

CADAL元数据整理方式及意义： ——以蒋介石研究资料库建设为例

龙向洋

复旦大学图书馆

2019.2.16

CADAL民国时期图书

CADAL收录民国时期图书17.3万册，资源比较丰富。有必要对其元数据进行整理，在统计分析、比较评估的基础上为特藏资源建设提供数据依据。

 [搜索](#)

CADAL图书

古籍	清朝	明朝	宋朝	编著	钦定	文集	先生	更多+
民国图书	民国	上海	三十年代	编著	四十年代	中国	清朝	更多+
民国期刊	公报	民国	三十年代	月刊	周刊	四十年代	周报	更多+
现代图书	九十年代	编著	当代	北京	中国	五十年代	八十年代	更多+
学位论文	九十年代	研究	当代	中国	论文	八十年代	硕士学位	更多+
报纸	福建日报	浙江日报	云南日报	江西日报	1975年	1977年	1973年	更多+
英文	volume	report	from	history	study	new	book	更多+
特色资源	满铁	地方志	侨批					

CADAL民国时期图书元数据（DC）

CADAL数字资源描述是采用通行**DC**元数据，具有较好的开放性和兼容性。其严谨和规范不如**MARC**数据，因而对其元数据的整理也是很有必要的。

id	字段	示例
1	DCTitle	科学的学庸
2	DCCreator	蒋中正
3	DCCContributor	
4	DCPublisher	黄埔出版社·[上海]
5	DCDate	1939-10(民国二十八年)
6	DCSubject	科学 学庸 学理
7	DCDescription	
8	DCType	民国图书
9	DCFormat	Image/Djvu(.djvu)
10	DCSource	
11	DCRights	
12	DCIdentifier	
13	DCLanguage	chi
14	DCRelation	
15	DCCoverage	

元数据整理方式

以朴素的观念和严谨的规范，在数据库内部，通过数据的聚合、规则的确立、文本的析分、数据的比较等方式予以实现。

ID	整理内容	数据来源
1	中图分类	MARC
2	书名索引	DCTitle
3	人名索引	DCTitle/DCCreator/DCCContributor
4	地名索引	DCPublisher/DCTitle/DCCreator/DCCContributor

元数据整理：中图分类

通过DC与MARC记录之间的数据比较，可以增加完整的“中图分类”字段内容，以便于按学科进行统计分析以及相关专题资料库的建设。

bib	中图分类	书名	字段	描述
02009152	D693.09	中国之命运	DCTITLE	中國之命運
09006820	D693.09	中国之命运	DCTITLE	中国之命运
02009151	D693.09	中国之命运	DCTITLE	中國之命運
09006818	D693.09	中国之命运	DCTITLE	中国之命运
02009150	D693.09	中国之命运	DCTITLE	中國之命運
11108328	D693.09	中国之命运	DCTITLE	中国之命运
16074290	D693.09	中正革命语录	DCTITLE	中正革命语录
13062755	D693.09	总裁建国言论选辑	DCTITLE	总裁建国言论选辑. 卷上
11109361	D693.09	总裁建国言论选辑	DCTITLE	总裁建国言论选辑
16072715	D693.09	总裁训词集	DCTITLE	总裁训词集
09006634	D693.09	总裁训词选辑	DCTITLE	总裁训词选辑
13052541	D693.09	总裁言论	DCTITLE	总裁言论
09002931	D693.09	总裁言论	DCTITLE	总裁言论
09002661	D693.09	总裁言论	DCTITLE	总裁言论· 第四册
09000000	D693.09	总裁言论	DCTITLE	总裁言论· 第一册

元数据整理：书名索引

通过文字统一、格式规范以及数据校对，编制“书名分类索引”，不仅可以增强浏览可读性，而且便于资源比较评估。

IDT	分类	书名	频率	
23501	D693.09	蒋介石之言论	1	
23502	D693.09	蒋介石最近论文集	1	
23503	D693.09	蒋介石最近言论	1	
23517	D693.09	蒋委员长阐述实施宪政真义	1	
23534	D693.09	蒋委员长抗建言论集	1	
23536	D693.09	蒋委员长抗战书告辑要	1	
23537	D693.09	蒋委员长抗战问答	1	
23538	D693.09	蒋委员长抗战言论集	3	
23542	D693.09	蒋委员长论抗战必胜训词释义	1	
23547	D693.09	蒋委员长为持久抗战告全国国民	1	
23558	D693.09	蒋委员长训词选辑	2	
DCTITLE		DCCREATOR	DCCONTRIF	DCPUBLISHER
蒋委员长训词选辑·第四册		[军事委员会委员长侍从室]	[NULL]	[军事委员会委员长侍从室]·[出版地不详]
蒋委员长训词选辑·第一册		[军事委员会委员长侍从室]	[NULL]	[军事委员会委员长侍从室]·[出版地不详]
23559		D693.09	蒋委员长言论集	4
DCTITLE		DCCREATOR	DCCONTRIF	DCPUBLISHER
蒋委员长言论集		中国国民党中央执行委员会	不详	不详
蒋委员长言论集		蒋介石	不详	光华图书公司
蒋委员长言论集		蒋介石	不详	光华图书公司
蒋委员长言论集		蒋介石	不详	中国国民党中央执行委员会宣传委员会

元数据整理：人名索引

通过对元数据文本析分，并参照《民国时期总书目》著者索引，整理人名资料。

词表	条目统计	频率统计	总目
蒋介石	357	492	456
林纾	189	339	317
郭沫若	375	429	299
鲁迅	582	1147	284
孙中山	333	428	250
毛泽东	213	266	225
高尔基	272	343	208
巴金	268	294	190

bib	字段	描述1
01011748	DCCREATOR	王尔德著 巴金译
01011786	DCCREATOR	巴金译
01012017	DCCREATOR	巴金著
01012031	DCCREATOR	克鲁泡特金著 巴金译
01012925	DCCREATOR	巴金译
02011844	DCCREATOR	巴金等译

元数据整理：地名索引一

从DC元数据中整理出地名资料，按照行政区划进行数据组织，形成地方文献资料库。

	浙江	地名	条目统计	频率统计
+	浙江省 杭州市	杭州	953	1091
+	浙江省 杭州市 淳安	淳安	2	4
+	浙江省 杭州市 富阳	富阳	2	2
+	浙江省 杭州市 临安	临安	12	14
+	浙江省 杭州市 桐庐	桐庐	9	10
+	浙江省 杭州市 余杭	余杭	12	12
+	浙江省 嘉兴市 海盐	海盐	8	10
+	浙江省 嘉兴市 桐乡	桐乡	24	25
+	浙江省 丽水地区 遂昌	遂昌	3	3
+	浙江省 宁波市 鄞县	鄞县	488	519
+	浙江省 衢州市 龙游	龙游	16	16
+	浙江省 绍兴市 上虞	上虞	299	316
+	浙江省 绍兴市 绍兴	绍兴	139	151

元数据整理：地名索引二

可以通过人名索引与人物传记资料关联方式整理元数据中地名资料。

通用名	籍贯	生年	卒年	条目统计	频率统计
张寿镛	浙江鄞县	1876	1945	1099	1099
刘承干	浙江湖州	1882	1963	713	855
罗振玉	浙江上虞	1866	1940	383	384
张元济	浙江海盐	1867	1959	241	241
王国维	浙江海宁	1877	1927	199	202
蔡元培	浙江绍兴	1868	1940	172	205
傅东华	浙江金华	1893	1971	171	184
何炳松	浙江金华	1890	1946	163	171
周作人	浙江绍兴	1885	1967	143	157
金兆梓	浙江金华	1889	1975	133	133
陆费逵	浙江嘉兴	1886	1941	119	119

bib	字段	描述1
11104987	DCCONTRIBUTOR	舒新城 陆费逵
11105407	DCCONTRIBUTOR	陆费逵 张相 戴克敦
11106949	DCCREATOR	陆费逵

CADAL民国时期图书：蒋介石研究资料

通过表中6个
检索词，可检
出357条记录。

通用名	词表	条目统计	频率统计
蒋中正	蒋介石	191	236
蒋中正	蒋委员长	100	110
蒋中正	蒋中正	91	96
蒋中正	蒋主席	37	38
蒋中正	蒋总裁	5	5
蒋中正	蒋总司令	7	7

bib	字段	描述1	描述
03030673	DCTITLE	民国二十年蒋总司令言论集	民国二十年蒋总司令言
13052768	DCTITLE	民国十九年蒋总司令言论之一	民国十九年蒋总司令言
16003513	DCTITLE	蒋总司令言论	蒋总司令言论
16003516	DCTITLE	蒋总司令最近言论	蒋总司令最近言论
16004790	DCTITLE	国民革命之领袖蒋总司令最近	国民革命之领袖蒋总司
16072194	DCDESCRIPTION	此书字迹不清晰，此书原书缺	此书字迹不清晰，此书
59238078	DCTITLE	民国十八年蒋总司令言论集	民國十八年蔣總司令言

检索词在元数据各字段的分布情况。

字段	检索词
DCTitle	蒋介石(60), 蒋委员长(84), 蒋中正(6), 蒋主席(37), 蒋总裁(5), 蒋总
DCCreator	蒋介石(95), 蒋委员长(14), 蒋中正(83),
DCSubject	蒋介石(65), 蒋委员长(8), 蒋中正(1), 蒋主席(1),
DCPublisher	蒋中正(1),
DCDescription	蒋介石(9), 蒋委员长(1), 蒋中正(3), 蒋总司令(1),
DCContributor	蒋介石(3),
DCRelation	蒋委员长(3),

元数据人名关联一：蒋中正档案(“国史馆”)

蒋中正档案
元数据人名
统计：与蒋
介石关系密
切的人名整
理。

阎锡山	6529	题名摘要
蒋中正电	阎锡山	如能来京出席中国国民党五中全会对党国折冲协调
阎锡山电	蒋中正	承示中国国民党北京市党部防共办法等已遵办
蒋中正电	阎锡山	国民政府决定改组请阎兼任内政部长
阎锡山电	蒋中正	诸事待理暂难成行内长可由赵丕廉或赵戴文改充
何成浚电	蒋中正言	阎锡山决于政府成立前入京
阎锡山电	蒋中正	已嘱赵戴文南下听候任命
蒋中正电	阎锡山	内政部长亟待决定请阎自兼任
蒋中正电	阎锡山	言赵戴文南来毋任欢迎决请兄长内政
阎锡山电	蒋中正	请仍以薛笃弼继任内政部长

蒋中正档案人名与CADAL元数据关联统计。

人名	频率	频率_CADAL	条目_CADAL
蒋中正	297913	357	492
何应钦	18817	13	12
宋美龄	13158	9	8
陈诚	12349	25	22
顾祝同	12100		
张学良	7458	5	4
蒋经国	6667	9	9
阎锡山	6529	30	22
刘峙	5836	7	6
汪兆铭	5474	150	120
何成浚	5290		
胡宗南	4973		
宋子文	4836	7	5
孔祥熙	4587	13	11
卫立煌	4479		
李宗仁	4479	36	28
朱培德	4464	1	1

元数据人名关联二： 《中国国民党职名录》

以刘维开先生所编
《中国国民党职名录》
全部文本为依据，建
立人名索引。

姓名	L4	L5	L6	页码
陈立夫	第三届中央执行委员会	秘书处	秘书长	66
陈立夫	第三届中央执行委员会	组织部	部长	67
陈立夫	第三届中央执行委员会	中央政治会议	候补委员	70
陈立夫	第三届中央执行委员会	三届五中全会之后	委员	71
陈立夫	第四届中央执行委员会	财务委员会	委员	79
陈立夫	第四届中央执行委员会	革命勋绩审查委员会	委员	80
陈立夫	第四届中央执行委员会	中央政治会议	委员	81
陈立夫	第四届中央执行委员会	组织委员会	委员	77

《中国国民党职名录》 与CADAL元数据关联 统计。

通过元数据人物关系
分析，可以丰富蒋介
石研究资料库若干文
献资源。

姓名	频率_职名	频率_蒋档	条目_CADAL	频率_CADAL
蒋中正	74	297913	357	492
叶楚伦	67	625	125	125
陈果夫	62	1135	24	24
孙科	62	967	15	15
居正	54	151	62	62
陈立夫	54	1259	66	66
丁惟汾	53	80		
胡汉民	52	755	70	70
于右任	50	338	21	21
甘乃光	46	82	34	34
张继	46	165	14	14
林森	43	857	13	13
谷正纲	41	222		
戴传贤	41	477	4	4
朱家骅	39	935	9	9
吴铁城	39	2387	7	7
邹鲁	36	115	37	37
何应钦	36	18817	12	12

谢谢！