

众智、众创、众享：开放数据竞赛作为公共文化服务的创新引擎

Digital Humanities | Value Co-Creation | Open Data Contest

赵宇翔

南京大学信息管理学院

Email: yxzhao@nju.edu.cn

AI重塑智慧图书馆未来服务新生态主题研讨会

暨上海图书馆开放数据竞赛十周年交流会

上海图书馆东馆 2025年11月27日



- 01. 背景
- 02. 众智：开放数据建设
- 03. 众创：开放数据实践中的价值共创
- 04. 众享：开放数据的社会共享与再利用

01 背景

数字化转型 & 价值共创

*Value
co-creation*

政策文件



中华人民共和国中央人民政府

www.gov.cn



简 | 繁 | EN | 注册 | 登录



国务院

总理

新闻

政策

互动

服务

数据

国情

国家政务服务

首页 > 新闻 > 政务联播 > 部门

科技部等六部门印发《关于促进文化和科技深度融合的指导意见》的通知

关于印发《全国一体化大数据中心协同创新体系算力枢纽实施方案》的通知

发改高技〔2021〕709号

中共中央办公厅 国务院办公厅印发《关于推进实施国家文化数字化战略的意见》



- ✓ 文化与科技深度融合
- ✓ 文化大数据协同创新
- ✓ 线上线下融合互动、立体覆盖的文化服务供给体系
- ✓ 文化大数据的价值共创

公共文化服务转型



- 公共文化服务机构正在经历从实体空间向**数字社区**的迁徙和转型
- 数字化的文化遗产能够跨越时空的桎梏，将文化资产关联重组形成庞杂的**数据海**
- 要求各类机构整合资源并进行有效联动，从数字化内容、数字化工具、数字化人才、数字化流程、数字化机构等不同层面开展工作

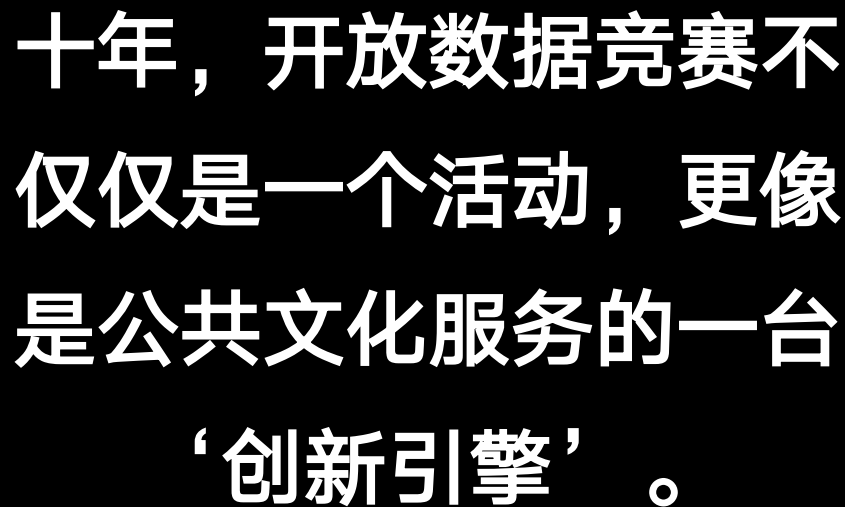
开放数据竞赛

2016年开始，上海图书馆以竞赛的形式联通社会各界力量，通过跨学科、跨领域合作共同释放开放数据的当代价值。



- 从最初开放家谱数据的探索起步，到逐步覆盖古籍、人物、建筑、红色文献、民俗、城市记忆等多主题资源。
- 已链接 28 家数据合作机构，开放数据规模累计超过 1.5 亿条，为公共文化领域的开放数据实践积累了坚实基础。

十年十届，竞赛以挖掘人文数据价值为起点。从家谱与人物谱系出发，逐步拓展至历史建筑、红色文化与地域民俗等多元领域；并始终紧扣数字人文与技术演进，探索多元赛道，持续深化新技术与人文故事的融合实践。



赛事



参赛者

观众



评委



研究者



合作者



我们团队在上图开放数据竞赛生态中的多重角色

02



众智：开放数据建设

众包 / 集体智慧 / 平台优化

*Value
co-creation*

研究背景



众包

文化遗产众包（Cultural Heritage Crowdsourcing, CHC）是指公众群体（志愿者）与文化遗产相关的公共文化服务机构及非营利性组织建立合作关系(通常由组织发起并主导，公众参与及配合)，协作完成工作量大且难以自动实现的任务，以践行文化遗产资源保护、传承和有效利用的共同目标，并达成供需两侧价值共创的一种开放协作模式。（赵宇翔&练靖雯，2021）

众包成为文化遗产机构开展开放数据建设的重要方式

作为合作者——举办众包竞赛

- 活动：文化遗产数字化竞赛
- 活动主题：

“对话盛宣怀，领略历史之美；古籍数字化，传承文化遗产”



作为合作者——举办众包竞赛

活动 | 文化遗产数字化竞赛来啦!

南大信管研会 2018-10-19 16:56

中华优秀传统文化具有五千年的悠久历史，那些古老的文明至今依然鲜活，并成为了永恒的经典。对中华优秀传统文化的挖掘、研究与传承，已经成为摆在我们面前的重大课题。尤其随着“云、物、移、大、智”与“互联网+”时代的到来，资源的数字化和有效整合无疑是传承优秀传统文化、实现文化创新的必要前提。

为响应“十九大”号召，将技术融入到文化遗产的研究中，激发高校学生的文化创新活力，感受传统文化的博大精深，提高新一代大学生的文化自信，南京大学信息管理学院现举办文化遗产数字化竞赛活动!!!

活动主题

对话盛宣怀，领略历史之美
古籍数字化，传承文化遗产

活动内容

本活动以上海图书馆历史文献众包系统 (<http://zb.library.sh.cn/>) 为平台，依托上海图书馆历史文献众包中心开展盛宣怀档案抄录项目——面向全校范围招募感兴趣的学生进行古籍手稿的线上抄录。



共有21个小组报名参加比赛，包括本科生、硕士和博士生。来自南京大学、复旦大学、南开大学、东南大学、四川大学、兰州大学、南京航空航天大学 and 南京师范大学等各学科和高校的学生参加了本次活动。

用户参与



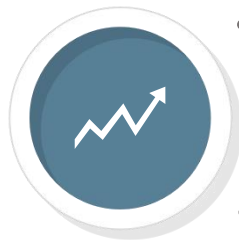
冷启动阶段

用户的广泛参与



持续发展阶段

用户的持续参与



用户参与

冷启动阶段

用户的广泛参与



研究方法：问卷调研、深度访谈、焦点小组

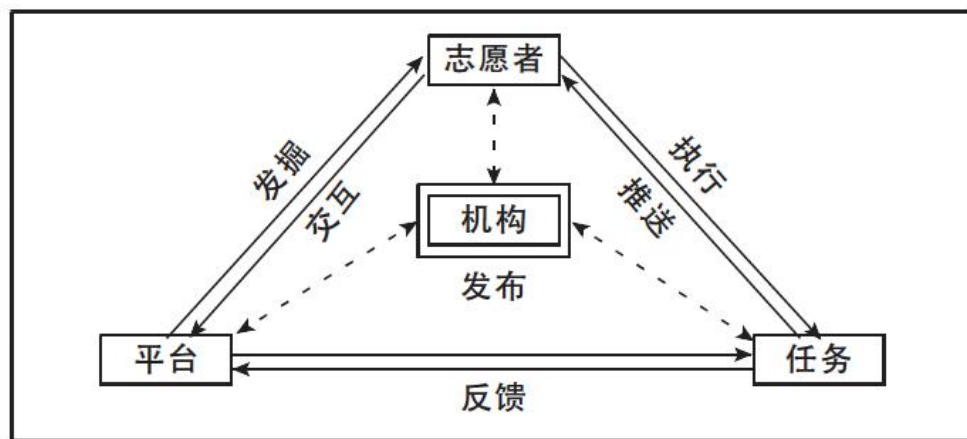


图1 数字人文类公众科学项目中影响志愿者参与的因素

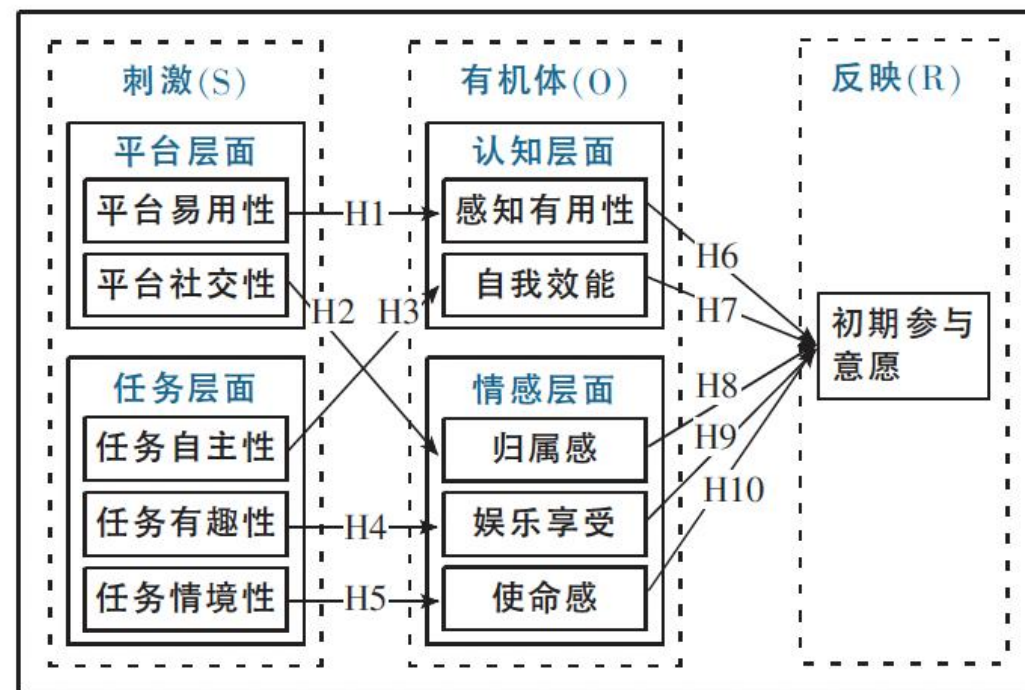


图2 数字人文类公众科学项目冷启动阶段的公众参与动因的实证模型

用户参与



持续发展阶段

用户的持续参与

研究方法：半结构化访谈、扎根理论、移动经验取样法（mobile experience sampling method）、行动研究（action research）

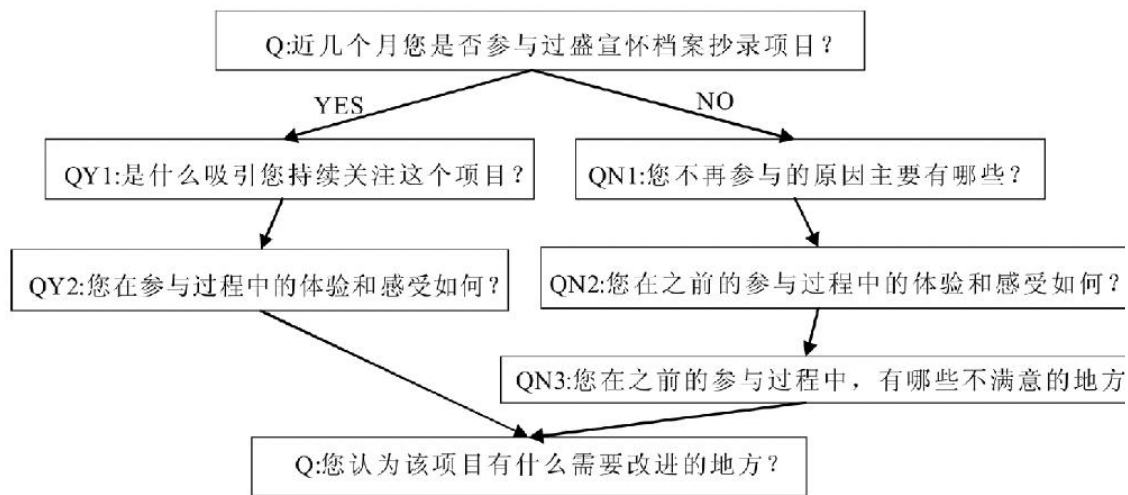


图2 访谈大纲设计

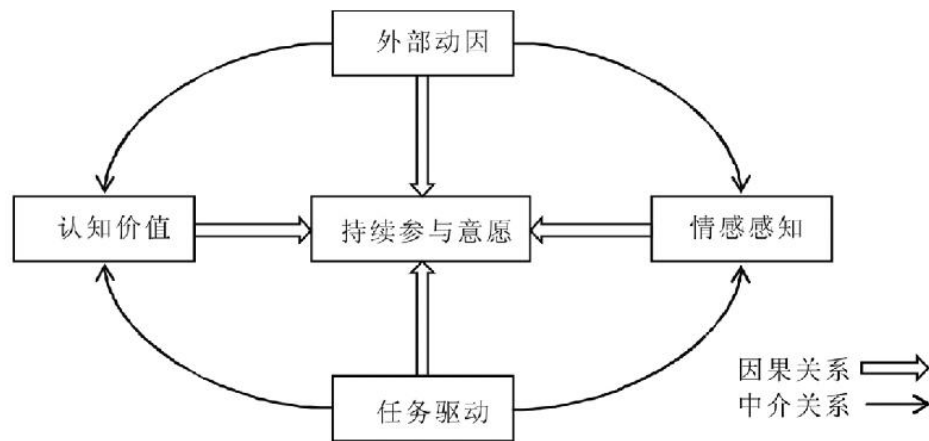


图3 数字人文类公众科学项目持续发展阶段的公众参与动因理论模型

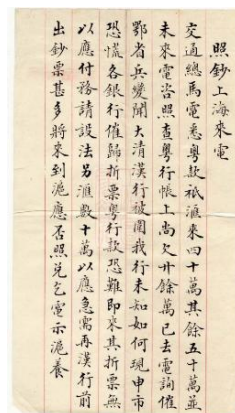
基于盛档的准实验研究

研究a

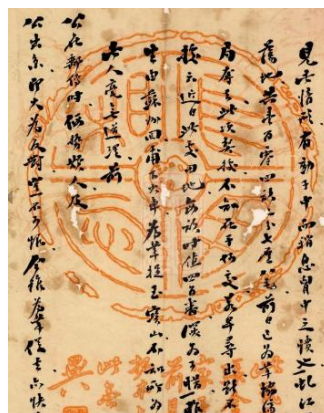
社会价值倾向和领域知识对手稿抄录型众包的作用

研究b

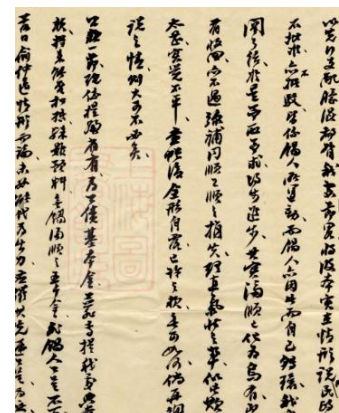
数字人文类众包抄录平台中任务绩效的影响因素研究
——基于任务复杂度与领域知识视角



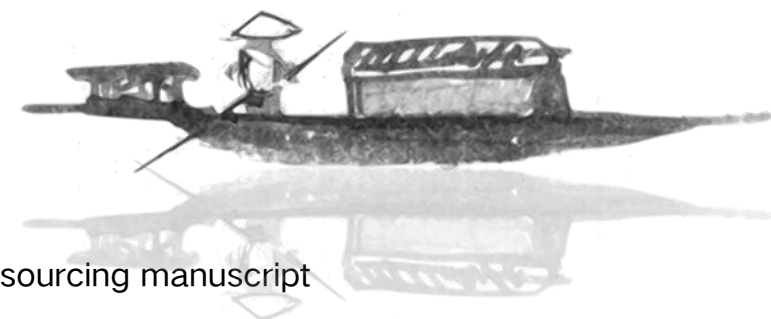
简单



中等

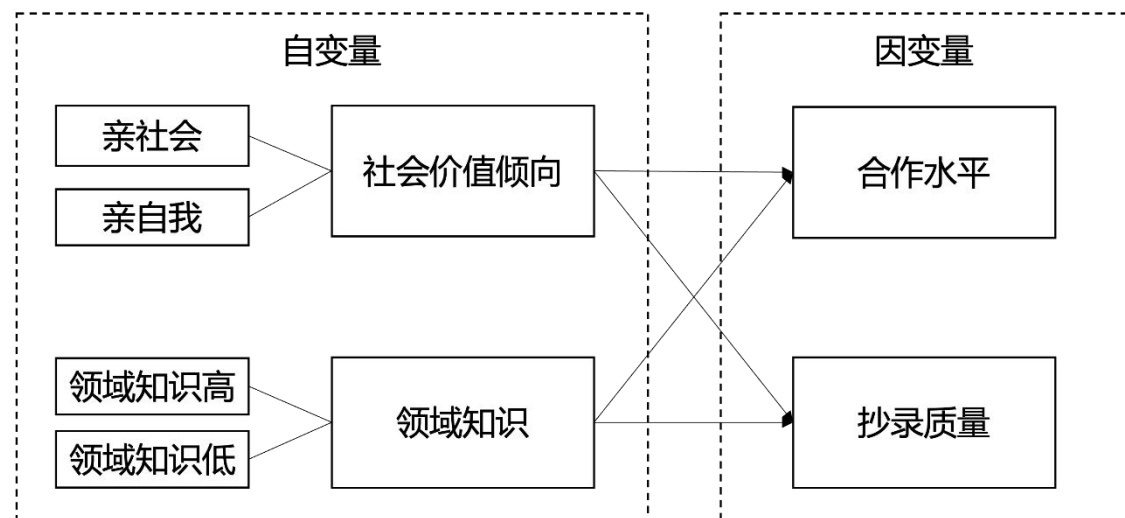


复杂



基于盛档的准实验研究a

- 研究问题：从个体差异性角度，什么样的志愿者在抄录型众包协作中的合作水平更高、完成质量更好？
- 研究意义：帮助组织机构在招募志愿者时，识别合格的志愿者，找到最合适的人来参加。



基于盛档的准实验研究a

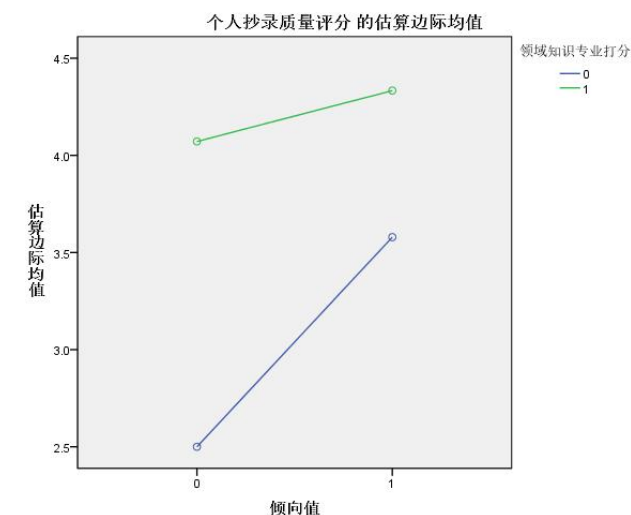
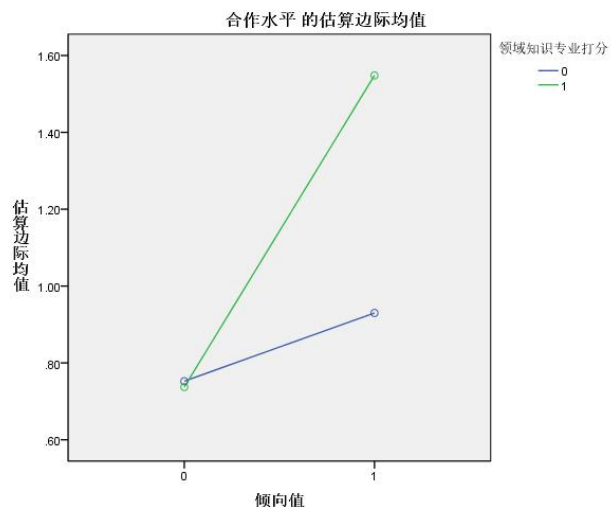
研究设计

- 准实验研究
- SVO: the Triple Dominance Measure (Van Lange et al., 1997)
- 领域知识(Domain Knowledge): 前测问卷
- 合作水平: $f(\text{Cooperation Level}) = \frac{\text{PersonalVolume}_{jk}}{\text{GrossVolume}_k} \times \text{TeamMembers}_k$
- 抄录质量: 专家打分

项目名称	任务题名	抄录人	导出时间	评分
【南大竞赛】盛宣怀档案抄录项目 (23)	SD019583	Chiaoe95	2018-12-26 14:09:02	★★★★★

项目名称	任务题名	抄录人	导出时间	评分
【南大竞赛】盛宣怀档案抄录项目 (16)	SD118680	l_a_n	2018-12-26 11:24:38	★★★☆☆

基于盛档的准实验研究a



主体间效应的检验

因变量: 合作水平

源	III 型平方和	df	均方	F	Sig.
校正模型	5.155 ^a	3	1.718	7.303	.000
截距	47.368	1	47.368	201.293	.000
倾向值	2.941	1	2.941	12.497	.001
领域知识专业打分	1.094	1	1.094	4.648	.036
倾向值 * 领域知识专业打分	1.208	1	1.208	5.134	.028
误差	11.531	49	.235		
总计	68.869	53			
校正的总计	16.686	52			

主体间效应的检验

因变量: 个人抄录质量评分

源	III 型平方和	df	均方	F	Sig.
校正模型	18.528 ^a	3	6.176	16.603	.000
截距	631.113	1	631.113	1696.650	.000
倾向值	5.409	1	5.409	14.541	.000
领域知识专业打分	16.274	1	16.274	43.750	.000
倾向值 * 领域知识专业打分	2.008	1	2.008	5.399	.024
误差	18.227	49	.372		
总计	769.000	53			
校正的总计	36.755	52			

研究结果:

- 社会价值倾向和领域知识分别对合作水平产生影响, 且存在交互影响作用。
(H1.H2.H3)

- 社会价值倾向和领域知识分别对抄录质量产生影响, 且存在交互影响作用。
(H4.H5.H6)

基于盛档的准实验研究

研究a

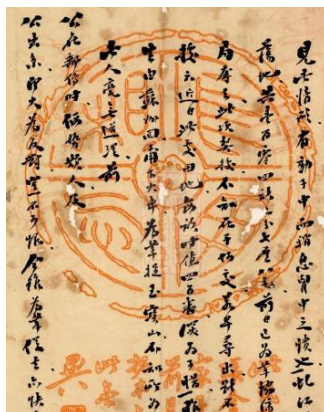
社会价值倾向和领域知识对手稿抄录型众包的作用

研究b

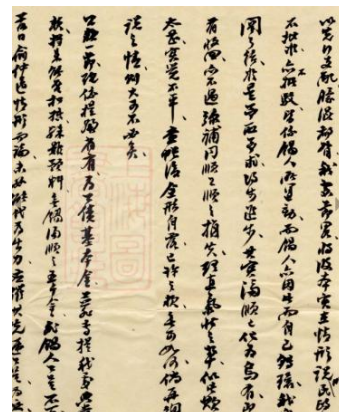
数字人文类众包抄录平台中任务绩效的影响因素研究
——基于任务复杂度与领域知识视角



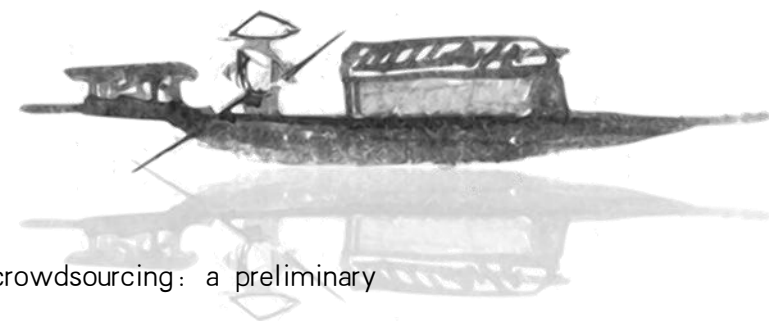
简单



中等



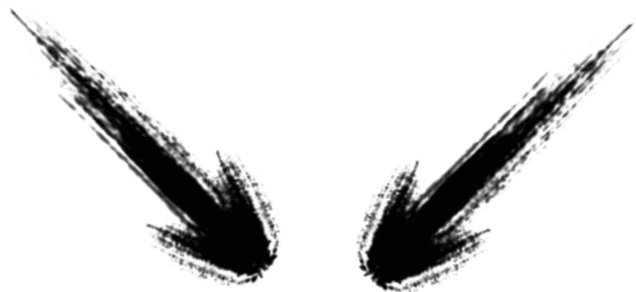
复杂



基于盛档的准实验研究b

任务设计

志愿者认知



考察志愿者进行抄录时的任务绩效

- 准实验研究
- 任务复杂度：20个低复杂度任务（档案页数在5页及以下、字体为正楷体、保存完好），20个高复杂度任务（档案页数在6-10页、字体为草书或行书、纸张有破损）
- 领域知识：前测问卷，20个高领域知识，20个低领域知识
- 任务绩效：专家打分

众包抄录平台优化

数字人文基础设施平台如何
实现技术与用户的匹配,
进而提升平台的利用率?



让人文数据活起来

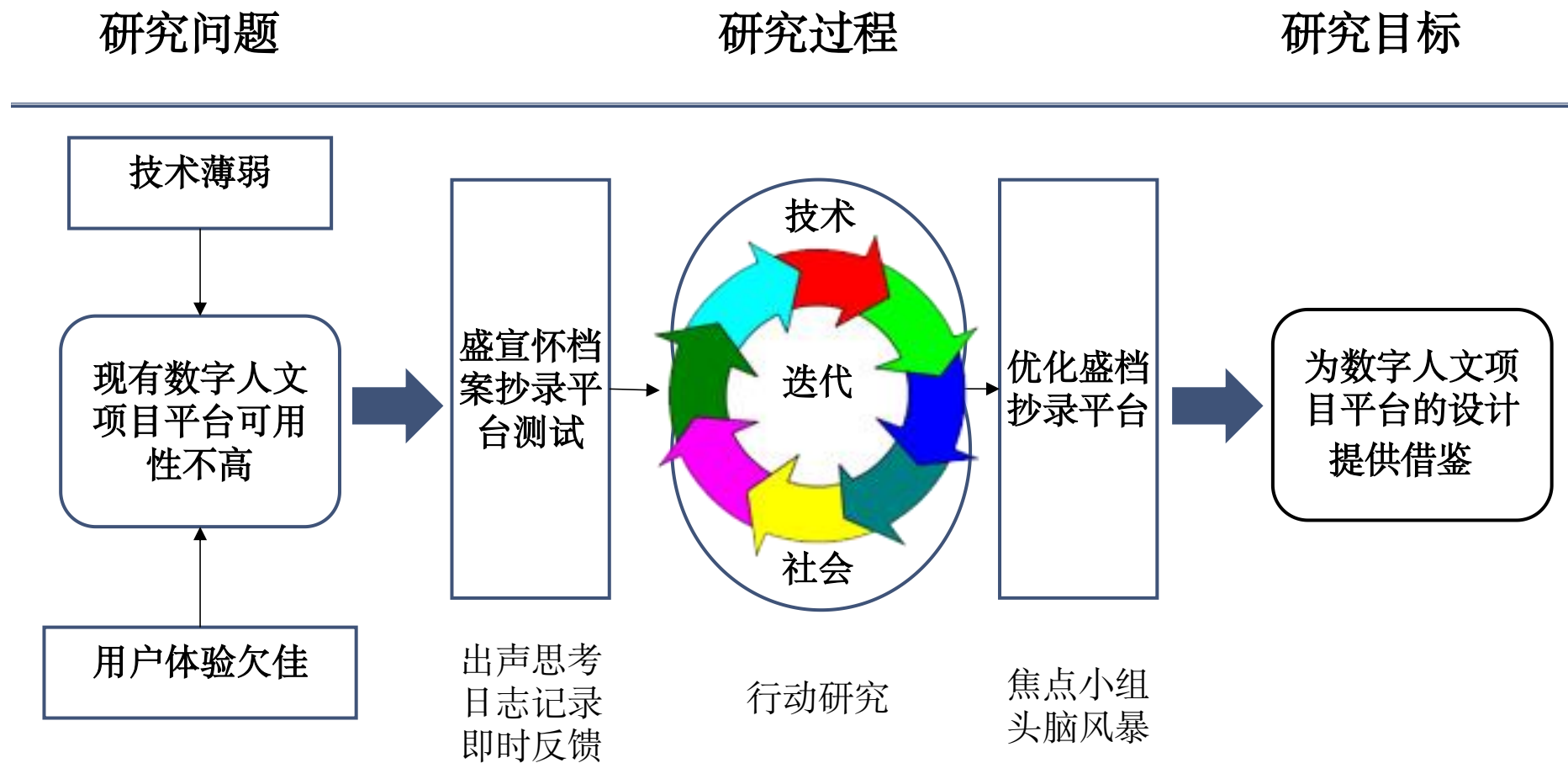
1

数字人
文项目
平台

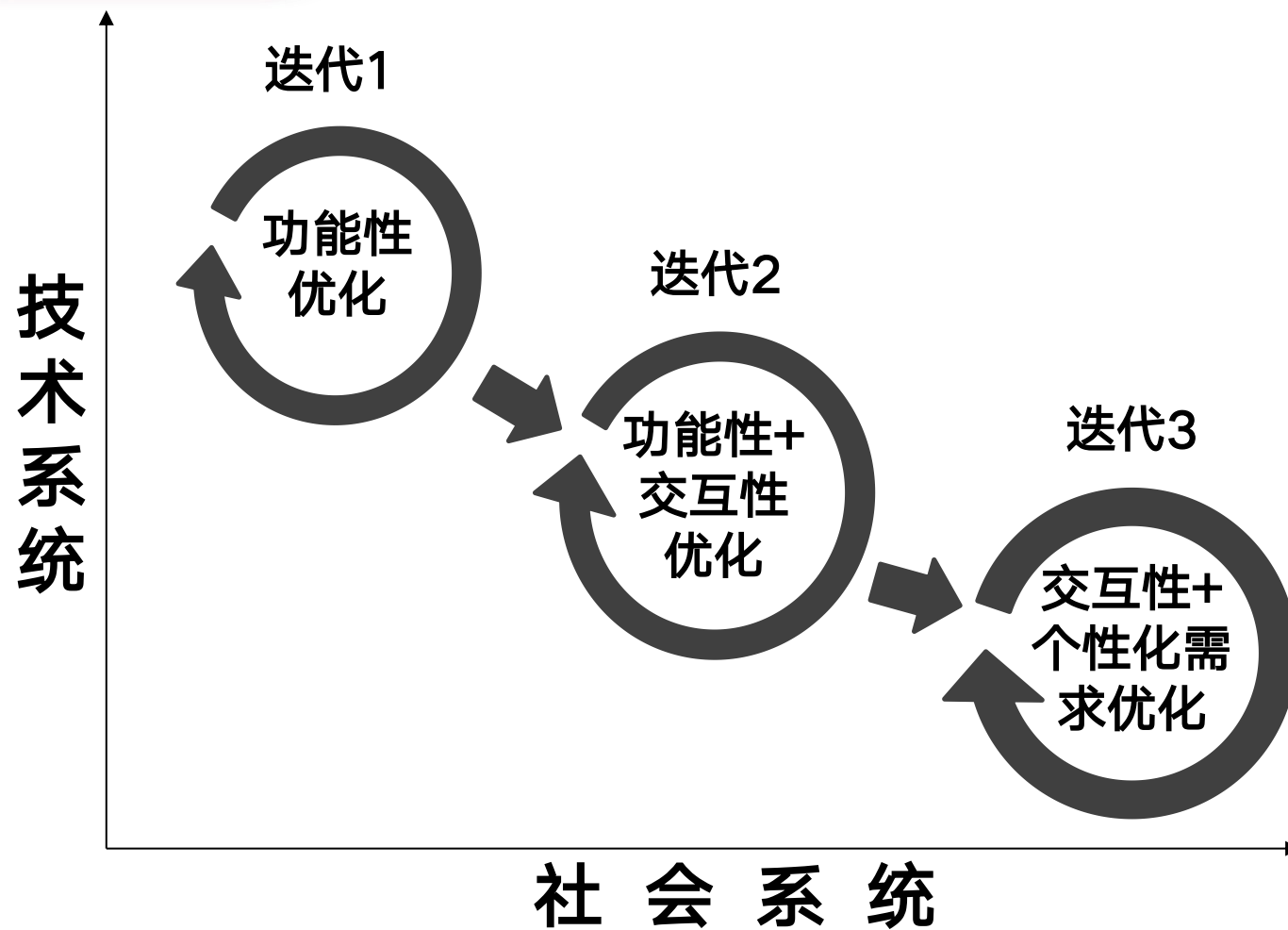
2

让人文数据被用起来

众包抄录平台优化



众包抄录平台优化



众包抄录平台优化

• 实践结果

- 盛宣怀档案抄录平台完成了从功能性优化到交互性优化再到个性化需求优化的迭代发展。
- 三轮测试结束后，平台共有注册用户115人，截至2019年5月底新增注册用户47人，且每天维持一定的访问量与浏览量。

• 理论结果

(1) 技术系统

- 引入游戏化元素，设计积分公式
- 任务细粒度，难度有区分

(2) 社会系统

- 线下-线上式的协作抄录模式
- 基于移动社交媒体的即时交流平台

03

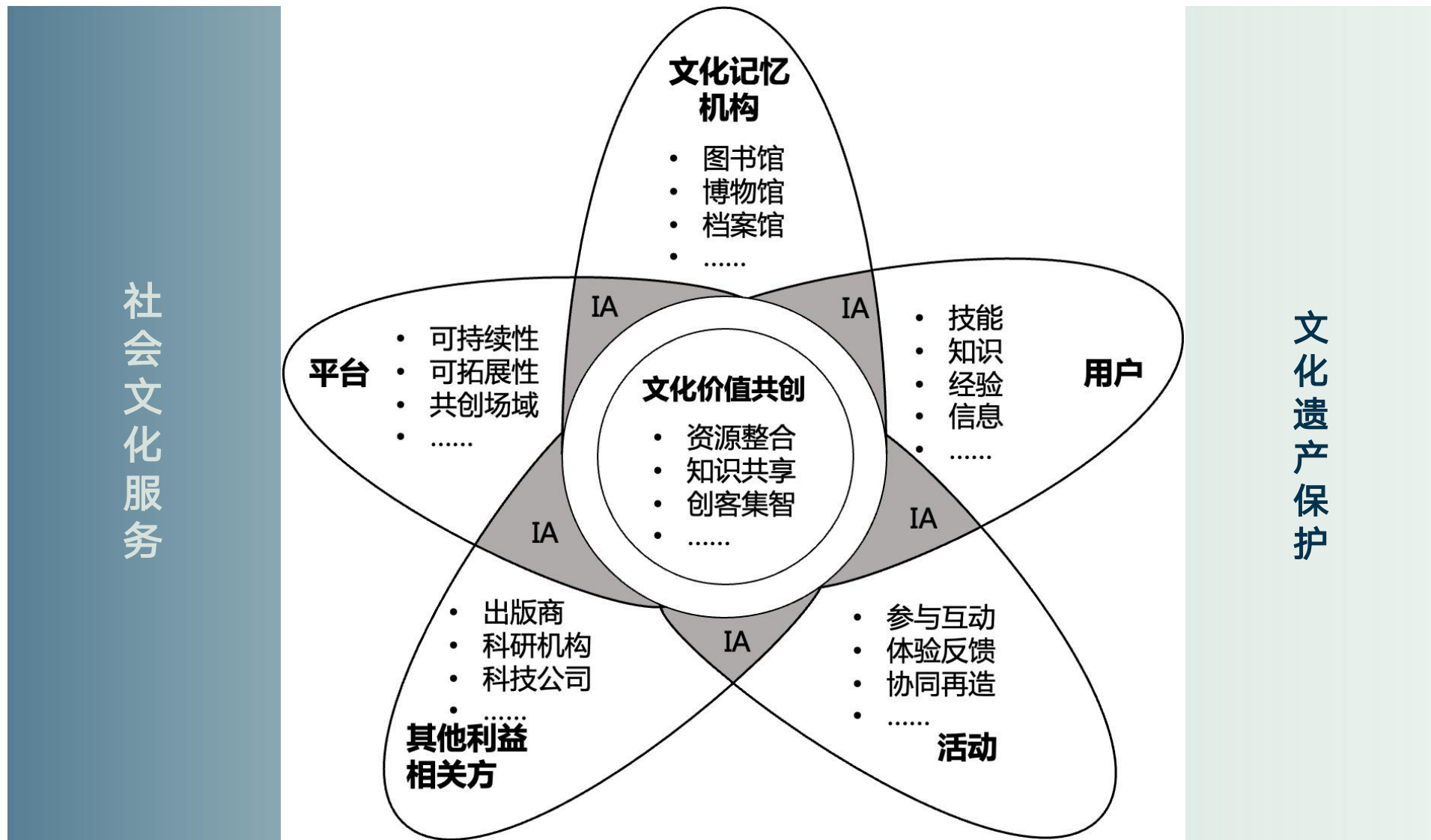


众创：开放数据的价值共创

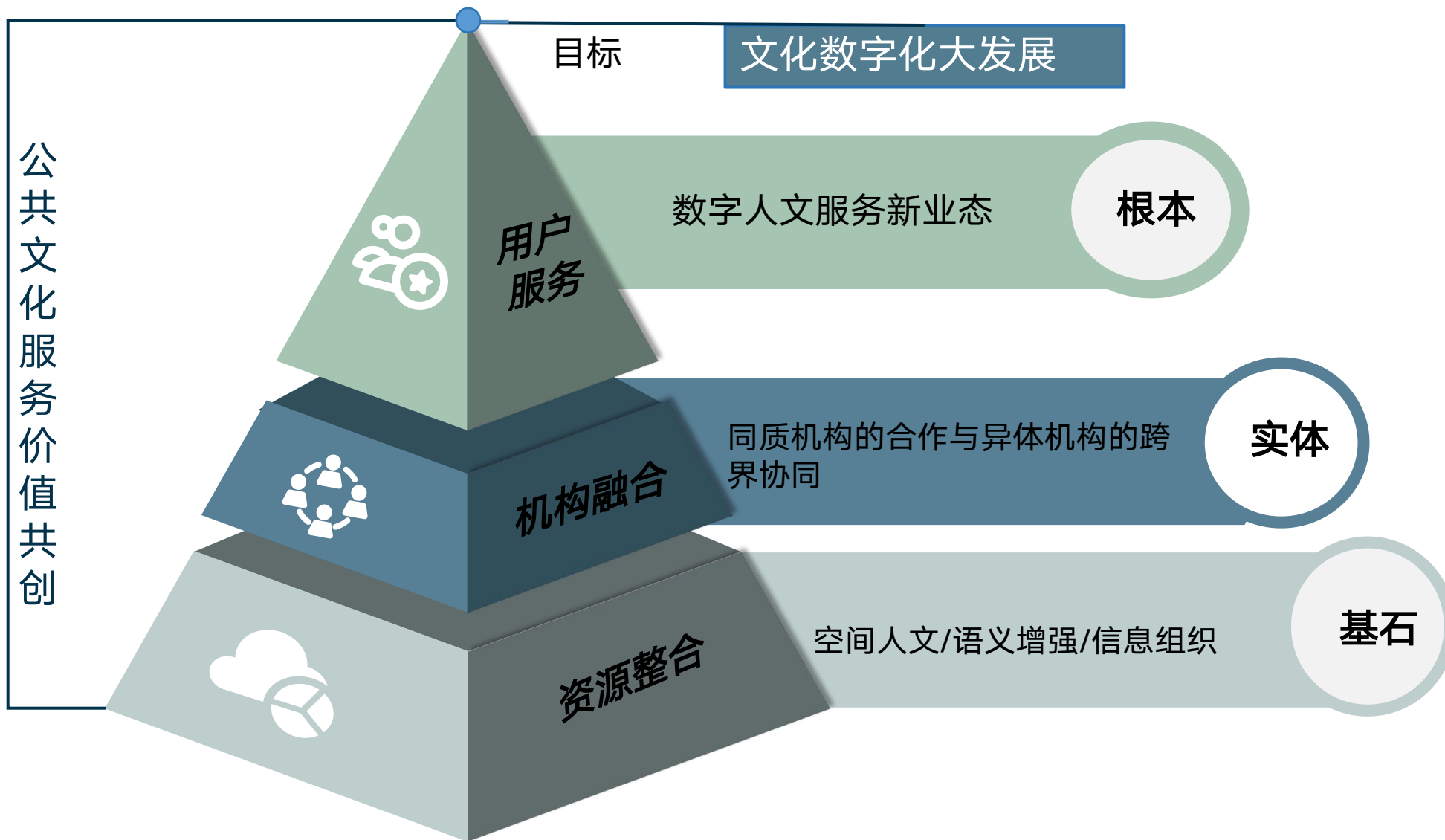
价值共创 / 开放数据竞赛 / 多主体参与

*Value
co-creation*

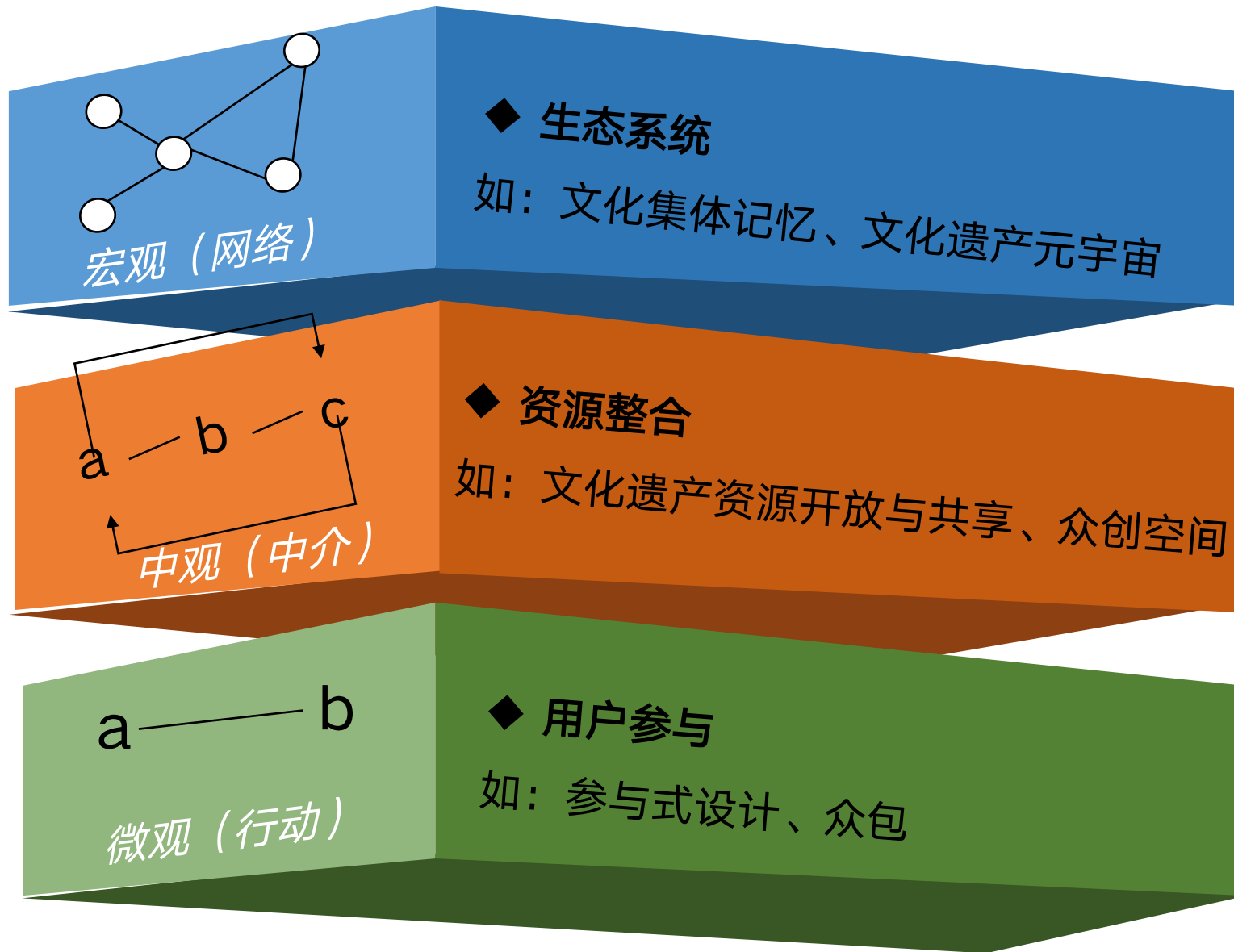
文化价值共创交互模式



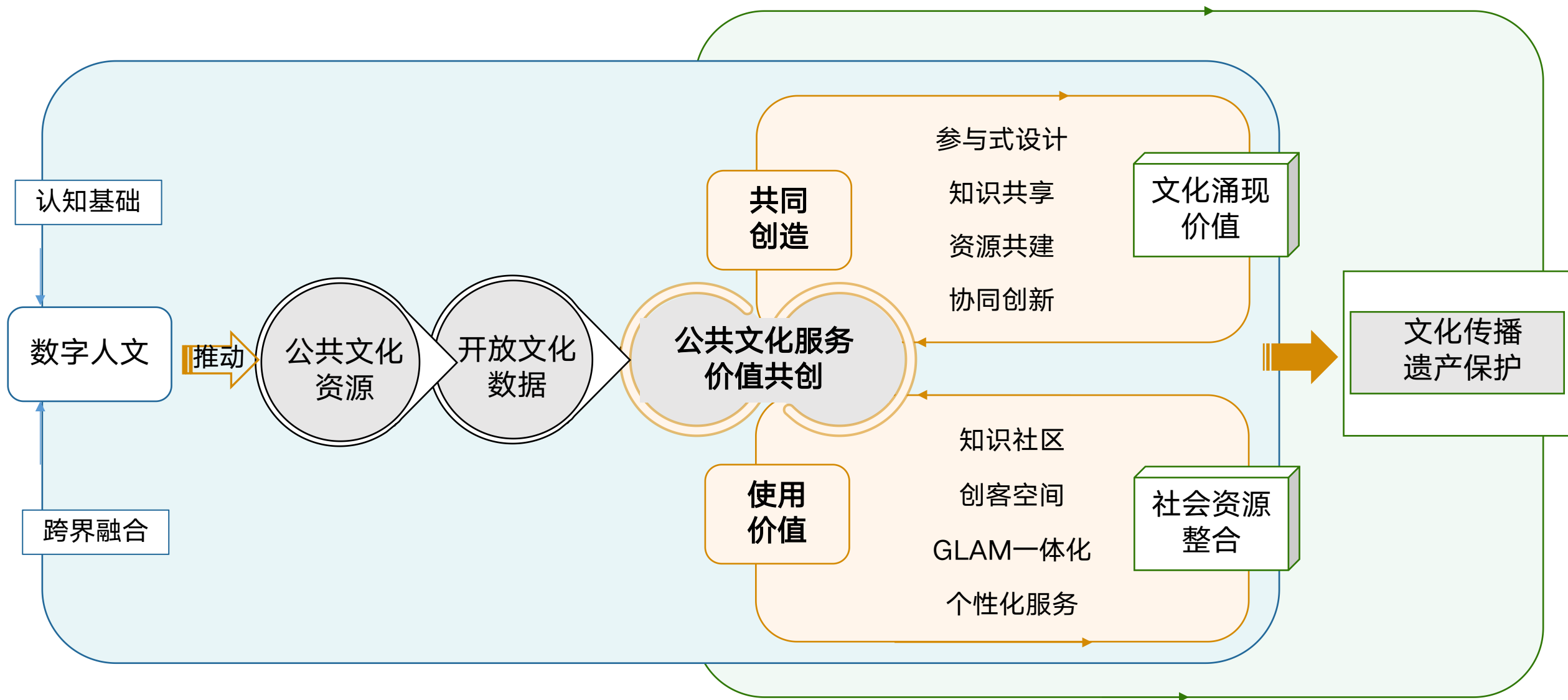
研究主题



基于价值共创的用户服务创新



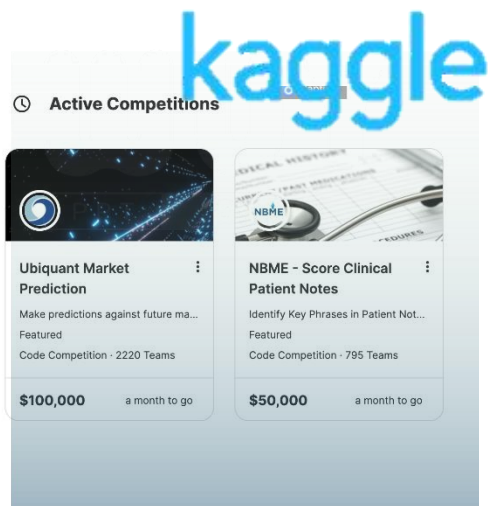
公共文化服务价值共创概念解析



开放数据竞赛



LinkedUp Challenge
系列竞赛



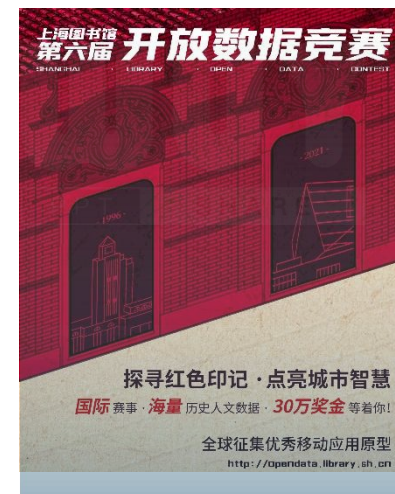
数据科学平台Kaggle
各类开放数据竞赛

挑战类开放数据竞赛

获取优秀的算法或计算模型



德国黑客马拉松
“编码达芬奇”



上海图书馆
开放数据竞赛

创意类开放数据竞赛

技术赋能数据的多样化呈现

研究现状

理论

- 多以单一竞赛要素或主体为切入点，鲜有研究从价值共创的整体视角，探寻各方参与主体特征的协同作用。
- 以竞赛形式本身作为发散点，但不同竞赛的目标和侧重点都存在差异，未能凸显数字人文特色。

实践

虽然国内外已有许多开放数据竞赛的成功应用实例，但囿于价值共创内在机理、框架的理论指导匮乏，致使文化记忆机构的探索成本增加且效果不佳。

🕒🕒 数字人文领域下创意类开放数据竞赛价值共创如何发生？

上海图书馆开放数据竞赛



上海图书馆
开放数据竞赛
SHANGHAI LIBRARY OPEN DATA CONTEST

首页 新闻公告 竞赛规则 开放数据 培训资料 获奖作品 Q&A 往期竞赛

第六届 开放数据竞赛

上海图书馆 · Shanghai Library

线上展览

<https://opendata.library.sh.cn>

- 上海图书馆开放数据竞赛是由上海图书馆主办的国际赛事，致力于历史人文数据的开放、利用以及多元数据的融合。
- 上海图书馆基于开放数据平台为科研人员、创客团队、文化科技企业、人文知识爱好者打造了一个良好的学习交流社区，共同创造历史人文数据的新兴价值。

一

开放数据应用
开发竞赛
2016



二

开放数据应用
开发竞赛
2017



三

开放数据应用
开发竞赛
2018



四

老建筑的故事



五

全媒融合，用
DH解构历史数
据之美



六

探寻红色印
记 · 点亮城市智
慧



行动者协作行为研究——案例简介



- 主题：老建筑的故事
- 目的：弘扬海派文化、江南文化和红色文化
- 推荐创意：阅读建筑、红色记忆、民国影事、故居旅游、名人掌故、民国书刊、寻根之旅、古今诗词、古籍探索
- 开放数据：家谱、古籍、书刊、中国历代人物传记资料库、中国近代报纸资源全库……

行动者协作行为研究——数据收集

竞赛从2019年4月23日开始到9月24日结束，总计154天，期间共涉及5项赛程。



赛程	时间	主体人员	客观日志
参赛报名	4.23-5.20	主办方、参赛者	系统报名日志
宣传	4.24-5.20	主办方、参赛者	宣讲材料、宣讲会日志等
专业培训	5.23	主办方、参赛者、开放数据提供方	培训材料、培训日志等
作品开发与提交	7.15-8.15	主办方、参赛者	作品提交日志
作品评审	8.16-9.24	主办方、评审方	评审日志

针对主体人员：访谈法+观察法

针对客观日志：日志法

行动者协作行为研究——研究发现

依据He对共创价值的定义，采用多人编码对主客观日志记录进行处理，得到数字人文领域创意类开放数据竞赛的价值共创过程涉及的三个价值：期望价值、新兴价值和承认价值。

期望价值是开放数据竞赛行动者在各阶段实施前期望通过协作创造的价值；

新兴价值主要是指在该阶段参与过程中，各行动者通过协作创造出的新价值；

承认价值是指协作后各行动者共同认可的价值。

行动者协作行为研究——研究发现

- 基于价值共创理论的数字人文领域创意类开放数据竞赛行动者协作行为流程分析

将期望、新兴、承认三种价值按照竞赛参与的前、中、后排序，得到数字人文领域创意类开放数据竞赛的价值创造阶段。

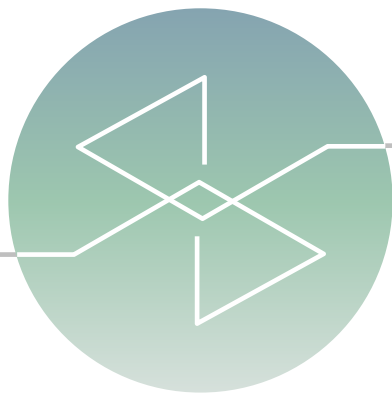
采用Gioia方法（Gioia Methodology），基于价值共创理论，对数字人文领域创意类开放数据竞赛不同价值创造阶段的行动者协作行为进行流程分析。

本研究最终构建包括价值预估、价值声明、价值贡献、价值诱导、价值拓展、价值评估与挖掘、价值认可价值生成路径。

行动者协作行为研究——小结

- 开放数据竞赛这一过程由一系列价值创造阶段组成
- 部分过程属于单实体进行价值创造的过程
- 部分过程属于多实体实现价值共创的过程

“多行动者之间协作实现价值共创”



“竞合关系”

- 各参赛团队作为开放数据竞赛的重要行动者，并非“孤军作战”，而是在整个阶段和其他参赛团队及主办方进行沟通与合作，共享某些技术问题的解决方案并从中受益。
- 行动者间的竞合（co-operative）水平会显著影响竞赛绩效。

价值共创影响因素研究——研究方法与数据收集



基本信息	类别	人数	占比 (%)
性别	男	20	54.1
	女	17	45.9
年龄	20—25岁	9	24.3
	26—30岁	18	48.7
	31—35岁	6	16.2
	35—40岁	4	10.8
受教育程度	本科	10	27.0
	硕士研究生	20	54.1
	博士研究生	7	18.9
职业	学生	15	40.5
	图书馆工作人员	8	21.6
	高校教师	3	8.1
	IT行业从业者	11	29.8
访谈方式	焦点小组访谈法	24	64.9
	深度访谈法	13	35.1

- 研究方法：关键事件技术
- 数据收集方法：焦点小组访谈法（参赛团队）、深度访谈法
- 访谈时间：2021年6月至9月
- 有效访谈对象：共37名受访者，包含开放数据提供方6人，主办方7人，参赛者24人

价值共创影响因素研究——研究发现

- 数字人文领域创意类开放数据竞赛行动者构成要素识别 ——行动者网络理论 (Actor-Network Theory)



- 关键事件主题分布

理论框架	主题	响应	
		频数	百分比 (%)
人类行动者	参与主体特征	146	26.3
	协作形式	127	22.9
非人类行动者	开放数据特征	113	20.4
	协作平台属性	94	16.9
	宏观环境相关因素	75	13.5

影响因素与价值共创机理

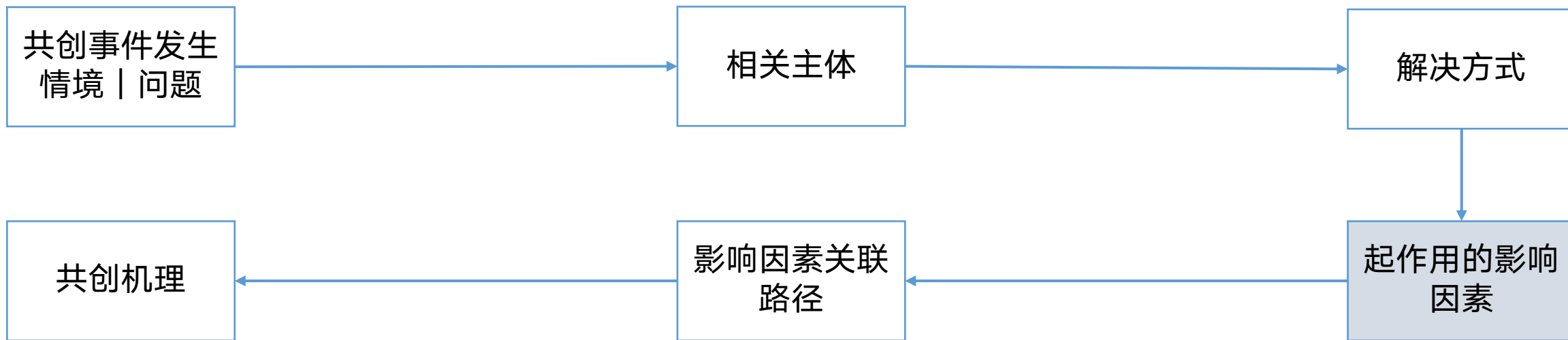
► 数字人文领域下的价值共创概念

多元行动者基于特定情境利用自身技能资源自发参与，通过中介平台进行实时协作，最终整合各类资源，共同创造价值的过程。



①事件的发生基于某一情境下的具体问题。

②共创的实质是各类行动者要素资源之间的碰撞与联系。



全国图书情报专业学位 研究生教育指导委员会

获奖通知

赵宇翔、张妍、刘炜:

为了推动高质量教学案例的开发和案例研究,加强图书情报专业教育和学术研究理论与实践的结合,促进图书情报事业的不断发展,由全国图书情报专业学位研究生教育指导委员会主办的“2022年全国图书情报案例学术研讨会暨图书情报创新案例大赛”已如期举行。

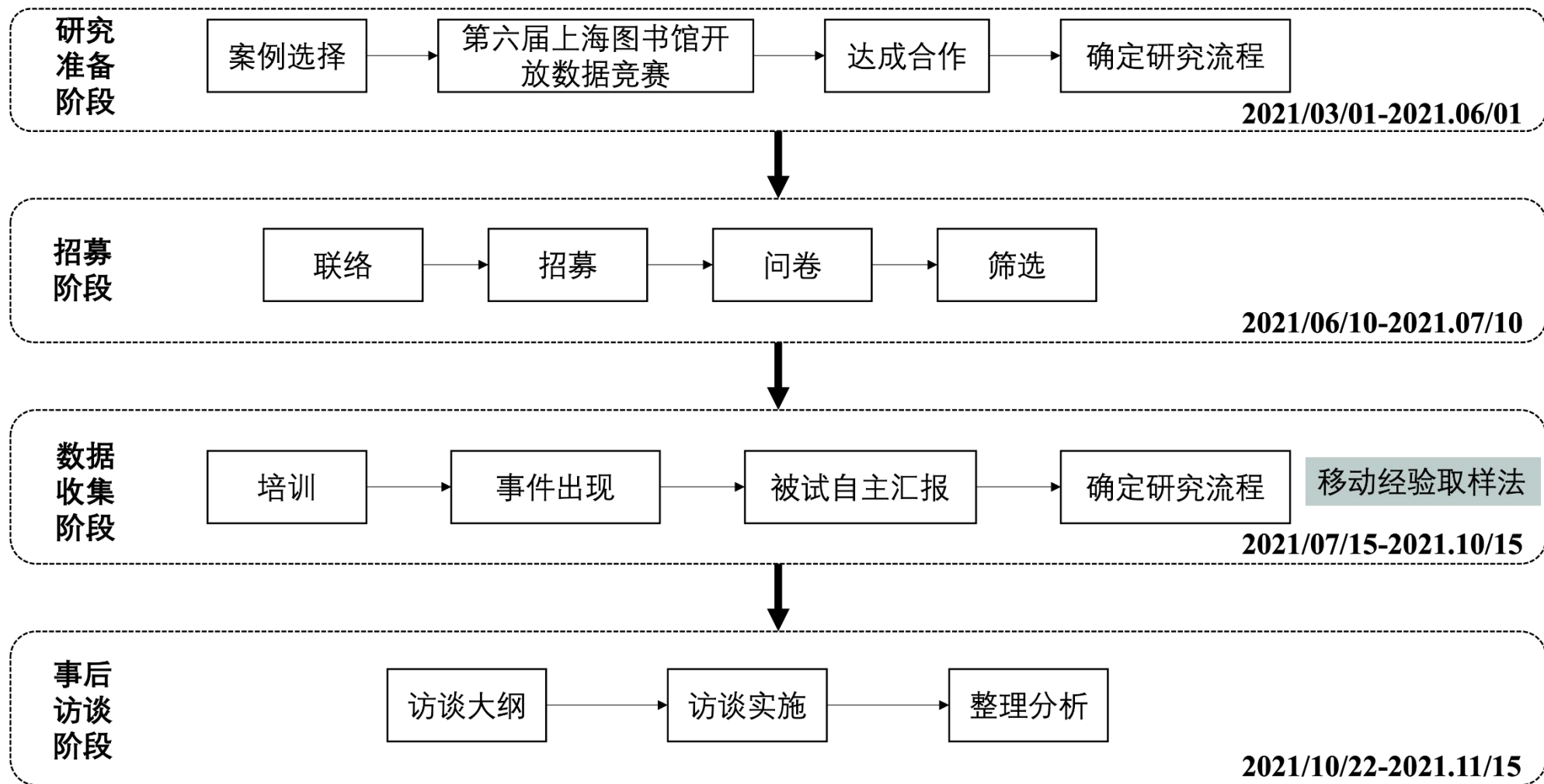
经全国图书情报专业学位研究生教育指导委员会评审,您们申报的从共识到共传:数字人文类开放数据竞赛价值共创的过程模型荣获“优秀案例奖”。

特此通知!

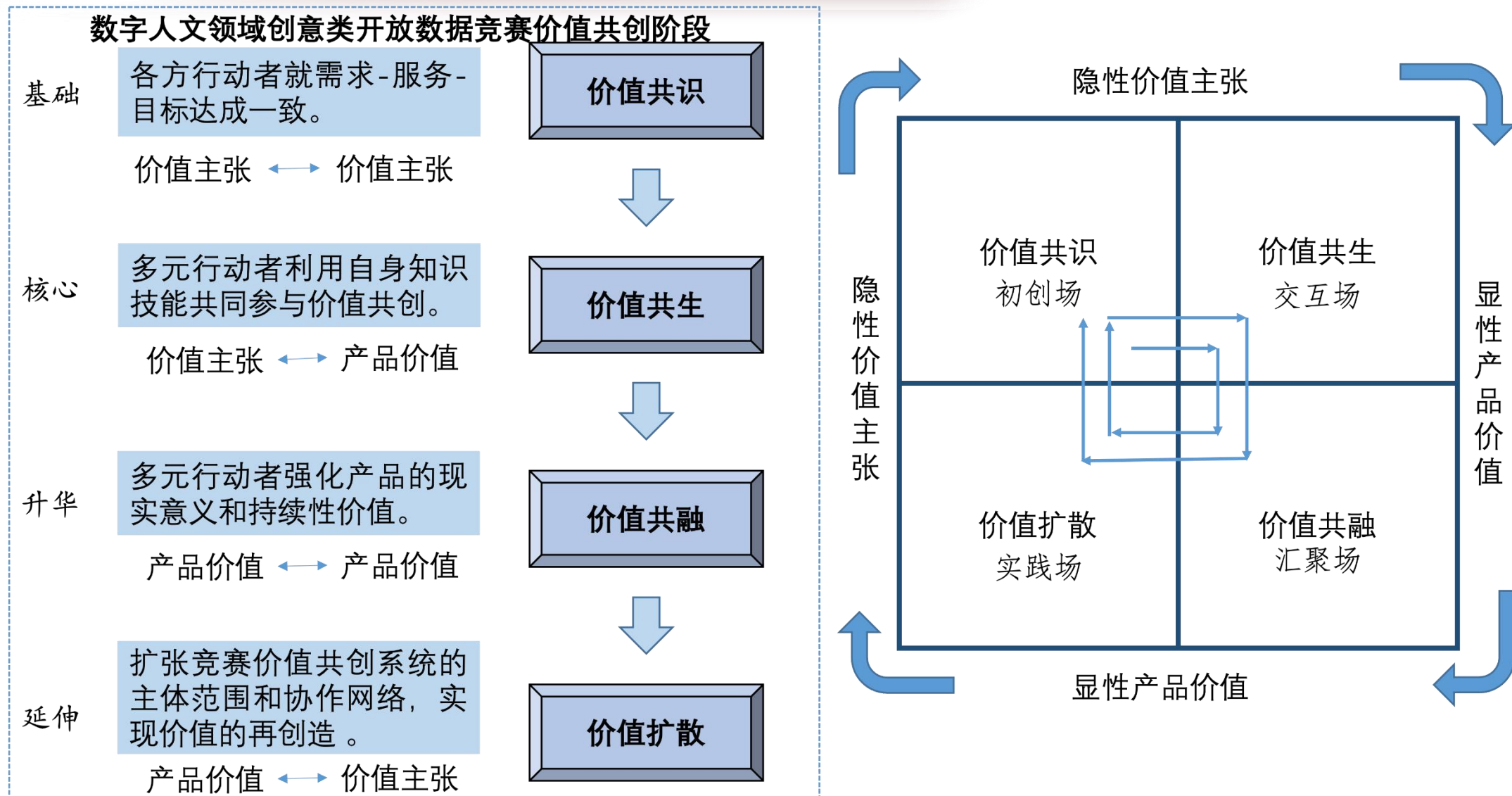
全国图书情报专业学位研究生教育指导委员会秘书处



价值共创机理探索研究——案例研究过程



价值共创机理探索研究——研究发现



多主体参与

开放数据实践中价值 共创的核心主体



04

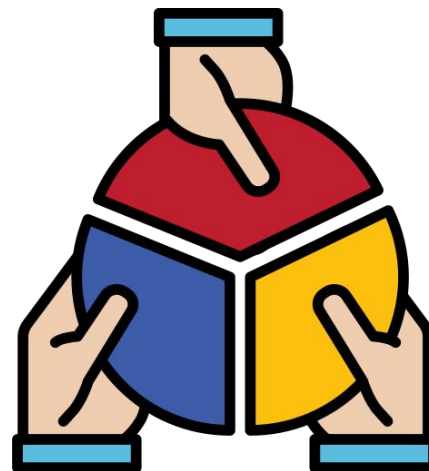


众享：开放数据的社会共享 与再利用

数据可获 / 知识可懂 / 成果可用

*Value
co-creation*

“众享”不只是数据开放



- **数据可获：** 更多人容易找到、下载、使用上图的开放数据
- **知识可懂：** 通过可视化、产品、故事，让复杂的档案/数据变得可理解
- **成果可用：** 竞赛成果有机会接入图书馆服务（如专题网站、展览互动等）；用于教育（课堂使用、学生项目、科普活动）；支撑后续研究和创新

用户搜索开放数据研究

以上图开放数据竞赛为场景，理解并优化用户的开放数据获取和使用

Participants ID	Gender	Age	Occupation	Discipline Background
P1	female	27	Product Manager	Library and Information Science
P2	male	23	Student	Software Engineering, Cultural & Creative Industries
P3	female	21	Student	Archival Science
P4	female	28	Student	Chinese Language and Literature
P5	male	21	Student	Computer Science
P6	female	31	Product Manager	Library and Information Science
P7	female	29	Librarian	Urban Science and Informatics
P8	male	26	Student	Cultural Heritage Informatics
P9	male	35	Researcher	Data Science
P10	male	32	Back-end Developer	Computer Science

Table 1. Basic information of participants



识别了参与者的数据需求、搜索策略、评估标准和相关挑战。研究结果将有助于文化遗产机构有效地组织和管理开放数据集，并倡导数字图书馆更好地整合开放数据资源和构建开放数据平台。

Lian, J., Zhao, Y., Li, X., & Zhu, Q. (2025). An Investigation of Searching Behavior for Open Datasets: Insights from Cultural Heritage Data Curation Competitions. *Proceedings of the Association for Information Science and Technology*, 62(1), 1007-1012.

用户使用场景与需求

表2 用户搜索文化遗产开放数据搜索的典型场景

场景类别	典型用途	搜索目的
学术研究	论文写作、科研项目、数据竞赛	以学术研究或数据分析为目的的搜索行为，反映开放数据在知识生产中的作用
教学与学习	教学展示、课程作业、学生项目	支持教学过程与学习任务，促进学生对文化遗产的理解与应用
专业实践	工作项目、行业咨询、委托研究	为职业任务或实际项目搜寻相关数据，用于规划、评估或设计
创意设计	艺术创作、数字展陈、产品开发	在文化创意、艺术设计中利用开放数据进行再创作与表达
个人探索	兴趣探索、旅行规划、文化体验	出于个人兴趣或休闲目的，对文化遗产资源进行自发搜索与浏览

研究发现，用户搜索文化遗产开放数据时呈现出多元化的应用场景，反映了数据在**学术、教育、创意**和个人领域的广泛应用。



多元化应用场景催生了用户
差异化、场景化的核心需求

根据访谈前问卷调查，用户最常需要的开放数据类型为文本（27%）与图像（23%），其次为视频（17%）、地图（16%）和统计数据（12%），音频数据需求相对较少（5%）。

任务
驱动



学术研究

教学与学习

专业实践



空间分析

量化研究

数据建模



创意设计

数字展陈

直观教学



促进开放数据的获取和使用

许多功能完备的专业数据平台使用率低下，其根源在于“重建设、轻运营”，公众与专业用户对其存在认知空白。这使得大量高质量数据未能进入主流搜索渠道，服务可及性严重不足。

“开放数据竞赛”帮助加强数据库宣传与公众认知
提升文化遗产开放数据服务可及性



作为参赛者生成创意作品



2020年开放数据竞赛参赛作品
“Good Omen (翰墨风华)” 优胜奖

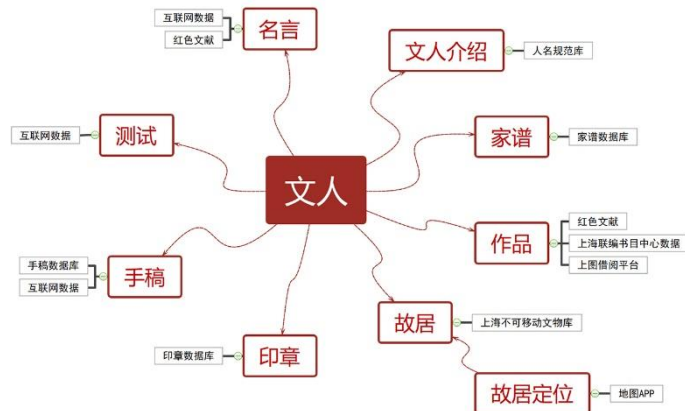


2021年开放数据竞赛参赛作品
“烽火青春” 一等奖

作为参赛者生成创意作品

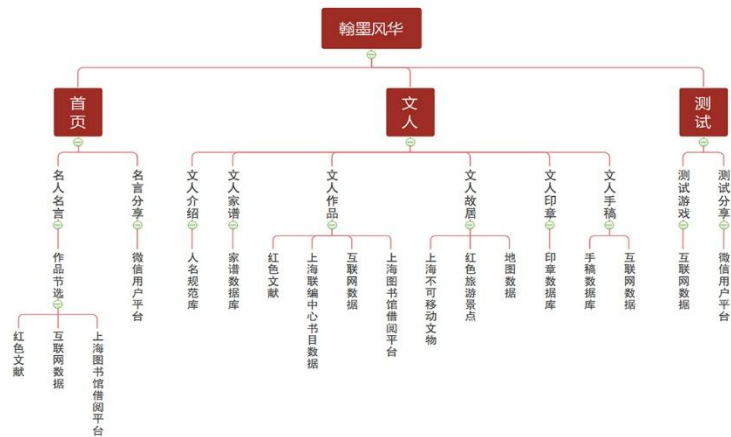
Linked data

关 联 数 据



Product structure

产 品 结 构



Team members

团 队 成 员



领队：赵宇翔
(理论指导)

教授，博士生导师，
研究方向：数字人文、
互联网群体协作、
网络信息资源管理



张轩慧(作品策划)

南京大学信息管理学院博士生
研究方向：数字人文、人机交互



刘野(视觉设计)

阅文集团 创意设计师



张继龙(技术支持)

拼多多 技术开发

Design ideas

设 计 思 路

开放公共文化资源+互联网海量数据
实现多数据资源融合

沉浸式
用户体验

游戏化设计+深度交互
打造沉浸式用户体验

多数据资源
融合

多主体
价值共创

微任务众包模式
+用户生成内容
共建多主体价值共创平台

作为参赛者生成创意作品



Interactive design

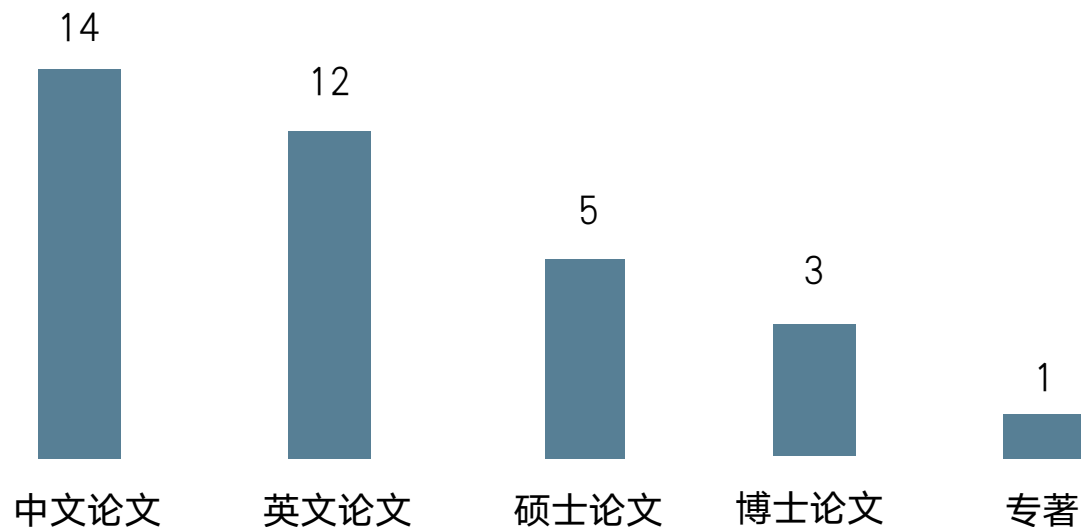
交 — 互 — 设 — 计

团队围绕开放数据竞赛生态的研究成果

- 国家自然科学基金面上项目“公共文化服务领域开放数据的价值共创机制及实现模式研究”（项目编号：72074112）
- 国家自然科学基金面上项目“基于科研众包模式的公众科学项目运作与管理机制研究”（项目编号：71774083）

国家自然科学基金面上项目支撑

成果情况



组织参赛两次：一等奖（陈琼），优胜奖（张轩慧）。

入选国家高层次人才青年项目“文化遗产数字化”

在过去十年的实践和研究中

我们以不同身份参与上图开放数据竞赛：参赛者、合作者、评委、研究者。

这些角色背后，是一个始终在一起学习和尝试的团队。

团队成员：

赵宇翔、张轩慧、刘周颖、练靖雯、张妍、高胜寒、
仝冲、齐云飞、宋小康、张亭亭、韩文婷、宋士
杰……



结语

十年前，开放数据竞赛帮助我们打开数据

十年后，它帮助我们重新想象公共文化服务

众智、众创、众享

是我们看得见的过去，也是可以共同推动的未来。





谢谢！

AI重塑智慧图书馆未来服务新生态主题研讨会
暨上海图书馆开放数据竞赛十周年交流会
上海图书馆东馆 2025年11月27日