社·上海

馆藏单位：四川大学

出版时间：1948-01(民国三十七年)

资源类型：民国图书

标签：民国;经济史;上海;四十年代;民国三十七年;专著 [添加标签](#)

主题：民国；经济史；金融

说明：冯子明 汤心儒 鲍文熙

[分享](#) [付费打印](#) [我是作者](#) [捐赠](#) [完善](#)

试读

章节阅读

收藏

id	userid	username	resourceid	resourcename	chapter	resource_cover_image	author	browsing_datetime	type	ip
5764119	1380000	已屏蔽	6397341	民族与社会		06/397/341/otiff/	费孝通(著)	2022/1/1 0:03		113.55.110.***
5764120	1370000	已屏蔽	1010781	房屋构造学		01/010/781/otiff/	唐英, 王	2022/1/1 0:06		121.249.15.***
5764121	1370000	已屏蔽	6350330	房屋构造学		06/350/330/otiff/	唐英(编著)	2022/1/1 0:07		121.249.15.***
5764122	1340000	已屏蔽	33136103	满族萨满教研究		33/136/103/otiff/	富育光 孟	2022/1/1 0:08		36.104.120.**
5764123	1440000	已屏蔽	6397172	社会研究方法		06/397/172/otiff/	(美)艾尔·	2022/1/1 0:08		113.55.110.***
5764124	1370000	已屏蔽	1000088	房屋工程		01/000/088/otiff/	华通斋著	2022/1/1 0:15		121.249.15.***
5764125	1370000	已屏蔽	11104009	混凝土工程施工细则		11/104/009/otiff/	江苏省公路	2022/1/1 0:15		121.249.15.***
5764126	1370000	已屏蔽	4101737	土木工艺		04/101/737/otiff/	科学书报编	2022/1/1 0:16		121.249.15.***
5764127	1370000	已屏蔽	8006346	建筑构造学		08/006/346/otiff/	唐英(编著)	2022/1/1 0:16		121.249.15.***
5764128	1370000	已屏蔽	4101773	建筑构造浅释		04/101/773/otiff/	盛承彦	2022/1/1 0:17		121.249.15.***
5764129	1200000	已屏蔽	6390555	瞿秋白文集：文学编·第三卷		06/390/555/otiff/	瞿秋白(著)	2022/1/1 0:23		10.186.42.***
5764130	1200000	已屏蔽	6374086	文学运动史料选·第二册		06/374/086/otiff/	北京大学	2022/1/1 0:26		10.186.42.***

名称

检索记录.csv

借阅记录.csv

浏览记录.csv

用户表.csv

资源表.csv

lid	username	password	status	role_id	real_name	balance	head_port	booklist	comment	rbooklist	hidden	braffiliati	email/邮箱	sex/性别	birthday	hometown	habitual_school	学major/专	telephone	brief	registrat	account_t
a014e7844	已屏蔽	已屏蔽	1	20	已屏蔽	99930.91	已屏蔽	no	no	no	no		*****@female	#####	东北	西京	哈哈	图书情报	188*****	一个爱	#####	F
c65cbde41	已屏蔽	已屏蔽	1	20	已屏蔽	9950	已屏蔽	yes	yes	yes	yes	浙江大学	*****@male	#####	北京	南京	cx	管理	150*****		#####	D
c289c37e1	已屏蔽	已屏蔽	1	18	已屏蔽	0	已屏蔽	yes	yes	yes	yes	浙江大学	*****@male	#####	其他-其他		浙江大学		135*****		#####	D
045b5532a	已屏蔽	已屏蔽	1	20	已屏蔽	0	已屏蔽	no	no	no	no	浙江大学	*****@male	#####			浙江大学		136*****		#####	F
b39e761e5	已屏蔽	已屏蔽	1	17	已屏蔽	0	已屏蔽	yes	yes	yes	yes		*****@female	#####	背景	北京	某学校	某专业	188*****	开心的	#####	D
ab3e9d721	已屏蔽	已屏蔽	1	18	已屏蔽	0	已屏蔽	yes	yes	yes	yes	浙江大学	*****@female	#####	浙江-杭州		浙江大学				#####	D
7d151cae5	已屏蔽	已屏蔽	1	18	已屏蔽	0	已屏蔽	yes	yes	yes	yes	浙江大学	*****@female	#####							#####	D
2a50b0314	已屏蔽	已屏蔽	1	19	已屏蔽	0	已屏蔽	yes	yes	yes	yes	江西师范大	*****@male	#####	其他-其他		浙江大学		135*****		#####	D
17ebf7a45	已屏蔽	已屏蔽	1	19	已屏蔽	0	已屏蔽	yes	yes	yes	yes	中北大学	*****@male	#####	山西-太原		其他		182*****		#####	D
a7062266a	已屏蔽	已屏蔽	1	18	已屏蔽	0	已屏蔽	yes	yes	yes	yes	中国人民大	*****@male	#####	北京-海淀区		人民大学				#####	D
51409fba8	已屏蔽	已屏蔽	1	18	已屏蔽	0	已屏蔽	yes	yes	yes	yes	复旦大学	*****@male	#####	上海-杨浦		复旦大学		151*****		#####	D
4142fae67	已屏蔽	已屏蔽	1	17	已屏蔽	0	已屏蔽	yes	yes	yes	no		*****@male	#####	广西-南宁		其他				#####	D
84647d5bt	已屏蔽	已屏蔽	1	18	已屏蔽	0	已屏蔽	yes	yes	yes	yes	四川大学	*****@male	#####	北京-东城区		四川大学				#####	D

名称

检索记录.csv

借阅记录.csv

浏览记录.csv

用户表.csv

资源表.csv

D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
enum/	cno	title/题名	title_ol	creator/	creator_c	subject/	subject_c	descripti	descripti	publisher	publisher	contribu	contribu	date/日期	date_ol	type/类
190	500111	御纂朱子全	御纂朱子全	宋朱熹 (著)	(宋)朱熹 (著)			本書包括卷	本書包括卷	5	苏州大学			1714	清康熙五十	古籍
215	200635	四書大成	四書大成	清沈器等 (纂訂)	(清)沈器等 (纂訂)			本書包括	本書包括	第一冊~第五冊。	原書(清)陆摺	(清)陆摺		1694	清康熙三十	古籍
86	701212	宋學士全	宋學士全	明宋濂 (著)	(明)宋濂 (著)			本書包括	本書包括	老南陽彭始	南陽彭始	搏		1709	清康熙四十	古籍
103	700744	百美新詠	百美新詠	第三冊	不详			本書包括	本書包括	第一冊~第三冊。				1804	清嘉慶九年	古籍
106	700032	宋詩鈔選	宋詩鈔選	清吳孟舉 (輯)	(清)吳孟舉 (輯)			本書包括	本書包括	三餘草堂	三餘草堂			1671	清康熙十年	古籍
92	701212	宋學士全	宋學士全	明宋濂 (著)	(明)宋濂 (著)			本書包括	本書包括	老南陽彭始	南陽彭始	搏	苏州大学	1709	清康熙四十	古籍
95	703812	皕鶴堂批	皕鶴堂批	明蘭陵笑笑生 (撰)	(明)蘭陵笑笑生 (撰)			本書包括	本書包括	老湖南	湖南	(清)張竹坡	(清)張竹坡	1911	清末(1644	古籍
133	200020	五經大全	五經大全	清來爾紉 (纂輯)	(清)來爾紉等 (纂輯)			本書包括	本書包括	卷一~卷十。					[清末]	古籍
90	700434	唐詩筌蹄	唐詩筌蹄	清黃大鴻 (註釋)	(清)黃大鴻 (註釋)			本書包括	本書包括	卷一~卷四。	原書部分頁殘損			1715	清康熙五十	古籍
124	200656	四書字詁	四書字詁	清段調廷 (原稿)	(清)段調廷 (原稿)			本書包括	本書包括	卷一~卷七十八。	原書(清)黃本	(清)黃本		1849	清道光二十	古籍
144	704021	鑑窗瑣話	鑑窗瑣話	清於源 (撰)	(清)於源 (撰)			本書包括	本書包括	卷一~卷八。				1847	道光二十七	古籍
123	701121	梁溪先生	梁溪先生	宋李剛 (撰)	(宋)李剛 (撰)			本書包括	本書包括	卷一~卷一百八十。	原書缺卷一			1644	清(1644-1	古籍

PART THREE

使用建议



本数据集涵盖2022年CADAL网站完整的用户行为数据，涉及用户信息、资源信息、检索信息、借阅信息、浏览信息相关字段150个。对探寻高校读者电子资源阅读动态、基于读者需求的电子资源采购、基于阅读行为的个性化推荐服务、基于读者交互行为的网站功能评估、不同院校/专业读者的电子书阅读喜好研究等众多方向都具有丰富的研究和应用价值。例如，可以使用数据分析软件或自主编程进行如下分析：

- （1）构建用户阅读素养模型。基于用户行为数据进行建模，以更加精准科学的方式对用户阅读进行分析和解读；
- （2）进行个体/群体偏好分析，如浏览、借阅、评论、收藏，根据用户的阅读行为进行个性化推荐服务；
- （3）进行用户共建参与度分析，如：捐赠副本、完善资源信息；





THANK YOU

谢谢！