

当代中国社会生活资料 数据化、共建与共享

复旦大学图书馆

张计龙



復旦大學圖書館
FUDAN UNIVERSITY LIBRARY

提纲

01

项目缘起

02

资料数字化与数据库建设

03

资料共建共享机制的探索

04

进行中的工作

项目缘起



张乐天：复旦大学社会发展与公共政策学院教授，博士生导师，当代中国社会生活资料中心主任，复旦大学-UCLA 社会生活比较研究中心中方主任，全球中国比较研究院中方院长。

—— **“社会生活资料”**：由个人、家庭、企事业单位、政府基层单位所制作，但没有纳入正式档案系统而流入社会的文字、数据等资料。

项目缘起

复旦大学当代中国社会生活资料中心收藏



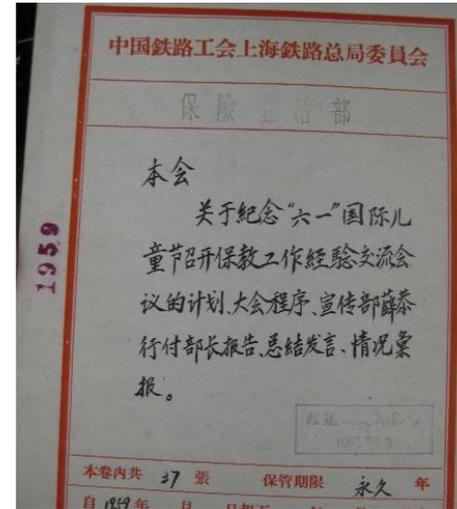
基层政府文书
约20万件



民间书信
超过50万封



日记和笔记
5000多册



企事业单位文书
约80万件

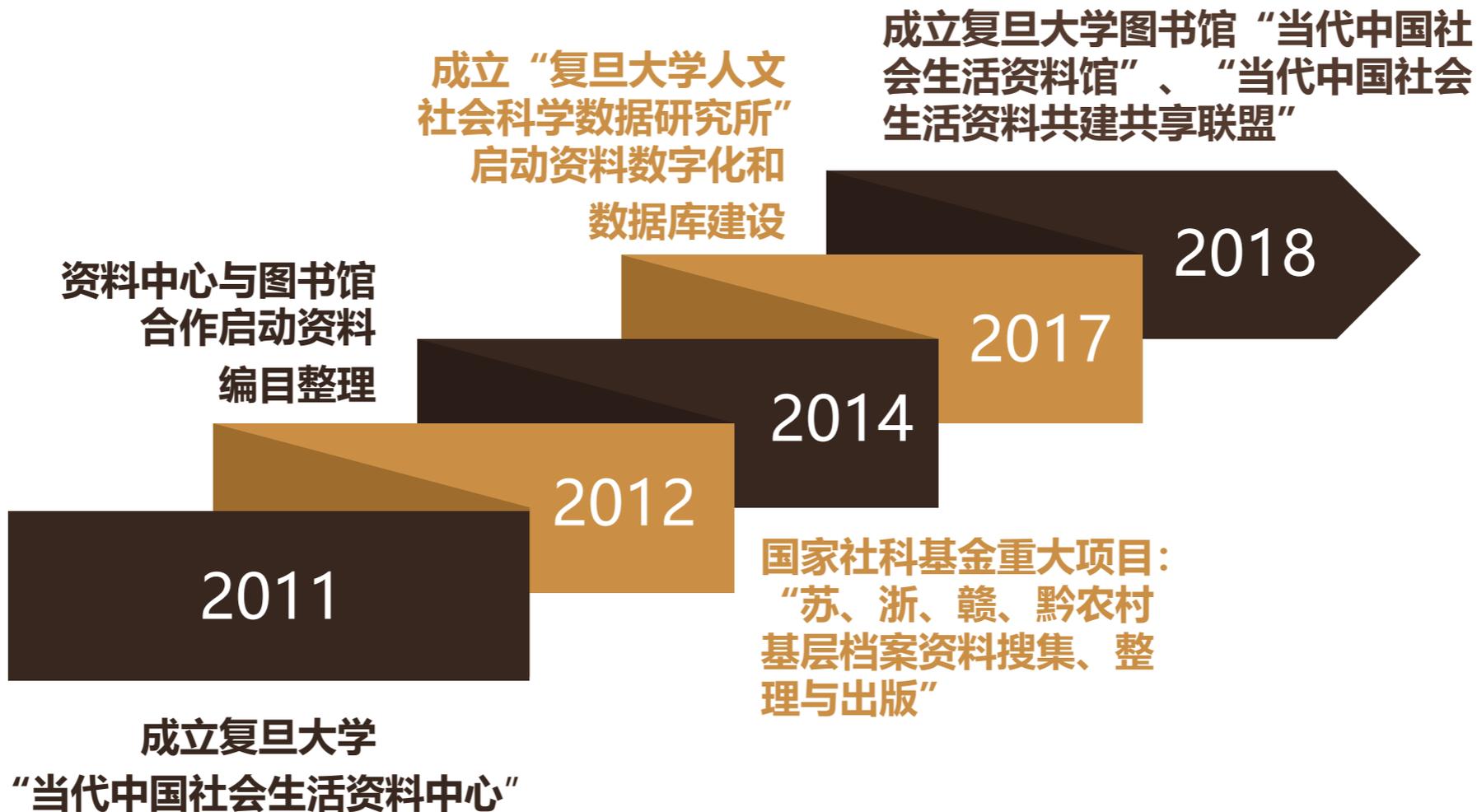
项目缘起



复旦大学当代中国社会生活资料的特点：

- **时间跨度**：1949年以后，以20世纪**50-80年代**为主；
- **地域分布**：集中于长江流域的上海、江西、江苏、安徽、四川、湖北等省市，部分资料涉及华北、东北地区；
- **主要内容**：家庭与个人生活史资料，企事业单位档案和基层政府资料；
- **资料总量**：约**150多万件**，其中90%以上是手写资料。

项目缘起



资料数字化与数据库建设

● 数据库建设概况 (2017.11—2019.05)

经费投入

50万页资料数字化
元数据标引
数据库系统软硬件采购和开发

人员投入

资料中心研究团队
图书馆技术、数字化、编目团队
软件开发厂商

成果

数据：50万页资料，部分资料完成标引

规范：书信、档案元数据规范等

平台：数据管理、发布、共享、数字版权保护

建设计划 (2019)

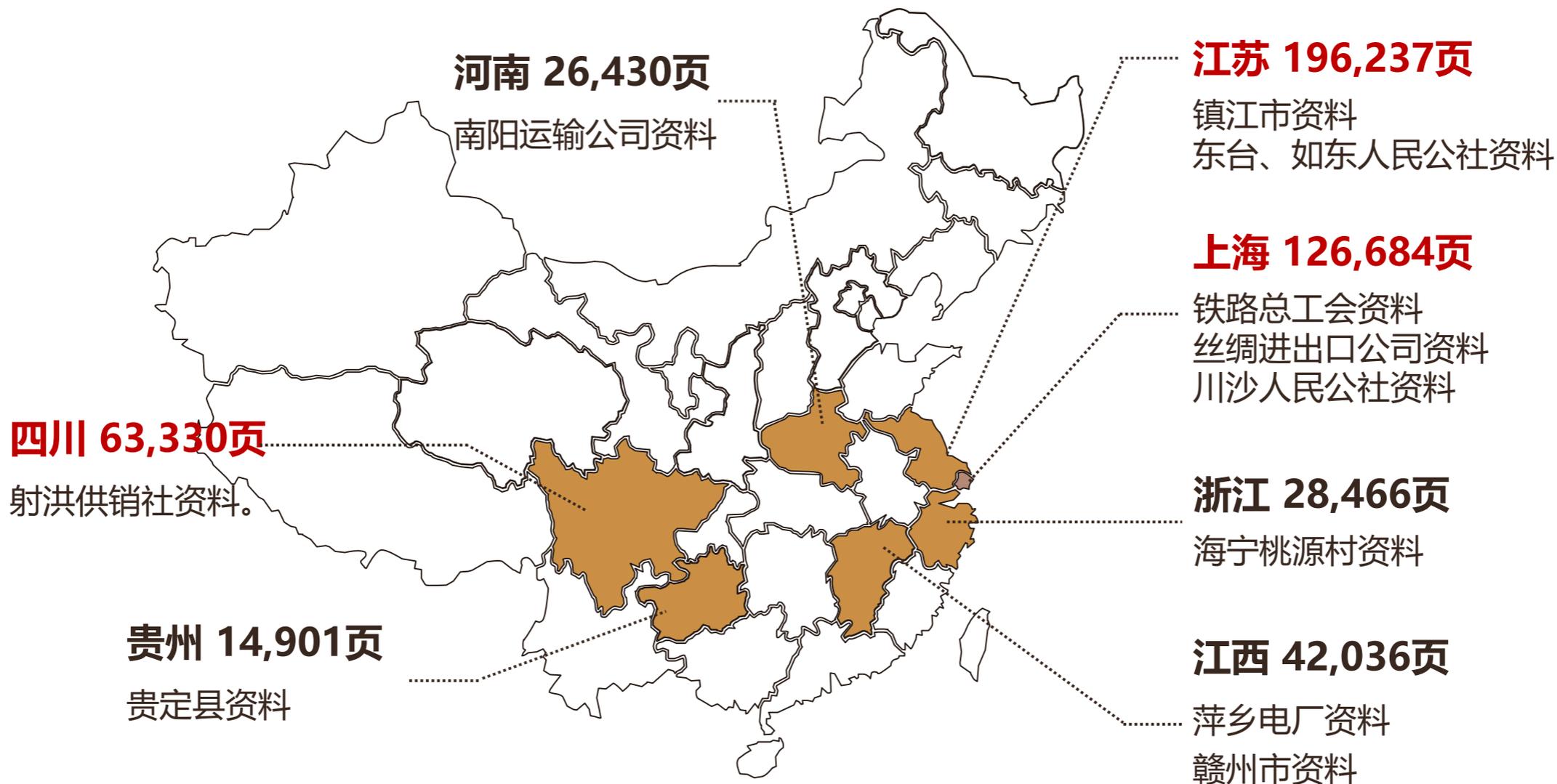
2019年5月上线，面向研究人员提供**部分资料和书信**的共享和利用

建立数据分级管理规范

依据数据安全级别，提供相应的数据共享服务

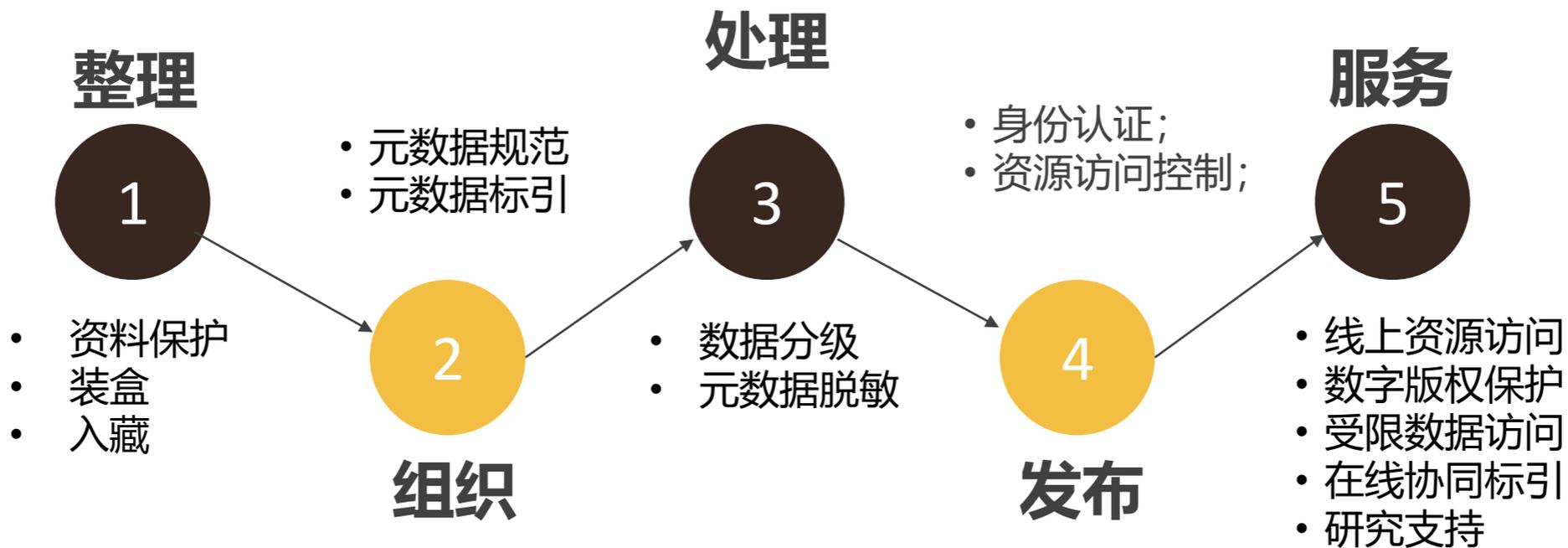
资料数字化与数据库建设

● 50万页数字化资料的地域和数量分布



资料数字化与数据库建设

● 数据库建设流程



资料数字化与数据库建设

● 资料的组织方式

1. 根据资料类型进行第一级的系列划分

DATA RESOURCES / 数据资源



文书资料

主要包括：各种团体机构的上级文件、同级文件及各类材料



数据资料

主要包括：各种团体机构的帐目、凭证、票据、合约、清单、报表等



书信资料

主要包括：个人的往来书信资料



日记笔记

主要包括：个人的日记、工作笔记、检讨书、报告等。

资料数字化与数据库建设

● 资料的组织方式

2. 根据“全宗原则”
进行第二级的资料组
划分



资料组列表

共 4 条结果

1. 中国铁路总工会上海区委员会资料

资料数量: 30000页以上

使用限制: 开放

访问方式: 在线

简介: 数量为500多份, 超过3万页。时段为1949-1991年。内容包括: 统计报表、工会工作文件、铁路工会明细分类账、工会工作收入日帐、涉及工会工作各个方面、宣传工作、工会筹备工作、集体合同、劳动保险、职工生活调查、工
[展开全部](#)▼

2. 四川射洪县供销社资料

资料数量: 50000页以上

使用限制: 开放

访问方式: 在线

简介: 系统性较强,涉及供销社所有工作, 数量为400多份, 超过5万页。时段为1951-1994年。内容包括: 青岗中心联社相关工作档案、青岗中心联社相关工作档案、金华供销社相关工作档案、柳树供销社相关工作档案、射洪县飞跃区供
[展开全部](#)▼

3. 河南省南阳汽车厂（河南省南阳市运输公司）资料

资料数字化与数据库建设

● 资料的组织方式

3. 根据资料主体进行第三级的资料盒划分

1. 上海铁路工会保险生活部

主要责任者：上海铁路工会保险生活部

数量：10卷

子单元题名：1.本会关于副业生产工作的情况报告、通知及江苏省工联批转浦口站通报。1957.9.17-1957.12.21;2.本会关于互储金、生活福利工作的训练班讲课教材。1957年;3.本部劳动保险金收支月报表。1957年;4.全委、关于生活、[展开全部](#)▼

2. 上海铁路工会生活部

主要责任者：上海铁路工会生活部

数量：3卷

子单元题名：1.南京、杭州、上海办事处、直属、工程处、医学院工会关于保险、生活工作的简报、汇报、报告、总结、通知、请示、经验介绍1961.4.25-1961.12.22;2.本会关于生活工作的计划、总结、报告、电报、汇报、指示、通[展开全部](#)▼

3. 上海铁路工会财务部

主要责任者：上海铁路工会财务部

数量：4卷

子单元题名：1.一九五四年推销公债铁路分会的总结。1954.3.2-1954.3.9;2.本会1954年福利金现金日记账1954.1.4-1954.12.31;3.本会1954年文教、工作费现金帐（八至十二月）1954.8-1954.12;4.上海铁路管理局基本建设[展开全部](#)▼

资料数字化与数据库建设

● 资料的组织方式

4. 每一个资料盒内包含若干资料

1. 各基层工会经费清查报告表及汇总

资料收录日期：19620331 至 19620627

地区：上海

2. 杭州办事处工委对所属各基层工会经费清查报告表及汇总

资料收录日期：19620331 至 19620704

地区：杭州

3. 全委通报西南工程区工会“关于精简工作中有关工会工作问题的意见”批复

资料收录日期：196204010 至 19621102

地区：上海

4. 上海市总、劳动局关于精简职工工作中做好职工退休工作的几点意见

资料收录日期：196204010 至 19621102

地区：上海

5. 上海市总工会办公室关于上海申新九厂精简职工工作经验的学习通知

资料收录日期：196204010 至 19621102

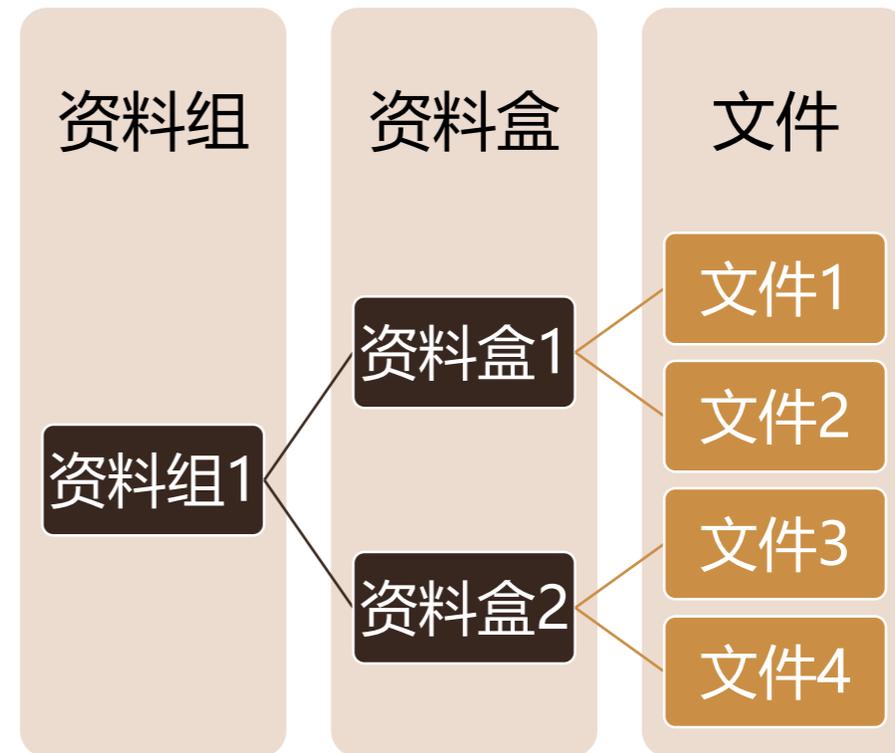
资料数字化与数据库建设

● 元数据规范

1. 依据EAD(机读档案著录标准, Encoded Archival Description),对档案资料的元数据描述分为资料组, 资料盒和文件三个层级, 对每一个层级都设计了相应的元数据规范

2. 元数据规范设计依据的相关标准

《中国文献编目规则 (第二版) 》
《新版中国机读目录格式使用手册》
《MARC21书目数据格式使用手册》
《资源描述与检索 (RDA) 》
《CADAL项目标准规范汇编 (三) 》



当代中国社会生活资料-资料组元素表		
元素	元素修饰词	必备性
题名/主要责任者 (Title/Creator)		必备
日期 (Date)	资料组起始日期 (Begin)	必备
	资料组截止日期 (End)	必备
其他责任者 (Contributor)		必备
所属地理区域 (Address)		必备
内容摘要 (Abstract)		必备
关键词 (Keywords)		必备
载体形态 (Manifestation)		必备
权限管理		必备

当代中国社会生活资料-资料盒元素表		
元素	元素修饰词	必备性
题名 (Title)	正题名 (Title proper)	必备
	副题名 (Other Title)	有则必备
主要责任者 (Creator)		有则必备
其他责任者 (Contributor)		有则必备
日期 (Date)	资料盒起始日期 (Begin)	有则必备
	资料盒截止日期 (End)	有则必备
格式化内容附注 (Formatted Contents Note)	子单元题名	必备
载体形态 (Manifestation)	资料盒规格 (Size)	必备
	文件数量 (Volume)	必备
	文件材质 (Material)	必备
收藏信息 (Collection)	典藏主标识 (Primary Identifier)	必备
	收藏单位 (Repositories)	必备
附注 (Notes)	相关信息补充标注	有则必备

当代社会生活资料-文件核心元素

当代社会生活资料-文件核心元素		语言 (Language)	必备				
题名 (Title)	元素	元素修饰	资源类型 (Type)	必备	书信资料个性元素		
	正题名 (Title Proper)	并列题名 (Parallel Title)	附注 (Notes)	印章附注 (Seal Note)			
	其他题名 (Other Title)	交替题名 (Alternate Title)	收藏信息 (Collection)	一般性附注 (General Note)	元素	元素修饰词	必备性
	主要责任者 (Creator)	次要责任者 (Contributor)	来源标识 (Source)	收藏单位 (Repositories)	信路 (Route)	寄信地 (From)	必备
关键词 (Keywords)	载体形态 (Physical Description)	权限管理 (Rights)	典藏主标识 (Primary Identifier)	析出寄信地 (Analyzed From)		可选	
材质 (Material)	数量 (Quantity)	受限信息标识 (Limited)	征集方式 (Ways)	寄信地邮政日戳 (Send Post Stamp)		可选	
计量单位 (Unit)	规格 (Size)	数据分级标识 (Level)	征集者 (Collected)	收信地 (To)		必备	
破损状况 (Condition)	破损程度 (Damage)	数据集全球唯一ID (Control ID)	征集日期 (Collection Date)	析出收信地 (Analyzed To)		可选	
日期类型 (Type of Date)	资料创建日期	数字化信息标识 (Digitalization)	征集地点 (Collection Site)	收信地邮政日戳 (Receive Post Stamp)		可选	
			扫描分辨率 (Scan Resolution)	可选			
			图片数量	必备			

资料数字化与数据库建设

● 元数据规范

- 采用**名称元素形式对资料的属性进行描述**，尽量复用MARC字段已有元素内容，便于今后与馆藏其他资源进行交互；
- 依据**FRBR实体-关系模型**，从内容、实体、数据和单册管理三个层次上进行描述，更加注重对资料内容的揭示；
- **核心元素与个性化元素结合**，为今后建立关联数据的核心本体和领域本体打下基础；
- 资料细节的描述，使用了**《张乐天关键词表》**，便于提炼特定时期的特定事件、人物类型，作为评价内容敏感程度的重要依据；
- 增加**人物、机构的属性信息**，作为数据分级的重要依据；

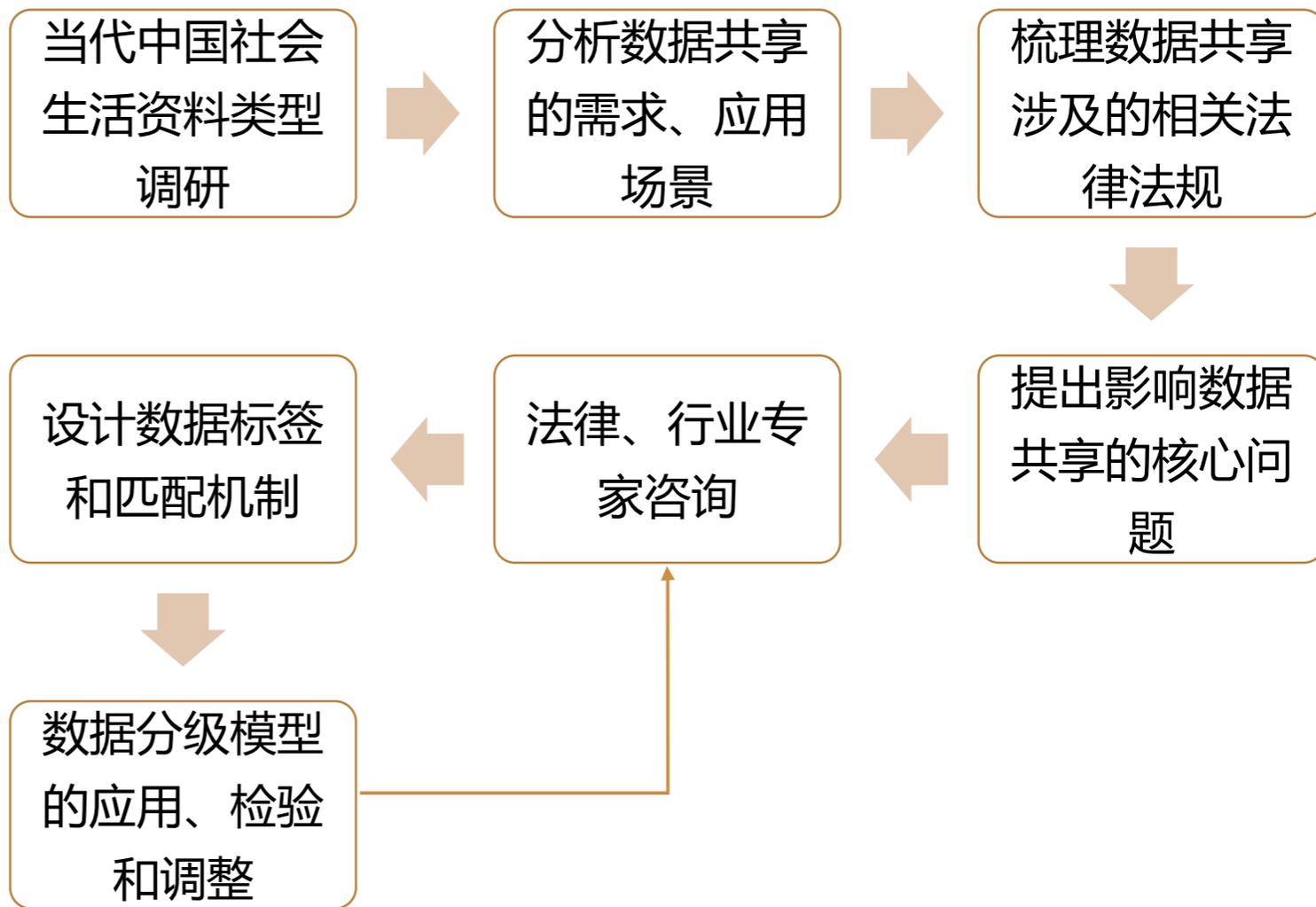
资料数字化与数据库建设

● 数据分级

引入哈佛大学
Datatags数据分级
模型，构建当代中国
社会生活资料数据分
级规范

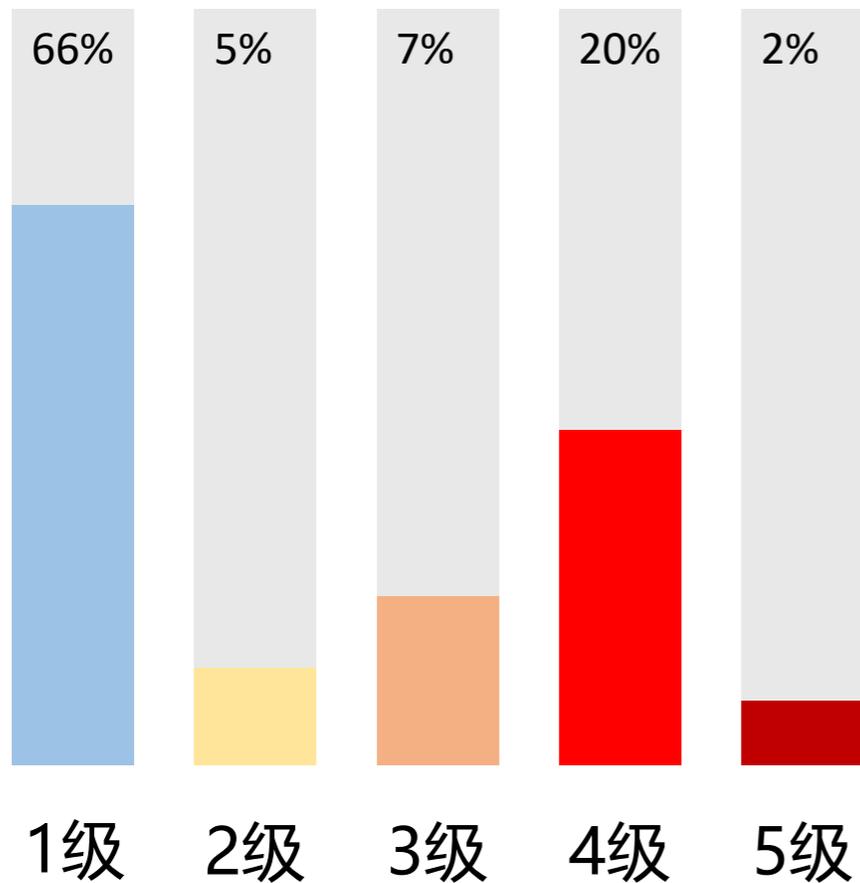
特点：

1. 结合法律、伦理和技术知识
2. 实现决策知识编码化
3. 以机器自动处理为目标



资料数字化与数据库建设

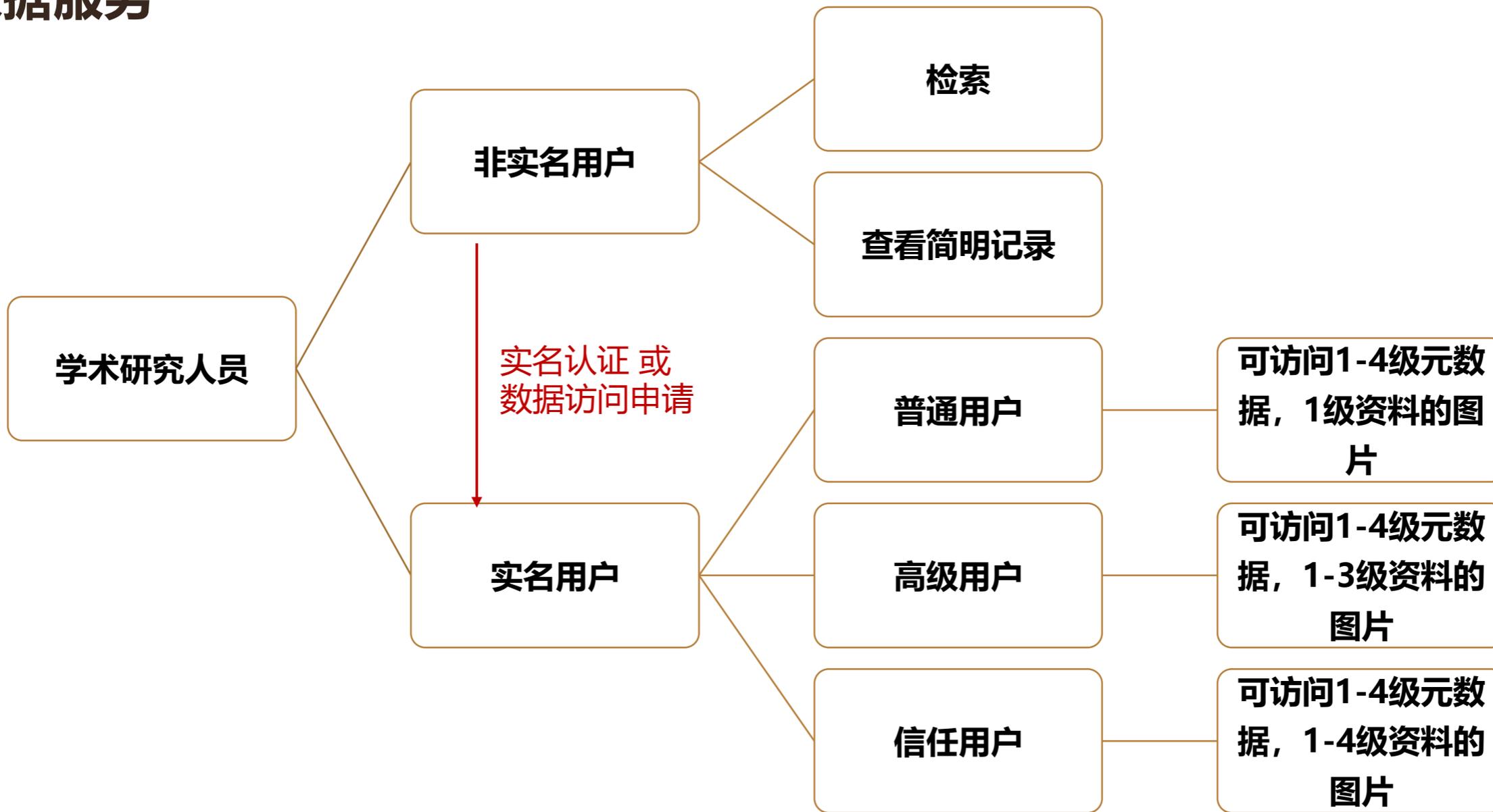
● 数据分级



数据标签	数据描述	访问凭证
1级	开放所有元数据 开放资料图片	实名认证用户
2级	开放所有元数据 部分开放资料图片	实名认证用户
3级	开放所有元数据 不开放资料图片	实名认证用户 需提交数据申请
4级	开放脱敏元数据 不开放资料图片	实名认证用户 需提交数据申请
5级	不开放元数据 不开放资料图片	线下使用

资料数字化与数据库建设

● 数据服务



资料数字化与数据库建设



JOIN US / 加入我们



约10万件

基层政府文书资料

约9万件

企事业单位文书资料

约1万件

群众组织文书资料

192本

会议记录

37万封

个人书信

4000本

日记与工作笔记

“社会生活资料中潜藏着另一个中国以及中国发展得以可能的道理，讲清楚这些道理是历史赋予我们的神圣使命。”

资料数字化与数据库建设

资料列表

已筛选条件：**农业** ×



射洪县大榆供销社第四届第三次社员代表大会开幕词

资料收录日期：196301 至 196312

地区：四川省射洪县

[查看详情](#)



射洪县大榆供销社第四届第三次社员代表大会议事规则

资料收录日期：196301 至 196312

地区：四川省射洪县

[查看详情](#)



射洪县大榆供销社第四届第三次社员代表大会议程

资料收录日期：196301 至 196312

地区：四川省射洪县

[查看详情](#)

主要年份

- 1966年 183
- 1964年 135
- 1963年 104

中文标签

农业(18824)	合作化运动
人民公社(13679)	社会主义教育运动
工会(7335)	铁路(7326)
交通运输(7325)	东台人民公社
东台县人民公社	上海铁路工会
供销社(5144)	射洪县供销社
统购统销(5144)	供销合作社(5143)
社会主义教育运动	退休纪念章(148)
俱乐部(76)	劳动保险费(75)
小组数(68)	竞赛(63)

自由标签

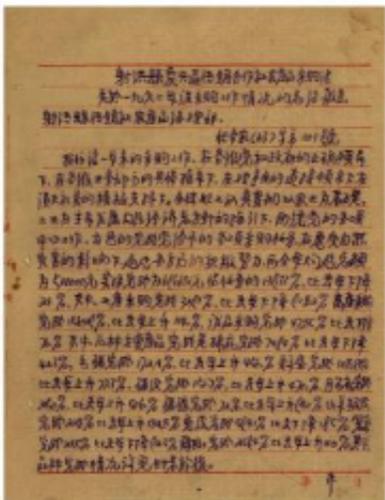
棉花(46)	生产资料(29)
帝国主义(28)	修正主义(26)
阶级斗争(24)	支援农业(21)
革命运动(18)	农业生产(17)
收购(17)	农具(15)
毛主席著作(14)	耕牛(14)
党领导(13)	农药(13)
政治挂帅(13)	采购(12)
增产节约(11)	棉花收购(11)
商业(10)	思想工作(10)

编目员根据内容提炼

来自张乐天《关键词汇总表》

资料数字化与数据库建设

资料详情



[射洪县复兴供销合作社]关于一九六二年度采购工作情况的总结报告

【资料收录日期】	1963-01-13
【中文关键词】	射洪县供销社; 农业; 统购统销; 供销社; 供销合作社
【自由关键词】	采购; 商业; 农产奖励; 党领导; 检查制度
【典藏主标识】	SA20170000194J
【主要责任者】	射洪县复兴区供销社
【载体形态】	纸
【数量】	7
【规格】	A4
【地区】	射洪; 复兴
【单位】	张
【数据集全球唯一ID】	20.500.12316/231010-chinalife-files-2369

在线阅读

Handle System唯一标识符

URL:

<http://hdl.handle.net/20.500.12316/231010-chinalife-files-2369>



缩略图



1



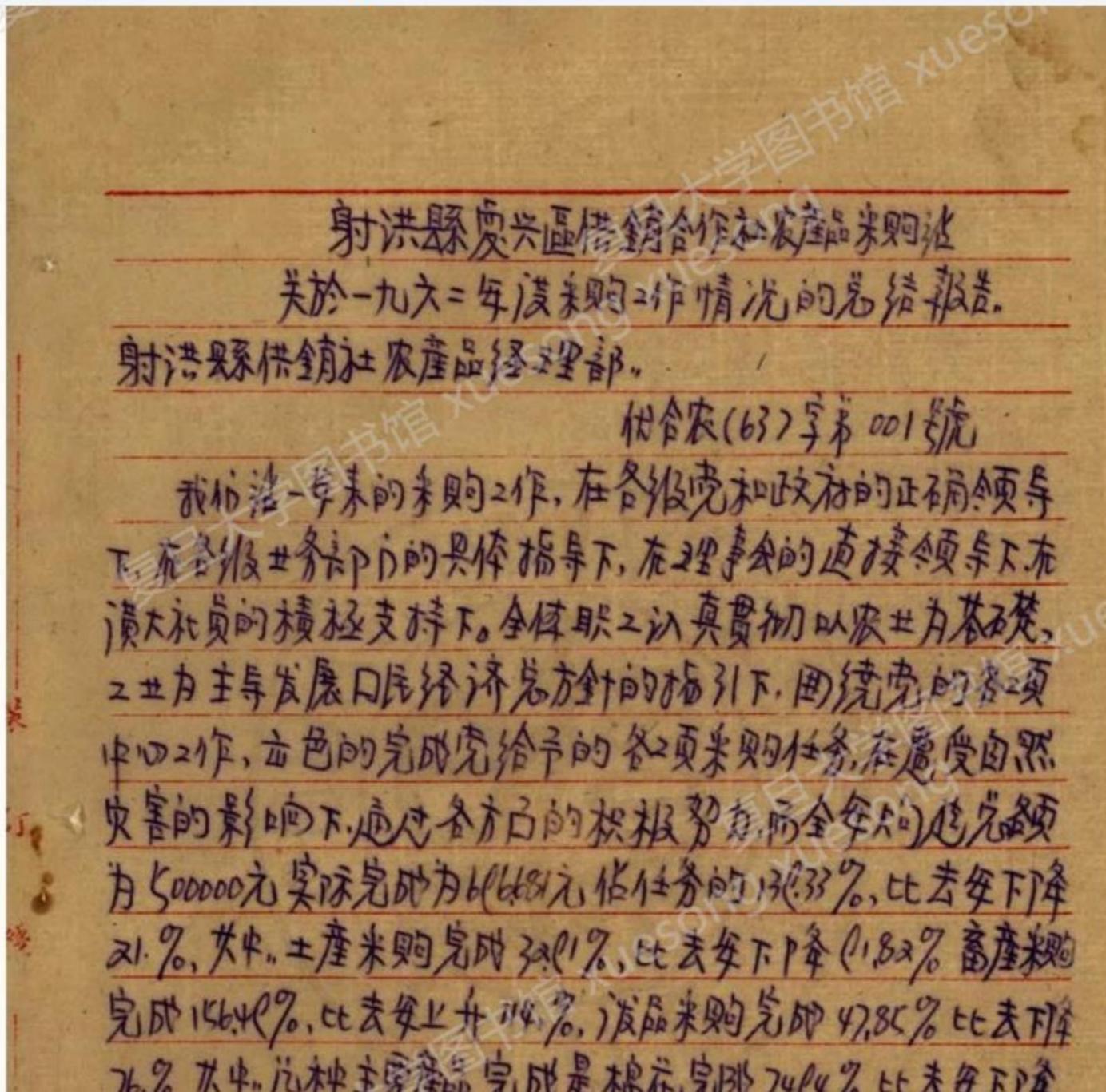
2



3



4



数字版权保护:

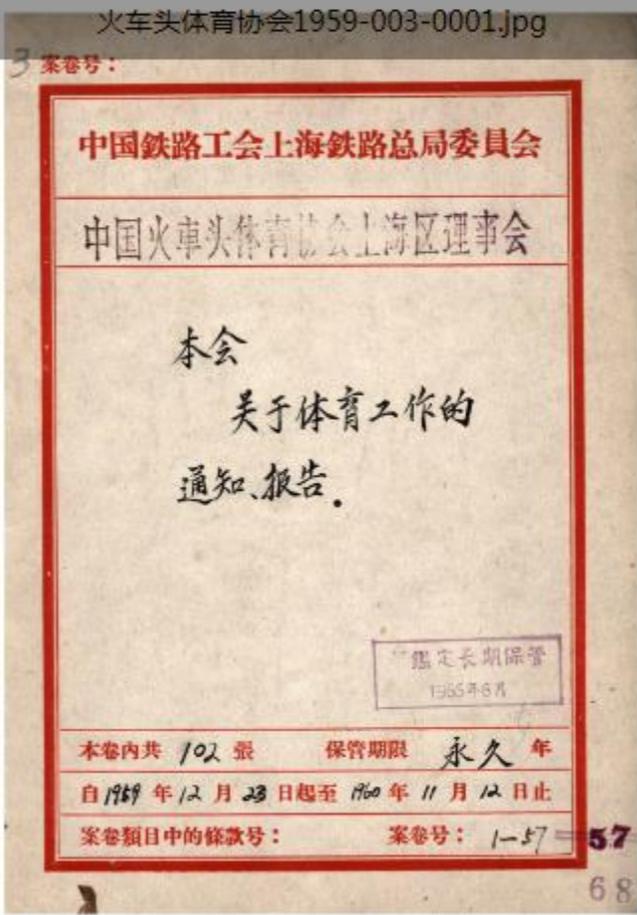
- 浏览器在线浏览，无需安装插件，图片不可保存，也无法从缓存中提取
- 显示低精度压缩图片
- 静态水印：复旦大学校徽
- 动态水印：用户所属机构名称和账号名称

当前位置： 档案->修改元数据

*正题名	<input type="text" value="体育工作报告"/>	添加
并列题名	<input type="text"/>	添加
副题名及说明题名文字	<input type="text"/>	添加
*主要责任者	<input type="text" value="上海铁路总局"/>	添加
次要责任者	<input type="text"/>	添加
主题	<input type="text"/>	添加
中文关键词	<input type="text" value="上海铁路工会; 工会; 铁路; 交通运输"/>	添加
自由关键词	<input type="text" value="体育工作; 广播体操; 全民皆兵; 民兵"/>	添加
英文关键词	<input type="text"/>	添加
附注	<input type="text"/>	添加
摘要	<input type="text"/>	添加
文件形成日期	<input type="text" value="1959"/>	添加
所属档案收录起始日期	<input type="text" value="1959-01-11"/>	添加
所属档案收录截止日期	<input type="text" value="1959-12-30"/>	添加

文件列表 + 添加文件

文件名	选中2条	操作	显示文件列表 ▾
-----	------	----	----------



在线协同标引：

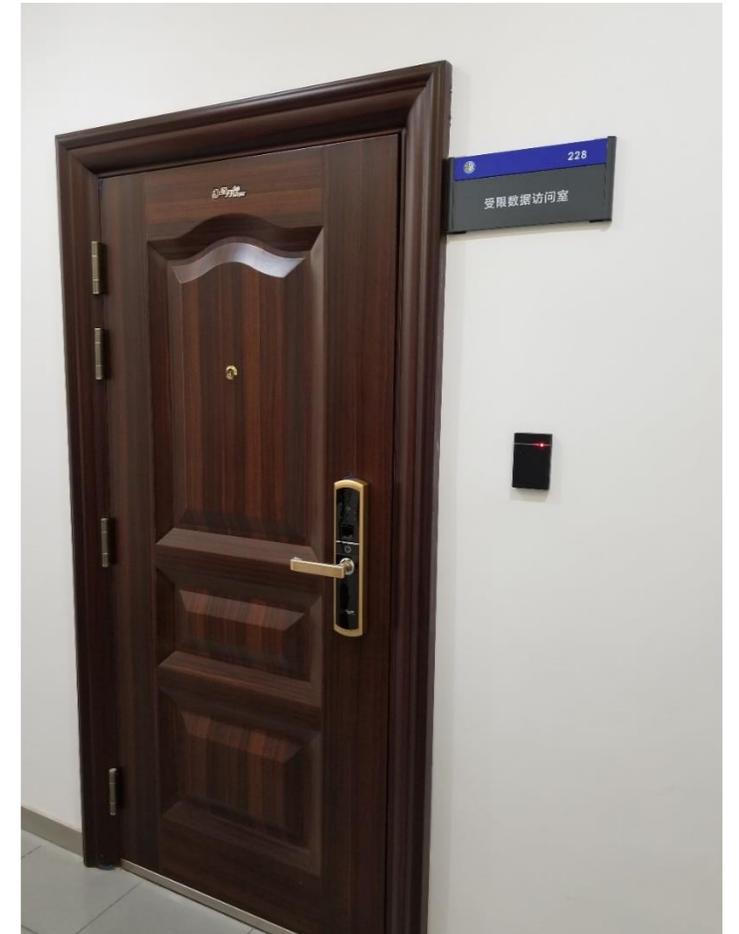
- 通过浏览器在线查看原始高精度图片，可放大缩小，便于识别手写文字；
- 完成元数据著录，并选取一份档案的多个图片文件后，系统自动将图片组合成一个PDF文件，并与元数据挂接；
- 支持任务分配、工作量统计等，可用于众筹标引。

资料数字化与数据库建设

● 如何访问数据分级为5级的数据？

受限数据访问室：

- 数据库服务器、保险箱、360度监视器、离线工作站；
- 数据沙箱系统实现数据不落地访问；
- 提供SPSS、SAS、Stata、ArcGis、图像处理等工具；



资料共建共享机制的探索

2018年10月16日，22家学术图书馆及研究机构共同发起成立“当代中国社会生活资料共建共享联盟”

(Coalition of Contemporary Chinese Society & Life Archives, CCCSLA)



资料共建共享机制的探索

联盟的宗旨

建立文献资源共建共享机制，促进此专题资源的收集、整理、保护与利用，并以服务全球学术研究为目标。

资源建设组组长单位

武汉大学图书馆

学术推广组组长单位

南京大学图书馆

平台建设组组长单位

浙江大学图书馆

标准规范组组长单位

复旦大学图书馆

资料共建共享机制的探索

依托联盟，探索成员之间的多种合作模式，加快资料整理、数字化和数据化建设



建立共享资料的元数据库

联盟成员提供在联盟内部可共享的纸质或数字化文献资料的元数据，并提供相应的访问方式。



联盟成员之间的项目合作

根据各自的收藏特色和资料利用需求，联盟成员之间通过项目合作方式开展资料整理、数字化及共享服务。（CADAL & FDULIB 2019合作项目）



专业整理与众筹标引的结合

探索建立激励机制，鼓励国内外对资料感兴趣的研究人员，在线上、线下阅读文献的同时，做一些标引和内容数据化工作，为数据库建设以及资料的内容揭示做贡献。

进行中的工作



申报上海市中国文化走出去重大项目

‘民心相通’——当代中国民间书信资料共享平台建设”项目计划以中心所收集、收藏的五十多万封民间书信为基础资料，由复旦大学图书馆牵头在目前初步建成的“当代中国社会生活资料数据平台”基础上，进行内容的升级和多语种配置，建设可面向海外的文化传播平台。



推进、落实联盟内部资源共建共享工作

- 面向联盟成员推出数据库试用；
- 收集成员单位的需求和试用反馈；
- 计划6月份召开工作组牵头单位第一次工作会议



数据库建设

- 开发与CADAL平台的数据接口
- 数字化资料的持续标引
- 完善数据分级规范
- 制定数据服务计划

下一步工作



推进、落实联盟内部资源共建共享工作

- 根据试用反馈意见，完善数据库系统功能，正式向联盟成员提供服务；
- 收集各成员单位资料的收藏及数字化建设情况；
- 调研各成员单位的共建共享需求；
- 规划联盟资源建设的阶段性方向；
- 规划联盟数据服务的网络基础设施；
- 以联盟共享、项目合作等方式推动资源共建共享。



推动基于资料的合作研究

立足联盟资源，通过组织一系列学术推广活动，推动形成国内外合作的社会生活资料研究网络。

人民日報

1957.1.18-4.16

(左西樹荷一疊)

谢谢

Questions?